Journal of Biostatistics and Biometric Applications
Volume 3 | Issue 2
ISSN: 2455-765X

Research article Open Access

Modeling of Longitudinal Factors on Children Body Mass Index at Bahir Dar Districts: Linear Mixed-Effects Model

Alemu AA*

Department of Statistics, College of Computing and Informatics, Haramaya University, Dire Dawa, Ethiopia

***Corresponding author:** Alemu AA, Department of Statistics, College of Computing and Informatics, Haramaya University, Dire Dawa, Ethiopia, Tel: +251948691620, E-mail: aleb.abebe@yahoo.com

Citation: Alemu AA (2018) Modeling of Longitudinal Factors on Children Body Mass Index at Bahir Dar Districts: Linear Mixed-Effects Model. J Biostat Biometric App 3(2): 202

Received Date: February 13, 2018 **Accepted Date:** June 12, 2018 **Published Date:** June 13, 2018

Abstract

Body mass index is calculated as weight in kilograms divided by square height in meters (kg/m^2). Its formula was developed by Belgium Statistician, and was known as the Quetelet Index [1]. It provides a reliable indicator of body fatness for most people and is used to screen weight categories that may lead to health problems. Body mass index is an internationally used measure of health status of an individual. This study was modeling of longitudinal factors on children body mass index at Bahir Dar districts. The study was based on data from 1900 pre four visits (475 per individual) children enrolled in the first 4 visits of the 4-year Longitudinal data of children at Bahir Dar districts of North West Ethiopia. Linear mixed effects model was used to describe the relationships between children body mass index and some covariates accounting for the correlation among the repeated observations for a given children. There were statistically significant (P-value<0.05) difference among children body mass index variation with respect to time, Sachet, age, residence, Antiretro-Viral Therapy and diarrhea. While fever, cough, Mid-Upper Arm Circumference and sex were statistically insignificant (P-value>0.05) effect on children body mass index. According to the findings of this study about 29.28% were normal weight, 67% were under weight, 2.52% were overweight and only 1.21% were obesity. Therefore, the study suggests that concerned bodies should focus on awareness creation to bring enough food to under-age five children at Bahir Dar districts especially in rural areas.

Keywords: Marginal Model; Hierarchical Model; Linear Mixed-Effects Model

Introduction

Body Mass Index (BMI) also known as Quetelet Index formula was developed by Belgium Statistician [1]. BMI is calculated as weight in kilograms divided by square height in meters. Centers of Disease Control and Prevention (CDC) and the American Academy of Pediatrics (AAP) recommend the use of BMI to screen for overweight and obesity in children beginning at the age of 2 years. BMI is used to screen for obesity, overweight, healthy weight or underweight. However, BMI is not a diagnostic tool. According to WHO technical report series 854:9, the BMI range less than 18.5 indicates underweight, 18.5 to 25 indicates normal weight, 25 to 30 indicates over weight and over 40 indicates obesity.

In recent years, overweight and/ or obesity among children have emerged as a global epidemic [2]. The fundamental cause of obesity and overweight is an energy imbalance between calories consumed and calories expended. Globally there has been an increased intake of energy-dense foods that are high in fat, salt and sugars but low in vitamins, minerals and other micronutrients and decrease in physical activity due to the increasingly sedentary nature of many forms of work, changing modes of transportation and increasing urbanization [3]. In 2005, the WHO reported that at least 400 million adults were obese and at least 20 million children under the age of 5 years are overweight globally. WHO further projects that by 2015 approximately 2.3 billion adults were overweight and more than 700 million were obese. There is still much uncertainty related to the causes and underlying physiological mechanisms of obesity. Recent scientific findings are able to show that long-term risk of obesity and related disorders begin very early in life [4].

In Africa, despite a high prevalence of under nutrition, the prevalence of overweight is increasing at an alarming rate. It is estimated that 25% to 60% of urban women are overweight [5]. According to Tanzania Health Research (2006) conducted in Simanjiro District, 82% of the adolescents had normal health status while 0.8% were overweight for their ages, 14.0% were moderately wasted and 3.2% were severely wasted. This could be explained by the change in the life style factors of the society. Moreover, the study was conducted on the prevalence of overweight and/ or obesity in Sudan which was 14% [6].

In Ethiopia, one study was conducted in Addis Ababa in 2007 reported that the prevalence of overweight and obesity on elementary school students were 7.6% and 0.9% respectively [7]. In 2014 in Addis Ababa conducted that prevalence of overweight and/ or obesity and associated factors among high school adolescents in Arada sub-city were 72.1% normal-weight, 18.5% underweight, 8.6% overweight and 0.8% obesity [1]. This might be due to the food eaten in Sudan were highly energy dense foods and there is also frequent eating habits while in Ethiopia mostly eaten foods are fibers and cereals, three times a day. Overweight and/ or obesity during childhood increases the risk for the development of non-communicable diseases and predisposes the individual to the development of overweight, obesity, cardiovascular disease and metabolic and other disorders in adulthood and childhood [1].

Children BMI are a serious threat to the governments' effort to meet the growth and transformation plan (GTP) two. Further, there are also few local researchers who have done on the issue, however, their method of analysis has descriptive in nature and limited to examine underline factors association between children BMI status with certain BMI-related covariate. Furthermore, most of those studies are based on small-scale survey data that obtained from certain districts. There had been no detailed scientific study on this thematic area especially at Bahir Dar districts, Ethiopia. The study will help to fill the gap of knowledge on BMI of children and the output will help to recommend bodies for primary prevention. The objective of this study is to model the longitudinal factors on children BMI by using the best robust linear mixed effects model was compressed the secondary data. In this study, model development procedures were Akaike Information Criteria (AIC) and Bayesian Information (BIC) and also forward selection were used for model selection.

