2012
2013

FACULTY OF SCIENCES
Master of Statistics: Biostatistics

Masterproef
Determinants of nutritional status among children under age 5 in mainland Tanzania

Promotor :
Prof. dr. Geert MOLENBERGHS

Promotor :
Prof. Y.C. MUZANILA

Pendael Zephania Machafuko
Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2012
2013

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Determinants of nutritional status among children under age 5 in mainland Tanzania

Promotor :
Prof. dr. Geert MOLENBERGHS

Promotor :
Prof. Y.C. MUZANILA

Pendael Zephania Machafuko

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Determinants of the Nutritional Status Among Children Under Age 5 in Mainland Tanzania

Pendael Zephania Machafuko

Supervisors

Prof. dr. Geert MOLENBERGHS

Prof. dr. Y.C. MUZANILA

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Statistics: Biostatistics

12 February 2013

Dedication

This thesis is dedicated to my beloved daughter Abigael P. Z. Machafuko. I shall also, in a very special way dedicate this piece of work to my wife Hilda E. Chigudulu and my parents Mrs. Machafuko and Mrs. Mattao for their unflinching moral support.

Acknowledgements

I would like to express my profound gratitude to my supervisors; Prof. dr. Geert MOLENBERGHS and Prof. dr. Y.C. MUZANILA who continually guided me through this work, by supplying thoughtful suggestions and reading materials for improving this report. In particular, I thank Stephano Cosmas from NBS for providing me with data and all necessary study materials that formed a back born of this thesis.

I'm deeply grateful to the Flemish Interuniversity Council (VLIR) for the scholarship which has enabled me to pursue this valuable Masters program. Not forgetting Martine Machiels who was not only a program manager but also a mother.

I acknowledge my instructors at Center for Statistics for imparting their statistical knowledge on me, especially Prof.dr. Geert Molenberghs who was not only my instructor but also a mentor in sampling techniques for Survey data which formed a back bone of the analysis in this report.

Not forgetting Abdalah Mtumwa: Josephine Shabani, Leacky Muchene, Helio Langa, Adelino Juga, Jose Parra and Ashiraf Mawanda.

Most of all, I thank God for giving me health, wisdom, strength and life to successfully complete this program.

Pendael Zephania Machafuko

University of Hasselt, Belgium, February, 2013

List of Abbreviations

ALR – Alternating Logistic Regression

FAO – Food and Agriculture Organization

GEE – Generalized Estimating Equations

GLMM – Generalized Linear Mixed Model

MAR – Missing at Random

MCAR – Missing Completely at Random

MDG – Millenium Development Goals

MoHW – Ministry of Health and Social Welfare

MI – Multiple Imputation

MNRA –Missing not at Random

NBS – National Bureau of Statistics

OR – Odds Ratio

POM – Proportional Odds Model

TDHS –Tanzania Demographic and Health Survey

TFNC – Tanzania Food and Nutrition Centre

VAD– Vitamini A Deficiency

WHO – World Health Organization

QIC – Quasi under Independence model Criterion

Abstract

Although the problem of malnutrition affects the entire population, children are more vulnerable because it prevents their physical growth, lower their intellectual quotient, cause deficiency of social skills and susceptibility of diseases. The importance of reducing malnutrition has been acknowledged in the MDGs as well and constitutes one of the prime targets of development processes globally. From this perspective, this project aimed to explore the determinants of nutritional status among under-five children in Mainland Tanzania.

A nationwide cross-sectional survey was conducted to record height and weight measurements from 6389 eligible children. These measurements were used to derive the indicators of nutritional status such as wasting and stunting using Z-scores based on WHO child growth standards. Two indicators were treated separately as responses due their biological differences. Alternating Logistic Regression (ALR), Generalized Estimating Equations (GEEs), a Proportional Odds Model (POM), and a baseline logit model were used for analysis. To account for clustering and unequal selection probability, multi-stage sampling and sampling weights were considered in analysis.

Results showed that children with at least 2.5kg birth weight, more than 24 months of age, male, and from Western and Southern highland zones are associated with low probabilities of wasting, whilst maternal health is associated with higher probability of wasting. Similar results were observed for stunting with birth weight, age, gender and maternal health. In addition, children with at least 2.5kg birth weight, more than 24 months of age and male are associated with decreasing stunting risk while maternal health is associated with increasing stunting risk.

In conclusion results suggest that birth weight, maternal health, age, gender, zones, and vitamin A are associated with wasting while birth weight, age, maternal health and gender are associated with stunting among children under-five in Mainland Tanzania.

Key Words: Alternating Logistic Regression (ALR), Generalized Estimating Equations(GEEs), Proportional Odds Model (POM), Baseline logit, Stunting, Wasting.

Contents

1 Introduction

1.1 Background

Reducing malnutrition among children under the age of five remains a huge challenge in developing countries of the world. An estimated 230 million under-five children are believed to be chronically malnourished in developing countries, Van de Poel *et al* [31]. Similarly, about 54 percent of deaths among children of this age group are believed to be associated with malnutrition in developing countries. In Sub-Saharan Africa, 41 percent of under-five children are malnourished and deaths from malnutrition are increasing on daily basis in the region, FAO [12]. Malnutrition is widespread in Tanzania, especially in the rural areas due to inadequate food and nutrition supply. The 2010 Tanzania demographic and health survey revealed that 45 percent of the children in rural areas and 32 percent in urban areas are stunted, and 18 percent for rural and 12 percent for urban areas are severely stunted, NBS [21].

Health and development are intimately interconnected. Meeting primary health care needs and nutritional requirements of children are fundamental to the achievement of sustainable development. The UN world summit, 2000 set a target to half the prevalence of underweight among children younger than 5 years between 1990 and 2015 through MDG number one. To address this problem, Tanzania government through the Ministry of Health adopted a national food and nutrition policy in 1992 and revised 2010. The policy provides guidance and coordination for food and nutrition programmes, specifically those related to food insecurity and micronutrient deficiencies [19]. The focus is on the prevention of nutrition-related diseases and conditions through the provision of adequate, nutritionally sound foods. However, for better planning and implementation of such programs continued efforts are required to identify the key determinants associated with malnutrition among children under-five.

Although these determinants have been highlighted by different stakeholders at international and national level, they are less explored across regions, at household and individual levels in Mainland Tanzania. This research project critically explores the determinants of the nutritional status for children under age 5 in Mainland Tanzania. Identification of these determinants can be used to guide strategic planning, address strength and weakness of current policy and calibrate estimates obtained from other sources for addressing this problem in Mainland Tanzania. Cross-sectional survey can serve as a tool to collect a subset of possible risk factors associated with the response variable of interest. The association established enables the researcher to determine evidence based in policy planning and resource allocations and in evaluating progress in policy implementations. This can be achieved by applying proper statistical techniques to quantify the impact of determinants, interventions coverage and other possible indicators are essential in the success of the policy.

1.2 Objective

The objective of this project is to study the risk factors associated with stunting and wasting among children of under-five in Mainland Tanzania.

2 Study Implementation

The 2010 Tanzania Demographic and Health Survey (TDHS) is the eighth in a series of national sample surveys conducted in Tanzania to measure levels, patterns, and trends in demographic and health indicators. The survey was implemented by the National Bureau of Statistics (NBS) and the Office of the Chief Government Statistician - Zanzibar (OCGS). The survey was funded by the Tanzania government through the MoHSW, Tanzania Food and Nutrition Centre (TFNC), Department for International Development (DFID), World Health Organization (WHO)/Zanzibar, United Nations Fund for Population Activities (UNFPA), United Nations Children's Fund (UNICEF), World Food Programme (WFP), United Nations Development Programme (UNDP), and Irish Aid. ICF Macro provided technical assistance for the survey through the MEASURE DHS programme. The United States Agency for International Development (USAID) funded the technical assistance.

