

Algebraic Approach to predict Diabetic Retinopathy: Evidence from Diabetes database

T. Subhramaniyan^{1*}, P. Sankar¹, P. Paulraj², M. G. Ragunathan³ and J. Jayanthi³

¹Department of Mathematics, Guru Nanak College, Chennai, India

²Department of Mathematics, A.M. Jain College, Chennai, India

³Department of Advanced Zoology and Biotechnology, Guru Nanak College, Chennai, India

*Corresponding Author: thangasubhramaniyan@gmail.com, Tel.: 9444896085

Available online at: www.isroset.org

Received:03/Oct/2018, Accepted:17/Oct/2018, Online: 31/Oct/2018

Abstract— The paper examines the factors that lead to retinopathy diabetes. A dataset of 1500 diabetic patients obtained from Madras Diabetes Research Foundation, Chennai was used in the study and 422 patients were found to be retinopathic. The study found that the family history and lipid profiles are insignificant factors causing retinopathy. The study reveals that smoking leads to retinopathy-diabetic rather than alcoholism.

Keywords— *Diabetics - Retinopathy - Smoking habit - Logit model.*

I. INTRODUCTION

Diabetes, a medical complication which is the result of lack of insulin production. There are several types of diabetes such as: Neuropathic diabetes; Nephropathic diabetes, Retinopathic diabetes, etc., Diabetic Retinopathy (DR) is the result of damage caused by diabetes to the small blood vessels located in the retina [1]. The damaged blood vessels in the retina due to diabetes can cause vision loss; the severity of retinopathy at large level affects working age group adults around the world [2-5]. In the global level, there are 37 million blind people due to the diabetic retinopathy. In India the proportion of type-2 diabetes is increasing at an alarming rate (WHO) [3]. Despite the significance of this problem, and the rising prevalence of diabetes particularly in Asian countries like India and China, there are few precise contemporary estimates of the worldwide prevalence of DR, particularly severe vision-threatening stages of the disease, including proliferative DR (PDR) and diabetic macular edema (DME). Despite the availability of many treatments, the problem remains the most common cause of blindness among people of 30 to 69 years of age in several countries. It is universally believed that the key to controlling diabetic retinal complications is to prevent them from happening. To achieve this target, it is necessary to identify the factors associated with it. Many patients are suffering from type-2 diabetes rather than type-1 diabetes. However, many studies have been conducted among type-1 diabetic patients, while studies targeting type-2 diabetic patients are relatively less. Moreover, many studies suggested that large differences may exist in DR onset among various populations. Earlier studies

were undertaken in western countries, whereas data regarding the incidence of DR in Indian, especially South Indian population is scarce. To the best of our knowledge, not even single study has been taken in the Algebraic Approach to predict Diabetic Retinopathy. During the hospital visit to collect the data, we had witnessed fully regressed retinopathy in several DR patients. As far as we know, no observational population study has reported the systemic factors associated with DR regression in a community population before. Although late stage DR can be corrected by photocoagulation, intraocular medicine or vitreoretinal surgery, these procedures is very expenses and mostly results in poor prognosis. By reviewing many studies the factors associated with DR, assessing and finding the early stage by available treatments procedures the progression of retinopathy is delayed but also to reverse the existing disease, thereby it is delaying the progression of late-stage retinopathy and achieving better prognoses. The results we report here are based on the study of 778 participants, who are taking five years treatment in Madras Diabetic Research Centre in Chennai. The aim of this study is to determine factors associated with diabetic retinopathy.

