

RESEARCH ARTICLE

Open Access



# A semi-parametric mixed models for longitudinally measured fasting blood sugar level of adult diabetic patients

Tafere Tilahun Aniley<sup>1\*†</sup> , Legesse Kassa Debusho<sup>2†</sup>, Zelalem Mehari Nigusie<sup>3</sup>, Wondwosen Kassahun Yimer<sup>4</sup> and Belay Birlie Yimer<sup>5</sup>

## Abstract

**Background:** At the diabetic clinic of Jimma University Specialized Hospital, health professionals provide regular follow-up to help people with diabetes live long and relatively healthy lives. Based on patient condition, they also provide interventions in the form of counselling to promote a healthy diet and physical activity and prescribing medicines. The main purpose of this study is to estimate the rate of change of fasting blood sugar (FBS) profile experienced by patients over time. The change may help to assess the effectiveness of interventions taken by the clinic to regulate FBS level, where rates of change close to zero over time may indicate the interventions are good regulating the level.

**Methods:** In the analysis of longitudinal data, the mean profile is often estimated by parametric linear mixed effects model. However, the individual and mean profile plots of FBS level for diabetic patients are nonlinear and imposing parametric models may be too restrictive and yield unsatisfactory results. We propose a semi-parametric mixed model, in particular using spline smoothing to efficiently analyze a longitudinal measured fasting blood sugar level of adult diabetic patients accounting for correlation between observations through random effects.

**Results:** The semi-parametric mixed models had better fit than the linear mixed models for various variance structures of subject-specific random effects. The study revealed that the rate of change in FBS level in diabetic patients, due to the clinic interventions, does not continue as a steady pace but changes with time and weight of patients.

**Conclusions:** The proposed method can help a physician in clinical monitoring of diabetic patients and to assess the effect of intervention packages, such as healthy diet, physical activity and prescribed medicines, because individualized curve may be obtained to follow patient-specific FBS level trends.

**Keywords:** Diabetes mellitus, Fasting blood sugar, Linear mixed model, Semi-parametric mixed model

## Background

Diabetes mellitus is a metabolic disorder of multiple etiology characterized by chronic hyperglycaemia with disturbances of carbohydrate, fat and protein metabolism resulting from defects in insulin secretion, insulin resistance, or both [1]. The long-term effects of untreated diabetes mellitus might results in health complications, such as visual disability and nerve disease [2–5], among

others. A person is considered to be diabetic if he or she has fasting blood sugar (FBS) level value of greater than or equal to 7.0 mmol/L (126 mg/dL) or 2-h blood sugar level of greater than or equal to 11.1 mmol/L (200 mg/dL) or glycated hemoglobin (HbA<sub>1c</sub>) level of 6.5% or higher [6].

There are three main types of diabetes, namely type 1 diabetes, type 2 diabetes and gestational diabetes. The type 1 diabetes is caused by an auto-immune reaction, in which the patient body defense system attacks the insulin producing beta cells in the pancreas and hence the body can no longer produce the insulin it needs. Whereas in type 2 diabetes, the body is able to produce insulin, however it becomes resistant so that the insulin

\*Correspondence: tafere27@yahoo.com

†Tafere Tilahun Aniley and Legesse Kassa Debusho contributed equally to this work.

<sup>1</sup>Department of Statistics, Jimma University, Jimma, Ethiopia  
Full list of author information is available at the end of the article



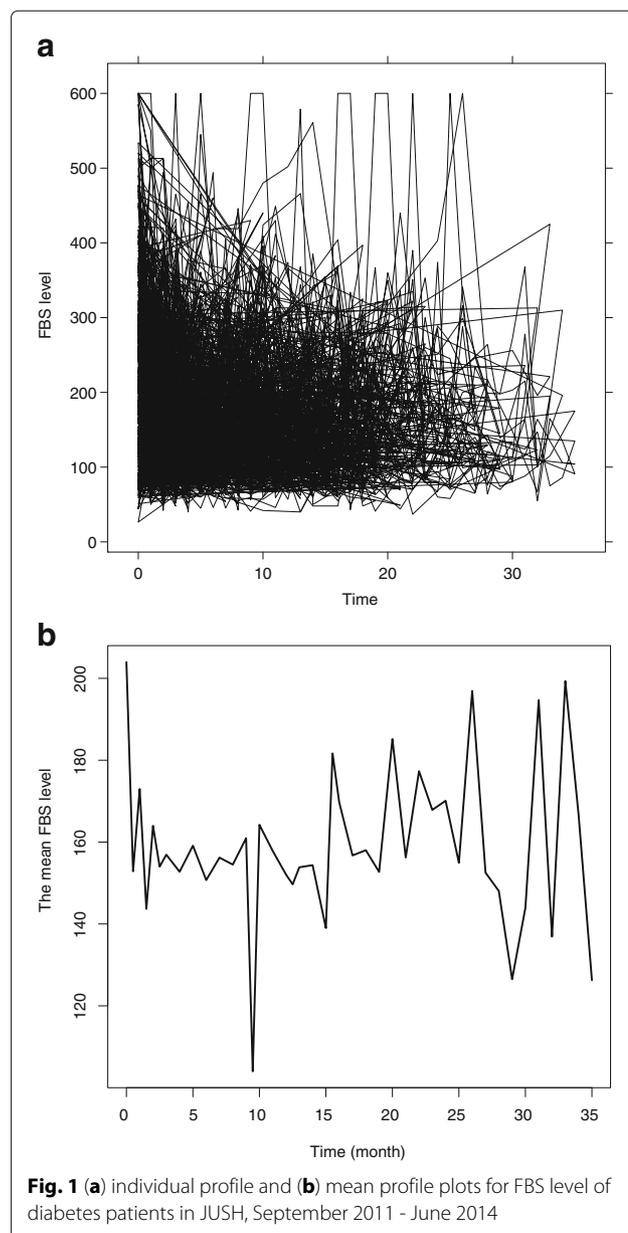
is ineffective. The type 2 diabetes is characterized by high levels of blood sugar or glucose resulting from defects in insulin production, insulin action, or both. The gestational diabetes is a form of diabetes that appears during pregnancy. It can lead to serious health risks for both the mother and child [7]. The risk factors that are associated with type 1 diabetes include family history of diabetes (diabetes history in one parent or both), infections and other environmental influences such as exposure to a viral illness, the presence of damaging immune system cells, i.e. autoantibodies and dietary factors low vitamin D consumption [8]. Whereas, for type 2 diabetes the risk factors are excess body weight, physical inactivity, poor nutrition, family history of diabetes, past history of gestational diabetes and older age [9]. The risk factors for increase or decrease in fasting blood sugar level of a patient include overweight, family history of diabetes, age, type of diabetes, blood pressure and gender [7]. The focus of this study however is on type 1 and type 2 diabetes.