Material and Methods

This study was based on data from 475 children enrolled in the first four visits of the four year longitudinal study of children BMI using the data of Bahir Dar districts working collaboration with Save the Children. The study area was located at Bahir Dar, Ethiopia and serve as a capital city of Amhara regional state, 563 kilometers far away from Addis Ababa in North West direction. This study was retrospective study on longitudinal data setting design that go back in time to assess exposure to known the trend and modeling of longitudinal factors on children BMI. The different socio-economic status, demographic, disease types and biological/ clinical characteristics were collected repeatedly in four waves between the years on January 2012 to 2016. Each repeated measures were conducted within one month interval in the study period. In this study both times invariant and variant covariates were employed. The first wave was conducted on January 2012 to 2013 within one month periods in 4 repetitions. On the contrary, the second, third and fourth waves were conducted on January 2013 to 2014, 2014 to 2015 and 2015 to 2016 were presented respectively.

Variables Considered in the Study

Children BMI health status was considered as the response variable. BMI (in a standardized form) was used as a continuous variable to maximize the amount of information available in the data set. And also, Explanatory variables (Covariates) are time of child treatment, age of child, child's sex (female, male), amount of sachet, place of residence (rural, urban), cough status (yes, no), diarrhea status (yes, no), fever status (yes, no), amount of Mid-Upper Arm Circumference (MUAC) and Antiretro-Viral Therapy (ART) treatment (on ART, on pre-ART and no ART).

Inferential Statistics

Methods of Statistical Analysis: Longitudinal data is a special case of repeatedly measured data, the observations are not independent and are characterized as having both between-subject and within-subject variation, time dependent covariates and missing data [8]. The variance covariance structure does not have to be independent. Data can be balanced or each subject does need to have the same number of observations per subject and repeated measurements have equal time intervals. Furthermore, mixed-effects modeling have become increasingly popular, more accessible and good in missing data handling through statistical software such as SAS Virson-9.2 [9].

Linear Mixed Effects Model: Linear Mixed-Effects Model can be used to accommodate complex features of longitudinal data where as traditional methods are limited by statistical assumptions, have become increasingly popular and more accessible for continuous response variable was proposed [10]. LMM assumes that the observations follow a linear regression where some of the regression parameters are fixed or the same for all subjects, while other parameters are random or specific to each subject [11]. While; mean population parameters, individual effects and within-person variation make up the first stage of the model [10]. The general form of the LMM after combining the two stages is approximately normal [8].

A two-stages of LMM was given by:

Stage 1: Linear regression model for each subject separately response Y_{ij} for i^{th} subject, measured at time t_{ij} :

$i = 1, - - -, j; j = 1, - - -, n_i$ *response vector* Y_i *for* i^{th} *subjects* : $Y_i = (Y_{i1}, Y_{i2}, - - -, Y_{in_i})'$

$$Y_i = Z_i\beta_i + \varepsilon_i \qquad (2.1)$$

Where, $Z_i = \begin{bmatrix} 1 & \cdots & t_{i1} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & t_{in_i} \end{bmatrix}$

Z_i is an $n_i \times q$ matrix of known covariates.

β_i is q dimensional vector of subject specific regression coefficients.

$\varepsilon_i \sim (0, \in_i)$, Often $\in_i = \alpha^2 \mathrm{I}_{n_i}$, this model describes the observed variability within subjects.

Stage 2: Describes between subject variability, that is, explains variability, in the subject specific regression coefficients using known covariates.

$$\beta_i = K_i\beta + b_i \tag{2.2}$$

K_i is a $q \times q$ matrix of known covariates.

β is a P dimensional vector of unknown regression parameter $b_i \sim \mathrm{N}(0, D)$.

Now, combining the two levels/ stages model (equation 3.1 and 3.2), we have:

$$\begin{cases} Y_i = Z_i\beta_i + \varepsilon_i \\ \beta_i = K_i\beta + b_i \end{cases} \Rightarrow \{ Y_i = Z_iK_i\beta + Z_ib_i + \varepsilon_i .$$

$$\text{Let } X_i = Z_iK_i, \text{ then } Y_i = X_i\beta + Z_ib_i + \varepsilon_i$$
$$\Rightarrow \begin{cases} b_i \sim N_q(0, D) \\ \varepsilon_i \sim N_{n_i}(0, \sum_i) \\ b_1, b_2, ---, b_n;\ \varepsilon_1, \varepsilon_2, ---, \varepsilon_n \text{ are independent} \end{cases} \tag{2.3}$$

Where:
y_i is the $n_i \times 1$ response vector for observations in the i^{th} group.
X_i is the $n_i \times P$ model matrix for the fixed effects for observations in group *i.*
β is the $P \times 1$ vector of fixed-effect coefficients.
b_i is the $q \times 1$ vector of random-effect coefficients for group *i.*
ε_i is the $n_i \times 1$ vector of errors for observations in group *i.*
D is the $q \times q$ covariance matrix for the random effects.
$\sum_i$ is the $n_i \times n_i$ covariance matrix for the errors in group *i.*

The above model can be rewritten as: The marginal density function of Y_i is then given by: $f(Y_i) = \int f(\frac{Y_i}{b_i}) f(b_i) db_i$, which can easily show the density function of n_i -dimensional normal distribution with mean vector $X_i\beta$ and with covariance matrix $V_i = Z_i\, D(Z_i)^t + \sum_i$. Hence, models makes very specific assumptions about the dependence of the mean structure and the covariance structure on the covariates X_i and Z_i respectively. It is therefore also called a Hierarchical model. A model for Y_i given b_i: $\frac{Y_i}{b_i} \sim \mathrm{N}(\mathrm{X_i}\beta_i + \mathrm{Z_i b_i})\mathrm{b_i} \sim \mathrm{N}(0,\ \mathrm{D})$. A model for b_i marginally, we have that Y_i is distributed as: $Y_i \sim N(X_i\beta,\ Z_iD(Z_i)^{'} + \sum_i)$. Note that, although the marginal model naturally follows from the hierarchical one, both models are not equivalent.

