Determinants of Diabetes Mellitus in the Pima Indian Mothers and Indian Medical Students

Rabindra Nath Das*

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

Abstract: Diabetes mellitus is a very common and serious disease in many American Indian tribes, Indians, and many other populations in the world. Several well-known risk factors such as parental diabetes, genetic markers, obesity, diet are considered as the main risk factors for diabetes mellitus, while the precise nature of the gene or genes remains unknown.

Objectives: The Pimas, Indians, and many other population in the world now suffer from the high prevalence rates of diabetes. Epidemiological research often seeks to identify a causal relationship between the risk factors and the disease. In the present article, two data sets from two different study groups (one for the Pima Indian mothers (768 subjects) and the other for the Indian medical students (64 subjects) are analyzed to determine the causal factors of diabetes mellitus. This article aims to identify the determinants of diabetes mellitus in the Pima Indian mothers and Indian young medical students.

Results: The causal factors for diabetes mellitus of the Pima Indian mothers and young Indian medical students are identified. Statistical *significant* causal factors, namely, triceps skin fold thickness (P-value < 0.01), serum insulin (P-value < 0.01), body mass index (or obesity) (P-value < 0.01), diabetes pedigree function (P-value = 0.06), age (P-value < 0.01) are identified as the determinants of diabetes mellitus in the Pima Indian mothers. In the young Indian medical students, age (P-value = 0.04), body mass index (P-value < 0.01), family history of diabetes mellitus (P-value < 0.01), sex (P-value < 0.01), low density lipoprotein (P-value = 0.01), total cholesterol (P-value = 0.11), serum triglyceride (P-value < 0.01), family blood pressure (P-value < 0.01), dietary habits like eating outside (P-value < 0.01) are identified as the determinants of diabetes mellitus. The effects of different causal factors on diabetes mellitus are explained based on probabilistic models.

Conclusions: Impacts of biochemical parameters, personal characteristics, family history, and dietary factors on human plasma glucose concentration are explained based on mathematical relationships. The results of the present analyses support many earlier research findings. However, these analyses also identify many additional casual factors that explain the mean and the variance of plasma glucose concentration, which earlier researches have not reported.

Keywords: Biochemical parameters, diabetes mellitus, gamma model, joint generalized linear model, non-constant variance.

1. INTRODUCTION

In recent years, the prevalence of diabetes has significantly increased over the world, and this has in turn impacted on the incidences associated with many complications, including heart failure. Diabetes mellitus (DM) is a chronic progressive disease, and is a well recognized risk factor for heart failure (HF). The disease DM is considered a world-wide epidemic and major chronic health problem. Diabetes is recognized as a silent disease, and in 70% of patients with type 2 diabetes, cardiovascular problem and coronary disease are the main cause of death [1-4].

The present article considers two study subject groups, namely non-diabetes and diabetes Type 2. A general concept of diabetes and its types is described as follows. Medically diabetes is known as *diabetes mellitus* which is a metabolism disorder. How the body uses and digests food for growth and energy is referred to metabolism. Most of our consumed foods are broken down into glucose which is a type of sugar in the blood. Glucose is the main source of food for our bodies (our cells). When food is digested it eventually enters our bloodstream in the form of glucose. Cells utilize the glucose for growth and energy. However, without the help of insulin, the glucose cannot enter our cells. Insulin, a hormone, is produced by Beta cells in the Islets of Langerhans, which are in the pancreas. After eating, the pancreas automatically release an adequate amount of insulin to transport the blood glucose into the cells, which results in lower blood sugar levels. If any one has diabetes, the glucose in the bloodstream does not enter the cells (at all or not enough), so glucose builds up until levels are too high, resulting in a condition called hyperglycemia. Consequently, excessive amounts of glucose accumulate in the blood. This blood glucose overload is eventually passed out of the body in urine. Even though the blood has plenty of insulin, the cells of a person with diabetes are not getting their crucial energy and growth requirements.

Address correspondence to this author at the Department of Statistics, University of Burdwan, Burdwan, West Bengal, India;
Tel: +91-9232638970; Fax: +91-(0342)-2530452;
E-mail: rabin.bwn@gmail.com

1876-5246/14

Hyperglycemia happens for one of two main reasons: (a) The body is producing no insulin which is known as "Diabetes Type 1" and (b) the cells do not respond correctly to the insulin which is known as "Diabetes Type 2". In Type 1 diabetes, the person's own body has destroyed the insulin-producing beta cells in the pancreas. When our own body destroys good stuff in our body it has what is called an *autoimmune disease*. Type 1 diabetes is known as an autoimmune disease. A person with diabetes Type 1 does not produce insulin. In the majority of cases this type of diabetes appears before the patient is 40 years old. That is why this type of diabetes is also known as *Juvenile Diabetes* or *Childhood Diabetes*. In order to stay alive, Type 1 diabetes patients have to take insulin regularly.

Diabetes Type 2 patients have one of following two problems, and sometimes both: (a) Not enough insulin is being produced and (b) the insulin is not working properly which is known as *insulin resistance*. The body produces insulin, but its insulin sensitivity is undermined and does not work as it should do, and glucose is not entering the body's cells properly. Consequently, blood sugar levels rise, and the cells are not getting their required nutrients for energy and growth. The problem of *insulin resistance* is with the cells as they are not responding to insulin like they used to. Medical experts are not sure what exactly is happening when cells stop responding well to insulin. A few simple explanations of why insulin resistance happens are given as follows**:** (a) Each time diabetes Type 2 patients cells are exposed to insulin and they build up a bit of resistance, (b) if a person is eating a lot he will be producing more insulin than somebody who doesn't, (c) if the exposure to insulin is high the cell will try to protect itself from intoxication, it will down-regulate its receptor activity and the number of receptors so that it does not have to be subjected to all that stimuli all the time, (d) if the cell's exposure to high insulin is frequent the insulin resistance will grow faster and (e) if the insulin is not doing its job properly the pancreas will put out more of it.