2.1 Survey Design

To assess the risk factors related to wasting and stunting prevalence, multi-stage and cluster sampling designs were used. This was done for the reasons of costs, logistics, facilitating field work, as well as sampling with certainty for a subgroup of units. The survey adapted household, women, and men questionnaires to reflect relevant issues related to population health in Tanzania. The main purpose of the household questionnaire was to identify women and men who were eligible for the individual interview, some basic household characteristics, and to record height and weight measurements for under-five children to evaluate nutritional status. Weight measurements were obtained using lightweight, SECA mother-infant scales with a digital screen, designed and manufactured under the guidance of UNICEF and height measurements were carried out using a measuring board. On the other hand women's questionnaire was used to collect from all women age 15-49 the background characteristics, infant feeding practices, fertility preferences, episodes of childhood illness. The sample excluded nomadic and institutional populations, such as persons staying in hotels, barracks, and prisons.

2.1.1 Multi-Stage Sampling

A direct selection of households or respondents from a list which sometimes is unknown would be too expensive because of the spread and therefore the interviewers travelling costs would be too high. Therefore, multi-stage sampling was introduced to facilitate the field work by providing the best way to access households and respondents. Within the region, clusters were selected from a list of enumeration areas demarcated for the 2002 Population and Housing Census. In the first stage, a total of 385 clusters (primary sampling units) were selected in 2010 TDHS. Twenty-five clusters were selected in Dar-es-Salaam, and 18 were selected in each of the other twenty regions in Mainland Tanzania. In the second stage, within each primary sampling unit, sample of households called secondary sampling units were then systematically selected to participate in the survey. Twenty-two households were selected from each of the primary sampling units in all regions except for Dar-es-Salaam where sixteen households were selected. A representative probability sample of 8320 households was selected. Due to the fact that interviewing more people is inefficient because members of the same family tend to resemble each other more closely than members from different households, Bruckers Liesbeth *et al* [7], only women of reproductive age (15-49) in a household and men in a subsample of one-third of all households selected who were either permanent residents or visitors present in the household on the night before survey were eligible to be interviewed.

2.1.2 Sampling Weights

In multi-stage sampling and clustering, it is rarely the case that sampled units (household members and households) have equal selection probability and that all desired information is obtained in a survey. Therefore, sampling weights are designed to adjust for differences in selection probability and corrections for differential response rates in a sample, either due to design or happenstance. In the DHS surveys, many times the sample is selected with unequal probability to expand the number of cases available for certain areas or subgroups to reduce sample variability. In DHS surveys, there are household weights and individual weights. The household weight for a particular household is the inverse of its household selection probability multiplied by the inverse of the household response rate of its household response rate group. The individual weight of a respondents case is the household weight multiplied by the inverse of the individual response rate of her individual response rate group.

2.2 Children Nutritional Status

Nutritional status is primarily determined by a child's growth in height and weight and is directly influenced by food intake and the occurrence of infections. Food intake is not only a result of food availability at the household level but also of dietary quality, quantity and feeding practices. Optimal infant feeding practices, which include breastfeeding and timely complementary feeding, contribute to the level of food intake in infants and young children, Brown *et al* [4]. Health and physical consequences of prolonged states of malnourishment among children are delay in their physical growth and motor development, lower intellectual quotient (IQ), greater behavioural problems, deficient social skills and susceptibility to contracting diseases, Black *et al* [6].

Nutritional status is evaluated using three primary indicators; weight for age, height for age, and weight for height. In turn poor nutritional status is determined by low weight for age (underweight), low heght for age (stunting), and low weight for height (wasting). However, stunting and wasting represent two different biological processes. Stunting or growth retardation or chronic protein-energy malnutrition is deficiency for calories and protein available to the body tissues and it is due to inadequate intake of food over a long period of time, or persistent and recurrent ill-health, WHO [27]. Wasting or acute protein-energy malnutrition reflect the failure to receive adequate nutrition during period immediately before the survey, resulting from recent episodes of illness, and diarrhoea in particular or from acute food shortage. Low weight for age can reflect both wasting and stunting and is thus not able to distinguish between long-term malnutrition and temporary undernourishment.

2.3 Data Description

The research project utilized the nationwide data of TDHS 2010, where completed and plausible anthropometric data were available for 6389 children under-five of the

interviewed mothers. Stunting and Wasting were calculated based on the growth standards published by WHO in 2006, and are expressed in standard deviation units. A child nutritional status was categorized into two groups for both indicators-stunted or wasted ($< -2Z-$ score) and non-stunted or non-wasted ($\geq -2Z-$ score). The indicators were further categorized into multicategory to assess the severity of the malnutrition status-severely stunted or wasted ($< -3Z-$ score), moderately stunted or wasted ($-3 \quad to-2Z-$ score) and non-stunted or wasted ($\geq -2Z-$ score). Several socio-economic and demographic characteristics were recorded as well as maternal health and nutritional information as predictor variables. The variables are age of child ($<$ 24month=breastfeeding, $\geq$ 24month=weaning), gender classified as 1 as males and 2 as females, duration of breastfeeding ($\leq$ 6month=1, $>$ 6month=2), highest educational level (1=No education, 2=Primary , 3=secondary, 4=higher),
birth weight($< 2.5kg$=underweight,$\geq 2.5kg$=normal), maternal health
($<$ 18.5 BMI=underweight, 18.5 to 24.5 BMI=normal, $\geq$ 25 BMI= overweight), vitamin A for past six months (0=No, 1=Yes, 8=Dont know, 9=others) and Zones
(1=Western, 2=Northern, 3=Central, 4=S. highland, 5=Lake, 6=Eastern, 7=Southern). Other variables such as cluster identification number (CLID) and household identification number (HhID) were considered to account for hierarchy of survey data in analysis

2.4 Response Variable

Binary and multicategory outcomes are very common in biomedical studies, for instance in the evaluation of nutritional status among children of under-five explained above (section 2.3). Based on these classifications, it is possible to employ plausible statistical tools for estimating the magnitude of the association between the response variable of interest as a function of predictor variables. Unlike continuous outcome variables, binary or multi-categorical outcome variables are often prone to loss of information. However, one very practical advantage of using statistical methods for binary or multi-categorical response over statistical methods for continuous response variable in epidemiologic research is that parameter estimates of the possible risk factors can be directly converted to an odds ratio and are easy to interpret.

$$\text{Stunted} = \begin{cases} 1 \text{ if low height for age } Z\text{-scores} < -2, \\ 0 \text{ if normal height for age } Z\text{-scores} \geq -2, \end{cases}$$

$$\text{Wasted} = \begin{cases} 1 \text{ if low weight for height } Z\text{-scores} < -2, \\ 0 \text{ if normal weight for height } Z\text{-scores} \geq -2, \end{cases}$$

$$\text{Stunted} = \begin{cases} 1 \text{ if normal height for age} \geq -2 \ Z\text{-score}, \\ 2 \text{ if moderate height for age} < -2 \text{ and} \geq -3 \ Z\text{-scores}, \\ 3 \text{ if very low height for age} < -3 \ Z\text{-score}, \end{cases}$$

$$\text{Wasted} = \begin{cases} 1 \text{ if normal weight for height} \geq -2 \ Z\text{-score}, \\ 2 \text{ if moderate weight for height} < -2 \text{ and} \geq -3 \ Z\text{-scores}, \\ 3 \text{ if very low weight for height} < -3 \ Z\text{-score}. \end{cases}$$

3 Literature Review

Nutritional status is the state of nutrition of an individual. It is dictated by the quality of nutrients consumed and the body's ability to utilize them for metabolic needs. An individual is said to have a good nutritional status if he or she shows no evidence of malnutrition, Laditan [13]. The essence of nutritional assessment is to identify nutritional disorders and determine which individuals need nutritional instruction and or nutritional support. Different age groups can be affected by nutritional disorders, especially children below 5 years old are most vulnerable to inadequate food intake due to increased physical growth.

