International Journal of Current Research and Modern Education (IJCRME)
Impact Factor: 6.925, ISSN (Online): 2455 - 5428
(www.rdmodernresearch.com) Volume 4, Issue 1, 2019

NON 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* & Dr. A. R. Muralidharan**
Department of Statistics, College of Natural and Computational Science, Debre Berhan University, Debre Berhan, Ethiopia

Cite This Article: Wudneh Ketema Moges & Dr. A. R. Muralidharan, "Non Gaussian Longitudinal Analysis of Progression of Diabetes Mellitus Patients Using Fasting Blood Sugar Level: A Case of Debre Berhan Referral Hospital, Ethiopia", International Journal of Current Research and Modern Education, Volume 4, Issue 1, Page Number 46-50, 2019.

Abstract:

Diabetes mellitus is a group of metabolic diseases characterized by hyperglycaemia resulting from defects in insulin secretion, insulin action, or both. 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 1st September, 2012 to 30th August 2015 in Debre Berhan referral hospital. The main objective of the study was Non 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 estimating equation would be used to estimate the progression of diabetic infection. The most appropriate working correlation structure was exchangeable correlation selected for this study. This study showed that age, male, primary, urban and the interaction effects of primary and age with time was statistically associated with the progression of fasting blood sugar level over time. Moreover, on average fasting blood glucose decreases in a quadratic pattern over time after patients initiated to insulin and metformin program. Finally we obtained generalized estimating equation model exhibited the best fit for this data with smaller disturbance for their estimated standard error.

Key Words: Diabetes, Fasting Glucose, DM & GEE Risk Factors

1. Introduction:

Diabetes mellitus is a group of metabolic diseases characterized by hyperglycaemia resulting from defects in insulin secretion, insulin action, or both. 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 [5]. 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% [8]. 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 [8]. 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 [6]. 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 [4].

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 [3] 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 [7]. 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 [9].

The general Objective of this study was Non 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 1st September, 2012- 30th August, 2015 G.C, To categorized fasting blood sugar level, that means under normal condition and below normal condition which was used to assess whether good control of fasting blood sugar level over time 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. But in this case the response variables are categorized, approaches were proposed to tackle this problem .i.e. GEE method was proposed.

2.2.1 Generalized Estimating Equations (GEE): Generalized Estimating Equations were introduced by [10] as a method of dealing with correlated data. The GEE approach is the most popular method seen in marginal models. GEE is an extension of GLM for the analysis of longitudinal data .In this method, the correlation between measurements is modelled by assuming a working correlation matrix. Estimating the correct working correlation matrix provides efficiency parameter estimates. Using the GEE method are marginal models that only estimate population average regression coefficients. They have consistent and asymptotically normal solutions by relying on the independence across subjects to estimate constantly the variance of the regression coefficient even when the assumed correlation structure is incorrect [12].

GEE analysis is generally valid only when the data are missing completely at random (MCAR) and it gives a biased estimator of the regression parameter in the mean model.

The Marginal Mean Model: Assumed that N patients measured repeatedly through time and let denote the response for i^{th} patient at j^{th} time y_{ij} is count response variable with non-negative integer values .The mean is related to X by a log link function.

$$g(\mu_{ij}) = \text{Log}(\mu_{ij}) = X_{ij}\beta \text{ .... } (1)$$

Where: μ_{ij} :The mean of $Y_{ij,}$ which is related to the covariates of X_{ij} by the link function

X_{ij}: Apx1vector of covariates

β: Apx1 vector of unknown regression coefficients of X, and

g(.): Log link function as Y_{ij} is count

Working Correlation Structures: GEE estimator for the regression parameter will be the most efficient if the working correlation matrix is correctly specified. Hence it is desirable to choose a working correlation matrix that is the closest to the underlying structure among a set of working structures. With GEE, this correction is carried out by assuming a priori certain working correlation structure for the repeated measurements of the outcome variable Y. Before carrying out a GEE analysis, the within-subject correlation structure was chosen based on the results of exploring correlation structure of the observed data. Accordingly two propose working correlations were compared.

I. **Independent Structure:** This is the correlation that GEE model assumes by default. With this structure the correlations between subsequent measurements are assumed to be zero or measurements are independent to each other within individuals.

II. Exchangeable Correlation Structure (Compound Symmetry): It assumes the correlations between subsequent measurements are assumed to be the same, irrespective of the length of the time interval. Generally, assuming no missing data, the JxJ covariance matrix
for y is modeled as:

$$V_i = \emptyset A_2^1 R_i A_2^1 \text{.. } (2)$$

Where $\emptyset$ is a GLM dispersion parameter which is assumed 1 for count data, A_i is a diagonal matrix of variance functions, and R_i is the working correlation matrix of Y.

2.2.2 Method of Estimation and Statistical Inference: Estimation is more difficult in the mixed model than in the general linear model. This is because in mixed model estimation of random effects and covariance structure of the random error is necessary besides to the fixed effect. Both the ML and REML were used for estimation of the parameters in this study. The ML method finds the parameter estimates that are most likely to occur given the data. Both are based on the likelihood principles, which have the properties of consistency, asymptotic normality, and efficiency. The difference between the two likelihood principles is:

- REML handles strong correlations among the responses more effectively
- The difference between ML and REML estimation increase as the number of fixed effects in the model increases .Instead inference can only use Wald statistics constructed with asymptotic normality of the estimators together with their estimated covariance matrix.

3. Results and Discussions:

3.1 Results: Fasting Blood Sugar level of Diabetes mellitus patients enrolled in the case unit of 1st September, 2012- 30th 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. The aim of exploratory data analysis is to understand the data structure and determine the relevant modelling approaches suitable for the data.

