Vol. 10(6), pp. 171-182, June 2018
DOI: 10.5897/JPHE2017.0987
Article Number: 4121F6157003
ISSN 2141-2316
Copyright © 2018
Author(s) retain the copyright of this article
http://www.academicjournals.org/JPHE

Journal of Public Health and Epidemiology

Full Length Research Paper

A longitudinal data analysis on risk factors for developing type-2 diabetes mellitus at the University of Gondar Comprehensive Specialized Hospital, Gondar, Ethiopia

Asrat Atsedeweyn Andargie[1*] and Melkamu Ayana Zeru[2]

[1]Department of Epidemiology and Biostatistics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia.
[2]Department of Statistics, Haramaya University, Harar, Ethiopia.

Received 10 February, 2018; Accepted 28 March, 2018

Diabetes mellitus is a group of metabolic chronic diseases characterized by high blood sugar levels with multi-system complications. The objective of this study was to assess the risk factors for developing type-2 diabetes mellitus over time in their random blood sugar and to obtain better predictive model for type 2 diabetes patients' random blood sugar (RBS) level in the University of Gondar Comprehensive Specialized Hospital. A retrospective cohort study with a total of 330 diabetic patients who have been active in the follow-up treatment for at least 3 times in three month interval in the hospital from February 2014 to February 2016 was conducted. Linear mixed effects model for longitudinal data were employed to measure the changes in RBS level. The results revealed that the linear distribution trend in the mean RBS level accounted for 79% of the variability in the data and the mean RBS level decreased over time. Age, residence, family history, alcohol intake, dietary type, BMI, treatment, exercise and education status were the significant factors for the change in mean RBS level of the diabetes patients over time. The study also confirmed that among the factors of RBS level included in the study, meat dietary type, patients who do not perform exercise, and body mass index (BMI) were positively correlated with the RBS level while the rest were negatively correlated. It was significant for the patients to do daily self-care activities to prevent long term complications. The government should also contribute to the education of communities to spread awareness creation and enhance prevention mechanisms of diabetes.

Key words: Diabetes mellitus, blood sugar, linear mixed effect, longitudinal data analysis, random blood sugar (RBS).

INTRODUCTION

Diabetes mellitus is a group of metabolic diseases characterized by high blood sugar levels that result from

*Corresponding author. E-mail: asrat07@gmail.com.

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

defects in either insulin secretion or its action (WHO, 2017). Diabetes mellitus is one of a number of chronic illnesses with multi-system complications. Diabetes is a chronic disease for which control of the condition demands patient self-management (MacPherson et al., 2004). Self-management of diabetes requires time and monitoring of blood-glucose levels. Taking control of diabetes to improve quality of life has put the spotlight on the need for additional support and education for patients with diabetes. Although new treatments and technology have aided in controlling the disease, diabetes can still influence every day social interactions in many ways.

The patient must be aware the types and amounts of food they ingest, they would have to monitor their blood glucose levels at specific times during the day, and medication would be necessary at times when the individual is engaging in social activities (Thorne et al., 2003). The current classification of diabetes is based upon the pathophysiology of each form of the disease. Type-1 diabetes results from cellular mediated autoimmune destruction of pancreatic *β*-cells, usually leading to loss of insulin secretion (Thorne et al., 2003).

Type-2 diabetes results from insulin resistance, which alters the use of endogenously produced insulin at the target cells. Type 2 patients have altered insulin production as well. However, autoimmune destruction of *β*-cells does not occur as it does in type-1, and patients retain the capacity for some insulin production. Because the type-2 patient still produces insulin, the incidence of ketoacidosis is very low compared to type 1, however, ketoacidosis can occur in association with the stress of another illness such as infection (Klinke, 2008). Type-2 patients can go undiagnosed for many years because the hyperglycemia appears gradually and often without symptoms. It is often accompanied by various chronic complications that may affect the productivity and quality of life (Hennekens, 1998). Nowadays, diabetes mellitus is becoming the leading cause of blindness, non-traumatic amputation, and chronic renal failure in the western world (Laing et al., 1999). Globally, the population of people affected by type 2 diabetes was 15.1 million in 2000 (IDF, 2013). The number of people with diabetes worldwide was projected to increase to 36.6 million by 2030 (CDC, 2007b). In 2007, it was indicated that 23.6 million people or 7.8% of the United States population had type 2 diabetes. People with diabetes have an increased risk of developing a number of serious health problems. Consequently, the economic and medical consequences of complications arising from diabetes is high blood glucose levels that can lead to serious diseases affecting the heart and blood vessels, eyes, kidneys, nerves and teeth. In addition, people with diabetes also have a higher risk of developing infections. In almost all high income countries, diabetes is a leading cause of cardiovascular disease and kidney failure (Kahn et al., 2006).

Like the rest of the world, sub-Saharan African countries are experiencing an increasing prevalence of diabetes along with other non-communicable diseases (WHO, 2008). In 2010, it was indicated that 12.1 million people were estimated to be living with diabetes in Africa, and projected to increase to 23.9 million by 2030 (Shaw et al., 2010).

In addition, the International Diabetes Federation (IDF) estimated that 19.8 million people have diabetes in Africa and approximately 75% are still undiagnosed (IDF, 2013). Countries with the highest estimated numbers of persons with diabetes include Nigeria (3.9 million), South Africa (2.6 million), Ethiopia (1.9 million), and Tanzania (1.7 million) (IDF, 2013). Type 2 diabetes contributes up to 90% of the cases (Feleke and Enquselassie, 2007). This spike is due to an aging population and lifestyle changes associated with rapid urbanization and westernization. Because of the high urban growth rate, unhealthy dietary changes, reduction in physical activity and increased obesity it is estimated that the prevalence of diabetes is going to triple within the next 25 years (IDF, 2013). Diabetes is common in Ethiopia but the incidence and prevalence of the disease is not well known ii the society. In recent studies, accesses to blood glucose monitoring and diabetes health education were found to be very low but overall burden of the disease in the country. These studies have a lack of comprehensiveness due to small sample sizes because most of them were limited to the capital city, Addis Ababa. The cost of inpatient diabetes management in the country is a high amount being significantly higher than the cost of other inpatient management categories (Feleke and Enquselassie, 2007). However, diabetes in Ethiopia has never been given the attention it deserves. Glycemic control and management of co-morbid conditions along with diabetes complications are alarmingly sub-optimal and perhaps one of the worst in the world (Abera, 2000).