Thus, being nutritionally vulnerable, under-5 children's nutritional status is generally accepted as an indicator of the nutritional status of any particular community, Davidson [9]. This is due to their easy susceptibility to malnutrition and infection [2]. Children in this age group require a high supply of nutrients since they are usually very active and their growth is rapid. Also during this period, under-nutrition in the form of kwashiorkor, marasmus, and anaemia are common, Ene-Obong [11]. Malnutrition is manifested in various forms such as stunting (short stature), underweight, muscle wasting, growth retardation, diminished subcutaneous fat and ill health with high mortality rate, Onimawo and Amangbangwu [22]. Thus, demographic and health related determinants of malnutrition for under-five have been narrated by several authors as follows;

From birth to 4-6 months of life, breast milk is the sole or prime source of nutrients and optimal breastfeeding practice becomes a critical factor in child survival and development, Onyesili [23]. Breast milk contains all nutrients, antibodies, hormones and antioxidants that an infant needs to thrive, UNICEF [30]. Early initiation within half an hour of birth will ensure that the protective antibodies in the colostrum are available rapidly to the infant, because after 24 to 48 hours, the level of antibodies in breast milk diminishes.

For social and biological reasons, women of the reproductive age are amongst the most vulnerable to malnutrition. A woman's nutritional status has important implications for her health as well as for the health of her children. It reduces productivity, increased susceptibility to infections, slowed recovery from illness, heightened risk of adverse pregnancy outcome. For instance, a woman who has a poor nutritional status as indicated by a low body mass index (BMI), short stature, micronutrient deficiencies has a greater risk of obstructed labour, a baby with low birth weight, low-quality breast milk, death from postpartum haemorrhage, and adverse effects upon both herself and her baby [21] . A child's birth weight is an important indicator of a child's vulnerability to the risk of childhood illness and the chances of survival [3]. Low birth weight (< 2.5kg), irrespective of gestational age has been associated with higher probabilities of infection, malnutrition and handicapped conditions during childhood, mental deficiencies and problems related to behaviour and learning during childhood, WHO [28]. Children who survive with low birth weight have higher incidence of diseases, retardation in cognitive development and undernourishment.

The role of parental education in determining children's health and nutritional status is two-fold. First, better education should translate into higher incomes, that is, better educated parents are likely be able to make better use of available information about child nutrition and health, partly as being educated themselves may increase their preference for child quality over quantity. Second, sufficient income lead to access of adequate food supplies, use of health services, availability of improved water sources, and sanitation facilities, which are prime determinants of child and maternal nutritional status, UNICEF [29]. Most likely, successful completion of primary schooling or functional literacy is sufficient in this context, and post-primary school education might only add limited benefits.

Vitamin A is an essential micronutrient for the immune system that plays an important role in maintaining the epithelial tissue in the body. Vitamin A deficiency(VAD) can also increase the severity of infections, such as measles and diarrhoeal diseases in children, and slow recovery from illness. Vitamin A is found in breast milk, other milks, liver, eggs, fish, butter, red palm oil, mangoes, papayas, carrots, pumpkins, and dark green leafy vegetables. The liver can store an adequate amount of the vitamin for four to six months. Periodic dosing (usually every six months) of vitamin A supplements is one method of ensuring that children at risk do not develop VAD, NBS [21].

The period from birth to age 2 is especially important for optimal growth. Unfortunately, this period is often marked by micronutrient deficiencies and is associated with childhood illnesses such as diarrhoea and acute respiratory infections that interfere with optimal growth and development [21].

4 Statistical Methodology

4.1 Exploratory Data Analysis (EDA)

The exploratory data analysis was performed using simple descriptive statistics as well as graphical techniques to explore the association between the response variable and covariates of interest. This was intended to provide an insight about the data structure and the most plausible aspects or implications to be considered during statistical analysis.

4.2 Missing Data

In many sample surveys, some of the units contacted do not respond to at least some items being asked, Little and Rubin [15]. Respondent's participation in most surveys is a voluntary decision, hence survey data collection activity may require special consent of the subject. The researcher has no influence to withhold a respondent who wants to withdraw from the course of the interview, mostly this happens when a respondent does not wish to answer sensitive or difficult questions. Such nonresponse called survey nonresponse arises from sample survey, is common in practice whenever the population consists of units such as individual people, households, or business, Rubin [25]. The problem created by survey nonresponse is that data value intended by survey design to be observed are in fact missing. These missing values lead to inefficient estimates because of the reduced size of the data and also standard complete data methods cannot be immediately used to analyse the data. Moreover, possible biases exists because the respondents are often systematically different from the nonrespondents; of particular concern, these biases are difficult to eliminate since the precise reasons for nonresponse are usually not known, Rubin [25]. In the TDHS for example, about 4% of eligible children were not measured, 1% had invalid values for height and weight and 1% had incomplete information on age. In order to incorporate incompleteness into the modelling process there is need to reflect on the nature of the missing value mechanism and its implications for the statistical inference, Molenberghs and Verbeke [20]. Little and Rubin [15], make important distinctions between different missing values process. A non-response process is said to be missing completely at random (MCAR) if the probability that a respondent did not report an item value is completely independent of the true underlying values of all of the observed and unobserved variables. Data are missing at random (MAR) if, conditional on the observed data, the missingness is independent of the unobserved values of the variables in the survey. Finally, data are not missing at random (MNAR) if missingness is neither MCAR nor MAR. For categorical outcomes the inferences with the GEEs and ALR approach are valid only under the strong assumption that the data are missing completely at random (MCAR), Molenberghs and Verbeke [20]. To relax the MCAR assumption and allow for a more flexible assumption by assuming that data is missing at random (MAR), one can use Multiple Imputation, Little and Rubin [15]. With this method, each missing value is replaced by two or more imputed values in order to represent the uncertainty about which value to impute, whose mean and variance can be determined to estimate the efficiency of the imputation procedure. When there is a combination of missing covariates and missing outcomes, as was the case for TDHS, multiple imputation can be a useful tool to deal with such a case. In this report, multiple imputation was used to account for missingness.

4.3 Marginal Models

In a survey, the units of analysis are organized as a hierarchy of successively higher level units forming clusters or multi-stage sample design. The hierarchical nature of survey data violates the assumption of independence, thereby requiring sophisticated approaches to account for the association between observations in the statistical modelling. For instance, in the TDHS the nutritional status was determined once for each eligible child (unit of analysis) in the selected household, and the units of analysis are grouped into, or nested within households (cluster of units), which are in turn nested within clusters (cluster of cluster) which were selected from each region. Due to the fact that children from the same household share behaviours, genetic traits, and environmental conditions are likely to be similar than unrelated children. Knowledge of nutritional deficiency clustering is important because it may provide insights into the etiology of the nutritional deficiency and risk factors operating within different levels of clusters. With respect to this study, there are two sources of clustering need to be considered in statistical modelling, there is households within a cluster and individual respondents within households. Different statistical methods have been proposed in the statistical literature such as Generalized Linear Mixed Models (GLMMs), Generalized Estimating Equations (GEEs), and Alternating Logistic Regression (ALR) which explicitly model the various levels in the data structure. It should be noted that, failure to accomodate the survey design might declare effects significant that, in fact are not. The fact that clusters' populations are not equal and members within a household have unequal selection probability, sampling weights were introduced in statistical modelling to account for the additional variability that may arise due to differences in selection probability. Weights tend to reduce precision of the estimates, this can be reflected in larger standard errors.

