Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 11 (12): 2076-2083, 2011
ISSN 1812-5654 / DOI: 10.3923/jas.2011.2076.2083

Classification and Regression Trees: A Possible Method for Creating Risk Groups for Progression to Diabetic Nephropathy

[1]Mehmood G. Sayyad, [2]G. Gopal and [3]Arjan K. Shahani
[1]Department of Statistics, Abeda Inamdar Senior College of Arts, Science and Commerce, Azam Campus, Camp, Pune-411 001, India
[2]Department of Statistics, University of Madras, Chepauk Campus, Chennai-600 005, India
[3]GeoData Institute, University of Southampton, Southampton SO17 IBJ, UK

Abstract: This study describes the use of classification and regression tree (CART) analysis for analysing the progression of diabetic nephropathy (damage to kidneys) among type 2 diabetics in India. We also discuss a modelling approach for the development of an operational model, at the level of individual patients, for an early detection and treatment of this common complication of diabetes in India. Data for this modelling work is taken from a prospective study "Wellcome Diabetes Study" undertaken by the Diabetes Unit, King Edward Memorial Hospital, Pune, India.

Key words: Diabetic nephropathy, risk grouping, CART analysis, classification technique, modeling

INTRODUCTION

Diabetes mellitus is a disorder of carbohydrate metabolism characterised by high levels of blood glucose resulting from insufficient insulin secretion, insulin action, or both. Excess glucose is passed out of the body in the urine. Over time, this excess glucose circulating through the body in the bloodstream can lead to a number of long-term complications including kidney failure, blindness, amputations and heart problems.

By 2025 the world's diabetic population is estimated todouble from 110 million in 1994 to 221 million (Amos *et al*., 1997). Prevalence varies widely by ethnic group and country (adult rates range from less than 2% in rural Bantu people in Tanzania to nearly 50% in US Pima Indians and South Pacific Naurauns (Amos *et al*., 1997). India has the highest number of diabetic patients in any one country (~25 million in 2000) this number is predicted to rise to 57 million by the year 2025 (King *et al*., 1998). Rates are also relatively high in "transplanted populations," such as Asians in Europe (McKeigue *et al*., 1991) and African Americans (Harris *et al*., 1998). The projected increase in rates, however, is universal. Diabetes causes a great deal of suffering in the community and the costs for the necessary health services are high.

Diabetes can remain undiagnosed for many years and this delay in the detection of diabetes has many serious consequences that include macrovascular complications (Titty *et al*., 2008) and microvascular complications (Behnam-Rassouli *et al*., 2010). Detection and treatment of diabetes and its complications is a lengthy process and in a developing country like India with resource constraints there are more undiagnosed than diagnosed cases. Care of people with diabetes requires a multidisciplinary team with an active participation of the patients. Diabetes can be prevented and managed effectively with the help of lifestyle changes, proper diet and antidiabetic medicinal plants (Haque *et al*., 2011).

In this study, we describe the use of Classification And Regression Tree (CART) technique for creating risk groups for the progression of diabetic nephropathy, a common complication that arises due to diabetes. We also discuss briefly a systems level approach using the modelling methods of Operational Research for the care of people with diabetes in a community.

Indian scenario: There is a rapidly escalating epidemic of type 2 diabetes in India. The prevalence has increased in adult urban Indians from <3% in early 1970's to over 13%; in addition about 15% have Impaired Glucose Tolerance (IGT) (Ramachandran *et al*., 1988; Ramachandran *et al*., 2001). This rapid rise is ascribed to changes in life style such as altered food habits, reduced physical activity and psychosocial stress. Migrant Indians also show a higher prevalence of diabetes compared to local populations.

Indian type 2 diabetic patients are usually younger at the time of diagnosis than their western counterparts;

Corresponding Author: Mehmood G. Sayyad, Department of Statistics, Abeda Inamdar Senior College of Arts, Science and Commerce, Azam Campus, Camp, Pune-411 001, India

Indians are thinner by the Body Mass Index (BMI) criteria but they may possess excess body fat percent and are centrally obese. In addition, Indians are more insulin resistant than their western counterparts. Insulin resistance in Indians is associated with high blood pressure and abnormal circulating lipid concentrations, suggesting the existence of an insulin resistance syndrome (Yajnik, 1998, 2001).

