ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2018-62156

A Multi-State Markov Model for the Progression of Chronic Kidney Disease

Kronik Böbrek Hastalığının Progresyonunda Çok Durumlu Markov Modeli

Gurprit GROVER[a],
Alka SABHARWAL[b],
Shrawan KUMAR[b],
Arpan Kumar THAKUR[a]

[a]Department of Statistics,
University of Delhi,
[b]Department of Statistics,
Kirori Mal College,
University of Delhi,
India

Received: 19.07.2018
Received in revised form: 06.12.2018
Accepted: 13.12.2018
Available Online: 19.04.2019

Correspondence:
Shrawan KUMAR
Department of Statistics,
Kirori Mal College,
University of Delhi,
INDIA/HİNDİSTAN
shrawan.kmc@gmail.com

ABSTRACT Objective: The main goal of this study is to develop a stochastic model for the progression of Chronic Kidney Disease (CKD) into different stages based on estimated Glomerular Filtration Rate (eGFR). **Material and Methods:** The present study is a retrospective study of 117 patients suffering from CKD during the period March 2006 to October 2016. The prognostic factors such as gender, age, body mass index, diabetes, hypertension, hemoglobin, urea, serum creatinine, albumin and duration of the disease were recorded for each patient. We have applied the continuous time homogeneous multistate model based on Markov processes. The deterioration of disease is continuous in time and the probability of transition from one state to another state depends on the length of time and is independent of time on which transition takes place. Also, Cox proportional hazard model has been used to examine the effects of prognostic factors on the transition rates. **Results:** The probabilities of staying in the same state in first five years i.e. stage 1, stage2, stage 3, stage 4 and stage 5 are 0.6126, 0.5508, 0.5631, 0.0596 and 1 respectively. The probabilities of moving to the next state are also computed for first five and ten years. The prognostic factors age, hypertension, diabetes, hemoglobin, urea, serum creatinine are significant factors for the progression of CKD into different stages. **Conclusions:** The mean sojourn times along with p-next probabilities provide more intuitive parametric information of continuous time multistate model based on Markov processes than crude transition intensities.

Keywords: Chronic kidney disease; multi-state markov model; end stage renal disease; glomerular filtration rate; transition intensity

ÖZET Amaç: Bu çalışmanın amacı tahmin edilen Glomerüler filtrasyon hızına (eGFR) bağlı olarak kronik böbrek hastalığının (CKD) farklı evrelerdeki progresyonu için stokastik model geliştirmektir. Yöntem: Bu çalışma Mart 2006 - Ekim 2016 döneminde CKD geçiren 117 hastaya ait retrospektif bir çalışmadır. Her bir hastaya ait cinsiyet, yaş, beden kitle indeksi, diyabet, hipertansiyon, üre, serum kreatinin, albümin ve hastalık süresi gibi prognostik faktörler kaydedilmiştir. Markov süreçlerine göre sürekli zamanlı homojen çok durumlu model uyguladık. Zamana bağlı olarak hastalık sürekli kötüleşmekte ve durumlar arası geçiş olasılığı geçişin gerçekleştiği zamandan bağımsızdır. Ayrıca geçiş hızlarındaki prognostik faktör etkilerini incelemek için Cox orantısal hazard modeli kullanılmıştır. **Bulgular:** İlk beş yılda aynı evrede kalma olasılıkları örneğin evre 1, evre 2, evre 3, evre 4 ve evre 5 sırasıyla 0.6126, 0.5508, 0.5631, 0.0596 ve 1'dir. Bir sonraki duruma ilerleme olasılıkları da ilk beş ve on yıl için hesaplanmıştır. CKD progresyonunun farklı evrelerinde yaş, hipertansiyon, diyabet, hemoglobin, üre ve serum kreatinin anlamlı faktörler olarak bulunmuştur. **Sonuç:** p-sonraki olasılıklarla birlikte ortalama konukluk süreleri, Markov sürecine dayalı sürekli zaman çok durumlu modelinin sezgisel parametrik bilgisini ham geçiş yoğunluğundan daha fazla sağlamaktadır.

Anahtar Kelimeler: Kronik böbrek hastalığı; Çok durumlu Markov modeli; Son dönem böbrek hastalığı; Glomerüler filtrasyon hızı; Geçiş yoğunluğu