Table 1: Descriptive Statistics of Continuous Covariates at Baselines

Variables	Age	Baseline Fasting Sugar Level	BMI
Mean	42.18	179.9	23.67
StdDev	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

4. Estimated Coefficients for GEE models for Fasting Data Taken in Debre Berhan Hospital, Ethiopia:

Parameter	Exchange		Independent		AR(1)	
	EST	S.E	EST	S.E	EST	S.E
Intercept	0.7702	0.0502	0.7688	0.0344	0.7627	0.0411
Baseline fasting	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BMI	-0.0002	0.0014	-0.0012	0.0014	0.00001	0.0013
Age	0.0007	0.0004	0.0007	0.0003	0.0007	0.0004
Time	-0.0036	0.0034	-0.0045	0.0035	-0.0043	0.0042
Male	0.0175	0.0080	0.0173	0.0049	0.0173	0.0067
Work	-0.0039	0.0110	-0.0041	0.0111	-0.0047	0.0111
Single	0.0027	0.0160	0.0026	0.0074	0.0033	0.0089
Illiterate	0.0036	0.0153	0.0025	0.0135	0.0046	0.0156
Primary	0.065	0.0141	0.0040	0.0126	0.0086	0.0141
Secondary	0.0188	0.0150	0.0177	0.0135	0.0222	0.0155
Urban	-0.0237	0.0117	-0.0244	0.0105	-0.0260	0.0122
Fulltime	0.0028	0.0080	0.0030	0.0056	0.0032	0.0067
Type I	-0.0014	0.0111	-0.0004	0.0078	0.0002	0.0098
Illitratextime	0.0062	0.003	0.0068	0.0033	0.0062	0.0033
Primaryxtime	0.0058	0.0030	0.0068	0.0031	0.0057	0.0030
Secondaryxtime	0.0039	0.0031	-0.0032	0.0032	-0.0041	0.0033
Urbanxtime	0.0065	0.0018	0.0069	0.0026	0.0072	0.0026
Agextime	-0.0001	0.0001	-0.0001	0.0001	-0.0001	0.0001

From the above Table 4: In GEE parameter estimate, the empirical standard error estimates are robust estimates that do not depend on the correctness of the structure imposed on the working correlation matrix. The model-based standard error estimates are based directly on the assumed correlation structure. The model-based standard errors are better estimates if the assumed model for the correlation structure is correct [13] and also the model based standard errors are obtained using the models option on the repeated statement. The GEE analysis estimates the scale parameter in empirical standard error estimates, but the scale parameter is not required with the GEE analysis this were shown in the (appendix B, Table A4, Table A5. Table 4.8, the intercept exp(0.7702) = 2.1601; is an estimate of the mean fasting blood sugar level ,at base line(Time=0)for which is significantly different from zero (p = 0:0001).

The estimated value for sex of male patients, exp(0.0175) = 1.0176; p = 0.0551,this implies that the mean of fasting blood sugar level for males 1.0176 higher than for the reference group. Now their difference was highly significant at 5% level of significance. Moreover, at base line the mean of fasting blood sugar level among working patients were 0.9961; p = 0:7353 times lower than the mean fasting blood sugar level among the ambulatory patients patients (reference group). The education levels of illiterate patients are 1.0036; p = 0.862 times higher than the other educational level groups, and also the primary and secondary patients were 1.0671; p = 0.0293 and 1.0189; p = 0.8315 counts per month higher than the rate of increase among subjects in the tertiary respectively, but there is no significant effect in the variable for illiterate and secondary of patients, but also the primary of patients are significant effect at 5% level of significance. From the above discussion depends on the GEE model for the exchangeable working correlation equation because the selected working correlation.

3.2 Discussions: In GEE model the appropriate working correlation structure was exchangeable correlation structure (compound symmetry) selected based on exploratory analysis result and independence correlation simply taken for the sake of comparison, for GEE model were compared in this paper and found that exchangeable working correlation structure fits the fasting data better than independence and AR(1). This study also compared LMM and GEE models using their standard error estimates ratio [11] and obtained GEE fits better than LMM with a small disturbance provided their marginal interpretations. Which support the findings of [11] .The term marginal means that in the model specification the expected value of the response variable log fasting blood sugar level, depends only on covariates (fixed effects) and does not depend on subject specific random effects nor directly on previous responses of the subject. Since the purpose is to describe the changes in population mean rather than changes within-subject correlation is regarded as a nuisance characteristic in GEE model [3].

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.

5. 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. G. M. Fizmaurice, et al., (2004). Applied Longitudinal Analysis, Wiley Inter-science, New Jersey.
4. 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.}.
5. 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
6. Hamil M.et al.,(1995).Diabetes in America. 2nd Edition. National Diabetes Data Group, 179-203.
7. Mckinly J.et al.,(2010).The Social Construction of Race/Ethnic Disparities in Diabetes-A Case of Misplaced Concreteness New England Journal of Medicine.
8. Zein, Za.et al., the Ecology of Health and Disease in Ethiopia.
9. Verbeke, G. and Molenberghs, G., (2000).Linear Mixed Models for Longitudinal Data. New York, Springer Verlag. No.2, pp.31-40.
10. Liang. K and Zenger, L., (1986). Longitudinal data Analysis for discrete and continuous outcomes .Biometrika, 42, 121-130.
11. Molenbergh G, Kenwards MG. (2007). Missing data in clinical studies.
12. L.D.Nshisso, et al., (2012).Prevalence of hypertension and diabetes among Ethiopian adults , Diabetes and Metabolic Syndrome, vol.6,no.1 pp.36-41.
13. Allison, P. D. (2009). Fixed Effects Regression Models (Quantitative Applications in the Social Sciences). SAGE.