

Longitudinal modeling of fasting blood sugar variation over time among adult diabetic patients in case of Adama hospital medical college

Abdulmenan Mohammed Abdulahi

Aksum University

Aragaw Eshetie Aguade

University Of Gondar

Hunachew Kibret Yohannis (✉ huneastat@gmail.com)

University Of Gondar

Research Article

Keywords: Diabetes mellitus, fasting blood sugar, linear mixed model, longitudinal data analysis

Posted Date: April 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-253888/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Longitudinal modeling of fasting blood sugar variation over time among adult diabetic patients in case of Adama hospital medical college

Abdulmenan Mohammed Abdulahi¹ Statistics Department, College of Natural and Computational Science, Aksum University, Aksum, Ethiopia.

Aragaw Eshetie Aguade² (Ph.D.) Statistics Department, College of Natural and Computational Science, University of Gondar, Gondar, Ethiopia.

Hunachew Kibret Yohannis^{2*} Statistics Department, College of Natural and Computational Science, University of Gondar, Gondar, Ethiopia.

Email addresses:

Abdulmenan Mohammed : abdike2435@gmail.com

Aragaw Eshetie: aragaw2018@gmail.com

Hunachew Kibret Yohannis: huneastat@gmail.com

Declarations

Ethics approval and consent to participate: All methods were carried out in accordance with relevant guidelines and regulations. Informed consent letter obtained from the University of Gondar College of Natural and Computational Science ethical committee and the committee has approved the study. The Adama Hospital Medical College ethical committee directly approved and waived the need of informed consent of the University of Gondar ethical committee letter for the present study due to secondary data analysis.

Consent for publication: Not applicable

Availability of data and materials: The datasets generated and/or analyzed during the current study are not publicly available due to the data is using for another analysis but are available from the corresponding author on reasonable request.

Competing interests: All authors declare no competing interest

Funding: Not applicable

Authors' contributions: A. M. A. contributed in the conceptualization of the research problem, study design, analysis of the data and interpretation of the final result; A. E. A. participated in, revision of the research, data analysis and editing of the manuscript and H. K. Y. contributed in guidance, consultation and continued follow up and encouragement from the beginning to the end of the study, revision of

the thesis and formulate the manuscript. Each and every authors of the paper carefully read, edited and finally approved the final manuscript.

Acknowledgement: We would like to thank Adama Hospital Medical College for providing us the required data for the completion of our work.

Abstract

Background: Diabetes Mellitus (DM) is a chronic, progressive disease characterized by elevated levels of blood glucose. Despite the fact that most international association/organization gave attention toward diabetes control and prevention by healthy professional, still diabetes and its complication such as cardio vascular, blood vessels, eyes, kidneys and nerves become a major cause of premature death and disability across the world. The overall aim of this study is assessing fasting blood sugar variation over time and its determinant among diabetic patients.

Methods: Data were obtained from Adama Hospital Medical College diabetic patients who have been active in the follow-up treatment from September 1, 2018 to August 30, 2019. The data consists of basic demographic and clinical characteristics of 312 DM patients were selected using simple random sampling techniques and of whom 177 were males and the rest 135 were females. The linear mixed effect model for longitudinal data analysis was used by taking the correlation between Fasting Blood Sugar (FBS) level of patients into account. Linear mixed model, random intercept and slope models were used for feting the data.

Results: The results from the linear mixed model with unstructured co-variance structure showed that for one-month change in time decreases log FBS level by 0.0111267 mg/dl. While a unit increase in Body Mass Index (BMI) of a patient on treatment, the log FBS level was increased by 0.0434 mg/dl. Similarly, a unit increase in Diastolic Blood Pressure (DBP), the log FBS level was increased by 0.0004749 mg/dl. Being tertiary and secondary level of education decreases logFBS level by 0.0058844 and 0.0055161 respectively compared with patients with no education.

Conclusion: Age, Educational status, Dietary type, Drug type, History of hypertension, BMI, DBP, Time, interaction effect of Age, history of hypertension, Dietary type, other comorbidity at baseline with time were the significant determinant for the change in mean FBS level of the diabetes patients over time. Based on the findings of our study and WHO recommendation, maintaining of healthy body

weight, by taking healthy diet along with lower blood glucose level is essential to control blood sugar in body and to prevent long term complication.

Key words: Diabetes mellitus, fasting blood sugar, linear mixed model, longitudinal data analysis.

List of Acronyms

AIC:	Akaike's Information Criteria
BIC:	Bayesian Information Criteria
DM:	Diabetes Mellitus
FBS:	Fasting Blood Sugar
IDFA:	International Diabetes Federation Atlas
ADA:	American Diabetes Association
WHO:	World Health Organization
LMM:	Linear Mixed-Effect Model
MAR:	Missing at Random
MCAR:	Missing Completely at Random
MNAR:	Missing Not at Random
MI:	Multiple Imputation
REML:	Restricted maximum likelihood
A1c:	Glycated Hemoglobin
OGTT:	Oral Glucose Tolerance Test

Background

Diabetes mellitus is a serious, chronic disease that occurs either when the pancreas does not produce enough insulin (a hormone that regulates blood glucose), or when the body cannot effectively use the insulin it produces. DM is a chronic, progressive disease characterized by elevated levels of blood glucose. Diabetes can lead to serious damage in many parts of the body such as heart, blood vessels, eyes, kidneys and

nerves and can increase the overall risk of dying prematurely (WHO, 2018; Nazir et al, 2018). Diabetes mellitus may exist with characteristic symptoms include excessive exertion of urine, thirst, constant hunger, weight loss, vision changes and fatigue. Early diagnosis can be performed through relatively in expensive testing of blood sugar (Ogurtsova et al, 2017).

There are three major categories of diabetes such as type I diabetes, type II diabetes and gestational diabetes based on etiology and cause of disease. Type I diabetes is happened when the body failed to produce the insulin required and type II diabetes is occurred when the body produce high level of blood sugar or glucose resulting from defects of insulin production or ineffective insulin. The gestational diabetes is a type of diabetes that occur during pregnancy (WHO, 2006). This study focused on the two type I and type II diabetic categories. The number of people with diabetes has steadily risen over the past few decades, due to population growth, the increase in the average age of the population, and the rise in prevalence of diabetes at each age (WHO, 2018).

Now a days there are 463 million people between 20 and 79 years of old living with diabetes world wide (Roglic, 2016). In 2019, 50% or 231.9 million of the 463 million adults living with diabetes, (overwhelmingly type 2 diabetes) are they didn't know that they have the disease. About 79.4% of diabetic patients in the world they are living in low and middle income countries and most of them are within the age group of 40 up to 59 years. In 2019 there were an estimation that around 4.2 million adults between the age of 20 and 79 years were died because of diabetes. Which means in every eight seconds there is a death of individuals due to diabetes. Among all death causes in the world diabetes scored 11.3%. Almost half (46.2%) of deaths associated with diabetes are in people under the age of 60 years the working age group (Federation, 2019). South-East Asia and Western Pacific Regions hold the largest numbers of diabetes patients and it accounts approximately half the diabetes cases in the world (Megerssa et al, 2013; Patterson et al, 2019).

In Africa based on recent report, 19.4 million adult people between the ages of 20 and 79 were living with diabetes. In this region, 45.9% of people with diabetes live in low-income countries and 54.1% of diabetic people live with in middle-income countries. In Africa 8.8% of the prevalence of diabetes is occurred among adults of the

ages 65-69 years and 59.7% of people affected by diabetes are unaware of their condition (Federation, 2019).

In Ethiopia, the prevalence of people living with diabetes increases from time to time. According to Global estimate prevalence of diabetes, there were an increase of diabetes prevalence among adult's Ethiopian population was 4.4% in 2013 (WHO, 2018). The International Diabetes Federation Atlas (Federation) in 2019 reported that, there were 1.9 million diabetes cases of 20–79 years of age and 1.2 million of them were diabetic patients from rural areas on the same way 34,262 diabetes related deaths were occurred (Federation, 2019). According to the International Diabetes Federation (IDF) report, there were 2,652,129 cases of diabetes in 2017 making it first among top five countries of Africa having number of people living with diabetes within the age 18-99 years (Ogurtsova et al, 2017).

Essentially, the development of diagnostic and therapeutic for diabetes patients has led to increasing the survival rates and decrease Diabetes related death and its complications such as heart disease, stroke, retinopathy, loss vision, chronic kidney disease and hypertension by controlling the blood sugar. However, the complete prevention of complications is not possible (Mahmoudi, 2006). Most of previous study focus on the incidence and prevalence of the diabetic disease and so far used cross-sectional data. However, unable to account correlation between data that leads to bias in inference and parameter estimation. Longitudinal studies are appropriate for modeling fasting blood sugar variation and to identify associated risk factors over time taking the correlation of FBS level within a patient into account. Missing data was observed in FBS level studies; by applying statistical methods the missing value problems are handled. Even if few longitudinal studies were conducted on fasting blood sugar change for instance, study from Ghana (Adampah et al, 2015) had done on Mixed-Effects Model for Longitudinal Study of Type-2-Diabetes, but they didn't address sufficient baseline clinical/laboratory characteristics in the study that can significantly affect fasting blood sugar change over time in diabetes patients with appropriate variance co-variance structures which is core of linear mixed model. The main objective of this study was modeling fasting blood sugar variation over time among diabetic patients. The results of this study provide information for public health practitioners, the development of an effective Diabetic care and DM patient

monitoring systems who are working in the areas of giving care, support and treatment for patients.