4.3.1 Generalized Estimating Equations (GEEs)

For binary clustered responses, the GEE approach take into account the correlation between responses of interest for subjects from the same cluster [10] by adopting working assumptions. GEE is a non-likelihood method that uses correlation to capture the association within the clusters or subjects in terms of marginal correlations [20]. GEE methodology for clustered or repeated data proposed by Liang and Zeger [34] requires only the correct specification of the univariate marginal distributions provided one is willing to adopt working assumptions about the correlations structure. The working assumptions as proposed by Liang and Zeger [34] include independence, unstructured, exchangeable, and first order auto-regressive AR(1). A detailed discussion of these assumptions can be found in Molenberghs and Verbeke [20]. When modelling the dependence structure is of scientific interest as well second order generalized estimating equations (GEE2) that use the correlation coefficient, Prentice [24] while Lipsitz, Laird, and Harrington [16] switched from marginal correlation to marginal odds ratio, as a measure of pairwise association. While GEE2 is nearly fully efficient as compared to a full likelihood approach, bias might occur in the estimation of the main effect parameters when association structure is incorrectly specified. The simultaneous estimation of marginal mean and dependence structure can become computationally prohibitive, especially when the cluster sizes become large. Alternating logistic regressions (ALR) overcome this limitation and can be used to model, simultaneously the marginal probability and pairwise association for large clusters as proposed by Carey, Zeger, and Diggle [8]. This method is limited to binary outcomes. Given multinomial response with K categories a vector of $K-1$ indicators of which the k^{th} level is 1 when the observation falls in category k and 0 otherwise, Agresti [1]. The usual approach for modelling such type of response data is to use logits of cumulative probabilities or proportional odds model given by:

$$\text{logit}\left(p(Y_{ij} \leq k|x_{ij})\right) = \alpha_k + X'_{ij}\beta \qquad (1) \qquad k = 1, 2, ..., K-1$$

Where Y_{ij} is a multinomial response with K categories on j^{th} child within i^{th} household, α_k are the intercepts for each logit, X'_{ij} and β are vectors of explanatory variables and slope parameters, respectively. The important assumption for the PO model of common slope for each logit should be tested before making inferences based on this model. Independence correlation structure was assumed, since the implementation of this model in SAS is restricted to this assumption with multinomial distribution. The benefits of utilizing the ordinality of a response variable are to improve model parsimony, that is, the same fixed effect can apply to each logit and power of the model.

4.3.2 Alternating Logistic Regression (ALRs)

Alternating logistic regression is applied to clustered binary data with multiple levels of nesting to model the pattern of clustering and the marginal probability at the population level. It models within cluster or between clusters association with pairwise odds ratio. To analyse the children's nutritional status at two levels of nesting multivariate approach is required. Following the notation of [14], let Y_{ijk} denote the binary response on the k^{th} observation within the j^{th} household of the i^{th} cluster. Let X_{ijk} denote a p-covariate vector including cluster, household and observation-levels covariates and let Z_{ijkl} $(k < l)$ denote q-covariate vector associated with the k^{th} and l^{th} response vector for pair of observations indexed by (k, l) in the i^{th} cluster and j^{th} household. The log odds ratio of any two observations from different clusters is assumed to be zero, implying independence among them. With Y_{ijk} being distributed as a Bernoulli variate, the marginal model for the probability of being stunted or wasted is modelled assuming a logistic representation:

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \text{logit}(Y_{ijk} = 1) = X'_{ijk}\beta. \qquad (2)$$

The dependence modelled by;

$$\text{logor}(Y_{ijk}, Y_{ijl}) = \alpha_0 + \alpha_1 Z_{ijkl} \tag{3}$$

where $Z_{ijkl} = 1$ if pair of observations is from the same sub-cluster and 0 otherwise. In model 3, α_0 is the log odds ratio for association among observations from sub-clusters, and $\alpha_0 + \alpha_1$ is the log odds ratio for within sub-cluster association.

4.4 Model Building

The purpose of building a statistical model is to find an optimal model characterized by principles of generalizability, goodness-of-fit and parsimony based on model selection criteria. Model selection criteria are statistical tools help to identify best fitting model among the set of candidates. Several criteria for selecting subset of covariates that describe the data optimally differs from likelihood to non-likelihood estimation methods. The Akaikes Information Criterion (AIC) is widely used model selection criterion when the likelihood function is fully specified. On the other hand, when the likelihood function is not fully specified, e.g., as in the GEE setup, the AIC cannot be used for model selection. A modified Akaikes Information Criterion QIC_u based on the quasi-likelihood function[17], can be used instead. For this report three different working correlation assumptions, that is,exchangeable, independence, and unstructured were used to select the best model with minimum QIC_u based on score statistics and QIC was applied to select a working correlation structure. In order to select important factors related to nutritional status backward selection procedure was used. With this procedure we start with a complex model containing all main effects and two-way interactions to a more parsimonious model.

5 Results

5.1 Exploratory Data Analysis

In Mainland Tanzania, 6389 children were eligible for height and weight measurements to evaluate their nutritional status. Of the eligible children, 663 (10.4%) were refused and some had invalid values for the height, weight, and age. Approximately equal numbers of eligible females and males were measured, 2903(50.32%) and 2823(49.68%) respectively. Table 1 presents risk factors associated with wasting. Overall, 30.77% of children are wasted while 11.07% and 9.12% of females and males are severely wasted respectively. At the zone level, 50.60% of eligible children are from Lake, Southern, and Western Zones, 16.92% of them are wasted whilst 13.17% are severely wasted. With respect to age, children with at most 24 months of age, 29.77% are wasted while 20.05% are severely wasted. About 44.78% of eligible children who had not received vitamin A in last six months, 18.40% are wasted while 14.32% of them are severely wasted. Moreover, it is worth noting that, one in four Tanzanian women is un educated, likewise 5.89% of children under age 5 whose mothers have little or no education, are severely wasted.

Table 8 (in the Appendex) presents summary of risk factors associated with stunting. The table indicates that, approximately equal number of eligible females and males were measured. Overall, 45.97% of children are stunted, out of these 22.61% and 20.97% of females and males are severely stunted respectively. At the zone level, 70.28% of eligible children are from Southern highland, Lake, Northern, and Western Zones, where 29.93% are stunted while 28.58% are severely stunted. With respect to age, children with at most 24 months of age, 41.17% are stunted while 40.02% are severely stunted. For children who did not receive vitamin A for the last 6 months, 23.90% are stunted while 23.02% are severely stunted. Moreover, 11.61% of children under age 5 whose mothers have little or no education are severely stunted. It should be noted that these frequencies reflect the sum of the weights, that is the weights are treated as the inverse of selection probabilities.

Figure 1 shows the distribution of number children in each category of stunting and wasting by gender and age. It can be seen that most of the stunting and wasting cases

Table 1: Risk factors associated with wasting for children under five in Mainland Tanzania

Variable	Total(%)	Non-wasted(%)	Moderate(%)	Severe(%)
Gender				
Female	2903(50.32)	1952(33.84)	309(5.41)	642(11.07)
Male	2823(49.68)	2016(35.38)	297(5.18)	510(9.12)
Education				
Higher	13(0.24)	12(16.12)	–	1(0.018)
No education	1477(25.78)	997(17.24)	151(2.65)	329(5.89)
Primary	3979(69.01)	2784(48.38)	421(7.27)	774(13.36)
Secondary	257(4.96)	175(3.38)	34(0.67)	48(0.92)
BMI				
Normal	4008(69.63)	2704(46.86)	426(7.36)	878(15.41)
Overweight	1165(21.27	892(16.30)	101(1.99)	172(2.98)
Underweight	553(9.10)	372(6.07)	79(1.24)	102(1.80)
Birth weight				
Above 2.5kg	5536(96.81)	3862(67.53)	576(10.08)	1098(19.20)
Below 2.5kg	190(3.20)	106(1.70)	30(0.51)	54(0.97)
Zone				
Central	619(9.92)	408(6.56)	65(0.98)	146(2.38)
Eastern	623(11.85)	448(8.37)	69(1.45)	106(2.04)
S.highland	776(14.25)	402(7.71)	22(0.44)	352(6.11)
Lake	1048(20.91)	725(14.41)	113(2.35)	210(4.16)
Northern	856(13.39)	606(9.34)	91(1.44)	159(2.60)
Southern	599(7.95)	327(4.38)	12(0.14)	260(3.42)
Western	1205(21.73)	833(14.88)	109(1.91)	263(4.94)
Age				
Breastfeeding	2560(45.10)	864(15.32)	553(9.72)	1143(20.05)
Weaning	3166(54.90)	3104(53.90)	53(0.87)	9(0.13)
vitamin A				
Dont know	19(0.35)	16(0.29)	2(0.05)	1(0.01)
No	2555(44.78)	1505(26.37)	237(4.08)	813(14.32)
Yes	3137(54.88)	2435(42.54)	365(6.46)	337(5.88)

fall under the severe category, with number of females and children with at most 24 months being higher for both indicators in severe category.

