17

Mathematical modelling of kidney disease stages in patients diagnosed with diabetes mellitus II

X. Escobar Carrera♭,[1] S. González de Julián♮ and I. Barrachina Martínez♭

(♭) Research Centre for Economics Engineering,
Universitat Politècnica de València
Camí de Vera s/n, Valencia, Spain.

1 Introduction

The direct costs associated with diabetes mellitus represent 8% of total healthcare expenditure in Spain, amounting to around 6 billion euros per year [1]. The overall prevalence of diabetes in people over 18 years of age (adjusted for age and sex) is estimated at 13.8% in 2010 [2], with type 2 (T2DM) being the most common type of diabetes, accounting for 85-95% of all diabetes cases in high-income countries [3]. T2DM is associated with multiple diseases such as chronic kidney disease [4], retinopathy, pyelonephritis, heart attack or stroke [5]. It is estimated that 35% of patients with T2DM develop diabetic kidney disease [6]. Treatment of end-stage renal disease requires expensive treatments such as haemodialysis and kidney transplantation.

The objective of this study is to evaluate in patients with T2DM the degree of renal damage and the risk of suffering complications according to their socio-demographic, clinical and morbidity characteristics and to obtain the weight of the variables that have most influence.

2 Methods

This was an observational, population-based, cross-sectional study in patients with T2DM aged over 18 years assigned to a health district in Spain in 2015.

Data for the study were provided by the Conselleria de Sanitat i Salut Publica of the Valencian Community and it contains: the information about the Population Information System (SIP), the Ambulatory Information System (SIA), the electronic outpatient clinical records (ABUCASIS), Pharmacy Prescriptions Manager (GAIA), the Hospital Minimum Data Set (MDS) and the clinical analytical laboratory information was Geslab.

Patient data were anonymized, and the study has strictly complied with the current personal data protection regulations, specifically Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, as well as the Organic Law 3/2018, of December 5 concerning protection of personal data and guarantee of digital rights.

Additionally, the study protocol was approved by the Ethical Review Board of the University Hospital Clinic of Valencia.

[1]silgonde@upv.es

The sociodemographic and clinical variables that were available are sex, age, morbidity (according to the Clinical Risk Groups classification), drug prescription, estimated glomerular filtration rate (eGFR), albuminuria, glycated haemoglobin (patient controlled or not) and cholesterol.
The renal impairment of patients is established using the KDIGO scale, that is based on the estimated glomerular filtration rate and albuminuria. The closest available values of these parameters to 31 December 2015 were used to classify each patient in a KDIGO group. According to the KDIGO scale, patients were classified in 4 groups in a graph with different colours (figure 1.1), where green means low risk, yellow means moderate risk, orange is high risk and red is a very high risk of developing kidney damage. However, in order to simplify the analysis, for this study patients have been classified just in two groups, where the first one includes patients that has any or little renal damage (green and yellow in the KDIGO scale) and the second group are patients with medium or high renal damage (orange and red in the KDIGO scale).
The value of the glycated haemoglobin is used to determine if the disease is controlled or not in the patient. In this sense, a diabetic patient is considered to be controlled if it is under 65 years old and has a glycated haemoglobin value equal or less than 7, and for patients above 65 years old, the glycated haemoglobin value is equal or less than 8. Pharmaceutical prescription has been considered by classifying patients according to the mechanism of action of the drug they are taking. The groups of drugs considered are: biguanides, oral hypoglucemiant drug combinations, intermediate and rapid-acting insulin and analogue combinations, dipeptidyl peptidase (DPP-4) inhibitors, alpha-glucosidase inhibitors, insulins, other hypoglucemiant drugs excluding insulins, sulfonylureas and thiazolidinediones.
We used real-world data including all T2DM patients with albuminuria and glomerular filtration rate measurements in one health department in 2015. This makes the number of patients included in each of the KDIGO groups differ significantly, finding that 85% of the patients have low or moderate risk of kidney damage, and only 15% have a high or very high risk of kidney damage. We used a decision tree function, fitctree to be exact, in Matlab program. This returns a fitted binary classification decision tree based on the input variables. Due to the different group sizes, some of them with a very small number of patients, a K-fold cross-validation in each of the models has been used, using the function named cvpartition, that defines a random partition on a data set. We used this partition to define training and test sets for validating a statistical model using cross-validation.
To evaluate each of the models we used the accuracy, precision, sensitivity, specificity, AUC ROC and ROC Curve. The accuracy is the percentage of correct predictions. For the total number of positive predictions, the precision evaluates the percentage of correct positive predictions. The sensitivity is the percentage of correct positive predictions out of the real number of positives and the specificity is the percentage of correct negative predictions out of the real number of negatives. The AUC ROC and the ROC Curve has been obtained by using the perfcurve function, that provides information about the relationship between the sensitivity and specificity.