There is little prospective data on the anthropometric and metabolic-endocrine predictors of diabetic tissue damage in Indians, most of the available data is cross sectional. However, there are a few prospective studies designed for this purpose, including United Kingdom Prospective Diabetes Study (UKPDS) and Diabetes Control and Complications Trial (DCCT) which are based on Western population (UKPDS group, 1991, DCCT Research Group 1998). In UKPDS study of 5102 type 2 diabetic subjects with average age 53 years were followed for 10-years and it is shown that tight glycaemic and blood pressure control can reduce the risk of microvascular complications (nephropathy is considered as one of the microvascular complications) significantly by 35% and 30% respectively. Adler *et al*. (2003) from UKPDS group have particularly studied the development and progression of diabetic nephropathy in type 2 diabetic patients. In the Wellcome Diabetes Study undertaken by the Diabetes Research Unit at King Edwards Memorial Hospital, Pune, India, newly diagnosed hyperglycemic subjects were followed for 10 years. The patients' anthropometric and metabolic-endocrine characteristics were measured during this follow up.

In this study, we analyse the predictors of diabetic nephropathy in Indian type 2 diabetic patients. We also discuss an approach for the development of an operational model for the early detection and the treatment of diabetic nephropathy.

Types of diabetes: There are two main types of diabetes: type 1 diabetes also known as Insulin Dependent Diabetes Mellitus (IDDM) and type 2 diabetes also known as Non-Insulin Dependent Diabetes Mellitus (NIDDM). In type 1 diabetes, there is very little or no production of insulin in the body and therefore insulin must be injected (it cannot be ingested as it would be digested by the body). Type 2 patients can produce some insulin themselves and their diabetes is controlled by diet or drugs. The commonest form of diabetes (95%) in a community is usually type 2 diabetes. Other forms of diabetes are Gestational Diabetes which appears during pregnancy and Impaired Glucose Tolerance (IGT) which may be thought of as a mild condition where the patient shows slightly higher than normal levels of glucose. Table 1 gives a brief description of type 1 diabetes and type 2 diabetes patients.

Table 1: Description of type 1 diabetes and type 2 diabetes patients

Details	Type 1 diabetes	Type 2 diabetes
Other names	IDDM	NIDDM
Age of onset	Usually less than 30 years	Usually over 40 years
Weight	Usually normal or underweight	Usually overweight
Onset	Usually over weeks	Often over months or years

Diabetic complications: Diabetic patients may suffer from a number of long-term complications. Over time, the excess levels of glucose in the bloodstream can damage the nervous system as well as damaging small blood vessels in the eyes and kidneys (Behnam-Rassouli *et al*., 2010). Three of the main diabetic complications can be summarised as:

- **Retinopathy:** It damage to the retina of eyes (Kohner, 1993). It can be treated successfully if detected in time. Around 20 percent newly diagnosed type 2 diabetic patients have retinopathy (background retinopathy). Average time to develop background retinopathy is 6 to 7 years from the diagnosis of type 2 diabetes (United Kingdom prospective diabetes study (UKPDS, 1990))
- **Neuropathy:** It damage to nerves. Nerve damage can eventually lead to foot ulceration and amputation. Around 15% of all people with diabetes eventually develop foot ulcers (United Kingdom prospective diabetes study (UKPDS, 1990))
- **Nephropathy:** It complications of kidneys. Damage to the small blood vessels in the kidneys can eventually lead to the renal failure (United Kingdom prospective diabetes study (UKPDS, 1990))

There are additionally many other complications, not specific to diabetes, where the risk of developing them increases if diabetes is present. Such complications include cardiovascular disease, cerebrovascular disease and peripheral vascular disease.

PATIENTS AND TREATMENT

The design of Wellcome Diabetes Study has been described earlier (Shelgikar *et al*., 1991; Yajnik *et al*., 2003). Briefly, 189 newly diagnosed type 2 diabetic patients were enrolled from outpatients and wards of the King Edward Memorial Hospital, Pune, India between February 1987 and December 1989 and were prospectively followed for ten years. These patients were studied at enrolment, five and ten years later. During 10-year follow-up 134 patients (Mean age at enrolment 45 years, 79 men and 55 women) were studied, 28 were lost to follow-up and 22 were dead. Following parameters were recorded at each time-point:

Table 2: Status of 10-year follow-up of wellcome diabetes study

Records	NGT	IGT	NIDDM
Initially enrolled (n)	133	79	189
Studied at 10 y follow-up (n)	94	60	134
Dead (n)	7	4	22
Lost to follow-up (n)	32	15	28
Followed-up (%)	76	81	80
Microalbuminurea (UAER 20 to 199 μg min^{-1}) (n, %)	-	-	18, 13.4%
Macroalbuminurea (UAER≥ 200 μg min^{-1}) (n, %)	-	-	5, 3.7%
Diabetic Nephropathy (n, %)	0	0	23, 17.2%