II. MATERIALS AND METHODS

The study has used the database collected and maintained by Madras Diabetes Research Foundation, Chennai. The collected information relating to diabetic patients gives details about individualized antihyperglycemia therapy, including metformin, sulfonylureas, meglitinides, glitazones and insulin injections, or behavioral therapy, such as

exercise, diet control and smoking cessation, being used to help control hyperglycemia. Patients' blood glucose levels were monitored by capillary blood glucose determination once a week, and their hemoglobin A1c (HbA1c) levels were measured once a year. Participants who had high blood pressure continued their antihypertensive therapy during follow-up using common drugs including diuretics, *b*-adrenoceptor blocking drugs, angiotensin- converting enzyme inhibitor (ACEI), angiotensin II receptor antagonist (ARB), and calcium channel blocking drugs. Participants with hyperlipemia or hypercholesterolemia took drugs such as statins to lower their lipid levels. As the clinical profile of DR suggests [22], most DR patients in this community with severe non-proliferative DR or proliferative retinopathy accepted photocoagulation, intraocular medicine injection or surgery to control DR progression. The database also provides information relating to anthropometric aspects, including age, gender, diabetic duration, diabetes onset age, occupation, education level and general ophthalmological medical history. The diabetic duration is the interval between the diagnosis of diabetes and the baseline examination. The patients' heights, weights, and systolic and diastolic blood pressures were measured by trained doctors. The blood pressure was measured in a supine position on the right arm using a standard mercury sphygmomanometer. Hypertension is defined as having a systolic blood pressure value 120 mmHg and/or a diastolic blood pressure value 80 mmHg. BMI was calculated as weight (kg)/height (m²). SPSS (version 25, IBM, USA) was used for all statistical analysis of the data. The baseline characteristics are presented as the means of 6 standard deviation for continuous variables and as rates (proportions) for categorical data. The data distribution was examined using the Kolmogorov- Smirnov Test. The categorical variables were compared using Chi-square test. From the univariate logistic analyses, variables with *p* values under 0.5 were considered for entry into multiple logistic regression analysis. Stepwise multiple logistic regression analysis was used to determine whether potential risk factors (including age, gender, diabetic duration, diabetes onset age, occupation, education level, BMI, baseline HbA1c, average HbA1c, high blood pressure, total cholesterol, serum creatinine and urinary microalbumin) Were associated with DR onset and regression. Statistical significance was defined as *p* (two tailed) at 5 percent level of significance. The study adopts Chi-Square Test to measure the association between the predictors. The test statistic for examining the independence is

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

It has an approximate $\chi^2_{(r-1)(c-1)}$ distribution where *r* is the number of rows and *c* is the number of columns. Under the test, it is hypothesized that the two variables are independent, the probability of a cell is the product of the probability of the column and the probability of row,

$n_{pl} = \frac{c_j}{n} \times \frac{r_i}{n} = \frac{c_j r_i}{n^2}$ Where *c_j* is the column total and *r_i* is the row total and *n* is the grand total or sample size. Then the expected value of the cell is

$n_{pl} = \frac{c_j r_i}{n} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$. Therefore

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^k \frac{(X_i - n_{pl})^2}{n_{pl}}$$

The number of cells *k* is the product of the number of columns and number of rows *k* = *rc*. The expected probabilities are found by estimating the row probabilities and the column probabilities. Since the row probabilities must sum to 1 and the column probabilities must sum to 1, there are (*r* - 1) and (*c* - 1) independent parameters that are being estimated as *m* = (*r* - 1) + (*c* - 1). Therefore, the degrees of freedom is

$$\begin{aligned} k - m - 1 &= rc - [(r - 1) + (c - 1)] - 1 \\ &= rc - r + 1 - c + 1 - 1 \\ &= (r - 1)(c - 1) \end{aligned}$$

Therefore, the Chi-square distribution with (*r* - 1)(*c* - 1) degrees of freedom can be used to find *p*-values using the Chi-square test for independence.

III. LOGISTIC REGRESSION

The study applies logistic regression to estimate the probability of existence of retinopathy in a sample.

The model of the study is

$$Y = \begin{cases} 1, & \text{if exposed to retinopathy} \\ 0, & \text{otherwise} \end{cases}$$

X(.) = Vectors of determining variables

Nature of Diabetics	Gender	
	Male	Female
Retinopathy	A	B
Non-Retinopathy	C	D

A hypothetical model of contingency table is illustrated above to understand the concept of Odd Ratio. The odd ratio can be computed by dividing cross-cells,

$$OR = \frac{AD}{BC}$$

If the estimated *OR* > 1 indicates risk factor and *OR* < 1 indicates protective factor.