According to the IDF report, in Ethiopia, about 1.9 million adults aged between 20 and 79 years were estimated to have diabetes in 2013. In addition to this, 2.9 million people living with impaired glucose tolerance are at higher risk of developing diabetes. With national diabetes prevalence of 4.36%, there was around 34,262 estimated diabetes related deaths in the same year.Presently, the incidence and prevalence of the disease has increasing in society. It is evident that, few studies have shown a significant increase in its prevalence over the last four decades. However, diabetes in Ethiopia has never been given the attention it deserves, despite the fact that the rate of incidence and prevalence has been increasing over time. The overall disease burden in the country is unknown because of the limited studies in the country (Abera, 2000). The burden of diabetes also presents a crisis in terms of health care costs, both direct and indirect, ranging from individual to national economy. According to International Diabetes Federation, an estimated average cost in USD was 1,437 per person with diabetes was spent globally on treating and managing the disease in 2013. Even though health professionals try to control random blood sugar levels, there are many questions which can be raised by individuals. For example, how is the change of RBS level over time? or does the change of RBS level have different patterns on different factors? what are the factors for controlling blood sugar levels of diabetes patients? In Ethiopia, there has not been known and well-developed longitudinal research conducted to know the mean evolution of random blood sugar. In addition, research has been conducted to assess whether blood sugar levels changed over time or whether change in blood sugar control varied by covariates. The empirical risk of having type 2 diabetes increases from 2 up to 6 fold, if a parents or siblings have the disease. Consequently, a positive family history is a practical, albeit a crude way, of figuring out if an individual is likely to have inherited susceptibility to the disease. On the other hand, familial aggregation may occur for non-genetic reasons. Family members often share a similar environment, particularly as children and in adulthood, thus familial aggregation alone is not definitive evidence of genetic determinants. Furthermore, with a disease as frequent as type 2 diabetes two or more family members may well have the disease by chance alone (Helgeson and Gottlieb, 2000).

A longitudinal mixed model project on a nurse-based diabetes management system from San Diego, California showed that multiple variables are associated with glycemic control. The project that collected information from July 18, 2000 to October 7, 2002 focused on a database containing demographics, health status, treatment, laboratory, and behavioral factors for each patient. Age, race, disease duration, medication, number of visits, total cholesterol, BMI, alcohol intake, dietary and insurance status were all significant. However, after controlling baseline A1C, time, other demographic implications, and disease severity factors; only age, insurance status, disease duration, pharmacotherapy, and total cholesterol were significant in the final model contributing significant effects (Philis-Tsimikas and Walker, 2001). The prevalence and incidence of type 2 diabetes vary to some extent between the sexes from one population to another. A longitudinal study of type-2-diabetes with the linear mixed effect model conducted in Ghana reported that, the change in RBS value for males and females were not the same (Timothy et al., 2015*)*. However, a study conducted in Japan states that both sexes have equal chance of being affected by diabetes. Therefore, being male or female may not have effects in the prevalence of type 2 diabetes (Baltazar et al., 2004). Another longitudinal study on type 2 DM reported that significant difference in hyperglycemia between insulin plus Oral Hyperglycemia Agents (OHA) and insulin ($p = 0.0321$) combinations contributes to better results (Yki-Jarvinen, 2001). The available literature provides little data on the quality of life of patients with type 2 diabetes depending on place of residence. Rural diabetics experience significant impairments or damages in their health or quality of life in relation to urban diabetics. A higher proportion of obesity was found among rural residents compared to urban residents. Since higher obesity is closely linked and the main cause of developing type 2 diabetes, rural patients have a higherrisk of developing the disease. In a study conducted among patients living in the Lublin Province, people with diabetes living in rural areas were more likely to perceive the use of insulin as being burdensome and believe that diabetes has hostile impact on their family life because of the stress than residents of urban areas (Graham et al., 1999). A report that compared the type 2 diabetes in adults and young adults in Europeans opposed to Native American tribes, Mexican Americans, African Americans, Chinese, Polynesians, Asian Indians and Arabs of the Gulf States showed that Europeans who are adults have a higher risk of developing type 2 diabetes than younger citizens. It was clear that young adults are better at controlling their blood sugar levels (Taylor and Lobel, 1989).

The nurses' health study suggests that the risk of type 2 diabetes among Europeans increases even within the normal BMI range. It also states that a BMI of 21 kg/m^2 might be an optimum level and that patients with type 2 diabetes had a higher prevalence with a large BMI because an overweight population has a greatest risk of developing diabetes (Tyler and Blader, 2001). A study has been carried out in the United States of America in 2001 when Miller et al. (2002) evaluated the impact of education intervention on blood glucose levels for 92 type 2 diabetic patients who were older than 45 years of old. Patients were put into literate and illiterate groups and introduced to a ten week follow up in life style changes, physical exercise, and proper dietary use. When the patients were evaluated, the literate group showed a greater improvement in fasting plasma glucose ($p=0.05$) and glycosylated hemoglobin ($p< 0.01$) than the illiterate group. So it is clear that the illiterate diabetic patients needed additional education to achieve metabolic control to reduce fasting (random) blood sugar levels and avoid mortality associated with diabetes (Miller et al., 2002).

The main objective of this study was to assess the risk factor for developing type-2 diabetes mellitus over time using the random blood sugar in a follow up study at University of Gondar comprehensive specialized hospital during treatment period of 2 years and to determine their relationship.

MATERIALS AND METHODS

Data description and study design

This study was a retrospective cohort study based on data from diabetic patients. The data used in this study were obtained from University of Gondar Comprehensive Specialized Hospital,

Table 1. Covariates used in the Linear mixed effect Model Analysis for DM Data.

S/N	Variable	Description
1	Gender	1=Female, 0= Male
2	Age	Year
3	Marital status	0 = Divorced, 1 =Married, 2 = Widowed, 3= Single
4	Family history	0 = no family history, 1= has family history
5	Place of residence	0 = rural, 1 = urban
6	Educational status	0 =illiterate, 1 =literate
7	Treatment	0 = insulin, 1 =OHA, 2= combination of both
8	Dietary type	0 =fruit, 1= vegetable, 2 = meat, 3 = others
9	Body mass index	Kilogram per meter square
10	Exercise activity	1 = do not perform exercise, 0= perform exercise
11	Alcohol intake	1 = drinks alcohol drinker, 0= does not drink alcohol

Ethiopia. The data comprised longitudinal measurements of diabetes mellitus type 2 (DMT2) risk factors. The most commonly used test of diabetes is random fasting blood glucose. The test is conducted before eating at least for 8 h, usually overnight. The target populations of this study were all diabetic patients who attended the Hospital and had been active in follow-up treatment for their diabetes for at least 3 times in three month interval in the hospital from February, 2014 to February, 2016. The patients also had a minimum of three and a maximum of eight repeated measurement values. The number of measurements in the data may not be equal for all patients due to the difference in the duration of the follow-up. All the patients of this study were those whose age was greater than 20 but less than 70 years. The patients under 20 and over 70 were not included in the study. Since the life expectancy in Ethiopia is low, those patients over 70 are difficult to include due to their scarcity. Since, there are very few individuals in that age range their contribution would have been minimal.

