**International Journal of Applied Mathematics**

**Volume 35 No. 5 2022, 787-794**
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
**doi:** http://dx.doi.org/10.12732/ijam.v35i5.10

# PREDICTIVE FACTORS FOR COMPLICATIONS IN PATIENTS WITH DIABETES MELLITUS

Sylvia V. Pashkunova

Military Medical Academy
3 St. Georgi Sofiiski str.
Sofia - 1606, BULGARIA

**Abstract:** A prospective study of the relationship between some clinical parameters and complications of the patients with diabetes is considered. About 200 patients (male and female) have been examined. The patients are classified into five groups subject to the type of the diabetes. The data obtained for each patient are related to the type of the complications – macro vascular, retina pathology, neuron pathology and nephrite pathology and 12 clinical parameters.

A mathematical model is build on the patients' data. A logistic regression to identify which factors are associated with the complications is performed. The software package used for statistical modeling of real data is STATISTICA 13.

**AMS Subject Classification:** 62P10
**Key Words:** diabetes-related complications; logistic regression; statistical analysis; probability density

## 1. Introduction

Diabetes mellitus (DM) is a chronic (long lasting), metabolic disease which is characterized by high blood sugar (glucose). Human metabolism consists in the simplest breakdown of food to basic molecules that are used to provide energy and building material for body cells. The main energy substrate for

Received: September 27, 2022 

humans is glucose. In a healthy person the level of blood glucose is regulated by several hormones, one of which is insulin. Insulin is produced by $\beta$-cells of the pancreas. High levels of blood glucose result from lack of production of the hormone insulin or lack of "response" of the cells to insulin. DM (also called diabetes only) is actually a lifelong disease that causes damage to blood vessels and nervous system after a longer period of time. These injuries are manifested as complaints from different organ systems such as the heart and circulatory system, retina and eye, kidney, peripheral nerves, and others (see [1]-[7]).

A survey of a four diabetics related complications in patients of the Department of Endocrinology of Military Medical Academy in Sofia is presented. Some clinical parameters are considered to be prognostic factors of these complications. The patients are classified into five groups subject to the type of the diabetes. A mathematical model build on the patient's data using logistic regression is performed in order to identify the clinical risk factors which were associated with the complications. The software package used for statistical modeling of real data was STATISTICA 13, [8].

Mathematical models have great potentialities as regards their utility in different disciplines of medicine and health.

## 2. Measurements

During the study period 94 healthy subjects (control group) and 206 patients (male and female) have been prospectively studied. The patients were divided into 5 groups depending on the mode of inheritance, the amount of insulin secretion, treatment and course of illness:

- Insulin-dependent type 1 diabetes (juvenile diabetes) – 43 patients;
- Insulin- not dependent type 2 diabetes mellitus (DM Age) – 60 patients;
- Insulin-dependent type 2 diabetes mellitus – 42 patients;
- Diabetes type MODY – 29 patients;
- Diabetes mellitus type LADA – 32 patients.

The data obtained for each patient included 12 clinical parameters as: Age (in years), Sex, Glucosed Hemoglobin (HbAl%), C-peptide, Creatinin, C-reactive Protein (CRP), Cholesterol, TRG, HDL, LDL, Duration (in years). The complications which are observed are retinopathy, neuropathy, nephropathy and macro vascular.

The primary data analysis is as follows: minimum age is 19 and maximum 81 years, 35% of patients were male and 65% women, 23% was observed retinopathy, minimum of duration is 1 year and maximum is 40 years, 75% had neuropathy, in 21% nephropathy is observed and 18% was observed with macro vascular complications. The sum of percentages is more than 100 because there are patients with 2, 3 or 4 complications.

## 3. Statistical methods and models

### 3.3 Dichotomous response mModels. Logistic regression model

Mathematical models have great potentialities as regards their utility in different disciplines of medicine and health, [9, 10, 11]. In this study the regression model is used to relate a categorical response (dependent variable $Y$) to the explanatory variables (predictors) $x_i$, $(i = 1, ..., c)$. We are interested of the presence (coded by 1) or the absent (coded by 0) of the four categorical responses:

- Retinopathy,
- Neuropathy,
- Nephropathy,
- Macro vascular complications.

At the first part predictors are 12 clinical parameters, and at the second part the explanatory variables are genotypes of the patients.

