# Application of semiparametric model in modelling diabetic retinopathy among type II diabetic patients

**Bezalem Eshetu Yirdaw** ( 12962805@mylife.unisa.ac.za )
University of South Africa
**Legesse Kassa Debusho**
University of South Africa

**Research Article**

**Keywords:** Covariate by factor interaction, diabetes mellitus, diabetic retinopathy, semiparametric model, tensor product interaction

**Posted Date:** April 13th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1532336/v1

RESEARCH

# Application of semiparametric model in modelling diabetic retinopathy among type II diabetic patients

Bezalem Eshetu Yirdaw[1*] and Legesse Kassa Debusho[2]

*Correspondence: 12962805@mylife.unisa.ac.za
[1]Department of Statistics, University of South Africa c/o Christiaan de Wet Road & Pioneer Avenue, Florida 1709, Johannesburg, South Africa
Full list of author information is available at the end of the article

**Abstract**

**Background:** The proportion of patients with diabetic retinopathy (DR) has grown with increasing number of diabetic mellitus patients in the world. It is among the top risk factors of blindness worldwide, especially those living in developing countries. The main objective of this study was to identify contributing risk factors of diabetic retinopathy among type II diabetic patients.

**Method:** A sample of 191 type II diabetic patients was selected from the Black Lion Specialized Hospital diabetic unit from 1 March 2018 to 1 April 2018. A multivariate stochastic regression imputation technique was applied to impute the missing values. The response variable, diabetic retinopathy is a categorical variable with two outcomes. Based on the relationship derived from the exploratory analysis, the odds of diabetic retinopathy were not necessarily linearly related to the continuous predictors for this sample of patients. Therefore, a semiparametric model was proposed to identify the contributing factors of diabetic retinopathy.

**Result:** From the sample of 191 type II diabetic patients, 98(51.3%) of them experienced diabetic retinopathy. The results of semiparametric regression model revealed that gender, hypertension, insulin treatment, and frequency of clinical visits had a significant linear relationships with the odds of diabetic retinopathy. In addition, the log- odds of DR has a significant nonlinear relation with the interaction of age by gender (for female patients), duration of diabetes, interaction of cholesterol level by gender (for female patients), haemoglobin A1c, and interaction of haemoglobin A1c by fasting blood glucose with degrees of freedom $3.2, 2.7, 3.6, 2.3$ and $3.7$, respectively. The interaction of age by gender and cholesterol level by gender appear non significant for male patients. The result from the interaction of haemoglobin A1c (HbA1c) by fasting blood glucose (FBG) showed that the risk of diabetic retinopathy is high when the level of HbA1c and FBG were simultaneously high.

**Conclusion:** Clinical variables related to type II diabetic patients were strong predictive factors of diabetic retinopathy. Hence, health professionals should be cautious about the possible effects and complications of diabetic mellitus which can be caused by the clinical variables. Furthermore, to improve intervention strategies similar studies should be conducted across the country.

**Keywords:** Covariate by factor interaction; diabetes mellitus; diabetic retinopathy; semiparametric model; tensor product interaction

## Introduction

Diabetes mellitus (DM) is one of the major noncontagious diseases that has a cumulative consequence on people in both developed and developing countries [1]. International diabetic federation in 2019 has reported that the estimated number of diabetic patients was 463 million, of which 19 million were from Africa including 14.2 million in sub-Saharan Africa [2]. Diabetic retinopathy (DR) is one of the microvascular complications of diabetes mellitus which is characterized by the presence of microaneurysms, haemorrhages, exudation, cotton wool spot, and/or new vessels in the peripheral retina, macula, or both [3, 4, 5]. The progression of DR has four stages; mild non-proliferative diabetic retinopathy (NPDR), moderate (NPDR), severe (NPDR), and proliferative diabetic retinopathy (PDR). In the first three stages, the formulation of microaneurysms may leak fluid and block blood vessels in the retina. On the other hand, PDR is the advanced stage which is characterized by the formulation of new blood vessels that can cause more leakage and bleeding of the retina, and can lead to permanent vision loss [6].

Among the total number of diabetic patients in the world about one third of them develop diabetic retinopathy. Furthermore, it has been the top cause of blindness in the middle age and elderly patients worldwide, of which more than 93 million patients suffer from DR [7]. Out of 10% adult patients worldwide, 7.5% are living in low- and medium-income countries, where healthcare resources are limited [7]. The estimated annual incidence and progression of diabetic-related eye disease ranged from 2.2% to 12.7% and 3.4-12.3%, respectively [5, 8]. Among 1.5 billion blind people in the world, 0.4 million were due to diabetic retinopathy [9]. There was a noticeable reduction in the number of blindness and vision loss in the world. However, the percentage of blindness and moderate to severe vision impairment due to DR increased by approximately 50% and 53%, respectively [9]. Further, the proportion of DR in Africa ranges from 7% to 62.4%, of which severe DR was observed in 15% of the patients. Ethiopia is one among the lower-income countries with a high percentage (3.8%) of adult diabetic community in sub-Saharan Africa [2, 10]. A study reveal that the prevalence of retinopathy among type II diabetic patients and in a group without diabetes was 34.6% and 8.8%, respectively [11].

Studies across the world have shown that, the most predictive factors of diabetic retinopathy are socio-demographic and clinical variables such as age, diabetic duration, lipid profiles of a patient and microalbuminuria [5, 12, 13, 14, 15, 16, 17]. Hussain et al. [15] reported that, gender and clinical variables have significant relationships with DR. Furthermore, glycaemic control and body-mass index have significant associations with diabetic retinopathy [4, 5]. A study from Ethiopia reported that, gender, haemoglobin a1c (HbA1c) and hypertension are predictive risk factors of diabetic retinopathy [17]. Some studies have also revealed that the odds of diabetic retinopathy is higher for a patient with higher HbA1c [4, 12, 15, 18], longer duration of diabetes [5, 12] and hypertensive patients [4, 5, 17]. A study based on data from a meta-analysis of seven cohort studies reported that insulin treatment has significant association with diabetic retinopathy in patients with type

II diabetes mellitus [13]. Another study based on 5.2 years follow up data indicates that variability of fasting plasma glucose (FPG) is a significant predictor of diabetic retinopathy [19]. Ten years follow up study also showed that, as compared to patients who did not experience DR, patients who experienced DR had a higher level of FPG and HbA1c [20].

At Black Lion Hospital (BLH), Addis Ababa, Ethiopia, health professionals at the diabetic unit are well trained in screening DR. Depending upon the stage of DR that a patient experiences, a clinician also provides appropriate treatment to prevent vision loss and blindness. Studies in Ethiopia used a parametric model, e.g., GLM which only identify the linear relationship between the link function and covariates to determine predictive factors of DR [5, 14, 17]. However, because of the incorrect functional form of the model some high risk covariates may be interpreted as no relationship with the response. In 2018, a study at BLH used binary logistic regression and identified gender, duration of diabetes, HbA1c, hypertension and frequency of clinical visit as predictors of diabetic retinopathy [17]. However, in this study all continuous predictors except duration of diabetes were categorized and considered as factors, and linear association between the response and predictors were considered via the logit link. For the current study, we used the same data from Shibru, Aga and Boka [17]. However, plots of continuous predictors from exploratory analysis (see Figure 1) shows nonlinearity and a logistic regression model may be too restrictive to analyse this data. Therefore, to develop effective strategies for the prevention of diabetic retinopathy, estimating the correct functional form of the model is important. Accordingly, the main aim of this research was to identify the contributing risk factors of diabetic retinopathy among type II diabetic patients at Black Lion Hospital and to estimate the data driven relationship between clinical variables, specifically continuous predictors and diabetic retinopathy using semiparametric models. According to [21], the functional form of a covariate in additive model varies across groups defined by levels of categorical variables. Further, the interaction between age and gender of a diabetic patient is epidemiologically plausible for consideration [22]. Therefore, this study was also motivated to estimate the functional form of the interaction effect of a covariate by the levels of categorical variables.

A study shows that there is a strong connection between HbA1c and FBG in a diabetic subject [23]. In a different study, it was reported that the interaction between mean HbA1c and FPG variability has no significant association with the odds of diabetic retinopathy [19]. However, Figure 1, which was generated from the current study data reveals each of the potential covariates HbA1c and FBG has a nonlinear relationship with the odds of diabetic retinopathy. This brings a question of what if the interaction of HbA1c and FBG has a significant nonlinear effect on being diabetic retinopathy. Therefore, we have also considered the interaction of HbA1c and FBG.

The remaining section of the paper is organized as follows. First, a description of the study data and semiparametric models for a binary response are discussed in the

methodology section. In the result section, the findings from applying these models on the study data are illustrated. Finally, discussion and conclusion with recommendation and pointers for future study are given in the discussion and conclusion sections, respectively.