IGT: Impaired glucose tolerance, NIDDM: Non-insulin dependent diabetes mellitus, NGT: Normal glucose toleraance

- Clinical history
- Anthropometry (height, weight, waist and hip circumferences)
- Resting supine blood pressure (systolic and diastolic)
- Biochemical parameters (hemoglobin, urine analysis, serum creatinine, plasma total cholesterol and triglyceride concentrations)
- Plasma glucose and immunoreactive insulin (IRI) during oral glucose tolerance test (WHO, 1985)
- Indices of pancreatic beta cell function and insulin resistance derived from homoeostasis model assessment (HOMA) (Matthews *et al*., 1985)
- Markers of diabetic tissue damage (Ophthalmoscopic examination for diabetic retinopathy, urinary albumin excretion rate on 8 h overnight urine collection for diabetic nephropathy, resting electrocardiogram analysed by Minnesota code, lower limb arterial blood pressure measurement to calculate ankle-brachial index and vibration sensory threshold in lower extremities).

All patients received appropriate treatment; including advice on diet, exercise and oral hypoglycaemic agents or insulin as appropriate. Table 2 gives the 10-year follow-up status of the Wellcome Diabetes Study.

DATA ANALYSIS FOR CREATING RISK GROUPS

The initial phase of the analysis attempted to classify patients into risk groups for progression of diabetes nephropathy in India. Quantitative measures of the risk of suffering from the said diabetes complication can help in the care of people with diabetes. The quantification of risk means attempting a statistical classification analysis and there are many statistical techniques available for this purpose (Mokeddem and Belbachir, 2009; Chandrasekaran *et al*., 2005). We analysed the data by using a non-parametric multivariate statistical technique called Classification And Regression Tree (CART) analysis with the help of Statistical Package for Social Sciences (SPSS) version 13.0. The reason for using CART analysis was its advantages over the other statistical classificatory techniques. Some of the advantages of CART analysis are (a) CART is relatively simple to understand and interpret by the non-technical people like Physicians (b) it is capable to handle both numerical and categorical data (c) it is possible to validate the model (d) it is robust (e) it can perform well with the large data in a short time.

Classification and regression trees (CART): CART has proved to be a highly useful tool in the work of the Institute of Modelling for Healthcare and in the work of many other people (Shahani *et al*., 1994; Ridley *et al*., 1998; Harper *et al*., 2003; Sayyad *et al*., 2002). Depending on the problem, the basic purpose of the CART analysis is to either produce an accurate classifier or to uncover the predictive structure in the available data. The main idea of this technique is the formation of subgroups of patients (nodes) within which the response variable is relatively homogeneous. The CART algorithm incorporates a binary splitting system that attempts to create groups of patients such that patients within the same group are as closely related as possible and statistically different from the other groups (a variance reduction algorithm is incorporated) (Breiman *et al*., 1984). The resulting tree clearly shows each risk group and can be easily interpreted.

CART methodology: The CART methodology involves a complete collection of rules that defines the tree which uses a process known as 'recursive partitioning'. A series of binary splits is made based on the answers to questions of the type 'Does observation or case $i \in A$', where A is a region of the covariate space. Answering such a question splits the covariate space. Cases for which the answer is yes are assigned to one group and those for which the answer is no to an alternative group. The implementation of tree modelling proceed by imposing the following constraints:

- Each split depends on the value of only a single covariate
- For ordered (continuous or categorical) covariate x_j, only splits resulting from the questions of the form 'Is $x_j<C$' are considered. Thus the ordering is preserved
- For categorical independent variables, all possible splits into disjoint subsets of the categories are allowed

A tree is grown using following steps:

Step 1: Examine every allowable split on each independent variable

Step 2: Select and execute (that is, create left and right daughter nodes) from the best of these splits

The initial or root node of the tree comprises the whole sample. Step 1 and 2 are then reapplied to each of the daughter nodes. To determine the best node to split into left and right daughter nodes at any stage in the construction of the tree involves the use of a numerical split function, Φ (s,g) often referred as *deviance*. This can be evaluated for any split 's' of node 'g'. The form of Φ (s,g) depends on whether the response variable is continuous or categorical. The usual split function chosen for a continuous response variable is based on the within-node sum of squares, that is for a node g with Ng cases, the term:

$$SS(g)=\sum_{i\in g}[yi-\bar{y}(g)]^2$$

where, yi denotes the response variable value for the 'ith' individual and $\bar{y}(g)$ is the mean of the responses of the Ng cases in node g. If a particular split, s of node g is into left and right daughter nodes, g_L and g_R, then the least-squares split function is given by:

$$\Phi(s,g) = SS(g)-SS(g_L)-SS(g_R)$$

and the best split of node g is determined as the one that corresponds to the maximum of the above function amongst all allowable splits.