In this study, 1,456 observations were considered to collect the random blood sugar which was evaluated at fixed time interval of 3 months. Measurements of all the patients were taken at 3, 6, 9, 12, 15, 18, 21, and 24 months which had an equal time interval with 3 months between all measurments. The random blood sugars (RBS) of the patients ranged from 78 to 600 mg/dl, with their mean, median and standard deviation values of 216.8 and 191.5 mg/dl respectively with a standard deviation of 83.35198 mg/dl. Individuals are not only said to be diabetic patients with large levels of RBS, but also for small levels. RBS level higher or lower than the standard value was considered as diabetic. Several potential explanatory variables were also considered in this study. The descriptions of these covariates are presented in Table 1.

Out of the total 1456 patients included in the study, 735 (50.5 %) were females. More than half of the patients (882, 60.6%) live in rural areas. Regarding education status, 688 (47.3%) were illiterate while 768 (52.7%) were literates. Regarding family history, 606 (41.6%) patients' family had a DM while 850 (58.4) had not. The treatment of patients who used both insulin and OHA were 134 (9.2%), OHA users were 538 (37.0%) and the rest used insulin. There were also 740 (50.8%) patients who did not perform an exercise to control their glucose levels but 716 (49.2%) patients were performed some sort of exercise to control their glucose. Regarding dietary type, among all the participants, 115 (7.9%) were fruit eaters, 85 (5.8 %) were meat eaters, 369 (25.3%) were vegetable eaters and 887 (60.9%) were people of other types of diets.

Similarly, 826 (56.7%) of the DM patients were married, 344 (23.6%) were single and the rest 186 (19.7%) of the patients were divorced and widowed. Regarding the continuous covariates, the mean of the baseline for age and body mass index were 46.7 years of age and 24.0886 kg/m^2 with the standard deviation of 14.737 years and 4.33084 kg/m^2 respectively. These variables were standardized to have a mean of 0 and variance of 1 so that their coefficients in the regression model represent the effect per a unit standard deviation change.

Data analysis

Data exploration

As a first step of the analysis, the data was explored in different ways in order to get details that may help to make decisions in the subsequent steps of the analysis. To determine the evolution and balances of the data, the individual and mean profiles with respect to time were plotted. The mean, the variance and the correlation structures were also explored through graphical techniques. In parallel to defining the fixed effects model, a random effects model was chosen to define a covariance model. After deciding the fixed effects, the study selected a set of random effects to be included in the model.

Longitudinal data modeling

The linear mixed effect model is the most widely used method for analyzing longitudinal data which could handle the complications of incomplete measurements in a very natural way. In this study, a linear mixed model was used with the assumption that the vector of repeated measurements in the original scale on each patient follows a linear regression model where some of the regression parameters are the same for all patients (that is population-specific), while others are different across patients (that is patient-specific). Thus, patient-specific parameters represent patients' variability which is random effects. The idea of randomly varying regression coefficients was also a common thread in the so-called two-stage approach to analyzing longitudinal data. In the two-stage formulation, the repeated measurements on each individual were assumed to follow a regression model with distinct regression parameters. The distribution of these individual-specific regression parameters, or random effects, is modeled in the second stage (Verbeke, 2000). Hence, simple explanatory tools using the two-stage approach were first employed in order to approximate each observed longitudinal profile (that is individual profiles based on the data) by an appropriate linear regression function. Other models are also fitted and compared via information criterion such as BIC and AIC to cross-check the model suggested by the two-stage approach. Often, subject-specific longitudinal profiles can be well approximated by linear regression functions and this leads to a 2-stage model formulation as shown below:

Stage 1: Linear regression model for each subject separately

Response Y_{ij} for i^{th} subject measured at time t_{ij}, $i=1,2,\ldots,n$, and $j=1,2,\ldots,n_i$.

Response vector Y_i for i[th] subject: $Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ . \\ Y_{ini} \end{bmatrix}$

The model of this stage was given as:

$$Y_i = z_i\beta_i + \varepsilon_i \qquad (1)$$

Where, Z_i is a (ni × q) matrix of known covariates; β_i is a q dimensional vector of subject-specific regression coefficients; $\epsilon_i \sim N(0, R)$, where R is the variance-covariance matrix of the error term.

Stage 2: In this stage, the interest is to study the between-subject variability and it can now be studied from relating the β_i to known covariates. Stage 2 model is also given as follows:

$$\beta_i = k_i\beta + b_i \qquad (2)$$

Where, K_i is a (q × p) matrix of known covariates; β is a p-dimensional vector of unknown regression parameters, $b_i \sim N(0,G)$.

A 2-stage approach can be performed explicitly in the analysis for which Yi is summarized by the estimated value of β and the summary statistic of β is analyzed in the second stage. Therefore, the associated drawbacks can be avoided by combining the two stages into one model.

$$\begin{cases} Yi = Zi\beta i + \varepsilon i \\ \beta i = Ki\beta + bi \end{cases} \Rightarrow Yi = ZiKi\beta + Zibi + \varepsilon i \qquad (3)$$

Say, $ZiKi = Xi$, from (3), then the linear mixed effect model is given by:

$$Y_i = X_i\beta + Z_ib_i + \varepsilon_i \qquad (4)$$

Where, Y_i is the (n_i x 1)-dimensional vector of random blood glucoses of the patient i (i = 1,2,…N), N is the number of subjects, X_i and Z_i are (n_i x p) and (n_i x q) dimensional matrices of known covariates, β is a p-dimensional vector containing the fixed effects, $b_i \sim N(0,G)$ is a q-dimensional vector containing the random effects, and $\varepsilon_i \sim N(0,R)$ is an (n_i x 1) dimensional vector of residual components. A key assumption for this model in the foregoing analysis is that ε and *b* are normally distributed with:

$$E\begin{bmatrix} b \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathrm{var}\begin{bmatrix} b \\ E \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

This indicates the random effect and the residual component are independent. The mean and the variance of Y is $X\beta$ and $ZGZ'+R$, respectively while the conditional mean and variance of Y/b is also $X\beta + Zb$, ZGZ' +R respectively. According to the Henderson (1984), the standard mixed model equations is obtained from the value $\hat{G}$, $\hat{R}$ and the known covariates X and Z. The estimate of the covariance matrices are determined based on the maximum likelihood and restricted maximum likelihood.

The maximum likelihood estimation includes both regression coefficients and the variance components, that is, both fixed-effects and random-effects terms in the likelihood function. It treats β as fixed but unknown quantities when the variance components are estimated, but does not take into account the degrees of freedom lost by estimating the fixed effects. This causes ML estimates to be biased with smaller variances. On the other hand, the restricted maximum likelihood estimation includes only the variance components, that is, the parameters that parameterize the random-effects terms in the linear mixed-effects model that accounts for the degrees of freedom lost by estimating the fixed effects. This made a less biased estimation of random effects variances. The estimates of *R* and G are invariant to the value of *β* and less sensitive to outliers in the data compared to ML estimates. However, if REML was used to estimate the parameters, only two models were compared that have the identical fixed-effects design matrices and are nested in their random-effects terms.