In year 2015, there were an estimated 415 million adults aged 20–79 years living with diabetes worldwide [10], including 193 million who are undiagnosed. There were approximately 5 million people estimated to have died from diabetes worldwide in the same year, and a majority of these were the result of cardiovascular complications. In Africa Region, the number of adults living with diabetes estimated at 14.2 million whereas in Ethiopia the number is estimated 1 to 10 million in year 2015. The Region has the highest proportion of undiagnosed diabetes, 9.5 million (about 66.7%) of people with diabetes are unaware they have the disease and in Ethiopia there are 500 thousand to 5 million such cases [11, 12].

At the diabetic clinic of Jimma University Specialized Hospital (JUSH), health professionals provide regular follow-up to help people with diabetes live long and relatively healthy lives. Depending on patients conditions, e.g. FBS level, they also provide interventions in the form of counselling to promote a healthy diet and physical activity and prescribing medicines.

The main objective of the current study is to assess the factors that affect the FBS level of adult diabetic patients. In addition to assessing the factors that affect the FBS level over time, we are also interested to estimate the rate of change of FBS profile experienced by patients over time. The change may help to assess the effectiveness of interventions taken by the clinic to regulate FBS level, where rates of change close to zero over time may indicate the interventions are good regulating the level. These changes are determined using first derivatives of penalized regression splines [13, 14].

The FBS level data of diabetic patients in this study are collected repeatedly over time hence the data have longitudinal time series profiles and the data also have continuous nature. For statistical inferences, therefore, it is necessary to capture properly the form of the evolution of profiles over time. In the analysis of longitudinal data, the mean profile is often estimated by parametric linear mixed effects model, for instance recently Mehari [15] analyzed the FBS level profiles of diabetic patients using parametric linear mixed effects model. However, the individual and mean profile plots of FBS level for diabetic patients (see Fig. 1) are nonlinear and imposing parametric models may be too restrictive and yield unsatisfactory results.



In the present paper, we propose a semi-parametric mixed model in particular using spline smoothing [16, 17] to efficiently analyze a longitudinal measured fasting blood sugar level of adult diabetic patients accounting for correlation between observations through random effects. The model assumes that the mean of FBS level is an arbitrary smooth function of time and parametric functions of other covariates. The link between mixed model and smoothing provides a flexible framework for estimating the patient profiles in a data driven way [13].

The rest of the paper is organized as follows. The data, some basic review of variance-covariance structure of the parametric linear mixed model, semi-parametric mixed models and inferences related them are introduced in “Methodology” section. The results from applying these methods on the the study data are discussed in “Results” section. Finally discussion, and conclusions and pointers for future study are given in “Discussion” and “Conclusion” sections respectively.

## Methodology

### Study data

The fasting blood sugar (FBS) level data used in this paper arises from a retrospective study conducted in Jimma University Specialized Hospital (JUSH) diabetic clinic. The hospital is located in Jimma town 352 km to the Southwest of Addis Ababa, the capital of Ethiopia. It is a teaching hospital and gives service to the southwestern part of Oromia region, some part of southern nations and nationalities and Gamella regions of Ethiopia. All diabetic patients aged 18 years or older, who were coming to JUSH diabetic clinic for their regular follow up during periods September 2011 and June 2014 were eligible for this study. During their follow up, patients FBS level along with other characteristics such as weight are measured and recorded in the individual follow up chart. The data in the chart include time (measured in months, where baseline or initial date was given a value 0), patient gender, age, type of diabetes ( Type 1 diabetes or Type 2 diabetes) and family diabetes history. The duration between initial and the last recorded visits ranged from one to 36 months. Patients with at least two observations were included in the analyses leading to a total of 534 patients and 4390 observations. Permission of the study was obtained from Postgraduate research office of Jimma University, College of Natural Sciences and JUSH.

### Variance-covariance structures and inference

#### Variance-covariance structures

The FBS level data of this study fall within the framework of continuous longitudinal data and hence can be modeled by use of a parametric linear mixed model. Let  $Y_{ij}$  denote the FBS level of the  $i$ th patient observed at time

$t_{ij}$ ,  $i = 1, \dots, n$  and  $j = 1, \dots, m_i$ . The parametric linear mixed model can be expressed as

$$Y_{ij} = \sum_{k=0}^p \beta_k t_{ij}^k + \sum_{l=1}^L \theta_l x_{ijl} + \sum_{u=0}^q b_{u_i} t_{ij}^u + \varepsilon_{ij}. \quad (1)$$

That is, the population level mean response is modeled as a polynomial function of time,  $t_{ij}$ , a linear function of covariates  $x_{ijl}$ ,  $l = 1, \dots, p$  where some of them may be time-varying covariates or interaction effects each has corresponding regression parameter coefficient  $\theta_l$ , a function of subject-specific random coefficient terms and measurement error  $\varepsilon_{ij}$ . The coefficients  $\beta_k$ ,  $k = 1, \dots, p$  and  $\theta_l$ ,  $l = 1, \dots, L$  are fixed effect parameters and  $b_{u_i}$ ,  $u = 0, \dots, q$  are subject-specific random coefficients. It is assumed that  $b_{u_i} \sim \mathcal{N}(0, \sigma_{b_u}^2)$ ,  $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ ,  $cov(b_{u_i}, b_{u'_i}) = \sigma_{b_u b_{u'}}$  and  $cov(b_{u_i}, \varepsilon_{ij}) = 0$ . We have examined models for  $p = 2$  which represents quadratic polynomial and  $b_{u_i}$  with  $u = 0, 1, 2$  represent a subject-specific random intercept, slope and quadratic coefficients, respectively for selection of a variance-covariance structure (see Table 1). The variance profile plot of FBS level shows (for the sake of brevity this plot is not reported) the variance changes overtime, therefore to allow for more flexibility to estimate between subject variability we have considered the above three variance-covariance structures.

In Table 1, for instance the subject-specific random intercept  $b_{0_i}$  in the quadratic random effects model (M<sub>3</sub>) is considered to capture correlation of the FBS level measurements over time within the patient and it is assumed that subject-specific random slopes for linear as well as for quadratic time effects to capture different evolution of FBS level over time. Note that these subject-specific random structures are different for each patient.

**Table 1** Linear mixed models for selection of variance-covariance structure for FBS level, JUSH, September 2011 - June 2014

Subject-specific random effect	Linear mixed model
M <sub>1</sub> : Random intercept	$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \sum_{l=1}^L \theta_l x_{ijl} + b_{0_i} + \varepsilon_{ij}$
M <sub>2</sub> : Linear random effects	$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \sum_{l=1}^L \theta_l x_{ijl} + b_{0_i} + b_{1_i} t_{ij} + \varepsilon_{ij}$
M <sub>3</sub> : Quadratic random effects	$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \sum_{l=1}^L \theta_l x_{ijl} + b_{0_i} + b_{1_i} t_{ij} + b_{2_i} t_{ij}^2 + \varepsilon_{ij}$

**Tests for zero variance components**

Adequate variance-covariance structure is essential to obtain valid model based inferences for the fixed effects or for parameters in the mean structure of the model [18]. Over-parametrization of the variance-covariance structure leads to inefficient estimation and potentially poor assessment of standard errors for the estimation of the mean structure, i.e. fixed effects, whereas a too restrictive specification invalidates inferences about the mean response profile when the assumed structure does not hold.