Split functions for categorical response variables (in particular, binary variables) are based on trying to make the probability of a particular category of the variable close to one or zero in each node. Most commonly used is a log-likelihood function defined for node 'g' as:

$$LL(g)=-2\sum_{i\in g}\sum_{k=1}^{K} y_{ik}\log(p_{gk})$$

where, 'K' is the number of categories of the response variable, y_{ik} is an indicator variable taking the value 1 if individual 'i' is in category 'k' of the response and zero otherwise and is the probability being in the kth category of the response in node 'g', estimated as n_{gk}/N_g, where n_{gk} is the number of individuals in category k in node 'g'. The corresponding split function Φ (s,g) is then calculated as:

$$\Phi(s,g) = LL(g)-LL(g_L)-LL(g_R)$$

and again the chosen split is that maximizing Φ (s, g).

Trees are grown recursively splitting nodes to maximize Φ, leading to smaller and smaller nodes of progressively increased homogeneity. A critical question is when tree construction should end and terminal nodes be declared. The two simples stopping rules are:

Table 3: Collinearity diagnostic statistics for the selected IVs (all initial), dependent variable is normalised# UAER at 10-years

Variables in the model (all measured initially)	Std. Beta	Collinearity statistics	
		Tolerance	VIF
Age (years)	-0.236	0.791	1.265
Sex (1: Male, 2: Female)	0.076	0.317	3.159
Smoking status (0:No, 1:Yes)	0.031	0.806	1.241
Body mass index (kg m^{-2})	0.121	0.622	1.607
Waist to hip ratio	0.122	0.362	2.763
Systolic blood pressure (mmHg)	0.212	0.342	2.928
Diastolic blood pressure (mmHg)	-0.375	0.287	3.479
HbA1C (mg%)	-0.099	0.453	2.209
Fasting glucose (mg%)*	0.138	0.100	9.953
120 min glucose (mg%)	0.125	0.200	4.988
Fasting insulin (mU L^{-1})	0.741	0.020	50.770
120 min insulin (mU L^{-1})	-0.310	0.256	3.911
Cholesterol (mg%)	0.261	0.695	1.439
Triglycerides (I mg%)	-0.141	0.682	1.467
HOMA-R*	-0.621	0.025	40.560
HOMA-B	0.073	0.111	9.011

#Natural logarithm (Log_e) transformation used. Std.Beta: Standardized beta coefficient, VIF: Varianve inflation factor. *The variables with relatively higher collinearity

- Node size-stop when this drops below a threshold value, for example when $N_g<10$
- Node homogeneity-stop when a node is homogeneous enough, for example, when its deviance is less than 1% of the deviance of the root node

Multicollinearity: When Independent Variables (IVs) are correlated, there are problems in estimating regression coefficients, as the impact of collinearity on regression analysis has been shown by several others including Mela and Kopalle (2002) and D'Ambra and Sarnacchiaro, (2010). Multicollinearity means that within the set of IVs, some of the IVs are nearly or totally predicted by the other IVs. Just like the other classificatory techniques, CART technique does not deal with multicollinearity precisely. So we obtained tolerance statistics with Variance Inflation Factor (VIF) as a diagnostic tool for multicollinearity among the IVs that we have chosen, using Multivariate Linear Regression Analysis (MLRA) with 10-year urinary albumin excretion rate (continuous variable as a surrogate for 10-yr nephropathy status) as a dependent variable. We found that the values of tolerance statistics for the variables fasting insulin and HOMA-R were very low (Fasting insulin 0.020, HOMA-R 0.025) with corresponding larger values of VIF (Fasting insulin 50.77, HOMA-R 40.56) (Table 3) suggested the presence of collinearity. We then excluded these two variables from the subsequent analysis.