Exploring Data Analysis: Data exploring was extremely helpful as additional tool in the selection of appropriate models [12]. The aspects of the data that should include: Individual Profiles, Exploring the Mean Structure, Exploring the Variance Structure, Exploring the Random Effects and Exploring the Correlation Structure were used for this study.

Methods of Parameter Estimation

In linear mixed effect model estimation of random effects and covariance structure of the random error is necessary besides to the fixed effect. Both the maximum likelihood (ML) and restricted maximum likelihood (REML) are the two common methods for parameter estimation in this study. These methods are based on maximizing the marginal likelihood function, which is a mathematical expression that describes the joint probability of obtaining the data expressed as a function of the parameter estimates. The algorithm for parameter estimation is usually done using a Newton-Raphson based procedure [11].

Maximum Likelihood (ML) Estimation: ML and Restricted Maximum Likelihood (REML) methods were used to estimate D *and* Σ. Let $V = Var(y) = \Sigma + ZDZ^{'}$. ML is the processes of finding the value of the parameter which maximize the likelihood function for a given data set. ML provides unbiased estimators under normal errors. The log-Likelihood function for observed y is:

$$ML: L(D, \Sigma) = -\frac{1}{2}\log|V| - \frac{1}{2}(m)^t V^{-1} m - \frac{n}{2}\log(2\Pi); where\, m = y - x(x^{'}V^{-1}x)^{-1}x^{'}V^{-1}y. \tag{2.4}$$

Restricted Maximum Likelihood Estimation: The REML estimation method applies ML estimation techniques to the likelihood function. i.e., REML is the processes of applying ML to the linearly transformed response data vector. We consider the estimation of σ^2 for the GLM. The MLE of σ^2 is $\hat{\sigma}^2 = \frac{(y-x)'(y-x\hat{\beta})}{n}$; where $\hat{\beta} = (x'x)^{-1}x'y$. The REML estimate of σ^2 is the minimum variance unbiased estimator $\tilde{\sigma}^2 = \frac{(y-x\hat{\beta})'(y-x\hat{\beta})}{n-p}$. The bias of the MLE is: $E(\tilde{\sigma}^2 - \hat{\sigma}^2) = -\sigma^2 \frac{p}{n}$, As P increase, bias of the MLE was negative and worsens. Let A be $n \times (n\text{-}p)$ matrix such that $A'A = I_{n-p}$ and $A'A = I_n - P_x$; where $P_x = x(x'x)^{-1}x'$; is the orthogonal projection matrix on to the column space of x. We can show $W=A'y$ is a vector of $n - p$ linearly independent error contrasts. The expected value of W is $E(A'y) = (I_n - p_x)x\beta = x\beta - x\beta = 0_{n-p}$. Each element of $A'y$ is the error contras $y \sim N_n(x\beta, V)$: Where, $V = ZDZ' + \Sigma$, *then* $W \sim N_{n-p}(0_{n-p}, A'VA)$. The REML approach applies ML estimator techniques to $W = A'y$ rather than y. The log-likelihood function is:

$$REML: L(D,\Sigma) = -\frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X| - \frac{1}{2}m'V^{-1}m - \frac{n-p}{2}\log(2\pi) \qquad (2.5)$$

Where, $m = y - x(x'V^{-1}x)^{-1}x'V^{-1}y$ and p is the rank of x estimating fixed effect (β) and random effect (b) parameters in the mixed model once getting estimates of D and Σ, where are denoted by, $\hat{D}$ and $\hat{\Sigma}$ respectively.

If $\hat{D}$ is nonsingular, we solve mixed model equations [13].

$$\begin{pmatrix} X'\hat{\Sigma}^{-1}X & X'\hat{\Sigma}^{-1}Z \\ Z^{-1}\hat{\Sigma}^{-1}X & Z'\hat{\Sigma}^{-1}Z + \hat{D}^{-1} \end{pmatrix}\begin{pmatrix} \hat{\beta} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} X'\hat{\Sigma}^{-1}y \\ Z'\hat{\Sigma}^{-1}y \end{pmatrix}$$

Then the solution is: $\hat{\beta} = (X\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y \; and \; \hat{b} = \hat{D}Z'\hat{\Sigma}^{-1}(y - x\hat{\beta})$

If $\hat{D}$ is singular, then the mixed model equations are modified [14] as follows:

$$\begin{pmatrix} X'\hat{\Sigma}^{-1}X & X'\hat{\Sigma}^{-1}Z\hat{L} \\ \hat{L}Z'\hat{\Sigma}^{-1}X & \hat{L}Z'\hat{\Sigma}^{-1}Z\hat{L} + I \end{pmatrix}\begin{pmatrix} \hat{\beta} \\ \hat{\tau} \end{pmatrix} = \begin{pmatrix} X'\hat{\Sigma}^{-1}y \\ \hat{L}Z'\hat{\Sigma}^{-1}y \end{pmatrix}$$

Where $\hat{L}$ is the lower-triangular Cholesky root of $\bar{D}$, satisfying $\hat{D} = \hat{L}\hat{L}$. Both $\hat{\tau}$ and a generalized inverse of the left hand side coefficient matrix are then transformed using $\hat{L}$ to determine $\hat{b}$. From this

$$Var(\hat{\beta}) = (X\hat{V}^{-1}X)^{-1}.$$

Results and Discussion

Results

Descriptive Statistics: Among the total of 475 children BMI under age five the minimum age was 0.17 month (5.1 day new born babies), maximum age was 5.00 years, and the average mean of these children was 2.0096 and the standard deviation of 1.16726. MUAC status had the mean of 12.197 and the standard deviation of 0.9015. In similar fashion, the amount of pump net (sachet) consumed by children was the mean of 36.78 Kcal and its standard deviation was 13.74 Kcal. As we know that the smallest standard deviation told us the preciousness of our research (Table 1).