## Methodology

### Study area and data

The data for this study was a secondary data obtained from BLH and the details on the data are available in [17]. BLH hospital is located at Addis Ababa, Ethiopia and it is the largest teaching hospital in Ethiopia. The diabetic unit at BLH gives a service provision for more than 200 individuals per week. For this study, a cross-sectional study design was used. The data was collected from March to April 2018 and all type II diabetic patients who had a follow up at BLH diabetic unit within the study period were eligible for this study. The study excludes critically ill patients who were very weak to give informed consent to participate in the study. A total of 191 patients were used for the analysis. The response variable, diabetic retinopathy is a categorical variable with two outcomes which is measured via direct eye examination using Topcon Camera [24]. Socio-demographic and treatment related variables were collected via face-to-face individual interview, and clinical variables, such as duration of diabetes since a patient confirmed to type II diabetes, frequency of clinical visits, fasting blood glucose, cholesterol level, Haemoglobin A1c and status of hypertension were extracted from patients record. To sum up, this study includes categorical and continuous variables as predictors of diabetic retinopathy, where gender, hypertension, insulin treatment, and frequency of clinical visits were considered as factors and age, duration of diabetes, total cholesterol level, haemoglobin A1c and fasting blood glucose were considered as covariates.

*Semiparametric model for binary response*

Given the exploratory plots in Figure 1, a semiparametric model is more reasonable for this data rather than assumptions based restrictive parametric models. Let a binary outcome variable $y_i$ denotes the diabetic retinopathy status of the $i^{th}$ patient, where $y_i = 1$ represents patient with diabetic retinopathy and $y_i = 0$ represents patient without diabetic retinopathy, let $z_m$ denotes the $m^{th}$ categorical variable, $m = 1, \cdots, M$ and let $x_j$ denotes $j^{th}$ continuous variable, $j = 1, \cdots, J$ then a semiparametric model for the outcome $y_i$ is given by:

$$g(\mu_i) = \alpha_0 + \sum_{m=1}^{M} \sum_{l=1}^{L_m} \alpha_{ml} z_{iml} + h_j(x_{ij}) + f_{z_i}(x_{ij}) + f_{ab}(x_a, x_b), \qquad (1)$$

where $\mu_i = E(y_i)$, $\alpha_0$ is the model constant, $\sum_{m=1}^{M} \alpha_{ml} z_{iml}$ is the parametric term of the model for the categorical variables (gender, hypertension, insulin treatment and frequency of clinical visit), $z_{iml}$ is the $l^{th}$ level of $m^{th}$ categorical variable measured on the $i^{th}$ type II diabetic patients and $\alpha_{ml}$ is the corresponding parameter, $M$ is total number of categorical variables, $L_m$ is number of categories/level of the $m^{th}$ categorical variable, $l = 1, \cdots, L_m$, e.g., when the $m^{th}$ categorical variable has two levels, we have one $\alpha_{ml}$, i.e. $L_m = 1$ because the first category is treated

as a reference category. For example, in this study frequency of clinical visit has three categories (every 1 month, every 3 month and every 6 month), where, every 1 month was treated as a reference category. Additionally, $h_j(x_{ij})$ is a smoothing function for the continuous clinical predictors, $f_{z_i}(x_{ij})$ is a smoothing function for the covariate by factor level interaction, $x_{ij}$ is the $j^{th}$ continious predictor measured on the $i^{th}$ type II diabetic patients and $f_{ab}(x_a, x_b)$ is a smoothing function for the tensor product interaction of two continuous clinical variables $x_a$ and $x_b$. In a semiparametric model, for each level of a factor we have one curve representing a covariate by factor interaction. For example, in the current study we have age by gender interaction which have two separate curves for male and female. To do this define:

$z_i \in \{1, \cdots, L_m\}$ and

$$z_{il} = \begin{cases} 1, & \text{if} \quad z_i = l \\ 0, & else, \end{cases}$$

Thus, the model in Expression (1) can be written as:

$$\begin{array}{rcl} g(\mu_i) & = & \alpha_0 + \sum_{m=1}^{M} \sum_{l=1}^{L_m} \alpha_{ml} z_{iml} + \beta_{1j} x_{ij} + \beta_{2j} x_{ij}^2 + \cdots + \beta_{pj} x_{ij}^p \\ & + & \sum_{k=1}^{K} b_{kj} (x_{ij} - \kappa_{kj})_+^p \\ & + & \sum_{l=2}^{L_m} z_{il} (\gamma_{0l} + \gamma_{1lj} x_{ij} + \gamma_{2lj} x_{ij}^2 + \cdots + \gamma_{plj} x_{ij}^p) \\ & + & \sum_{l=1}^{L_m} z_{il} \{ \sum_{k=1}^{K} c_{kj}^l (x_{ij} - \kappa_{kj})_+ + f_{ab}(x_a, x_b, ) \end{array} \tag{2}$$

$$f_{ab}(x_a, x_b) = \sum_{s_1=0}^{p} \sum_{s_2=0}^{p} \delta_{s_1 s_2} x_{ia}^{s_1} x_{ib}^{s_2} + \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} b_{k_1 k_2} (x_{ia} x_{ib} - \kappa_{k_1 k_2})_+^p,$$

$w_+ = max\{0, w\}$, $\beta_{1j}, \beta_{2j}, \cdots, \beta_{pj}$ are fixed effect parameters for the main effect smoothing functions, $(\gamma_{0l}, \gamma_{1lj}, \cdots, \gamma_{plj})$ are fixed effect parameters for the smoothing function of an interaction of $x_j$ by $L_m$ levels of a factor $z_i$, $x_{ia}$ and $x_{ib}$ are two continuous predictors measured on the $i^{th}$ type II diabetic patients which are considered to have a tensor product interaction effect on the response, $\sum_{s_1=0}^{p} \sum_{s_2=0}^{p} \delta_{s_1 s_2}$ are fixed effect parameters for the tensor product smoothing interaction $x_a \odot x_b$. Finally, $\kappa_{kj}$ are knots where the $p^{th}$ degree spline evaluated at a covariate $x_j$ and covariate by factor interaction of the smoothing term, and $\kappa_{k_1 k_2}$ are knots where the $p^{th}$ degree spline evaluated at the tensor product $x_a \odot x_b$ for the tensor product interaction of the smoothing term, $\sum_{k=1}^{K} b_{kj} (x_{ij} - \kappa_{kj})_+^p$ is the over all smooth term for the main effect, $\sum_{l=1}^{L_m} z_{il} \{ \sum_{k=1}^{k} c_{kj}^l (x_{ij} - \kappa_{kj})_+ \}$ is the deviation from the over all smooth term of the covariate by factor interaction and $\sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} b_{k_1 k_2} (x_{ia} x_{ib} - \kappa_{k_1 k_2})_+^p$ is the overall smooth term for the tensor product smoothing function. According to [25], a Penalized cubic regression spline allows to retain the good properties of splines and has good computational efficiency. Therefore, we have considered a penalized cubic regression spline ($p = 3$) to model nonlinearity of the covariates. The respective random effect coefficients $b_{kj}$, $c_{kj}^l$ and $b_{k_1 k_2}$ were assumed to follow a gaussian distribution, i.e. $b_{kj} \sim N(0, \sigma_{bj}^2)$, $c_k^l \sim N(0, \sigma_{cl}^2)$ and $b_{k_1 k_2} \sim N(0, \sigma_{b_{ab}}^2)$, respectively.