DIABETIC NEPHROPATHY

Clinical definition: Diabetic nephropathy is a generic term referring to any deleterious effect on kidney structure

and/or function caused by diabetes mellitus. More specifically, diabetic nephropathy is thought of in stages, the first being that characterized by microalbuminuria (urinary excretion of 20 to 199 μg of albumin min^{-1}). This may progress to macroalbuminuria, or overt nephropathy (urinary excretion of over 200 μg of albumin min^{-1}). Later on, renal function decline progressively which is characterized by decreased Glomerular Filtration Rate (GFR) results in clinical renal insufficiency and End-Stage Renal Disease (ESRD). We calculated the Urinary Albumin Excretion Rate (UAER) using the formula:

$$\text{UAER} = \frac{\text{Urinary albumin} \times \text{Urine volume}}{\text{Time in min}}$$

What causes diabetic nephropathy?: A variety of factors contribute to the renal damage seen in diabetes. By definition, hyperglycemia is a common etiological factor in diabetic patients with nephropathy but a genetic predisposition and smoking may contribute as well. Most significant, however, is the presence of hypertension, not only before and after the onset of microalbuminuria but probably also as another familial marker of risk, since patients with diabetes and a positive family history of hypertension are at higher risk of nephropathy. It has also been shown that, microalbuminurea is a powerful predictor of cardiovascular disease in both type 1 and type 2 diabetes (ADA: Position Statement, 1999; Agius *et al*., 2009).

Diabetic nephropathy as an outcome variable: The dependent variable was a nominal 0-1 variable, clinically defining the absence (UAER<20 μg min^{-1}) and presence (UAER≥20 μg min^{-1}) of diabetic nephropathy at 10-year follow-up. The list of independent variables included age, sex, smoking status, body mass index, waist to hip ratio, systolic blood pressure, diastolic blood pressure and circulating concentrations of fasting glucose, two h glucose, two-h insulin, glycocylated haemoglobin (HbA_{1C}) cholesterol, triglycerides and beta cell sensitivity (HOMA) (Matthews *et al*., 1985). All these independent variables are measured at enrolment. Figure 1 shows the resulting classification tree. Table 4 gives a summary of the data for the patients in the various nodes.

From the node tree and node summary Table it is clear that overall 17.2% type 2 diabetic patients had diabetic nephropathy at 10-year follow-up (Node = 0). The most important predictor of nephropathy is the 2h insulin level as indicated by the first binary split (Node = 1 for 2 h insulin ≤45.5 mU L^{-1} and Node = 2 for 2 h insulin >45.5 mU L^{-1}). Among the patients who had 2 h insulin level ≤45.5 mU L^{-1} (hypoinsulanemic) the incidence of nephropathy is 30.2% while those who had 2 h insulin level >45.5 mU L^{-1} (hyperinsulanemic) it falls down to 8.6% which indicates that insulin deficiency at diagnosis is a most important predictor of diabetic nephropathy. The next important determinant of diabetic nephropathy among hypoinsulanemic patients is the central obesity as measured by waist to hip ratio. Hypoinsulanemic patients, with higher central obesity (waist to hip ratio>0.88) had 44.8% chances of developing nephropathy. The other and perhaps the most important predictor of diabetic nephropathy among hypoinsulanemic and centrally obese patients is the systolic blood pressure with a highest incidence of 85.7%. In the hyperinsulanemic patients the higher fasting glucose level (>133.0 mg%) predicts the diabetic nephropathy with 15.4% incidence.

From an overall incidence of 17.2%, the CART algorithm has managed to find five risk groups (End nodes 3, 5, 6, 7, 8) with incidence of nephropathy varying from as little as 2.4% to as much as 85.7%. The risk factors such as two-h insulin, fasting glucose and systolic blood pressure levels produced by the CART analysis can be clinically controlled by the appropriate interventions. While the risk factor of central obesity can be controlled by the appropriate diet and exercise. The possible interventions for all or high-risk groups (Nodes 1, 4, 7, 8 as indicated by CART analysis) could be diet, Oral Hypoglycaemic Agents (OHAs) insulin injections, exercise and dialysis etc., or several combinations of these interventions.

In a similar study done by a Nigerian group (Ejuoghanran *et al*., 2009); the incidence of diabetic nephropathy is 72.63%, about 4 times higher than the incidence of our study. This might be due to relatively higher age distribution (older) of Nigerian patients at

Table 4: Node summary for diabetic nephropathy complication

Node	n	Field description	Range (Min.-Max.)	Deviance	Probability of getting nephropathy
0	134	Group of patients studied at 10 yrs (Type 2 diabetes)	-	53.36	0.172
1	53	120 min Insulin (mU L^{-1})	1.00-45.50	28.19	0.302
2	81	120 min Insulin (mU L^{-1})	45.50-300.00	20.70	0.086
3	24	Waist to hip ratio	0.70-0.88	7.85	0.125
4	29	Waist to hip ratio	0.88-1.03	17.32	0.448
5	42	Fasting glucose (mg%)	77.5-133.02	4.10	0.024
6	39	Fasting glucose (mg%)	133.02-333.99	14.54	0.154
7	7	Systolic BP (mmHg)	100-113	2.49	0.857
8	22	Systolic BP (mmHg)	113-182	11.95	0.318

