



# Gaussian Longitudinal Analysis of Progression of Diabetes Mellitus Patients Using Fasting Blood Sugar Level: A Case of Debre Berhan Referral Hospital, Ethiopia

Wudneh Ketema Moges\*, A. R. Muralidharan, Haymanot Zeleke Tadesse

Department of Statistics, College of Natural and Computational Science, Debre Berhan University, Debre Berhan, Ethiopia

## Email address:

wudnehketema@gmail.com (W. K. Moges), mdharan75@gmail.com (A. R. Muralidharan),

haymiezt@gmail.com (H. Z. Tadesse)

\*Corresponding author

## To cite this article:

Wudneh Ketema Moges, A. R. Muralidharan, Haymanot Zeleke Tadesse. Gaussian Longitudinal Analysis of Progression of Diabetes Mellitus Patients Using Fasting Blood Sugar Level: A Case of Debre Berhan Referral Hospital, Ethiopia. *American Journal of Theoretical and Applied Statistics*. Vol. 7, No. 1, 2018, pp. 21-28. doi: 10.11648/j.ajtas.20180701.13

Received: November 21, 2017; Accepted: December 4, 2017; Published: January 9, 2018

---

**Abstract:** Diabetes mellitus is a metabolic disorder where by glucose cannot effectively get transported out of the blood. It is a chronic disease with a high prevalence and growing concern in world wide. There are two Types of diabetes, which are Type I and Type II. A longitudinal data analysis retrospective based study was conducted between 1<sup>st</sup> September, 2012 to 30<sup>th</sup> August 2015 in Debre Berhan referral hospital. The main objective of the study was Gaussian longitudinal analysis of progression of Diabetes mellitus patients using fasting blood sugar level count following insulin, metformin and to identify factors predicting the progression of diabetic infection. A total of 248 Diabetes mellitus patients were included in the study whom 111 (44.8%) were females and the rest 137 (55.8%) were males. The generalized linear mixed model would be used to model the progression of diabetic infection. The appropriate variance covariance structure was Compound symmetry selected for this study. This study showed that age, sex, time, illiterate with time, primary with time, address with time, age with time and time with time were statistically significant factors for the progression of fasting blood sugar level at a logarithmic fasting sugar level over time in generalized linear mixed model. The mean fasting blood sugar level showed an increasing progress over time after patients were initiated on insulin and metformin. The statistical modelling approaches linear mixed model and generalized linear mixed model have been compared for the analysis of fasting data and we obtained generalized linear mixed model exhibited the best fit for this data with smaller disturbance than linear mixed model for their estimated standard error.

**Keywords:** Diabetes, Fasting Glucose, GLMM, LMM, Risk Factors

---

## 1. Introduction

Diabetes mellitus is a metabolic disorder where by glucose cannot effectively get transported out of the blood. The hormone insulin facilitates glucose from the blood and into tissues, decreasing blood sugar concentration. In diabetics, insulin is not produced either in adequate amounts or the body cannot effectively respond to insulin produced, chronically high blood glucose concentration can cause damage to capillaries, inhibiting the efficiency of blood circulation. This can lead to severe ailments such as kidney disease, limb amputations, glaucoma and bacterial infection [7]. It is also a group of metabolic disorders characterised by

hyperglycaemia.

Diabetes mellitus is a catabolic multisystem disease with both biochemical and anatomical consequences. It is a chronic disease of carbohydrate, fat and protein metabolism caused by either absolute lack of insulin or insulin resistance or secretors defects. Diabetes mellitus may present with characteristic symptoms such as thirst, polyuria, blurring of vision and weight loss according to [1].

According to International Diabetes Federation, 2011 reports of, the number of adults living with diabetes in Ethiopia was 3.5% [10]. Even though the national prevalence of diabetes in Ethiopia is estimated to be 2%, evidence suggests that its prevalence could be more than 5% in those

older than 40 years of age in some setting [10]. A study by Watkins and Alemu conducted in Gondar found out most of the rural patients (77%) had Type I diabetes whereas in urban areas only 29% had Type I and 71% of them Type II diabetes [8]. Generally, the global burden of Diabetes mellitus has been increasing radically. The impact is high especially in developing countries in which resource is limited to identify the problem and develop need based clinical and community intervention [2].