Methods and Material

Study Area and Study Population Description

The study area of this research was diabetes clinic of Adama Hospital Medical College which is located 99 km to southeast Addis Ababa, Ethiopia. The target populations of this study was all newly diagnosed type I and II DM patients at Adama Hospital who had been active in follow-up treatment with three-month interval in the hospital from September 1, 2018 to August 30, 2019. The medical records of the diabetes patients within the specified study periods were extracted from patient's chart. All diabetic patients aged 18 years or older, who were coming to attend their treatment at Adama Hospital Medical college for their regular follow up during study periods and only Patients with at least two visits of longitudinal response were eligible for this study.

Sample Size and Sampling Procedure

The samples were selected using simple random sampling technique through sampling frame of the identification number of DM patients. The number of measurements were not equal for all patients due to the difference in the duration of the follow-up. Measurements of all patients were taken at baseline (time zero), 3, 6, 9, 12, 15, 18, 21 and 24 months which had an equal time interval with three months between all measurements. Sample size for longitudinal response is calculated by using Diggle formula (Diggle et al, 2002).

$$N = \frac{4(Z_{\alpha/2} + Z\beta)^2 \sigma^2 (1 + (m-1)\rho)}{md^2} = \frac{4(1.96 + 0.842)^2 12.72(1 + (9-1)0.5)}{9 * 0.8^2} = 312$$

Where, N is the total sample size, d is effect size sample (0.8), m is number of time points repeated measurement was taken, ρ is correlation between repeated measurements and σ^2 is variance of outcome variables all taken from previous study. In our case assuming significance level of 0.05, power of the study 0.8, $m = 9$, which is the number of time points repeated measurements was taken, $\rho = 0.5$ and effect size of 0.8, from study conducted in Jimma we have $\sigma^2 = 12.72$ (random

intercept model) (Aniley et al, 2019) Using $z_{\alpha/2} = 1.96$, $Z\beta = 0.842$ and inserting all quantities in the formula:

Study Variables

Outcome Variable

The outcome variable is continuous longitudinal repeated measurement per individuals, which is fasting blood sugar level measured in milligram per deciliter.

Independent Variables

The predictor variables which assumed to influence the fasting blood sugar of diabetes patients were age, sex, marital status, residence, educational status, exercise activity, frequency of meals, dietary type, history of hypertension, alcohol use, body mass index of patients, family history of DM, systolic blood pressure, diastolic blood pressure, co-morbid condition and initial drug given for patients (Islam, 2017; Taylor & Lobel, 1989; Ikezaki et al, 2002; Baltazar et al, 2004; Miller et al, 2002; Alemu, 2015).

Data Collection Procedure

Data for this study was secondary data routinely recorded from patients' chart in the study area. The health management information system (HMIS) card number used to identify individual patient cards.

Method of Data Analysis

Longitudinal data is defined as the data in which repeated measurements are taken on a subject through time (Carey & Wang, 2011). For longitudinal data, there are two sources of variations: within-subject variation' the variation in the measurements within each subject and this variation allows studying changes over time, and between-subject variation' the variation in the data between different subjects (Laird, & Ware, 1982). In longitudinal data analysis, observations of individual over time are correlated of each other this is because multiple measurements are taken on the same subject at different points in time (Diggle et al, 2002) and; from clustering, where measurements are taken on subjects that share a common category or characteristic that leads to correlation. Longitudinal data can be either continuous or

discrete binary (Diggle et al, 2002). This study, focused on both cases of binary and continuous repeated measurements.

Exploring Mean Structure

The average evolution describes how the profile for a number of relevant sub populations (or the population as a whole) evolve over time and the results of this exploration are used to choose a fixed-effects structure for the linear mixed model and for the detection of inclusion of different form of time effects like linear, quadratic and so on (Shek & Ma, 2011).

Exploring the Correlation Structure

It helps to describe how measurements within an individual correlate. Pair-wise scatter plots matrix and pair-wise correlation matrix are used for exploring the correlation structure (Pusponegoro, 2017; Liu et al, 2012).

Exploring the Variance Structure

To explore the variance structure of the data three plots are used. The first one shows the average evolution of the variance as a function of time and the second produces the individual profile plots of the data which shows whether there is a considerable within and between subject variability. The third is the interaction plot which is used to plot the variance functions separated for different groups (sex: male, female etc.) as a function of time. The covariance structure selected reflects the correlation between successive FBS levels. The four most commonly used covariance structures are compound symmetry (CS), Toeplitz (TOEP), unstructured (UN) and autoregressive (1) (AR (1)) (Liu et al, 2012; Shek and Ma, 2011).

Linear Mixed Model for Longitudinal Data

Mixed models provide a flexible and powerful tool for the analysis of data with complex covariance structure, such as longitudinal correlated data (Guo, 2004). Linear mixed effects models for repeated measures data formalize the idea that an individual's pattern of responses is likely to depend on many characteristics of that individual, including some that are unobserved, where these unobserved effects are then included in the model as random variables, or equivalently called, random effects (Der & Everitt, 2005; Van Montfort, et al., 2010). The name mixed model indicates that the model contains both the fixed or the mean model component and the random

component. It can use for variable effects either fixed or random depending on how the levels of the variables that appear in the study are selected (Laird and Ware, 1982). It can also use for data with unequal number of measurements per subjects. A fixed effect is considered to be a constant which we wish to estimate, but the random effect is considered as just an effect coming from a population of effects (Fitzmaurice et al, 2008).

Generally, the linear mixed model is defined as

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

Where Y_i is the response vector for the i^{th} subject, X_i is the $n_i \times p$ fixed between-subject design matrix, β is the $p \times 1$ vector of fixed effects (population averaged) assumed common for all subjects, Z_i is the $n_i \times q$ random within-subject design matrix (subject specific design matrix). b_i is a q -dimensional vector of random-effect parameters $b_i \sim N(0, D)$, b_1, \dots, b_N are independent, ε_i is $n_i \times 1$ vector of random errors and $\varepsilon_i \sim N(0, \Sigma_i)$, D and Σ_i are variance components (Hallahan, 2003).

In case of our data, if we model with response variable fasting blood sugar (FBS) over time, it can be modeled as

$$FBS = \beta_{10} + \beta_{11}(\text{Age}) + \beta_{12}(\text{Educational status}) + \dots + \beta_{1p}(\text{p}^{\text{th}} \text{covariates}) + b_{10i} + b_{11i}(\text{Time}) + \varepsilon_{1ij}$$

Random intercept model: A random-intercept model, consists of augments, the linear predictor and a single random effect for subject i (Kuznetsova et al., 2017), it is expressed as

$$FBS = \beta_0 + \beta_1(\text{Age}) + \beta_2(\text{Educational status}) + \dots + \beta_p(x_{ijp}) + b_{10i} + \varepsilon_{2ij}$$

Where, β 's is a fixed effect of interest,

$b_{10i} \sim N(0, \sigma^2 b)$ are random effects

$\varepsilon_{1ij} \sim N(0, \sigma^2 \varepsilon)$ are independent measurement error

Random intercept and slope model: The vector of random effects are assumed to follow normal distribution with mean vector 0 and variance covariance matrix D , the linear predictor is written as (Kuznetsova et al., 2017; Zeger and Huang, 2014).

$FBS = \beta_{10} + \beta_{11}(\text{Age}) + \beta_{12}(\text{Educational status}) + \dots + \beta_{1p}(x_{ijp}) + b_{10i} + b_{11i}(t_{ij}) + \varepsilon_{1ij}$
 Fixed effects are the same as before in the random intercept model and b_{10i} & b_{11i} assumed to have a bivariate normal distribution as follows

$$\begin{bmatrix} b_{10i} \\ b_{11i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \right)$$

Fixed Effect Variables: Fixed effects are the covariate effects that are fixed across subjects in the study sample. These effects are the ones of our particular interest for example, Sex where male and female, Age group, Marital Status, Baseline fasting blood sugar, baseline hemoglobin level and Baseline BMI are considered as fixed effect.

Random Effect Variables: Random effects are the covariate effects that vary among subjects and the model parameters are random variables. A random effect model is generally something that can be expected to have a nonsystematic, unpredictable, or “random” influence in our data.

Methods of Parameter Estimation

The estimation methods in longitudinal studies are maximum likelihood (ML) and restricted maximum likelihood estimation (REML) techniques (Guo, 2002). The two estimations are asymptotically equivalent and often give very similar results. Both ML and REML are based on the likelihood principle, which has the properties of consistency, asymptotic normality, and efficiency. The difference between ML and REML is the construction of the likelihood function. The REML estimation method applies ML estimation techniques to the likelihood function. The difference is that, the REML estimation method is associated with a set of “error contrasts” rather than associated with the original observations. Therefore, it will lose degrees of freedom and give less biased estimates of the variance components. The bias issue cannot be neglected, especially when the number of parameters is not small relative to the total number of observations (Verbeke, 1997).

Missing Data in Longitudinal Studies

Missing data is common in longitudinal studies (Ibrahim & Molenberghs, 2009). In longitudinal measurements, missing data are the main problem of analysis. Since

missing data may hide the true values that are important in making an actual estimation. Therefore, ignoring missing value leads to biased estimate.

Methods for handling missing data

There are different methods and strategies available to handle missing data. Graham (2009) defines three conditions that should be satisfied in a good method. First, the method should yield unbiased estimates of a variety of different parameters. Second, the method should include a way to assess the uncertainty about the parameter estimates, and third, the method should have good statistical power.

Multiple imputation is the popular method for handling missing values in a data. It replaces each missing value with five or more acceptable values, representing a distribution of possibilities (Rubin, 2000). The multiple imputation inference involves three distinct phases (Rubin, 2000): The first step is the missing data are filled m times to generate m complete data sets, then the m complete data sets are analyzed by using appropriate procedures and finally the results from the m complete data sets are combined for the inference.