The insulin resistance will reach a point in which the amount of insulin produced by the pancreas is not enough to make up for the cells lower response. At this point the person will have to take additional insulin. Lack of physical activity, being overweight, and some genetic factors make it much more likely that the cells build up insulin resistance more quickly. It is important to remember that *insulin resistance is not the insulin not responding properly, it is the cells not responding properly to insulin.*

In this article, one group of subjects is taken from the longitudinal study of diabetes that has been conducted among the residents of the Gila River Indian Community of Arizona, USA, since 1965 [5]. The Pimas now suffer from high prevalence rates of Type 2 diabetes and obesity. The Pima Indian Community in the desert of central Arizona have participated in a comprehensive longitudinal study of the epidemiology of diabetes mellitus, for which they have been examined at intervals of two years since 1965. Several research articles provide much of the information on the prevalence, incidence, risk factors, and pathogenesis of diabetes in the Pima Indian population (Review articles - [6-8] and research articles [9-12]).

Diabetes in the Pimas is strongly familial, and probably of genetic origin [7, 13]. The incidence of diabetes in the Pima has increased during the last several decades providing further evidence for environmental genetic interaction. Longitudinal studies of the Pimas suggest that the progression from normal to diabetes can be considered to involve two stages. The first, primarily attributable to insulin resistance, leads to impaired glucose tolerance, and the second which depends on insulin secretory failure, leads to worsening hyperglycemia and evert diabetes [7-8]. Many research articles investigate the relative influences of birth weight and maternal diabetes on the development of obesity in the offsprings of Pima Indian women [14-16].

Evidence shows that diabetes is one of the main risk factors for developing many diseases. Many researches focus on causal factors of diabetes mellitus based on some simple statistical techniques [17-21], which are *inappropriate* in many cases [22, 23]. Thus, the earlier findings invite some doubts and debates. In general, any physiological data set is positive, and the variance may be non-constant as it may have relationship with the mean. In this situation, an appropriate method is the transformation of the response variable which is used to stabilize the variance, making the distribution of the response variable closer to the Normal distribution, and improving the fit of the model for the given data set. However, in practice the variance may not always be stabilized under a proper (seems to be suitable) transformation [22, 23]. In analyzing diabetes mellitus data sets, many authors [17] feel that variances are non-constant, thus the earlier researchers have used "log transformation", which may *not* always be appropriate [23, 24]. Generally, some positive data sets (variance with mean relationship) are analyzed by log-normal or gamma models [22-27]. The problem of non-constant variance (for the response variable) in linear regression is a departure from the standard least squares assumptions. The inequality of variance problem occurs relatively often in practice, frequently in conjunction with a non-Normal response variable. These observations (non-constant variance and non-Normal distribution) have motivated us to identify the causal factors of diabetes mellitus using joint modeling of mean and variance simultaneously.

The main goal of this article is to identify the causal factors of diabetes mellitus. The present article considers two different study groups, namely, the Pima Indian mothers and young Indian medical students. For Pima Indian mothers, it has been detected that triceps skin fold thickness, serum insulin, body mass index, diabetes pedigree function and age are the causal factors for developing diabetes mellitus. In case of Indian young medical students, it has been identified that age, body mass index, family history of diabetes mellitus, sex, low density lipoprotein, total cholesterol, serum triglyceride, family blood pressure, dietary habits like eating outside are the causal factors for developing diabetes mellitus. This study supports some of earlier research findings. However, it identifies many new causal factors which are little known in the literature. Many significant factors have been identified in the mean and the variance models. The present research also identifies the causal relationships between the risk factors and the disease diabetes mellitus for the Pima Indian mothers, and young Indian medical students. The effects of different causal factors on diabetes mellitus are illustrated.

2. METHODOLOGY: JOINT GENERALIZED LINEAR GAMMA MODELS

For the diagnosis of diabetes, recent criteria suggested by American Diabetes Association (ADA 2013) are described as follows. It is well known that the Haemoglobin A1c (HbA1c) is the pre-eminent factor for quantifying the risk of complications in patients with diabetes and for monitoring glycemia. For a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, (a) if HbA1c ≥ 6.5% (the test should be performed in a laboratory using a method that is NGSP certified and standardized to the Diabetes Control and Complications Trial (DCCT) assay), (b) if fasting (fasting is defined as no caloric intake for at least 8 hours) plasma glucose (FPG) ≥ 126 mg/dL (7.0 mmol/L), (c) if 2-hour plasma ≥ 200 mg/dL (11.1 mmol/L) during an oral glucose tolerance test (OGTT) (the test should be performed as described by the World Health Organization, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water), or (d) if random plasma glucose ≥ 200 mg/dL (11.1 mmol/L), the patient is identified as diabetic. Note that the diabetes markers HbA1c, FPG and 2-hour plasma glucose levels are all positive characteristics (values are always positive). A positive data set is generally analyzed by log-normal or gamma models. Below we describe gamma joint generalized linear models.

In practice, the class of generalized linear models includes distributions useful for the analysis of some continuous positive characteristic data sets which have non-Normal error distributions. The simplest examples are perhaps the exponential and gamma distributions, which are often useful for modeling positive data that have variance with mean relationship, and the variance of the response is non-constant. Nelder and Lee [28] proposed a modeling approach for the analysis of positive data y_i. They advocated the use of joint generalized linear model (JGLM):

$$E(y_i)=\mu_i \text{ and } Var(y_i)=\sigma_i^2 V(\mu_i),$$

where $V(\cdot)$ is the variance function and σ_i^2's are the dispersion parameters. In GLMs, the variance consists of two components, one is $V(\mu_i)$ which depends on mean changes, and the other is σ_i^2 which is independent of mean adjustment. In GLMs, the variance function characterizes the distribution of GLM family. For example, the distribution is normal if $V(\mu)=1$, Poisson if $V(\mu)=\mu$, gamma if $V(\mu)=\mu^2$, etc. Joint models for the mean and the dispersion parameters are

$$\eta_i=g(\mu_i)=x_i^t\beta \text{ and } \varepsilon_i=h(\sigma_i^2)=w_i^t\gamma,$$

where $g(\cdot)$ and $h(\cdot)$ are GLM link functions (i.e., the relationship between the mean and the linear predictors or the relationship between the variance and the linear predictors) for the mean and the dispersion, respectively; and x_i^t, w_i^t are respectively, the row vectors for the regression models of mean and dispersion, based on the levels of control variables. Maximum likelihood (ML) method is used to estimate the mean model parameters, and restricted ML (REML) estimators are used for the dispersion model [28]. A detailed discussion on GLMs approach is given in [23, 29-34].