The *logit* model: *p*(*x*) is the probability of the Retinodiagnostics for a given value of *x* and

$$\text{logit}(p(x)) = \log \frac{p(x)}{1 - p(x)} = \alpha + \beta$$

x = 0 (*unexposed*), *logit* (*p*(*x*)) = *logit* (*p*(0)) = $\alpha + \beta(0) = \alpha$

x = 1 (*exposed*), *logit* (*p*(*x*)) = *logit* (*p*(1)) = $\alpha + \beta(1) = \alpha + \beta$

$$\text{Odd Ratio} = \begin{cases} \frac{p(0)}{1-p(0)}, & \text{if unexposed} \\ \frac{p(1)}{1-p(1)}, & \text{if exposed} \end{cases}$$

Also,

$$OR = \frac{p(1)/(1 - p(1))}{p(0)/(1 - p(0))}$$

Now,

$$\begin{aligned} \text{and } \beta &= \text{logit } p(1) - \text{logit } p(0) \\ &= \log \left[\frac{p(1)/(1-p(1))}{p(0)/(1-p(0))} \right] \\ &= \log \left[\frac{p(1)/(1-p(1))}{p(0)/(1-p(0))} \right] \\ &= \log \text{ Odd Ratio} \end{aligned}$$

The regression coefficient in the estimated model is the *log* of *OR*, hence the *OR* is obtained by exponentiation of β , then

$$e^\beta = e^{\log(OR)} = OR$$

Results:

There were 422 diabetes patients with existing Retinopathy out of 1500 diabetes patients, and 72.75% ($n = 307$) of them were male. Their ages ranged from 40 years of age to 70 years of age, and 84.65% were more than 50 years of age. Their diabetes duration ranged from 1 year to 33 years, and 43.42% of them had diabetes for more than ten years. All participants had micro albuminuria values less than 30 mg/L. One hundred and five patients exhibited *DR* regression in both eyes, and another 5 patients had *DR* regression in one eye and no *DR* worsening of the other eye.

DR often occurred in patients with shorter diabetes duration, normal serum triglyceride levels, lower baseline HbA1c levels and lower average HbA1c levels (Table 2). From the univariate logistic regression, participant age, diabetes duration, serum creatinine, serum triglyceride, serum total cholesterol, baseline HbA1c and average HbA1c were entered into the multivariate logistic analysis framework (Methodology described in section-II).

Table 1: Association between Social variables and DR

Variable(s)	Indicators	Retinopathy	Non-Retinopathy	Chi-Square
Gender	Male	307	670	14.996***
	Female	115	408	
Paternal Diabetes	Yes	180	242	1.471 ^{NS}
	No	423	655	
Maternal Diabetes	Yes	151	271	0.888 ^{NS}
	No	414	664	
Sibling Retinopathy	Yes	234	188	9.776***
	No	501	577	
Smoking Habit	Yes	109	313	60.414***
	No	214	864	
Alcoholism	Yes	102	320	0.009 ^{NS}
	No	258	820	

Source: Computed. Note: *** $p < 0.01$; NS-Insignificant

Based on the methodology described in section-II, the chi-square test results are presented in the Table1. Diabetic retinopathy is one of the most worrisome complications. The concept of diabetic retinopathy has been studied by several authors. The association between LDL cholesterol and severity of *DR* has been studied by T.L. Dorman et al. [10]. In this study, the variables such as OCT values, age, duration of diabetics, level of hyperglycemia (HbA1c), family

history, hypertension, LDL play significant role in the occurrence of OCT. The findings from chi-square analysis reveals that there is no significant association that exists between retinopathy and hereditary. The computed chi-square value of paternal and maternal diabetes is not significantly related with Retinopathy patients. However, the inborn do have significant association with their Retinopathy parents. A significant finding highlighted in the study is that alcohol habit does not induce retinopathy, while smoking habit is a severe cause of retinopathy.

