Diabetes Disease Progression Determinants

Rabindra Nath Das*

Department of Statistics, The University of Burdwan, Burdwan, West Bengal, India

Abstract

The incidence and prevalence of diabetes mellitus (DM) have significantly increased worldwide. Consequently, the prevalence of coronary heart disease (CHD) and chronic kidney disease (CKD) have been increased significantly, as they are highly associated with DM. The present report derives the determinants of the DM disease progression, based on ten baseline covariates (such as sex, age, average blood pressure, body mass index and six blood serum measurements) on 442 diabetes patients. It also identifies the effects of the factors or covariates on DM disease progression. The current report has examined the objectives based on the derived probabilistic models of a quantitative measure of DM disease progression one year after the baseline. Moreover, these models can be used for predicting the disease progression for future patients.

Keywords: Average Blood Pressure; Blood Glucose Random; Blood Serum: Body Mass Index; Diabetes Mellitus; Gamma Model; Non-Constant Variance

Introduction

In the present world, the DM is recognized as a worldwide epidemic. The DM is considered as the risk factor for many diseases such as CHD & CKD [1-7]. It is well-known that the DM is considered as a silent disease, and the principal causes of death of more than 70% type 2 diabetes patients are chronic kidney disease and cardiovascular diseases [2-15]. Recently some articles [2-8] have clearly described the relationships between the CHD and DM, and the association between the DM and the chronic kidney disease. The determinants of Type 2 diabetes mellitus (T2DM) for two different group of subjects are described in [1,5,7,9]. The world prevalence of diabetes among adults (aged 20-79 years) will be 6.4%, affecting 285 million adults, in 2010, and will increase to 7.7%, and 439 million adults by 2030 [16]. T2DM is a strong risk factor for CHD, CKD and many other diseases [3,7]. Some diabetes and coronary heart disease patients have associated with obesity and cardiac risk factors such as hypertension and hypercholesterolemia [3,4,5,12,15-17]. Now it is very important how to identify the determinants of T2DM. Equivalently, it is important to identify the risk factors of the DM disease progression.

The present report aims to identify the risk factors of the DM disease progression. Now we have the following hypotheses (or queries): What are the risk factors of the DM disease progression? What are the effects of the risk factors on DM disease progression? How do we identify the risk factors of DM?

Answer of these queries are derived from a data set of 9 variables and 1 factor on 442 diabetes patients, with all non-missing information. These 9 variables are the age, body mass index [BMI], average blood pressure, and six blood serum measurements such as Total cholesterol (TC), Low density lipoproteins (LDL), High density lipoproteins (HDL), Triglyceride (TG), Serum concentration of lamorigine (LTG), Glucose (GLU). Only one factor is sex (male=0, female=1). In addition the data set contains a quantitative measure of DM disease progression one year after the baseline. Source of the data set is given in [18].

In Table 1, a part of the data set is given, as an example for illustrating the theory of “least angel regression” [18]. In the article [18], the quantitative measure of DM disease progression one year after the baseline is considered as the response variable, and the remaining ten variables are considered as the explanatory variables. The article [18] aimed to develop a model that would derive accurate baseline predictions of the dependent variable for future patients. In addition, it also aimed to identify the factors which were significant in disease progression. In practice, the article [18] just used the data set to illustrate the theoretical steps, but the final model has not been derived.

Based on our knowledge, there is no study about the determinants of the DM disease progression. Note that the data set is a physiological positive data which generally belong to exponential family distribution, and its variance may be non-constant. In this case, the variance may have relationship with the mean. The non-constant variance problem of the dependent variable in linear regression is a departure from the standard least squares assumptions. It generally occurs for a non-normal dependent variable distribution. Under this situation, an appropriate transformation of the dependent variable is used to stabilize the variance. However, in practice, the variance is not stabilized always [19, p. 36-38]. For positive data analysis of a non-normal dependent variable with non-constant variance, it is more appropriate to apply joint generalized linear

*Corresponding author: Rabindra Nath Das, Department of Statistics, The University of Burdwan, Burdwan, West Bengal, India, Email: rabin.bwn@gmail.com

Sub Date: November 28, 2016, Acc Date: December 12, 2016, Pub Date: December 12, 2016.