The log likelihood function of the covariance matrix of R and G is in the case of ML, and REML were given as follows:

$$\begin{cases} ML: l(G,R) = \frac{-1}{2}\log|V| - \frac{1}{2}rV^{-1} - \frac{n}{2}\log(2\pi) \\ REML: l(G,R) = \frac{-1}{2}\log V - \frac{1}{2}\log\left|X'V^{-1}X\right| - \frac{1}{2}r'V^{-1}r \end{cases} \qquad (5)$$

Where, $r = y - x(x'v^{-1}x)^{-1}x'v^{-1}y$

When minimize equation (5) two times by using a ridge-stabilized Newton-Raphson algorithm, one can estimate the value of *V*.

After estimating V, the parameters of the model can be obtained from the normal equation given below:

$$\begin{bmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X'\hat{R}^{-1}Y \\ Z'\hat{R}^{-1}Y \end{bmatrix} \qquad (6)$$

By simplifying Equation 6, one can get the estimate of,

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y \quad \text{and} \quad \hat{b} = \hat{G}Z'\hat{V}^{-1}(Y - X\hat{\beta})$$

Where, V=Var(Y).

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used for selecting the better model. The criteria select the model that minimizes:

AIC = -2(maximized log likelihood - number of parameters in the model)

Thus, a model with the smallest AIC value would be taken as a best among the candidate models. AIC penalizes a model for having many parameters. The Likelihood-Ratio test is employed to assess the appropriateness, adequacy and usefulness of the model. Moreover, the Wald test were also used to test whether the

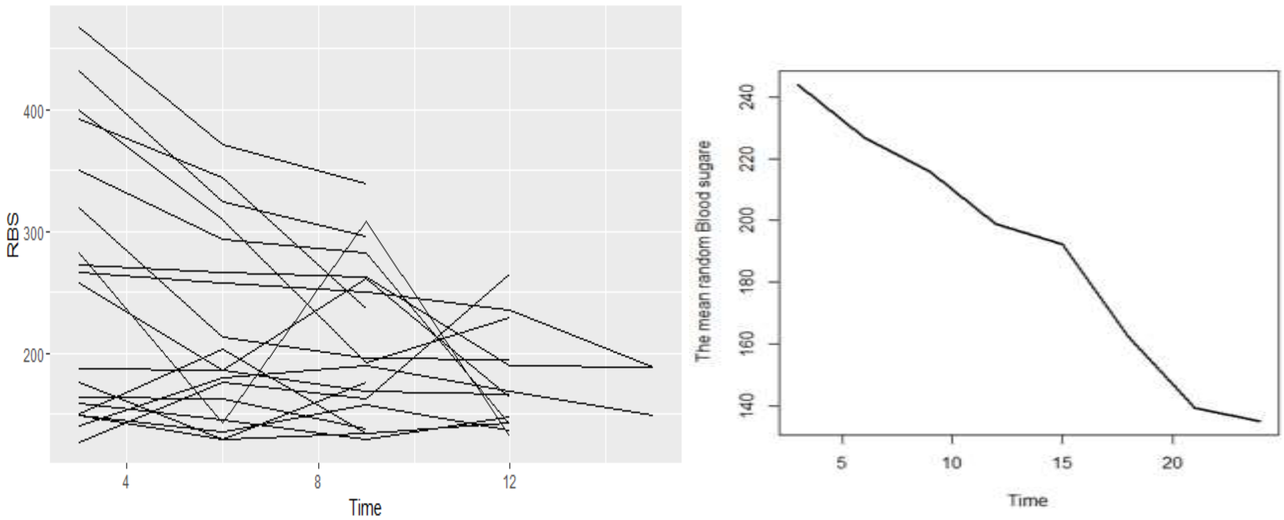

Figure 1a. Over all individual profile of RBS; **b** Over all mean profiles of DM patients.

parameter associated with the fixed effect explanatory variable is zero or not and would be assessed by carrying out statistical tests of the significance of the coefficients (Agresti, 1996, 2008). Some graphical techniques were used to assess peculiarities or the distinctive feature of the model with regard to the data. For detecting outliers, histograms and scatter plot of Empirical Bayes (EB) estimates were used. The EB residuals are defined as the conditional mean of the vector of random effects given the data and the estimated parameter values. A procedure to obtain the empirical Bayes estimates is presented in (Verbeke, 2000). Deletion diagnostics are statistics is used to measure the change in a parameter estimate when some subsets of the data are deleted. Cook's distance is the commonly used statistic for measuring changes in fixed effects. It measures the distance between the fixed effects estimates obtained from the full data and those obtained from the reduced data. To evaluate the effect of measurements of RBS on the variance components, the relative variance change (RVC), which measures the change in variance components with and without deleting a RBS, was employed.

RESULTS AND DISCUSSION

In this study, 1456 observations were considered to collect the random blood sugar levels which was evaluated at fixed time points and measurements. All the patients were taken at 3, 6, 9, 12, 15, 18, 21, and 24 months, which had equal time intervals of 3 months between all measurements. The individual and the mean profile of the patients for their sugar levels starting their follow up and treatment are presented in Figure 1.

The subject profile plots of Figure 1a, was obtained from a randomly selected 20 DM patients. Figure 1a shows that there was a decrease in RBS over time. It is also observed from this plot that there was much variability between patients but less variability within patients over time. From the mean profiles analysis in Figure 1b, the variability of the DM patients was observedand the mean RBS of the individual patients at its initial time was high. After they start the treatment, the RBS level decreased over the time. This is due, to in many cases, the correlation between two repeated measurements decreasing as the time span between those measurements increases. This was very important to determine the type of progression rate for DMT2 disease in terms of RBS over the linear effect of time. In assessing the adequacy of the first-stage linear model to the observed longitudinal profiles, subject-specific coefficients of multiple determinations R^2i (i = 1, 2… N) was used, where, a scatter plot of the subject-specific coefficients of multiple determinations versus the numbers n_i of repeated measurements was conducted. Two such plots based on the first-stage models for linear and quadratic were presented in Figure 2.

Figure 2a and b showed that the values for both linear and quadratic first-stage models were 0.79 and 0.61 respectively. These results revealed that 79 and 61% of the variability in RBS value in the DM patients was explained by the variables included in the model respectively while the remaining variability was described by other factors which are not included in the model. To identify whether the linear or the quadratic time effect model is better, fit the linear mixed model with the same covariance structure but additional quadratic time effect and its interaction with factors in the fixed and the random part was fitted with the value of (AIC, BIC) = (1631.6, 1644) for the linear time effect and (AIC, BIC) = (1836.5, 1761.3) for the quadratic time effect.