About the Dichotomous Response Models, [9, 10]: Let us define the dummy random variable to indicate the two categories by $Y = 1$ for category $A$ and $Y = 0$ for $B$. The probability density for $Y$ given the parameter $p$ is therefore point binomial

$$f(Y/p) = p^Y (1-p)^{(1-Y)}. \tag{1}$$

We assume that the probability $p$ depends on a linear function

$$d(x_1, ..., x_c) = \beta_0 + \sum_{i=1}^{c} \beta_i x_i, \tag{2}$$

where $x_i$ $(i = 1, ..., c)$ are the explanatory variables (the independent variables), $\beta_i$ are the constants. So, the joint conditional density is

$$f(y_1, ..., y_n/p(d_1), ..., p(d_n)) = \prod_j^n [p(d_j)]^{y_j} [1 - p(d_j)]^{(1-y_j)}, \tag{3}$$

where $n$ is the size of a random simple of data $y$ for response variable $Y$. To be able to relate value of $y$ to the value of $d$ a most specific assumptions about the form of $p(d)$ is required. In so called logit or logistic model the distribution function of logistic density is:

$$p(d) = \frac{\mathrm{e}^d}{1 + \mathrm{e}^d}. \tag{4}$$

The shape of $p(d)$ (logistic distribution) is quite similar to the shape for normal distribution. The odds of a dichotomous response is given by

$$o\,d\,d\,s = \left[\frac{p(d)}{1 - p(d)}\right]. \tag{5}$$

The logit transformation

$$\ln\left[\frac{p(d)}{1 - p(d)}\right] = d = \beta_0 + \sum_{i=1}^{c} \beta_i x_i, \tag{6}$$

gives an important advantage of the model because (2) is the linear function of the explanatory variables. The Newton-Raphson iterative procedure is usually used to make maximal likelihood estimator $\hat{\beta}$ of coefficient vector $\beta$ in the logistic model. The procedure is based on a preliminary estimator of $\beta$ given by $\hat{\beta} = (X^T X)^{-1} X^T Y$ where $Y$ is the vector of $y_i$ response values $(i = 1, ..., n)$ and $X(n \times c)$ is the matrix of observations. The maximum likelihood is obtained by solving the system of $(c + 1)$ equations:

$$\sum p_i x_i = \sum_{i=1}^{n} y_i x_i, \tag{7}$$

where

$$p_i = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}. \tag{8}$$

The solutions to these equations given by $\hat{\beta}$ can be used to obtain the estimator $\hat{p}_i$ for each of $n$ observations and hence the fitted sum $\sum_{i=1}^{n} \hat{p}_i x_i$ is equal to the observed sum in the right side. In comparison to the multiple linear regression model, the coefficient vector $\hat{\beta}$ must be interpreted differently:

- The coefficients $\hat{\beta}$ were interpreted as estimates of *log odds*;
- A marginal one unit increase in $x_j$ brings an increase in $d$ (i.e. in *log odds*) of the amount of $\hat{\beta}$;
- The confidence intervals were calculated for the odds estimates by taking the exponent of upper and lower endpoints of the asymptotic confidence interval for the *log odds*.

Testing of hypothesis concerning the regression parameters can include test of single parameter, test involving several parameters from the same regression, and joint tests involving parameters from different regressions. In polychotomous logistic regression, tests for contribution of one or more parameters from the same regression are usually constructed with a large sample Wald test, with test statistic

$$Q_W = \hat{\beta}^T \left[\mathrm{Var}(\hat{\beta})\right]^{-1} \hat{\beta}, \tag{9}$$

where $\mathrm{Var}(\hat{\beta})$ is the estimated covariance sub-matrix for the relevant parameters. This statistics is approximately distributed as a $\chi^2(r)$ random variable with $r$ degrees of freedom under the null hypothesis that $r$-dimensional vector $\hat{\beta}$ is equal to $\vec{0}$. When there is a single parameter of interest the test statistic is

$$Q_W = \left[\hat{\beta}_j / SE(\hat{\beta}_j)\right]^2. \tag{10}$$

$(SE(\hat{\beta}_j))$ is standard error of $\hat{\beta}_j$) and its distribution is $\chi^2(r = 1)$.

## 4. Application to the real data. Results and conclusions

Four logistic regressions, to identify which of 12 clinical parameters (factors) were associated with the complications (macro vascular, retinopathy, neuropathy and nephropathy), have been performed. As an example, the resulting models for retinopathy (for the different types of diabetics) are presented in Table 1.