Citation: Rabindra Nath Das (2016) Diabetes Disease Progression Determinants. BAOJ Diabet 2: 015.

Copyright: © 2016 Rabindra Nath Das. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
models (JGLMs) (modeling of mean and variance simultaneously) [20,21]. Then, joint generalized linear gamma and Log-normal models are used [20-24]. Below we shortly describe the joint Log-normal and gamma models in the methodology section.

**Methodology: Joint Generalized Gamma and Log-Normal Models**

The response that a quantitative measure of DM disease progression [in 18] is a positive, continuous, heterogeneous and non-normal distributed random variable. Based on our knowledge, its analysis has not been clearly done in the literature [18]. It has been considered as an example for illustrating 'least angle regression' [in 18]. In order to confirm its analysis, we have examined the analysis of the quantitative measure of DM disease progression using both the Log-normal and gamma model analysis, as it is well known that a positive continuous random variable may be well modeled by these two models [21,25]. For ready reference, both these two models are shortly described as follows.

For the positive random variable \( Y_i \)’s with variance \( \sigma_i^2 \), if \( E(\gamma) = \mu_i \) and \( Var(\gamma) = \sigma_i^2 \mu_i^2 \)

the transformation \( Z_i = \log(Y_i) \) is generally used to stabilize the variance \( Var(Z_i) = \sigma_i^2 \).

Note that the variance is not always stabilized under the log transformation [19, p. 36]. In that case, if we need a parsimonious model, a better transformation is applied. Due to transformation, many model assumptions may be failed. Under these situations, a joint generalized linear models (JGLMs) for the mean and dispersion may be used [26].

Generally, the log transformation \( Z_i = \log(Y) \) is used for the positive response \( Y_i \). For the response \( Z_i \), with log-normal distribution, a joint modelling of the mean and dispersion is given by

\[
E(\mu_i) = \mu_i \quad \text{and} \quad Var(\mu_i) = \sigma_i^2 \mu_i^2 \]

\[
\mu_i = x_i^t \beta \quad \text{and} \quad \log(\sigma_i^2) = g_i^t \gamma ,
\]

where \( x_i \) and \( g_i \) are the row vectors for the regression coefficients \( \beta \) and \( \gamma \) in the mean and dispersion model, respectively.

The JGLMs is used for analyzing of positive data \( y_i \), if

\[
E(\gamma_i) = \mu_i \quad \text{and} \quad Var(\gamma_i) = \sigma_i^2 V(\mu_i) ,
\]

where \( V(\gamma) \) is the variance function and \( \sigma_i^2 \)’s are the dispersion parameters. In GLMs, the variance has two parts. A part is \( V(\mu_i) \) which is dependent on the mean changes, and the other is \( \sigma_i^2 \) which is independent of mean adjustment. Note that the variance function identifies the GLM family distribution. For example, the GLM distribution is gamma if \( V(\mu_i) = \mu_i^2 \), Poisson if \( V(\mu_i) = \mu_i \), normal if \( V(\mu_i) = 1 \), etc. Therefore, the joint mean and the dispersion models are

\[
\eta_i = g(\mu_i) x_i^t \beta \quad \text{and} \quad \epsilon_i = h(\sigma_i^2) = w_i^t \gamma ,
\]

where \( g(\cdot) \) and \( h(\cdot) \) are GLM link functions between the mean or the variance with the linear predictors, respectively, for the mean and the dispersion, and \( x_i^t \), \( v_i^t \) are the row vectors respectively, for the regression models of mean and dispersion. Generally, the maximum likelihood (ML) method and the restricted ML (REML) are applied respectively, for estimating the mean and dispersion parameters [21, 26]. Using iterative method, unknown mean and dispersion parameters are estimated. A detailed discussion is given in [26].

