

Using Markov Chain to Predict by the Natural Progression of Diabetic Nephropathy at Diabetes and Endocrinology Hospital

Mohammedelameen Eissa Qurashi¹, Abdellateef Khalifa Hamid², and Mubarak H.Elhafian^{3,1}
wedalameen81@gmail.com abdu-53@hotmail.com hafian10@yahoo.com

¹ Sudan University of Science and Technology, Department of Statistics, College of Science

²Shendi University. Department of Statistics, Faculty of Science and Technology

³King Abdulaziz University, Department of Mathematics, College of Science and Arts, Jeddah, Kingdom of Saudi Arabia

Summary

The study deals with one techniques of the stochastic process types called Markov chains in estimating movement of diabetic nephropathy patients in type1 and type2 diabetes for three years from 2015 to 2017 for the stages of diabetic nephropathy (patients pass protein in their urine after intense physical activity, patients would have advanced condition of proteinuria, Patients are suffer from severe form of edema, Hypertension and their condition is not sensitive to diuretics, It is called renal failure uremia and patient's condition is critical) in Diabetes and Endocrinology Hospital to predict by the natural progression of Diabetic Nephropathy. The most important results movements of Diabetic Nephropathy patients represent Markov chains; there is a relationship between duration of infection with diabetes and diabetes kidney disease, for a long time period diabetic nephropathy patients can be infection by the renal failure, or damage kidney disease.

Key words:

Stationary distribution, Transition probability matrix, Diabetic nephropathy, Markov Chains.

1. Introduction

Diabetic nephropathy(DN) is one of most common complication of diabetes disease in spite of the availability of effective treatment it can be causes renal failure for a long times of infection with diabetes [1].

Diabetes is a chronic disease that costs countries a lot for health care and results from poor lifestyle choices. For example, spending on diabetes treatment alone will average A \$ 300 million a year over the next 30 years. Reducing these costs is done by changing healthy living habits in at-risk populations [2].

The diagnosis of the DN head is in the matured or forward stage so the focus is on illustrating the progress process to more advanced stages and helping patients to maintain a healthy lifestyle that may delay DN progression [3].

Some previous studies of the Markov modeling method and the use of the Bayesian rule to predict patients at high risk of Diabetic nephropathy based on their physiological criteria (blood pressure, sugar, urea) [4].

another study have found a relationship between the duration of diabetes and nephropathy.[5]

There are many studies show that patients with diabetes are more likely to develop the disease to other diseases compared to patients without diabetes.[6]

One of the most important study which related to our study used Markov model to the diabetes kidney disease clinical cycles. Therefore, from 1995 to 2015, predict future changes in the number of patients with diabetes nephropathy (DN) and end-stage kidney disease (ESRD).[7]

2. Study Problem and Solving

Modeling is an important statistical method. The calculated models are used to identify the future behavior of the phenomenon. The m problem of this study is to find a model to determine the distribution of Diabetic nephropathy in the four states ((loss little amount of protein in urine, loss more amount of protein in urine, Hypertension and Renal failure).in addition to determining the movements during this situation.

In order to achieve the objective of this study and solve the study problem the Markov chain and its methodology are used throwing to afford the natural progression of DN towards patients through estimating transition probabilities, prediction of probabilities in stationary case, prediction of reachable DN patients and prediction of expected total number of infection with renal failure during the period 2015-2017

3 Markov Chains

The applications of Markov Chains in medicine fields are quite common and have become standard tool of medical decision making. the probability of transition from state I to state j $P(X_{t+1} = j / X_t = i)$

$$\text{where: } P(X_{t+1} = j / X_t = i) = P_{ij} \quad (1)$$

this means the transition depend only on current state and independent from the previous states [8]

$$\sum_{j=1}^m P(X_{t+1} = j / X_t = i) = 1 \tag{2}$$

$$\sum_{j=1}^m P_{ij} = 1 \tag{3}$$

the transition probabilities was displayed as matrix T or Pij, where [9]:

$$T = [P_{ij}] = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1m} \\ P_{21} & P_{22} & \dots & P_{2m} \\ \dots & \dots & \dots & \dots \\ P_{m1} & P_{m2} & \dots & P_{mm} \end{pmatrix} \tag{4}$$

Any row of Pij matrix represent the probability distribution

$$\sum_i P_{ij} = 1 \tag{5}$$

In order to test that any stochastic process Xt represent M.C or not, we use χ^2 test, the step of test hypothesis [10]
 H0: The process doesn't represent M.C.
 H1: The processes represent M.C.