From these results and according to the remarks in Subsection 3.3 (4) we can infer that the probability $p(d)$ for retinopathy is higher if:

- The values of "Creatinin" and "Duration" tend to minimum in the patient with type 1 DM;
- The values of "C-peptide" tend to minimum in the patient with type 2-isulin DM;

Table 1: Statistical significant clinical factors associated with the complication retinopathy identify by the logistic regression

| Logistic regression models: The complication Retinopathy as a function of predictive clinical factors | | | | | | |
|---|---|---|---|---|---|---|
| Type Diabetics | Factor | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $p$-level (Wald's $\chi^2$) | $p(d)_{\min}$ | $p(d)_{\max}$ |
| TYPE1 | creatinin | 1.487 | -0.166 | 0.016 | 0.79 | 0.004 |
| | duration | 3.218 | -0.133 | 0.033 | 0.96 | 0.11 |
| TYPE 2 - insulin | c-peptide | 1.117 | -0.767 | 0.045 | 0.74 | 0.27 |
| MODY | cholesterol | -8.135 | 0.822 | 0.022 | 0.001 | 0.64 |
| LADA | duration | -4.062 | 0.341 | 0.034 | 0.02 | 0.93 |

- The values of "Cholesterol" tend to maximum in the patient with type MODY DM;

- The values of "Duration" tend to maximum in the patient with type LADA DM.

For the factor "Duration", using minimum and maximum values in the corresponding samples, we estimate that for type 1 DM with increase of years, the probability of complication "retinopathy" decreased from 0.96 to 0.11, as for type LADA this probability increases from 0.02 to 0.93.

The part of analysis for macro vascular, neuron pathology and nephrite pathology have been carried out and published in [12].

Similar prognostic factors for the complications are published in other authors (see [1]-[7] and [13]).

## References

[1] R. Diez, L. Gorrono, L. Senatus, A. Shekhman, R. Ramasamy, A. Schmidt, Diabetes and cardiovascular complications: The epidemics continue, *Current Cardiology Reports*, **23**, No 7 (2021), 74–84; doi:10.1007/s11886-021-01504-4.

[2] B.M. Leon, T.M. Maddox, Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment recommendations and future research, *World J. Diabetes*, **6**, No 13 (2015), 1246–1258; doi:10.4239/wjd.v6.i13.1246.

[3] K. Papatheodorou, M. Banach, E. Bekiari, M. Rizzo, M. Edmons, Complications of Diabetes 2017, *J. Diabetes Res.*, (2018), Art. ID 3086167; doi:10.1155/2018/3086167.

[4] T. Pham, T. Nguyen, H. Truong, C. Trinh, Ha Thi Du, Tam Thi Ngo, L. Nguyen, Effects of diabetic complications on health-related quality of life impairment in Vietnamese patients with Type 2 Diabetes, *J. Diabetes Research*, (2020), Art. ID 4360804; doi:10.1155/2020/4360804.

[5] E.H. Sohn, H.W. van Dijk, C. Jiao, M.D. Abramoff, Retinal neurodegeneration may precede microvascular changes characteristic of diabetic retinopathy in diabetes mellitus, *PNAS*, **113**, No 19 (2016), E2655–E2664; doi:10.1073/pnas.1522014113.

[6] M. Tarek, E. Masry, A.M. Mona, Z. Abou, M. Mohamed, E. Tawil, A. Rawhia, A. Khalifa, Manganese Superoxide Dismutase Alanine to Valine Polymorphism and risk of neuropathy and nephropathy in Egyptian Type 1 Diabetic patients. *Review of Diabetic Studies*, **2** (2005), 70–74.

[7] I. Larson I, M.M. Hoffmann, J.M. Ordovas, E.J. Schaefer, W. Marz, J. Kreuzer, The lipoprotein lipase HindIII polymorphism: association with total cholesterol and LDL-choleterol, but not with HDL and triglycerides in 342 females, *Clin. Chem.*, **45** (1999), 963–968.

[8] StatSoft, Inc., *STATISTICA Manual (Data Analysis Software System)*, Version 13.0 (2019).

[9] J.D. Jobson, *Applied Multivariate Data Analysis*, Vol.II, Springer Verlag (1991), 278–303.

[10] Y. Uzunova, K. Prodanova, Comparison of binary regression models for the outcome of pediatric liver transplantation, *International Journal of Applied Mathematics*, **35**, No 4 (2022), 601–609; doi:10.12732/ijam.v35i4.8.

[11] L. Lazarova, N. Stoikovikj, A. Stojanova, M. Mileva, M. Ljubenovska, Mathematical model for prediction of Covid-19 dynamics, *International Journal of Applied Mathematics*, **35**, No 1 (2022), 119–133; doi:10.12732/ijamv35i1.9.

[12] S. Pashkunova, *Associative Investigations of Genetic Markers Determing Type of Dibetes and Chronic Diabetes-Related Complications*, PhD Thesis, Medical Academy of Bulgarian Army, Sofia, 2012, pp. 1–35 [in Bulgarian].

[13] I. Larson I, M.M. Hoffmann, J. M. Ordovas, E.J. Schaefer, W. Marz, J. Kreuzer, The lipoprotein lipase HindIII polymorphism: association with total cholesterol and LDL-choleterol, but not with HDL and triglycerides in 342 females, *Clin. Chem.*, **45** (1999), 963–968.