The likelihood ratio test for testing, for example  $H_0 : \sigma_{b_0}^2 = 0$  versus  $H_1 : \sigma_{b_0}^2 > 0$  for model  $M_1$ , has an asymptotic  $0.5 \chi_0^2 + 0.5 \chi_1^2$  mixture distribution under  $H_0$  [19], if the vector of FBS level can be divided into a large number of independent and identically distributed sub-vectors both under  $H_0$  and  $H_1$ . However, this assumption usually does not hold, for example in linear mixed models or for unbalanced data [20–22]. Note that the FBS level data are unbalanced in the sense that all patients do not have equal number of measurements, hence the independent and identically distributed assumption can be violated in the linear mixed models used in this paper. Therefore, we used the exact finite sample null distribution of the restricted likelihood ratio test (RLRT) statistic derived by Crainiceanu and Ruppert [22] to test a zero random effect variance in  $M_1$ . However, since models  $M_2$  and  $M_3$  contain more than one random effect, the tests for a zero random effect variance in these models were done using the exact finite sample null distribution of the RLRT statistic derived by Greven et al. [21].

**Semi-parametric mixed effects model**

Given the mean profile plots over time in Fig. 1b, imposing parametric functions to describe the mean FBS level evolution may not be easy and also too restrictive [17]. As an alternative, we can model the mean profiles over time with a semi-parametric smooth function,  $f(t_{ij})$ . Using the  $p$ th degree truncated power basis,  $f(t_{ij})$  can be written as

$$f(t_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \dots + \beta_p t_{ij}^p + \sum_{l=1}^K b_l (t_{ij} - \kappa_l)_+^p, \tag{2}$$

here  $z_+ = \max\{0, z\}$ . The function  $f(t_{ij})$  is a combination of fixed effects parameters  $\beta_0, \beta_1, \dots, \beta_p$  and  $p$ th degree splines evaluated at time  $t_{ij}$  with knots at distinct locations  $\kappa_1, \kappa_2, \dots, \kappa_K$  in the range of  $t_{ij}$  and corresponding coefficients  $b_1, b_2, \dots, b_K$ . The function  $f(t_{ij})$  can be estimated among others, with penalized splines. The coefficients of spline basis functions  $b_l$  are assumed to follow a Gaussian distribution such that  $b_l \sim \mathcal{N}(0, \sigma_b^2)$ , where  $\sigma_b^2$  is a variance component controlling the smoothness of

$f(t_{ij})$ . Then, incorporating  $f(t_{ij})$  in model (1), the general semi-parametric mixed effects model can be expressed as

$$Y_{ij} = f(t_{ij}) + \sum_{l=1}^L \theta_l x_{ijl} + \sum_{u=0}^q b_{u_i} t_{ij}^u + \varepsilon_{ij}. \tag{3}$$

**Estimation of parameters**

Let  $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{im_i})'$  be the  $m_i \times 1$  vector of responses for the  $i$ th patient,  $i = 1, \dots, n$ . Under the linear mixed model formulation, model (3) with subject-specific quadratic random effects can be expressed succinctly in matrix form as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_{i(f)} \mathbf{v} + \mathbf{Z}_{i(u)} \mathbf{u}_i + \mathbf{e}_i \tag{4}$$

where  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p, \theta_1, \dots, \theta_L)'$  is a  $(p + L + 1) \times 1$  vector of fixed effects which is common to the  $n$  individuals,  $\mathbf{X}_i$  is an  $m_i \times (p + L + 1)$  design matrix associating  $\boldsymbol{\beta}$  to  $\mathbf{y}_i$ ,  $\mathbf{v} = (b_1, b_2, \dots, b_K)$  is a  $K$ -dimensional vector of random coefficients in the summand in Eq. (2),  $\mathbf{Z}_{i(f)}$  is the  $m_i \times K$  matrix for the  $p$ th-degree spline basis functions,  $\mathbf{u}_i = (b_{0_i}, b_{1_i}, b_{2_i})'$  is subject-specific vector of random effects,  $\mathbf{Z}_{i(u)}$  is an  $m_i \times 3$  design matrix which relates  $\mathbf{u}_i$  to the response  $\mathbf{y}_i$  and  $\mathbf{e}_i = (e_{1i}, e_{2i}, \dots, e_{im_i})'$  is an  $m_i$ -dimensional vector of within-individual errors. Furthermore, it is assumed that  $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma_b^2 \mathbf{I}_K)$ ,  $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$ ,  $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$ ,  $\mathbf{v}$ ,  $\mathbf{u}_i$  and  $\mathbf{e}_i$  are assumed to be pairwise independent with and between subjects for  $i = 1, 2, \dots, n$ . Note that  $\mathbf{G}$  and  $\mathbf{R}_i$  are  $3 \times 3$  and  $m_i \times m_i$  variance-covariance matrices, respectively.

The overall model for  $n$  individuals has the form

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{b} + \mathbf{e}$$

where

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix},$$

$$\mathbf{X}_i = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 & \dots & t_{i1}^p & x_{i11} & \dots & x_{i1L} \\ 1 & t_{i2} & t_{i2}^2 & \dots & t_{i2}^p & x_{i21} & \dots & x_{i2L} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{im_i} & t_{im_i}^2 & \dots & t_{im_i}^p & x_{im_i1} & \dots & x_{im_iL} \end{pmatrix},$$

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_{1(f)} & \mathbf{Z}_{1(u)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{Z}_{2(f)} & \mathbf{0} & \mathbf{Z}_{1(u)} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_{n(f)} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_{n(u)} \end{pmatrix},$$

$$\mathbf{Z}_{i(u)} = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 \\ 1 & t_{i2} & t_{i2}^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{im_i} & t_{im_i}^2 \end{pmatrix},$$

$$\mathbf{Z}_{i(f)} = \begin{pmatrix} (t_{i1} - \kappa_1)_+^p & (t_{i1} - \kappa_2)_+^p & \dots & (t_{i1} - \kappa_K)_+^p \\ (t_{i2} - \kappa_1)_+^p & (t_{i2} - \kappa_2)_+^p & \dots & (t_{i2} - \kappa_K)_+^p \\ \vdots & \vdots & \ddots & \vdots \\ (t_{im_i} - \kappa_1)_+^p & (t_{im_i} - \kappa_2)_+^p & \dots & (t_{im_i} - \kappa_K)_+^p \end{pmatrix},$$

$$\mathbf{e} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{pmatrix}$$

and  $\mathbf{b} = (b_1, b_2, \dots, b_k, b_{0_1}, b_{1_1}, b_{2_1}, \dots, b_{0_n}, b_{1_n}, b_{2_n})'$ . Estimation of the coefficients of penalized and unpenalized terms in model (4) was done using a penalized iteratively reweighted least squares (P-IRLS) based on 20 equidistant knots in the range of FBS level and a smoothing parameter selection was done by REML [23].