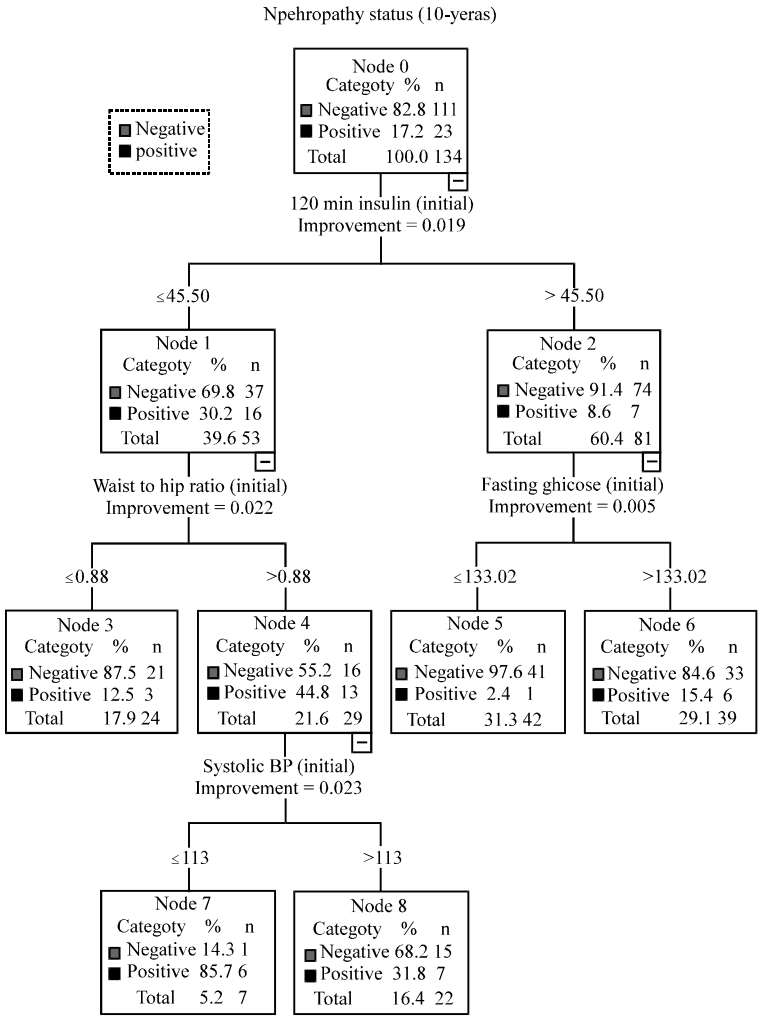

Fig. 1: Risk groups for diabetic nephropathy

baseline. Consequently, they have higher duration of diabetes compared to the group of patients in the current study.

Validation of CART analysis: We validated the CART analysis in order to assess its predictability to a larger diabetic community. As our sample size was small we performed validation using cross-validation technique by specifying 10 sample folds. The cross-validated risk estimate is a measure of tree's predictive accuracy for the final tree. The risk estimate is calculated by averaging the risks for all the trees generated during validation. The risk estimate was equal to 0.209 with its standard error being 0.035.

MODELLING APPROACH

It is hoped that future work will involve building an operational model for the patients with type 2 diabetes. This will involve development of detailed models, at the level of individual patients, for an early detection and treatment of diabetic nephropathy.

This work would incorporate the evolved risk groups in a community, together with the natural history of this complication and options for early detection and the treatment. Further, it has been shown that the operational models can be powerful tools for making effective decisions about effective and efficient health care (Shahani, 1996). Figure 2 illustrates a model for the progression of patients with diabetic nephropathy.

The information about transition times and the probability of transition for the different risk groups may be obtained from the Wellcome diabetes study, other relevant databases and expert opinion. There is no clear record in the Wellcome diabetes study about the cost of the different treatment options for a given risk group; however, this information can be obtained by using expert opinion (Indian practising Diabetologist).

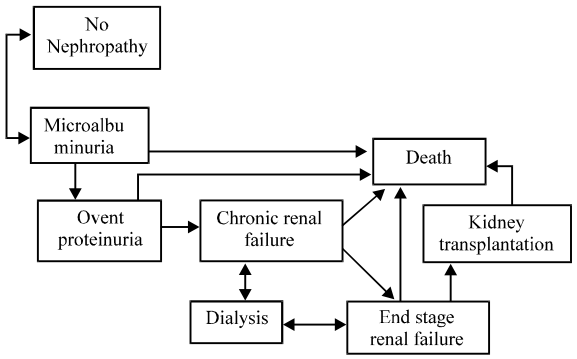

Fig. 2: A model for the progression of patients with diabetic nephropathy

A simulation model of the natural history incorporating the risk groups, transition times, intervention options and the costs of the interventions should provide a powerful tool for evaluating various options for cost effective interventions (Shahani *et al*., 1994).