Model Comparison Techniques

The aim of model comparison is to choose the most parsimonious model that provides best fit to the data. Even though, there are different techniques that uses to select the best-fitted model on this paper we used Akaike's information criterion (AIC), Bayesian information criterion (BIC) and Likelihood ratio test .

Akaike's information criterion (AIC) and Bayesian information Criterion (BIC) are given by the formula

$$AIC = -2\log L + 2k$$

$$BIC = -2\log L + k\log(N)$$

Where, k denotes the number of parameters in the model and N the total number of observations used to fit the model. If we use AIC and BIC to compare two or more models for the same data, we prefer the model with the lowest AIC and BIC values (Zhang and Davidian, 2001; Pourahmadi and Daniels, 2002; Lin and Lee, 2008; Akaike, 1974; McQuarrie & Tsai, 1998; Schwarz, 1978).

Likelihood-ratio test is constructed by comparing the maximized log likelihoods for the Saturated (full) model and reduced models, respectively, and the test statistic is expressed as:

$$\text{LRT} = -2\ln(L_0 - L_m)$$

Where, L_0 and L_m are the maximum likelihood estimates that maximize the likelihood functions of the reduced and full or saturated model, respectively (Shek and Ma, 2011).

Model Diagnostics

In linear mixed effects model, the assumption is that, residuals and random effects are normally distributed and uncorrelated with the error term and violation of this assumption does affect the parameter estimates and standard errors of the residual effects. Residual plots can be used visually to check normality of these effects and to identify any outlying effect categories. Examining the plot of the standardized residuals versus fitted values by any covariates of interest can give a better feeling (Lange & Ryan, 1989; Zhang & Davidian, 2001). The assumption of normality for the within-group error was assessed with the normal quantile plot of the residuals by covariates.

Results and Discussion

Distribution of Fasting Blood Sugar Level Data

The normality of the data was checked by the boxplot. This study was found that the actual fasting blood sugar level were not normal at all-time points as the test revealed that significant deviation from the assumption of normality. From the below Figure 1, boxplot of the actual FBS data shows that distribution of FBS is skewed to the right (high FBS level) especially for minor points. Hence, transformation of actual fasting blood sugar is needed. After the logarithmic transformation, right-side plot, the outlier observations observed were minimized in logFBS than actual fasting blood sugar level, hence the data attained slight normality for logarithmic transformation when we compare to actual fasting blood sugar. Therefore, logarithmic transformed fasting blood sugar level is better than actual FBS.

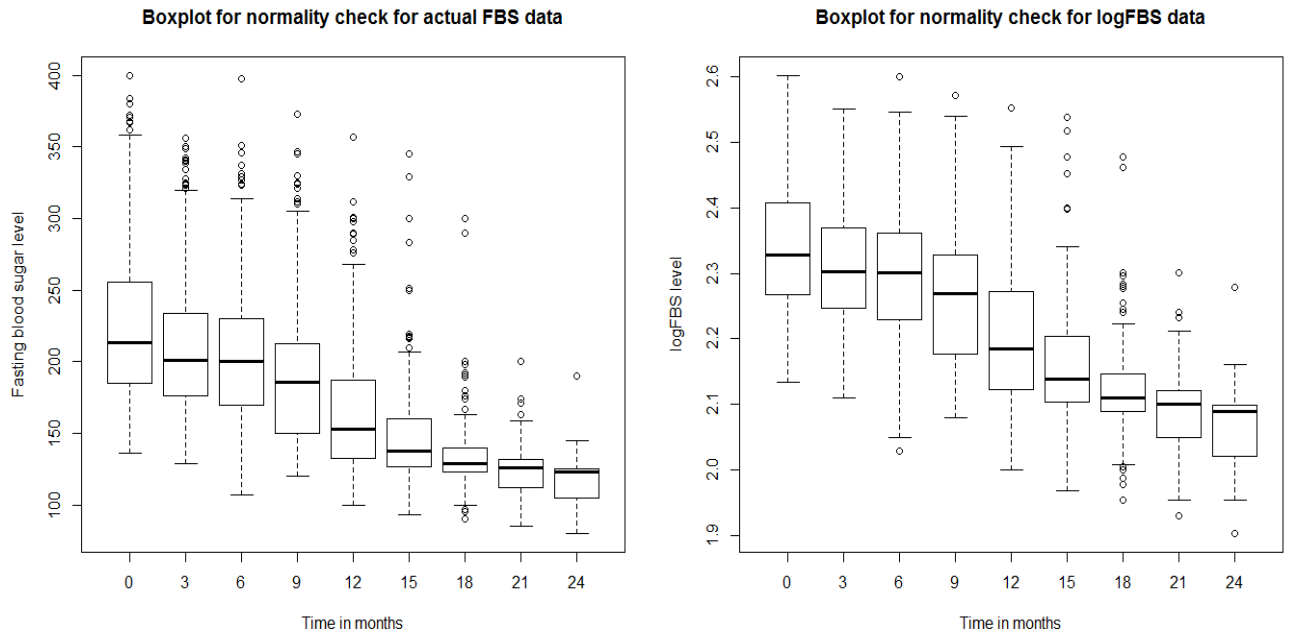


Figure 1: Box plot of fasting blood sugar and logarithmic of FBS of DM patients

Exploratory Data analysis

Exploring Individual Profiles

Exploring individual profiles over time is used to identify general trends within patients and may detect change over time that gives information about the variability at given times. Figure 2 indicate that almost decreasing linear time effect and there is no need of considering other forms of time effects, we observed that there was high variability between patients but less variability within patients over time. It is also observed from this plot that, there was a fluctuation in FBS of patients over time. The variation at the beginning was higher than at the end of the follow up time for patients. Therefore, to fit the data which have variability in the intercept and slope of trajectories very well, we have used a mixed model.

Individual profile plot by follow

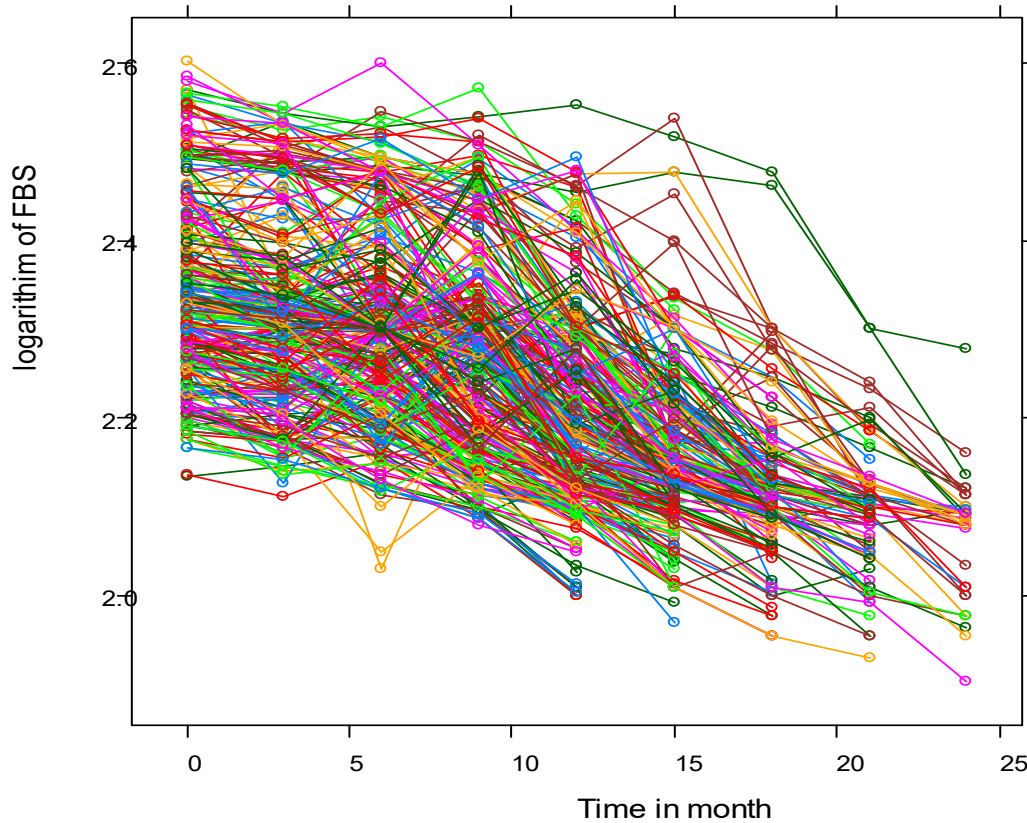


Figure 2: Individual profile for logarithmic transformed FBS level of DM patients

Exploring the Mean Profile

Figure 3 below revealed that the mean profile plot of log FBS level over time using multiple imputation and by LOWESS smoothing respectively. Since all measurements in the data have no similar time of measurements, we have used locally weighted scatter plot smoothing (LOWESS) to determine FBS variation over time. Generally, as one can easily understand from the mean profile plot, the log FBS level of these patients showed a linear decrement over the treatment time. The mean log FBS level of patients seems to show decreasing and shows linear pattern over time for all covariates even if the amount varies from group to group with in a covariate.