The correspondence between the penalized spline smoother and the optimal predictor in a mixed model framework enables us to take advantage of the existing methodology for mixed model analysis and the use of mixed model software, such as the function `gamm` in `mgcv` R package, for fitting the penalized spline model and the MIXED and GLIMMIX procedures in SAS [24]. This implementation of penalized smoothing in the linear mixed model framework also provides an automated approach to obtain a smoothing parameter and flexibility to extend the models [17].

In this paper, parameters in the fitted models are estimated by restricted maximum likelihood (REML) method because the statistical hypotheses that were considered have the same mean structures between models under the null and alternative hypotheses. Furthermore, maximum likelihood estimators of variance components are biased downward as they do not take into account the degrees of freedom lost in the estimation of fixed effects (e.g. see Ruppert et al. [16]).

**Model selection and inference**

The model building process of this work includes selection of suitable variance-covariance structure for random effects, testing whether the inclusion of spline effects in the parametric model improves model fit or not and also selection of covariates. The linear mixed model framework provides a unified approach to do all these [25]. In the parametric cases, the best fitting model can be selected by employing a commonly used selection criteria, Akaike’s Information Criterion (AIC) and Bayesian Information Criterion (BIC) or by a likelihood ratio test. However, since the semi-parametric mixed models that we considered here are differ in both the fixed effects and the nonparametric part, model selection is done via adjusted Akaike’s information criterion, abbreviated  $AIC_{adj}$ , using the effective number of parameters in the model [16, 26]. Let  $\mathbf{C} = [\mathbf{X} \ \mathbf{Z}_f]$  be the design matrix

with appropriate fixed effects components and the corresponding smoothing matrix,  $\mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}$  where  $\mathbf{G}$  is the variance-covariance matrix of random effects used in the model and  $\mathbf{R} = \text{diag}\{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n\}$ , i.e.  $\mathbf{R}$  is the block diagonal variance-covariance matrix of error terms with blocks  $\mathbf{R}_i$  on the main diagonal and zeros elsewhere. Then the effective number of parameters and  $AIC_{adj}$  may be computed as

$$E_p = \text{trace} \left\{ (\mathbf{C}' \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{C}' \mathbf{R}^{-1} \mathbf{C} \right\}$$

and  $AIC_{adj} = -2 \log(Lik) + 2 E_p$ , respectively. Unlike the marginal AIC which penalizes only for the number of parameters in fixed effects vector and variance components, the penalty of  $AIC_{adj}$  takes into account the additional parameters introduced into a model via  $f(t_{ij})$  or smoothing by including the design matrix  $\mathbf{Z}_f$  in  $\mathbf{C}$  [17]. Like the marginal AIC, the smaller the  $AIC_{adj}$  value, the better the model.

Testing whether the inclusion of spline effects in the parametric model improves model fit or not is equivalent to testing  $H_0 : \sigma_b^2 = 0$  versus  $H_1 : \sigma_b^2 > 0$ . In this paper, due to the second objective of the study, a quadratic penalized spline was added in Eq. (1), therefore neither of the two methods discussed in “Variance-covariance structures and inference” section can be used to test  $H_0 : \sigma_b^2 = 0$  [27] instead an approximate  $F$ -test of Hastie and Tibshirani [28] was applied. For Hastie and Tibshirani approximate  $F$ -test, residual degrees of freedom for the null and alternative model fits are used in the place of the number of parameters in each model.

**Rate of change over time and simultaneous confidence bands**

The change in smoothing function  $f(t)$  overtime, for selected semi-parametric mixed model, can be estimated by taking the derivative of  $f(t)$  with respect to time  $t$ . For example, let  $f(t)$  be a quadratic penalized spline, that is

$$f(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{l=1}^K b_l (t_{ij} - \kappa_l)_+^2.$$

Taking the first derivative with respect to time  $t$  yields

$$f'(t) = \beta_1 + 2 \beta_2 t + 2 \sum_{l=1}^K b_l (t_{ij} - \kappa_l)_+.$$

An estimate of  $f'(t)$ , denoted  $\hat{f}'(t)$ , is obtained by substituting the quadratic fit parameter estimates  $\hat{\beta}_1, \hat{\beta}_2$ , and  $\hat{b}_1, \hat{b}_2, \dots, \hat{b}_K$ . However, the construction of simultaneous confidence bands requires the variance-covariance matrix for the vector of contrasts between the estimated and true parameters for the fixed and random effects. Let  $\mathbf{C} = [\mathbf{X} \ \mathbf{Z}_f]$  be a design matrix containing

quadratic time effects and a truncated quadratic basis,  $\mathbf{B}$  is a matrix constructed from variance components corresponding to smoothing, i.e.  $Var(\mathbf{v})$  in model (4). Then, a variance-covariance matrix for the vector of contrasts is given by

$$Var\left(\begin{bmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{v}} - \mathbf{v} \end{bmatrix}\right) \simeq (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{B})^{-1}$$

Ruppert et al. [16], where  $\mathbf{R}$  is the block diagonal variance-covariance matrix of error terms defined in “Semi-parametric mixed effects model” section. Let  $\mathbf{g} = (g_1, g_2, \dots, g_T)$  be a grid of equally spaced time points. Define

$$\hat{\mathbf{f}}_g - \mathbf{f}_g = \mathbf{C}_g \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{v}} - \mathbf{v} \end{pmatrix}$$

where  $\mathbf{C}_g$  is  $\mathbf{C}$  with design matrices  $\mathbf{X}$  and  $\mathbf{Z}_f$  are evaluated over  $\mathbf{g}$ . Assuming the vector of contrasts have approximately multivariate distribution with mean vector  $\mathbf{0}$  and variance-covariance matrix  $(\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{B})^{-1}$  [16, 29], i.e.

$$\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{v}} - \mathbf{v} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{B})^{-1}) \tag{5}$$

a  $100(1 - \alpha)\%$  simultaneous confidence bands for  $\mathbf{f}_g$  is given by

$$\hat{\mathbf{f}}_g \pm h_{(1-\alpha)} \mathbf{s}_g \tag{6}$$

where  $\mathbf{s}_g = (\widehat{SD}(\hat{f}_{g_1} - f_{g_1}), \widehat{SD}(\hat{f}_{g_2} - f_{g_2}), \dots, \widehat{SD}(\hat{f}_{g_T} - f_{g_T}))'$  with

$$\widehat{SD}(\hat{f}_{g_m} - f_{g_m}) = \sqrt{\text{the } (m, m)\text{th diagonal element of } Var(\hat{\mathbf{f}}_g - \mathbf{f}_g)}$$

and  $Var(\hat{\mathbf{f}}_g - \mathbf{f}_g) = \mathbf{C}_g (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{B})^{-1} \mathbf{C}'_g$ , and  $h_{(1-\alpha)}$  is the  $(1 - \alpha)$  quantile of

$$\sup \left| \frac{\hat{f}(t) - f(t)}{\widehat{SD}\{\hat{f}(t) - f(t)\}} \right| \approx \max_{1 \leq m \leq T} \left| \frac{\left(\mathbf{C}_g \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{v}} - \mathbf{v} \end{bmatrix}\right)}{\widehat{SD}\{\hat{f}(g_m) - f(g_m)\}} \right| \tag{7}$$

The quantile  $h_{(1-\alpha)}$  can be approximated using simulations. First we simulate from realization of (5) and computation of (7) can be repeated for a large number of times, say  $N$  times, to obtain  $\tilde{h}_{1-\alpha}^1, \tilde{h}_{1-\alpha}^2, \dots, \tilde{h}_{1-\alpha}^N$ . The value with rank  $N \times (1 - \alpha)$  is used as  $h_{1-\alpha}$ .