The non-communicable and chronic diseases are one of the leading causes of the morbidity and mortality all over the world.[1] The word chronic means ongoing, persistent and long-standing. The kidney disease is characterized by its inability to filter blood properly, inability to eliminate wastes effectively, not able to balance fluids, non-adjustment of minerals and inability to azzzctivate vitamin D in the body. The kidney diseases are of two types: Acute Kidney Injury (AKI) and Chronic Kidney Disease (CKD). AKI is the sudden abrupt loss of kidney functions whereas CKD is a progressive decline in kidney functions over months or years. There is a decrease in excretory, endocrine and metabolic function in most of CKD patients. The complications associated with CKD are drug toxicity, metabolicz and endocrine complications, and enhanced risk for Cardio Vascular Disease (CVD), infection, frailty and cognitive impairment.[2] CKD tends to worsen over time. Hence, the risk of adverse outcomes and disease severity increases over time. The adverse outcomes of CKD such as kidney disease, cardiovascular disease, premature death and progression into severe stages can be prevented or delayed by early intervention and take appropriate measures consisting of medications and lifestyle changes. If a person is diagnosed with CKD early enough, it may help in inhibiting the disease so that it never debilitates the person. The worldwide prevalence of CKD is varying between 10.5% and 13.1%.[3] In Western countries, 2/3rd cases of CKD are due to diabetes and hypertension.[4] The prevalence of CKD in India is approximately 800 per million populations.[5] Approximately 60% CKD patients have diabetes and hypertension. The prevalence of CKD is increasing rapidly with the rising prevalence of diabetes and hypertension. By 2030, India will have the highest population of diabetics in the world.[6] Gender, advancing age, abdominal obesity diabetes, hypertension and smoking have a significant relationship with chronic kidney disease.[7] Usually, kidney function is determined by Glomerular Filtration Rate (GFR). GFR is generally considered as the best overall index of kidney function. GFR is computed on the basis of gender, age, community and creatinine levels in blood. CKD is defined as either kidney damage or GFR < 60 ml/min/1.73m^2 for more than 3 months.[8] Kidney damage is defined as pathological abnormalities or markers of damage including abnormalities in blood and urine tests or in imaging studies. The different stages of CKD are irreversible and represent the degree of deterioration of kidney function in nature as defined by Kidney Disease Outcomes Quality Initiative (KDOQI)are as follows:[8]

Staging of Chronic Kidney Disease		
STAGE	**GFR**	**DESCRIPTION**
1	≥ 90	Kidney damage with normal GFR
2	60 – 89	Kidney damage with mild reduction in GFR
3	30-59	Moderate reduction in GFR
4	15-29	Severe reduction in GFR
5	<15	Kidney failure

GRF: Glomerular Filtration Rate.

The risk factors for kidney disease depend on individual's genetic and phenotypic make-up. The prominent risk factors for chronic kidney disease are race, gender, age and family history of an individual. For example a person of African-American origin, older age, low birth weight and having a family history of kidney disease is at more risk for CKD.[9] CKD is more common in men than women.[10] People with age between 40 to 60 years are at higher risk for CKD.[10] Smoking, obesity, hypertension, and diabetes mellitus are some other important risk factors for kidney disease. Hypertension and / or uncontrolled diabetes will hasten the progress of chronic kidney disease and lead to end stage renal disease very easily and quickly. Other factors for CKD are exposure to heavy metals, excessive consumption of alcohol, and use of analgesic medications, history of cardiovascular disease, metabolic syndrome, HIV infection, hepatitis C virus and malignancy. CKD progresses gradually and is usually asymptomatic in its earlier stages. Early stages

are diagnosed by a urine test, blood test and imaging. Symptoms develop slowly and are not specific to the disease. Some people have no symptoms at all and are diagnosed by a lab test. Generally, 50% of the patients with advanced stage of CKD are diagnosed for the first time. Flank pain, sleep problems, nausea, and vomiting, metallic taste in the mouth, loss of appetite, numbness or tingling in the toes or fingers edema of extremities, fatigue and dark orange, brown, tea-colored urine are some of the common symptoms associated with stage 4 of CKD. AKI hasten the progression of CKD to end stage renal disease (ESRD).[11] Early stage detection is more amenable to treatment, thereby reducing the economic burden on CKD and reducing the mortality from the disease. Systematic screening is recommended for early detection.

A multi-state Markov model is a well established and widely used model for describing an individual's movement between a series of states in continuous time. It is used to compute the transition probabilities from any state *i* to any other state *j*. We have used this model to compute the transition probabilities from lower state to higher state as chronic kidney disease is irreversible in nature. We have also computed the mean time for remaining in a particular state. The state (stage) of chronic kidney disease indicates the health condition of the kidney. The five stages of CKD are stage 1, stage 2, stage 3, stage 4 and stage 5 and they reflect the degree of severity of the disease. To the best of our knowledge, multistate Markov modeling for the progression of CKD has not been done in India.

MATERIAL AND METHODS

DATA DESCRIPTION

Different registered laboratories and hospitals were approached to obtain the records of CKD patients (hospitalized and non-hospitalized). Almost 550 patients were approached for sharing their data. Only 248 patients responded positively and were ready to share their data. Informed consent was taken from all of them. The data of only 117 patients were found to be in accordance with the present study. The present study is a retrospective study of 117 patients suffering from CKD during the period March 2006 to October 2016.

The CKD patients with co -morbidities like cardiovascular disease, HIV, Pneumonia, obstructive respiratory disease were excluded from the study on account of their insufficient number. The vital information such as gender, age, body mass index was noted for each patient. Also the information related to diabetes, hypertension, hemoglobin, urea, serum creatinine, albumin and history of the disease were recorded for each patient. The patients were under medical supervision. For the present study, WHO or Internationally accepted standards were taken as a reference value for the categorical variables as well as continuous variables sex, hypertension and diabetes are taken as categorical variables. SBP > 130mmHg and DBP > 90mmHg for hypertension (HTN), FPG≥ 126 mg/dl for diabetes. The continuous variables are age, body mass index (BMI), hemoglobin (Hb), serum creatinine (Cr), urea (Urea), albumin (Alb). Serum creatinine ≥ 1.4mg/dl for the onset of nephropathy and control limits for albumin in blood are 4.0 to 5.4 g/dl and hemoglobin for men are 13.5 to 17.5 g/dl and for female are 12.0 to 15.5 g/dl, urea is 42.02 to 131.13 mg/dl.