Variables	Minimum	Maximum	Mean	Std. Deviation
Age	.17	5.00	2.0096	1.16726
MUAC status	4.2	14.6	12.197	.9015
Plump net	15	90	36.78	13.740

Table 1: Descriptive statistics for continuous covariates

Of the total 475 number of children, at the first time of measurement cough status indicates that 41.46% of children were coughed and 58.95% of children were non-coughed responses recorded in the BMI data. The covariate had 34.81% of children were have diarrhea and the remaining 65.19% of children were haven't diarrhea. Moreover, 51.11% of children have fever and 48.89% haven't fever in the first four visits. Similarly, ART status of children who had no ART were 96.58%, pre-ART were 1.01% and on ART were 2.41%. In addition, 54.73% of the children were living in urban area and the remaining 45.27% were living in rural area. The number of female children was relatively higher than that of males (Table 2).

	Sex		Cough		Diarrhea		Fever		Residence		ART		
Visits	**Male**	**Female**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Urban**	**Rural**	**On**	**Pre**	**No**
Visit 1	202	295	204	293	173	324	254	243	272	225	12	5	480
(%)	40.64	59.36	41.46	58.95	34.81	65.19	51.11	48.89	54.73	45.27	2.41	1.01	96.58
Visit 2	202	295	204	293	173	324	254	243	272	225	12	5	480
(%)	40.64	59.36	41.46	58.95	34.81	65.19	51.11	48.89	54.73	45.27	2.41	1.01	96.58
Visit 3	202	295	204	293	173	324	254	243	272	225	12	5	480
(%)	40.64	59.36	41.46	58.95	34.81	65.19	51.11	48.89	54.73	45.27	2.41	1.01	96.58
Visit 4	202	295	204	293	173	324	254	243	272	225	12	5	480
(%)	40.64	59.36	41.05	58.95	34.81	65.19	51.11	48.89	54.73	45.27	2.41	1.01	96.58

Table 2: Summary statistics of children at the first 4-visits on their covariates

The variable sex had 26.49% of males and 31.19% females were normal weight status; 69.93% of males and 65% of females were underweight status; where by 1.98% males and 2.88% females were overweight and 1.61% of males and 0.93 of females were obesity. Moreover, children were grouped according to their place of residence; 65.37% rural children and 68.35% urban children found to have underweight; while 2.66% rural children and 2.39% urban children were with overweight and 0.89% of rural children and 1.47% of urban children were with obesity (Table 3). Similarly, when children grouped according to cough; 66.3% of children coughed and 67.49% of children non-coughed were underweight; while 2.33% of children coughed and 2.65% non-coughed children were overweight and 1.72% of children coughed and 0.85% of children were obesity. Although, when children were grouped according to diarrhea status; 69.22% of children have diarrhea and 65.82% of children haven't diarrhea were underweight status, 2.31% of children have diarrhea and 2.62% of children haven't diarrhea were overweight, where by 0.72% of children have diarrhea children and 1.47% of children haven't diarrhea were obesity. In similar fashion other covariates have the same fashion of interpretation (Table 3).

	Sex		Cough		Diarrhea		Fever		Residence		ART		
BMI	**Male**	**Female**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Urban**	**Rural**	**On**	**Pre**	**No**
Under	565	767	541	791	479	853	667	665	743	589	33	18	1281
(%)	69.93	65	66.3	67.49	69.22	65.82	65.59	68.79	68.35	65.37	68.75	90	66.72
Normal	214	368	242	340	192	390	306	276	302	280	12	2	568
(%)	26.49	31.19	29.66	29.01	27.75	30.09	30.09	28.42	27.78	31.08	25	10	29.58
Over	16	34	19	31	16	34	29	21	26	24	3	0	47
(%)	1.98	2.88	2.33	2.65	2.31	2.62	2.85	2.16	2.39	2.66	6.25	0	2.45
Obesity	13	11	14	10	5	19	15	9	16	8	0	0	24
(%)	1.61	0.93	1.72	0.85	0.72	1.47	1.47	0.93	1.47	0.89	0	0	1.25
Sub total	808	1180	816	1172	692	1296	1017	971	1087	901	48	20	1920
(%)	40.64	59.36	41.05	58.95	34.81	65.19	51.16	48.84	54.68	45.32	2.41	1.01	96.58

Table 3: Assessing health status of children BMI measured by different Covariates

Based on WHO criteria from 475 number of children examined around 29.28% were normal weight status, 67% were under weight, 2.52% were overweight and only 1.21% were obesity. As we observed from Table 4: if the time of treatment for children increased from time 1 to time 4, then the number of children who had normal weight status were increased over time and the number of children who had underweight status were decreased over time. Generally, this study seems to have a problem of underweight than overweight and obesity in percentage. Mean of children BMI tends to increase over time and similarly variance of children BMI tends to increase over time (Table 4).

Time	**Health Status**				**BMI**	
	Normal W.	**Obesity**	**Over W.**	**Under W.**	**Mean**	**Variance**
Time1	24	0	2	471	14	6.94
(%)	4.12	0	4	35.36		
Time2	179	1	6	311	17.19	11.53
(%)	30.76	4.17	12	23.35		
Time3	186	5	18	288	17.82	14.88
(%)	31.96	20.83	36	21.22		
Time4	193	18	24	262	18.81	24.69

Time	Health Status				BMI	
	Normal W.	Obesity	Over W.	Under W.	Mean	Variance
(%)	33.16	75	48	19.67		
Sub total	582	24	50	1332		
(%)	29.28	1.21	2.52	67		

Table 4: Assessing health status of children BMI measured over time

Explanatory Data Analysis

Individual Profile Plots: The individual profiles can also provide some information between children BMI variability and illustrate that there is change among children BMI over time. Correspondingly, it appears that most of the children are gaining BMI over time and the variability of the children BMI seems smaller at the beginning compared to the end (Figure 1).