The proposed semi-parametric mixed models were fitted with the `gamm` function available in R package `mgcv` [29] and the linear mixed models using the `lme` function available in R package `nlme`.

## Results

### Patients baseline characteristics

A total of 534 adult diabetic patients were in the study, of which 342 (64.04%) were male, 399 (74.72%) were Type 2 diabetic patients and 417 (78.09%) didn't have family history of diabetes. The patients mean (SD) age at the first visit (or baseline) was 45.40 (14.62) years and ranges between 18 and 93 years, weight was 62.83 (13.36) kgs and FBS level was 164.72 (86.20) mg/L. There were significant differences of these means between Type 1 and Type 2 diabetic groups (Table 2). The results in Table 2 also show that, at baseline there was a significant association between family history of diabetes and type of diabetes ( $p$ -value  $< 0.0001$ ). However, the association between patient gender and type of diabetes was nonsignificant ( $p$ -value = 0.9935). The median (first quartile - third quartile) time between first and last clinic visits of patients was 15.25 (7.25 - 24.75) months and ranged from as few as 0.5 month between visits to as much as 6 months between visits.

### Parametric mixed models

#### Mean structure

The main interest of this study is to apply semi-parametric mixed models, however for comparison purpose here we start the analysis by fitting parametric mixed models. Scatter plot smoothing was used to examine changes in FBS level over time and also to assess the interactions of each categorical covariate with time [30, 31]. The smoothing plots suggest the changes in FBS can be

**Table 2** Baseline characteristics of adult diabetic patients in JUSH, September 2011 - June 2014

Characteristics		Type of diabetes		$p$ -value	Overall
		Type 1	Type 2		
Gender	Male, N (%)	87 (16.29%)	255 (47.75)	0.9935	342 (64.04%)
	Female, N (%)	48 (8.99%)	144 (26.97%)		
Family history	No, N (%)	37 (6.93%)	380 (71.16%)	$< 0.0001$	417 (78.09%)
	Yes, N (%)	98 (18.35%)	19 (3.56%)		
Age, mean (SD)		34.55 (11.92)	48.63 (13.78)	$< 0.0001$	45.4 (14.62)
Weight, mean (SD)		58.83 (11.10)	64.02 (13.74)	$< 0.0001$	62.83 (13.36)
FBS, mean (SD)		171.38 (102.39)	162.73 (80.66)	0.0139	164.72 (86.20)

described by quadratic trend. Furthermore, due to the non-crisscrossing of trends representing Type 1 and Type 2 diabetes groups, and with family history and no family history groups Type  $\times$  time and Family history  $\times$  time were not included in the mixed models. However, the trend representing male and female crossing at one time point. Therefore, we start with very general model that includes time (in quadratic form), other fixed effects and the necessary interactions, that is

$$E(Y_{ij}) = \beta_0 + \beta_1 \text{time} + \beta_2 \text{time}^2 + \beta_3 \text{Age} + \beta_4 \text{Gender} + \beta_5 \text{Gender} \times \text{time} \\ + \beta_6 \text{Type} + \beta_7 \text{F.History} + \beta_8 \text{Weight} + \beta_9 \text{Weight} \times \text{time}, \quad (8)$$

where Type and F.History represent diabetes type and family history of diabetes, respectively.

#### Variance-covariance structure for random effects

The above mean structure fitted with subject-specific random intercepts, linear random time effects and quadratic random time effects. For each of the models, the independent error structure is assumed and the results are given in Table 3.

The fixed effect estimates were consistent in sign but have slight differences in magnitude across the three different variance-covariance structures. The variables age, gender, diabetes type, family history, and time

by weight and gender by time interactions were statistically nonsignificant in all models, except for time by weight interaction where its  $p$ -value marginally significant for subject-specific random intercept and slope model (i.e. linear random effects model). The covariates that were statistically significant at 5% level, i.e. *Time*, *Time*<sup>2</sup> and *weight* and the time by weight interaction were retained for the subsequent analysis.

The Crainiceanu and Ruppert [22] RLRT statistic for testing  $H_0 : \sigma_{b_0}^2 = 0$  against  $H_1 : \sigma_{b_0}^2 > 0$  in model  $M_1$  takes the value RLRT = 738.24 with  $p$ -value < 0.0001. The large value of the test statistic or a very small  $p$ -value strongly suggests a rejection of the null hypothesis (i.e.  $H_0 : \sigma_{b_0}^2 = 0$ ) that no subject-specific random effects should be included in the model. Similar tests were conducted using the exact finite sample null distribution of the RLRT statistic of Greven et al. [21] to test  $H_0 : \sigma_{b_1}^2 = 0$  against  $H_1 : \sigma_{b_1}^2 > 0$  and  $H_0 : \sigma_{b_2}^2 = 0$  against  $H_1 : \sigma_{b_2}^2 > 0$  in models  $M_2$  and  $M_3$ , respectively. The RLRT statistic is 3.944 with  $p$ -value = 0.0207 for  $H_0 : \sigma_{b_1}^2 = 0$  indicating rejection of the null hypothesis which implies the need for subject-specific random slopes. Whereas the RLRT statistic for  $H_0 : \sigma_{b_2}^2 = 0$  is 0.639 with  $p$ -value = 0.1859 suggesting a non-rejection of the null hypothesis  $H_0 : \sigma_{b_2}^2 = 0$  which implies no quadratic random effect should be included in the model. Therefore, in the subsequent

**Table 3** Parameter estimates (standard errors, s.e.),  $p$ -values for associated  $t$ -tests and model fit criteria, FBS level of diabetes patients in JUSH, September 2011 - June 2014