3. ANALYSIS AND INTERPRETATION OF DIABETES MELLITUS DATA

3.1. Description of Pima Indian Diabetes Mellitus Data

This study is based on the data of Pima Indians diabetes database. The source of the data set is National Institute of Diabetes and Digestive and Kidney Diseases (Donor of database: Dr. Vincent Sigillito (vgs@aplcen.apl.jhu.edu), Research Center, RMI Group Leader, Applied Physics Laboratory, The Johns Hopkins University, Johns Hopkins Road, Laurel, MD 20707). This database is formed from all female patients (768 subjects) at least 21 years old of Pima Indian heritage. The population lives near Phoenix, Arizona, USA. For each study unit, the following characteristics have been recorded: 1. Number of times pregnant (NTPREG or x_1), 2. Plasma glucose concentration level (PGL) a 2 hours in an oral glucose tolerance test (y), 3. Diastolic blood pressure (mm Hg) (DBP or x_3), 4. Triceps skin fold thickness (mm) (TSFT or x_4), 5. 2-Hour serum insulin (mu U/ml) (SRMI or x_5), 6. Body mass index (weight in kg/(height)^2 in m) (BMI or x_6), 7. Diabetes pedigree function (DPF or x_7), 8. Age (years) (AGE or x_8), and 9. Class variable (non-diabetes = 0 or diabetes =1)(CV or x_9). According to World Health Organization criteria, a patient shows signs of diabetes if the 2 hour post-load plasma glucose is at least 200 mg/dl at any survey examination or if it is found during routine medical care.

The above data set is for Type 2 DM patients. Usually in clinical practice, 2-hour post prandial glucose levels are used, not serum insulin. In Type-2 DM, the balance (harmony) between insulin secretion and glucose suppression is lost, that is why both variables (blood glucose and insulin) were measured at the same time. This can be used in otherwise normal subjects, but not in diabetics. Furthermore, it cannot be concluded that persons were normal, as they might have underlying impaired glucose tolerance (not diabetic, but borderline) and which can also produce imbalance between insulin secretion and glucose suppression. The only fact which can be utilized to identify (not 100%) normal subjects is by negative family history of diabetes, which was not considered in this data set. This could introduce bias in the data and statistical analysis. Note that for this data set the only DM marker is plasma glucose concentration level (PGL) a 2 hours in an oral glucose tolerance test (y). Smith *et al.* [35] used this data to forecast the onset of diabetes mellitus. These researchers were unable to identify that the variance of the response is non-constant. Thus, their forecasting may mislead the factors.

3.2. Analysis of Pima Indian Mothers Diabetes Mellitus Data

For the analysis of positive data, log-normal and gamma models are generally used [22-27]. In the present section, the above mentioned diabetes data set is analyzed by using the gamma JGLMs as given in Section 2. Both the log-normal and the gamma model analyses [23] are examined for this data set, and It is found that the gamma model analysis is much better than the log-normal, so only the results of gamma model analysis are reported in Table **1**.

Table 1. Results for mean and dispersion models of plasma glucose concentration data of Pima Indian mothers from gamma fit.

	Covariate	Estimate	Standard Error	T	P-Value	95% C.I.
Mean Model	Const.	4.3746	0.0374	116.99	< 0.01	[4.3013, 4.4479]
	TSFT	–0.0020	0.0006	-3.31	< 0.01	[–0.003, –0.001]
	SRMI	0.0007	0.0001	11.63	< 0.01	[0.0005, 0.0009]
	BMI	0.0064	0.0011	6.01	< 0.01	[0.0042, 0.0086]
	DPF	0.0429	0.0229	1.87	0.06	[–0.0020, 0.0878]
	AGE	0.0054	0.0007	7.60	< 0.01	[0.0040, 0.0068]
Dispersion Model	Const.	–3.7984	0.2733	-13.90	< 0.01	[–4.3341, –3.2627]
	SRMI	–0.0018	0.0004	-4.10	< 0.01	[–0.0026, –0.0010]
	BMI	0.0143	0.0069	2.07	0.04	[0.0008, 0.0278]
	AGE	0.0125	0.0046	2.72	0.01	[0.0035, 0.0215]

The present analysis treats 'plasma glucose concentration a 2 hours in an oral glucose tolerance test' as the response variable (y). A diabetic or non-diabetic patient is determined based on 'plasma glucose concentration level (PGL)', thus, we treat 'PGL' as the response variable. A patient shows the sign of diabetes if PGL is high. Therefore, it is interested to identify the factors which have any effect on PGL. So, PGL is treated as the response variable and the remaining other variables are considered as the explanatory variables. It is identified that PGL is a non-constant variance response. Thus, we fit the data using joint gamma models as in Section 2, and the results are displayed in Table **1**. The selected models have the smallest Akaike information criterion (AIC) value in each class. It is well known that AIC selects a model which minimizes the predicted additive errors and squared error loss ([36], p. 203-204). It is not necessary that *all* the selected effects are significant [36].

Fig. (**1a**) presents the histogram of residuals, which is almost symmetric. It does not show any lack of fit due to missing variables. In Fig. (**1b**), we plot the absolute residuals, with respect to the fitted values. Fig. (**1b**) is completely a flat diagram with the running means, an indication that the variance is constant under the gamma fitting [31]. Fig. (**2a**, **b**) display respectively the normal probability plot for the mean and the variance model (Table **1**). Figs. (**1a**, **2a**, **b**) do not show any lack of fit due to missing variables and outliers, as there is not any gap in any one of these figures. Neither figure shows any systematic departures, indicating no lack of fit of the selected final models (Table **1**).

The fitted mean and variance models (Table **1**) for 'plasma glucose concentration level (Pima Indian mothers data)', respectively, are

$$\hat{y} = e^{4.3746 - 0.0020\,x4 + 0.0007\,x5 + 0.0064\,x6 + 0.0429\,x7 + 0.0054\,x8},$$

and

$$\widehat{\sigma^2} = e^{(-3.7984 - 0.0018\,x5 + 0.0143\,x6 + 0.0125\,x8)}.$$

Interpretation of the regression coefficients for the mean and the variance models can be drawn from these above two models. Note that the interpretation is same as usual regression analysis if we consider the response as 'log $\hat{y}$' and 'log $\widehat{\sigma^2}$'.

