International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 3, Issue 4(April 2014), PP.50-58

Estimation of Survival Function and Probability of Onset of Diabetic Nephropathy using Path Analysis and Analysis of Covariance

Gurprit Grover[1], Alka Sabharwal[2*], Juhi Mittal[1]

[1]*Department of Statistics, University of Delhi, Delhi, 110007, India* [2]*Department of Statistics, Kirori Mal College, University of Delhi, Delhi, 110007, India*

Abstract:- The objective of this paper is to estimate the probability of onset of diabetic nephropathy (DN) arising out of type 2 diabetic patients. To achieve our goal, firstly we have analyzed the association of occurrence of renal complication with the available risk factors by applying path analysis viz., is an extension of multiple regression model that helps in better prediction, since it can model the impact of each factor on an outcome variable. The results of the path analysis showed that serum creatinine (SrCr), number of successes ((NOS), i.e. SrCr $\geq$ 1.4mg/dl) and age at onset of diabetes are the most appropriate predictors for estimating the progression of DN. Thus, using these significant predictors we have estimated the probability of onset of DN for two different levels of age at onset of diabetes diagnosis (Age at onset of diabetes $\leq$ 40 years and Age at onset of diabetes > 40 years) by applying ANCOVA with two covariates SrCr and NOS. The results of the analysis indicates that the probability of the onset of DN differ among the two levels of age at onset of diabetes. And SrCr and number of successes are the significant contributors for predicting probability.

Keywords:- Analysis of covariance, diabetic nephropathy, path analysis, serum creatinine, Sobel test

I. INTRODUCTION

The incidence of diabetes mellitus (DM) and diabetic nephropathy (DN) have risen rapidly in the past few decades and have become an economic burden to the healthcare system worldwide. DN, also known as diabetic glomerulosclerosis or diabetic kidney disease, is a major complication of DM and is considered to be an irreversible and progressive disease [1]. Approximately 40% of people with type 2 diabetes develop nephropathy and it is a leading cause of end-stage renal disease (ESRD). Due to the growing disease burden of diabetes and its complications, and due to the fact that early detection has been associated with improved outcomes, it is important to identify DN predictors, in order to facilitate its diagnosis and treatment [2]. With the development of kidney complications, Glomerular Filtration Rate (GFR) starts to fall and SrCr level starts to increase. Various studies have shown the importance of measurement of Estimated Glomerular Filtration Rate (eGFR) and SrCr level for predicting the development of DN. The rate of rise in SrCr, a well accepted marker for the progression of DN, (creatinine value 1.4 mg/dl to 3.0 mg/dl) is the indicator for impaired renal function [3] and the normal level of creatinine is 0.8 mg/dl to 1.4 mg/dl [4].

Multiple regression analysis deals with the model development of relationship between two or more explanatory variables and a response variable by fitting a linear equation to the observed data. There are times when we like to see the effects of predictor variables on a number of dependent variables or outcome variables. Path analysis is a highly flexible and comprehensive methodology used in such situation. This methodology is appropriate for investigating achievement, economic trends, medical health issues, family and peer dynamics, self-concept, exercise, self-efficacy, depression, psychotherapy, and other phenomenon. Path analysis is an extension of multiple regression analysis. It is a multivariate technique specifying relationships between observed (measured) variables. Variables in path analysis could be independent and dependent whereas, variables in regression analysis are either independent or dependent. Multiple, related equations are solved simultaneously to determine parameter estimates. It is a decision support tool that helps researchers to determine the contribution of each variable to the dependent variable and each variable via other variables to that dependent variable. Also path analysis model is not a substitute of regression analysis; rather it is a complementary methodology to regression analysis. Because of this complexity, it is always accompanied by a path diagram which helps in displaying all causal relationships between variables under study [5-6].

Mosteller applied path analysis to find the relationship between genetic, environmental, and age-related components for cardiovascular diseases in adult's female twin pair's population [7]. Wheelwright , Birchall , Boaden, Pearce , Lennon also applied path analysis to the process of care for patients with head and neck cancer [8]. Tae, Heitkemper, Kim applied the same to test a model of depression in Korean women with breast cancer and to test the mediating effects of self-esteem and hope by considering the variables depression, self-

esteem, hope, perceived health status, religious beliefs, family support, economic status, and fatigue [9]. Path analysis helps in selecting the most appropriate variables from a set of variables for estimating a particular response variable. The target response variable can be estimated by applying any statistical techniques used in medical field. Analysis of variance (ANOVA) and Analysis of covariance (ANCOVA) are one of the most frequently used techniques in medical data analysis.