A plot of the OLS residual profiles over time, based on the mean structure suggested by the two-stage approach, was employed to check the adequacy of the model.

The plot OLS residual versus time in Figure 3, gave

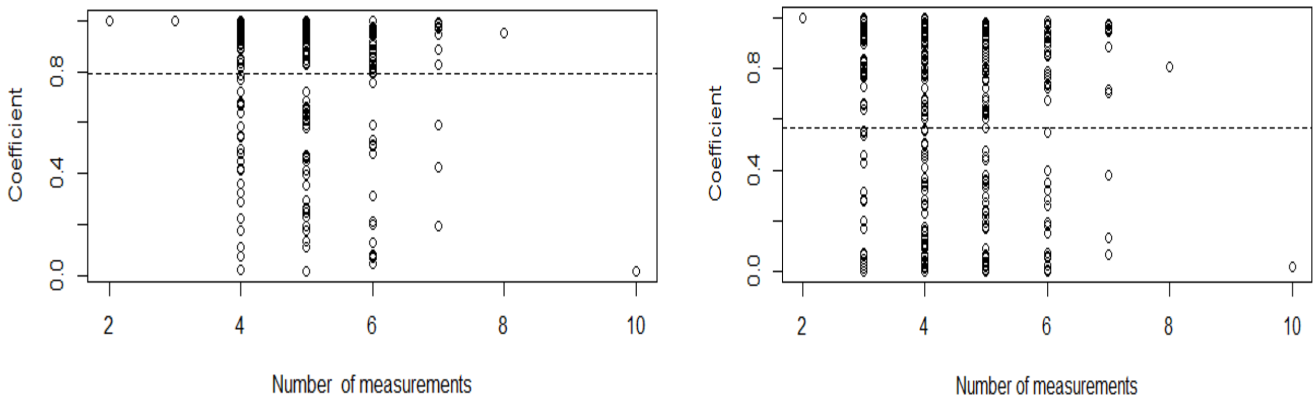

Figure 2. Subject-specific coefficients R^2_i of multiple determinations and the overall coefficient R^2_{meta} of multiple determinations which are shown by dashed line. 2a Linear subject specific profiles; b. Quadratic subject specific profiles.

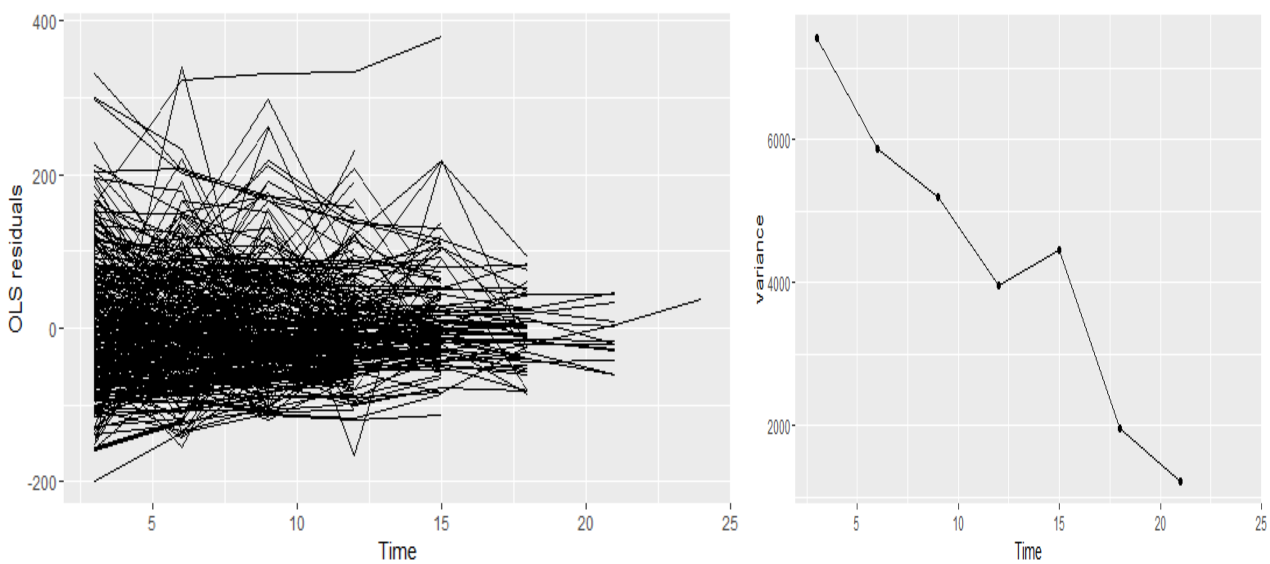

Figure 3. OLS residual profile (a) and variance function of OLS residuals (b).

visual proof that the linear regression trend model provides a good fit to the data. The variance function is shown in Figure 4 which was clearly suggested as non-stationary since the variability varies over time. This also implies the existence of some remaining systematic structure in the residual profiles. By considering Figure 3, it was assumed that the remaining structure in the OLS residual profile might be described by a higher order function over time. Depending on the variance function graph both random intercept and slope were included in the model as a preliminary random effect structure.

By combining all the above explorations, a mixed effect random intercept and random slope with linear time effect was considered for this study.

The significant variables in this study were selected by using backward elimination techniques. The marital status and sex with p value of 0.3315 and 0.4422 respectively were removed from the full model. Also, interactions of main effects over time were insignificant except for the exercise, alcohol, dietary type and BMI. Since these covariates have no other importance in the model improvement, it has to be excluded from the final reduced linear mixed model. The p values above were taken from the model compassion but not from the estimate of the full model. Hence, the final was reduced to a more parsimonious model and was fitted on the data.