Therefore, the objective of this study was to test statistical modelling in longitudinal analysis and identifies associated factors of fasting blood sugar level count of diabetic patients among outpatients of Debre Berhan referral hospital [6].

Fasting Blood Glucose assessment is a tool used to help diagnose diabetes where glucose concentration is measured using venous or capillary blood. After a period of fasting, a healthy individual would exhibit a glucose [5] concentration, between 70-100mg/dl. However, even after a period of fasting, a diabetic would exhibit an abnormally high concentration of glucose in the blood, (126 + mg/dl) providing evidence for diabetes [4]. To study the progression of fasting blood glucose level in diabetic patients, fasting blood glucose level should be measured repeatedly per individual what is called longitudinal data, since the measurements are correlated within individuals, the classical regression techniques couldn't use rather the most flexible and powerful models were employed to handle such types of data [9]. The main aim of data analysis using the linear mixed model is to define an adequate error covariance structure in order to obtain efficient estimates of the regression parameters. The statistical software now includes the covariance structure as part of the statistical model and thus the covariance matrix can be used to estimate the fixed effects of treatment and time by means of the generalized least squares method [11].

The general Objective of this study was Gaussian longitudinal analysis of progression of Diabetes mellitus Patients using fasting blood sugar level following insulin and metformin follow up and identify associated factors of fasting blood sugar level in Debre Berhan referral hospital, Ethiopia.

## 2. Data and Methodology

### 2.1. Data

All Diabetes mellitus patients who were both Type I and Type II, and placed under insulin and metformin follow up the case unit of 1<sup>st</sup> September, 2012- 30<sup>th</sup> August, 2015 G. C, in Debre Berhan referral hospital for a period of three years. The total number of patients included in this study was 248 of whom 111 (44.8%) were females and the rest 137 (55.2%) were males.

### 2.2. Methodology

The data set was a longitudinal observational study and

the data also unbalanced, since some patients do not have data until the end of the study. One of the major objectives of statistical analysis is to address variations in the data. Measurements within each subject and between-subject variation; the variation in the data between different subjects. Modelling within-subject variation allows studying changes over time, while modelling between-subject variation allows understanding differences between subjects.

The generalized linear mixed model can be fitted to longitudinal data by assuming that the vector of repeated measurements on each subject follows a linear regression model where some of the regression parameters are population-specific (fixed-effects), but other parameters are subject-specific (random-effects).

#### 2.2.1. Random Coefficient Model

Random coefficient models are useful for highly unbalanced data with many repeated measurements per subject. In random coefficient models, the fixed effect parameter estimates represent the expected values of the population of intercepts and slopes. The random effect intercept represent the difference between the intercept for the  $i^{\text{th}}$  subject and the overall intercept. The plot of the confidence intervals of the random effects (subject specific intercepts and slopes) gives an indication of what random effects structure to use in a generalized linear mixed model.

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + b_{0i} + b_{1i} Z_{ij} + \epsilon_{ij} \quad (1)$$

Where:  $\beta_0, \beta_1$  are population intercept and slope  
 $b_{0i}$  and  $b_{1i}$  are subject-specific intercept and slope

#### 2.2.2. Generalized Linear Mixed Model

The subject-specific regression coefficients reflect how the response evolves over time for each subject. Let  $Y_{ij}$  denote the count response of fasting blood sugar level for  $i^{\text{th}}$  patient at  $j^{\text{th}}$  time, taking positive integers [3].

$$Y_{ij} = X_{ij} \beta_i + Z_{ij} b_i + \epsilon_{ij} \quad (2)$$

Where:  $X_{ij}$ : Covariates of the  $i^{\text{th}}$  patient at  $j^{\text{th}}$  time

$\beta$ : Regression coefficients of  $X_{ij}$

$\mu_{ij}$ : The mean of  $Y_{ij}$ , which is related to the covariates of  $Y_{ij}$  by the link function