Test the statistic

$$Q = 2 \sum n_{ij} \ln \left[\frac{n_{ij}}{n_i \cdot n_j} \right] \tag{6}$$

Where:

- nij the no. of transition from state "i" to state "j"
- ni. the no. of transition from state "i" to any other state .
- n.j the no. of transition from any other state to state "j".

The tabulated value is $\chi^2_{(n-1), \alpha}$, The decision if $Q > \chi^2_{(n-1), \alpha}$ Reject H0, if $Q \leq \chi^2_{(n-1), \alpha}$ Accept H0

Stationary Distribution of Markov Chains

The Stationary probability distribution represent that the system state j in the long run or after large number of steps $n \rightarrow \infty$, if we have M.C with transition probability Pij then the probability vector "π" is called stationary distribution such that: [11] $\pi = \pi P$

$$(\pi_1 \ \pi_2 \ \dots \ \pi_m) = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1m} \\ P_{21} & P_{22} & \dots & P_{2m} \\ \dots & \dots & \dots & \dots \\ P_{m1} & P_{m2} & \dots & P_{mm} \end{pmatrix} \tag{7}$$

Canonical Form

Pij's canonical form, which means that the transient states must come first.. If the matrix includes m transient states and m-s recurrent states, then the canonical form as follows [12]:

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$$

where Q represent the (T) transition probability matrix
 R represent the $m \times (m-s)$ matrix for transition from transient states to other recurrent states,
 0 matrix represent $(m-s) \times m$ matrix with zero elements.

The matrix I is a $(m-s) \times (m-s)$ matrix for transition from recurrent states to another recurrent state .

The advantage of Q is to know the progression in DN.**

The Fundamental Matrix

After calculating the Q matrix we can obtained another matrix N where $N = (I-Q)^{-1}$ the matrix N called fundamental matrix for p.[13]

The sum of the ith row of the $(I-Q)^{-1}$ is s the time required for the system to stay in transient states until transferring to a recurrent state if the process begins in a transient state.[14]

4. Applied Aspect

4.1 Data Set

The data of this study have been collected from Diabetes and Endocrinology Hospital, department of statistics and information during period (2015-2017). It's a patients of DN in four cases which are (loss little amount of protein in urine, loss more amount of protein in urine, Hypertension and Renal failure).

4.2 Data Description

Table1: The descriptive statistics of key attributes recorded in this study.

Attributes	Mean	Std	Minimum	Maximum	C.V
Age	57	18.14	13	96	0.32
Duration	14	9.6	1	40	0.69
RBG	279.41	133.33	27	653	0.48
Urine	59.47	51.95	12	310	0.87
Creatinine	3.05	3.65	10	12.40	1.2

Table 2: Gender of patients

Gender	Count
Male	132
Female	78
Total	210

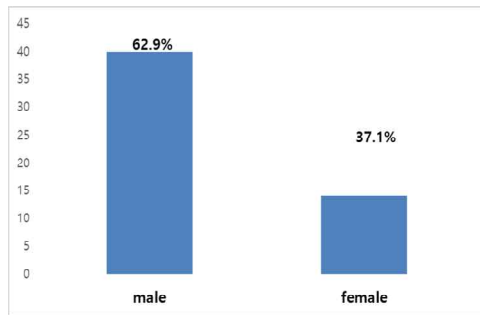


Fig. 1 Gender of patients

From the above Table 2 and Figure 2 we found that the male are more exposure to the kidney disease (or DN) than the female.

Table 3: patient organization according to gender (male, female) and four stage of DN.

Gender	DN patients stages				Total
	loss little protein	loss more protein	Hypertension	Renal failure	
Male	40	15	53	34	142
Female	14	8	20	17	59
Total	54	23	73	51	201

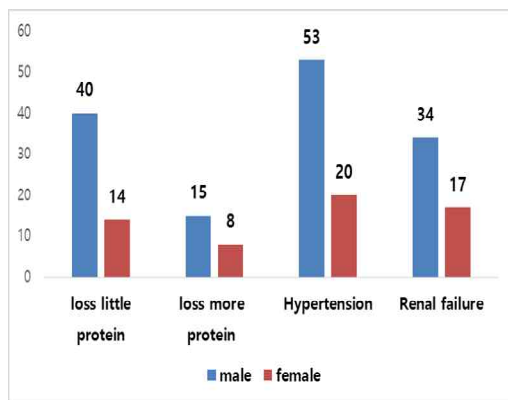


Fig. 2 patient representation according to gender (male, female) and four stage of DN .

Source: the researcher from applied study, Excel.
From the above table 3 and figure 3, we observed that at all the cases of DN patient the number of male more than the number of female.