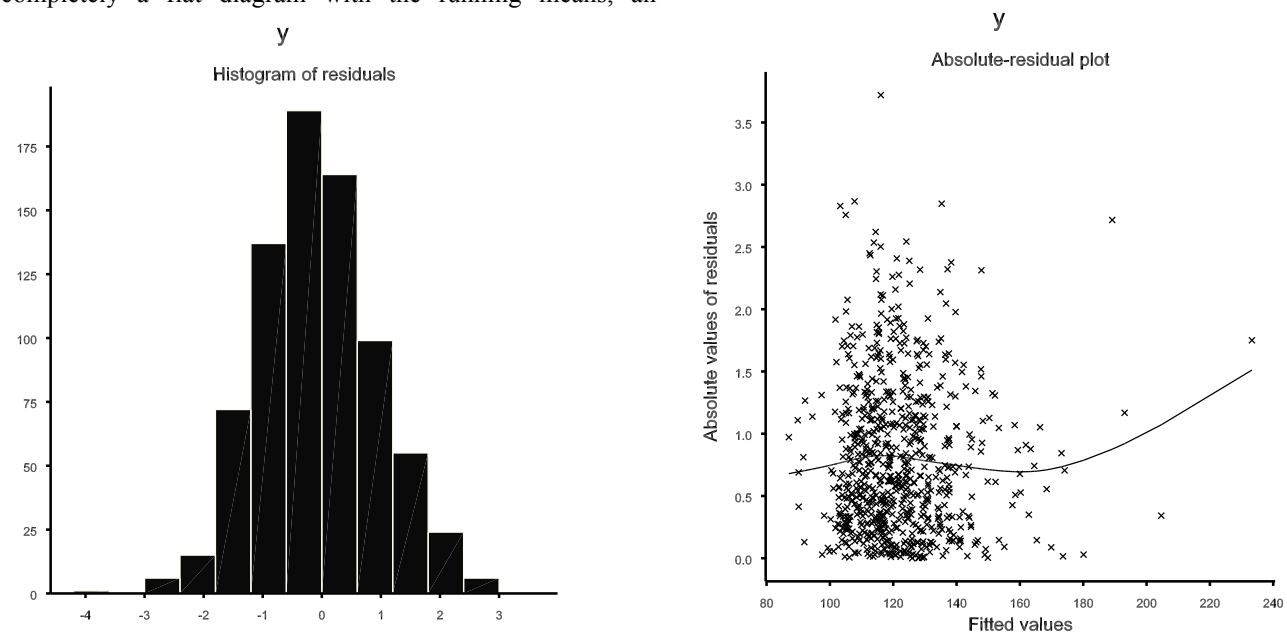

Fig. (1). (**a**) The histogram plot of residuals and (**b**) the absolute residuals plot with respect to the fitted values (Table **1**) of plasma glucose concentration level for Pima Indian mothers.

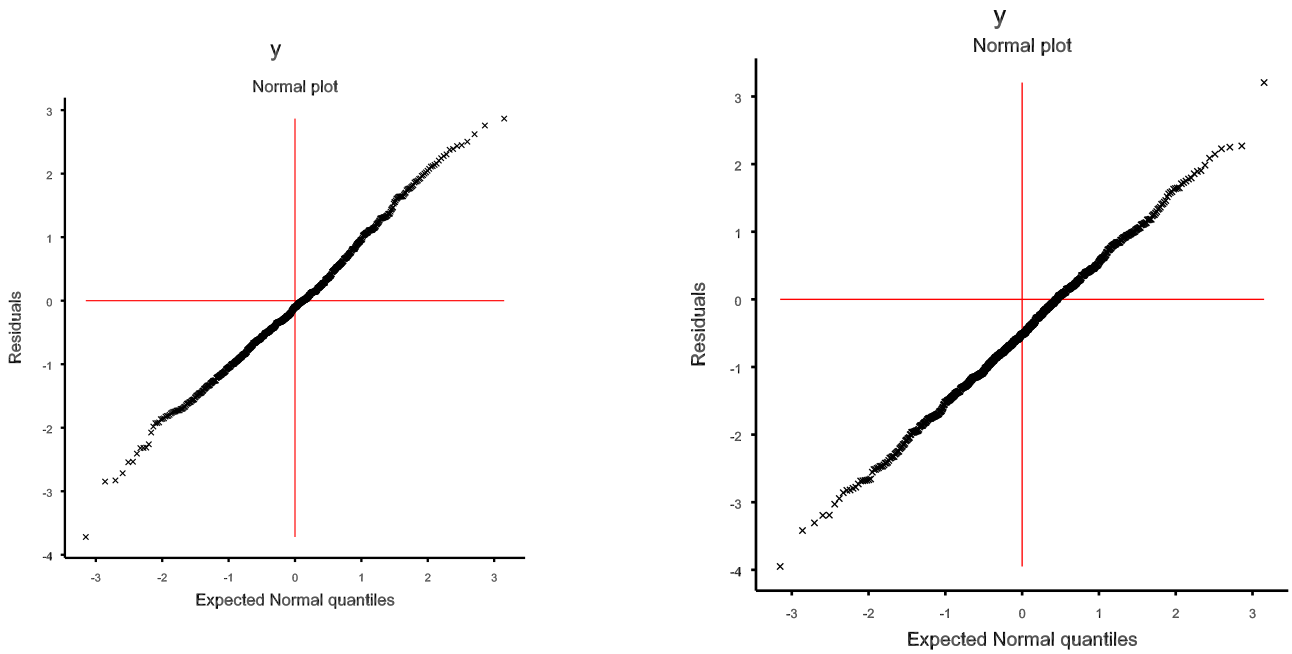

Fig. (2). The normal probability plot for the (**a**) mean and (**b**) variance model (Table **1**) of plasma glucose concentration level for Pima Indian mothers.