$Z_{ij}$ : The covariates of the random effect of the  $i^{\text{th}}$  subject at  $j^{\text{th}}$  time

$b_i$ : The random effect which are assumed to be multivariate normal distribution having mean vector 0 and covariance matrix  $G$ , i.e  $b \sim N(0; G)$

## 3. Results and Discussions

### 3.1. Results

Blood Glucose fasting blood sugar level of Diabetes mellitus patients enrolled in the case unit of 1<sup>st</sup> September, 2012- 30<sup>th</sup> August, 2015 G. C, in Debre Berhan referral

hospital for a period of three years. The total number of patients included in this study was 248 of whom 111 (44.8%) were females and the rest 137 (55.2%) were males.

**Table 1.** Descriptive Statistics of Continuous Covariates at Baselines.

Variables	Age	Baseline Fasting Sugar level	BMI
Mean	42.18	179.9	23.67
Std Dev	15.78	94.97	3.72

Depicted that the mean of age at baseline was 42.18 and a standard deviation of 15.78, the mean of baseline fasting blood sugar level of diabetic patients was 179.9 and with its standard deviation of 94.97 and the Body mass index of Diabetes mellitus patients of the mean was 23.67 and standard deviation was 3.72.

**Table 2.** Descriptive Statistics of Continuous Covariates Response by Time.

Time	Baseline	1	2	3	4	5	6	7
N	248	232	239	236	237	221	193	182
Mean of fasting	179.53	189.3	190.38	184.87	180.9	175.79	172.97	171.84
Std Dev	94.8	92.95	96.85	89.99	92.92	91.1	92.23	86.59
Maximum	600	574	600	587	522	493	589	456
Minimum	22	32	29	31	28	31	44	46
Missing value	0	16	9	12	11	27	55	66

From Table 2: the mean of fasting blood sugar level of patient's increases with an increasing rate until at time two, then decreases slowly until at time five. The largest value of standard deviation with at time two, the value was 96.85, so the number of measurements for fasting blood sugar level count showed increasing and decreasing observations between follow up times for the response indicating that the data have both intermittent and dropout missing observations and also, missing value was increasing over time.

**Table 3.** Shapiro Wilk Test of Normality.

Follow up time	Actual fasting. B			Square root Fasting. B			Logarithm of fasting. B		
	Statistics	Df.	Sign.	Statistics	Df	Sign.	Statistics	Df	Sign.
0	0.887	248	0.000	0.940	248	0.000	0.984	248	0.398
1	0.893	232	0.000	0.973	232	0.0001	0.987	232	0.079
2	0.925	239	0.000	0.981	239	0.004	0.988	239	0.074
3	0.923	236	0.000	0.962	236	0.016	0.992	236	0.518
4	0.917	237	0.000	0.972	237	0.0001	0.991	237	0.474
5	0.922	221	0.000	0.964	221	0.001	0.991	221	0.049
6	0.869	193	0.000	0.939	193	0.000	0.993	193	0.305
7	0.920	182	0.000	0.954	182	0.004	0.991	182	0.221

In any data analysis, before going to make analysis, first we have to check the assumption of data. In fasting blood sugar level of data we must using Shapiro wilk test, box plots and q-q plots, were used for checking the normality of the data. The Shapiro-Wilk test for normality is available when using the distribution platform to examine a continuous variable. The null hypothesis for this test is that the data are normally distributed. Count data can be well approximated by a normal distribution when the number of the continuous becomes large, so to normalize the data using logarithmic and square root transformation were carried out.

Now identify which condition is more to normal approximation satisfied. Now this study was found that the actual fasting blood sugar level count were not normal at all-time points as the test showed significant deviation from normality. Likewise, the square root fasting blood sugar level was not normal by the Shapiro wilk test. But, the test approves normality of logarithm transformation fasting blood sugar level at all-time were normality satisfied and also when to compare the significance level at each time point greater than 0.05, except at time five And also, the q-q plots of the

overall of the data signifying at logarithm transformation of the data to satisfied normality by follow up time.