DEVELOPMENT OF MODELS

The five stages of CKD are stage 1, stage 2, stage 3, stage 4 and stage 5. All the stages 1 to 4 are transient states and stage 5 is an absorbing state. A patient may make forward transition only among different transient states continuously. The arrows show the possible transition between stages. The time of movement between the different states and the state occupancy in between the observation times are not known. The severity of CKD determines the frequency of visits to the doctor. Numbers of visits of CKD patients with stage 4 is more than the numbers of visits of stage 1, stage 2 and stage 3 patients of CKD.

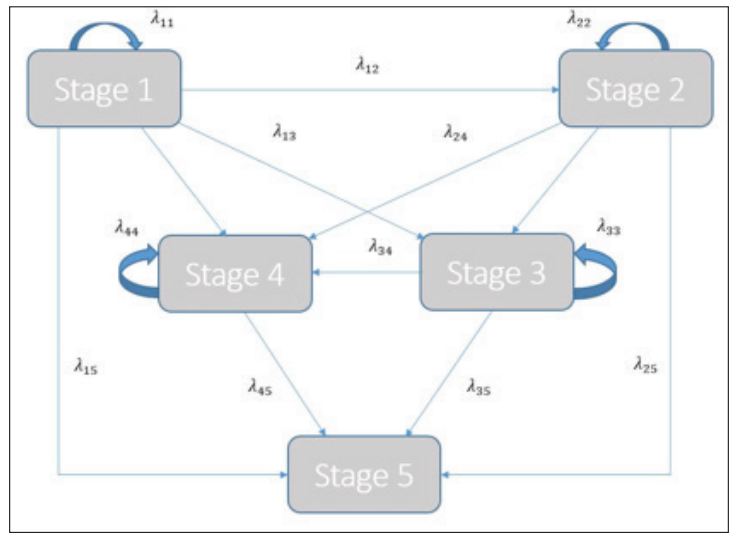

FIGURE 1: State Transition Diagram.

Jackson developed a general model consisting of a number of transient states and an absorbing state for the progression of chronic diseases.[12] The multi-state models based on Markov processes are extensively used to model the progression of disease involving a number of movements of an individual to various states in continuous time.[13] The progression of CKD is often described by stages of severity. The progression of CKD into different stages is irreversible in nature as shown in the Figure 1. The progression of CKD into different stages is continuous in time whereas the state spaces are discrete in nature. We have considered the continuous time homogeneous multistate model based on Markov processes as deterioration of disease is continuous in time and the probability of transition from one state to another state depends on the length of time and as such is independent of time on which transition takes place.[14]

Let $S(t) = k$ be the state of a patient at a time t then the transition intensity with which the patient moves to the state l during the time interval $(t+\delta t)$ is given by

$$\lambda_{kl} = \lim_{\delta t \to 0} \frac{P(S(t+\delta t) = l \mid S(t) = k)}{\delta t}.$$

The schematic diagram (Figure 1) for the progression of CKD suggests the following transition intensity matrix:

$$Q = \begin{pmatrix} -(\lambda_{12}+\lambda_{13}+\lambda_{14}+\lambda_{15}) & \lambda_{12} & \lambda_{13} & \lambda_{14} & \lambda_{15} \\ 0 & -(\lambda_{23}+\lambda_{24}+\lambda_{25}) & \lambda_{23} & \lambda_{24} & \lambda_{23} \\ 0 & 0 & -(\lambda_{34}+\lambda_{35}) & \lambda_{34} & \lambda_{35} \\ 0 & 0 & 0 & -\lambda_{45} & \lambda_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Q is a 5 X 5 matrix with transition elements λ_{kl} (k = 1,2,...,5 and l = 1,2,...,5). λ_{kl} denotes the transition intensity or instantaneous rate of transition from state k to state l. λ_{kl} (k = 1,2,...,5 and l = 1,2,...,5) are the model parameters and they control the movement to next possible state. The two properties of the transition intensity matrix are (i) the sum of elements of each row of the transition matrix Q is zero and (ii) diagonal element of Q is

$\lambda_{kk} = -\sum_{k \neq l} \lambda_{kl}$.

Let P(s,t) denotes the transition probability matrix with transition probabilities $p_{kl}(s,t)$, $k = 1,2,...,5$ and l= 1,2,...,5, as its elements. Where

$p_{kl}(s,t)$ = Pr {an individual in state k at a time s will be in state l at the time t}.