V.RESULTS FROM LOGISTIC REGRESSION

The study has applied logistic regression analysis to estimate the likelihood to become a retinopathic diabetes among the diabetic patients. The model fit statistics shows that the full model is significant (*Chi-Square* = 157.112, $p < 0.05$).

Table 2: Model fit Statistics

Model	Chi-Square	Df	Sig.
Step	157.112	25	0.000
Block	157.112	25	0.000
Model	157.112	25	0.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1463.590*	.509	.557

* Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	13.524	8	.000

Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; and many people have been trying to come up with one. There are a wide variety of pseudo-R-squared statistics. Because this statistics does not mean what R-squared means in OLS regression (the proportion of variance explained by the predictors), therefore, it should be carefully interpreted. The H-L Test also shows that the model is significant.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for Exp(B)		
							Lower	Upper	
Step 1*	Age	.015	.008	3.956	1	.047	1.015	1.000	1.030
	Gender	.635	.172	13.595	1	.000	1.888	1.347	2.646
	BMI	.060	.017	12.610	1	.000	.942	.911	.974
	Sys BP	.017	.005	13.922	1	.000	1.017	1.008	1.026
	Dias BP	.022	.009	5.830	1	.016	.979	.962	.996
	Paternal	.261	.136	3.722	1	.054	1.299	.996	1.694
	Maternal	.108	.140	.590	1	.443	.898	.682	1.182
	Sibling	.343	.137	6.291	1	.012	1.409	1.078	1.842
	Smoking	.317	.180	3.092	1	.079	1.373	.964	1.954
	Alcohol	-.325	.182	3.181	1	.074	.723	.506	1.033
	FBS	.001	.002	.648	1	.421	.999	.995	1.002
	PPB	.000	.001	.111	1	.739	1.000	.998	1.003

HBA	.230	.046	24.661	1	.000	1.259	1.149	1.378
CHO	.017	.010	2.892	1	.089	.984	.965	1.003
TRI	.000	.002	.005	1	.944	1.000	.996	1.004
HDL	.011	.012	.752	1	.386	1.011	.986	1.036
LDL	.011	.010	1.328	1	.249	1.011	.992	1.031
Max_per	.131	.057	5.248	1	.022	1.140	1.019	1.276
Max_alk	.000	.001	.141	1	.707	1.000	.999	1.002
Max_sgot	.002	.010	.040	1	.841	1.002	.982	1.023
Max_sgpt	-	.008	1.234	1	.267	.992	.977	1.006
Max_wbc	.000	.000	1.712	1	.191	1.000	1.000	1.000
Max_pc	.000	.000	.108	1	.742	1.000	1.000	1.000
Max_sod	-	.023	1.689	1	.194	.970	.927	1.015
Max_pot	.325	.172	3.558	1	.059	1.385	.987	1.942
Constant	.350	3.537	.010	1	.921	1.419		

* Variable(s) entered on step 1: age, Gender, BMI, Systolic BP, Diasystolic BP, father, mother, sibling, Smoking, Alcohol, Max_fbs, Max_ppb, Max_hba, Max_cho, Max_tri, Max_hdl, Max_ldl, Max_pcr, max_alk, max_sgot, max_sgpt, max_wbc, max_pc, max_sod, max_pot.

Table 3 reports the result of logit analysis. For the study, anthropometric information, family history, habits, cholesterol level, Kidney Function Test (KFT), Liver Function Test (LFT) values were taken into account. The study reports the beta values of logistic regression equation for predicting the dependent variable from the independent variable. They are in log-odds units. Similar to OLS regression, the prediction equation is