The box plot shows that, outlier observations observed were minimized in the logarithm of fasting blood sugar level of patients than square root transformed and actual fasting blood sugar level counts of patients. This that means that in logarithm of fasting blood sugar level there was small outlier presented to compare to the other, so logarithm of fasting blood sugar level is the most appropriate for this study.

The average evolutions fasting blood sugar level counts of patients initiated on insulin and metformin over time. This study was observed that there is no presence a sharp linear increase at each time points. The mean profiles of fasting blood sugar level of patients of male patients is higher than female at baseline, this is same at time 2. Generally they does not shows linear growth pattern in both groups. The average loess smoothing fasting blood sugar level profile plot by marital status, it has been observed that the mean profile for the marital status of patients does not show linear growth pattern in married, but the mean profile for single patients seems like linear pattern.

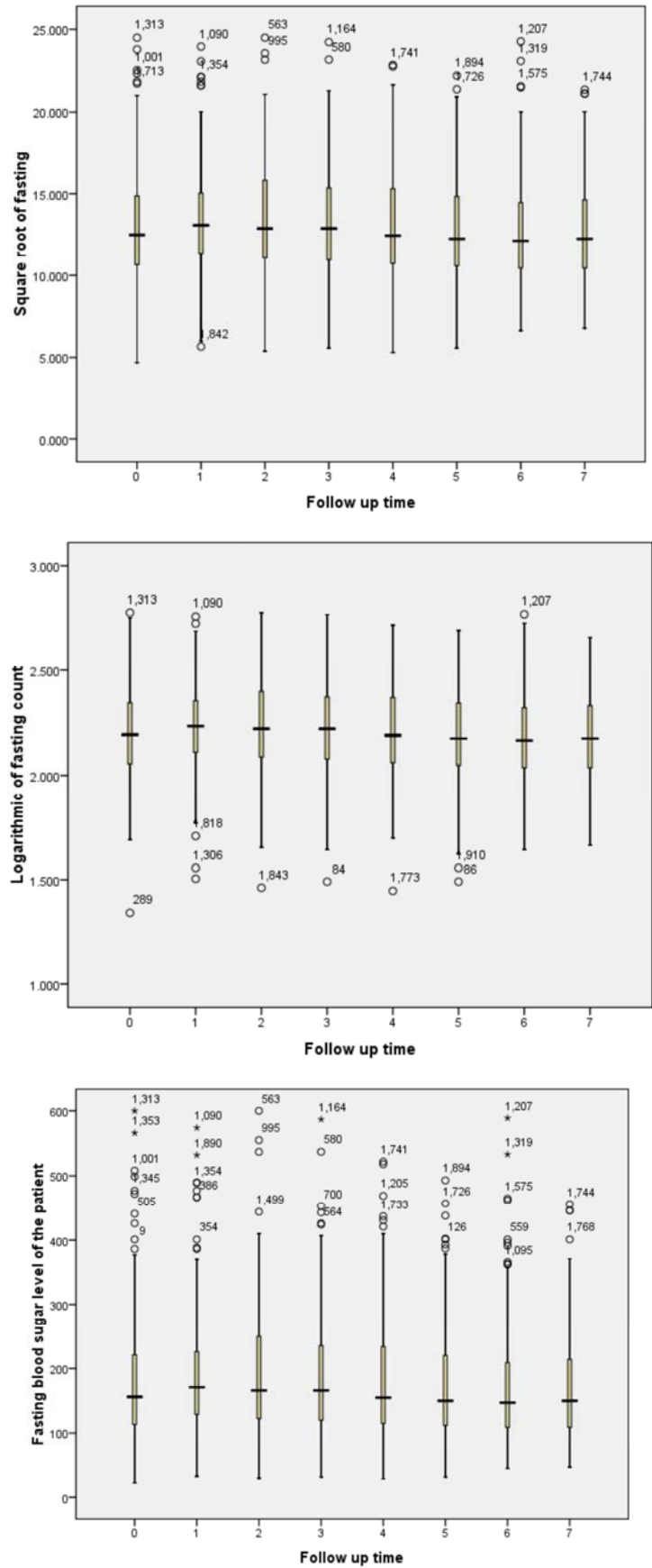
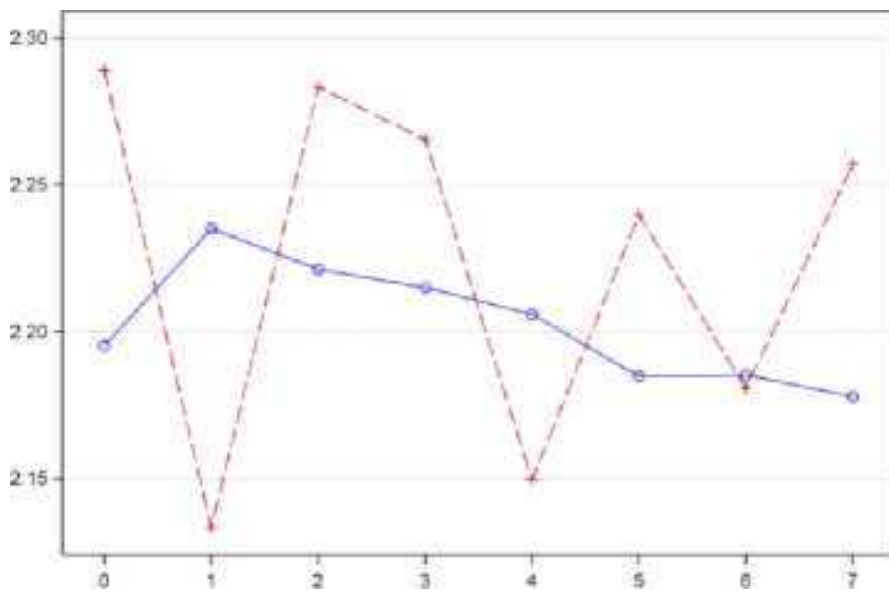
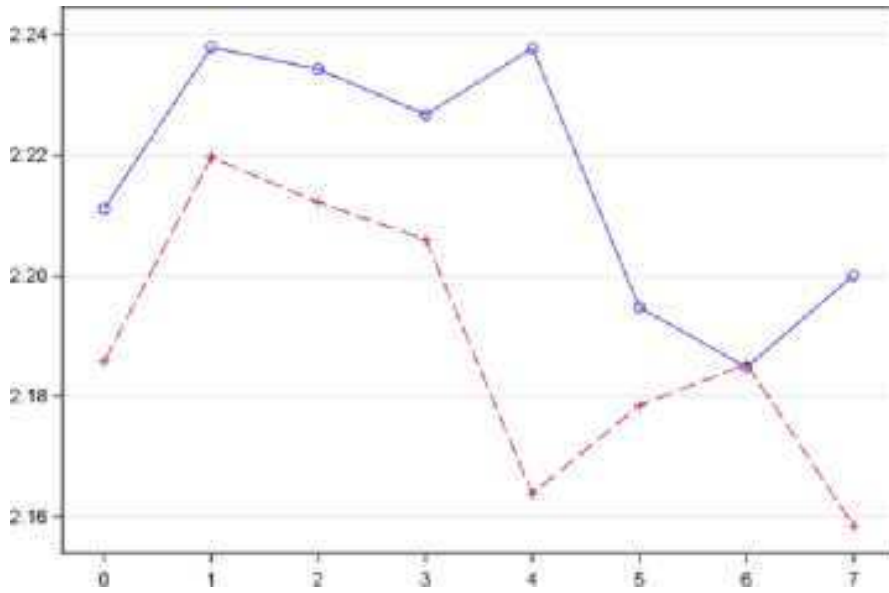
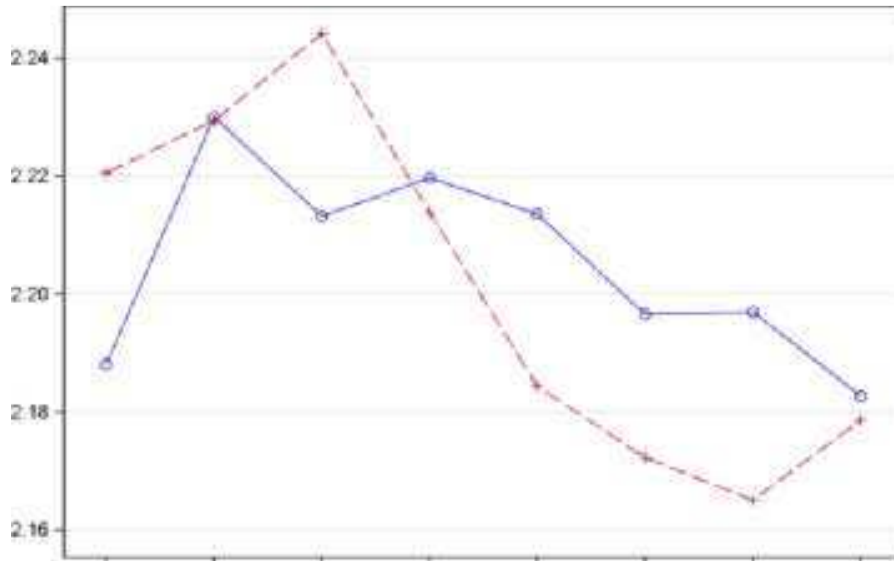