The transition matrix is given as

$$P(s,t)=\begin{pmatrix} p_{11}(s,t) & p_{12}(s,t) & p_{13}(s,t) & p_{14}(s,t) & p_{15}(s,t) \\ p_{21}(s,t) & p_{22}(s,t) & p_{23}(s,t) & p_{24}(s,t) & p_{25}(s,t) \\ p_{31}(s,t) & p_{32}(s,t) & p_{33}(s,t) & p_{34}(s,t) & p_{35}(s,t) \\ p_{41}(s,t) & p_{42}(s,t) & p_{43}(s,t) & p_{44}(s,t) & p_{45}(s,t) \\ p_{51}(s,t) & p_{52}(s,t) & p_{53}(s,t) & p_{54}(s,t) & p_{55}(s,t) \end{pmatrix}$$

The progression of CKD into different stages is irreversible and hence $p_{kl}(s,t) = 0$ for all $k > l$. Stage 5 of CKD is an absorbing state and therefore $p_{55}(s,t) = 1$. The sum of the probabilities in each row of the transition probability matrix equals one. A continuous time multistate model based on Markov process is completely specified if the transition intensity between different states are known or transition probability matrix is completely specified along with the expected sojourn time of each state.[15] The average period in a single stay in each transient state before moving to a higher state is called mean Sojourn time. In the case of CKD, the expected duration of a single stay in a state or mean sojourn time is equal to the total length of stay in that state as the progression of CKD is irreversible. In the Markov process, there is an assumption that the future transition of an individual depends on the current state only and is independent of his transition in the past. In the case of time-homogeneous continuous-time Markov model, the Sojourn time in the state k has an exponential distribution with the rate $-\lambda_{kk}$. The probability of an individual moving to the next state l from the current state k is given by $-\frac{\lambda_{kl}}{\lambda_{kk}}$. The other elements of the k^{th} row of the transition matrix are proportional to $-\frac{\lambda_{kl}}{\lambda_{kk}}$. Forward Kolmogorov differential equations can model the progression of CKD.

$$\frac{\partial}{\partial t} P(s,t) = P(s,t).Q$$

With the initial condition

$$P(s,s) = I$$

Where I is the identity matrix. The solution of the system of forward differential equations provides the transition probabilities in terms of transition intensities. The solution is

$$p_{11}(s,t) = e^{\lambda_{11}(t-s)}$$

$$p_{12}(s,t) = \frac{\lambda_{12}}{\lambda_{22} - \lambda_{11}}(e^{\lambda_{11}(t-s)} - e^{\lambda_{22}(t-s)})$$

.

.

.

$$p_{44}(s,t) = e^{\lambda_{44}(t-s)}$$

$$p_{45}(s,t) = 1 - e^{\lambda_{44}(t-s)}$$

$$p_{55}(s,t) = 1.$$

The likelihood function is formed using transition probabilities in terms of transition intensities. The likelihood estimates of model parameters are obtained from likelihood function. A general method for evaluation of likelihood of a general Multi-state Markov model in continuous time has been explained by Kalbfleisch and Lawless.[16] Kay had used a Markov model for analyzing cancer markers and disease states in survival studies and established that it can be used for any form of the transition matrix.[17] The matrix exponential of scaled transition matrix provides the transition matrix $P(t) = Exp(tQ)$ where, $P(t) = P(s,t)$ Cox and Miller.[18] The exponential of matrix is defined by matrix products in the power series. Each element of probability transition matrix can be represented in terms of the elements of transition intensity matrix. For simple model, it can be done analytically or by making use of symbolic algebra software such as Mathematica. When eigen values are distinct, the msm package for R uses eigen system decomposition and in case of repeated eigen values, msm package uses the method of Pade approximants.[19] The likelihood function of transition intensities is the product of probabilities of transition between observed states over all individuals and observation times. Let $S_i(t)$ denotes the state of each individual i=1,...,M and is known for all the times t of the entire period of study.

$$\text{L(Q)}=\prod_i L_i = \prod_{i,j} L_{i,j} = \prod_{i,j} P_{S(t_{ij})S(t_{i,j+1})}(t_{i,j+1} - t_{ij})$$

Where $L_{i,j}$ is the element corresponding to $S(t_{ij})^{\text{th}}$ row and $S(t_{i,j+1})^{\text{th}}$ column of the transition matrix $P(t)$ evaluated at $t = t_{i,j+1} - t_{ij}$. The likelihood L(Q) is maximized in term of $\log(\lambda_{kl})$ therefore the estimates and standard error of λ_{kl} are obtained from $\log(\lambda_{kl})$ using optimization technique. The msm package for R computes transition matrix analytically. In multi-state model, our interest lies in computing the changes in transition rates in the presence of covariates. The effect of explanatory variables on particular transition intensity can be computed by modeling the transition intensity as a function of these covariates. In such case the new transition matrix Q is used in the likelihood function for estimating the transition intensities. Marshall and Jones used the proportional hazards model for studying the effect of a vector of explanatory variables on transition intensity for the individual i at a time j by replacing the transition intensity element λ_{kl} by $\lambda_{kl}(z_{ij}) = \lambda_{kl}^{(0)} \exp(\beta_{kl}^T z_{ij})$.[20] The msm package has been used to fit the individual specific covariates to transition intensities.