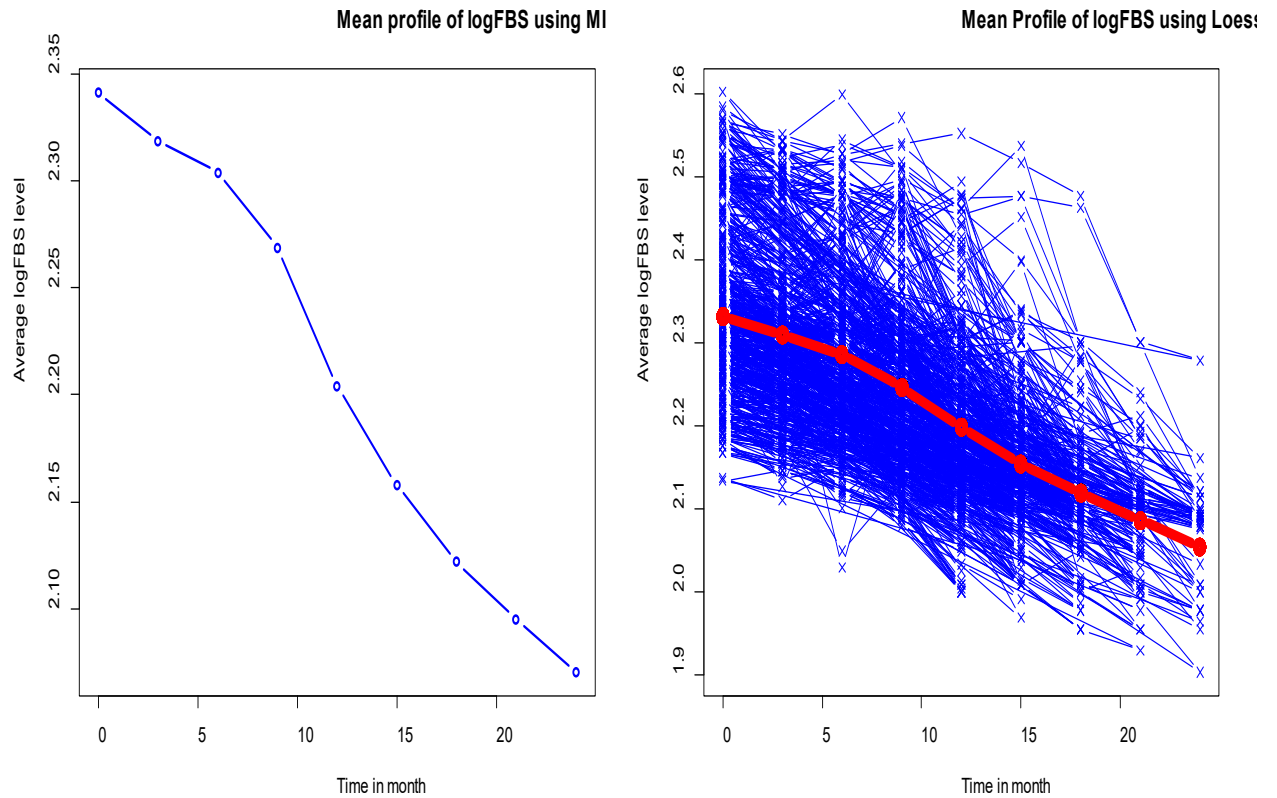


Figure 3: Mean profile plot of logarithm of FBS level over time using Loess smoothing and MI

Exploring Variance Profiles

The variability structure plays a great role in identifying the pattern of outcome variable and gives information on what variance-covariance structure to be expected. The variance structure for log FBS shows an irregular pattern over time. As Figure 4 showed there was high variability at the baseline and low variation between patients at the end of follow up time. The variance of patients fasting blood sugar level over time fluctuates through follow up time and changed by categorical covariates over time.

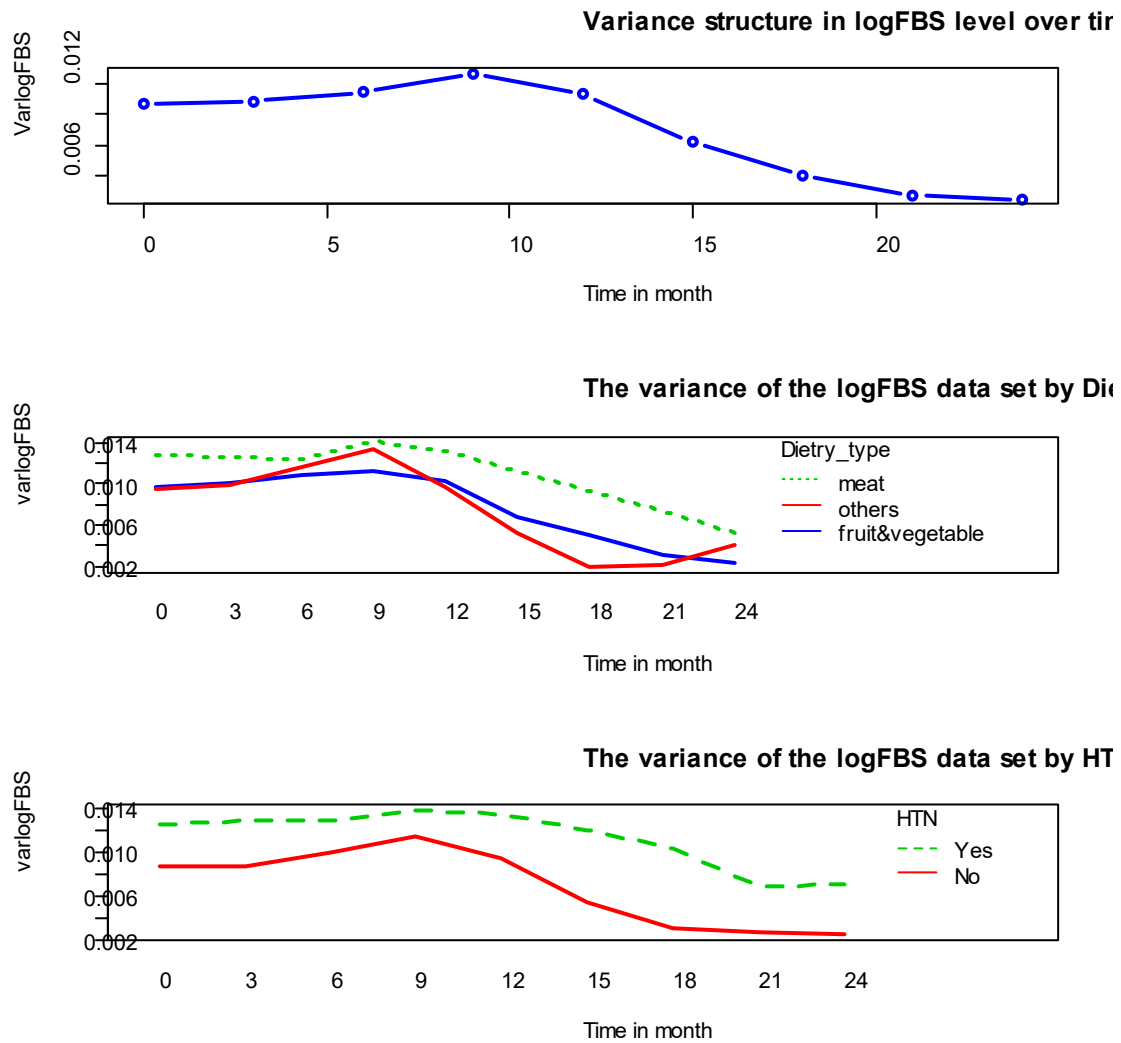


Figure 4: Variance profile plot of log FBS level by follow up time for some categorical covariates.

Exploring Correlation Structure Matrix

The correlation matrix is used to show the dependence between repeated measurements and the correlation between repeated measurements of the responses over time. The correlation matrix shown in Table 1 revealed a positive correlation between any two repeated measurements which relate the correlation deemed decreasing over time. Since the off-diagonal correlation has no constant value over time, it gives the clues on correlation structure may be unstructured.

Table 1: Correlation structure matrix

	logFBS0	logFBS3	logFBS6	logFBS9	logFBS12	logFBS15	logFBS18	logFBS21	logFBS24
logFBS0	1.0000								
logFBS3	0.9303	1.0000							
logFBS6	0.8051	0.8617	1.0000						
logFBS9	0.8676	0.8819	0.8118	1.0000					
logFBS12	0.7766	0.8294	0.8006	0.8394	1.0000				
logFBS15	0.6097	0.6662	0.6484	0.7311	0.7685	1.0000			
logFBS18	0.5911	0.6574	0.6068	0.6885	0.7662	0.8585	1.0000		
logFBS21	0.5516	0.5990	0.5571	0.6154	0.6776	0.7803	0.8870	1.0000	
logFBS24	0.3817	0.4376	0.3764	0.4048	0.5183	0.6210	0.7434	0.8238	1.0000

Linear Mixed Model

Variance Covariance Structure

The variance covariance structure is the important issues to be included in the longitudinal data analysis. Comparing with Akaike information criteria (AIC) and Bayesian Information criteria (BIC) the correlation structure with the smaller value of AIC and BIC was chosen, so, the unstructured correlation structure best explained the model under the study which is displayed on Table 2. Therefore, unstructured variance covariance was used in identifying the correlation structure

Table 2: Correlation structure checking

	Compound symmetry	Unstructured	AR (1)
AIC	5831.52	5783.03	5912.65
BIC	5713.93	5689.13	5782.60

Random effect model

In Longitudinal data analysis, random effect should be included to the model in order to account between individual variability. The need for model with random intercept and slope were checked by AIC and likelihood ratio test. So, as we can see from Table 3 the inclusion of random intercept and random slope is reasonable, so in the final model both random intercept and random slope with linear time effect was considered for this study.

Table 3: Comparison of random effect model

Random effects model	AIC	Likelihood ratio-test
Model 1 Intercept	5534.4	2808.2
Model 2 Intercept, time	5277	2677.7

Multi-variable Linear Mixed Model

After we chose the reasonable model, the variables are selected in the univariate and fixed effects model using backward elimination method were incorporated into multivariable linear mixed effect model and then the model with estimated value of significant covariates were fitted. The restricted maximum likelihood estimates of covariates and standard errors with its corresponding significance value (P-value) is found in the following Table 4.