Exploring Mean Structure: The mean profile plot gives clue to determine the type of time effect on BMI and we observed that the time seems to have almost linear effect on BMI (Figure 2).

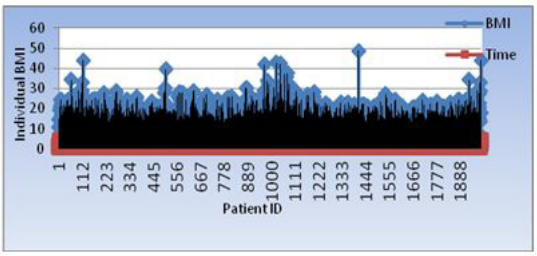

Figure 1: Individual profile plot

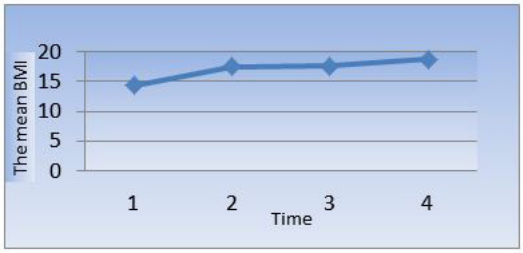

Figure 2: Mean profile plots

Exploring Variance Structure: Plot of variance appears that the observed variance was not constant through time evolution and seems variability of children BMI tends to increase over time (Figure 3a). The variability of BMI for males looked higher than that of females through time. The slope of female children BMI had higher rate than that of male children through time evolution (Figure 3b).

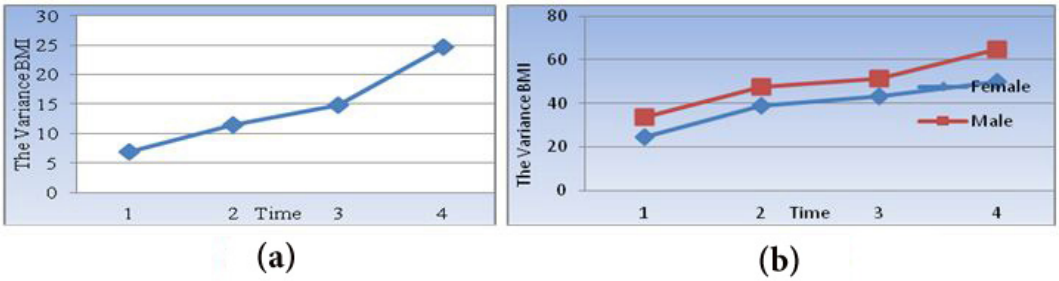

Figure 3: Observed variance and variability of children BMI by sex

Checking Assumption for the Final Model: Individual specific residual plots for fitted model designates that the residuals are centered at zero. That is, $E(\varepsilon_{ij})=0$, but the variability changes with group (Appendix A). Since there were only four observations per individual, we couldn't rely too much on the individual box plots for inference about the within-group variances. Examining the plot of the standardized residuals versus fitted values by gender appears to show that the variability in BMI measurements were slightly greater among males than among females (Appendix B). However, within each gender the variability was somewhat constant which might imply heteroscesdasticity. The adequacy of the heteroscedastic fit has been assessed by examining plots of the standardized residuals versus the fitted values by sex. The standardized residuals in each sex now have about the same variability, but not exactly the same and the qq normal plot had shown that random errors were approximately normally distributed and symmetric with respect to zero. Therefore, the normality assumption was shown (Appendix C).

Assessing the Assumption on the Random Effects: Basically, qq-norm normal plot of estimated random effects were used for checking marginal normality and identifying outliers whereas pairs scatter plot matrix of the estimated random effects were used for identifying outliers and checking the assumption of homogeneity of the random effects covariance matrix (Appendix D and F). The heteroscedastic model accommodates the impact of the outlying observations in the within-group variances estimation and this accommodation reduces the estimated between group variability, thus increasing the degree of shrinkage in the random effects estimates. Box plots of Sex for heterosedastic model did not suggest any departures from the assumption of homogeneity of the random effects distribution (Appendix E).

Linear Mixed-Effects Model: Mean profile plots have suggested that time has nearly linear effect on BMI progression over time (Figure 2). Hence, Linear Mixed Effects Model with linear time effect was fitted as:

$$BMI_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 A_i + \beta_3 Sa_i + \beta_4 M_i + \beta_5 S_i + \beta_6 C_i + \beta_7 D_i + \beta_8 F_i + \beta_9 AR_i + \beta_{10} R_i + \beta_{11} P_i * T_{ij} + \beta_{12} S_i * T_{ij} + \beta_{13} P_i * T_{ij} * S_i + \beta_{14} M_i * T_{ij} + \beta_{15} A_i * T_{ij} + \beta_{16} D_i * R_i + b_{i0} + b_{i1} + \varepsilon_{ij}.$$

Where, BMI_{ij} = Body Mass Index on i^{th} children on j^{th} measurement
T_{ij} =Time at which i^{th} children on j^{th} measurement, $i = 1, ---, 1900$ *and* $j = 1, 2, 3, 4.$

A_i =Age at i^{th} children
Sa_i=Amount of sachet given for i^{th} children
M_i =MUAC status for i^{th} children
S_i =Sex of i^{th} children (Male, Female)
C_i =Cough status i^{th} children (Yes, No)
D_i =Diarrhea status i^{th} children (Yes, No)
F_i =Fever status i^{th} children (Yes, No)
AR_i =ART treatment i^{th} children (On ART, on pre-ART, No ART)
R_i =Residence of i^{th} children (Urban, Rural)
$Sa_i * T_{ij}$, $S_i * T_{ij}$, $M_i * T_{ij}$, $A_i * T_{ij}$ =Interaction terms with time
$Sa_i * T_{ij} * S_i$ =Interaction terms between Sachet and sex with time
$D_i * R_i * T_{ij}$ =Interaction between diarrhea and residence over time
β_0 =Over all intercepts
$\beta_1, ---, \beta_{16}$ =Coefficients of fixed effect
b_{i0} =Intercept of random effect part
b_{i1} =Coefficient of random time effect
ϵ_{ij} =Random error term