Figure 1. Boxplots of Fasting Blood Sugar Level Count by Follow up Time.



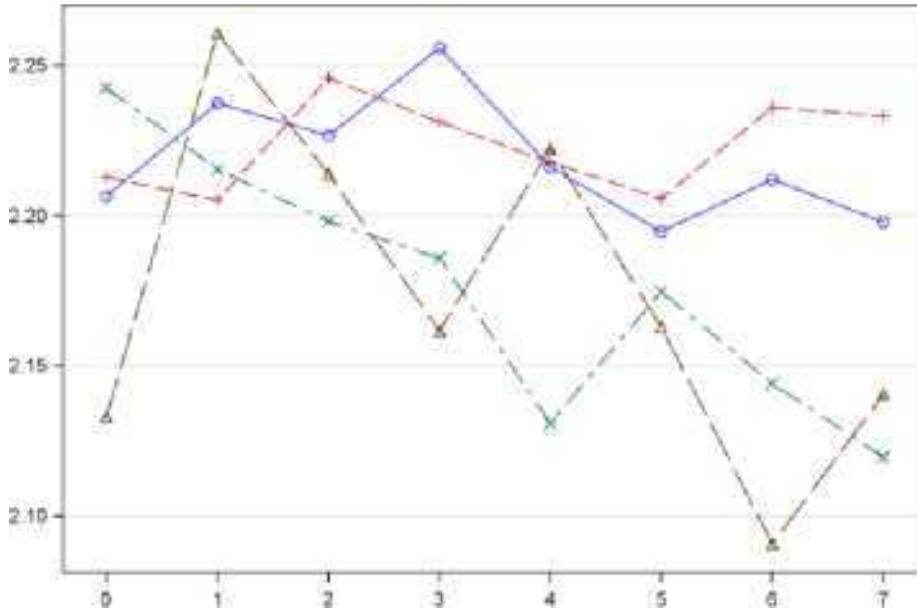


Figure 2. Mean Profile Plots of Fasting Blood Sugar Level for Categorical Covariates.