Analysis of covariance; a general linear model which blends ANOVA and regression; evaluates whether population means of a dependent variable are equal across levels of a categorical independent variable, in the presence of some continuous variables known as covariates. Ludman et. al. applied ANCOVA to determine if the overall number of diabetes symptoms was related to having major depression with covariates such as age, gender, marital status, education, racial ethnicity, clinic, HbA1c, number of diabetes complications, medical co-morbidity, duration of diabetes in years, type 1 vs. type 2 diabetes and diabetes treatment intensity [10]. Shrivanova applied ANCOVA analysis for detecting the effect of co-factors (such as age, BMI, sex, and the presence of co-morbidities, including rheumatic diseases and invalidity) on patients' quality of life scores with chronic myeloid leukemia [11].

The objective of the study is to estimate the probability of onset of DN in type 2 diabetic patients. To achieve our goal we have first analyzed the occurrence of renal complication in type 2 diabetic patients with the risk factors such as Fasting Blood Glucose (FBG), Diastolic Blood Pressure (DBP), Low Density Lipoprotein (LDL), age at onset of diabetes, Serum Creatinine (SrCr) and number of successes ((NOS), defined as the number of times SrCr value was recorded greater than 1.4 mg/dl out of total number of times test is recommended for a particular patient). The relation between these variables is analyzed by conducting path analysis, under which we have considered two endogenous variables namely: survival function of DN patient (survival is defined as the occurrence of renal complication in a type 2 diabetic patient) and SrCr values and five exogenous variables namely: FBG, DBP, LDL, age at onset of diabetes and NOS. As mentioned earlier path analysis studies the relationship between a set of variables and hence helps in indicating the most appropriate predictors to a estimate particular dependent variables. The results of the path analysis showed that SrCr, NOS and age at onset of diabetes are the most appropriate predictors for estimating the progression of DN. Thus using these three variables we have estimated the probability of onset of DN for two different levels of age at onset of diabetes diagnosis (Age at onset of diabetes $\leq$ 40 years and Age at onset of diabetes > 40 years) by applying ANCOVA with two covariates SrCr and NOS.

To the best of our knowledge this is the first investigation about the estimation of probability of onset of DN in diabetic patients using path analysis and ANCOVA. This study would help the medical fraternity to guide the patients about the development of renal complication on the basis of the limited information available in terms of their age at onset of diabetes and SrCr values. The remainder of this paper is organized as follows. In section 2 development of model is discussed. Section 3 describes the data used and section 4 applies the model to the data of type 2 diabetic patients. And some concluding remarks are made under section 5.

II. METHODOLOGY

2.1 Path Analysis

Path analysis is an extension of multiple regression analysis and is very easy to understand. They rely very heavily on pictures called path diagrams. These path diagrams help in studying the relation between a set of variables. All the variables are represented by rectangles, and each path is represented by a straight line with an arrow head at one end. In the path analysis, the variables can be an exogenous or an endogenous variable. An exogenous variable has paths coming from it and none leading to it while an endogenous variable has at least one path leading to it. A standard statistical method in the path analysis literature is to treat the variables as having a normal distribution and to estimate paths using several least squares regression equations. Standardized regression coefficients are taken as the estimated path coefficients. The standardized regression coefficients allow researchers to compare the relative magnitude of the effects of different explanatory variables in the path model by adjusting the standard deviations such that all the variables, despite different unit of measurements, have equal standard deviations. These standardized path coefficients measure the relative strength and sign of the effect from a causal variable to an endogenous or outcome variable in the model. Indirect paths coefficients are a product of the direct path from the exogenous variable to the mediating variable and the direct path of the mediating variable to the dependent variable.

Consider a simple path analysis model with three variables Y , X_1 & X_2 . Where, Y & X_2 are the endogenous variables and X_1 is an exogenous variable. The relation between these three variables can be explained by the path diagram presented in figure 1:

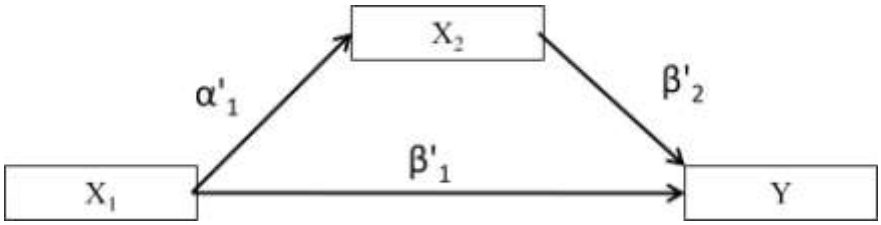

Figure 1. Path diagram with three variables

For conducting the path analysis the following two regression equations are fitted by applying the method of least square:

$$X_2 = \alpha_0 + \alpha_1 X_1 + \in \quad (1)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \in' \quad (2)$$

Where, α_0 & α_1 unstandardized regression coefficients for model (1) and β_0 , β_1 & β_3 are the unstandardized regression coefficients for model (2). And $\in_i$ & $\in'_i$ are independent random error terms assumed to be normally distributed with zero means and constant variances. The direct effect between the variables X_1 & X_2 is given by the standardized regression coefficient α'_1 obtained by fitting equation (1). And similarly, the direct effect of X_1 on Y and X_2 on Y is given by the standardized regression coefficients β'_1 and β'_2 respectively. The indirect effect of X_1 through X_2 on Y is given by the product $\alpha'_1 \times \beta'_2$.