Table 4: Parameter estimates for full linear mixed-effects model with both random intercept and slope model

Effect	Estimate	St. Error	DF	t-value	P-value
Intercept	1.9439499	0.018031159	1795	107.8105	<0.0001**
Age	-0.0002012	0.000212467	292	-0.94697	0.0344*
SexMale	-0.0025840	0.004581226	292	-0.56404	0.5732
Educational_statusPrimary	-0.0018078	0.006890196	292	-0.26238	0.7932
Educational_statusSecondary	-0.0055161	0.006625390	292	-0.83257	0.0405*
Educational_statusTertiary	-0.0058844	0.006963855	292	-0.84500	0.0398*
Marital_statusMarried	0.0121576	0.005556352	292	2.18805	0.2951
Marital_statusDivorced	0.0049914	0.007617271	292	0.65528	0.5128
Marital_statusWidowed	0.0024773	0.010917391	292	0.22692	0.8206
ResidenceUrban	0.0048482	0.004746779	292	1.02138	0.3079
Dietry_typemeat	0.0027650	0.004710933	292	0.58694	0.0470*
Dietry_typeothers	-0.0029909	0.005119327	292	-0.58423	0.5595
Exercise_activityperform exercise	-0.0036818	.005100623	292	-0.72184	0.4710
Drug_typeOHA	0.0018733	0.005500709	292	0.34056	0.7337
Drug_typeboth	-0.0069765	0.004634013	292	-1.50549	0.0133*
HTNYes	0.0062799	0.005408709	292	1.16107	0.0246*
Alcohol_statusYes	0.0056387	0.004998443	292	1.12808	0.2602
DBP	0.0004749	0.000208397	292	2.27878	0.0234*
BMI	0.0009951	0.000731718	1795	1.35992	0.0434*
Time	-0.0111267	0.002707189	1795	-4.11007	<0.0001**
Other_CommorbidityYes	0.0043492	0.004075806	292	1.06709	0.2868
DBP*Time	0.0000064	0.000032421	1795	0.19597	0.8447
Dietry_typemeat*Time	0.0023280	0.000738460	1795	3.15252	0.0016**
HTNYes*Time	0.0020705	0.000832589	1795	2.48683	0.0130*
Time*Other_CommorbidityYes	0.0011170	0.000631890	1795	1.76766	0.0077**

Note: ** is significant at 1% level of significant and * is significant at 5% level of significant

Based on estimated value of covariates and its significance level obtained from Table4, linear mixed effect model was modeled as:

$$\text{Log(FBS)} = 1.944 - 0.0002012(\text{Age}) - 0.0055161(\text{Educational_statusSecondary}) - 0.0058844(\text{Educational_statusTertiary}) + 0.0027650(\text{Dietry_typemeat}) - 0.0069765(\text{Drug_typeb oth}) + 0.0062799(\text{HTNYes}) + 0.0004749(\text{DBP}) + 0.0009951(\text{BMI}) - 0.0111267(\text{Time}) + 0.0043492(\text{Other_CommorbidityYes}) + 0.0023280(\text{Dietry_typemeat} * \text{Time}) + 0.0020705(\text{HTNYes} * \text{Time}) + 0.0011170(\text{Time} * \text{Other_CommorbidityYes}) + b_{10i} + b_{11i}$$

Where; b_{10i} is the random intercept and b_{11i} is the random slope for the linear effect of time, represents how the subsets of regression parameter for the i^{th} individuals deviates from population average and it is difficult to derive all coefficients of random effect for all subjects.

Based on the multivariable linear mixed-effects regression output Age, Educational status, dietary type, drug type, history of hypertension, Diastolic blood pressure, BMI, time and interaction effect of history of hypertension, Dietary type, other comorbidity at baseline with time were found to be significant determinant of logarithm change in FBS level of DM patients.

The results of statistical analysis of linear mixed model was discussed depending on the model fitted below. For a given patient, keeping the effect of other covariates constant, for one-month change in time decreases log FBS level by 0.0111267 mg/dl. A unit increase in the age of a patient on treatment, the expected value of logFBS level reduced by 0.0002012 mg/dl while a unit increase in BMI of a patient on treatment, the log FBS level was increased by 0.0434 mg/dl. Similarly, a unit increase in DBP, the log FBS level was increased by 0.0004749 mg/dl when the effect of the other factors were kept constant. Being tertiary and secondary level of education decreases logFBS level by 0.0058844 and 0.0055161 respectively compared with patients with no education (reference group). Drug type has undeniable effect on the change of logarithm of FBS level, patients who take a combination of (insulin +OHA) drug lowers their logarithm of FBS level by 0.0069765 amount as compared to patients who take insulin drug (reference group) keeping the effect of other covariates constant.

The interaction effect of time with meat as dietary type, patients with history of hypertension and comorbid with other disease shows that the rate of change in the fasting blood sugar of the DM patients over time. For a unit increment of the visiting time of the patient, the average rate of change of log (FBS) level of patients who usually uses meat was 0.0023280 increment over treatment time as compared to the patient who uses fruit and vegetable dietary types by fixing the effect of other covariate. Also, interaction effect of history of hypertension with time also significant effect on the log FBS Level change, which means that patients who have history of hypertension increased an average log FBS level by 0.0063 over time relative to patients without history of hypertension keeping the effect of other covariates constant. In addition, for a unit increment of the visiting time of the patient, the expected or average rate of log FBS level of patients who were comorbid with other disease at baseline increased by 0.001117 over time as compared to the patients who do not have con-infected with other disease.

Table 5: Random parameter estimates for FBS data set

Groups	Name	Standard deviation	Intra class corr.
Subject	(Intercept)	0.105	
	Time	0.004	0.687
	Residual	0.04943	

From Table 5, the variances of the intercepts and linear effects of time were significantly different from zero. This showed that the log FBS level values at baseline vary across subjects and the change of log FBS level over time vary within subjects. This means that the highest variability came from the random intercepts. It also shows that the standard deviation of the random intercepts was higher than that of the random slopes, showing to higher between patient variability than the within patient variability. The value of the intra-class correlation which is the ratio of the between-patients variance to the total variance is presented in Table 4. It tells us the proportion of the total variance in log FBS level (68.7%) that is accounted by variance among patients. The value of the intra class correlation revealed that mixed model is necessary to fit the data.

Model Diagnostic Checking

In model diagnostic checking for longitudinal data analysis, we use residual plots to evaluate the validity of model assumption as shown on Figure 5. The normal quantile plot shows that the residuals do not exhibit departure from normality. The “Residual fit” concentrated around zero implies that linear mixed effect model were well fitted the FBS data.

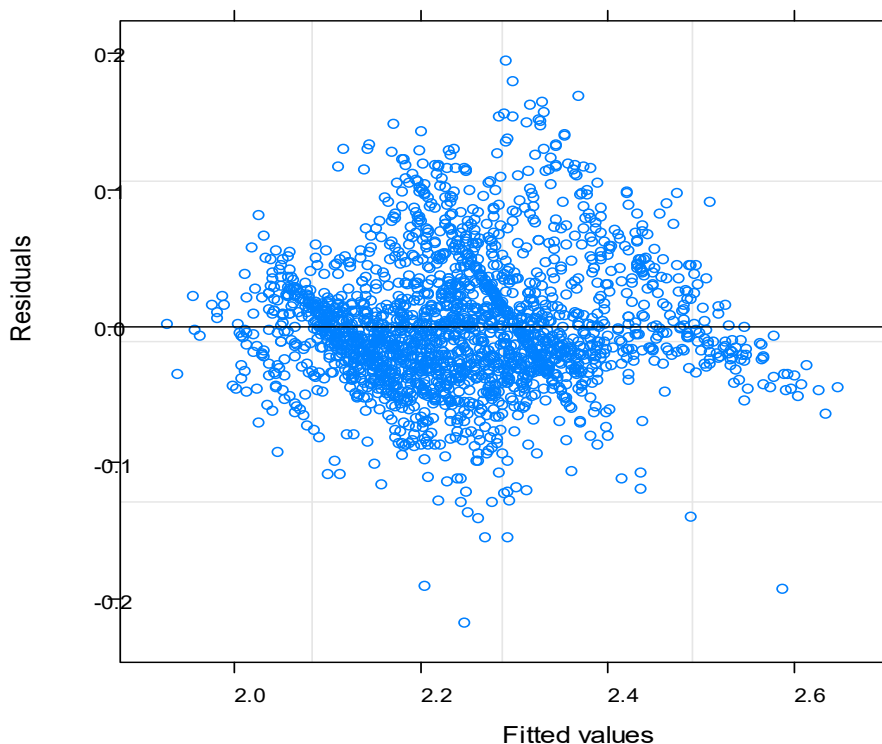


Figure 5: Residuals plot for logarithm of FBS level data

Discussion

From linear mixed model the study revealed that people who mostly eat meat were exposed to DM disease and could not control the FBS to attain a normal fasting blood sugar due to lack of mineral contents for disease protection. To control and manage blood sugar, Diet which related with cereals, vegetables, and fruits are recommended while other dietary types related with eggs, meat, proteins are increases blood sugar level which is related to diabetes and its complication. This indicated that the combination of the two drug has more important in reduction and control of fasting blood sugar as compared to insulin and OHA drug. When compared with patients who had no history of hypertension, those having history of hypertension increases fasting

blood sugar level. The interaction effect of history of hypertension with time was also significant effect on fasting blood sugar variation over time that means that there were significant difference in variation of log FBS level between patients who had history of hypertension and had not. The interaction effect of comorbidity at the baseline with time was found to be significant effect on fasting blood sugar variation over time. There is significant difference of fasting blood sugar change between patients who comorbid with other disease and who are free from comorbidity of other disease at baseline. The rate of reduction in fasting blood sugar level change was higher for those patients with comorbidity at baseline when compared to those who had no other comorbidity diseases at baseline. Being a retrospective study, this study has limitations of secondary data collection; unequal number of measurement per subjects and we were unable to find some variable like income, side effect from other medication, illness and stress that may have influenced the rate of change in FBS level. Therefore, more public health and epidemiology researchers are needed to examine the impact of those predictors on population health, people living with diabetes to avoid its complications over time and to identify new risk factors for diabetes.