Random Effect Term Selection: To select random effect to the model with only intercept, with only slope and with intercept and slope have been fitted and compared. An appropriate random effect to the model was selected by using likelihood ratio test. The small p-value indicates that, we can reject model 2 in favor of model 1; we prefer the more parsimonious first model. This conclusion is consistent with the AIC and the BIC values for which smaller value is considered as better. That is, the AIC information criterion increased from 9504.7 to 9640.8, which indicates that model with intercept and slope (both) was a better fitted model (Table 5).

	Model	df	AIC	BIC	loglik test	L.ratio	p-value
Both	1	19	9504.7	9505.1	9464.7	847.76	<0.0001
Slope	2	19	9683	9685.3	9607.7		
Intercept	3	19	9640.8	9685.4	9640.8		

Table 5: Selection of best random effects based on intercept, slope and both cases

Model Selection: To select the best model is not possible using only the best mean and variance structure, but also correlation structure. In order to select best variance covariance structure for the final model, first deals with the variance structure by using different variance functions. The combined result was equivalent to dealing with variance covariance structure.

Selecting Variance Function: The variance functions were used to model the variance structure of the within group errors using covariates. The primary tool for investigating within-group Heteroscesdasticity was plots of residuals against the fitted values. Independent variance type of variance functions was used by default for this study.

Selecting Correlation Structure Function: The correlation functions were used to model dependence among observations. Among different correlation structure classes/ functions in this study unstructured, compound symmetry, Toeplitz and autoregressive [1]. covariance models were used and compared. An information criterion was important to select a responsible covariance structure in order to obtain valid inferences for fixed effects. The small AIC value and the corresponding significant p-value indicated that the model with unstructured covariance function is preferable (Table 6).

Fits	Model	Df	AIC	BIC	loglik	Test	L.Ratio	P-value
Model. AR(1)	1	19	10005.3	9997.3	9998.3			
Model. CS	2	19	10106.4	10094.4	10001.6			
Model. Toep	3	19	10013.7	10013.7	9993.6			
Model. UN	4	19	9504.7	9505.1	9464.7	4Vs3	847.76	<0.0001

Table 6: Selection of information criteria to fit responsible structure

After selected the best information criteria based on the smallest AIC with its corresponding p- value, the next step is selection of the best fitted model from homoscedastic and heteroscedastic model. Hence, very small p-value corresponding to the likelihood

ratio statistics confirmed that the heteroscedastic model explained the data significantly better than the homoscedastic model (Table 7). The assumption of normality for the within-group errors were assessed with the normal probability plot of the residuals, produced by the qq normal method. Therefore, for the data set of these types of study heteroscedastic model with unstructured correlation was considered as best final model.

	Model	Df	AIC	BIC	loglik	Test	L.Ratio	P-value
Homoscedastic	1	18	9684.8	9685.4	9640.8			
Heteroscedastic	2	19	9504.7	9505.1	9464.7	1Vs2	84.76	<0.0001

Table 7: Selection of best fitted model from homoscedastic and heteroscedastic model

Random Effect with positive intercept indicates an increase in BMI of each child provided that time is included in the model. Intra-class correlation coefficient (ICC) = $\frac{5.4086}{3.1071+5.4086} = 0.635,$ indicates that the random effects affects the children BMI data by 63.5%. The residual term indicates that variation with in children in different time of measurements. The term labeled residual is the estimate of σ^2 Additionally, random effect with positive intercepts indicates that an increase in BMI each child provided that time was included in the model (Table 8).

$$\omega = V\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} = \begin{bmatrix} \omega_0^2 & \omega_{01} \\ \omega_{01} & \omega_1^2 \end{bmatrix} = \begin{bmatrix} 7.6189 & 2.1994 \\ 2.1994 & 5.4086 \end{bmatrix}$$

Random effects	S.Dev	Corr
Intercept	7.6189	ICC
Time	5.4086	0.635
Residual	3.1071	

Table 8: Estimates of random effects

The final fitted model was suggested as:

$$BMI_{ij} = 15.514 - 4.085T_{ij} - 0.048A_i + 0.226Sa_i - 0.216M_i + 0.311SF_i + 0.895DNO_i - 2.315ARON_i$$
$$-0.047Sa * T_{ij} + 0.009T_{ij} * SF_i * Sa + 0.458T_{ij} * M_i - 0.009T_{ij} * A_i - 1.816D_i * R_i * T_{ij} + b_{i0} + b_{i1}T_{ij} + \varepsilon_{ij}.$$

The intercept coefficient, $\hat{\beta}_0$ =15.514 represents an estimate of the average level of children BMI for all covariates. Coefficient of age is $\hat{\beta}_A$ =-0.048-0.009=-0.057 indicates that children in the reference group of mean of children BMI decreased by 0.057 units per month. Coefficient of Sachet is $\hat{\beta}_{Sa}$ =0.226-0.047=0.179 indicates that the children in the reference group of the mean children BMI increased by 0.179 units per month. $\hat{\beta}_T$ = -4.085 is the rate of change of the average children BMI for a unit change of time by considering the other variables constant. Likewise, $\hat{\beta}_{DNO}$ =0.895 is the average difference of children BMI between having diarrhea disease cases and non-diarrhea. $\hat{\beta}_{on\text{ART}}$ = -2.315 is the average difference of children BMI between on ART and non-ART treatment. The rate of change of average children BMI difference between female who eat sachet and male who eat sachet for a unit change of time by considering the other variables constant is $\hat{\beta}_{S*Sa*T}$ =0.009. The rate of change of average children BMI difference between no responses who live in rural and positive responses who live in urban through time by considering the other variables constant is $\hat{\beta}_{DNO*Rrural*T}$ =-1.816 (Table 9).