3.3. Interpretation of Pima Indian Mothers Diabetes Mellitus Data

Table **1** presents the following facts for the mean and the variance models of plasma glucose concentration level (PGL):

1. Triceps skin fold thickness (mm) (x_4) is statistically significant. It is negatively associated with the PGL. This indicates that if (x_4) decreases, PGL increases, and vice versa.

2. A 2-Hour serum insulin level (mu U/ml) (x_5) is positively (significant) associated with the PGL. It implies that if serum insulin level increases, PGL also increases, and vice versa. If serum insulin is at low level, PGL is also at low level, indicating that patient is non-diabetic.

3. Body mass index (x_6) is positively associated with the PGL. It indicates that if body mass index increases, PGL also increases, and vice versa. This is observed in practice for Type 2 diabetes mellitus patients [3, 15, 17]. Obesity, insulin resistance, and hyperinsulinemia are common features of non-insulin-dependent diabetes mellitus (NIDDM) [17].

4. Diabetes pedigree function (x_7) is positively associated with the PGL. This implies that if diabetes pedigree function increases, PGL also increases. This is practically true as diabetes pedigree function is determined based on PGL.

5. Age (x_8) is positively associated with the PGL. It indicates that if age increases, PGL increases, and vice versa. In general, Type 2 DM is observed at higher ages [8]. This data set is for Type 2 DM women. Therefore, this relation supports the earlier research findings [8].

6. Variance of the response PGL is non-constant. The variance depends on (x_5), (x_6) and (x_8). These are all statistically significant factors in the variance model. Thus, (x_5), (x_6) and (x_8) have dual effects (in the mean and the variance models) on PGL. Note that (x_6) and (x_8) have the *same* association with the PGL in the mean and the variance models, while (x_5) has opposite association. Therefore, the functional relationship of PGL with the explanatory variables is very complicated.

This analysis also shows that PGL (i.e., diabetes) is independent of number of times pregnant (x_1) and diastolic blood pressure (x_3). Standard error of all the estimated factors are very small (Table **1**), indicating that estimates are stable [31]. The present analysis identifies many causal factors of PGL in the mean and the variance models. These results will be helpful to medical practitioners for taking fruitful action.

3.4. Description of Indian Medical Students Case Study Data

Bhattacharyya *et al.* [37, 38] conducted a study on the students of a tertiary level referral and teaching hospital--RG Kar Medical College, Kolkata, West Bengal, India. The subjects of the study were selected by simple random sampling method from the students of RG Kar Medical College, Kolkata. The students of this Medical College were taken up in the study; those having extreme dietary habits or any facets influencing the lipid profile were excluded from the study; then 10% of the students were selected by simple random sampling method. Thus, 64 subjects were taken up as sample in the study after they were informed of the nature of the study and then getting consent from them for inclusion in the study.

The data regarding their identity, sex (SEX) (female = 1, male = 2), age (AGE), height (HT), weight (WT), obesity (OB) (under weight =1, all others = 2), body mass index

(BMI), lifestyle, dietary habits like eating in outside (EO) (i.e., eating of ready food habit), types of oil consumption (OC) (mustered oil = 1, all others = 2), smoking habit (SH), family history of diabetes mellitus (FDM) (nil = 1, all others =2), family blood pressure (FBP) (nil = 1, all others =2), hypertension (HT) (nil = 1, all others =2), and coronary heart disease (CHD) (nil = 1, all others =2), history of past illness, history of any drug intake, and biochemical parameters like fasting plasma glucose level (PGL), serum triglyceride (STG), total cholesterol (TC), high density lipoprotein (HDL), low density lipoprotein (LDL), fasting serum insulin (FI), were collected in a predesigned and pretested proforma. The method of collection of this data and a part of it is given in [38]. We have not included the data set in this paper as it would increase the length of the paper, but we can submit the data on requirement for verification of the present derived results. The BMI was calculated by the formula 'weight (in kg)/height^2 (in metre)'.

3.5. Analysis of Indian Medical Students Case Study Data

Indian medical students case study data contains two types of characteristics, namely, factors and variables. In the present analysis, each factor is leveled in two groups, namely, low and high, where high group includes all the levels except the low level. For factors, we accept the constraint that the effects of the first levels are zero. That is, we take the first level of each factor as the reference level by estimating its as zero. Suppose that $\mathbf{a}_i$ for $i = 1,2,3$ represents the main effect of $\boldsymbol{A}$. We take $\hat{\mathbf{a}}_1 = 0$, so that $\hat{\mathbf{a}}_2 = \hat{\mathbf{a}}_2 - \hat{\mathbf{a}}_1$. For example, the estimate for the effect *A2* (second level of $\boldsymbol{A}$) means the effect of difference between the second and the first levels in the main effect $\boldsymbol{A}$, i.e., $\hat{\mathbf{a}}_2 - \hat{\mathbf{a}}_1$.

In this subsection, the response fasting plasma glucose level (PGL) has been predicted by the remaining other explanatory variables, using the gamma JGLMs as in Section 2, and the results of the analysis are displayed in Table **2**. Only the variable 'low density lipoprotein (LDL)' is significant in the mean model, whereas many variables and factors, namely, age, BMI, serum triglyceride (STG), sex, dietary habits like eating in outside (EO) (i.e., eating of ready food habit), family history of diabetes mellitus (FDM), family blood pressure (FBP), are significant in the dispersion model.

In Fig. (**3a**, **b**), we plot respectively the absolute residuals with respect to the fitted values and the normal probability plot for the gamma fitted mean models (Table **2**). Fig. (**3a**) is a flat diagram with the running means, indicating that the variance is constant under the gamma JGLMs fitting. Fig. (**3b**) does not show any systematic departures, indicating no lack of fit of the selected final models (Table **2**). Partially insignificant effects (in epidemiology known as confounder) total cholesterol (TC) and serum triglyceride (STG) are included in the mean model for better fitting.