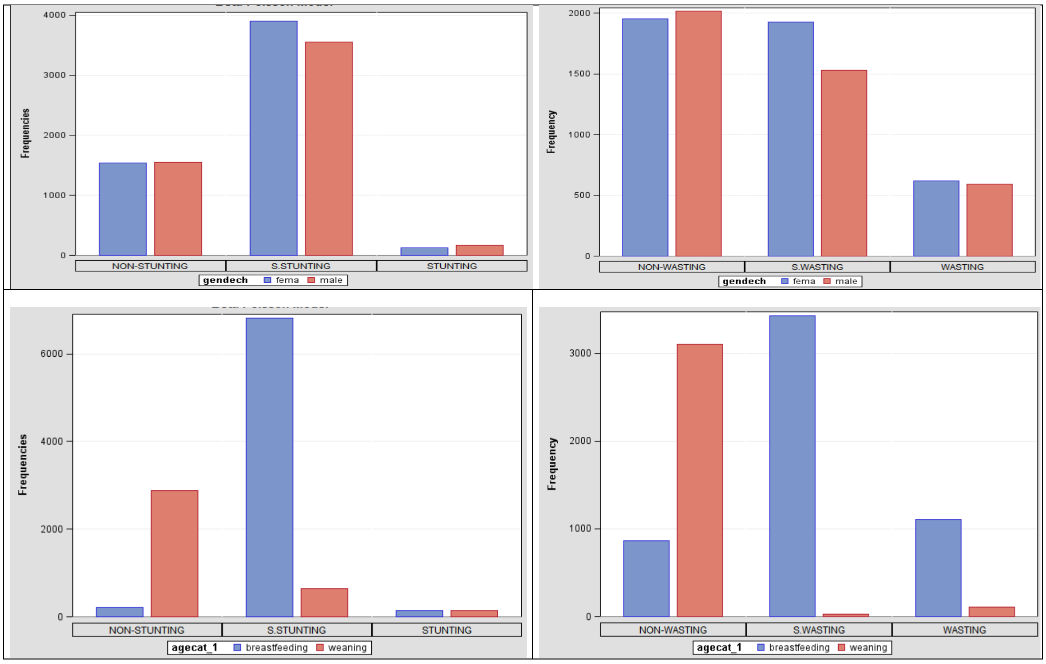

Figure 1: Distribution of stunting and wasting by age and sex

5.2 Marginal Models

The important covariates related to malnutrition was selected based on backward selection procedure. With backward selection, the terms for which its removal has the least damaging effect on the model are sequentially removed based on score statistics ($p > 0.05$), Agresti [1]. This means that the variables with least contribution to the model based on the highest p-value were eliminated sequentially and each time a new model with the remaining covariates was refitted, until we remained with covariates fit well the data for both wasting and stunting responses. However, none of the interaction terms were found to be significant. It turned out that the model with wasting as response and age, gender, zone, birth weight, maternal nutrition status and vitamin A as covariates was found to be the most parsimonious model. This model had the smallest QIC_u value for all three correlation structures. Finally, the comparison of empirical and model based standard errors for the parameter estimates obtained based on the three working correlation assumptions (in this study exchangeable, independence and unstructured) was performed. Exchangeable working correlation assumption was found to be more plausible since the two standard errors were close. The final model for wasting as a response is given as:

$$\text{Logit}(\pi_{ij}) = \beta_0 + \beta_1 \text{weight}_{ij} + \beta_2 \text{WBM1}_{ij} + \beta_3 \text{agecat}_{ij} + \beta_4 \text{vitA}_{ij} + \beta_5 \text{gender}_{ij} + \beta_6 \text{zonal}_{ij}.$$

Finally, using the selected covariates , a two-level ALR model which provides information about pairwise association of observations between two different individuals within the same household was fitted. Later this model was extended to a three-level ALR to accommodate the association of pairs of responses from two different households within the same cluster. For wasting as response, the QIC_u values of 4559.98 and 4560.80 for the two and three-levels ALR models respectively, it was concluded that the two-level ALR model was a better model in explaining the population-averaged association between wasting and the selected predictor variables. Thus, interpretation rely on the two level ALR model. Table 2 presents parameter estimates and their corresponding empirically corrected standard errors alongside the p-values from GEE and two-level ALR. Each parameter reflects the effect of factor on the log odds of having low weight for height (wasted), statistically controlling for all other factors. Overall, parameter estimates under ALR are slightly less than or equal to those of GEE. The slight differences in parameter estimates could be attributed to the fact that ALR takes the association into account, whereas GEE treats the association as a nuisance parameter.

Table 2: Parameter estimates and empirical standard errors for two-level models

Marginal Model		**GEE**		**ALR**	
Effect	**Par.**	**Est.(s.e.)**	p	**Est.(s.e.)**	p
Intercept	β_0	2.094(0.993)	0.035	2.128(1.006)	0.034
Birth weight					
$\geq 2.5kg$	β_1	-1.1323(0.309)	0.0002	-1.112(0.310)	0.0003
Maternal Health					
Normal	β_2	0.375(0.173)	0.030	0.385(0.174)	0.027
Underweight	β_3	0.5004(0.145)	.0001	0.497(0.116)	.0001
Age					
> 24 *month*	β_4	-4.876(0.167)	.0001	-4.876(0.167)	.0001
Vitamini A					
No	β_5	0.027(0.927)	0.977	-0.027(0.927)	0.9772
Yes	β_6	-0.946(0.927)	0.327	-0.946(0.927)	0.3077
Don't know	β_7	0.069(1.328)	0.988	-0.070(1.328)	0.9583
Gender					
Male	β_8	-0.322(0.089)	0.0003	-0.327(0.089)	0.0002
Zone					
Western	β_9	-0.517(0.166)	0.002	-0.516(0.166)	0.0018
Northern	β_{10}	-0.081(0.194)	0.676	-0.082(0.194)	0.6744
S . highland	β_{11}	-0.415(0.193)	0.032	-0.409(0.195)	0.0359
Lake	β_{12}	-0.089(0.181)	0.623	-0.088(0.181)	0.6253
Eastern	β_{13}	-0.234(0.202)	0.247	-0.231(0.202)	0.2537
Southern	β_{14}	0.004(0.188)	0.983	0.005(0.188)	0.9779
Correlation	ρ	0.00000967			
Alpha	α			-0.046(0.027)	0.0910

The analysis under ALR with two-levels suggest that birth weight is significantly related to wasting, it was observed that children who had more than 2.5kg birth weight had exp(-1.112)=0.329 times lower odds of wasting than children who had less than 2.5kg birth weight. With respect to maternal health effect, it was observed that maternal health is significantly related to wasting. This implies that, the estimated odds of wasting for children belong to mothers who are underweight was 11.85% more than the estimated odds for normal weight mothers. Furthermore, the analysis suggest that age is significantly related to wasting, this means that estimated odds of wasting for children with more 24 months of age is 99.23% less than the estimated odds for those with less than 24 months of age. For the gender effect, male children had exp(-0.327)=0.721 times lower odds of wasting than their female counterparts. Moreover, it was observed that, the estimated odds of wasting for children from Western and Southern Highland zones are respectively 40.3% and 33.6% less than the estimated odds of children from central zone. Table 2 also presents the estimated constant log odds ratios (α_1), which provides information about association between individuals within a household. This means that, the estimated pairwise odds ratio relating two responses from the same household was found to have a small negative association i.e. -0.046, this underscoring the weak association in wasting within a household.