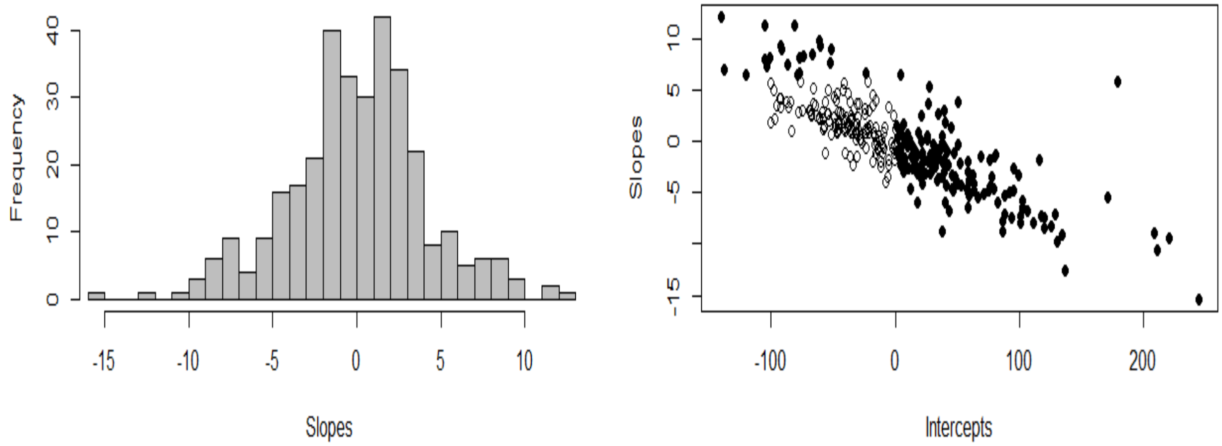

Figure 4. Histograms (a) and scatter plot (b) of empirical Bayes estimates.

Figure 5. Dot plots of Cook's distance.

The graph of histograms and scatter plots of the Empirical Bayes estimates of the random effects were used to detect model deviations or subject's evolutions over the study time (Verbeke, 2000). Therefore, histograms and scatter plots of the Empirical Bayes estimates for the random effects were employed (Figure 4).

From the plot in Figure 4, both scatter plots and the histogram suggested the presence of some outliers. But still the histogram and the scatter plot did not help to determine which observations are outliers. Cook's distance was used to identify outlier observations with any value which was greater than the cutoff value. This is presented in Figure 5.

The plot in Figure 5 shows that the patient with ID number indicated on the graph had Cook's distance value greater than unit, since the cutoff value is one. Thus, the final linear mixed effect model was fitted by excluding these individual outlier observations.

After we chose the appropriate model, the linear mixed model of the data with estimated value of significant covariates was fitted. The restricted maximum likelihood estimates of covariates and the standard error with its corresponding significance value (p value) is found in the Table 2. Based on the estimated values of the parameters (Table 2) and the corresponding significance values, the following linear mixed effect model was modeled as:

$$Y = 373.29 - 12.43 * Time - 37.3x_1 - 26.55x_2 - 19.01x_3 + 41.71x_4 + 84.45x_5 - 17.97x_6 - 27.27x_7 - 0.44x_8 + 2.59x_9 + 3.44x_{10} - 2.39x_{11} - 4.35x_{12} + 0.24x_{13}$$

Table 2. Restricted maximum likelihood parameters estimates of the linear mixed effect model for reduced model.

Effect	Estimate	Std. error	t-value	p-value
Intercept	373.2917	40.83663	9.141098	0.0000
Time	-12.4279	3.81744	-3.255568	0.0012
Non Alcoholic	-37.2972	11.89526	-3.135468	0.0018
No Family History	-26.5524	10.79766	-2.459088	0.0141
Urban	-19.0089	11.58260	-1.641161	0.0101
No Exercise	41.7051	11.46286	3.638277	0.0003
Meat	84.4490	29.60675	2.852358	0.0046
Insulin + OHA	-17.9721	6.69143	-2.685836	0.0073
Literate	-27.2744	11.31629	-2.410190	0.0161
Age	-0.4438	0.21408	-2.073115	0.0384
BMI	2.5992	1.26768	2.050347	0.0406
Time*No Exercise	-2.3923	1.10298	-2.168983	0.0303
Time*Non-Alcoholic	3.4371	1.15745	2.969550	0.0030
Time*Meat	-4.3528	2.83009	-1.538028	0.01243
Time*BMI	0.2361	0.12027	1.963035	0.0499
Likelihood ratio test	**43.93224**			**< 0.0001**

Where, Y is the Random Blood Sugar (RBS) of DM patients and X_1=Non- alcoholic patient; X_2=No-family history patent; X_3= Urban residence; X_4= patients do not perform exercise; X_5= Mostly Meat dietary eater patients; X_6=Treatment of both insulin and OHA; X_7 = Literate patients; X_8 = Age of patients at diagnosis; X_9 = Body mass index (BMI); X_{10}= Interaction of alcohol with time; X_{11}= Interaction of exercise with time; X_{12} = Interaction of meat with time and X_{13}=Interaction of BMI with time.

The statistical analysis results of the linear mixed model were discussed depending on the model fitted above. The average mean value random sugar level of patients was 373.29 keeping the effect of other factors at zero. As one unit increased, the average rate of change in RBS levels was 12.43 mg/dl per unit increased over time. This implied that the rate of change in the mean RBS level decreased by 12.43 mg/dl, keeping constant the other explanatory variables. There was a significant interaction between DM status and time (p=0.0012) such that RBS levels of the patients decreased over time. The other estimated value of covariates were also interpreted and discussed by keeping constant the effect of the remaining factors or taking the constant over time. For a one year increase of age of a patient, the expected value of RBS level reduced by 0.44 mg/dl when the effect of the other factors were kept constant. This supported the findings of (Taylor and Lobel, 1989) that the RBS levels tended to correlate negatively with age. In this study most of the samples had an age range greater than 40. Hence, this result was in line with those of Taylor and Lobel, (1989). In addition, a unit increase in BMI of a patient increases the RBS level by 2.59 mg/dl; this also supported the finding of Tyler and Blader (2001) that a population that had more weight were said to be more obese. This obesity highly correlated with BMI. BMI were calculated based on the value of weight. Since obesity is a main factor for diabetes, and hence the more the obesity of a person is also the greater the BMI. This ultimately shows that with the increase of weight of patients, they are more exposed to greatest risks of having diabetes. The average rate of change of RBS for literate patients decreased by the amount of 27.3 as compared to the illiterate patients keeping the effect of other variables constant. This result clearly showed that literate patients managed the condition of diabetes better, because they were able to understand the basic disease management and treatment plans. This result is in line with the finding of Miller et al. (2002).