*Proposed semiparametric models*

The locally estimated scatter-plots smoothing presented in Figure 1 suggest that the relationship between the log odds of being DR and each of the continuous clinical variables is nonlinear. Moreover, Figure 3 revealed that there is a variation between the total cholesterol levels of male and female. Therefore, it is worthy to investigate the interactions of age by gender and HbA1c by FBG ($HbA1c \times FBG$. Thus, we proposed five different semiparametric models. We start with a more general model ($M_1$) which includes gender, hypertension, frequency of clinical visit (FCV) and insulin treatment (IT) as a linear term and interactions of age by gender, $HbA1c \times FBG$, total cholesterol level (CL) by gender, and duration of diabetes (DD) as nonlinear terms and $M_1$ therefore defined as

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1\, gender + \beta_2\, hypertension + \beta_3\, IT \\ &+ \beta_4\, FCV + f_{gender}(age) + f(DD) + f(HbA1c) \\ &+ f(FBG) + f(HbA1c, FBG) + f_{gender}(CL) \end{aligned} \tag{3}$$

where, using Expression (2) presentation, for example

$$f(DD_i) = \beta_0 + \beta_1(DD_i) + \beta_2(DD_i)^2 + ... + \beta_p(DD_i)^p + \textstyle\sum_{k=1}^{K} b_k((DD_i) - \kappa_k)_+^p$$

and

$$\begin{aligned} f(HbA1c_i, FBG_i) &= \textstyle\sum_{s_1=0}^{p} \sum_{s_2=0}^{p} \delta_{s_1 s_2} (HbA1c_i)^{s_1} (FBG_i)^{s_2} \\ &+ \textstyle\sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} b_{k_1 k_2} ((HbA1c_i)(FBG_i) - \kappa_{k_1 k_2})_+^p. \end{aligned}$$

The second model $M_2$ was proposed to test the nonlinearity of $HbA1c \times FBG$ and it is given by

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1\, gender + \beta_2\, hypertension + \beta_3\, IT \\ &+ \beta_4\, FCV + \beta_5\, HbA1c + \beta_6\, FBG + \beta_7\, HbA1c \times FBG \\ &+ f_{gender}(age) + f(DD) + f_{gender}(CL) \end{aligned} \tag{4}$$

The third model $M_3$ was proposed to test the nonlinearity of age by gender interaction and it is given by

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1\, gender + \beta_2\, age + \beta_3\, age \times gender \\ &+ \beta_4\, hypertension + \beta_5\, IT + \beta_6\, FCV \\ &+ f(DD) + f(HbA1c, FBG) + f_{gender}(CL) \end{aligned} \tag{5}$$

The fourth model $M_4$ was proposed to test the nonlinearity of total cholesterol level by gender interaction and it is given by

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1\, gender + \beta_2\, CL + \beta_3\, CL \times gender \\ &+ \beta_4\, hypertension + \beta_5\, IT + \beta_6\, FCV \\ &+ f_{gender}(age) + f(DD) + f(HbA1c, FBG) \end{aligned} \tag{6}$$

The fifth model $M_5$ was proposed to test the nonlinearity of duration of diabetes and it is given by

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1\, gender + \beta_2\, hypertension \\ &+ \beta_3\, IT + \beta_4\, FCV + \beta_5\, DD + f_{gender} age \\ &+ f(HbA1c, FBG) + f_{gender}(CL) \end{aligned} \tag{7}$$

*Estimation of parameters*

Estimation of both penalized and unpenalized coefficients in the above models was done using penalized iterative reweighted least squares (PIRLS). We have used evenly spaced knots with $k = 10$ in the ranges of the covariate $x_j$ for main effect and for covariate by factor interaction of the smoothing functions, and $k = 8$ for the tensor product interaction [25]. Since under finite sample size, prediction error criteria, such as generalized cross validation (GCV) (for the known scale parameter) and unbiased risk estimator (UBRE) (for the unknown scale parameter), is more likely to develop multiple minima which undersmooth the function $f_j$ relative to restricted maximum likelihood (REML), therefore the smoothing parameter selection in the analyses was done using REML and data analysis was done using `gam` function from `mgcv` package in R statistical software. For the detailed information on parametric estimation and modeling of semiparametric model see [25] and for covariate by factor interaction of a smoothing function see [21].

*Test of nonlinearity*

The hypothesis test for a statistically significance of a nonlinear effect of a continuous covariate $x_j$ was done using the likelihood ratio test by fitting two models, that is, we fit first a model where $x_j$ has a linear relationship and then a second model with a nonlinear relationship. Then the hypothesis is, there is a linear relationship between the covariate $x_j$ and the response against there is no linear relationship between the covariate $x_j$ and the response.

## Results

*Missing data imputation*

The presence of missing observations in some of the variables in a data has an effect on statistical inference, such as poor precision on confidence intervals and biased on parameter estimates, which may result poor statistical power [26]. Therefore, we imputed the missing values of variables with more than 5% missing values using multivariate stochastic regression imputation technique [27]. Furthermore, missing observations in two variables, cholesterol level and HbA1c which had 9% and 50% missing values, respectively were imputed using the above technique. As it can be seen in Figure 2, the distribution for the imputed values and observed values are similar.

*Descriptive statistics*

There were a total of 191 type II diabetic patients in the study, of which 98 (51.3%) experienced diabetic retinopathy. Among the total sample, 114 (59.7%) were female, of which 51 (26.7%) of them experienced diabetic retinopathy. More than half (54.8%) of the total patients in the study were hypertensive, of which 76 (39.8%) of them experienced diabetic retinopathy. Out of patients who used insulin treatment, 40 (20.9%) of them experienced diabetic retinopathy. The total number of patients whose clinical visit was within 1 month interval was 43 (22.6%), of which 7.4% of them experienced diabetic retinopathy. Moreover, of those patients whose clinical visit were within 3 and 6 months interval, 38 (20%) and 46 (24.2%) of them experienced diabetic retinopathy, respectively. The mean age of patients who experienced diabetic retinopathy was 58.5 years with a standard deviation of 10.1 years (Table 1). The average diabetic duration since a patient was confirmed to type II diabetes mellitus till the patient experienced diabetic retinopathy was 15.2 years with a standard deviation of 10.9 years. In this study, HbA1c is measured in percentage form known as Diabetes Control and Complications Trial (DCCT) units. The mean HbA1c for a patient with diabetic retinopathy was 9.3% with a standard deviation of 2.8%. In addition, the average total cholesterol level for a patient who experienced diabetic retinopathy was 187.6 mg/dL. The mean FBG for a patient who experience diabetic retinopathy was 170.9 mg/dL with a standard deviation of 66 mg/dL (Table 1).

As it can be seen from the box plot in Figure 3, the median total cholesterol level for female was around 186 mmol/L and the median cholesterol level for male patient was around 171 mmol/L.

*Test of multicollinearity and nonlinearity*

The covariates were checked for multicollinearity using the variance inflation factor (VIF) before adding them to the model. None of these VIFs (the values are between 1.08 and 1.21) were greater than 5 suggesting the collinearity is not strong to affect the statistical inference in the analysis. Next, the five proposed models in the methodology section were fitted and a likelihood ratio test was used to test the nonlinearity of continuous covariates. As it can be seen from Table 2, the deviance for testing the nonlinearity of the interaction of age by gender is 11.98 with $p - value = 0.0461$, indicating that there was a significant nonlinear relationship between the odds of diabetic retinopathy and the interaction of age by gender. Similarly, the deviance for the nonlinearity test of cholesterol level by gender is 37.20 with $p-value = 0.0012$. Thus, there was a significant nonlinear relationship between the odds of diabetic retinopathy and the interaction of cholesterol level by gender. The likelihood ratio test for the relationship between the odds of diabetic retinopathy and duration of diabetes has deviance equals to 13.02 with $p - value = 0.0228$ (Table 2), therefore, the relationship was significantly nonlinear. The nonlinearity test for the interaction of HbA1c and FBG was also significant ($p-value = 0.0157$) supporting the nonlinear relationship.

*Model selection*

In this section, we are focusing in selecting the best model which fits the data very well using Akaike's Information Criterion (AIC). As it can be seen from Table 3, $M_1$ is a model with the smallest AIC value (163.64) which supports the nonlinearity test in Table 2. Therefore, the final model which best explains the diabetic retinopathy data for a patient at Black Lion Hospital during the study period was $M_1$ which includes gender, hypertension, insulin treatment and frequency of clinical visit as a linear term and the interaction of age by gender, duration of diabetes, tensor product or interaction of HbA1c and FBG, and cholesterol level by gender as smooth functions with the cubic spline bases. Therefore, the result in the next section is based on $M_1$.

*Semiparametric multivariable analysis*

The results from fitting $M_1$ are displayed in Table 4, Figure 4 and Figure 5. Keeping the effects of being hypertensive, insulin treatment, frequency of clinical visit, interaction of age by gender, duration of diabetes, $HbA1c \times FBG$ and interaction of total cholesterol level by gender constant, the odds of diabetic retinopathy for a male patient was 3.5 (95% CI:1.14-11.09) times higher than that of female patients. Keeping the effect of other covariates constant, the odds of diabetic retinopathy was significantly higher for the hypertensive patient (adjusted odds ratio (AOR)=38.9, 95% CI: 9.85-153.23). The odds of diabetic retinopathy for a patient who used insulin treatment was 6.2 (95% CI: 1.81-13.84) times higher than the odds of diabetic retinopathy for a patient who did not use insulin treatment to control their blood glucose level keeping the effect of other covariates constant. Keeping the effect of other covariates constant, the odds of diabetic retinopathy for a patient whose clinical visit was every 3 months was 8.7 (95% CI: 2.13-35.9) times higher than the odds of retinopathy for a patient whose clinical visit was every 1 month. Similarly, the odds of diabetic retinopathy was higher for a patient who had follow-up every 6 months (AOR=6.7, 95% CI: 1.63-27.41) as compared to a patient who had follow-up every one month keeping the effect of other covariates constant.