For stunting as a response, the model with birth weight, maternal health, age, and gender as covariates was found to be the most parsimonious model with the smallest QIC_u value. Additionally, a two and three-levels ALR model was fitted and their corresponding QIC_u values of 4145.52 and 4146.44 was observed, similar conclusion as earlier was reached. Thus, interpretation rely on the two level ALR model. Table 3 presents parameter estimates and their corresponding empirically corrected standard errors alongside the p-values from GEE and two-level ALR. Each parameter reflects the effect of factor on the log odds of stunting, statistically controlling for all other factors. Finally, the comparison of empirical and model based standard errors for the parameter estimates obtained based on the three working correlation assumptions (in this study exchangeable, independence, and unstructured) was performed. An exchangeable working correlation assumption was found to be more plausible since the two standard errors were very close.

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 \text{weight}_{ij} + \beta_2 \text{WBM1}_{ij} + \beta_3 \text{agecat}_{ij} + \beta_4 \text{gender}_{ij}.$$

Table 3: Parameter estimates and empirical standard errors for two-level models

Marginal Model		**GEE**		**ALR**	
Effect	**Par.**	**Est.(s.e.)**	p	**Est.(s.e.)**	p
Intercept	β_0	2.719(0.259)	$< .0001$	2.720(0.260)	$< .0001$
Birth weight					
$\geq 2.5kg$	β_1	-0.803(0.228)	0.0004	-0.807(0.228)	0.0004
Maternal Health					
Normal	β_2	0.708(0.195)	0.0003	0.710(0.194)	0.0003
Underweight	β_3	0.623(0.141)	$< .0001$	0.625(0.141)	$< .0001$
Age					
> 24 *month*	β_4	-4.749(0.107)	$< .0001$	-4.749(0.107)	$< .0001$
Gender					
Male	β_5	-0.207(0.101)	0.0392	-0.208(0.101)	0.0395
Correlation	ρ	-0.00249145			
Alpha	α			-0.002(0.045)	0.9638

Table 3 presents the analysis under ALR with two-levels suggest that birth weight is significantly related to stunting, it was observed that children who had more than 2.5kg birth weight had exp(-0.807)=0.446 times lower odds of stunting than children who had less than 2.5kg birth weight. With respect to maternal health effect, it was observed that, the estimated odds of stunting for children belong to underweight mother is 8.15% less than the estimated odds for children belong to normal weight mothers. Moreover, the analysis suggest that age is significantly related to stunting, this implies that, children with more 24 months had exp(-4.749)=0.00866 times lower odds of stunting than those with less 24 months. For gender effect, male children had exp(-0.208)=0.812 times lower odds of stunting than their female counterparts. Table 3 also presents the estimated constant log odds ratios (alpha) which provide information about association between individuals within a household. he estimated pairwise odds ratio relating two responses from the same household was found to have a small negative association i.e. -0.002, this in turn implies that there is a weak association between children in a household in terms of stunting.

5.3 Proportional Odds Model

The binary outcome for the response of interest (stunting, not stunting) can be extended to accommodate the ordering of severity of the two indicators. Hence a commonly used extension of logistic regression to accommodate the ordering of levels is proportional odds model. Generalized estimating equations (GEE) is extended to accommodate ordering of the levels in stunting or wasting. Stunting was categorized using a three-point ordinal scale: 1 = non-stunted, 2 = moderately stunted, 3 = severely stunted.

5.3.1 Proportional Odds Assumption

Before inference could be made, the assumption of common slope for proportional odds model had to be tested first. The score test for the null hypothesis of constant slope gave a χ^2 test statistic of 8.53, on 5 degree of freedom, leading to a p-value of 0.13 for stunting as a response. There is no statistical evidence to reject the null hypothesis (i.e., that the PO assumption holds) and can proceed to examine the model output. Table 4 summarizes the results of the fitted marginal ordinal logistic regression model. The estimated cumulative odds ratios and 95% confidence intervals for the odds ratios for the selected predictor variables, the results suggest that birth weight, maternal health, gender and age are significant predictors in the cumulative logit model for stunting as their 95% confidence intervals for the odds ratios do not include 1.

Table 4: Estimated Cumulative Logit Model

Effect	Par.	Est.(s.e.)	p	OR[95% C.I]
Intercept	α_0	2.472(0.247)	$< .0001$	7.737[6.48,9.24]
Intercept	α_1	2.781(0.249)	$< .0001$	10.538[8.75, 12.69]
Birth weight				
$\geq 2.5kg$	β_1	-0.754(0.225)	0.0008	0.470[0.303,0.731]
Maternal Health				
Normal	β_2	0.063(0.142)	0.656	1.065[0.81,1.41]
Underweight	β_3	0.573(0.126)	$< .0001$	1.774[1.386,2.268]
Age				
> 24 *month*	β_4	-4.752(0.105)	$< .0001$	0.0086[0.007,0.011]
Gender				
Male	β_5	-0.244(0.096)	0.011	0.783[0.649,0.945]

From Table 4 the coefficient for birth weight is negative, suggesting that birth weight is negatively related to stunting. This implies that holding other covariates in the model constant, the estimated cumulative odds of children with birth weight above 2.5kg is 53.05% less than estimated cumulative odds for children having less than 2.5kg birth weight. For maternal health effect, the coefficient for maternal health (underweight) is positive, suggesting that underweight is positively related to stunting. This means that, the estimated cumulative odds for children belong to underweight mothers being in normal category relative to severe stunting category increases by 77.4% more than cumulative odds for children belong to overweight mothers. Moreover, the age was found to be negatively related to stunting. This means that, the estimated cumulative odds for children with more than 24 months being in the severe category relative to the normal category decreases by approximately 99.14% less than cumulative odds for children with less than 24 months. The analysis revealed that gender is negatively related to stunting. This in turn implies that the cumulative odds that males are in the severe category of stunting relative to normal category decreases by approximately 21.7% less than the cumulative odds for their females counterpart. It is worth noting that, this interpretation holds across the entire range of stunting status, that is from severe stunting to non-stunting. This follows from the fact that the assumption of common slope was not rejected.

5.4 Baseline Logit Model

A similar score test for the null hypothesis of common slope was performed on wasting as a response and gave a χ^2 test statistic of 76.38, on 7 degree of freedom, leading to a p-value of 0.0001. Therefore, the null hypothesis of common slope is rejected, that is, the proportional odds assumption does not hold. Therefore, a baseline-category logit model for nominal responses which uses separate binary logit for each pair of response to describe the effects of predictors on these logits, Agresti [1] is considered. The effects of predictors vary according to the response paired with the baseline. baseline category logit model was fitted to investigate the effects of birth weight, maternal health, age, gender, zone, and vitamin A on wasting.

Table 5 presents the analysis under a baseline logit model. The results suggest that birth weight is significantly related to wasting. Given that other predictors in the model are constant, it was observed that for children with more than 2.5kg birth weight, the estimated odds of severe wasting instead of moderate wasting are $\exp(-0.106) = 0.899$ times lower the estimated odds of children with less than 2.5kg birth weight. Moreover, maternal health was found to be significantly related to wasting. This in turn implies that, the estimated odds of severe wasting instead of moderate wasting for children belong to normal weight mothers are 37.03 % more than the estimated odds for children belong to overweight mothers. With respect to age, it was found to be significantly related to wasting. This implies that, the estimated odds of severe wasting instead of moderate wasting for children with more than 24 months are 92.94% less than estimated odds of children with at most 24 months. Moreover, the gender effect indicate that the estimated odds of severe wasting instead of moderate wasting for male children are 17.47% less than the estimated odds for female children.