The significance of direct effects is tested by applying the student's t test. And, the significance of indirect effect is tested by applying Sobel test. The null hypothesis under this test is $H_0 : \alpha_1 . \beta_2 = 0$. The test statistics is:

$$Z = \frac{\alpha_1 . \beta_2}{\sqrt{\alpha^2_1 . S^2_{\beta_2} + \beta^2_2 . S^2_{\alpha_1}}} \quad (3)$$

Where, S_{α_1} and S_{β_2} are the standard errors of unstandardized regression coefficients α_1 and β_2 respectively and Z follows standard normal distribution [12].

2.2 Analysis of Covariance

Analysis of Covariance is a technique that combines analysis of variance and regression. The idea behind ANCOVA is to include a quantitative variable x (called a covariate or concomitant variable) that is linearly related to the response variable. Furthermore, x cannot be controlled by the experimenter but can be observed along with y. The inclusion of the covariates is designed to reduce the mean square error (MSE), making the analysis more precise. An ANCOVA will be superior to its ANOVA counterpart in two distinct respects (i.e., increased statistical power and control), so long as a good covariate is used. It in fact can be viewed as a special type of multiple regressions.

The appropriate statistical model for one way ANCOVA with `ytwo covariates can be defined as follows,

$$y_{ij} = \mu + \tau_i + \beta_1 (x_{ij} - \bar{x}_{..}) + \beta_2 (z_{ij} - \bar{z}_{..}) + \in_{ij} ; \; i = 1, 2, ..., v \; \& \; j = 1, 2, ...r_i \quad (4)$$

Where , y_{ij} is jth observation of response variable under ith classification of the fixed factor, τ_i is the fixed effect due to ith classification of the fixed factor, x_{ij} and z_{ij} are jth value of the covariates under ith classification of fixed factor, $\bar{x}_{..}$ and $\bar{z}_{..}$ are the mean of x_{ij} and z_{ij} values respectively and $\in_{ij}$ are the random error terms assumed to be normally distributed with mean zero and constant variance. β_1 *and* β_2 are regression coefficients corresponding to x_{ij} and z_{ij} respectively. ANCOVA model assumes a linear relationship between the covariates and the mean response, with the same slopes for each classification. Applying the least square method, the normal equations can be written as,

$$\frac{\delta SSE}{\delta \hat{\mu}} = -2 \sum_{i=1}^{v} \sum_{j=1}^{r_i} \left(y_{ij} - \hat{\mu} - \hat{\tau}_i - \beta_1 (x_{ij} - \bar{x}_{..}) + \beta_2 (z_{ij} - \bar{z}_{..}) \right) = 0$$

$$\frac{\delta SSE}{\delta \hat{\tau}_i} = -2\sum_{j=1}^{r_i}\left(y_{ij} - \hat{\mu} - \hat{\tau}_i - \beta_1\left(x_{ij} - \bar{x}_{..}\right) + \beta_2\left(z_{ij} - \bar{z}_{..}\right)\right) = 0$$

$$\frac{\delta SSE}{\delta \beta_1} = -2\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \hat{\mu} - \hat{\tau}_i - \beta_1\left(x_{ij} - \bar{x}_{..}\right) + \beta_2\left(z_{ij} - \bar{z}_{..}\right)\right)\left(x_{ij} - \bar{x}_{..}\right) = 0$$

$$\frac{\delta SSE}{\delta \beta_2} = -2\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \hat{\mu} - \hat{\tau}_i - \beta_1\left(x_{ij} - \bar{x}_{..}\right) + \beta_2\left(z_{ij} - \bar{z}_{..}\right)\right)\left(z_{ij} - \bar{z}_{..}\right) = 0$$

Solving the (v+3) normal equations simultaneously the estimates of μ, τ_i, β_1 and β_2 under the restriction $\sum_{i=1}^{v}\tau_i = 0$ are:

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\tau}_i = \left(\bar{y}_{i.} - \bar{y}_{..}\right) - \hat{\beta}_1\left(\bar{x}_{i.} - \bar{x}_{..}\right) - \hat{\beta}_2\left(\bar{z}_{i.} - \bar{z}_{..}\right)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{i.}\right)\left(x_{ij} - \bar{x}_{i.}\right) - \beta_2\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(x_{ij} - \bar{x}_{i.}\right)\left(z_{ij} - \bar{z}_{i.}\right)}{\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(x_{ij} - \bar{x}_{i.}\right)^2}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{i.}\right)\left(z_{ij} - \bar{z}_{i.}\right) - \beta_1\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(x_{ij} - \bar{x}_{i.}\right)\left(z_{ij} - \bar{z}_{i.}\right)}{\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(z_{ij} - \bar{z}_{i.}\right)^2}$$