The result in Table 4 illustrates that holding the effects of other covariates constant, there was a significant nonlinear relationship between the log odds of diabetic retinopathy and age of female patients ($p-value = 0.0357$) with estimated degrees of freedom 3.2. Furthermore, visual inspection of Figure 4(a) shows that the log odds of diabetic retinopathy for female patients increase slightly with age at the begning, but it shows a gradual decline after the age of 65 years: the confidence band is very wide in this age range, it could be because of a few number of patients older than 65 years who experienced diabetic retinopathy. Table 4 also reveals that, duration of diabetes had a significant nonlinear relationship ($p - value = 0.0059$) with the log odds of diabetic retinopathy. Moreover, according to Figure 4(b), the functional relationship between duration of diabetes and log odds of diabetic retinopathy looks inverted U-shape with estimated degrees of freedom 2.7. However, the confidence band after 30 years of duration of diabetes becomes notably wider, indicating greater variability which may be due to a small number of observations in that interval. As it can be seen in Figure 4(c), the finding of this study also indicates that the

relationship between the log odds of diabetic retinopathy and female cholesterol level was initially flat, but a moderate increment in the log odds of DR for a female patient was observed for a total cholesterol level $> 250mg/dL$. Table 4 also shows that there was a significant nonlinear relationship between the log odds of diabetic retinopathy and female cholesterol level ($p-value = 0.0166$) with degrees of freedom 3.6. Similarly, there was a significant nonlinear relationship between the log odds of diabetic retinopathy and HbA1c ($p-value = 0.0020$) with estimated degrees of freedom 2.3. As it can be seen in Figure 4(d), the log odds of diabetic retinopathy has an increasing pattern when the patient HbA1c is between $6\%-11\%$ and flat pattern was observed for HbA1c greater than 11%. However, the confidence band at the initial (for HbA1c between (0-4)%) and at the end (for HbA1c $> 11\%$) was wide, which may be due to greater variability at these intervals.

There was a significant nonlinear relationship between the log odds of diabetic retinopathy and $HbA1c \times FBG$ ($p-value = 0.0500$) with degrees of freedom 3.7 (Table 4). The 3D contour plot in Figure 5 indicates that, the risk of diabetic retinopathy increases with increasing HbA1c slowly for the patient with low FBG and the risk was higher for high FBG-HbA1c combinations. Furthermore, the 2D contour plot also shows that the risk of diabetic retinopathy was higher when both FBG and HbA1c were simultaneously high. The darker red region indicates that the risk of being diabetic retinopathy was minimum for the low percentage of HbA1c. Moreover, the combination of $HbA1c \geq 6\%$ and $FBG \geq 150mg/dL$ shows a relatively high risk of diabetic retinopathy. Generally, the distribution of numerical values (value of linear predictor) on the contour lines in the three regions; dark red (low risk), light red (intermediate risk), and yellow (high risk) of the plots tell the nonlinear relationship between the linear predictor measuring the risk of diabetic retinopathy and $HbA1c \times FBG$.

## Discussion

This study was aimed to identify the contributing factors of diabetic retinopathy using data collected from Black Lion Hospital at Addis Ababa, Ethiopia. In the current study, rather than using statistical methods which impose some parametric assumptions, we focused on the data-driven relationship. The results from applying semiparametric regression analysis on the data showed that the odds of diabetic retinopathy had a significant linear association with gender, hypertension, insulin treatment and frequency of clinical visit. In addition, the log odds of diabetic retinopathy had a significant nonlinear association with the interaction of age by gender (for female patients), duration of diabetes, interaction of cholesterol level by gender (for female patients) and the interaction of HbA1c by FBG.

It was observed that being hypertensive was a strong predictive factor of DR. This finding agrees with previous studies. For example, a cross-sectional study conducted by [28] showed that systolic blood pressure was significantly associated with DR. The result of a case-control study [29] in 1000 patients in two groups, patients with retinopathy and patients without retinopathy revealed that those with DR had high

systolic blood pressure than patients without DR indicating that systolic blood pressure was one of the contributing factors of DR. Another cross-sectional study [4] using univariate logistic regression analysis reported hypertension as a major risk factor of DR. Tilahun et al. [5], identified hypertension as one of the predictors of diabetic retinopathy. In their study, the odds of being diabetic retinopathy for hypertensive patients appeared 3.39 times higher than that of non-hypertensive patients.

According to the current study, insulin treatment was one of the contributing factors of DR. Further, patients who used insulin treatment were more likely at risk than patients who did not use insulin treatment. This result is similar to a study that was based on data from a meta-analysis of seven cohort studies [13]. Their pooled result of the seven-cohort studies showed that insulin use increases the risk of being diabetic retinopathy, where the estimated relative risk was equal to 2.3. One of the other findings of this study was the odds of DR significantly differ for a patient with their number of clinical visits, i.e., patients who visited the diabetic clinic every 3 months and 6 months had a higher risk of being diabetic retinopathy than patients who visited the diabetic clinic every one month.

In a nonlinear terms of a semiparametric analysis, some interaction terms were incorporated based on scientific literature and exploratory analysis, i.e., age by gender, cholesterol level by gender, and the tensor product or interaction of HbA1c and FBG. As it was discussed in the descriptive statistics section, almost half of the study participants experienced diabetic retinopathy. Therefore, understanding both additive and interaction effects of those socio-demographic and clinical variables is crucial to prevent the progression of DR. The result of this study show that the interaction of age by gender had a significant nonlinear association with being DR, i.e., the log odds of diabetic retinopathy have a significant nonlinear relationship with age of female patients. However, sex and age based stratified analysis showed that the incidence rate of sight-threatening diabetic retinopathy had a decreasing trend for women as compared to men [30]. Despite this, several studies reported the marginal effect of age and gender on being diabetic retinopathy. For example, Hussain et al.[15] show that the risk of diabetic retinopathy was higher for a male patient and a patient with age $\geq$ 49 years. Tan et al. [18] showed that for a patient with mild DR, older age was associated with lower retinal capillary density, while the result in [31] reveal that sex is not a risk factor for being diabetic retinopathy, but they identified age as a strong predictor of diabetic retinopathy.

In the current study, the duration of diabetes since a patient confirmed type II diabetes was appeared as one of the risk factors of diabetic retinopathy. We identified a nonlinear relationship between duration of diabetes and the log odds of diabetic retinopathy. This result agrees with previous studies [4, 12, 14, 15, 17]. However, these studies used a generalized linear model which can only identify a linear association between duration of diabetes and linear predictor rather than using a

data-driven relationship like a semiparametric model. Furthermore, the interaction between total cholesterol level and gender had a significant nonlinear association with the log odds of diabetic retinopathy. Though, Hanai et al. [32] investigated the progression of diabetic kidney disease and found that those lipid profile parameters are correlated with gender as a predictor of kidney disease progression. Further, Kaewput et al. [33], conducted a nationwide cross-sectional study in Thailand showing that DR had a significant association with renal function. Therefore, these two studies indirectly revealed that the interaction between the lipid profile of a patient and gender had a significant effect on being diabetic retinopathy.

The other interesting finding of our study was the significant nonlinear relationship between interaction $HbA1c \times FBG$ and the log odds of diabetic retinopathy. Despite the nonlinear relationship, a semiparametric model based on the tensor product of HbA1c and FBG suggested that the combination of a high level of HbA1c and a high level of FBG resulted in a higher risk of being diabetic retinopathy. Our study finding agrees with a study that used 10-year follow-up data [20]. Their finding suggested that patients with diabetic retinopathy at the baseline had a high level of FBG and a high level of HbA1c. However, our finding contradicts some of the previous studies. For example, Gimeno-Orna et al. [19] conducted a cohort study with a mean follow-up period of 5.2 years to examine whether FBG variability determines the onset of DR irrespective of HbA1c. Their finding from univariate logistic regression analysis showed that the interaction of mean HbA1c and FBG variability was not a significant risk factor of DR, however, they concluded that FBG variability can determine the onset of diabetic retinopathy. However, several studies showed the marginal effects of FBG and HbA1c on diabetic retinopathy [12, 14, 18].

## Conclusion

This study identified the possible risk factors of diabetic retinopathy based on data obtained from BLH using a semiparametric model. The results from this study indicate that clinical variables related to patient characteristics were strong predictors of diabetic retinopathy. Therefore, the results of a semiparametric analysis reveal evidence that being hypertensive, insulin treatment, 3 and 6-months clinical visits were strong predictive factors of diabetic retinopathy. Moreover, duration of diabetes, interaction of age by gender, and cholesterol level by gender had significant nonlinear relationships with diabetic retinopathy. Additionally, the nonlinear relationship between the interaction $HbA1c \times FBG$ and the linear predictor suggested that the risk of diabetic retinopathy was higher when the value of both HbA1c and FBG high.