Table 5: Estimated Baseline Logit Model

	Logit(moderate wasted vs. non-wasted)		logit(severe wasted vs. non-wasted)	
Effect	**Par.**	**Est.(s.e.)**	**Par.**	**Est.(s.e.)**
Intercept	α_1	1.178 (1.047)	α_2	1.290(1.250)
Birth Weight				
($\geq 2.5kg$)	β_{02}	-1.096(0.305)*	β_{03}	-1.202 (0.365)*
Maternal Health				
Underweight	β_{12}	0.496 (0.201)*	β_{13}	0.403 (0.211)
Normal	β_{22}	0.368 (0.150)*	β_{23}	0.683(0.136)*
Age				
($> 24months$)	β_{02}	-3.742 (0.168)*	β_{03}	-6.393(0.356)*
Vitamin				
Yes	β_{12}	-0.725 (0.955)	β_{13}	-0.616(1.183)
No	β_{02}	-0.585 (0.962)	β_{03}	0.876(1.188)
Don't Know	β_{82}	0.345 (1.285)	β_{83}	-1.156(1.676)
Gender				
Male	β_{12}	- 0.233 (0.114)*	β_{13}	-0.425(0.103)*
Zone				
Western	β_{42}	-0.110 (0.229)	β_{43}	-0.584(0.0.224)*
Northern	β_{22}	0.106 (0.242)	β_{23}	-0.184(0.213)
S. Highland	β_{52}	0.017 (0.226)	β_{53}	-0.229(0.203)
Lake	β_{12}	-0.412 (0.226)	β_{13}	-0.666(0.188)*
Eastern	β_{62}	-0.042(0.252)	β_{63}	-0.419(0.234)
Southern	β_{72}	0.174 (0.245)	β_{73}	0.0150(0.209)

* significant with $p < 0.05$

5.5 Missing Data

Table 7 (in the Appendix) shows the parameter estimates and their standard errors of the fitted two-level ALR and GEE models using imputed dataset. These results from Table 7 (appendix) can be compared with the results from Table 2. From these two tables, it can be observed that the estimated standard errors for the MI estimation of the ALR and GEE models are generally less than or equal to the standard errors of the coefficients from the analysis of the standard ALR and GEE. The slightly smaller standard errors from the MI analysis reflect the fact that the imputation has recovered additional statistical information from the cases that were excluded by procedure from the analysis of the non-imputed data. Furthermore, models fitted using standard ALR and GEE as well as ALR and GEE with Multiple Imputation approaches provide comparable parameter estimates. That is, in this particular model, the differences are not large enough to require remedial action to assess the statistical significance of the individual predictors in the final model selection.

6 Discussion and Conclusion

This project utilized the 2010 demographic and health survey data to identify the risk factors associated with stunting and wasting among children of under-five in Mainland Tanzania. The two indicators might be differently distributed within as well as across households, since each indicates a different mechanism by which such nutritional deficits are acquired. To take into account clustering and unequal selection probability, multi-stage sampling design and sampling weights were considered in analysis. The marginal models were employed to account for clustering as well as unequal selection probability whilst the proportional odds model was employed to study the severity of stunting.

For wasting as a response, marginal models led to the same conclusion that age, gender, birth weight, maternal health, and zone were found to be significantly related to wasting (low weight for height). Age, birth weight, gender and zone had a negative effect indicating that male, children with more than 2.5kg birth weight, more than 24 months, Western, and Southern highland zones are associated with low probabilities of wasting whilst maternal health had positive effect indicating that children who belong to underweight and normal weight mothers are associated with higher probabilities of wasting.

With respect to stunting, gender, age, maternal health and birth weight were found to be significantly related to stunting (low height for age). Birth weight, age and gender had negative effect indicating that male, children with more than 2.5kg birth weight, and more than 24 months are associated with low probability of stunting whilst maternal health had positive effect indicating that children belong to underweight mothers are associated with high probability of stunting. Proportional odds model led to the conclusion that male, children with more than 2.5kg birth weight, and more than 24 months are associated with decreasing stunting risk whilst maternal health are associated with increasing stunting risk.

Furthermore, the two-level alternating logistic regression under the marginal model family indicated a small negative association between any two pairs of responses from the same household. The small negative association between any pair of responses is an indication of weak association between children within a household in terms of stunting and wasting.

The problem of missing data is one that is almost unavoidable. Incorrectly accounting for missing data can lead to invalid inference and misleading study conclusions. This may in turn have detrimental effects in real life decision making. In this study, multiple imputation technique was used to take into account the missingness. However, the differences of results from models fitted with and without Multiple Imputation were not large enough to require remedial action to reverse decision concerning the statistical significance of the individual predictors in the final model. The same results and conclusion was observed for stunting with multiple imputation

.

7 Recommendations

Based on the highly significant effect of birth weight and maternal health on stunting and wasting, it would then be important for government and other stakeholders to improve primary health care facilities and strengthening nutritional programmes related to food insecurity and micronutrients. The intervention would reduce the burden of child mortality due to stunting and wasting as well as risk of low birth weight. Additionally, in this analysis, it is observed that the risk of being stunted or wasted depends on age of a child, gender, maternal health, geographical zones. However, it is worth noting that the probability of being stunted or wasted could also be associated with other factors such as, birth interval, sanitation, household size and culture. Investigation of such factors could be recommended in future studies. However, challenges may lie on the side of resources made available and possibly means of collecting these factors.

References

[1] Agrest, A. (2002). *Categorical Data Analysis* (2nd ed.). New York: John Wiley and Sons.

[2] Akinlosotu, R.A. and Hassain, M.A. (1985). The usefulness of height at first year of schooling as an index of nutritional status of a community. *Journal of Nutrtional Science*, **6(2)**, 131-138.

[3] Berkowitz, G.S. and Papiemik, E. (1993). Epidemiology of preterm birth. 4. *Epidemiology Review*, **15**, 414-443.

[4] Brown, K.H., Dewey, K.G., and Allen, L.H. (1998). Complementary Feeding of Young Children in Developing Countries: *A Review of Current Scientific Knowledge.* Geneva Switzerland: World Health Organization.

[5] Brown, K.H., et al. (1990). *Effects of common illnesses on infants' energy intakes from breast milk and other foods during longitudinal community-based studies in Huascar* (Lima), Peru. Am J Clin Nutr, **52**, 1005-1013.

[6] Black, R., Morris, S., and Jennifer, B. (2003). *Where and Why Are 10 Million Children Dying Every Year?*, **361**, 2226-2234.

[7] Bruckers Liesbeth., Molenberghs, G., Tibaldi Fabian.,and Van Steen Kristel. (2001). *The Health Interview Survey* : Protocol for the sampling design . Limburgs Universitair Centrum Center for Statistics; Scientific Institute of Public Health.

[8] Carey, V.C., Zeger, S.L., and Diggle, P.J. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, **80**, 517-526.

[9] Davidson, S. (1975). *Human Nutrition and Dietetics.* ELBS Church Living Stone Publishers, Great Britain, pp: 20.

[10] Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data.* Oxford: Claredon Press.

[11] Ene-Obong, H.N. (2001). *Eating right (A Nutrition Guide).* The University Press, Calabar.

[12] FAO. (2008). *The State of Food Insecurity in the World.* Food and Agriculture Organization, Rome. Retrieved on December 29,2012

[13] Laditan, A.A., (1983). Nutrition and physical growth in children Nigeria. *Journal of Nutrtional Science*, **4**, 5-10.

[14] Liang, K.Y., Zeger, S.L.,and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B*, **54**, 3-40.

[15] Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). New York: Wiley.

[16] Lipsitz, S.R., Laird, N.M., and Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: using odds ratio as measure of association.*Biometrika*, **78**, 153-160.

[17] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*(2nd edition). London: Chapman and Hall.

[18] Myers, R.H., Montgomery, D.C., Vining, G.G., and Robinson, T.J. (2010). *Generalized Linear Models with Application in Engineering and the Sciences*(2nd ed.). New Jersey: John Wiley and Sons.

[19] Ministry of Health and Social Welfare (MOHSW) [United Republic of Tanzania]. (2010). *National Food and Nutrition Policy.* Dar es Salaam, Tanzania:

[20] Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* New York: Springer.

[21] National Bureau of Statistics (NBS) [Tanzania] and ICF Macro. (2010) *Tanzania Demographic and Health Survey 2010, Preliminary report.* Calverton, Maryland: NBS and ICF Macro.S

[22] Onimawo, I.A., Amangbangwu, K.E. and Eluwa, M.A. (2006). Nutritional Status of school aged children assessed by methods. *Nigeria Journal of Nutrtional Science*, **1(2)**: 27-34.