Effect	Estimate	St.Error	DF	t-value	P>|t|	95 % Conf.Interval
Intercept	15.5140	1.8699	474	8.30	<.0001*	(13.146, 17.039)
Sex Female(SF)	0.3110	0.6895	474	0.45	0.6522	(-0.367, 0.541)
Cough No(CNO)	0.6821	0.6391	474	1.07	0.2864	(-0.7534, 0.7021)
Diarrhea No(CNO)	0.8954	1.3573	474	-0.66	0.0398*	(0.7009, 0.9902)
Fever No(FNO)	-0.1092	0.6350	474	-0.17	0.8636	(-0.2093, 0.2609)
ART on(on AR)	-2.3145	0.9994	474	-2.32	0.0210*	(-3.0781, -2.0901)
ART pre(Pre-AR)	-2.2646	2.1618	474	-1.05	0.2954	(-2.8809, 2.1609)
RES Rural(Rrural)	-1.4139	0.7669	474	1.84	0.0659	(-1.8970, 2.0912)
Age(A)	-0.04837	0.009640	1479	-5.02	<.0001*	(-0.1408, -0.0672)
MUAC(M)	-0.2159	0.1204	1479	-1.79	0.0732	(-0.3098, 0.0983)
Sachet(Sa)	0.2255	0.01503	1479	15.00	<.0001*	(0.1278,0.3009)

Effect	Estimate	St.Error	DF	t-value	P>|t|	95 % Conf.Interval
Time(T)	-4.0847	0.5636	1479	-7.25	<.0001*	(-4.2970, -3.9982)
Sa*T	-0.04732	0.006584	1479	-7.19	<.0001*	(-0.0509, -0.0398)
T*SF	-0.3207	0.2968	1479	-1.08	0.2800	(-0.3612, 0.0125)
M*T	0.4576	0.04125	1479	11.09	<.0001*	(0.4011, 0.5431)
A*T	-0.00913	0.004497	1479	-2.03	0.0424*	(-0.01023, -0.00601)
SF*Sa*T	0.00948	0.006073	1479	1.56	0.0118*	(0.008012, 0.01030)
D*T*Rrural	-1.8157	0.6456	1479	-2.81	0.0051*	(-1.9001, -1.7901)

Table 9: Linear Mixed Effects model with main and interaction effect

- Statistically significant at 95% level of confidence denoted by star sign (*)
- The remaining category of each covariates are reference group

The type-III tests of hypotheses deals about a nice method to decide the final significance covariates in the model. The results was retained the significance levels, but it is based on by considering more data. Generally, ART, residence, diarrhea, age, Sachet, time, Sachet over time, MUAC over time, age over time, interaction between Sachet and sex over time and diarrhea and residence over time were statistically significance effect on average BMI of children. The remaining covariates were statistically insignificant. However, there is the variation within the subjects. Therefore, some of the random slops are statistically significance within subjects (Table 10).

Effect	NumDF	DenDF	F-Value	Pr>|F|
Sex	1	474	0.19	0.6655
Cough	1	474	0.03	0.8703
Diarrhea	1	474	1.23	0.0287*
Fever	1	474	0.19	0.6597
ART	2	474	3.09	0.0463*
Residence	1	474	8.05	0.0048*
Age	1	1479	25.18	<.0001*
MUAC	1	1479	3.21	0.0732
Sachet	1	1479	225.14	<.0001*
Time	1	1479	66.37	<.0001*
Sachet*Time	1	1479	62.45	<.0001*
Sex*Time	1	1479	1.17	0.2800
MUAC*Time	1	1479	123.07	<.0001*
Age*Time	1	1479	4.12	0.0424*
Sachet*Time*Sex	1	1479	2.44	0.0128*
Diarrhea*time*Residence	1	1479	4.53	0.0338*

Table 10: Linear Mixed Effects model type-III tests of hypotheses

Discussion

The results of the analysis presented in this study was modeling of longitudinal factors on children BMI to identify factors that are statistically significantly associated with children health status depend on the criteria of WHO. The results can be useful in an improving the series problem of abnormal health status of children BMI. Sachet, diarrhea, ART, residence and having healthy MUAC status have been established as contributing factors for higher rate of children BMI. In developing country of the world like Ethiopia such as poorly developed, children BMI especially underweight remained to be a big challenge in public health. To address these issues different stakeholders at international, national and regional levels have been implementing different strategies.

In this study the factors: amount of Sachet(p-value<0.0001: $\hat{\beta}_{Sa}$ =0.2255), visiting time(p-value<0.0001: $\hat{\beta}_{\text{time}}$ =-4.0847), age(p-value=<0.0001: $\hat{\beta}_{\text{age}}$ =-0.04837), ART (p-value=0.0463: $\hat{\beta}_{\text{on ART}}$ =-2.3145, $\hat{\beta}_{\text{pre-ART}}$ =-2.2646), residence(p-value=0.0048: $\hat{\beta}_{\text{res}}$ =-1.4139), diarrhea(p-value=0.0287: $\hat{\beta}_{\text{diarrhea}}$ =0.8954), Sachet over time(p-value<0.0001: $\hat{\beta}_{\text{Sa*T}}$ =-0.04732), age over time(p-value=0.0424: $\hat{\beta}_{\text{A*T}}$ =-0.00913), MUAC over time(p-value<0.0001: $\hat{\beta}_{\text{MUAC*T}}$ =0.4576), diarrhea and residence through time(p-value=0.0338: $\hat{\beta}_{\text{D*res}}$ =-1.8157) and interaction between Sachet and sex over time(p-value<0.0001: $\hat{\beta}_{\text{Sa*S*T}}$ =0.009482) were statistically significantly associated with children BMI. Prior research based on rural longitudinal survey in four provinces of Pakistan showed that in

households having incomes at the lowest quintile, both males and females had lower BMIs than those belonging to the highest quintile income groups in all provinces except Baluchistan (Garcia & Alderman, 1989). Consistency in the relationship between BMI and an independently assessed measure of socio-economic status was also seen in communities in India, Ethiopia and Zimbabwe despite there being a relatively small range of BMIs among the rural populations of these countries [15]. This study, for particular area of Bahir Dar Districts, did support the prior research results related to rural area.