Effects	Variance-components					
	Random intercept		Linear random effects		Quadratic random effects	
	Estimate (s.e.)	$p$ -value	Estimate (s.e.)	$p$ -value	Estimate (s.e.)	$p$ -value
<i>Fixed effects</i>						
Intercept	304.362 (14.616)	< 0.0001	306.756 (15.743)	< 0.0001	303.139 (15.678)	< 0.0001
Age	0.252 (0.183)	0.1693	0.212 (0.179)	0.2362	0.197 (0.179)	0.2699
Gender, Male	-2.605 (5.487)	0.6352	-1.968 (5.983)	0.7424	-2.609 (5.933)	0.6603
Diabetes type, Type 2	-9.758 (8.697)	0.2624	-10.553 (8.814)	0.2317	-10.581 (8.852)	0.2325
Family history, Yes	-12.763 (8.478)	0.1328	-12.335 (8.606)	0.1523	-12.593 (8.643)	0.1457
Time	-4.462 (0.870)	< 0.0001	-5.614 (1.071)	< 0.0001	-5.549 (1.116)	< 0.0001
Time <sup>2</sup>	0.123 (0.018)	< 0.0001	0.135 (0.020)	< 0.0001	0.153 (0.025)	< 0.0001
Weight	-1.981 (0.196)	< 0.0001	-1.991 (0.216)	< 0.0001	-1.906 (0.215)	< 0.0001
Time $\times$ Weight	0.016 (0.013)	0.2139	0.032 (0.016)	0.0439	0.025 (0.016)	0.1162
Gender, Male $\times$ Time	-0.412 (0.363)	0.2563	-0.482 (0.443)	0.2761	-0.425 (0.444)	0.3390
<i>Variance components</i>						
$var(b_0)$	2135.023		2797.766		3352.606	
$var(b_1)$			4.575		40.343	
$var(b_2)$					0.048	
Residual	5023.386		4873.227		4723.609	

analysis we use the following parametric linear mixed model, called  $M_4$ :

$$y_{ij} = \beta_0 + \beta_1 \text{weight} + \beta_2 t_{ij} + \beta_3 t_{ij} \times \text{weight} + \beta_4 t_{ij}^2 + b_{0i} + b_{1i} t_{ij} + e_{ij}. \quad (9)$$

The analysis results for model  $M_4$  are presented in Table 4. Except the *time*  $\times$  *weight* interaction effect, which is marginally non-significant at 5% level, all the fixed effects are highly significant.

### Semi-parametric mixed model

The observed mean FBS level profile of patients is shown in Fig. 1b. The plot in this figure shows that the linearity assumption is not reasonable. Therefore, the analysis had to account for the longitudinal data structure and the observed nonlinearity of FBS level estimated with smooth effects in the mixed model framework. Given our specific interest in rate of change in FBS level due to clinic interventions, its functional form (over time) can affect the rate of change. Since the rate of change involves taking derivatives of the smooth function  $f(t_{ij})$ , we choose to use quadratic penalized spline to model the FBS level mean response [31]. Following results from the previous section, we propose the following semi-parametric mixed model with linear random effects structure, called  $M_5$

$$y_{ij} = \beta_0 + \beta_1 \text{weight} + \beta_2 t_{ij} + \beta_3 t_{ij} \times \text{weight} + \beta_4 t_{ij}^2 + \sum_{l=1}^K b_l (t_{ij} - \kappa_l)_+^2 + \sum_{u=0}^1 b_{u_i} t_{ij}^u + e_{ij}. \quad (10)$$

Using appropriately constructed matrices this model can be represented using a matrix notation of “Tests for zero variance components” section. This model is fitted using the random intercept and linear random

**Table 4** Parameter estimates (standard errors, s.e.) and  $p$ -values for associated  $t$ -tests for model  $M_4$ , FBS level of diabetes patients in JUSH, September 2011 - June 2014

Effects	Estimate (s.e.)	$p$ -value
<i>Fixed effects</i>		
Intercept	302.931 (13.330)	< 0.0001
Time	-5.815 (1.061)	< 0.0001
Weight	-1.968 (0.212)	< 0.0001
Time $\times$ Weight	0.031 (0.016)	0.0509
Time <sup>2</sup>	0.134 (0.020)	< 0.0001
<i>Variance components</i>		
$var(b_0)$	2797.887	
$var(b_1)$	4.601	
Residual	4877.259	

effects variance structures of the previous section and the results are displayed in Table 5.

The results in Table 5 show that the fixed effects estimates were consistent in sign but have slight difference in magnitude in both semi-parametric and parametric mixed models (see Table 4), except for the effect of time where both the sign and magnitude of its coefficient estimates were different in the two models and the effect of “time square” was nonsignificant in the semi-parametric mixed models. Further, the interaction of weight with time was not statistically significant in any of the semi-parametric mixed model. Except for the subject-specific random slope variance component, there is a slight decrease in subject-specific random intercept and residual variance components in the semi-parametric model compared to variance components in the linear mixed model  $M_4$  (see Table 4).

To compare the two variance structures under the semi-parametric mixed model given in Eq. (10), we computed AIC, BIC and adjusted AIC (see Table 6). Adjusted AIC shows that the semi-parametric mixed model with subject-specific intercepts as well as slopes (or random linear effects) value is smaller than that of the random intercept. Therefore, the semi-parametric model with random linear effects is the preferred model.

### Model selection

In this section we are focusing on assessing whether the inclusion of spline effects improves model fit compared to parametric counterpart. This is equivalent to testing  $H_0 : \sigma_b^2 = 0$  versus  $H_1 : \sigma_b^2 > 0$  in model  $M_5$ , where  $\sigma_b^2$  is a variance component controlling the smoothness of

$$f(t_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \sum_{l=1}^K b_l (t_{ij} - \kappa_l)_+^2.$$

The approximate  $F$ -test statistic for testing the above hypotheses, i.e. quadratic form of  $f(t_{ij})$  against a quadratic penalized splines, is 83.63 with  $p$ -value < 0.0001. This strongly suggests a rejection of the null hypothesis  $H_0 : \sigma_b^2 = 0$ . Thus, the shape of the function  $f(t_{ij})$  is statistically different from a quadratic trend.

Furthermore, consider the semi-parametric mixed model  $M_5$  in Eq. (10) with random linear effects variance-covariance structure and the linear mixed model  $M_4$  in Eq. 9. The fit statistics from fitting these two models are displayed in Table 6. The  $-2 \log(Lik)$ , AIC and BIC values indicate a substantial improvement in the fit of  $M_5$  compared to  $M_4$ , implying model with penalized spline representation of FBS level was preferred over its parametric counterpart.

The overall results show that, out of the models evaluated, FBS level of diabetes patients at the JUSH diabetic clinic during the study period best characterized by a

**Table 5** Parameter estimates (standard errors, s.e.),  $p$ -values for associated  $t$ -tests and variance components estimates of semi-parametric models under various variance structures, FBS level of diabetes patients in JUSH, September 2011 - June 2014

Effects	Variance structures			
	Random intercept		Linear random effects	
	Estimate (s.e.)	$p$ -value	Estimate (s.e.)	$p$ -value
<i>Fixed effects</i>				
Weight	-1.908 (0.191)	< 0.0001	-1.899 (0.212)	< 0.0001
Time	28.264 (6.087)	< 0.0001	26.742 (6.359)	< 0.0001
Time × Weight	0.017 (0.013)	0.1837	0.031 (0.016)	0.0536
Time <sup>2</sup>	0.408 (0.402)	0.3095	0.448 (0.421)	0.2875
s(Time)Fx1	-2971.649 (551.992)	< 0.0001	-3014.737 (579.734)	< 0.0001
<i>Variance components</i>				
Standard deviation				
Intercept	2104.479		2796.166	
Linear			4.814	
Residual	4919.429		4762.647	
s(Time)	13.287	< 0.0001	13.939	< 0.0001

penalized spline model with truncated quadratic basis, with subject-specific random intercept and slope effects and with linear function of weight and time, called the final model,  $M_6$ .