MODEL COMPARISON CRITERIA

The various stochastic models for progression of CKD stages have been compared on the basis of the Akaike Information Criterion (AIC) value. The quality of statistical model relative to other model is measured by AIC value and is computed as $AIC = 2K - 2\ln(\hat{L})$, where K denotes the number of unknown parameters in the model and $\hat{L}$ denotes the maximum value of the likelihood function of the model. The model with the minimum value of AIC is preferred over other models. The second criterion for the model selection is based on likelihood ratio test. It is a statistical test for comparing the goodness of fit of two statistical models, in which the null model is a particular case of alternative model. It is based on the likelihood ratio

indicating how many times more likely the data are less than one model than the other. Log likelihood ratio statistic (LR statistic) is used for testing of hypothesis. LR statistic is given by,

$$LR\ Statistic = -2(\ln \text{ reduced model}) - (-2(\ln \text{ current model}))$$

A larger value of $LR\ Statistic\ (-2\ln L)$ is in favour of the current model. Larger value leads to a smaller p value which is an indication against the reduced model in favour of current model. For a large number of observations, $LR\ Statistic$ follows a chi-square distribution with K degrees of freedom, where K is the number of predictors being assessed.

RESULTS

It is evident from Table 1 (a); there were 66 males and 51 females in the recorded data set of 117 CKD patients. The number of patients in stage 1, stage 2, stage 3 and stage 4 were 17, 20, 37 and 43 respectively. The numbers of patients with hypertension were75 and the numbers of patients who have diabetes were 62. There is a significant relationship between stages of CKD and prognostic factors HTN and Diabetes.

Table 1 (b) shows the descriptive statistics for continuous variables. The minimum and maximum age of the patient was 18 years and 78 years respectively. The mean age of the patients was 48.08 years with a standard deviation of 14.0 years. The median age of the patients was 49 years. Similar descriptive statistics are computed for other variables and are summarized in Table 1(b).

Table 2 shows the state transition of CKD patients in their subsequent visits. A number of visits of CKD patients depend on the criticality of disease. CKD patients with stage 4 will visit more frequently than CKD patients in stage 1, stage 2 or stage 3. Multistate data has been summarized by counting the transition of stages for each patient in their subsequent visits. The total number of transitions of stage 1 of CKD patient to stage 1 is 75 for all the patients under study.

It means that there were 75 instances when a patient of stage 1 remained in stage 1 in his subsequent visits. A number of transitions to stage 5 from stage1, stage 2, stage 3 and stage 4 are 0, 1, 2 and 51 respectively. Reverse (backward) transition is not possible so none of the patients had moved from a higher stage to lower stage in the subsequent visit. The numbers of transitions at diagonal places are very high as the progression of CKD is very slow. The numbers of transitions from stage 4 to stage 5 is highest in compari-

TABLE 1 (a): The results of Chi-square analysis for categorical variables.

Variables			Stages				Chi-square	p value
			Stage 1	Stage 2	Stage 3	Stage 4		
Sex	Male	Count	8	13	21	24	1.213	0.750
		percentage	(12.1%)	(19.7%)	(31.8%)	(36.4%)		
	Female	Count	9	7	16	19		
		percentage	(17.6%)	(13.7%)	(31.4%)	(37.3%)		
HTN	No	Count	11	6	14	11	8.483	0.037
		percentage	(26.2%)	(14.3%)	(33.3%)	(26.2%)		
	Yes	Count	6	14	23	32		
		percentage	(8%)	(18.7%)	(30.7%)	(42.7%)		
Diabetes	No	Count	17	5	15	18	24.131	< 0.001
		percentage	(30.9%)	(9.1%)	(27.3%)	(32.7%)		
	Yes	Count	0	15	22	25		
		percentage	(0%)	(24.2%)	(35.5%)	(40.3%)		
Total		Count	17	20	37	43		
		percentage	(14.5%)	(17.1%)	(31.6%)	(36.8%)		

HTN: Hypertension.

TABLE 1 (b): Descriptive statistics for continuous variables.

Stage	N		Age	BMI	Hb	Urea	Cr	Alb
		Mean	47.47	22.0324	11.6588	120.0588	1.7529	5.0276
		Standard Deviation	10.961	3.97205	1.64395	11.46446	.35243	.56237
Stage 1	17	Median	47.00	20.6000	11.5000	124.0000	1.7500	4.8000
		Minimum	23	17.80	8.70	98.00	1.20	4.10
		Maximum	63	31.00	13.80	137.00	2.30	5.90
		Mean	38.35	23.7430	12.3450	54.7000	1.6850	3.8750
		Standard Deviation	15.069	5.05893	1.47772	15.05114	.37874	.70103
Stage 2	20	Median	39.50	22.1500	12.3500	51.5000	1.6000	3.9000
		Minimum	18	16.56	9.80	38.00	1.20	2.40
		Maximum	67	34.90	16.60	86.00	2.50	5.60
		Mean	47.49	23.9297	10.8297	76.9459	2.4730	3.4973
		Standard Deviation	13.291	5.13257	1.80029	20.57634	.74746	.70769
Stage 3	37	Median	47.00	23.2000	10.9000	77.0000	2.3000	3.6000
		Minimum	22	11.70	6.90	39.00	1.30	2.20
		Maximum	72	36.20	13.90	145.00	4.10	4.80
		Mean	53.35	23.9786	9.2279	113.2791	4.1953	3.1395
		Standard Deviation	12.958	4.23011	2.03637	23.80709	.90972	.52651
Stage 4	43	Median	55.00	23.3000	9.0000	113.0000	4.3000	3.1000
		Minimum	25	16.50	5.60	66.00	2.30	2.30
		Maximum	78	33.10	13.30	158.00	6.20	4.60
		Mean	48.08	23.6401	10.6205	92.7607	2.8667	3.6527
		Standard Deviation	14.006	4.63381	2.15437	31.41097	1.28064	.87796
Total	117	Median	49.00	22.8000	10.7000	92.0000	2.5000	3.5000
		Minimum	18	11.70	5.60	38.00	1.20	2.20
		Maximum	78	36.20	16.60	158.00	6.20	5.90