To test the null hypothesis: $H_0 : \tau_i = 0 \ \forall \ i = 1, 2$, the F statistics is defined as follows,

$$F = \frac{\left(SSE' - SSE\right)/(v-1)}{\left(SSE\right)/(n-v-2)} \sim F_{[v-1,\, n-v-2]}$$

Where,

$$SSE = \sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{i.}\right)^2 - \beta_1\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{i.}\right)\left(x_{ij} - \bar{x}_{i.}\right) - \beta_2\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{i.}\right)\left(z_{ij} - \bar{z}_{i.}\right)$$

$$SSE' = \sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{..}\right)^2 - \beta_1\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{..}\right)\left(x_{ij} - \bar{x}_{..}\right) - \beta_2\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{..}\right)\left(z_{ij} - \bar{z}_{..}\right)$$

Further, whether a particular covariate contributes significantly to the model can be tested by defining the following F statistics. For the null hypotheses $H_{0X} : \beta_1 = 0$ and $H_{0Z} : \beta_2 = 0$ the statistics are

$$F_X = \frac{\beta_1\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{i.}\right)\left(x_{ij} - \bar{x}_{i.}\right)}{\left(SSE\right)/(n-v-2)} \sim F_{[1,\, n-v-2]} \text{ and}$$

$$F_Z = \frac{\beta_2\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{i.}\right)\left(z_{ij} - \bar{z}_{i.}\right)}{\left(SSE\right)/(n-v-2)} \sim F_{[1,\, n-v-2]}$$

respectively [13-14] .

III. DATA

The methods discussed in section 2 are applied to the data obtained through house to house survey of type 2 diabetic patients who were referred for pathological tests to Dr. Lal path lab, Delhi, India. Retrospective study has been conducted on the collected data. The data regarding the duration of diabetes and other factors like age at which diabetes was diagnosed; FBG, DBP, SBP, LDL and values of SrCr were recorded for each patient. Using results of Grover, Gadpayle and Sabharwal [18] we have also considered an additional factor namely number of successes, which measures the number of time SrCr value was recorded greater than 1.4 mg/dl out of total number of times test is recommended for a particular patient, as a factor for predicting the progression of DN in type 2 diabetic patients.

IV. APPLICATION

4.1 Path Analysis

For conducting the path analysis we have considered seven variables namely: survival function of DN patient (($S_{DN}(t)$), survival is defined as the occurrence of renal complication in a type 2 diabetic patient), SrCr, FBG, DBP, LDL, age at onset of diabetes and NOS. The variable $S_{DN}(t)$ i.e survival function is first estimated by applying Kaplan Meier method. Then according to the causal ordering of variables, we may divide these variables into the following two groups:

Exogenous Variable	**FBG, DBP, LDL, Age at onset of diabetes, NOS**
Endogenous Variable	SrCr, $S_{DN}(t)$

We wish to estimate the direct and indirect effect of the exogenous on endogenous variables. The hypothesized relation between these variable can also be depicted by the following path diagram:

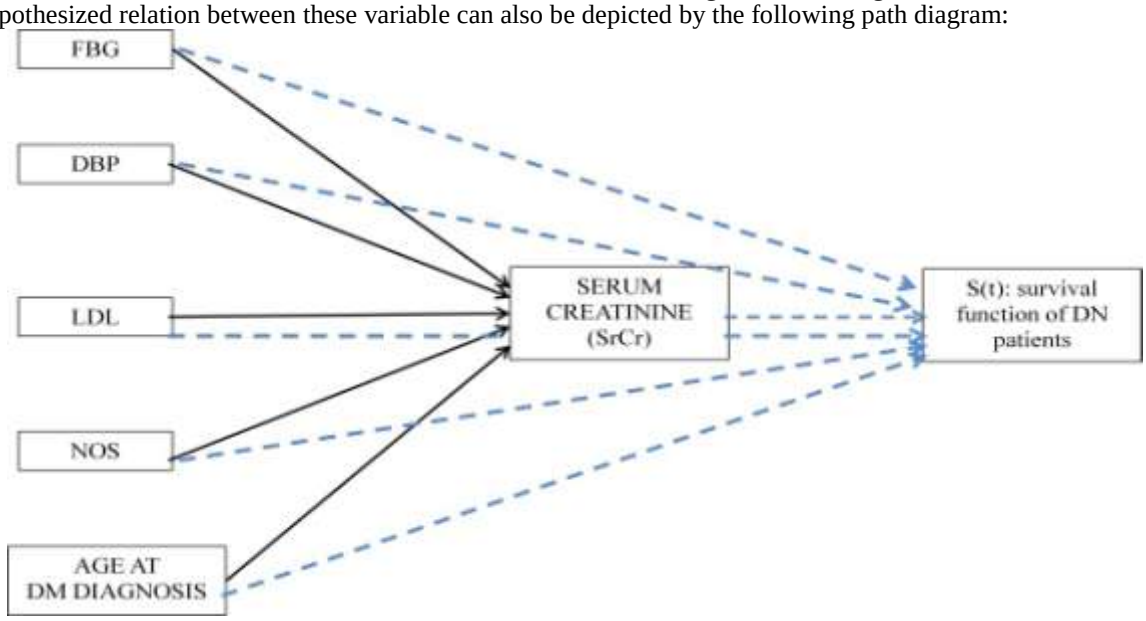

Figure 2. Path diagram is representing a proposed model for estimating the occurrence of nephropathy in type 2 diabetic patients.

The path analysis involves the fitting of two regression models presented in the following equations:

$$(SrCr)_i = \beta_0 + \beta_1 (FBG)_i + \beta_2 (DBP)_i + \beta_3 (LDL)_i + \beta_4 (Ageatonset)_i + \beta_5 (NOS)_i + \in'_i \quad (5)$$

$$(S_{DN}(t))_i = \alpha_0 + \alpha_1 (FBG)_i + \alpha_2 (DBP)_i + \alpha_3 (LDL)_i + \alpha_4 (Ageatonset)_i + \alpha_5 (NOS)_i + \alpha_6 (SrCr)_i + \in_i \quad (6)$$

Where, $\beta_0, \beta_1, ..., \beta_5$ are the regression coefficients for the first model and $\alpha_0, \alpha_1, ...\alpha_6$ are the regression coefficients for the second model and. And $\in'_i$ & $\in_i$ are independent random error terms assumed to be normally distributed with zero means and constant variances.

Table 1. Path analysis results of model with SrCr as endogenous variable.

Model

$$(SrCr)_i = \beta_0 + \beta_1 (FBG)_i + \beta_2 (DBP)_i + \beta_3 (LDL)_i + \beta_4 (Ageatonset)_i + \beta_5 (NOS)_i + \epsilon'_i$$

Dependent Variable : **Serum Creatinine (SrCr)**

Variable	**Standardized Parameter estimates**	**t for H_0:β_i=0**	**P-value**
FBG	0.213	3.673	<0.001
DBP	0.228	4.602	<0.001
LDL	0.075	1.699	0.092
Age at onset	-0.001	-0.041	0.968
NOS	0.530	9.729	<0.001

Table 2. Path Analysis results of model with survival function as endogenous variable.

Model

$$(S_{DN}(t))_i = \alpha_0 + \alpha_1 (FBG)_i + \alpha_2 (DBP)_i + \alpha_3 (LDL)_i + \alpha_4 (Ageatonset)_i + \alpha_5 (NOS)_i + \alpha_6 (SrCr)_i + \epsilon_i$$

Dependent Variable : Survival function of DN

Variable	**Standardized Parameter estimates**	**t for H_0:α_i=0**	**P-value**
FBG	**0.053**	**0.316**	**0.753**
DBP	**0.251**	**1.390**	**0.171**
LDL	**0.180**	**1.132**	**0.263**
Age at onset	**-0.225**	**-2.186**	**0.021**
NOS	**-0.712**	**-3.673**	**0.001**
Sr Cr	**0.560**	**2.496**	**0.016**