**Analysis and Interpretations of Dm Disease Progression Response Data**

**Analysis**

To obtain the better and confirmed model for the present data, we have analyzed the response by using both the joint Log-normal & gamma models [20, 21]. Here the quantitative measure of DM disease progression has been modeled based on the remaining ten explanatory variables using both the Log-normal and gamma models [20], and the derived results are displayed in Table 1. The final models have been selected based on the smallest Akaike information criterion (AIC) value in each class. Note that the AIC selects a model which minimizes the predicted additive errors and squared error loss [27, p. 203-204]. It can be examined (based on the data set given in the above mentioned site) that the selected gamma fit (Table 1) (AIC= 4715.708) gives better results than the Log-normal fit (AIC=4734). Therefore, only the diagnostic plots of gamma fitted joint models [Table 1] have been examined. Some insignificant factors are included in the variance model for better fitting [27]. In epidemiology, the included insignificant factors in the model are known as confounder. Note that some insignificant factors are very important in the model fitting, and they should be included in the model [27].

In order to examine the proper fitting of the gamma fitted model (Table 1), two diagnostic plots, namely, the absolute residuals plot and the normal probability plot are shown in Figure 1. In Figure 1(a), the absolute residual values are plotted with respect to fitted values. It is almost a flat diagram with the running means, indicating that the variance is constant for the fitted model. Figure 1(b) displays the normal probability plot of the gamma fitted mean model (Table 1), which does not show any lack of fit for outliers, or variables as there is not any gap in the figure.

**Results and Interpretations**

The results and interpretations of Table 1 are described as follows:

- The mean quantitative measure of T2DM disease progression (QMDMDP) is negatively associated with the sex (0=male, 1=female) (P=0.0003). It indicates that the QMDMDP is higher for male than the female patients.
- The mean QMDMDP is positively associated with the body mass index (BMI) (P<0.0001), indicating that the QMDMDP is higher of the diabetes patients having higher BMI.
- The mean QMDMDP is positively associated with the average blood pressure (ABP) (P<0.0001), indicating that the QMDMDP...
Citation: Rabindra Nath Das (2016) Diabetes Disease Progression Determinants. BAOJ Diabet 2: 015.

The mean QMDMDP is negatively associated with the total cholesterol (TC) \((P<0.0001)\), indicating that the QMDMDP increases as the TC decreases. Note that TC has the same association with both the mean and variance of QMDMDP.

The QMDMDP variance is positively associated with the LDL \((P=0.0041)\), indicating that the QMDMDP variance increases as the LDL increases. Also, LDL has the same association with both the mean and variance of QMDMDP.

The QMDMDP variance is negatively associated with the triglyceride (TG) \((P=0.0705)\), indicating that the QMDMDP variance increases as the TG decreases.

The QMDMDP variance is negatively partially associated with the age \((P=0.1500)\), indicating that the QMDMDP variance increases for the younger diabetes patients. Note that the minimum age

### Table 1: Joint Log-normal & Gamma fitted results for determinants of diabetes mellitus disease progression

<table>
<thead>
<tr>
<th>Model</th>
<th>Covariate</th>
<th>Gamma fitted results</th>
<th>Log-normal fitted results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Constant</td>
<td>1.7651, 0.1699, 10.388</td>
<td>1.4619, 0.1756, 8.324</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>-0.1396, 0.0380, -3.669</td>
<td>-0.1198, 0.0392, -3.053</td>
</tr>
<tr>
<td></td>
<td>Body mass index (BMI)</td>
<td>0.0325, 0.0046, 6.995</td>
<td>0.0351, 0.0048, 7.263</td>
</tr>
<tr>
<td></td>
<td>Average blood pressure (ABP)</td>
<td>0.0081, 0.0014, 5.604</td>
<td>0.0081, 0.0015, 5.427</td>
</tr>
<tr>
<td></td>
<td>Total cholesterol (TC)</td>
<td>-0.0085, 0.0014, -5.883</td>
<td>-0.0078, 0.0015, -5.248</td>
</tr>
<tr>
<td></td>
<td>Low density lipoproteins (LDL)</td>
<td>0.0077, 0.0014, 5.299</td>
<td>0.0071, 0.0015, 4.732</td>
</tr>
<tr>
<td></td>
<td>Serum concentration of lamorigine (LTG)</td>
<td>0.5355, 0.0502, 10.657</td>
<td>0.5502, 0.0518, 10.616</td>
</tr>
<tr>
<td>Variance</td>
<td>Constant</td>
<td>0.7886, 0.6967, 1.132</td>
<td>1.1348, 0.6981, 1.625</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>-0.0083, 0.0058, 1.142</td>
<td>-0.0092, 0.0058, 1.596</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>-0.1903, 0.1583, 1.142</td>
<td>-0.2790, 0.1583, 1.763</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>-0.0299, 0.0200, 1.142</td>
<td>-0.0275, 0.0203, 1.135</td>
</tr>
<tr>
<td></td>
<td>TC</td>
<td>-0.0135, 0.0046, 2.912</td>
<td>-0.0149, 0.0047, 3.177</td>
</tr>
<tr>
<td></td>
<td>LDL</td>
<td>0.0164, 0.0057, 2.882</td>
<td>0.0168, 0.0058, 2.881</td>
</tr>
<tr>
<td></td>
<td>Triglyceride (TG)</td>
<td>-0.1537, 0.0848, -1.813</td>
<td>-0.1335, 0.0856, -1.559</td>
</tr>
</tbody>
</table>