3.6. Interpretation of Indian Medical Students Case Study Data

Table **2** presents the analysis of fasting plasma glucose level (PGL), and it presents the following facts for the mean and the variance models of PGL of Indian medical students:

1. In the mean model, low density lipoprotein (LDL) is positively associated with the mean level of PGL. This indicates that if LDL increases, mean value of PGL increases, and vice versa.
2. In the dispersion model, age, BMI, sex and family history of diabetes mellitus (FDM) are positively associated with the variance of PGL. This implies, if age or BMI increases, the variance of PGL increases. Sex is coded as female = 1 and male = 2. Sex is positively associated with the variance of PGL, indicating that the variance of PGL is high for male sex. In other words, the variability of PGL is higher for a male than a female. Family history of diabetes mellitus (FDM) is coded as high group = 2 (includes all the levels except the low level) and low group = 1 (only the low level). FDM is positively associated with the variance of PGL, indicating that higher FDM level increases the variance of PGL. That is, the variability of PGL is higher for an individual with the presence of FDM.
3. Serum triglyceride (STG), dietary habits like eating in outside (EO), family blood pressure (FBP) are negatively associated with the variance of PGL, indicating that the variability of PGL is lower for an individual with higher STG, dietary habits, or family

Table 2. Results for mean and dispersion models of fasting plasma glucose level (PGL) data of Indian medical students from gamma fit.

	Covariate	Estimate	Standard Error	T	P-Value	95% C.I.
Mean Model	Const.	4.2071	0.0146	288.82	< 0.01	[4.1175, 4.2367]
	LDL	0.0012	0.0002	2.67	0.01	[0.0008, 0.0016]
	TC	0.0004	0.0003	1.61	0.11	[–0.0002, 0.0010]
	STG	0.0002	0.0001	-1.17	0.25	[0.0000, 0.0004]
Dispersion model	Const.	-13.133	1.826	-7.19	< 0.01	[–16.611, 9.553]
	AGE	0.202	0.094	2.14	0.04	[0.018, 0.386]
	BMI	0.189	0.046	4.15	< 0.01	[0.099, 0.279]
	STG	-0.020	0.005	-4.13	< 0.01	[–0.029, –0.011]
	SEX2	3.831	0.453	8.45	< 0.01	[2.943, 4.719]
	EO	-0.009	0.002	-4.44	< 0.01	[–0.013, – 0.005]
	FDM2	2.242	0.418	5.36	< 0.01	[1.423, 3.063]
	FBP2	-1.203	0.402	-2.99	< 0.01	[–1.991, –0.415]

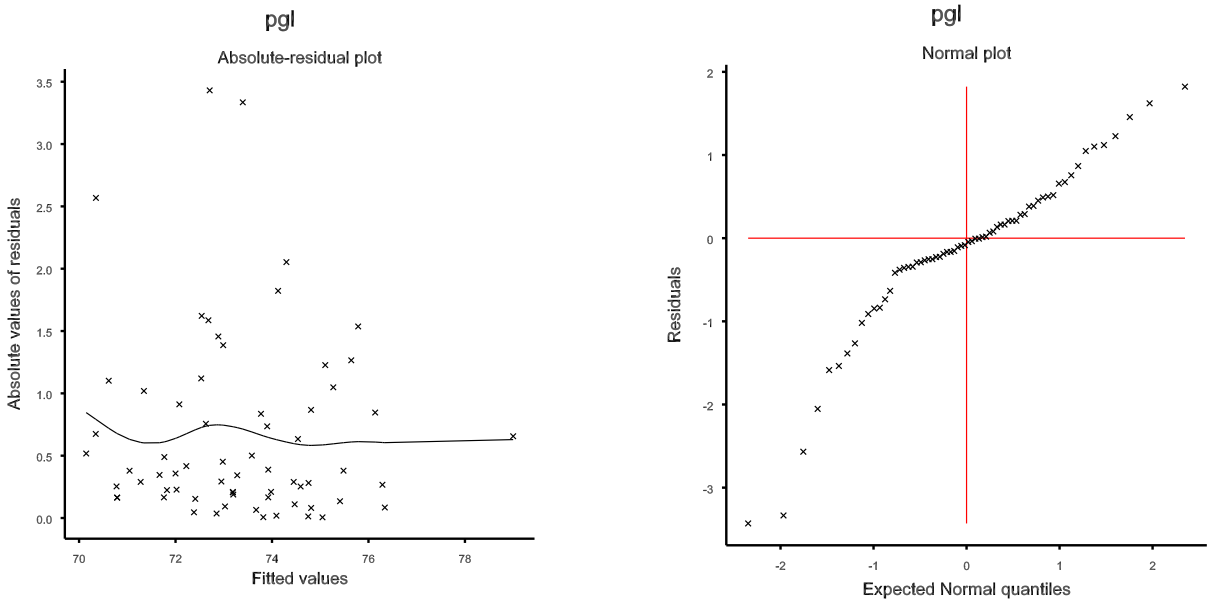

Fig. (3). (**a**) The absolute residuals plot with respect to the fitted values and (**b**) the normal probability plot for the mean model (Table **2**) of PGL data for Indian medical students.

blood pressure.

4. CONCLUSION

In the present article, the determinants of DM are identified from two data sets. One data set contains the diabetes marker '2-hours plasma glucose concentration level', and the other data set contains the diabetes marker 'fasting plasma glucose level'. Response '2-hours plasma glucose concentration level' has been modeled for the Pima Indian mothers and 'fasting plasma glucose level' has been modeled for the Indian medical students, using gamma JGLMs. Fitted models are given in Section 3. Final models are selected based on AIC and graphical analysis. Analyses show that both the responses '2-hours plasma glucose concentration level' and 'fasting plasma glucose level' fit well the gamma JGLMs. At present we have no data set with HbA1c diabetes marker. It is interesting to examine the determinants of DM with HbA1c. It will be more better if the same data set contains the above three DM markers along with many interested covariates. It is expected that the analysis of these three DM markers will give the same determinants. Hope that future research on DM will consider all the three DM markers along with the explanatory variables lifestyle characteristics, dietary habits, family history, blood biochemical parameters, and consequently many interesting results will be focused to the literature.