CONCLUSION AND SCOPE FOR FURTHER WORK

This study has described the use of CART technique in forming the risk groups for diabetic nephropathy in India based on a prospective study done for 10-years. The quantitative assessment of risk of developing diabetic nephropathy will definitely provide the health care professionals the options for its better management. A major limitation of present findings is that they come from a relatively smaller sample size, therefore the minimum number of patients in each node was set at 15 for the parent node and 7 for the daughter node. However, the strength of our results lies in the prospective design of underlying study and the use of independent variables measured at enrolment. The size of the tree that we obtained after cross validation is relatively smaller and can be easily interpreted, so we did not opt for finding the optimal tree size. A simple plot of deviance against tree size helps to decide the optimal tree size in a complex tree. The size of the tree at which minimum deviance is obtained is taken as the optimal tree size. Our results are comparable with the findings by Adler *et al*. (2003) from UKPDS group. Diabetic Nephropathy (microalbuminurea) is also considered as an important marker of greatly increasing cardiovascular morbidity and mortality among the patients with type 1 and type 2 diabetes. Thus, our findings based on CART analysis serve as an indication for screening for possible vascular disease and aggressive intervention to reduce all cardiovascular risk factors (e.g., antihypertensive therapy, tight glycaemic control therapy, diet and exercise etc.,) in Indian type 2 diabetic patients.

In this study, we also demonstrate the potential value of a multidisciplinary study of the care of people with diabetes nephropathy. The methodology with appropriate databases, multivariate statistical analysis including CART analysis, mathematical modelling, together with the development of easy to use models on personal computers will be a very powerful one.

It is hoped that the further work will involve the construction of detailed operation model at the level of individual patients for the effective and efficient patient management for diabetic nephropathy.

ACKNOWLEDGMENTS

We thank Dr. C.S. Yajnik, Director of Diabetes Unit, King Edward Memorial Hospital, Pune, India for providing the dataset for the statistical analysis. We also thank the referees for their valuable suggestions.

REFERENCES

ADA: Position Statement, 1999. Diabetic nephropathy. Diabetes Care, 22: 66-69.

Adler, A.I., R.J. Stevens, S.E. Manley, R.W. Bilous, C.A. Cull and R.R. Holman, 2003. Development and progression of nephropathy in type 2 diabetes: The United Kingdom Prospective Diabetes Study (UKPDS 64). Kidney Int., 63: 225-232.

Agius, E., G. Attard, P. Shakespeare, P. Clark, M.A. Vidya, A.T. Harttersley and S. Fava, 2009. Familial factors in diabetic nephropathy: An offspring study. Diabetic Med., 23: 331-334.

Amos, A.F., D.J. McCarty and P. Zimmet, 1997. The rising global burden of diabetes and its complications: Estimates and projections to the year 2010. Diabet. Med., 14: S1-S85.

Behnam-Rassouli, M., M.B. Ghayour and N. Ghayour, 2010. Microvascular complications of diabetes. J. Biol. Sci., 10: 411-423.

Breiman, L., J.H. Friedman, R.A. Olshen and J. Stone, 1984. Classification and Regression Trees. 1st Edn., Wadsworth International Group, Belmont, CA, ISBN: 978-0412048418, pp: 102-116.

Chandrasekaran, V., G. Gopal and A. Thomas, 2005. Summarizing data through a piecewise linear growth curve model. Stat. Med., 24: 1139-1151.

D'Ambra, A. and P. Sarnacchiaro, 2010. Some data reduction methods to analyze the dependence with highly collinear variables: A simulation study. Asian J. Math. Stat., 3: 69-81.

DCCT Research Group, 1998. The effect of intensive diabetes therapy on measures of autonomic nervous system function in the Diabetes Control and Complications Trial (DCCT). Diabetologia, 41: 416-423.

Ejuoghanran, O.S.O., O.E. Chukwu and O.E. Fidelis, 2009. Incidence of diabetic nephropathy in Southern Nigeria. J. Med. Sci., 9: 264-269.

Haque, N., U. Salma, T.R. Nurunnabi, M.J. Uddin, M.F.K. Jahangir, S.M.Z. Islam and M. Kamruzzaman, 2011. Management of type 2 diabetes mellitus by lifestyle, diet and medicinal plants. Pak. J. Biol. Sci., 14: 13-24.