Conclusion

In this research, we demonstrate that the use of linear mixed effect model including (random intercept and random slope) with unstructured covariance structure was preferred for estimation of the rate of change of log (FBS) level experienced by patients over treatment time. With respect to time, we found that the pattern of average of fasting blood sugar level revealed a linear decrement over time that was also confirmed with the model that the estimate of time was negative. This study illustrated that, 68.7% of variability in log FBS levels were accounted by variance among patients.

Finally, we conclude that time (duration of follow up), BMI, diastolic blood pressure, dietary type, education status, age, history of hypertension, drug type and interaction of dietary type, history of hypertension and other comorbidity with time were found to be significant predictors on the FBS level variation over time (p-value <0.05). From the result of the study, patients who had no history of hypertension, have lower BMI, higher educational attainment, and who used a combination of

insulin and OHA drug were better to manage and control their fasting blood sugar level in their body. However, the determinant factors of marital status, Alcohol status, Residence and sex were an insignificant variable which showed that there was no significance difference in reduction of FBS level by their category over time.

Reference

- Adampah, T., Nawumbeni, D. N., Nyadanu, S. D., & Polishuk, R. (2015). Mixed-Effects Model for Longitudinal Study of Type-2-Diabetes.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Alemu, F. (2015). Prevalence of diabetes mellitus disease and its association with level of education among adult patients attending at Dilla Referral Hospital, Ethiopia. *J Diabetes Metab*, 6(4), 1-5.
- Aniley, T. T., Debushe, L. K., Nigusie, Z. M., Yimer, W. K., & Yimer, B. B. (2019). A semi-parametric mixed models for longitudinally measured fasting blood sugar level of adult diabetic patients. *BMC medical research methodology*, 19(1), 1-11.
- Baltazar, J. C., Ancheta, C. A., Aban, I. B., Fernando, R. E., & Baquilod, M. M. (2004). Prevalence and correlates of diabetes mellitus and impaired glucose tolerance among adults in Luzon, Philippines. *Diabetes research and clinical practice*, 64(2), 107-115.
- Carey, V. J., & Wang, Y. G. (2011). Working covariance model selection for generalized estimating equations. *Statistics in medicine*, 30(26), 3117-3124.
- Der, G., & Everitt, B. S. (2005). *Statistical analysis of medical data using SAS*. CRC Press.
- Federation, I. D. (2019). IDF diabetes atlas ninth. *Dunia: IDF*.
- Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (Eds.). (2008). *Longitudinal data analysis*. CRC press.
- Diggle, P., Diggle, P. J., Heagerty, P., Liang, K. Y., Heagerty, P. J., & Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, 60, 549-576.
- Guo, W. (2002). Functional mixed effects models. *Biometrics*, 58(1), 121-128.

- Guo, W. (2004). Functional data analysis in longitudinal settings using smoothing splines. *Statistical methods in medical research*, 13(1), 49-62.
- Hallahan, C. (2003). Longitudinal Data Analysis with Discrete and Continuous responses using Proc Mixed. *Maintained at: <http://www.cpcug.org/user/sigstat/PowerPointSlides>.*
- Ibrahim, J. G., & Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. *Test*, 18(1), 1-43..
- Ikezaki, A., Hosoda, H., Ito, K., Iwama, S., Miura, N., Matsuoka, H., ... & Sugihara, S. (2002). Fasting plasma ghrelin levels are negatively correlated with insulin resistance and PAI-1, but not with leptin, in obese children and adolescents. *Diabetes*, 51(12), 3408-3411.
- Islam, M. R. (2017). Association between socio-demographic factors and blood sugar levels in type 2 diabetes mellitus patients in Bangladesh. *Journal of diabetes mellitus*, 7(03), 151.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82(13), 1-26.
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.
- Lange, N., & Ryan, L. (1989). Assessing normality in random effects models. *The Annals of Statistics*, 624-642.
- Lin, T. I., & Lee, J. C. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in medicine*, 27(9), 1490-1507.
- Liu, S., Rovine, M. J., & Molenaar, P. (2012). Selecting a linear mixed model for longitudinal data: Repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological methods*, 17(1), 15.
- Mahmoudi, A. (2006). Effects of self care planning on reduction of A1C hemoglobin in adults with diabetes mellitus. *Medical Science Journal of Islamic Azad Univesity-Tehran Medical Branch*, 16(3), 171-176.
- McQuarrie, A. D., & Tsai, C. L. (1998). *Regression and time series model selection*. World Scientific.

- Megerssa, Y. C., Gebre, M. W., Birru, S. K., Goshu, A. R., & Tesfaye, D. Y. (2013). Prevalence of undiagnosed diabetes mellitus and its risk factors in selected institutions at Bishoftu Town, East Shoa. *Ethiopia J Diabetes Metab.* 2013.
- Miller, C. K., Edwards, L., Kissling, G., & Sanville, L. (2002). Nutrition education improves metabolic outcomes among older adults with diabetes mellitus: results from a randomized controlled trial. *Preventive medicine*, 34(2), 252-259.
- Nazir, M. A., AlGhamdi, L., AlKadi, M., AlBeajan, N., AlRashoudi, L., & AlHussan, M. (2018). The burden of diabetes, its oral complications and their prevention and management. *Open access Macedonian journal of medical sciences*, 6(8), 1545.
- Ogurtsova, K., da Rocha Fernandes, J. D., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N. H., ... & Makaroff, L. E. (2017). IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes research and clinical practice*, 128, 40-50.
- Patterson, C. C., Karuranga, S., Salpea, P., Saeedi, P., Dahlquist, G., Soltesz, G., & Ogle, G. D. (2019). Worldwide estimates of incidence, prevalence and mortality of type 1 diabetes in children and adolescents: Results from the International Diabetes Federation Diabetes Atlas. *Diabetes research and clinical practice*, 157, 107842.
- Pourahmadi, M., & Daniels, M. J. (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics*, 58(1), 225-231.
- Pusponegoro, N. H., Notodiputro, K. A., & Sartono, B. (2017). Linear mixed model for analyzing longitudinal data: A simulation study of children growth differences. *Procedia computer science*, 116, 284-291.
- Roglic, G. (2016). WHO Global report on diabetes: A summary. *International Journal of Noncommunicable Diseases*, 1(1), 3.
- Rubin, D. B. (2000). Causal inference without counterfactuals: comment. *Journal of the American Statistical Association*, 95(450), 435-438.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of statistics*, 6(2), 461-464.
- Shek, D. T., & Ma, C. (2011). Longitudinal data analyses using linear mixed models in SPSS: concepts, procedures and illustrations. *TheScientificWorldJOURNAL*, 11, 42-76.

- Taylor, S. E., & Lobel, M. (1989). Social comparison activity under threat: downward evaluation and upward contacts. *Psychological review*, 96(4), 569.
- Van Montfort, K., Oud, J. H., & Satorra, A. (Eds.). (2010). *Longitudinal research with latent variables*. Springer Science & Business Media.
- Verbeke, G. (1997). Linear mixed models for longitudinal data. In *Linear mixed models in practice* (pp. 63-153). Springer, New York, NY.
- World Health Organization. (2006). Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation.
- World Health Organization. (2018). Noncommunicable diseases country profiles 2018.
- Zeger, S. L., Diggle, P. J., & Huang, W. (2014). Generalized linear models for longitudinal data. *Wiley StatsRef: Statistics Reference Online*.
- Zhang, D., & Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57(3), 795-802.