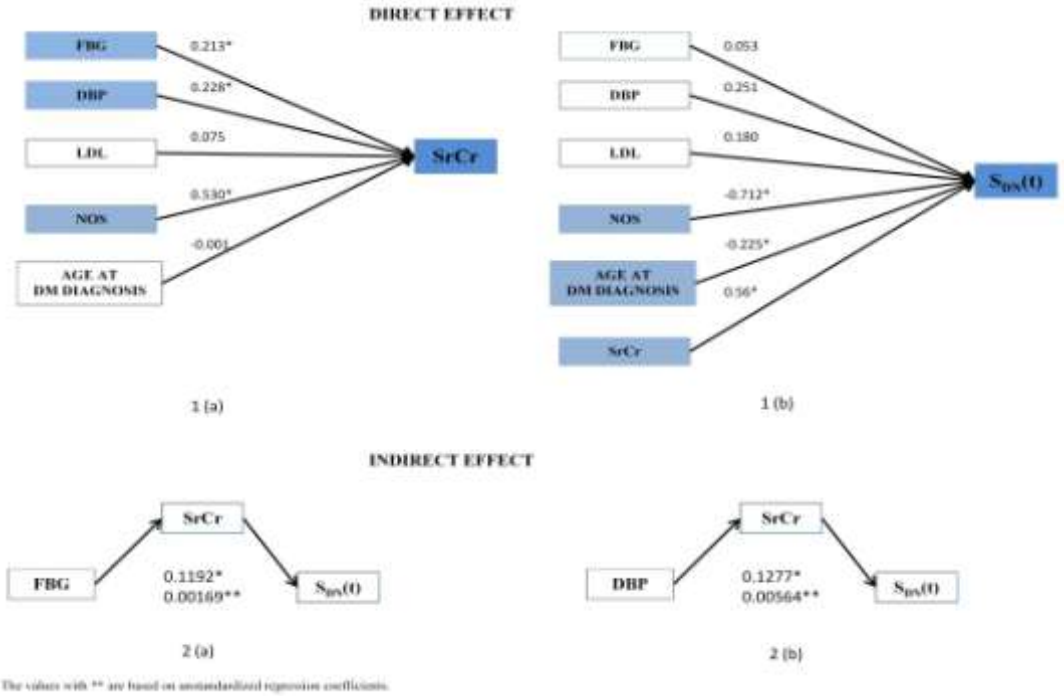

Figure3: Results of path analysis with direct and indirect effects.

The results of the path analysis presented in table 1 and 2 and in figure 3 can be summarized as follows. The variable NOS has significant direct effects on both SrCr and $S_{DN}(t)$. FBG and DBP have significant direct effects on SrCr. Age at onset of diabetes has only direct effect on $S_{DN}(t)$. Also, SrCr influences the survival function of DN patients. There were two significant indirect effect in our model: FBG on $S_{DN}(t)$ and DBG on $S_{DN}(t)$ through SrCr. These effects represents the mediating effect, for example the indirect effect of FBG on $S_{DN}(t)$ represents the effect of FBG on $S_{DN}(t)$ through the mediating variable SrCr. This effect is obtained by multiplying the standardized regression coefficients corresponding to FBG in model (1) and SrCr in model (2) and is found to be 0.1192. Similarly, indirect effect of DBP on $S_{DN}(t)$ through SrCr is found to be 0.1277. There statistical significance is tested by applying Sobel test which showed that, both the effects are statistically significant (Z_{FBG}= 2.03 and Z_{DBP}= 2.35). It was also observed that LDL is insignificant for both SrCr and $S_{DN}(t)$. Hence, the reduced path model can be presented by figure 4.

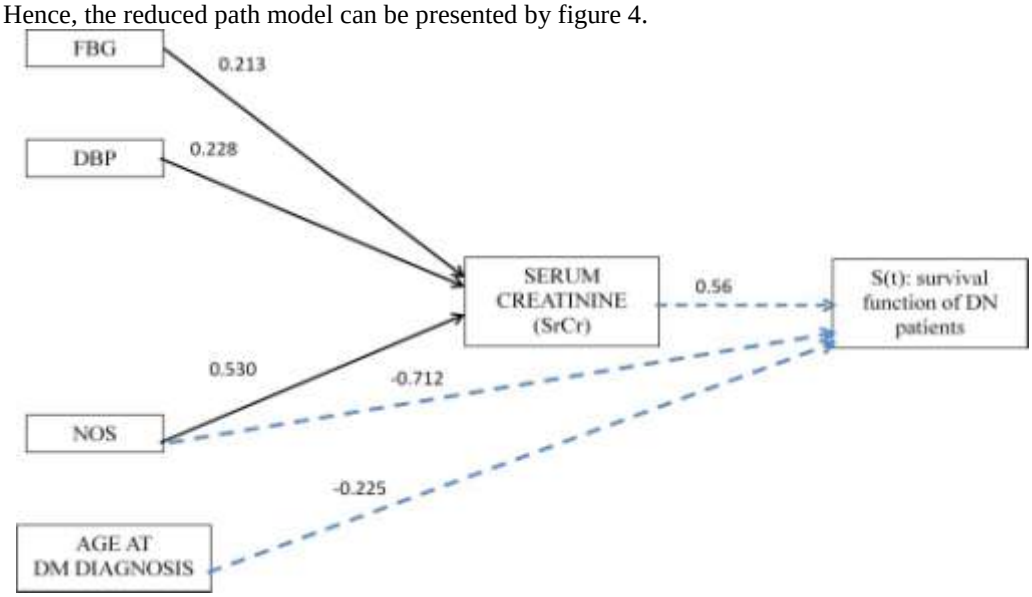

Figure 4: Reduced Path Model for estimating the occurrence of nephropathy in type 2 diabetic patients