This article examines two study groups. One group is the Pima Indian mothers and the other is Indian young medical students. There are some diabetic and non-diabetic patients among the Pima Indian mothers. The subjects Indian medical students (both male and female) are very young (18-33 years). None of them is a diabetic patient. These two group of subjects are completely different from different origin, region, age, sex and disease distributions. Yet, there are two covariates AGE and BMI are common to both the groups. But these two groups contain mainly different covariates with different interest of study. The present study aims to identify the causal factors of DM from different source of studies and to compare the effects of the common causal factors to the disease DM for different studies. It is expected that the common covariates (causal factors) should have similar effects on the disease even though the study groups are completely different. Here it is identified that the common causal factors AGE and BMI have similar effects on the disease DM for both the groups. Moreover, the present study identifies the different causal factors of DM for these two study groups. Based on our knowledge, this is the first study on DM for the Indian medical students for this data set. Only BMI of these medical students has been studied by Bhattacharyya *et al*. [37, 38]. Das [39] has derived the relationships of human blood biochemical parameters for these Indian medical students. These authors did not study DM for the Pima Indian mothers or any other population with the Indian medical students.

Earlier researches have identified only the parental diabetes, genetic makers, obesity, diet, age etc., as the causal factors for diabetes mellitus. However, the present study identifies many additional new causal factors (Tables **1** and **2**). This study supports the earlier research findings that the parental diabetes (Table **2**), obesity (or body mass index (BMI)), age (Tables **1** and **2**) and diet (Table **2**) are the causal factors for diabetes mellitus. Parental diabetes (or family history diabetes mellitus (FDM)) (Table **2**) and BMI (Tables **1** and **2**) are positively associated with the variance of Plasma glucose level (PGL). Also BMI is positively associated with the mean of PGL (Table **1**). Dietary habits like eating in outside (EO) (Table **2**) is negatively associated with the variance of PGL. Moreover, the present study identifies many additional new causal factors. Age (AGE) (Table **1**), serum insulin (SRMI) (Table **1**), diabetes pedigree function (DPF) (Table **1**), low density lipoprotein (LDL) (Table **2**) are positively associated with the mean of PGL. Triceps skin fold thickness (TSFT) (Table **1**) is negatively associated with the mean value of PGL. Earlier researches

assume that the variance of the response PGL is constant. However, the present study identifies that the variance of PGL is non-constant. This article identifies many causal factors for explaining the variance of PGL. BMI (Tables **1** and **2**), AGE (Tables **1** and **2**), family history of diabetes mellitus (FDM) (Table **2**) and SEX (Table **2**) are positively associated with the variance of PGL. Also serum insulin (SRMI) (Table **1**), serum triglyceride (STG) (Table **2**), dietary habits like eating in outside (EO) (Table **2**) and family blood pressure (FBP) (Table **2**) are negatively associated with the variance of PGL. Note that total cholesterol (TC) and serum triglyceride (STG), *partially significant* effects, known as *confounders* in epidemiology, are positively associated with the mean of PGL (Table **2**).

Table **2** shows that family history of diabetes mellitus (FDM) and family blood pressure (FBP) have significant effects on the variance of PGL. Thus, we can conclude that parental diabetes and blood pressure are responsible for the development of diabetes. The present research identifies the genetic and other risk factors (Tables **1** and **2**) of DM through the mathematical relationships. The causal factors of diabetes mellitus in the variance models (Table **1** and **2**) are believed to be new in the literature. Moreover, mean models (Tables **1** and **2**) provide many new additional causal factors of diabetes mellitus. This particular research presents the association and the effects of different risk factors on DM, which will be helpful to medical practitioners for taking fruitful action.

To fill the gaps in the DM research literature, this study derives the relationships of '2-hours plasma glucose concentration level' and 'fasting plasma glucose level' with different risk factors. The mathematical models (in Tables **1** and **2**) in this report show the relationships of these two DM markers. The models reported here illuminate the complex relationships. Fortunately, a true mathematical model can open the truth that is covered by the complex relationships. Our results, though not completely conclusive, are revealing –

*Our findings confirm many previous research findings (Section 3).

*An important conclusion has to do with the use of earlier used statistical models. While further research is called for, we find that the joint gamma models are much more effective than either traditional simple, multiple regression or Log-Gaussian models (with constant variance), because they better fit the data. In short, research should have greater faith in these results than those emanating from the simple, multiple regression or Log-Gaussian (with constant variance) models.

To prevent diabetes, this study suggests the following. BMI of an individual should be in normal (Tables **1** and **2**). Diastolic blood pressure (Table **1**) and family blood pressure (Table **2**) should not be high. DM markers should be measured at regular intervals at higher age (Tables **1** and **2**). Every adult should measure fasting lipid profile at least annually (Table **2**). Individuals with family history of DM (or without), should modify their lifestyle focusing on the reduction of fat and cholesterol intake (Table **2**). To reduce the risk of DM, optimize blood pressure control.

CONFLICT OF INTEREST

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

ACKNOWLEDGEMENTS

The author is very much indebted to the Editor and the referees who have provided valuable comments to improve this paper. The author also thanks Dr. Vincent Sigillito (vgs@aplcen.apl.jhu.edu), Research Center, RMI Group Leader, Applied Physics Laboratory, The Johns Hopkins University, Johns Hopkins Road, Laurel, MD 20707) and Dr. Bhattacharyya and others who generously provided the data sets to freely distribute and use for non-commercial purposes.

REFERENCES

[1] Gottdiener JS, Tracy RP, Arnold AM. Predictors of congestive heart failure in the elderly: the cardiovascular health study. J Am Coll Cardiol 2000; 35 (B): 1628-37.

[2] Nichols GA, Gullion CM, Koro CE, Ephross, SA, Brown, JB. The incidence of congestive heart failure in 2 diabetes: An update. Diabetes Care 2004; 27: 1879-84.

[3] Shamseddeen H, Getty JZ, Hamdallah IN, Ali MR. Epidemiology and economic impact of obesity type 2 diabetes. Surg Clin North Am 2011; 91: 1163-72.

[4] Kannel WB. Vital epidemiologic clues in heart failure. J Clin Epidemiol 2000; 53: 229-35.

[5] Bennett PH, Burch TA, Miller M. Diabetes mellitus in American (Pima) Indians. Lancet 1971; 2: 125-8.

[6] Bennett PH, Rushforth NB, Miller M, LeCompte PM. Epidemiologic studies of diabetes in the Pima Indians. In: Greep RO, Eds. Recent Progress in Hormone Research. New York: Academic Press 1976; pp. 333-76.

[7] Knowler WC, Pettitt DJ, Saad MF, Bennett PH. Diabetes mellitus in Pima Indians incidence risk factors and pathogenesis. Diabetes Metab Rev 1990; 6: 1-27.