[23] Onyesili, F.N., (2000). *Nutrition Education in Child Development and Eradication of Disease.* Invited Paper for the 31st Annual Conference and Scientific Meeting of the Nutrition Society of Nigeria Nov. 7 2000, Banquet Hall, National Centre Women for Development, Abuja.

[24] Prentice, R.L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033-1048.

[25] Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

[26] WHO. (2006). *WHO child growth Standards: Length/height for age, weight for age, weight for length, weight for height and body mass index for age. Methods and development.* Geneva, Switzerland.

[27] World Bank. (2009). *World Development Indicators.* Washington DC, World Bank.

[28] WHO. (1992). *Low birth weight.* Geneva, Switzerland.

[29] UNICEF. (1990). *Strategies of improving nutrition of children and women in developing countries.* New York: UNICEF.

[30] UNICEF. (1998). *The State of the Worlds Children.* Oxford University Press, Oxford.

[31] Van de Poel, E., Hosseinpoor, A., Jehu-Appiah, C. and Speybroeck, N. (2007). *Malnutrition and Socioeconomic Gap in Malnutrition in Ghana.* Mimeo.

[32] Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York: Springer.

[33] Vice-Presidents Office [Tanzania] (2005). *National Strategy for Growth and Reduction of Poverty (NSGRP).* Dar es Salaam, Tanzania.

[34] Zeger, S.L., Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrika* **42**, 121-130.

8 Appendix

Table 6: Parameter estimates and empirical standard errors for imputed data

Marginal Model		**GEE**		**ALR**	
Effect	**Par.**	**Est.(s.e.)**	p	**Est.(s.e.)**	p
Intercept	β_0	1.272(0.926)	0.181	1.271(0.928)	0.182
Birth weight					
$\geq 2.5kg$	β_1	-0.680(0.245)	0.008	-0.676(0.244)	0.007
Maternal Health					
Normal	β_2	0.531(0.102)	$< .0001$	0.532(0.103)	$< .0001$
Underweight	β_3	0.589(0.147)	$< .0001$	0.593(0.147)	$< .0001$
Age					
> 24 *month*	β_4	-3.645(0.095)	$< .0001$	-3.646(0.095)	$< .0001$
Vitamini A					
No	β_5	0.099(0.831)	0.906	0.094(0.827)	0.910
Yes	β_6	-0.489(0.816)	0.553	-0.493(0.813)	0.549
Don't know	β_7	0.481(0.798)	0.549	0.476(0.797)	0.553
Gender					
Male	β_8	-0.315(0.074)	$< .0001$	-0.316(0.074)	$< .0001$
Zone					
Western	β_9	-0.574(0.151)	0.002	-0.573(0.151)	0.0002
Northern	β_{10}	-0.074(0.148)	0.619	-0.074(0.148)	0.617
S . highland	β_{11}	-0.482(0.154)	0.002	-0.481(0.154)	0.002
Lake	β_{12}	-0.158(0.146)	0.283	-0.156(0.146)	0.286
Eastern	β_{13}	-0.456(0.163)	0.0.006	-0.455(0.163)	0.006
Southern	β_{14}	0.063(0.177)	0.723	0.065(0.178)	0.720
Correlation	ρ	-0.000245			
Alpha	α			-0.014(0.026)	0.0910

Table 7: Risk factors associated with stunting for children under five in Mainland Tanzania

Variable	Total(%)	Normal(%)	Moderate(%)	Severe(%)
Gender				
Female	2903(50.32)	1539(26.66)	62(1.04)	1302(22.61)
Male	2823(49.68)	1554(27.35)	84(1.35)	1185(20.97)
Education				
Higher	13(0.24)	10(0.17)	–	3(0.08)
No education	1477(25.78)	781(13.61)	35(0.56)	661(11.61)
Primary	3979(69.01)	2169(37.68)	106(1.78)	1704(29.56)
Secondary	257(4.96)	133(2.57)	5(0.07)	119(2.33)
BMI				
Normal	4008(69.63)	2097(36.37)	105(1.76)	1806(31.49)
Overweight	1165(21.27	712(13.12)	31(0.51)	422(7.65)
Underweight	553(9.10)	284(4.52)	10(0.15)	259(4.43)
Birth weight				
Above 2.5kg	5536(96.81)	3009(52.67)	142(2.35)	2385(41.79)
Below 2.5kg	190(3.20)	84(1.34)	4(0.07)	102(1.79)
Zone				
Central	619(9.92)	331(5.23)	13(0.21)	275(4.48)
Eastern	623(11.85)	361(6.64)	19(0.31)	243(4.88)
S.highland	776(14.25)	545(10.32)	87(1.53)	144(2.40)
Lake	1048(20.91)	576(11.40)	27(0.50)	445(9.02)
Northern	856(13.39)	469(7.33)	29(0.43)	358(5.63)
Southern	599(7.95)	403(5.35)	72(0.93)	124(1.67)
Western	1205(21.73)	627(11.31)	24(0.39)	544(10.03)
Age				
Breastfeeding	2560(45.10)	216(3.92)	72(1.16)	2272(40.01)
Weaning	3166(54.90)	2877(50.10)	74(1.25)	215(3.56)
Vitamin A				
Dont know	19(0.35)	14(0.24)	–	5(0.11)
No	2555(44.78)	1192(20.87)	54(0.88)	1309(23.02)
Yes	3137(54.88)	1875(32.83)	92(1.54)	1170(20.51)

```
/*FINAL MODELS FOR WASTING RESPONSE*/;
proc genmod data=mydta_new descending;
title 12. GEE logistic regression, for Tanzania;
title2 weighted + clustered;
weight wfin;
class HID region CLID weight1 (ref="1") WBM1 (ref="3") agecat (ref="1")
VitA (ref="9") zonal(ref="3") gender (ref="2")/param=ref;
model wasted = weight1 WBM1 agecat vitA gender  zonal/ dist=b
link=logit type3;
repeated subject = HID / type=exch modelse;
run;quit;

/*FINAL MODEL FOR STUNTING RESPONSE*/;
proc genmod data=mydta_new descending;
title 12. GEE logistic regression, for Tanzania;
title2 weighted + clustered;
weight wfin;
class ClID HID region weight1 (ref="1") WBM1 (ref="3")
 agecat (ref="1") gender (ref="2")/param=ref;
model stunted = weight1 WBM1 agecat gender/ dist=b link=logit type3;
repeated subject =HID / type=exch modelse;
run;quit;

/*FITTING ALR MODEL FOR STUNTING RESPONSE-TWO LEVEL*/;
proc genmod data=mydta_new descending;
title 12. GEE logistic regression, for Tanzania;
title2 weighted + clustered;
weight wfin;
class  HID weight1 (ref="1") WBM1 (ref="3") agecat (ref="1")
 gender (ref="2")/param=ref;
model stunted = weight1 WBM1 agecat gender/ dist=b link=logit type3;
```

```
repeated subject =HID / logor=exch modelse;
run;quit;

/*FITTING PROPORTIONAL ODDS MODEL FOR STUNTING*/
proc genmod data=mydta_new descending;
title " Wasting, GEE Ordinal";
weight wfin;
class  HID stuntingcases(ref="1") weight1(ref="1") WBM1(ref="3")
agecat(ref="1") gender (ref="2")/param=ref;
model stuntingcases = weight1 WBM1 agecat gender/dist=multinomial
link=clogit lrci type3;
repeated subject=HID/type=ind modelse;
run;

 /*BASELINE LOGIT MODEL*/;
 proc surveylogistic data=baseline1;
class weight1 (ref='1') WBM1(ref='3') agecat(ref='1') gender(ref='2')
vitA(ref='9')  zonal(ref='3')/order=data param=ref ref=first;
weight wfin;
cluster HID;
model mulcat2(ref="1") = weight1 WBM1 agecat gender vitA  zonal/ link=glogit ;
run;quit;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Determinants of nutritional status among children under age 5 in mainland Tanzania

Richting: **Master of Statistics-Biostatistics**
Jaar: **2013**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Machafuko, Pendael Zephania

Datum: **4/02/2013**