Based on the findings we recommend that health care professionals should give more attention to the possible effect of clinical variables which can lead a type II diabetic patient to diabetic retinopathy. Since our study was based on one hospital, we recommend that a similar study should be conducted across the country to get more information to improve intervention strategies.

**Acknowledgements**
We thank Shibru T, Aga & Boka for providing the data which was originally collected from the Diabetic clinic of Black Lion Hospital, Addis Ababa, Ethiopia for another study [17]. We also thank the National Research Foundation (NRF) of South Africa for the partial scholarship and the University of South Africa (UNISA) for the M & D bursary to top-up the NRF partial scholarship of BEY

**Funding**
There was no funding to conduct this research.

**Abbreviations**
AIC: Akaike information criteria; BLH: Black lion hospital; CL:Cholesterol level; DD:Duration of diabetes; DM: Diabetes Mellitus; DR: Diabetic retinopathy; FBG: Fasting blood Glucose;FPG:fasting plasma glucose; FCV: Frequency of clinical visits; GCV:generalized cross validation; HbA1c: Hemoglobin A1c; IDF: International diabetes federation; IT: Insulin treatment; NPDR: Non-proliferative diabetic retinopathy; PDR: proliferative diabetic retinopathy; PIRLS: penalized iterative reweighted least squares; REML: Restricted maximum likelihood; VIF: variance inflation factor; UBRE: Unbiased risk estimator.

**Availability of data and materials**
The data sets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Ethics approval and consent to participate**
Ethical clearance and approval was obtained from the Institutional Review Board (IRB) of the College of Health Sciences of Addis Ababa University by the data providers [17]. After explaining the purpose and possible benefit of the study, oral and written informed consent was obtained from each patient before starting the procedure. All methods were carried out in accordance with relevant guidelines and regulations/Declaration of Helsinki.

**Competing interests**
The authors declare that they have no competing interests.

**Consent for publication**
Not applicable

**Authors' contributions**
BEY reviewed literature, performed the statistical analyses and drafted the manuscript. LKD supervised and reviewed the findings of data analyses and compilation of the manuscript.

**Author details**
[1]Department of Statistics, University of South Africa c/o Christiaan de Wet Road & Pioneer Avenue, Florida 1709, Johannesburg, South Africa. [2]Department of Statistics, University of South Africa c/o Christiaan de Wet Road & Pioneer Avenue, Private Bag X6, Florida 1710, Johannesburg, South Africa.

**References**
1. Obirikorang, Y., Obirikorang, C., Anto, E.O., Acheampong, E., Batu, E.N., Stella, A.D., Constance, O., Brenya, P.K.: Knowledge of complications of diabetes mellitus among patients visiting the diabetes clinic at sampa government hospital, ghana: a descriptive study. BMC public health **16**(1), 1–8 (2016)
2. Atlas IDF: International Diabetes Federation 9th Edition. (2019). Atlas IDF. http://www.diabetesatlas.org
3. Alam, S., Hasan, M., Neaz, S., Hussain, N., Hossain, M., Rahman, T., *et al.*: Diabetes mellitus: insights from epidemiology, biochemistry, risk factors, diagnosis, complications and comprehensive management. Diabetology **2**(2), 36–50 (2021)
4. Amer, J., Suboh, R., Abualrob, M., Shaheen, A., Abu Shanab, A.: Risk factors associated with diabetic retinopathy: A cross-sectional study within palestinian patients in northern west bank. Front. Clin. Diabetes Healthc. 2: 736715. doi: 10.3389/fcdhc (2021)
5. Tilahun, M., Gobena, T., Dereje, D., Welde, M., Yideg, G.: Prevalence of diabetic retinopathy and its associated factors among diabetic patients at debre markos referral hospital, northwest ethiopia, 2019: Hospital-based cross-sectional study. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy **13**, 2179–2187 (2020)
6. Boles, S.F., Center, A.A.E.: Diabetic retinopathy: What you should know. US Department of health and Human services, National Eye Institute National Institutes of Health, NIH Publication No: 03-2171 (2020)
7. Mwangi, N., Gachago, M., Gichangi, M., Gichuhi, S., Githeko, K., Jalango, A., Karimurio, J., Kibachio, J., Muthami, L., Ngugi, N., *et al.*: Adapting clinical practice guidelines for diabetic retinopathy in kenya: process and outputs. Implementation science **13**(1), 1–9 (2018)
8. Sabanayagam, C., Banu, R., Chee, M.L., Lee, R., Wang, Y.X., Tan, G., Jonas, J.B., Lamoureux, E.L., Cheng, C.-Y., Klein, B.E., *et al.*: Incidence and progression of diabetic retinopathy: a systematic review. The lancet Diabetes & endocrinology **7**(2), 140–149 (2019)
9. Flaxman, S.R., Bourne, R.R., Resnikoff, S., Ackland, P., Braithwaite, T., Cicinelli, M.V., Das, A., Jonas, J.B., Keeffe, J., Kempen, J.H., *et al.*: Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. The Lancet Global Health **5**(12), 1221–1234 (2017)
10. Abebe, N., Kebede, T., Addise, D.: Diabetes in ethiopia 2000–2016 prevalence and related acute and chronic complications; a systematic review. Afr J Diabetes Med **25**(2), 7–12 (2017)
11. Olafsdottir, E., Andersson, D.K., Dedorsson, I., Stefánsson, E.: The prevalence of retinopathy in subjects with and without type 2 diabetes mellitus. Acta Ophthalmologica **92**(2), 133–137 (2014)