Patients who would predominantly eat meat dietary type have been shown to have a significantly higher increase in the mean change value of RBS levels than patients who eat other types of dietary meals. More simply, the mean rate of change of RBS for meat eating patients increased by the value of 84.45 as compared to that of the fruit eaters. The influence on RBS level when people consume more meat may be due to the lack of mineral contents for disease protection. Hence, naturally the RBS level increased by unexpected means. Therefore, people who mostly eat meat were exposed to DM disease and could not with ease control the RBS to attain a normal condition. To reduce the blood sugar, patients would have to abide by a strict diet. Foods associated with cereals, vegetables, and fruits are advised while other dietary methods associated with

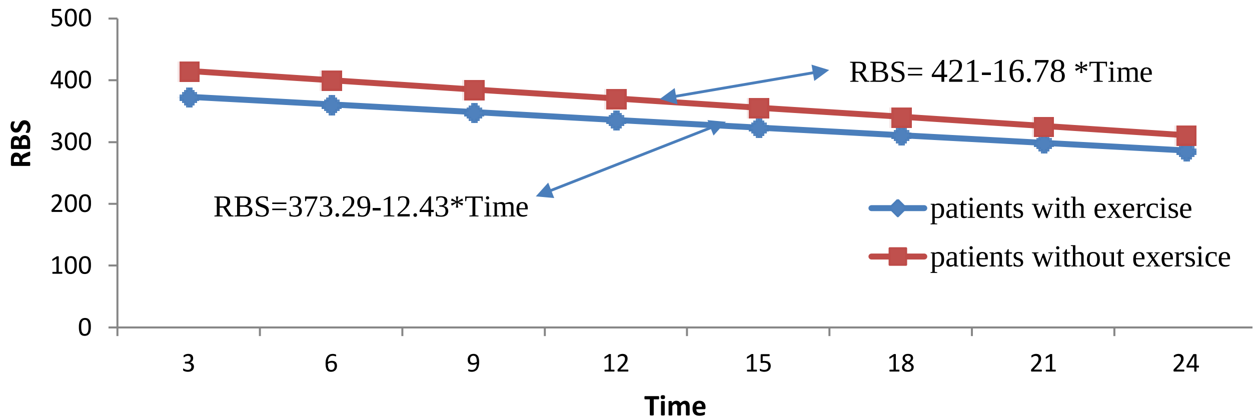

Figure 6. Interaction of activity with time.

eggs, meat, proteins, and starch need to be lessened or even avoided.

Patients use the treatment of insulin, OHA or the combination of them. Among these three treatments, the estimated value for combined treatment was 17.97 which showed that the mean change of RBS level of the patient who uses the combination of the treatment insulin and OHA decreased by 17.97 as compared to the patient who used the insulin treatment. That means that the combination of the two treatments has a greater contribution to the reduction of the random blood sugar of the diabetes mellitus patient opposed to the patient who uses the treatment of insulin. Patients who perform an exercise frequently, no-alcoholic and had no family aggregation (family history) can monitor their RBS well. This indicated that the rate of change in the random blood sugar in their body was higher than that of the alcoholic, had family history and who did not perform regular exercise. An individual with inherited susceptibility to the disease and who did not exercise was negatively correlated with the change in the RBS. These results were also consistent with finding of the study on Diabetes mellitus in the literature of this paper.

The interaction effect of time with exercise, alcohol use, meat as a dietary type and with BMI shows the rate of change in the random blood sugar of the DM patients over time.

The line equations of exercise activity with time given was as follows:

$$Y = \begin{cases} 373.29 - 12.43 * Time, & patient \text{ who perform exercise} \\ 415\text{-}14.82\text{*Time}, & \text{patient who not perform exercise} \end{cases}$$

From the mentioned two line equations, the rate of change can be determined based on the difference of the slopes. This showed that as the visiting time increased by one unit, the average rate of RBS for patients who perform exercise was decreased by the amount of 2.4 as compared to that of patients who did not perform exercise. And hence performing exercise was important for controlling the random blood glucose level in the body. The line graph was presented in Figure 6.

A line equation of alcohol user having interaction over time was given as follows:

$$Y = \begin{cases} 373.29\text{-}12.43\text{*Time}, & \text{for alcolic patients} \\ 335.99\text{-}8.99\text{*Time}, & \text{for non-alcolic patients} \end{cases}$$

According to Figure 6, for a one unit increment of time, the average rate of change in the RBS level for DM patients who did not use alcohol, decreased by the rate of 3.44 as compared to patients who use alcohol. Therefore, this revealed that using alcohol increased the patients RBS level in higher rates compared to the non- alcoholic patients.

The line graph of the interaction effect of time with the factor of alcoholic patients was also given in the Figure 7.

A line graph equation of the interaction of dietary type with time is given below:

$$Y = \begin{cases} 421 - 16.78 * Time, & \text{patients who abid by a meat diet} \\ 373.29 - 12.43 * Time, & \text{patients who abid by a fruit diet} \end{cases}$$

Similarly, as can be seen from Figure 7, for a one unit increment of the visiting time of the patient, the expected or average rate of RBS level of patients who mostly eat meat increased by 4.35 as compared to the patient who abides by a fruit diet.

Therefore, as can be seen from Figure 8, the patient must restrict the types and amounts of food they consume. They might also have to monitor their blood glucose levels at specific times during the day, and medication might be necessary at times when the individual is engaged in

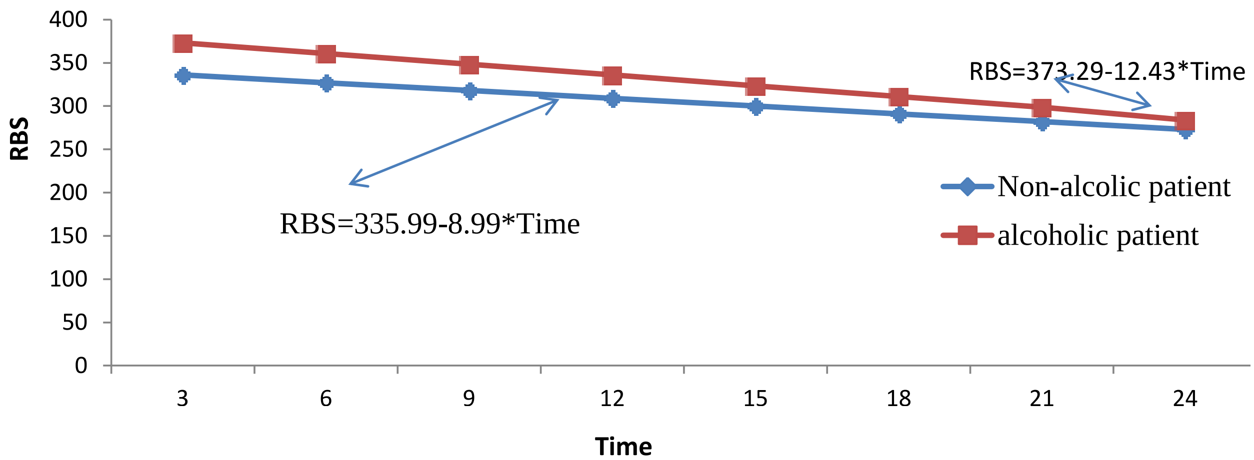

Figure 7. Interaction of alcohol use with time.