BMI: Body mass index, Hb: Hemoglobin, Cr: Serum creatinine, Alb: Albumin.

TABLE 2: Number of state transitions.

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Stage 1	75	8	4	3	0
Stage 2	0	116	6	3	1
Stage 3	0	0	293	29	2
Stage 4	0	0	0	145	51

TABLE 3: Estimated transition intensities.

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Stage 1	-0.098	0.0599	0.0266	0.0115	0
Stage 2	0	-0.119	0.0725	0.0467	0
Stage 3	0	0	-0.115	0.115	0
Stage 4	0	0	0	-0.564	0.564

son to transition from other stages to stage 5 as the deterioration of CKD in stage 4 is very fast compared to other stages of CKD.

The likelihood estimates of transition intensities between various stages are shown in the above Table 3. The transition intensity of moving from stage 1 to stage 1, i.e. retaining the stage 1 is –0.098 whereas the transition intensity of transition from stage 1 to stage 2 is 0.0599. The transition intensities are zero for backward movement i.e. transition from a higher stage to a lower stage is not possible. This proves that the progression of CKD is irreversible.

The transition probability matrix over the period of 1 year, 5 years and 10 years after fitting the continuous multistate Markov model has been shown in Table 4. The transition probability matrix for one year period shows that a chronic kidney patient remains in the same stage with very high probability. It indicates that the progression of ckd is very slow except for stage 4. The table suggests that in 5 years' time a patient of CKD with stage 1 will remain in stage 1 with probability 0.61. The probabilities of his moving from stage 1 to stage2, stage 3, stage 4 and stage 5 are 0.174, 0.109, 0.040 and 0.064 respectively. It is more likely that the CKD patient of stage 1 will remain in stage 1 over five years of time than moving to higher stages. The probability of moving from stage 1 to stage 5 in 10 years is more in comparison to 5 year time. The probability of a CKD patient moving from a higher stage to a lower stage is zero for all time periods. This is because of the fact that CKD is irreversible in nature. Similar interpretations can be derived for the other elements of the transition probability matrix.

Table 5 shows the mean sojourn times of CKD patients in various stages. Mean sojourn time for stage 1 is highest. Stage 2 and stage 3 also have considerably high sojourn time. This proves the slow movement of progression of CKD in earlier stages. The mean sojourn time for stage 4 1.1733 years which is quite low in comparison to other stages. This suggests the kidney failure within 2 years. The standard error along with 95 % confidence interval for mean sojourn time for each stage is displayed in Table 5.

TABLE 4: Estimated transition probabilities after one year, five years and ten years.

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
after one year					
Stage 1	0.9067	0.0538	0.0258	0.0106	0.0031
Stage 2	0	0.8876	0.0645	0.0367	0.0112
Stage 3	0	0	0.8915	0.0825	0.0260
Stage 4	0	0	0	0.6689	0.3311
Stage 5	0	0	0	0	1
after five years					
Stage 1	0.6126	0.1741	0.1094	0.0401	0.0638
Stage 2	0	0.5508	0.2019	0.0828	0.1645
Stage 3	0	0	0.5631	0.1288	0.3082
Stage 4	0	0	0	0.0596	0.9404
Stage 5	0	0	0	0	1
after ten years					
Stage 1	0.3753	0.2026	0.1638	0.0554	0.2029
Stage 2	0	0.3033	0.2249	0.0765	0.3952
Stage 3	0	0	0.3170	0.0802	0.6028
Stage 4	0	0	0	0.0036	0.9964
Stage 5	0	0	0	0	1

TABLE 5: Mean sojourn times at different stages.