[8] Knowler WC, Nelson RG, Saad MF, Bennett PH, Pettitt DJ. Determinants of diabetes mellitus in the Pima Indians. Diabetes Care 1993; 16: 216-27.

[9] Parks JH, Waskow E. Diabetes among the Pima Indians of Arizona. Ariz Med 1961; 18: 99-106.

[10] Dort A, Ballintine EJ, Bennett PH, Miller M. Retinopathy in Pima Indians relationships to glucose level duration of diabetes age at diagnosis of diabetes and age at examination a population with a higher prevalence of diabetes mellitus. Diabetes 1976; 25: 554-60.

[11] Pettitt DJ, Knowler WC, Lisse JR, Bennett PH. Development of retinopathy and proteinuria in relation to plasma-glucose concentrations in Pima Indians. Lancet 1980; 4: 1050-2.

[12] Jarrett RJ, Keen H. Hyperglycemia and diabetes mellitus. Lancet 1976; 2: 1009-12.

[13] Knowler WC, Pettitt DJ, Nelson RG, Howard BV, Bogardus C, Bennett PH. Obesity in the Pima Indians its magnitude and relationship with diabetes. Am J Clin Nutr 1991; 53: S1541-3.

[14] Reid JM, Fullmer SD, Pemgrew KD, *et al*. Nutrient intake of Pima Indian women relationships to diabetes mellitus and gallbladder disease. Am J Clin Nutr 1971; 24: 1281-9.

[15] Pettitt DJ, Baird HR, Aleck KA, Bennett PH, Knowler WC. Excessive obesity in offspring of Pima Indian women with diabetes during pregnancy. N Eng J Med 1983; 308: 242-5.

[16] Pettitt DJ, Knowler WC, Bennett PH, Aleck KA, Baird HR. Obesity in offspring of diabetic Pima Indian women despite normal birth weight. Diabetes Care 1987; 10: 76-80.

[17] Odeleye OE, Courten M, Pettitt DJ, Ravussian E. Fasting hyperinsulinemia is a predictor of increased body weight and obesity in Pima Indian children. Diabetes 1997; 46: 1341-5.

[18] Bogardus C, Lillioja S, Howard BV, Reaven G, Mott D. Relationship between insulin secretion, insulin action, and fasting plasma glucose in non-diabetic and non-insulin-dependent diabetic mellitus in Pima Indians. J Clin Invest 1985; 74: 1238-46.

[19] Lillioja S, Mott D, Spraul M, *et al*. Insulin resistance and secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. N Engl J Med 1993; 329: 1988-92.

[20] Valdez R, Mitchelj BD, Haffner SM, *et al*. Predictors of weight change in a bi-ethnic population: the San Antonio Heart study. Int J Obes 1994; 18: 85-91.

[21] Reaven GM, Chen YDI, Hollenbeck CB, Sheu WHH, Ostraga D, Polonsky KS. Plasma insulin, C-peptide, and proinsulin concentrations in obese and non-obese individuals with varying degrees of glucose tolerance. J Clin Endocrinol Metab 1993; 476: 44-4.

[22] Myers RH, Montgomery DC, Vining GG. Generalized Linear Models with Applications in Engineering and the Sciences. New York: John Wiley & Sons 2002.

[23] Das RN, Lee Y. Log normal versus gamma models for analyzing data from quality-improvement experiments. Quality Engineering 2009; 21(1): 79-87.

[24] Das RN, Park JS. Discrepancy in regression estimates between Log-normal and Gamma: Some case studies. J Appl Stat 2012; 39(1): 97-111.

[25] Firth D. Multiplicative errors: log-normal or gamma?. J R Statist Soc 1988; 50: 266-8.

[26] McCullagh P, Nelder JA. Generalized linear models. London: Chapman & Hall 1989.

[27] Das RN. Discrepancy in fitting between log-normal and gamma models: An illustration. Model Assisted Stat Appl 2012; 7(1): 23-32.

[28] Nelder JA, Lee Y. Generalized linear models for the analysis of Taguchi-type experiments. Appl Stochastic Model Data Anal 1991; 7: 107-20.

[29] Lee Y, Nelder JA. Generalized linear models for the analysis of quality improvement experiments. Can J Stat 1998; 26: 95-105.

[30] Lee Y, Nelder JA. Robust design *via* generalized linear models. J Qual Tech 2003; 35: 2-12.

[31] Lee Y, Nelder JA, Pawitan Y. Generalized Linear Models with Random Effects (Unified Analysis *via* H-likelihood). London: Chapman & Hall 2006.

[32] Lesperance ML Park S. GLMs for the analysis of robust designs with dynamic characteristics. J Qual Tech 2003; 35: 253-63.

[33] Qu Y, Tan M, Rybicki L. A unified approach to estimating association measures *via* a joint generalized linear model for paired binary data. Commun Stat Theory Meth 2000; 29: 143-56.

[34] Wolfinger RD, Tobias RD. Joint estimation of location, dispersion, and random effects in robust design. Technometrics 1998; 40: 62-71.

[35] Smith JW, Everhart JE, Dickson WC, Knowler WC, Johannes RS. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. Proceedings of the Symposium on Computer Applications and Medical Care, IEEE Computer Society Press, 1988; pp. 261-5.

[36] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. USA: Springer-Verlag 2001.

[37] Bhattacharyya S, Mukhopadhyay M, Bhattacharyya I, Lahiri SK, Mitra PK, Dutta UK. A study on body mass index, serum insulin level and lipid profile in normoglycemic young adults. Ind J Med Biochem 2006; 10(2): 33-7.

[38] Bhattacharyya S, Mukhopadhyay M, Bhattacharyya I, Lahiri SK, Mitra PK, Dutta UK. A study on body mass index (BMI) and some biochemical parameters of the medicos with family history of diabetes mellitus, hypertension and coronary heart disease. J Indian Med Assoc 2007; 107(7): 370-9.

[39] Das RN. Modeling of Biochemical Parameters. Model Assisted Stat Appl 2011; 6(1): 1-12.

Received: November 28, 2013 Revised: February 1, 2014 Accepted: February 27, 2014