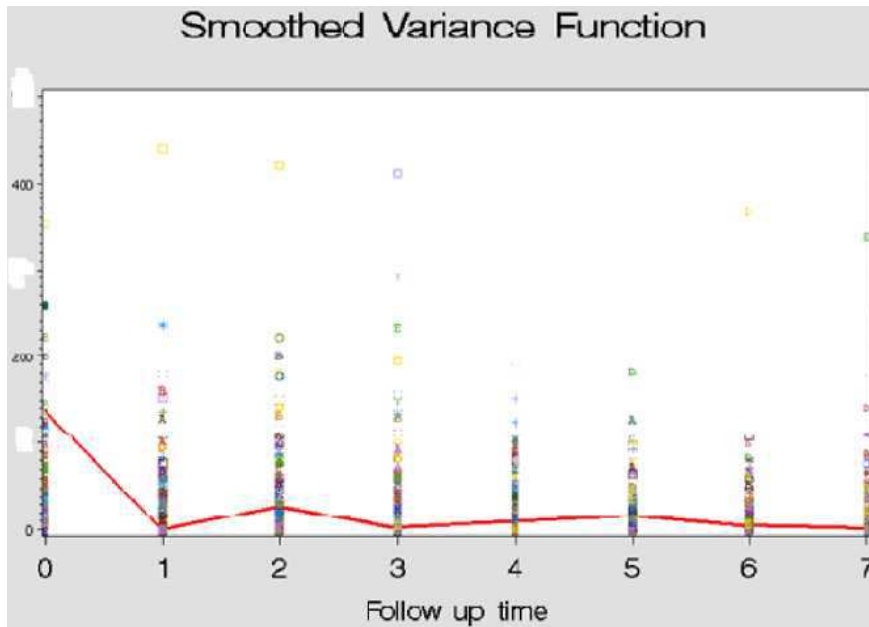


Figure 3. Variance Function Structure of Fasting Blood Sugar Level.

From Figure 3: Explored that the variances of patients fasting blood sugar level count, shows that the observed variance function, which is changing over time, this implies that, the need for other random effects in addition to the random intercepts would be needed in modelling patients of fasting blood sugar level over time. And also it shows that the variability of the observation is not constant over time. Generally, the variance of patients fasting blood sugar level count over time fluctuates from time to time in different covariates.

Table 4. Generalized Linear Mixed Model for both Fixed and Random Effect Models.

Random Coefficient	Estimate
$b_{11}$	0.0425
$b_{21}$	-0.00027
$b_{22}$	0.0076
$\delta^2$	0.0384

Table 4. Continued.

Effect	Estimate	S. E	P-value	95% CL	
				Lower	Upper
Intercept	2.356	0.1367	0.0001	2.088	2.624
Base fasting	-0.000001	0.00001	0.8969	0.000	0.0000
BMI	0.0015	0.0041	0.7099	-0.0065	0.0095
Age	0.0016	0.0008	0.0405	0.0001	0.0031
Time	0.0036	0.0104	0.0001	0.0167	0.0241
Male	0.0524	0.0232	0.024	0.0069	0.0979
Work	-0.0068	0.0292	0.8156	-0.064	0.0504
Married	0.0069	0.0225	0.7561	-0.0372	0.051
Illiterate	0.0015	0.0383	0.683	-0.0736	0.0766
Primary	0.0188	0.0334	0.5723	-0.1147	0.0261
Secondary	0.0443	0.0359	0.2181	-0.0033	0.0015
Urban	-0.0484	0.0291	0.0962	-0.1054	0.0086
Occupation	0.0089	0.0185	0.6277	-0.0274	0.0452
Clinical D	-0.0084	0.0285	0.7679	-0.0001	0.0275
Time *time	-0.0017	0.00096	0.0001	0.0004	0.0036
Illiterate*time	0.0138	0.007	0.0504	-0.0295	-0.0025
Primary*time	-0.016	0.0069	0.021	-0.0206	-0.0076
Secondary*time	-0.0065	0.0072	0.3642	-0.0123	0.0351
Age*time	-0.0003	0.0002	0.0372	0.0235	0.456
Urban*time	0.0237	0.0058	0.0188	-0.0008	-0.0005