	Estimates	SE	CI
Stage 1	10.2030	2.6395	(6.1450, 16.9407)
Stage 2	8.3829	2.6501	(4.5112, 15.5772)
Stage 3	8.7051	1.5645	(6.1206, 12.3808)
Stage 4	1.7734	0.2456	(1.3517, 2.3265)

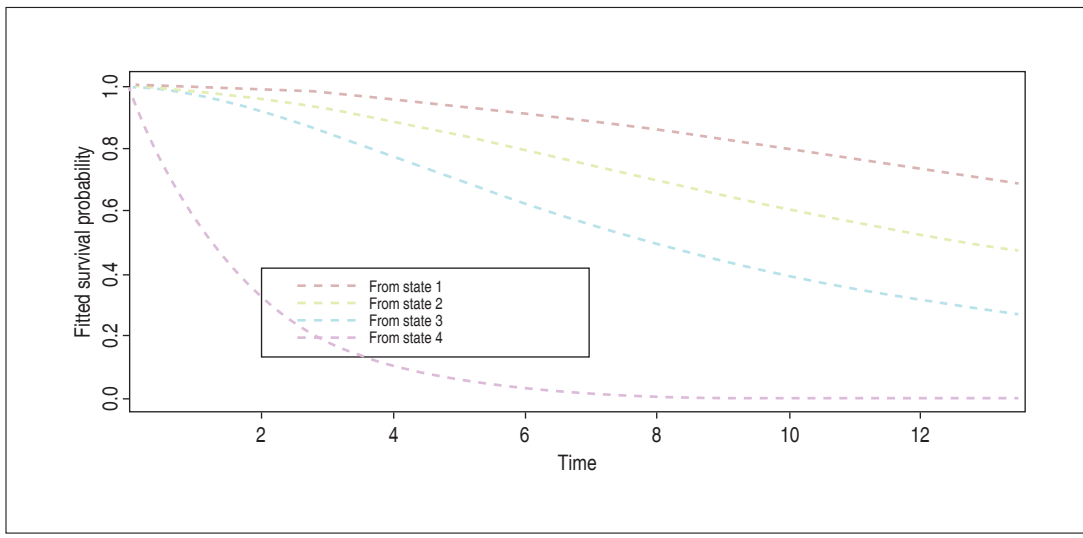

FIGURE 2: Plot of fitted survival probability.

The survival probabilities of stage 1, stage 2, stage 3 and stage 4 have been plotted against time (years) and have been shown in Figure 2. It is evident from the figure that the progression of CKD to end-stage (ESRD), i.e. stage 5 is very slow for stage 1 and stage 2 patients. There is a sharp decline in survival probability curve near the 1-year period for stage 4 patient. This indicates the fast deterioration of kidney function of stage 4 CKD patient.

The Table 6 shows the matrix of probabilities corresponding to each state k and l, $k=1,2,...,5$ and $l=1,2,...,5$. Each probability denotes the probability of next state l from state k and is equal to $-\frac{\lambda_{kl}}{\lambda_{kk}}$. These probabilities are also called p-next probabilities. They are different from the transition probabilities. The mean sojourn times along with p-next probabilities provide more intuitive parametric information of continuous time multi-state model based on Markov processes than crude transition intensities λ_{kl}.

TABLE 6: Probability of the next higher state from each state.

	State 1	State 2	State 3	State 4	State 5
State 1	0	0.6110	0.2716	0.1170	0.0000
State 2	0	0	0.6079	0.3919	0.0002
State 3	0	0	0	1.0000	0.0000
State 4	0	0	0	0	1.0000
State 5	0	0	0	0	0

TABLE 7: Values of likelihood ratio test statistics, p values and AIC values of the selected models involving covariates.

Covariates	-2*Log-Likelihood	LR test stats.	D.F.	p value
No covariates	572.08	-	-	-
Age	535.42	36.66	9	<0.0001
Sex (cat.)	554.57	17.51	9	0.0413
BMI	559.72	12.36	9	0.1938
HTN (cat.)	550.52	21.56	9	0.0104
Diabetes (cat.)	547.34	24.74	9	0.0033
Hb	544.31	27.77	9	0.0010
Urea	543.47	28.61	9	0.0008
Cr	544.68	27.4	9	0.0012
Alb	547.66	24.42	9	0.0037
HTN + Diabetes	520.17	51.91	18	<0.0001
HTN+Diabetes+Cr+Age+Hb+Urea	507.33	64.75	54	0.1501
All covariates	494.02	78.06	81	0.57192

BMI: Body mass index, Hb: Hemoglobin, Cr: Serum creatinine, Alb: Albumin, HTN: Hypertension, LR test stats.: Likelihood ratio test statistic.

Table 7 shows the values of -2 Log Likelihood values, likelihood ratio test statistics and p value of the model comprising of various covariates. On comparing the likelihood ratio test statistic of 36.66 corresponding to a covariate Age to a chi square distribution with 9 degrees of freedom, it is observed that model with the covariate Age fits significantly better than the base model having no covariates as p value is less than .0001. On the other hand, on comparing the likelihood ratio test statistic of 12.36 corresponding to a covariate BMI to a chi square distribution with 9 degrees of freedom, it is observed that it is not a significant factor as the as p value is more than 0.05 even though the model with the covariate BMI has lesser AIC value than the base model having no covariates. Similar interpretation can be given for the other models as well. Firstly we have fitted the model with the univariate covariate then higher number of covariates has been considered. The prognostic factors age, sex, hypertension, diabetes, hemoglobin, urea, serum creatinine, albumin are significant factors for the progression of CKD into different stages as the models based on these covariates have a p value less than 0.05. However, the covariate BMI is non-significant for our data. The p values of the model based on the significant covariates age, sex, hypertension, diabetes, hemoglobin, urea, serum creatinine and the model based on all the covariates considered are more than .05. The similar result holds for the other models based on the combination of a number of covariates as well. It suggests that the model with all the significant covariates is not significant. However, the model based on the covariates hypertension and diabetes is a significant one and is the best one. It further emphasizes the role of hypertension and diabetes in the progression of ckd. Hazard ratios and 95% confidence interval for the covariates hypertension (HTN) and diabetes associated with the best model have been shown in Table 8.