3 Results

The total population of the Health District was 263,334 adults in 2015, of these 28,345 patients had type 2 Diabetes Mellitus, and only 14,935 patients had available the determinations of the eGFR and albuminuria. Among these patients, almost the 86% of the patients are classified on the 2 first groups of the KDIGO scale (green and yellow), which includes our first group of analysis, and the rest of them belong to the last 2 groups (orange and red) (Figure 1).
Different models with the available variables were considered, on an individual basis and in different combinations, with the aim to obtain the model that can better differentiate the characteristics of patients at higher risk of complications.

Low risk	66.19%
Moderately increased risk	20.63%
High risk	7.94%
Very high risk	5.24%

				Albuminuria			
				A1	A2	A3	
				Normal to mildly increased	Moderately increased	Severely increased	
				< 30 mg/g	30 - 300 mg/g	≥ 300 mg/g	Total
eGFR	G1	Normal or High	>90	4,508	688	45	5,241 (35.1%)
	G2	Mildly decreased	60-89	5,377	1,210	113	6,700 (44.9%)
	G3a	Mildly to Moderately decreased	45-59	1,183	458	62	1,703 (11,4%)
	G3b	Moderately to severely decreased	30-44	570	323	75	968 (6.5%)
	G4	Severely decreased	15-29	113	134	40	287 (1.9%)
	G5	Kidney failure	<15	6	17	13	36 (0.2%)
			Total	11,757 (78.7%)	2,830 (18.9%)	348 (2.3%)	14,935

Figure 1: Distribution of the 14,935 patients according to eGFR and albuminuria categories according to the KDIGO 2012 classification.

In the section 3, it is shown that, depending on which variables or combinations or variables are considered, different results were obtained, but the accuracy, precision and sensitivity are almost identical in all and very high in most of the cases. However, the indicator that allows for a better comparison of the different models is the AUC ROC and denotes that the variables that individually get higher results are age (0.72) and cholesterol (0.60). When two variables are combined, the best model includes age and cholesterol, with an AUC ROC of 0.85. The best model with 3 variables is the one including age, cholesterol, and drugs, with an AUC ROC of 0.90. Finally, when all the variables are included (age, cholesterol, sex, drugs, CRG, and controlled glycated haemoglobin) the best AUC ROC value is obtained (0.93), but it shows the same result if glycated haemoglobin is excluded. These two models that show the best results are very complex, and all the necessary data is not always available. It is possible to obtain similar results with simpler models that only include the combination of three variables (age, cholesterol and drugs or age, cholesterol and CRG).
As mentioned before, the parameter used to compare the different models evaluated is the AUC ROC, and it is represented graphically in Figure 3 with the ROC Curve. The models that combined 5 or 6 variables get the best results, but similar values can be obtained with less complex models.

VARIABLES	ACCURACY	PRECISION	SENSITIVITY	SPECIFICITY	AUC ROC
AGE	0.87	0.87	0.99	0.02	0.72
CONTROLLED*	0.87	0.87	1.00	0.00	0.50
SEX	0.87	0.87	1.00	0.00	0.50
CRG	0.87	0.87	1.00	0.00	0.50
DRUGS*	0.87	0.87	1.00	0.00	0.50
CHOLESTEROL	0.87	0.87	1.00	0.01	0.60
AGE- CHOLESTEROL	0.84	0.88	0.95	0.17	0.85
AGE CHOLESTEROL-CONTROLLED	0.84	0.88	0.94	0.18	0.87
AGE-CHOLESTEROL-CRG	0.84	0.89	0.93	0.22	0.89
AGE-CHOLESTEROL-DRUGS	0.84	0.89	0.93	0.24	0.90
AGE-CHOLESTEROL-SEX	0.84	0.88	0.94	0.17	0.86
AGE-CHOLESTEROL-SEX-DRUGS	0.84	0.89	0.93	0.24	0.91
AGE-CHOLESTEROL-SEX-DRUGS-CRG	0.83	0.89	0.92	0.27	0.93
AGE-CHOLESTEROL-SEX-DRUGS-CRG-CONTROLLED	0.83	0.89	0.92	0.28	0.93

*Controlled: if glycated haemoglobin is controlled or not. Drugs: Mechanism of action of drugs.