From the above Table 4: Emphasized that the covariates for each individual values of the log fasting blood sugar level count at baseline rather than the other covariates of individual measurements. The intercept represents the estimated mean for the last level of log fasting blood sugar level counts. The variables sex of male is about  $\exp(0.0524) = 1.0537$  was higher than that of female (the reference group)  $p = 0.0240$  which coincides with the mean profile plot Figure 2: and the independent variables, address with time interaction of patients was higher by  $\exp(0.0237) = 1.0239$ , compared to the references group of urban with time interaction patients  $p = 0.0188$ , the independent variables for age of patients are a significant effect of log fasting blood sugar level and also the independent variables for sex of male patients are a Significant effect at 5% level of significant. Now to compare for the above Table 4, the only fixed effect of log fasting blood sugar level that means there is no random effect value was included. After included the random effect independent variables, quadratic of time there are slightly difference for the parameter estimate of the covariates and their significance, but also the parameter interpretation is different. Because, in linear mixed model the marginal (population average) interpretation of the independent variables, whereas for generalized linear mixed model for the subject Specific interpretation. Moreover, age, sex, educational status with time, address with time and age with time interaction were shows significant covariates in generalized linear mixed model. Lastly, this study was conducted PROC MI for analyzed missing value for identifying missing mechanism. The Generalized estimate equation approach assumes missing completely at random while PROC MIXED in generalized linear mixed model assumes missing at random. Therefore, the missing value assumption is more restrictive in PROC GENMOD compared to PROC MIXED. MI is an

alternative to direct likelihood and the MI value of some of the covariates that were significant.

### 3.2. Discussions

The result of this study has indicated that the education level of patients is a key determinant of fasting blood sugar level. The result obtained in this study showed that illiterate and primary educational level of patients is a significant effect ( $p=0.0504$  and  $0.021$  respectively) interaction with time in generalized linear mixed model. Education exposed to information empowers to makes them more aware of their own health for fasting blood sugar level. Therefore, supports the research findings of [8] who found that the risk of diabetes decreases with increasing education level. From the final model results of generalized linear mixed model; age, sex, time by address and time by age interactions ( $p=0.0405$ ,  $0.024$ ,  $0.0188$  and  $0.0372$  respectively) were significant independent variables for fasting glucose level count progression at 5%.

## 4. Conclusions

This study evaluated the relationship between the progressions of diabetic infection using longitudinally measured fasting blood sugar level count and its possible covariates using longitudinal analysis methodologies.

The results of this study showed that the main effect independent variables age, male, urban and time and the interaction effects of age, illiterate, primary, secondary and urban with time are significantly associated with the progression of fasting blood sugar level count over time in linear mixed model. In generalized linear mixed model the results showed that the main effect independent variables

age, male and time, and the interaction effects of illiterate, primary, urban and age with time are significantly associated with the progression of fasting blood sugar level over time.

---

## References

- [1] DIABETES ATLAS, sixth edition. (2013). *International Diabetes Federation*. DIGGLE, *et al.*, (2009). *Analysis of Longitudinal Data*. 2nd ed. Oxford: Oxford University Press.
- [2] D. HEDEKER and R. D., (2006). Gibbons, *Longitudinal Data Analysis*, Wiley-Inter science, New Jersey.
- [3] DONALD R. HEDEKER and ROBERT D. GIBBONS., (2006). *Longitudinal Data Analysis*. J. Wiley and Sons.
- [4] D. R. WHITING, L. GUARIGUATA, C. WEIL, and J. SHAW., (2011 and 2030). IDF Diabetes Atlas, Global Estimates of the Prevalence of diabetes, *Diabetes Research and Clinical Practice*, vol. 94, no. 3, pp. 311–321.
- [5] G. M. FIZMAURICE, *et al.*, (2004). *Applied Longitudinal Analysis*, Wiley Inter-science, New Jersey.
- [6] T. A. HARRISON, *et al.*, (2003). Family History of Diabetes as a Potential Public health tool, *American Journal of Preventive Medicine*, vol. 152{15924, no. 2, pp.}.
- [7] CDC., (2010). Healthy People; National Center for Health Statistics Healthy People Stat Notes U. S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, (2001-2011). Retrieved from.
- [8] HAMIL M. *et al.*, (1995). Diabetes in America. 2<sup>nd</sup> Edition. National Diabetes Data Group, 179-203.
- [9] MCKINLY J. *et al.*, (2010). The Social Construction of Race/Ethnic Disparities in Diabetes-A Case of Misplaced Concrete New England Journal of Medicine.
- [10] ZEIN, ZA. *et al.*, the Ecology of Health and Disease in Ethiopia.
- [11] VERBEKE, G. and MOLENBERGHS, G., (2000). *Linear Mixed Models for Longitudinal Data*. New York, Springer Verlag. No. 2, pp. 31-40.