4.2 Analysis of Covariance

The main objective of the study is to estimate the probability of occurrence of DN by applying one way ANCOVA with two covariates. The probabilities are estimated on the basis of one independent variable namely age at onset of diabetes (levels included: age at onset of diabetes ≤ 40 and age at onset of diabetes >40). And two covariates namely SrCr and number of successes are considered, as they were found to be most significant factors from path analysis. Also, the data for the observed probabilities are taken from the paper of Grover, Gadpayle and Sabharwal [18]. The ANCOVA model for our study can be defined as,

$$P[Onset\ of\ DN]_{ij} = p_{ij} = \mu + (Ageatonset)_i + \beta_1 \left(SrCr_{ij} - SrCr_{Mean}\right) + \beta_2 \left(NOS_{ij} - NOS_{Mean}\right) + \in_{ij} \quad ;i = 1,2 \& j = 1,2,....r_i \tag{7}$$

Where , p_{ij} denotes the probability of onset of DN for the j^{th} patient under i^{th} classification of age at onset of diabetes, $(Ageatonset)_i$ is the fixed effect due to i^{th} classification of age at onset of diabetes, $SrCr_{ij}$ is j^{th} the value of SrCr for the i^{th} classification of age at onset of diabetes, NOS_{ij} is the j^{th} value of number of successes for the i^{th} classification of age at onset of diabetes, $SrCr_{Mean}$ and NOS_{Mean} are the mean of $SrCr_{ij}$ and NOS_{ij} values respectively and $\in_{ij}$ are the random error terms assumed to be normally distributed with mean zero and constant variance. β_1 *and* β_2 are regression coefficients corresponding to SrCr and number of successes respectively. Also, $r = r_1 + r_2$ is the total number of individuals whose onset of DN is observed.

Before conducting an ANCOVA – the homogeneity-of-regression (slope) assumption should first be tested. The test evaluates the interaction between the covariate and the factor (independent variable) in the prediction of the dependent variable. If the interaction is significant – the results from an ANCOVA are not meaningful – and ANCOVA should not be conducted. For our study results suggested that the interactions between age at onset of diabetes with SrCr and number of successes are not significant (p-values 0.775, 0.959).

The fitting of the model is tested through R^2, which came out to be 0.7560 i.e. 75.60% of variation in response variable is explained by the fitted model. The results of the analysis indicates that the probability of the onset of DN differ among the two levels of age at onset of diabetes (F-value = 4.5030, p-value = 0.0384). The F test is applied to find the independent contribution of each covariate in estimating the probability of onset of DN. The results showed that SrCr and number of successes are the significant contributors for predicting probability. Detailed description of results is given in table 3.

Table 3. ANACOVA for estimating probability of onset of DN for two levels of age at onset of diabetes and two covariates: SrCr and NOS.

Source	SS	df	MS	F	p-value
Intercept	**1.1140**	**1**	**1.1140**	**7.7820**	**0.0070**
Age at onset of diabetes	**0.6440**	**1**	**0.6440**	**4.5030**	**0.0384**
SrCr	**0.7460**	**1**	**0.7460**	**5.212**	**0.0260**
Number of Successes	**1.8160**	**1**	**1.8160**	**12.6810**	**0.0010**
Error	**8.0080**	**56**	**0.1430**		
Total		**59**			

V. DISCUSSION

Diabetic nephropathy occurs in approximately one third of individuals with type 2 diabetes [15]. Diabetic nephropathy is a clinical syndrome characterized by persistent albuminuria, a relentless decline in GFR, raised arterial blood pressure and increased relative mortality for cardiovascular diseases. This follows with a more rapid progression of other secondary complications, (retinopathy, neuropathy, diabetic foot and blood pressure)[16]. While considerable advances have been achieved in slowing the progression of diabetic nephropathy, the ultimate goal of arresting or reversing disease development remains unfulfilled. Therefore, proper statistical research is required to study the effectiveness of early intervention of renal diseases in high risk diabetic patients for conventional risk factors. The aim of this study is to estimate the probability of onset of DN in type 2 diabetic patients. The estimation of probability of a disease has a fundamental importance in medical science, as it provide greater sensitivity for detecting patients with diabetic nephropathy for future studies.

Firstly we have analyzed the association of occurrence of renal complication with the risk factors namely FBG, DBP, LDL, Age at onset of diabetes, number of successes and SrCr, by applying path analysis. Path analysis model are the extension of multiple regression model which are helpful in better prediction, since they can model the impact of each factor on an outcome. They can add the causal relationship structure to the regression analysis and capture both indirect and direct effects [17]. The results of the path analysis showed that FBG, DBP and NOS have direct effect on SrCr. And NOS, age at diabetes diagnosis and SrCr has direct effect in estimating the survival function of DN patients. Because of the lack of availability of data on factors like body mass index, weight, height, HbA1c, calorie intake etc. the number of variables considered in path analysis are not large. This could be taken as one of the limitation of our study, as path analysis is more effective when number of variables is large in number.