Harper, P.R., M.G. Sayyad, V. de Senna, A.K. Shahani, C.S. Yajnik and K.M. Shelgikar, 2003. A systems modelling approach for the prevention and treatment of diabetic retinopathy. Eur. J. Operat. Res., 150: 81-91.

Harris, M.I., K.M. Flegal, C.C. Cowie, M.S. Eberhardt and D.E. Goldstein *et al*., 1998. Prevalence of diabetes, impaired fasting glucose and impaired glucose tolerance in U.S. adults. The Third National Health and Nutrition Examination Survey, 1988-1994. Diabetes Care, 21: 518-524.

King, H., R.E. Aubert and W.H. Herman, 1998. Global burden of diabetes, 1995-2025: Prevalence, numerical estimates and projections. Diabetes Care, 22: 1414-1431.

Kohner, E.M., 1993. Diabetic retinopathy. Br. Med. J., 307: 1195-1199.

Matthews, D.R., J.P. Hosker, A.S. Rudenski, B.A. Naylor, D.F. Treacher and R.C. Turner, 1985. Homeostasis model assessment: Insulin resistance and betta-cell function from fasting plasma glucose and insulin concentration in man. Diabetologia, 28: 412-419.

McKeigue, P.M., B. Shah and M.G. Marmot, 1991. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. Lancet, 337: 382-386.

Mela, C.F. and P.K. Kopalle, 2002. The impact of collinearity on regression analysis: The asymetric effect of negative and positive correlations. Applied Econ., 34: 667-677.

Mokeddem, D. and H. Belbachir, 2009. A survey of distributed classification based ensemble data mining methods. J. Applied Sci., 9: 3739-3745.

Ramachandran, A., M.V. Jali, V. Mohan, C. Snehalatha and M. Viswanathan, 1988. High prevalance of diabetes in an urban population in South India. Br. Med. J., 297: 587-590.

Ramachandran, A., C. Snehalatha, A. Kapur, V. Vijay and V. Mohan *et al*., 2001. High prevalence of diabetes and impaired glucose tolerance in India: National urban diabetes survey. Diabetologia, 44: 1094-1101.

Ridley, S., S. Jones, A. Shahani, W. Brampton, M. Nielsen and K. Rowan, 1998. Classification trees: A possible method for iso-resource grouping in intensive care. Anesthesia, 53: 833-840.

Sayyad, M.G., A.K. Shahani, P.R. Harper, V.de Senna, K.M. Shelgikar and C.S. Yajnik, 2002. A system modelling approach for the care of diabetic retinopathy in India. Quantitative apporaches in health care management. Proceedings of 27th Meeting of European Working Group on Operational Reaserch Applied in Health Services, (ORAHS'02), Vienna, pp: 271-280.

Shahani, A.K., N. Korve, K.P. Jones and D.J. Paynton, 1994. Towards an operational model for prevention and treatment of asthama attacks. J. Operat. Res. Soc., 45: 916-926.

Shahani, A., 1996. Commentary: Models can be powerful tools for making effective decisions about effective and efficient health care. Br. Med. J., 313: 1057-1057.

Shelgikar, K.M., C.S. Yajnik and T.D.R. Hockaday, 1991. Central rather than generalised obesity is related to hyperglycaemia in Asian Indian subjects. Diabetes Med., 8: 712-717.

Titty, F.V.K., W.K.B.A. Owiredu and M.T. Agyei-Frempong, 2008. Prevalence of metabolic syndrome and its individual components among diabetic patients in Ghana. J. Boil. Sci., 8: 1057-1061.

UKPDS group, 1991. UK Prospective Diabetes Study VIII: Study design, progress and performance. Diabetologia, 34: 877-890.

UKPDS, 1990. Complications in newly diagnosed type 2 diabetic patients and their association with different clinical and biochemical risk factors. Diabetes Res., 13: 1-11.

World Health Organization, 1985. Diabetes mellitus: Report of WHO study group. WHO Technical Report Series, 727. Geneva.

Yajnik, C.S., 1998. The insulin resistance epidemic in India: Small at birth, big as adult. IDF Bull., 43: 23-28.

Yajnik, C.S., 2001. The insulin resistance epidemic in India: Fetal origins, later lifestyle, or both. Nutr. Rev., 59: 1-9.

Yajnik, C.S., K.M. Shelgikar, S.S. Naik, M.G. Sayyad and K.N. Raut *et al*., 2003. Impairment of glucose tolerance over 10 years in middle aged normal glucose tolerant Indians. Diabetes Care, 26: 2212-2213.