#### Simultaneous confidence band

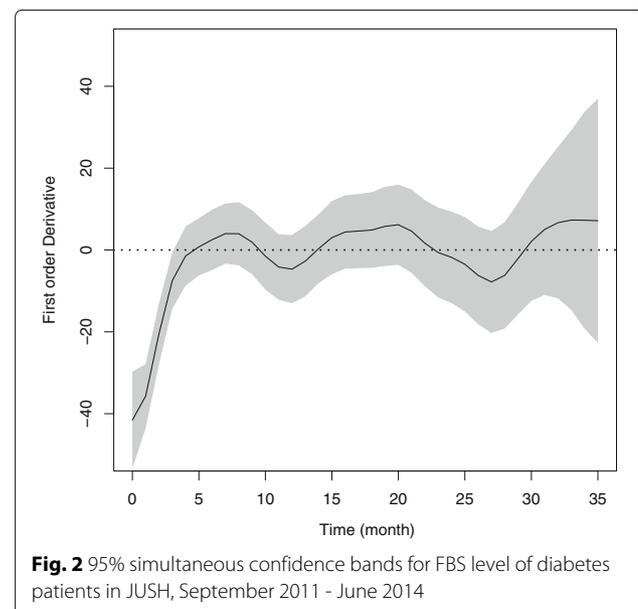
The first derivative of mean response function, i.e.  $\hat{f}'(\cdot)$ , with respect to time was estimated for the final model,  $M_6$  holding weight constant. The rate of change in mean response of FBS level then investigated using the 95% simultaneous confidence bands for the model. The confidence bands were constructed following the discussion in “[Estimation of parameters](#)” section. A gride  $\mathbf{g}$  of time points (0, 35) were defined by increments of one month such that there are  $T = 36$  equally spaced time points. The resulting simultaneous confidence bands displayed in Fig. 2 where the solid line and shaded region represent, respectively, the mean predicted FBS level and the confidence bands. Visual inspection indicates that on average diabetes patients were able to decrease or control their FBS level, due to JUSH clinic interventions, in the first five months period after their initial visit. However, after

month 5, the slope of the curve starts changing it signs, this might imply that patients do actually not follow-up the intervention packages properly or not come to the clinic for treatment due to some unknown reason.

The confidence bands become noticeably wider after 27 months of follow-up period, demonstrating the increased variability. This increase may be due to a smaller number of FBS level recordings being observed at the later period of the study or a potential artifact induced by the spline smoothing [32]. In practice spline smoothing creates a challenge in semi-parametric regression settings through the inherent bias from using truncated basis functions.

**Table 6** Fit statistics for model  $M_5$  and  $M_4$ , FBS level of diabetes patients in JUSH, September 2011 - June 2014

Fit statistics					
Variance structure	$-2 \log(Lik)$	AIC	BIC	$E_p$	$AIC_{adj}$
$M_5$					
Random intercept	50538.54	50554.54	50605.63	7.087	50545.627
Random linear	50507.09	50527.09	50590.96	7.260	50514.350
$M_4$					
	50583.51	50601.51	50658.98		

**Fig. 2** 95% simultaneous confidence bands for FBS level of diabetes patients in JUSH, September 2011 - June 2014

The confidence bands obtained for FBS level does not account for this function bias. However, this bias could be corrected, e.g. using bootstrapping methods [33].

## Discussion

This study focused on longitudinal data analysis of fasting sugar level of adult diabetic patients at Jimma University Specialized Hospital diabetic clinic using an application of semi-parametric mixed model. The study revealed that the rate of change in FBS level in diabetic patients, due to the clinic interventions, does not continue as a steady pace but changes with time and weight of patients. Furthermore, it clarified the associations between FBS level and some characteristics of adult diabetic patients that weight of a diabetes patient has a significant negative effect whereas patient gender, age, type of diabetes and family history of diabetes did not have a significant effect on the change of FBS level. The result on gender agrees with the findings of [34] where the gender effect on fasting blood glucose level of type 2 diabetes was statistically nonsignificant.

Under the two variance-covariance structures of subject-specific random effects, the semi-parametric mixed models had better fit than their parametric counterparts. This was likely due to the localized splines which captured more variability in FBS level than the linear mixed models. The methodology used in the analysis has implications for clinical monitoring in regular followup of diabetic patients and to assess the effect of intervention packages, such as healthy diet, physical activity and prescribed medicines, because individualized curve may be obtained to follow patient-specific FBS level trends [31].

The main limitation of the study is the limited information on important predictors such as type of interventions including treatment types and nutritional status of a patient that may have influenced the rate of change in FBS level. Due to lack of data on these potential predictors for most of the patients involved in the study, we were unable to include them in the analyses. Therefore, more public health and epidemiology researches are needed to examine the impact of treatments and interventions on population health in general and in particular, people living with diabetes to avoid its complications over time and to identify new risk factors for diabetes.

## Conclusion

In this paper, we demonstrate the use of semiparametric mixed effect model for estimation of the rate of change of fasting blood sugar (FBS) level experienced by patients over time. The proposed method can help a physician in clinical monitoring of diabetic patients and to assess the effect of intervention packages such as healthy diet, physical activity.

## Abbreviations

AIC: Akaike's information criteria; BIC: Bayesian information criteria; FBS: Fasting blood sugar level; IDF: International diabetes federation; JUSH: Jimma University specialized hospital; P-IRLS: Penalized iteratively reweighted least squares; REML: Restricted maximum likelihood; RLRT: Restricted likelihood ratio tests; SD: Standard deviation

## Acknowledgements

We thank the Diabetic Clinic of Jimma University Specialized Hospital, Ethiopia for giving access to the data and the staff members of the clinic for their support in extracting the information from patient's medical card.

## Funding

This study project was awarded funding from College of Natural Sciences Research and Postgraduate Office. The second author's Jimma University visit, when this study started, was financially supported by the Africa Interaction, Knowledge, Interchange and Collaboration (KIC) funding (UID: 105298) from the National Research Foundation (NRF) and the University of South Africa under the Foreign general research grant.

## Availability of data and materials

The data sets analyzed in this study available from the corresponding author on reasonable request.

## Authors' contributions

BBY and WKY contributed to the study concept and design. ZMN participated in the data collection and checked the data. TTA performed the statistical analyses and drafted the manuscript. LKD reviewed the findings of data analyses and was a major contributor in writing the manuscript. All authors approved the final manuscript.

## Ethics approval and consent to participate

Ethical approval to conduct the study and human subject research approval for this study was received from Jimma University, College of Natural Sciences, Research Ethics Committee and the medical director of the Hospital. As the study was retrospective, informed consent was not obtained from the study participants, but data were anonymous and kept confidential.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Department of Statistics, Jimma University, Jimma, Ethiopia. <sup>2</sup>Department of Statistics, University of South Africa c/o Christiaan de Wet Road & Pioneer Avenue, Private Bag X6, Florida 1710 Johannesburg, South Africa.