4.3 Applied Markov Chains Analysis

- The four stages used in our study are
- L: Loss little amount of Protein in urine,
 - M: Loss More amount of Protein in urine,
 - H: Hypertension.
 - R: Renal Failure.

The model assumes a unidirectional Stream of progression from loss of Small protein NPDN to renal failure for the purpose of this analysis. It should be remembered, however, that patients can miss one or more stages to pass to advanced stages. We also did not observe patients regressing in this study (moving to previous stages) and no patient reported progressing to death in this study.

$$n = \begin{pmatrix} 54 & 0 & 12 & 0 \\ 0 & 23 & 5 & 0 \\ 0 & 0 & 73 & 18 \\ 0 & 0 & 0 & 51 \end{pmatrix}$$

First we Test if the data represent Markov Chain or not:

We must test the following hypothesis:

H0: Xn doesn't represent Markov Chain.

H1: Xn represent Markov Chain.

Test the statistic:

We applied the data of the above matrix on the equation (6), that implies the following:

$$Q = 7$$

$\chi^2(n-1)2, \alpha, \chi^2(4-1)2, 0.05 = \chi^2 9, 0.05 = 3.325$ Shows value Q test (7) is significant, Reject H0 and accept H1 that means The Process (number of DN Patients) Represent M.C.

Initial probability distribution

$$\pi^{(0)} = (\pi_1^{(0)}, \pi_2^{(0)}, \pi_3^{(0)}, \pi_4^{(0)})$$

$$LLP = \pi_1^{(0)} = 66/236 = 0.28$$

$$LMP = \pi_2^{(0)} = 28/236 = 0.12$$

$$HT = \pi_3^{(0)} = 91/236 = 0.38$$

$$RF = \pi_4^{(0)} = 51/236 = 0.22$$

Then the initial probability distribution will be write as the following form:

$$\pi^{(0)} = (0.28 \quad 0.12 \quad 0.38 \quad 0.22)$$

The transition probability matrix

Here we have one absorbing state (Renal Failure [RF])

The reachable of LLP, LMP and HT:

$$LLP \rightarrow LLP = 54/66 = 0.8$$

$$LLP \rightarrow LMP = 0/66 = 0$$

$$LLP \rightarrow HT = 12/66 = 0.2$$

$$LLP \rightarrow RF = 0/66 = 0$$

$$LMP \rightarrow LLP = 0/28 = 0$$

$$LMP \rightarrow LMP = 23/28 = 0.82$$

$$LMP \rightarrow HT = 5/28 = 0.18$$

$$LMP \rightarrow RF = 0/28 = 0$$

$$HT \rightarrow LLP = 0/91 = 0$$

$$HT \rightarrow LMP = 0/91 = 0$$

$$HT \rightarrow HT = 73/91 = 0.8$$

$$HT \rightarrow RF = 18/91 = 0.2$$

Then p matrix can be represented as follows:

$$P = \begin{matrix} LLP & LMP & HT & RF \\ \begin{pmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0.82 & 0.18 & 0 \\ 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Stationary Distribution of Markov Chain

$$\pi = \pi P$$

$$\pi = P^{(0)} = (0.28 \quad 0.12 \quad 0.38 \quad 0.22)$$

Then:

$$\begin{aligned} \pi &= \pi P = \\ &= (0.28 \quad 0.12 \quad 0.38 \quad 0.22) \begin{pmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0.82 & 0.18 & 0 \\ 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= (0.224 \quad 0.098 \quad 0.382 \quad 0.296) \end{aligned}$$

After n-steps ($n \rightarrow \infty$) the probability of diabetic nephropathy patients LLP is 0.224, LMP is 0.098, HT is 0.382, and RF is 0.296.

Canonical Form

$$P = \begin{pmatrix} Q & R \\ 0 & 1 \end{pmatrix} \quad Q = \begin{pmatrix} 0.8 & 0 & 0.2 \\ 0 & 0.82 & 0.18 \\ 0 & 0 & 0.8 \end{pmatrix}$$

$$(I - Q)^{-1} = \begin{pmatrix} 5 & 0 & 5 \\ 0 & 5.6 & 5 \\ 0 & 0 & 5 \end{pmatrix}$$

by finding the sum of each row of the above matrix we will get a vertical vector as follows:

$$\sum_{i=1}^3 (I - Q)^{-1} = \begin{pmatrix} 10 \\ 10.6 \\ 5 \end{pmatrix}$$

5. Results

From the applied Markov Chain on number of DN patients in the Diabetes and Endocrinology Hospital at Khartoum Bahri We modeled the transition of 210 patients from NPDR to Renal Failure on an annual basis. We acquire the following result

Depending on the stationary process, it can be said that in the long term for any group of patients, the distribution of the patient among the four cases will be as follows:22.4%: Loss little amount of Protein in urine ,9.8%: Loss More

amount of Protein in urine,38.2%: Hypertension,29.6%: Renal Failure.