12. Ferm, M.L., DeSalvo, D.J., Prichett, L.M., Sickler, J.K., Wolf, R.M., Channa, R.: Clinical and demographic factors associated with diabetic retinopathy among young patients with diabetes. JAMA network open **4**(9), 2126126–2126126 (2021)
13. Zhao, C., Wang, W., Xu, D., Li, H., Li, M., Wang, F.: Insulin and risk of diabetic retinopathy in patients with type 2 diabetes mellitus: data from a meta-analysis of seven cohort studies. Diagnostic pathology **9**(1), 1–7 (2014)
14. Tsegaw, A., Alemu, S., Dessie, A., Patterson, C.C., Parry, E.H., Phillips, D.I., Trimble, E.R.: Diabetic retinopathy in type 2 diabetes mellitus patients attending the diabetic clinic of the university of gondar hospital, northwest ethiopia. Journal of Ophthalmology **2021** (2021)
15. Hussain, S., Qamar, M.R., Iqbal, M.A., Ahmad, A., Ullah, E.: Risk factors of retinopathy in type 2 diabetes mellitus at a tertiary care hospital, bahawalpur pakistan. Pakistan journal of medical sciences **29**(2), 536 (2013)
16. Atkin, S.L., Butler, A.E., Hunt, S.C., Kilpatrick, E.S.: The retinopathy-derived hba1c threshold of 6.5% for type 2 diabetes also captures the risk of diabetic nephropathy in nhanes. Diabetes, Obesity and Metabolism **23**(9), 2109–2115 (2021)
17. Shibru, T., Aga, F., Boka, A.: Prevalence of diabetic retinopathy and associated factors among type 2 diabetes patients at tikur. Journal of Diabetes & Metabolism (2019)
18. Tan, F., Chen, Q., Zhuang, X., Wu, C., Qian, Y., Wang, Y., Wang, J., Lu, F., Shen, M., Li, Y.: Associated risk factors in the early stage of diabetic retinopathy. Eye and Vision **6**(1), 1–10 (2019)
19. Gimeno-Orna, J.A., Castro-Alonso, F.J., Boned-Juliani, B., Lou-Arnal, L.M.: Fasting plasma glucose variability as a risk factor of retinopathy in type 2 diabetic patients. Journal of diabetes and its complications **17**(2), 78–81 (2003)
20. Massin, P., Lange, C., Tichet, J., Erginay, A., Cailleau, M., Eschwège, E., Balkau, B., an Epidemiological Study on the Insulin Resistance Syndrome) Study Group, D.D.F., *et al.*: Hemoglobin a1c and fasting plasma glucose levels as predictors of retinopathy at 10 years: the french desir study. Archives of ophthalmology **129**(2), 188–195 (2011)
21. Coull, B.A., Ruppert, D., Wand, M.: Simple incorporation of interactions into additive models. Biometrics **57**(2), 539–545 (2001)
22. Chen, H.-F., Ho, C.-A., Li, C.-Y.: Age and sex may significantly interact with diabetes on the risks of lower-extremity amputation and peripheral revascularization procedures: evidence from a cohort of a half-million diabetic patients. Diabetes care **29**(11), 2409–2414 (2006)
23. Ghazanfari, Z., Haghdoost, A.A., Alizadeh, S.M., Atapour, J., Zolala, F.: A comparison of hba1c and fasting blood sugar tests in general population. International journal of preventive medicine **1**(3), 187–194 (2010)
24. Davila, J.R., Sengupta, S.S., Niziol, L.M., Sindal, M.D., Besirli, C.G., Upadhyaya, S., Woodward, M.A., Venkatesh, R., Robin, A.L., Grubbs Jr, J., *et al.*: Predictors of photographic quality with a handheld nonmydriatic fundus camera used for screening of vision-threatening diabetic retinopathy. Ophthalmologica **238**(1-2), 89–99 (2017)
25. Wood, N.S.: Generalized Additive Models. University of Bristol, UK, Boca Raton (2017)
26. Soley-Bori, M.: Dealing with missing data: Key assumptions and methods for applied analysis. Boston University **23**, 20 (2013)
27. Allison, P.D.: Multiple imputation for missing data: A cautionary tale. Sociological methods & research **28**(3), 301–309 (2000)
28. Meng, X., Zhang, Y., Kong, Q., Lv, Y., Hu, H., Chen, T., Tang, Z.: Interaction analysis of systolic blood pressure and glycosylated hemoglobin in diabetic retinopathy: A chinese sample. Traditional Medicine and Modern Medicine **2**(03), 119–125 (2019)
29. Goyal, M., Kamboj, P., Behgal, J., Rathee, S., Lather, T.: Risk factors of diabetic retinopathy in patients with type 2 diabetes mellitus. Diabetes Manage **7**(6), 408–411 (2017)
30. Lin, J.-C., Shau, W.-Y., Lai, M.-S.: Sex-and age-specific prevalence and incidence rates of sight-threatening diabetic retinopathy in taiwan. JAMA ophthalmology **132**(8), 922–928 (2014)
31. Magliah, S.F., Bardisi, W., Al Attah, M., Khorsheed, M.M.: The prevalence and risk factors of diabetic retinopathy in selected primary care centers during the 3-year screening intervals. Journal of family medicine and primary care **7**(5), 975–981 (2018)
32. Hanai, K., Babazono, T., Yoshida, N., Nyumura, I., Toya, K., Hayashi, T., Bouchi, R., Tanaka, N., Ishii, A., Iwamoto, Y.: Gender differences in the association between hdl cholesterol and the progression of diabetic kidney disease in type 2 diabetic patients. Nephrology Dialysis Transplantation **27**(3), 1070–1075 (2012)
33. Kaewput, W., Thongprayoon, C., Rangsin, R., Ruangkanchanasetr, P., Mao, M.A., Cheungpasitporn, W.: Associations of renal function with diabetic retinopathy and visual impairment in type 2 diabetes: A multicenter nationwide cross-sectional study. World journal of nephrology **8**(2), 33–43 (2019)


## Figures

**Figure 1** A loess fit to the log odds of diabetic retinopathy and continuous predictors separately

**Figure 2** Goodness of fit of the imputed observation relative to the observed data for HbA1c and total cholesterol level

## Tables

**Figure 3** A Boxplot for cholesterol level by gender

**Figure 4** Estimate of smooths in a semiparametric model.

**Figure 5** Estimated effects for the tensor product smooth interaction $HbA1c \times FBG$ in a semiparametric model.

**Table 1** Summary statistics of diabetic retinopathy status of type II diabetic patients vs socio-demographic and clinical variables

| | | Diabetic Retinophaty(DR) | | |
|---|---|---|---|---|
| Characteristics | Levels | Patient with DR,N(%) | Patient without DR, N(%) | Total |
| Gender | M | 47 (24.6) | 30 (15.9) | 77 (40.5) |
| | F | 51 (26.7) | 63 (32.8) | 114 (59.5) |
| Hypertension | No | 22 (11.5) | 63 (32.9) | 85 (44.4) |
| | Yes | 76 (39.8) | 30 (15.8) | 106 (54,8) |
| IT | No | 58 (30.4) | 79 (41.4) | 137 (71.7) |
| | Yes | 40 (20.9) | 14 (7.3) | 54 (28.3) |
| FCV | every 1 month | 14 (7.4) | 29 (15.2) | 43 (22.6) |
| | every 3 month | 38 (20.0) | 34 (17.9) | 72 (37.9) |
| | every 6 month | 46 (24.2) | 29 (15.3) | 75 (39.5) |
| | | mean(sd) | mean(sd) | $n = 191$ |
| Age | | 58.5(10.1) | 56 (10.4) | |
| DD | | 15.2(10.9) | 9.2 (8.8) | |
| HbA1c | | 9.3 (2.8) | 7.09 (2.3) | |
| CL | | 187.6(56.2) | 179.3(41.5) | |
| FBG | | 170.9 (66.0) | 160.7 (60.6) | |

**Table 2** Test of nonlinearity for the continuous covariates

| nonlinearity test | Models | Resid.Df | resid.Dev | Df | Deviance | $pr(> chi)$ |
|---|---|---|---|---|---|---|
| $HbA1c \times FBG$ | $M_2$ | 158.20 | 113.21 | | | |
| $M_2$vs$M_1$ | $M_1$ | 153.43 | 99.62 | 4.77 | 13.59 | $0.0157^*$ |
| Age by gender | $M_3$ | 158.99 | 111.61 | | | |
| $M_3$vs$M_1$ | $M_1$ | 153.43 | 99.62 | 5.53 | 11.98 | $0.0461^*$ |
| CL by gender | $M_4$ | 168.56 | 136.83 | | | |
| $M_4$vs$M_1$ | $M_1$ | 153.43 | 99.62 | 15.13 | 37.20 | $0.0012^*$ |
| DD | $M_5$ | 158.41 | 112.64 | | | |
| $M_5$vs$M_1$ | $M_1$ | 153.43 | 99.62 | 4.97 | 13.02 | $0.0228^*$ |

**Table 3** Model comparison using AIC

| Models | $M_1$ | $M_2$ | $M_3$ | $M_4$ | $M_5$ |
|---|---|---|---|---|---|
| AIC | 163.64 | 169.44 | 166.52 | 176.32 | 168.61 |

**Table 4** Semi parametric estimate of socio-demographic and clinical variables that have a significant effect on DR

| predictors | levels | df | $\hat{\beta}$ (se) | $p - value$ | AOR | 95%CI |
|---|---|---|---|---|---|---|
| Intercept | | 1 | -4.30 (0.83) | $< 0.0001$ | | |
| Gender | Male | 1 | 1.27 (0.58) | $0.0280^*$ | 3.5 | [1.14, 11.09] |
| Hypertension | yes | 1 | 3.66 (0.70) | $< 0.0001$ | 38.9 | [9.85, 153.23] |
| IT | yes | 1 | 1.84 (0.63) | $0.0040^*$ | 6.2 | [1.81, 13.84] |
| FCV | every 3 month | 1 | 2.17 (0.73) | $0.0020^*$ | 8.7 | [2.13, 35.9] |
| | every 6 month | 1 | 1.91 (0.72) | $0.0080^*$ | 6.7 | [1.63, 27.41] |
| nonlinear Terms | | | | | | |
| $f_{gender}(age)$ | Female | 3.2 | | $0.0357^*$ | | |
| $f_{gender}(age)$ | Male | 1.0 | | 0.2386 | | |
| f(DD) | | 2.7 | | $0.0059^*$ | | |
| $f_{gender}(CL)$ | Female | 3.6 | | $0.0166^*$ | | |
| $f_{gender}(CL)$ | Male | 3.7 | | 0.1321 | | |
| $f(HbA1c)$ | | 2.3 | | $0.0020^*$ | | |
| $f(FBG)$ | | 1.0 | | 0.2784 | | |
| f(HbA1c, FBG) | | 3.7 | | $0.0500^*$ | | |