Secondly, we have applied ANCOVA for studying effect of the age at diabetes onset on the development of DN, as type 2 diabetes most often occurs in middle age. ANCOVA is applied, as it is the only statistical technique that compares a variable in 2 or more groups taking into account the variability of other variables, called covariates. For studying the effect of age we have estimated the probability of onset of DN at two different levels of age at onset of diabetes in the presence of two covariates namely SrCr and number of successes. The estimated probabilities are consistent with the previous studies [18].

In conclusion, the present study uses a powerful statistical technique viz. path analysis that allows for more complicated and realistic models than multiple regressions with its single dependent variable. The study helps us to understand the progression of type 2 diabetes on the development of nephropathy, by considering six different variables. The findings of our study extend previous studies and add new evidence to literature on the estimation of probability of onset of DN in type 2 diabetic patients by considering a new factor i.e age at onset of diabetes.

REFERENCES

[1] Alwakeel, J.S., Isnani, A.C., Alsuwaida, A., AlHarbi, A., Shaikh, S.A., AlMohaya, S., and Ghonaim, M.A., Factors affecting the progression of diabetic nephropathy and its complications: A single-center experience in Saudi Arabia, Ann Saudi Med,31(3), 2011, 236-242. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3119962/

[2] Zhuo L., Zou G., Li W., Lu J., Ren W. Prevalence of diabetic nephropathy complicating non-diabetic renal disease among Chinese patients with type 2 diabetes mellitus. European Journal of Medical Research, 2013, 1-8. http://www.eurjmedres.com/content/pdf/2047-783X-18-4.pdf

[3] Adler AI, Stevens RJ, Manley SE, Bilous RW, Cull CA, Holman RR. Development and Progression of nephropathy in type 2 diabetes: Observation and modeling from the United Kingdom Prospective Diabetes Study. Kidney International , 2003, 63,225-32 http://dx.doi.org/10.1046/j.1523-755.2003.00712.x.

[4] Dabla PK. Renal function in diabetic nephropathy. World J Diabetes. 2010, 1(2): 48-56. http://dx.doi.org/10.4239/wjd.v1.i2.48

[5] Streiner D.L. Finding your way: An introduction to path analysis. Can J Psychiatry, 2005, 50(2), 115-122.

[6] Ahn J. Beyond single equation regression analysis: Path analysis and multi-stage regression analysis, American Journal of Pharmaceutical Education, 2002, 66, 37-42.

[7] Mosteller M. A genetic analysis of cardiovascular disease risk factor clustering in adult female twins. Genetic Epidemiology 1993, 6:569–574.

[8] Wheelwright R , Birchall MA, Boaden R, Pearce G, Lennon A. Critical path analysis in head and neck cancer: a management technique for surgical oncology, Eur J Oncol Nurs 2002, 6(3),148-54.

[9] Tae YS, Heitkemper M, Kim MY. A path analysis: a model of depression in Korean women with breast cancer-mediating effects of self-esteem and hope. Oncol Nurs Forum, 2012, 39(1), E49-57. doi: 10.1188/12.ONF.E49-E57.

[10] Ludman E J., Katon W, Russo J, Korff M V, Simon G, Ciechanowski P, Lin E, Bush T, Walker E, Young B. Depression and diabetes symptom burden. General Hospital Psychiatry , 2004, 26, 430– 436. http://download.journals.elsevierhealth.com/pdfs/journals/0163-8343/PIIS0163834304001094.pdf.

[11] Skrivanova K, Bendova M, Dusek L, Zackova D, Racil Z, Mayer J. The Effect of Imatinib Treatment Duration on the Quality of the Life of Patients with Chronic Myeloid Leukemia , J Blood Disorders Transf, 2013, 4-6. http://dx.doi.org/10.4172/2155-9864.1000167.

[12] http://www.sagepub.in/upm-data/47570_ch_16.pdf

[13] Das MN, Giri NC, Design and analysis of experiments (2nd edition, 1999, New Age International)

[14] Dean A, Voss D, Design and analysis of experimentsI(Springer, 1999, New York)

[15] O'Bryan GT, Hostetter TH. The renal hemodynamic basis of diabetic nephropathy, Semin Nephrol ,1997,17(2): 93-100

[16] Lovell HG. Angiotensin converting enzyme inhibitors in normotensive diabetic patients with microalbuinuria. Cochrane Review In; The Cochrane Library, 2000, 2, 1-13.[17] http://www.education.umd.edu/EDMS/fac/Hancock/writings/Hancock%20&%20Mueller%20Path%20Analysis.pdf

[17] Grover G, Gadpayle AK, Sabharwal A. Identifying patients with diabetic nephropathy based on serum creatinine under zero truncated model, Electron. J. App. Stat. Anal, 2010, 3(1), 28 – 43