- Patients stay for long time periods in Hypertension stage before transitioning to renal failure.
- At least the diabetes patient spend 5 years after that it can be able to infection with diabetic nephropathy if that patient has some complications or reasons causes diabetic nephropathy.

6. Conclusion

This study was based on the use of Markov chains in the analysis of the behavioral change of diabetic patients. Through the analysis, very good results were obtained which could be used in the awareness work of the patients This study provides interested individuals with an excellent database through which to prevent the spread of this disease, as well as work to stop the development and transition to dangerous situations.

References

- [1] Lim, A. K. (2014). Diabetic nephropathy–complications and treatment. International journal of nephrology and renovascular disease, 7, 361.
- [2] Knezevic, M. (2009). Estimating the long-term costs of diabetic kidney disease: an economic approach. Applied Economics Letters, 16(10), 1059-1064.
- [3] Hasslacher, C. (Ed.). (2001). Diabetic nephropathy (Vol. 6). John Wiley & Sons..
- [4] Sasikumar, R., & Kalpana, V. Applications Perspective Predictions on Markov Modelling in Medical Diagnosis Systems.
- [5] Srikanth, P. (2015). Using Markov chains to predict the natural progression of diabetic retinopathy. International journal of ophthalmology, 8(1), 132.
- [6] Begun, A., Icks, A., Waldeyer, R., Landwehr, S., Koch, M., & Giani, G. (2013). Identification of a multistate continuous-time nonhomogeneous markov chain model for patients with decreased renal function. Medical Decision Making, 33(2), 298-306.
- [7] Murakami, Y., & Ohashi, Y. (2001). Projected number of diabetic renal disease patients among insulin-dependent diabetes mellitus children in Japan using a Markov model with probabilistic sensitivity analysis. International journal of epidemiology, 30(5), 1078-1083.
- [8] Ross, S. M. (2014). Introduction to probability models Academic press.
- [9] Parzen, E. (1960). Modern probability theory and its applications. John Wiley & Sons, Incorporated.
- [10] Medhi, J. (2002). Stochastic models in queueing theory. Elsevier.:
- [11] Hiscott, R. N. (1981). Chi-square tests for Markov chain analysis. Journal of the International Association for Mathematical Geology, 13(1), 69-80.:

- [12] Chung, K. L. (1960). Transition probabilities. In Markov Chains with Stationary Transition Probabilities (pp. 5-11). Springer, Berlin, Heidelberg.
- [13] Brémaud, P. (2013). Markov chains: Gibbs fields, Monte Carlo simulation, and queues (Vol. 31). Springer Science & Business Media.
- [14] Bailey, N. T. (1990). The elements of stochastic processes with applications to the natural sciences (Vol. 25). John Wiley & Sons.



Mubarak Hassan Mubarak Elhafian

graduated with a Bachelor of Science (Applied Statistics) from Sudan University of Science and Technology, Department of Statistics, College of Science Sudan. He obtained his MSc Degree in Statistics from Sudan University of Science and Technology, Sudan. His received PhD degree in Statistics from Sudan University

of Science and Technology, Sudan. He is currently work as Assistant Professor in King Abdulaziz University, Department of Mathematics, College of Science and Arts, Jeddah, Kingdom of Saudi Arabia. His current research interest includes Multivariate Analysis, Time Series Analysis, and Stochastic Process



Mohammed Elammeen Eissa Qurashi

graduated with a Bachelor of Science (Applied Statistics) from Sudan University of Science and Technology, Department of Statistics, College of Science Sudan. He obtained his MSc Degree in Statistics from Sudan University of Science and Technology, Sudan. His received PhD degree in Statistics from Sudan University

of Science and Technology, Sudan. He is currently work as Assistant Professor in Statistics from Sudan University of Science and Technology, Applied Statistics Department, College of Science, Jeddah, Sudan. His current research interest includes Multivariate Analysis, Time Series Analysis, and Stochastic Process.



Abdellateef Khalifa Hamid Ali

graduated with a Bachelor of Science (Statistics) from Shendi University, Department of Statistics, College of Science and Technology, Sudan. He obtained his High Diploma of Science in Statistics from Shendi University. He obtained his MSc Degree in Statistics from Sudan University of Science and Technology, Sudan. He is currently work

as Lecturer in Shendi University, Statistics Department, College of Science and Technology, Sudan. His current research interest includes, Stochastic Process.