Figure 1(a): For the fitted gamma models of DM disease progression (Table 1), the (a) absolute residuals plot with respect to fitted values, and the (b) normal probability plot of the mean model.
of the diabetes patient is 19 year, while the maximum age is 79 year.

- The QMDMDP variance is negatively partially associated with the sex (0=male, 1=female) \( (P=0.2300) \), indicating that the QMDMDP variance increases for the male diabetes patients than the female. Sex has the same association with both the mean and variance of QMDMDP.

- The QMDMDP variance is negatively partially associated with the BMI \( (P=0.1346) \), indicating that the QMDMDP variance increases for diabetes patients having lower BMI. Note that BMI has opposite association between the mean and variance of QMDMDP.

From Table 1, the gamma fitted mean and variance models of the quantitative measure of DM disease progression \( (Y) \), respectively, are

\[
\hat{Y} = \exp(1.7651 - 0.1396 \text{sex} + 0.0325 \text{BMI} + 0.0081 \text{ABP} - 0.0085 \text{TC} + 0.0077 \text{LDL} + 0.5355 \text{LTG}),
\]

and

\[
\hat{\sigma}^2 = \exp(0.7865 - 0.0083 \text{age} - 0.1903 \text{sex} - 0.0299 \text{BMI} - 0.0135 \text{TC} + 0.0164 \text{LDL} - 0.1537 \text{LTG}).
\]

Conclusions and Discussions

The present report has identified many significant factors (such as sex, body mass index, average blood pressure, total cholesterol, low density lipoproteins, and serum concentration of lamorigine) as the determinants of the quantitative measure of DM disease progression. Age is identified as a partially significant factor, which is a confounder in the derived variance model. Note that high density lipoproteins is insignificant in both the mean and the variance models. Here the final gamma fitted models have been selected based on AIC values in comparison of both the fitted models. Note that the standard errors of the mean and variance estimated parameters for the gamma fitted model are smaller than the respective estimates of the Log-normal fitted model. This also indicates that the gamma fitted model is better than the Log-normal fitted model. Also the diagnostic plots do not show any discrepancy of the gamma fitted models. Note that the interpretations of all the estimates from the two fitted model are identical, even though the estimates are little different. Then it may be concluded that the distribution of the DM disease progression is gamma. All the conclusions have been derived herein based on satisfying the above data analysis criteria. Thus, the research should have greater faith in the present reported results.

The present data set contains only a few covariates. It does not contain the separate systolic, diastolic blood pressure, heart rate, values of HbA1c, lifestyle characteristics, smoking status, and many other family history values. Also the quantitative measure of DM disease progression has not been clearly described. Data collection method has not been given in [18]. Future similar researchers are advised to collect as many information as possible. The above results may help the medical practitioners for better medical treatment.

Conflict of Interest

The author confirms that this article content has no conflict of interest.

Acknowledgement

The author is very much indebted to the referees who have provided valuable comments to improve this paper.

References


Citation: Rabindra Nath Das (2016) Diabetes Disease Progression Determinants. BAOJ Diabet 2: 015.