# Figures

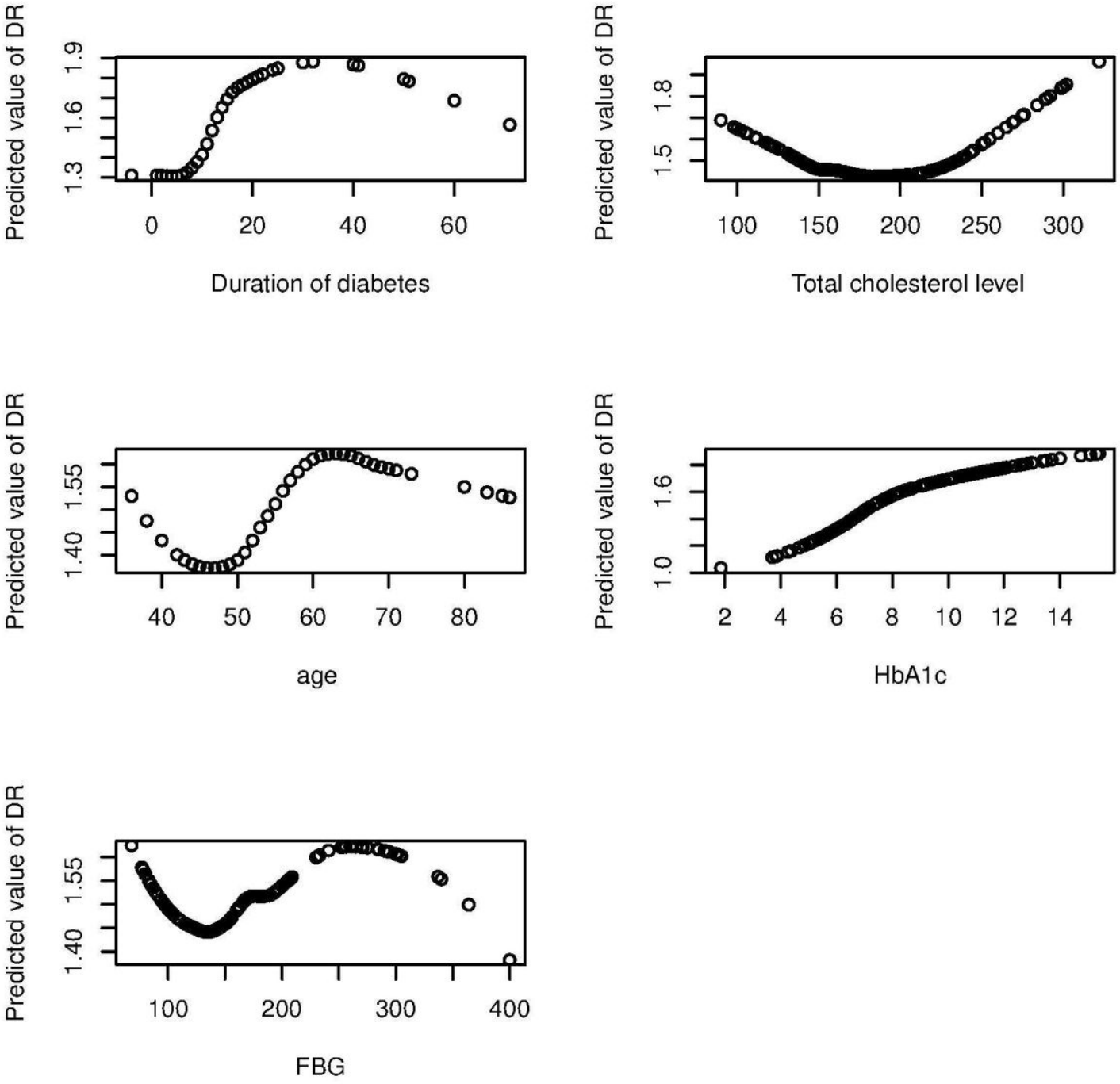
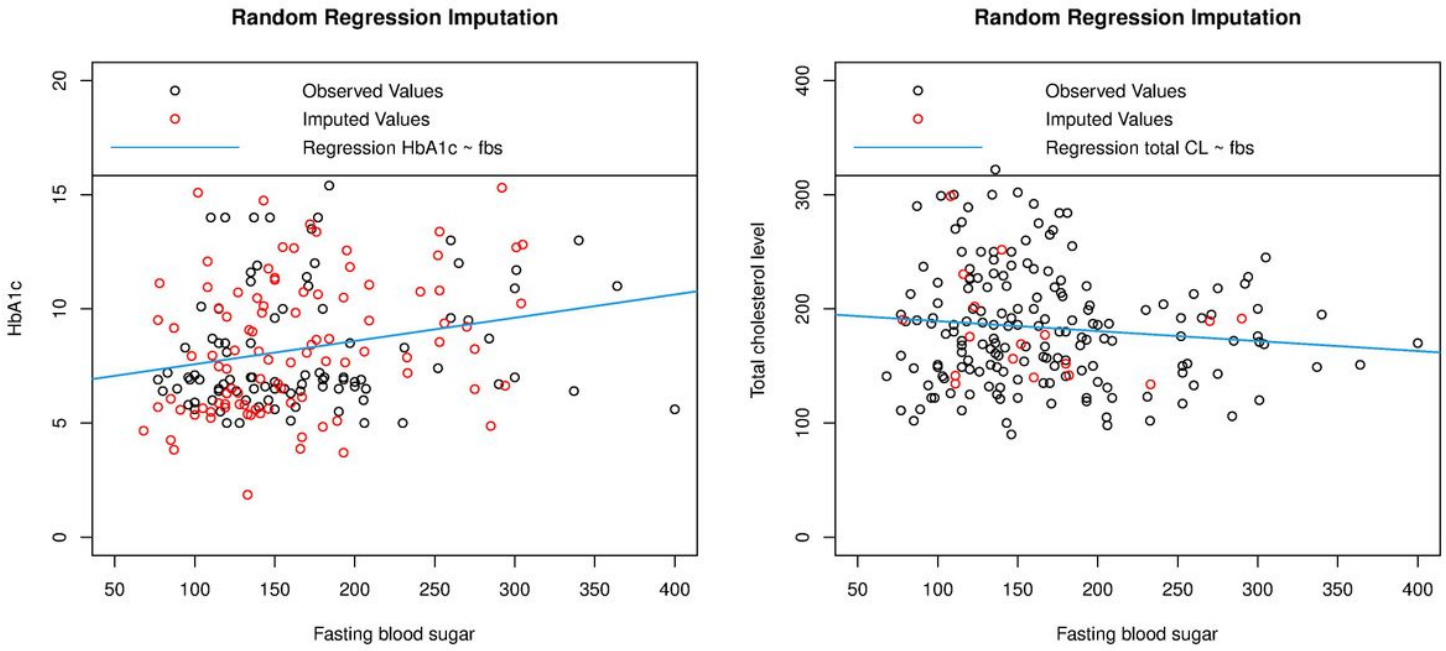



**Figure 1**

A loess fit to the log odds of diabetic retinopathy and continuous predictors separately




## Figure 2

Goodness of fit of the imputed observation relative to the observed data for HbA1c and total cholesterol level