Figures

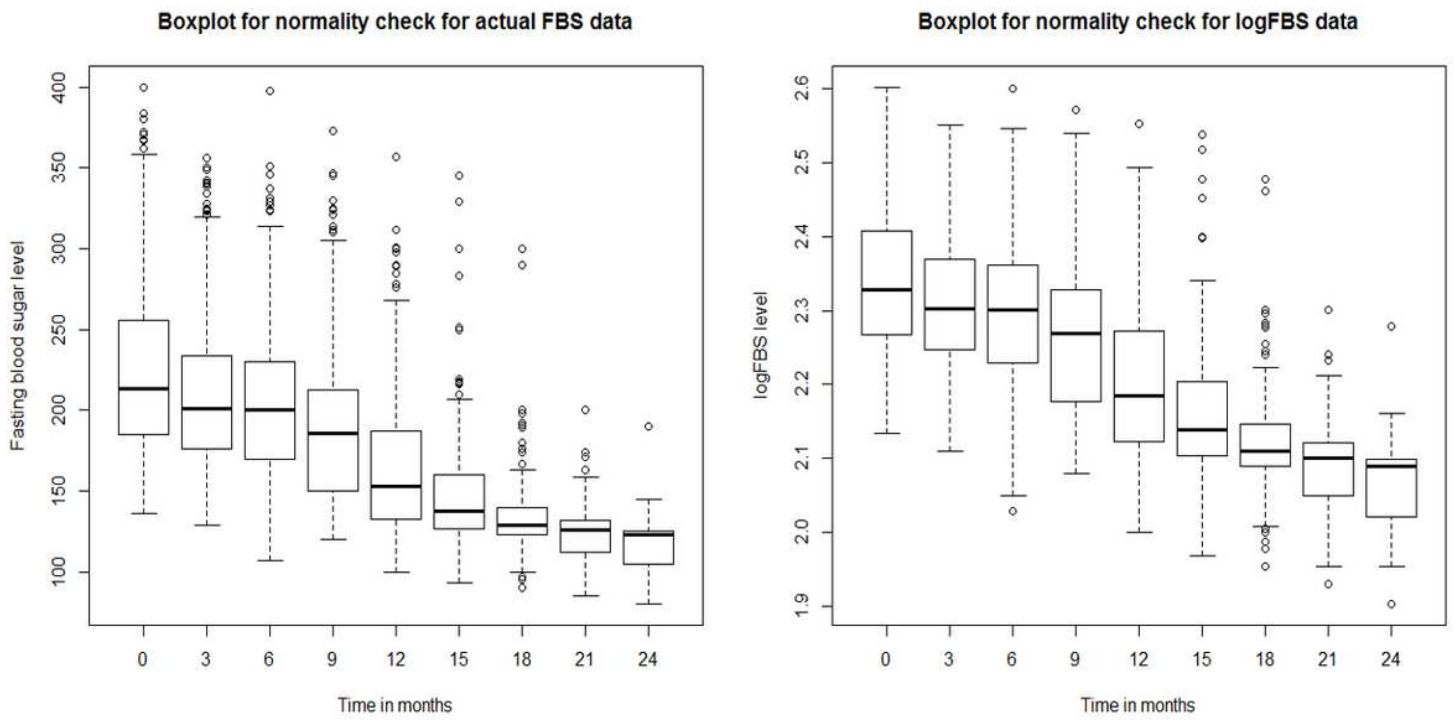


Figure 1

Box plot of fasting blood sugar and logarithmic of FBS of DM patients

Individual profile plot by follow up time

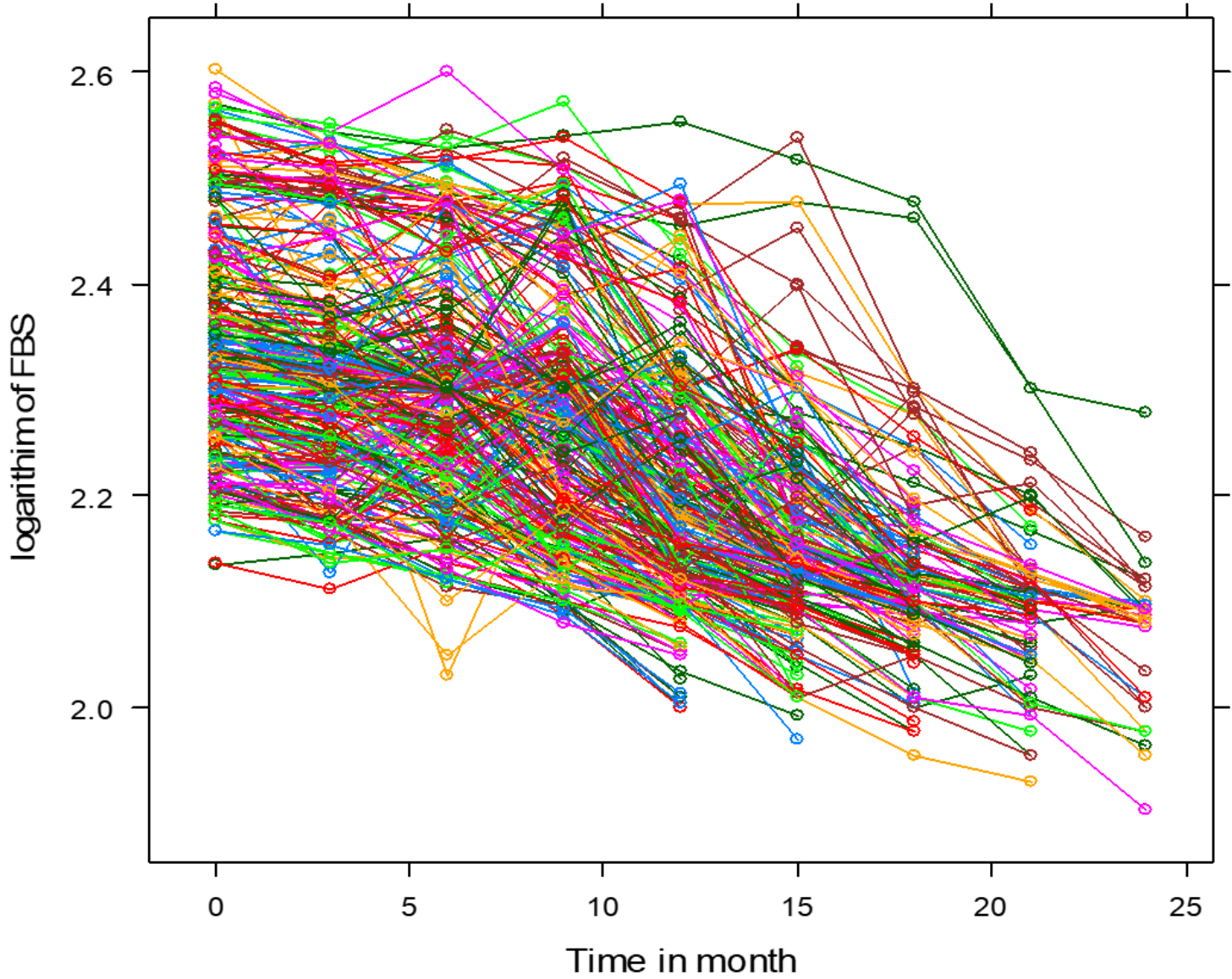


Figure 2

Individual profile for logarithmic transformed FBS level of DM patients

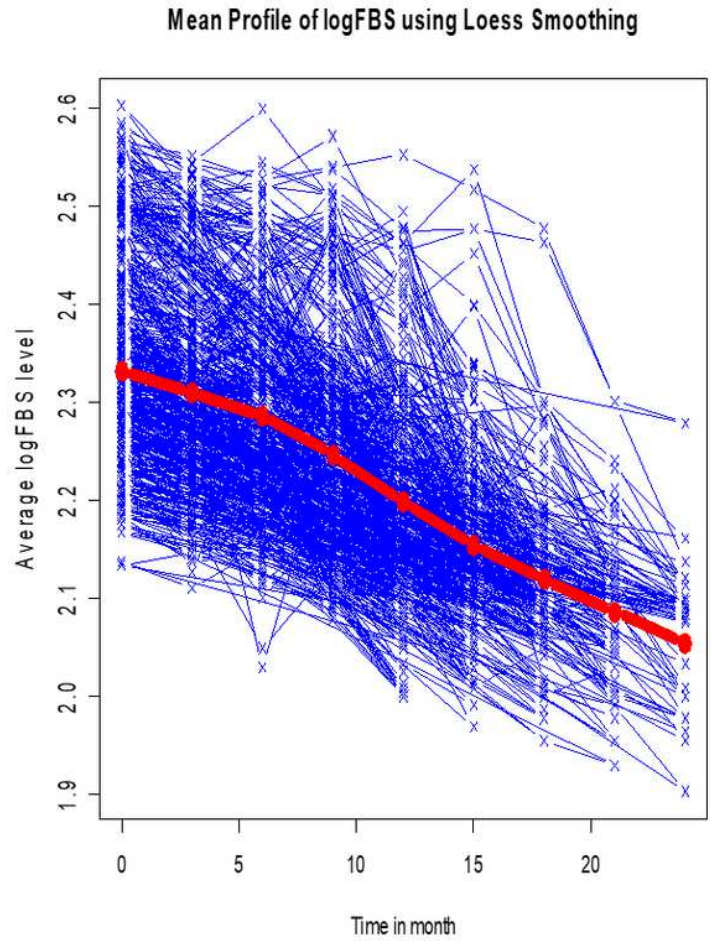
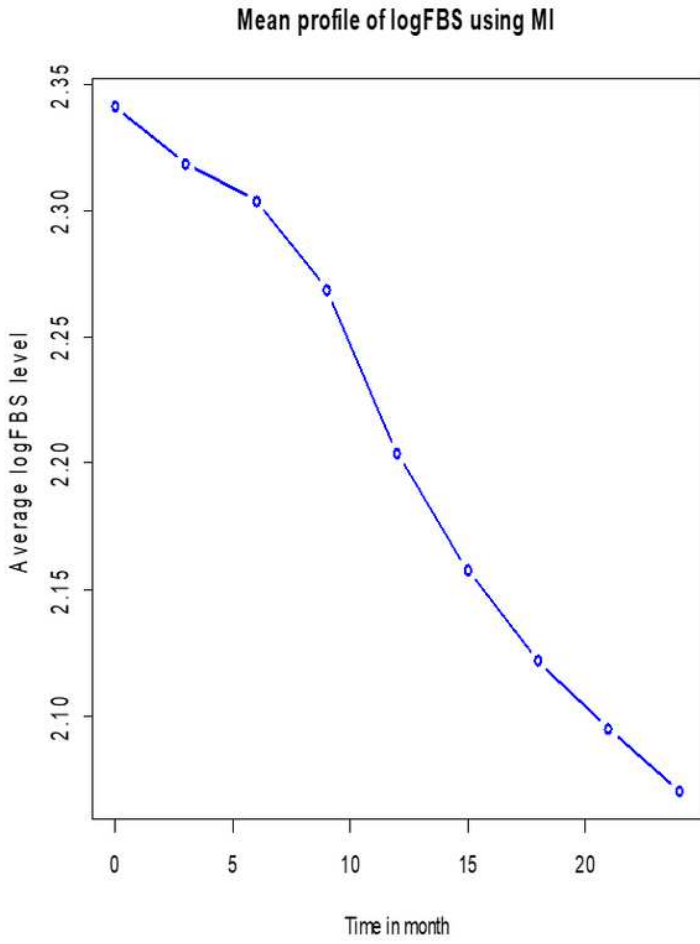