<sup>3</sup>Department of Epidemiology and Biostatistics, College of Medicine and Health Sciences, Bahir Dar university, Bahir Dar, Ethiopia. <sup>4</sup>Department of Data science, The university of Mississippi Medical center, 2500 North State, 39216 Jackson, MS, USA. <sup>5</sup>Arthritis Research UK Centre for Epidemiology, Division of Musculoskeletal and Dermatological Sciences, The University of Manchester, Manchester, UK.

Received: 30 May 2018 Accepted: 17 December 2018

Published online: 10 January 2019

## References

- World Health Organization and others: Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications: Report of a WHO Consultation. Part 1, Diagnosis and Classification of Diabetes Mellitus. 1999. World Health Organization and others.
- Campos C. Chronic hyperglycemia and glucose toxicity: pathology and clinical sequelae. *Postgrad Med.* 2012;124(6):90–7.
- Rahman S, Rahman T, Ismail AA-S, Rashid ARA. Diabetes-associated macrovasculopathy: pathophysiology and pathogenesis. *Diabetes Obes Metab.* 2007;9(6):767–80.

4. Lu C-C, Jiann B-P, Sun C-C, Lam H-C, Chu C-H, Lee J-K. Association of glycemic control with risk of erectile dysfunction in men with type 2 diabetes. *J Sex Med.* 2009;6(6):1719–28.
5. Enzlin P, Rosen R, Wiegel M, Brown J, Wessells H, Gatcomb P, Rutledge B, Chan K-L, Cleary PA, et al. Sexual dysfunction in women with type 1 diabetes: long-term findings from the dcct/edic study cohort. *Diabetes Care.* 2009;32(5):780–5.
6. World Health Organization. Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus. Abbreviated report of a WHO consultation. Geneva: World Health Organization; 2011, p. 25. Available from: (<http://www.ncbi.nlm.nih.gov/books/NBK304267/>).
7. World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: Report of a WHO/IDF consultation (PDF). Geneva: World Health Organization; 2006, p. 21. ISBN: 978-92-4-159493-6.
8. Smith DE. Type 1 Diabetes Risk Factor. 2016. <https://www.endocrineweb.com/conditions/type-1-diabetes/type-1-diabetes-risk-factors>.
9. Leontis LM, Hess-Fischl AM. Type 2 Diabetes Risk Factor. 2014. <https://www.endocrineweb.com/conditions/type-2a-diabetes/type-2-diabetes-risk-factors>.
10. Atlas IDF. International Diabetes Federation 7th Edition, p. 2015. <http://www.diabetesatlas.org>.
11. Worku D, Hamza L, Woldemichael K. Patterns of diabetic complications at jimma university specialized hospital, southwest ethiopia. *Ethiop J Health Sci.* 2010;20(1):33–39.
12. SM A, Y B, A W, S A. Increasing trends of diabetes mellitus and body weight: A ten year observation at gondar university teaching referral hospital, northwest ethiopia. *PLoS ONE.* 2013;8(3):60081.
13. Thilakarathne PJ, Clement L, Lin D, Shkedy Z, Kasim A, Talloen W, Versele M, Verbeke G. The use of semiparametric mixed models to analyze pamchip@peptide array data: an application to an oncology experiment. *Bioinformatics.* 2011;27(20):2859–65.
14. Szczesniak RD, Li D, Amin RS. Semiparametric mixed models for nested repeated measures applied to ambulatory blood pressure monitoring data. *J Mod Appl Stat Methods.* 2016;15(1):255–75.
15. Mehari Z. Application of Longitudinal Data Analysis on FBS of Adult Diabetes: Statistical Analysis of Fasting Blood Sugar Level of Adult Diabetic Patients in Jimma University Specialized Hospital. Saarbrücken, Germany: LAP Lambert Academic Publishing 2014; 2014.
16. Ruppert D, Wand M, Carroll R. Semiparametric regression. Cambridge, UK: cambridge University Press; 2003.
17. Maringwa JT, Geys H, Shkedy Z, Faes C, Molenberghs G, Aerts M, Ammel KV, Teisman A, Bijmens L. Application of semiparametric mixed models and simultaneous confidence bands in a cardiovascular safety experiment with longitudinal data. *J Biopharm Stat.* 2008;18(6):1043–62.
18. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* New York: Springer; 2000.
19. Stram D, Lee J. Variance components testing in the longitudinal mixed effects model. *Biometrics.* 1994;50(3):1171–7.
20. Scheipl F, Greven S, Küuchenhoff H. Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comput Stat Data Anal.* 2008;52(7):3283–3299.
21. Greven S, Crainiceanu C, Küuchenhoff H, Peters A. Restricted likelihood ratio testing for zero variance components in linear mixed models. *J Comput Graph Stat.* 2008;17(4):870–91.
22. Crainiceanu CM, Ruppert D. Likelihood ratio tests in linear mixed models with one variance component. *J R Stat Soc Ser B (Statistical Methodology).* 2004;66(1):165–185.
23. Wang Y. Smoothing spline models correlated random errors. *J Am Stat Assoc.* 1998;93:341–8.
24. Durbán M, Harezlak J, Wand M, Carroll R. Simple fitting of subject-specific curves for longitudinal data. *Stat Med.* 2005;24(8):1153–67.
25. Ni X, Zhang D, Zhang HH. Variable selection for semiparametric mixed models in longitudinal studies. *Biometrics.* 2010;66(1):79–88.
26. Wager C, Vaida F, Kauermann G. Model selection for penalized spline smoothing using akaike information criteria. *Aust N Z J Stat.* 2007;49(2): 173–90.
27. Crainiceanu CM, Ruppert D, Claeskens G, MP W. Exact likelihood ratio tests for penalized splines. *Biometrika.* 2005;92(1):91–103.
28. Hastie T, Tibshirani R. *Generalized Additive Models.* New York: Wiley Online Library; 1990.
29. Wood S. *Generalized Additive Models: an Introduction with R.* Boca Raton, Florida: CRC press; 2006.
30. Hastie T, Tibshirani R. Generalized additive models for medical research. *Stat Methods Med Res.* 1995;4(3):187–96.
31. Szczesniak RD, McPhail GL, Duan LL, Macaluso M, Amin RS, Clancy JP. A semiparametric approach to estimate rapid lung function decline in cystic fibrosis. *Ann Epidemiol.* 2013;23(12):771–7.
32. Szczesniak RD, Li D, Raouf AS. Semiparametric mixed models for medical monitoring data: An overview. *J Biom Biostat.* 2015;6(2):234. <https://doi.org/10.4172/2155-6180.1000234>.
33. Hardle W, Bowman AW. Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J Am Stat Assoc.* 1988;83(12):102–110.
34. Khan HA, Sobki SH, Alhomida AS. Regression analysis for testing association between fasting blood sugar and glycated hemoglobin in diabetic patients. *Biomed Res.* 2015;26(3):604–6.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