$$\log = \beta_0 + \beta_1 x_1(\text{age}) + \beta_2 x_2(\text{gender}) + \beta_3 x_3(\text{bmi}) + \beta_4 x_4(\text{diasbp}) + \beta_5 x_5(\text{fatherbp}) + \beta_6 x_6(\text{motherbp}) + \beta_7 x_7(\text{siblingbp}) + \beta_8 x_8(\text{smoking}) + \beta_9 x_9(\text{alcohol}) + \beta_{10} x_{10}(\text{fbs}) + \beta_{11} x_{11}(\text{ppb}) + \beta_{12} x_{12}(\text{hba}) + \beta_{13} x_{13}(\text{cho}) + \beta_{14} x_{14}(\text{tri}) + \beta_{15} x_{15}(\text{hdl}) + \beta_{16} x_{16}(\text{ldl}) + \beta_{17} x_{17}(\text{pcr}) + \beta_{18} x_{18}(\text{alk}) + \beta_{19} x_{19}(\text{sgot}) + \beta_{20} x_{20}(\text{sgpt}) + \beta_{21} x_{21}(\text{wbc}) + \beta_{22} x_{22}(\text{pc}) + \beta_{23} x_{23}(\text{sod}) + \beta_{24} x_{24}(\text{pot}) + \beta_{25} x_{25}(\text{sys_bp}) + \text{error}$$

The se estimates tell you about the relationship between the independent variables and the dependent variable, where the dependent variable is on the *logit* scale. These estimates give the extent of increase (or decrease, if the sign of the coefficient is negative) in the predicted log odds of age = one that would be predicted with one year increase (or decrease) in the predictor, holding all other predictors constant. Because these coefficients are in log-odds units, they are often difficult to be interpreted. Thus they are often converted into odds ratios. If age increases by a year, we expect a 0.015 increase in the log-odds of retinopathy, holding all other independent variables constant. Compared to female, male diabetic patients do have 0.635 expected probability to be a retinopathic. The estimated coefficient of BMI is positively associated with retinopathy. When BMI value is increased by one unit, expected probability for being retinopathy is 0.060. The blood pressure level both Systolic and Diastolic have significant effect on retinopathy. Family background: Father or Mother of a diabetic patient is supposed to be diabetic, then there is high likelihood to be diabetic. But, whether there is a likelihood to be a

retinopathic diabetes. In the study, there is high likelihood to have retinopathy when the patients' father has diabetics. Personal Habits: If a respondent has smoking and alcohol habits, there is a chance to be diabetic. An interesting finding is that if a patient has smoking habits, there is high likelihood to be retinopathic whereas not so with alcohol consumption. Lipids level: The lipids considered in the study are Serum Cholesterol (CHO), Triglycerides (TRI), HDL and LDL. These parameters do not influence retinopathy as expected. The kidney function test and liver function test parameters values have also not induced retinopathy except potassium level which is a significant factor in diabetes.

VI.CONCLUSION

Diabetes is an incurable but controllable disease. Diabetic retinopathy is a vascular disorder affecting the microvasculature of the retina. It is estimated that diabetes mellitus affects more than 6 per cent of the world's population, almost half of whom have some degree of retinopathy. In India with the epidemic increase in type-2 diabetes mellitus as reported by the World Health Organization (WHO), diabetic retinopathy is fast becoming an important cause of visual disability. Visual disability due to diabetes is a significant public health problem; however this morbidity is largely preventable and treatable, if managed with timely intervention, the quality of life can be preserved. In India, out of 31.7 million diabetic patients, the ratio of retinopathic diabetics is 5.6 million. This underscores the need for routine retinal screening of diabetic individuals annually to detect DR and prevent visual impairment. In addition, optimized control of systemic considerations, which affect onset and/or progression of Retinopathy, through an intensive, multidisciplinary, healthcare team-based approach, can markedly reduce impairment of vision due to DR. This study reports that smoking habit is a significant factor in retinopathy along with age factor. On the other hand, there is no significant effect of lipid profiles and family history on the progression of diabetics retinopathy.