Figure 3

Mean profile plot of logarithm of FBS level over time using Loess smoothing and MI

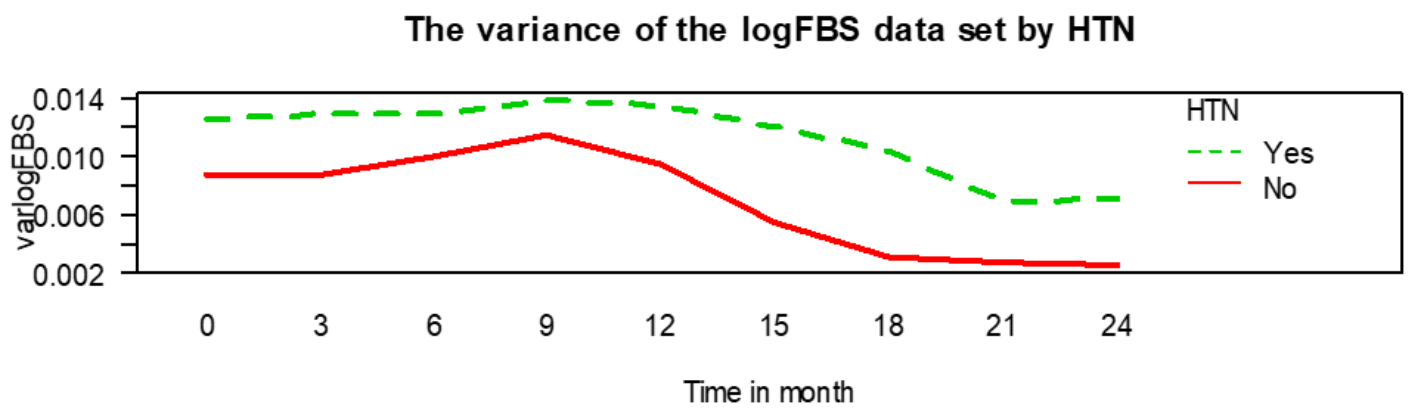
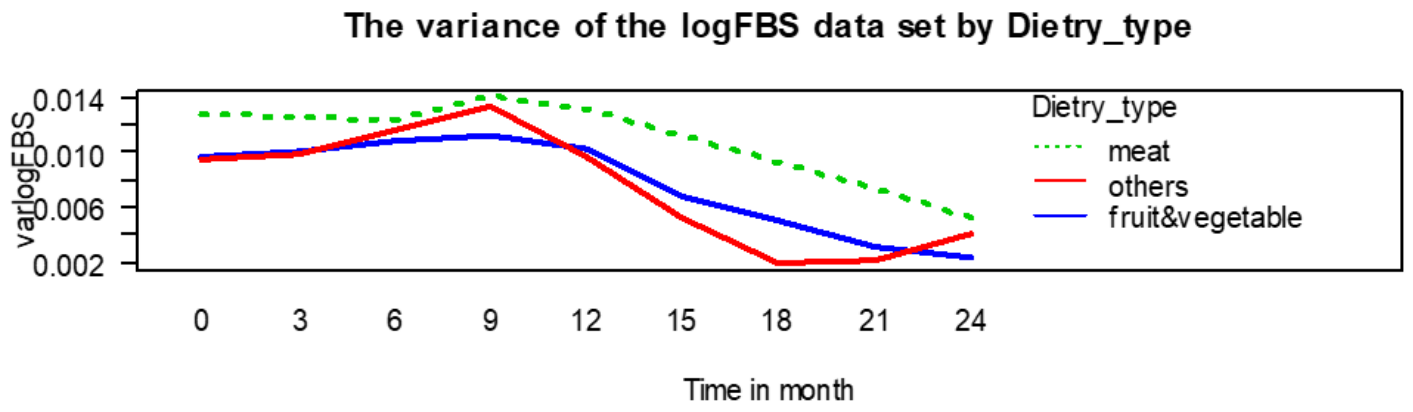
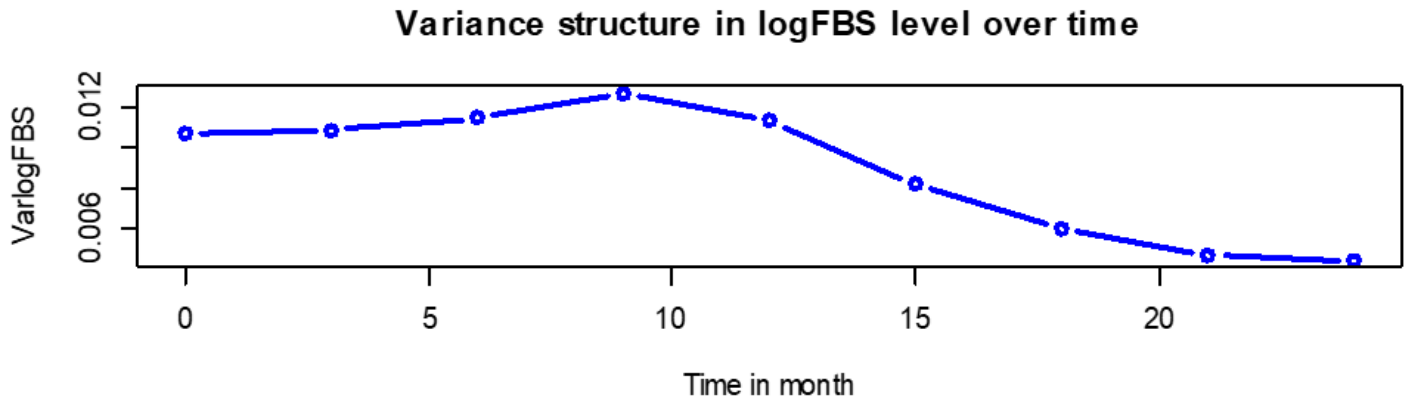


Figure 4

Variance profile plot of log FBS level by follow up time for some categorical covariates.

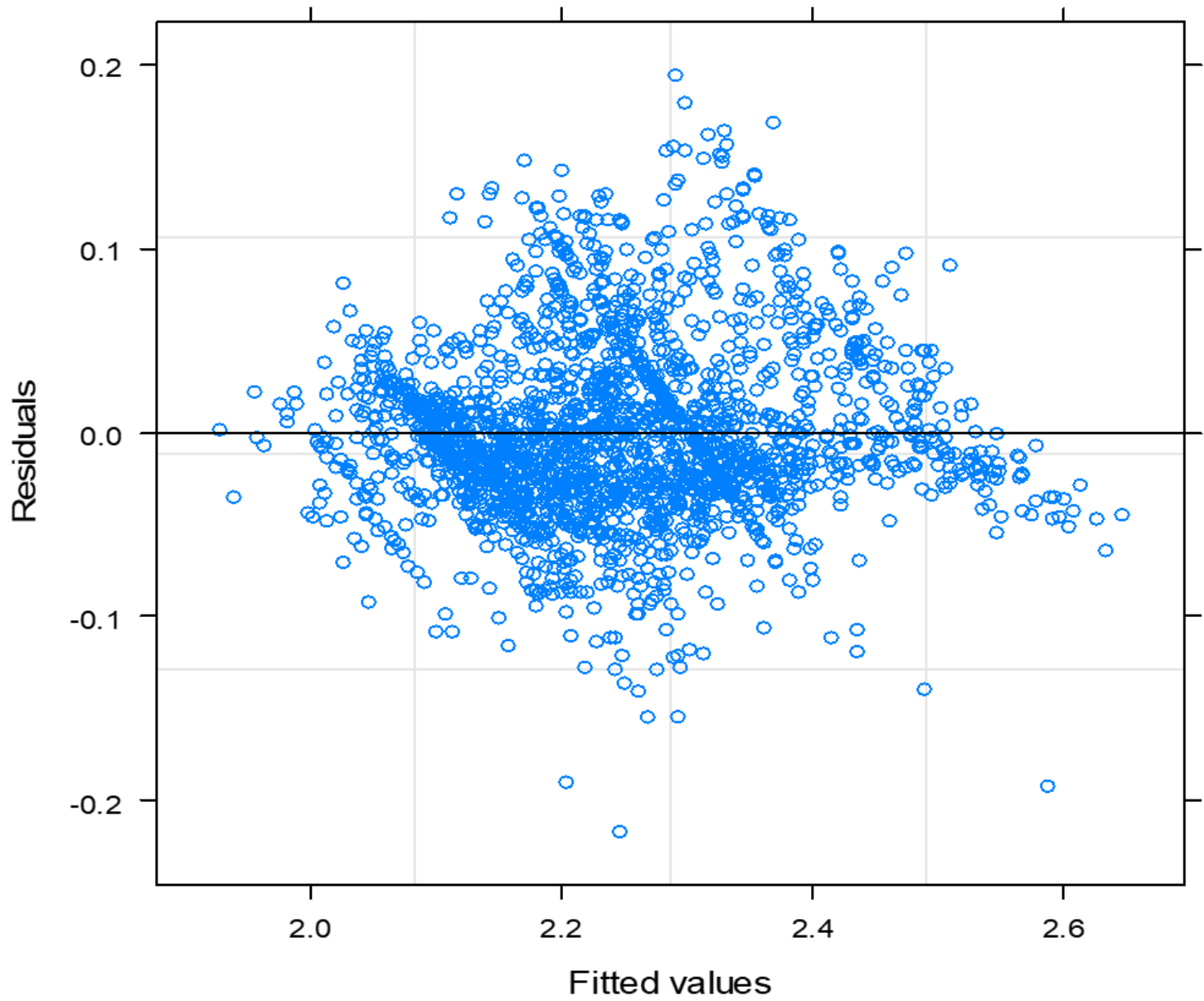


Figure 5

Residuals plot for logarithm of FBS level data