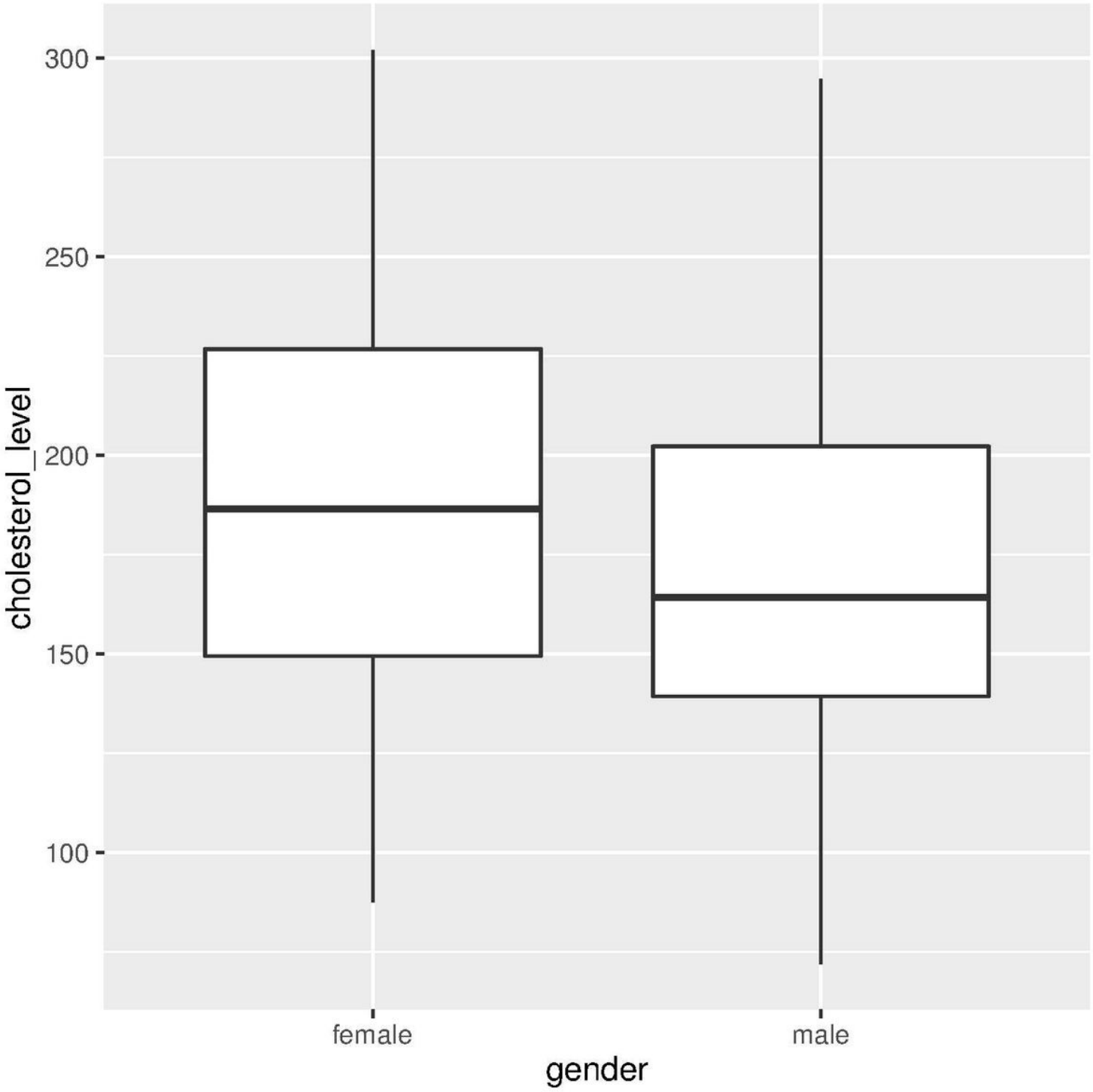



**Figure 3**

A Boxplot for cholesterol level by gender

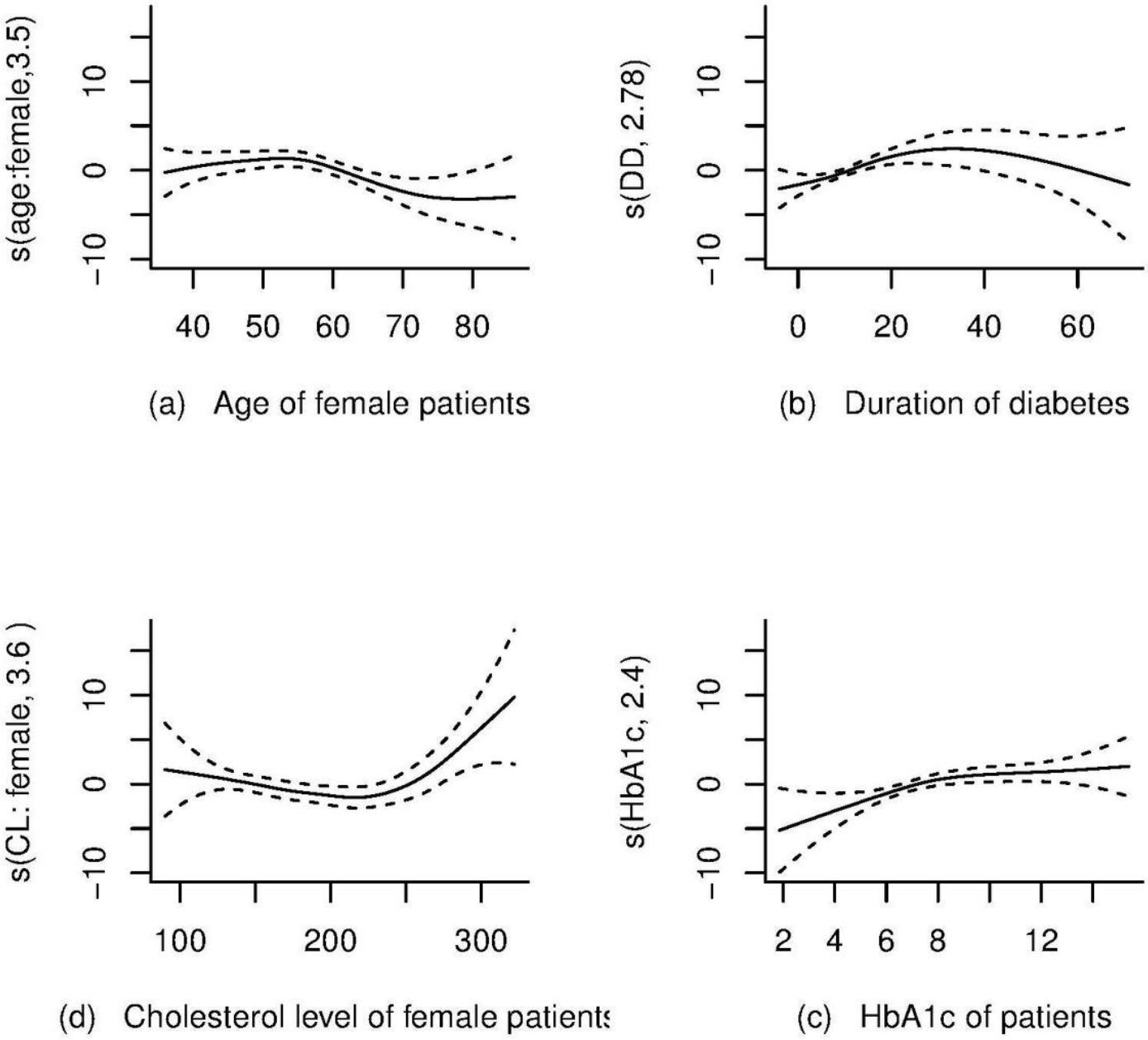
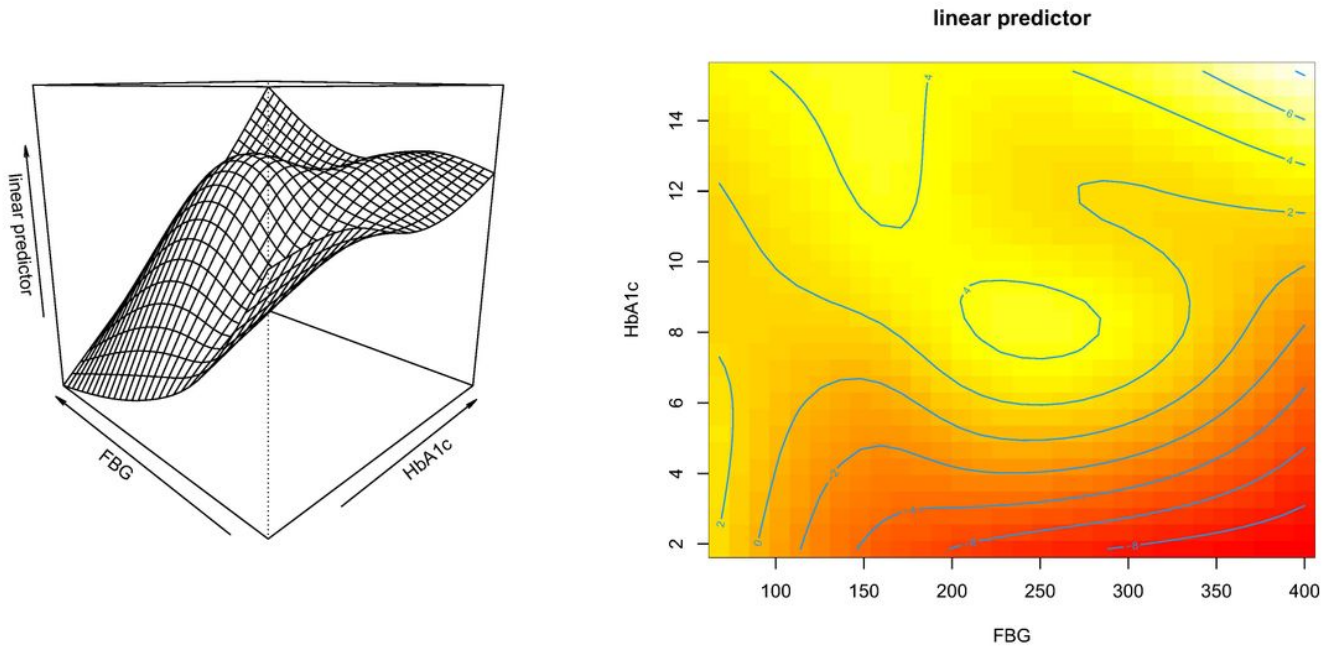



**Figure 4**

Estimate of smooths in a semiparametric model.




## Figure 5

Estimated effects for the tensor product smooth interaction *HbA*1c × *FBG* in a semiparametric model.