REFERENCES

- [1] Rema, M and R.Pradeepa, (2006). Diabetic Retinopathy: An Indian Perspective. Indian Journal of Medical Research, 125, 297-310.
- [2] Eye Diseases Prevalence Research Group. Causes and prevalence of visual impairment among adults in the United States. Arch Ophthalmol 2004; 122: 477- 485.
- [3] Eye Diseases Prevalence Research Group. The prevalence of diabetic retinopathy among adults in the United States. Arch Ophthalmol 2004; 122:552- 563.
- [4] Wild S, Roglic G and Green A, Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004; 27:1047-1053.
- [5] Ramachandran A, Jali MV and Mohan V, High prevalence of diabetes in an Urban population in South India. BMJ 1988; 297:587-590.

- [6] Agarwal S, Raman R and Paul PG, Sankara Nethralaya- Diabetic Retinopathy Epidemiology and Molecular Genetic Study (SN-DREAMS 1): study design and research methodology. *Ophthalmic Epidemiol* 2005; 12:143–153.
- [7] Oakes JM and Rossi PH, The measurement of SES in health research: current practice and steps towards a new approach. *Soc Sci Med* 2003; 56:769–784.
- [8] Sharma T, How necessary is the second estimation of fasting plasma glucose level by laboratory venous blood to diagnose type 2 diabetes, particularly in epidemiological studies [letter] *Ophthalmic Epidemiol* 2006;13:281–282.
- [9] Raman Kutty V, Joseph A and Soman CR, High prevalence of type 2 diabetes in an urban settlement in Kerala, India. *Ethn Health* 1999; 4:231–239.
- [10] Gupta A, Gupta R and Sarna M, Prevalence of diabetes, impaired fasting glucose and insulin resistance syndrome in an urban Indian population. *Diabetes Res Clin Pract* 2003; 61:69–76.
- [11] Zargar AH, Khan AK and Masoodi SR, Prevalence of type 2 diabetes mellitus and impaired glucose tolerance in the Kashmir Valley of the Indian subcontinent. *Diabetes Res Clin Pract* 2000; 47: 135– 146.
- [12] Ramachandran A, Snehalatha C and Kapur A, Diabetes Epidemiology Study Group in India (DESI). High prevalence of diabetes and impaired glucose tolerance in India: National Urban Diabetes Survey. *Diabetologia* 2001; 44: 1094–1101.
- [13] Mohan V, Deepa M and Deepa R, Secular trends in the prevalence of diabetes and glucose tolerance in urban South India—the Chennai Urban Rural Epidemiology Study (CURES-17). *Diabetologia* 2006; 49:1175– 1188.
- [14] Kuzuya T, Akanuma Y, Akazawa Y and Uehata T, Prevalence of chronic complications in Japanese diabetic patients. *Diabetes Res Clin Pract* 1994; 24 (suppl):S159–164.
- [15] Dandona L, Dandona R and Naduvilath TJ, Population based assessment of diabetic retinopathy in an urban population in southern India. *Br J Ophthalmol* 1999; 83:937– 940.
- [16] Cugati S, Kifley A, Mitchell P and Wang JJ, Temporal trends in the age-specific prevalence of diabetes and diabetic retinopathy in older persons: population-based survey findings. *Diabetes Res Clin Pract* 2006; 74:301– 308.
- [17] Kumaramanickavel G, Ramprasad VL and Sripriya S, Association of Gly82Ser polymorphism in the RAGE gene with diabetic retinopathy in type II diabetic Asian Indian patients. *J Diabetes Complications* 2002; 16:391– 394.
- [18] Van Leiden HA, Dekker JM and Moll AC, Blood pressure, lipids, and obesity are associated with retinopathy: the Hoorn Study. *Diabetes Care* 2002; 25:1320–1325.
- [19] Perusicova J, Prevalence of dyslipidemia, hypertension and vascular complications in newly diagnosed diabetics (prospective study: part 2) [in Czech]. *Vnitr Lek* 2001; 47:146–150.

AUTHORS PROFILE

T. Subhramanian, Associate Professor & Head, Department of Mathematics, Guru Nanak College, Chennai has an in depth knowledge in Mathematics and a well experienced teacher with 30 years of experience. His field of specialization is Algebra and Graph Theory.