Figure 2: Accuracy, precision, sensitivity, specificity and AUC ROC results according to the variables introduced in the models

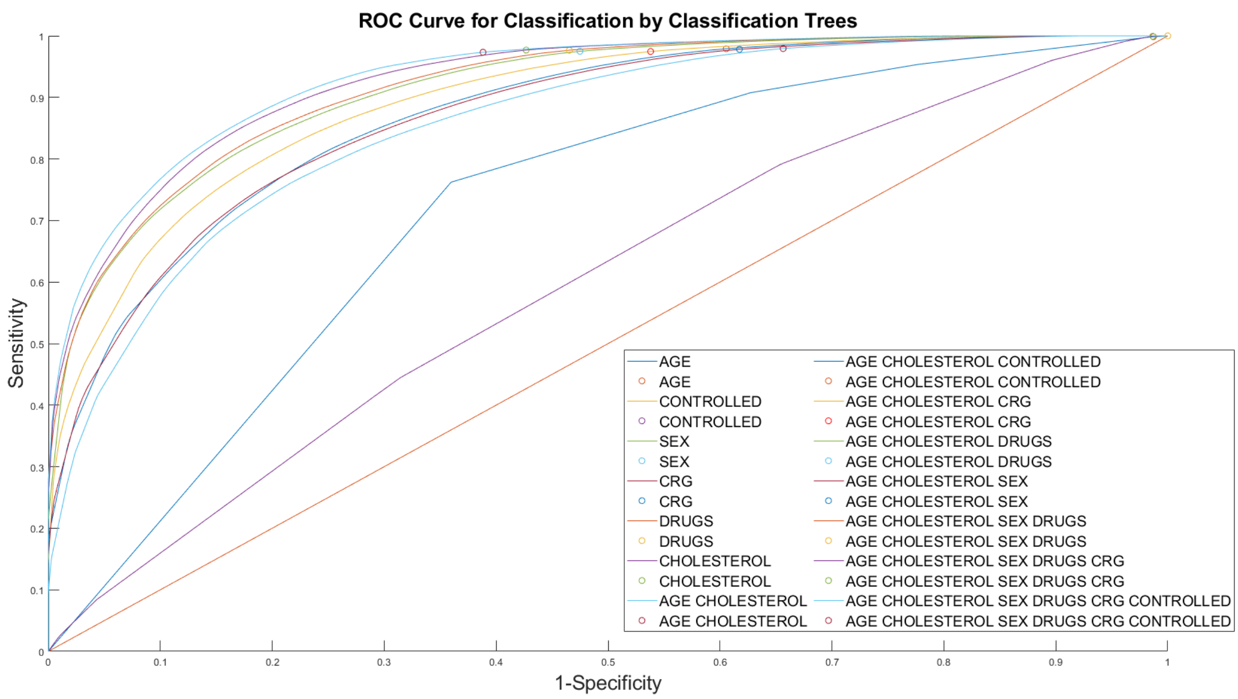

Figure 3: ROC Curve.

4 Conclusions

Considering that chronic kidney disease is highly prevalent in patients with T2DM in Spain and the progression of the disease is correlated with an increase of healthcare resource use, it's important to identify the characteristics of patients most likely to suffer kidney damage in order to implement early management strategies.

Different models were elaborated to analyse the degree of renal impairment of patients with T2DM according to the socio-demographic, clinical and morbidity variables, that it was determined the weight of each variable and obtained the most influential one.

The variables with a higher discrimination power in our model in an individual basis are age and cholesterol. The best model is obtained with the combination of all variables, but the simpler and the best option is the model that includes the variables age, cholesterol and drugs, that gets an AUC ROC value of 0.90.

Next steps include the design of a logistic regression model and its representation with a nomogram, which make it easier to know the weight of each variable. In addition, it is important to obtain more years of data to make a follow up and improve the models.

References

[1] Crespo C, Brosa M, Soria-Juan A, Lopez-Alba A, López-Martínez N, Soria B. Direct cost of diabetes mellitus and its complications in Spain (SECCAID Study: Spain estimated cost Ciberdem-Cabimer in Diabetes). Av en Diabetol. 2013 Nov 1;29(6):182–9.

[2] Soriguer F et al. Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es Study. Diabetologia. 2012;55(1):88-93.

[3] International Diabetes Federation. IDF Diabetes Atlas, 6th ed. Brussels, Belgium: International Diabetes Federation, 2013.

[4] de Boer IH et al. Temporal trends in the prevalence of diabetic kidney disease in the United States. JAMA. 2011 Jun 22;305(24):2532-9.

[5] Ray JA et al. Review of the cost of diabetes complications in Australia, Canada, France, Germany, Italy and Spain. *Curr Med Res Opin*. 2005;21(10):1617-1629.

[6] Wanner C, Inzucchi SE, Lachin JM, Fitchett D, Von Eynatten M, Mattheus M, et al. Empagliflozin and progression of kidney disease in type 2 diabetes. *N. Engl. J. Med. 2016;375(4):323–34.*