Childhood overweight and obesity status tend to share a relationship with household food insecurity. This relationship was first raised in 1995 when Dietz investigated a case about a young, 7 year-old African-American girl who belonged to a household with limited amounts of food during certain points each month and the girl's mother would feed her inexpensive, high fat/ high caloric foods in order to compensate for the lack of resources to costlier, more nutritious foods [16]. Dietz then speculated that this act may be due to an "adaptive response to food shortages when the family lacked financial resources where in turn there was an increase in the consumption of low-cost, high fat foods resulting in increased body mass" (1995, p766). Similarly, Sachet is one types of dietary so, in this study if the amount of Sachet taken by children increase the children BMI also increased.

According to Haramaya study that show the prevalence of underweight on children nutrition were 36.4% among the 8200 cases examined were 38% of male and 34.8% of female children were underweight [17]. In this study the sever problem is under weight, so we can make comparison with the previous researcher result. The present study deals about 67% of the children were underweight among the 497 per individual cases examined were 65% of male and 69.93% of female children were underweight. The study conducted in Addis Ababa, the prevalence of associated factors among high school adolescents students in Arada sub city were 72.1% normal-weight, 18.5% underweight, 8.6% overweight and 0.8% obesity [18]. In the present study, 29.28% of the children were normal health status while 67% were underweight, 2.52% were overweight and 1.21% was obesity in Bahir Dar Districts.

Conclusion

Despite the progress that has been made in the country to improve the problem of children BMI remain high at Bahir Dar districts. According to this study, there were statistically significant difference among children BMI variation with respect to time, Sachet, age, residence, ART and diarrhea. While; fever, cough, MUAC and sex were statistically insignificant effect on children BMI. In the profile analysis, the mean evolution and variability of BMI were higher on average in females than males. The findings further suggest that if the amount of Sachet were taken by children increase, so the children BMI also increased. Equally, the interaction between Sachet and female over time had greater BMI than that of the interaction between Sachet and male over time. Additionally, children BMI were higher in average haven't diarrhea than have diarrhea through time. Finally, children who live in rural area had lower body mass index than that of who live in urban area. Even if, children who were on ART had lower BMI than that of who were no ART and children who were pre-ART had similarly lower BMI that of who were no ART children.

Acknowledgement

I would like grateful to thank Amhara regional state health office and Bahir Dar health administrative office for permission to use the children body mass index (secondary data) set for this study. It's also my great pleasure to thank Bahir Dar University for providing me the financial support for this study.

Appendixe

References

1. Adolphe Q (1796-1874) BMI formula was developed. Belgium Statistician, USA.

2. WHO (2013) Media Center fact sheet. Obesity and overweight, Geneva, Switzerland.

3. WHO (2012) Info base data on overweight and obesity, mean BMI, healthy diets and physical inactivity, Geneva, Switzerland.

4. Freedman DS, Dietz WH, Srinivasan SR, Berenson GS (1999) The relation of overweight to cardiovascular risk factors among children and adolescents: The Bogalusa Heart study. Pediatrics 103: 1175-82.

5. Caleyachetty R, Richards N (2012) Prevalence of overweight, obesity and thinness in 9-10 year old children in Mauritius. Global Health 8: 28.

6.Salman Z, krik GD,Deboer MD (2011) High Rate of Obesity-Associated Hypertension among Primary Schoolchildren in Sudan. Int J Hypertens 5: 169-73.

7. Alebachew Z (2009) Prevalence of childhood and adolescent overweight and obesity among elementary school students in Addis Ababa: Double burden of malnutrition in Ethiopia. Addis Ababa University digital library.

8. Davis CS (2002) Statistical Methods for the Analysis of Repeated Measurements. Springer-Verlag, New York.

9. Singer JD (1998) Using SAS Proc Mixed to fit multilevel models, hierarchical models, and individual growth models. J Edu Behv Stat 23: 323-55.

10. Laird NM,Ware JH (1982) Random-effects models for longitudinal data. Biometrika 38: 963-74.

11. Verbeke G (2009) Linear Mixed Models for Longitudinal Data. Springer, New York.

12. Diggle PJ (2002) Analysis of Longitudinal Data, 2nd Edn. Oxford University Press, New York.

13. Lyle C Gurrin et al. (2005) Solving mixed model equations. John Wiley and Sons, New York.

14. Henderson CR (1984) Applications of Linear Models in Animal Breeding. Guelph Canada: University of Guelph Press.

15. Ferro-Luzzi A, Sette S, Franklin M, James WP (1992) A simplified approach of assessing adult chronic deficiency. Eur J Clin Nutr 46: 173-86.

16. Dietz W (1995) Does Hunger Cause Obesity? Pediatrics 95: 766-7.

17. Kassahun T (2013) Semi-parametric analysis of children nutrition status in Ethiopia. Intern J Statist Appli 3: 141-54.

18. Alemu E, Atnafu A, Yitayal M, Yimam K (2014) Prevalence of Overweight and/or Obesity and Associated Factors among High School Adolescents in Arada Sub city, Addis Ababa, Ethiopia. J Nutr Food Sci 4: 261.