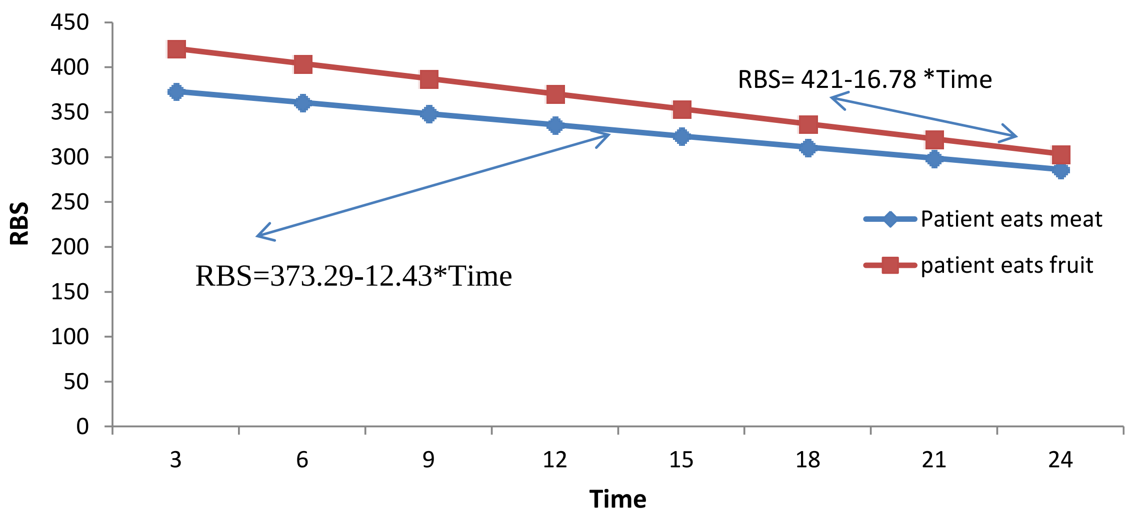

Figure 8. Interaction of dietary type with time.

social activities.

Conclusion

The mixed effects model developed was confirmed to be adequate for the prediction of RBS levels based on the available variable of health determinants. The pattern of mean change in RBS levels revealed a linear distribution that decreased over time. The coefficient of determination explained about 79% of variability in RBS level accounted by the predictor variables. From the individual profile of RBS, there was high variability between subjects and less variability within subjects. And also from the mean profile, the RBS level of the patients decreased over time which was also confirmed with the model that the estimate of time was negative. Among the indicator factors of RBS level, meat as a dietary type, patients who do not perform exercise and body mass index (BMI) correlated positively to the RBS level while the rest are negatively correlated. Hence, more attention was given to control the RBS level of the patients in relation to these factors. The linear mixed-effects model showed that time (duration of follow up), BMI, alcohol use, diet, exercise, education status, residence, age, family history and treatment type have significant influence on the RBS level ($p<0.05$). From the result of the study, patients who live in urban areas, has lower BMI levels, who had no family history, who were educated, did not drink alcohol and who used a combination of insulin and OHA treatment were better suited to control and reduce their RBS in their body over time. But the determinant factors of marital status and sex were an insignificant variable which showed that there was no significance reduction between males and females over time but it does not mean that RBS was not decreased over time by sex. Similarly, the mean RBS reduction over time was not different among the category of marital status.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGMENTS

The authors are grateful to the Editor of JPHE, and anonymous referees for their helpful comments and suggestions on the earlier version of this article.

REFERENCES

Abera E (2000). Patterns of chronic complications of diabetic patients in Menelik II Hospital, Ethiopia. Ethiop. J. Health Dev. 14(1):113-116.

Agresti A (1996, 2008). An Introduction to Categorical Data Analysis. John Wiley and Sons, New York.

Baltazar JC, Fernando RE, Baquilod MM (2004). Prevalence and correlates of diabetes mellitus and impaired glucose tolerance among adults. Diabetes Res. Clin. Pract. 64(2):107-115.

Center for Disease Control (CDC) (2007b). Take Charge of Your Diabetes. 4th edition. Centers for Disease Control and Prevention, Alanta, GA. US.

Feleke Y, Enquselassie F (2007). Cost of Hospitalization of Diabetic patients admitted at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia. Ethiop. Med. J. 45(3):275-282.

Graham CN, Herbut D, Quandt SA (1999). Quality of care of patients with diabetes mellitus living in a rural area of Germany. Med. Klin. 94(3):127-132.

Helgeson VS, Gottlieb BH (2000). Social support measurement and intervention: A guide for health and social scientists. New York: Oxford University Press. pp. 221-245.

Hennekens C (1998). Increasing burden of cardiovascular disease. Biomed. Center. 97:1095-1102.

International Diabetes Federation (IDF) (2013). Diabetes atlas. 6th ed. Belgium.

Kahn S, Hull R, Utzschneider K (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature Int. Weekly J. Sci. 444:840-846.

Klinke DJ (2008). Extent of Beta Cell Destruction Is Important but Insufficient to Predict the Onset of Type 1 Diabetes Mellitus. PLoS ONE 3(1):e1374.

Laing S, Swerdlow A, Slater S, Botha J, Burden A, Waugh N (1999). The British diabetic association cohort study ii: cause-specific mortality in patients with insulin treated diabetes mellitus. Diabetic Med. 16:466-471.

MacPherson SL, Joseph D, Sullivan E (2004). The benefits of peer support with diabetes. Nurs. Forum. 39(4):5-12.

Miller CK, Edwards L, Kissling G, Sanille L (2002). Nutrition Education Improves Metabolic Outcomes among Older Adults with Diabetes Mellitus: results from a Randomized Controlled Trial. Preventive Med. 34(2):252-259.

Philis-Tsimikas A, Walker C (2001). Improved care for diabetes in underserved populations. J. Ambulatory Care Manage. 24:39-43.

Shaw JE, Sicree RA, Zimmet PZ (2010). Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res. Clin. Pract. 87(1):4-14.

Taylor S, Lobel M (1989). Social comparison activity under threat: downward evaluation and upward contacts. Psych Rev. 96(4):569-575.

Thorne S, Paterson B, Russell, C. (2003). The structure of every day self-care decision making in chronic illness. Qual. Health Res. 13(10):1337-1352.

Timothy A, Derek N, Nawumbeni S, Dodzi N, Ruth P (2015). Mixed-Effects Model for Longitudinal Study of Type-2-Diabetes, Gahana. Int. J. Stat. Appl. 5(2):47-55.

Tyler TR, Blader SL (2001). Identity and cooperative behavior in groups. Group processes and intergroup relations. 4(3):207-226.

Verbeke G, Molenberghs G (2000). Linear Mixed Models for Longitudinal Data. Springer Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA.

World Health Organization (WHO) (2008). The global burden of disease: 2004 Update. WHO Press. pp. 112-117.

World Health Organization (WHO) (2017). Diabetes fact sheet. Available from: http://www.who.int/mediacentre/factsheets/fs312/en/

Yki-Jarvinen H (2001). Combination therapy with insulin in type 2 diabetes. Diabetes Care 24(4):758-767.