TABLE 8: Hazard ratio and 95% confidence interval.

Stage		HTN			Diabetes		
		HR	L.L.	U.L.	HR	L.L.	U.L.
Stage 1	Stage 2	1.3708	0.545	2.20	1.0141	0.171	1.86
Stage 1	Stage 3	1.0211	0.195	1.85	1.009	0.166	1.85
Stage 1	Stage 4	1.0041	0.178	1.83	1.0003	0.157	1.84
Stage 2	Stage 3	1.8904	01.06	2.72	2.7412	1.90	3.58
Stage 2	Stage 4	1.3715	0.545	2.20	2.0126	1.17	2.86
Stage 2	Stage 5	1.2103	0.384	2.04	1.0472	0.204	1.89
Stage 3	Stage 4	2.4561	1.63	3.28	3.9114	3.07	4.75
Stage 3	Stage 5	1.9807	1.15	2.81	2.8624	2.02	3.71
Stage 4	Stage 5	4.2816	3.46	5.11	3.9706	3.13	4.81

HTN: Hypertension.

DISCUSSION

Chronic diseases have become one of the major concerns for the national public health policy makers, especially in developing countries. Early detection of the disease helps in keeping the patients at primary stages and delaying its progression to more severe stages due to its amenability to treatment Jackson.[12] Expected burden of the disease for the future can be computed from the knowledge of the progression of the disease. It helps the health policymakers to evaluate the cost-effectiveness of alternative interventions and formulate the national health policy.

Prevalence of CKD is more in men than in women, Amin.[10] There are 56.4% males in the present dataset of CKD patients. CKD is asymptomatic in its earlier stages. There are only 17 patients of stage 1 whereas a number of patients in stage 4 is 43. Diabetes and hypertension are significant factors for the progression of CKD[5] and is evident from the Cox-Proportional model in the present study as well. The mean age of the data set is 48.8 years consistent with the fact that people between the age group 40 to 60 years are more likely to be affected with CKD.

Multi-state model based on Markov processes is often used to model and analyze the course of CKD.[21,22] The rates of transition between different stages can be estimated. Also, covariates can be fitted to the transition rates. The state transition table under result section reveals an important fact that people are more worried towards their health as the number of transition in the same state is very high. The transition to stage 5 from stage 4 of CKD patient indicates the ultimate failure of kidney function. The transition intensity and transition probability of stage 4 to stage 5 is very high. The mean sojourn time for stage 4 is least indicating the rapidness in deterioration of the disease. The Survival probability curve between expected survival probability and time also reveals the sharp decline in the survivability of stage 4 patient.

Our health policymakers should emphasize the need for controlling the factors diabetes and hypertension along with other prognostic factors like serum creatinine, urea, albumin, body mass index and hemoglobin. People should be made aware of the catastrophic effects of these on their lives. People should be encouraged to have a regular screening so that early detection of CKD can be made. The government should ensure sufficient treatment to confine the disease at primary stage and delay or slow down its progression to the more severe stage.

The present study has some limitations. The records of patients were not maintained properly. The records were incomplete with respect to follow up data on clinical parameters of the patients. Some patients were left out from the present study on account of this deficiency. Some people did not respond positively and were against sharing their data and hence could not be included in the study. The present data have been collected from the patients of Delhi and adjoining areas only. Hence, the utmost care should be taken before generalizing it. The result should be substantiated from the data collected from the other parts of the country. The assumption of the fixed value of covariate for a continuous prognostic factor may not be a valid assumption as time homogeneity.

CONCLUSION

The results of the likelihood ratio tests are consistent with the earlier findings.[3] The prognostic factors age, sex, hypertension, diabetes, hemoglobin, urea and serum creatinine are significant factors. The likelihood test statistics value suggests that the model with diabetes and hypertension as a covariate is the most appropriate model for the progression of CKD. The progression of chronic kidney disease is very slow at

the initial stage of the disease but is quite fast in severe stages. The probability of transition from stage 3 to stage 4, stage 3 to stage 5 and from stage 4 to stage 5 is very high. Our findings from the present study reiterate the fact that hypertension and diabetes are the most important factors for the onset and progression of CKD.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***Design:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***Control/Supervision:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***Data Collection and/or Processing:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***Analysis and/or Interpretation:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***Literature Review:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***Writing The Article:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***Critical Review:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***References and Fundings:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur;* ***Materials:*** *Gurprit Grover, Alka Sabharwal, Shrawan Kumar, Arpan Kumar Thakur*

REFERENCES

1. Atkins RC. The epidemiology of chronic kidney disease. Kidney Int Suppl. 2005;(94):S14-8. [Crossref] [PubMed]
2. Hailpern SM, Melamed ML, Cohen HW, Hostetter TH. Moderate chronic kidney disease and cognitive function in adults 20 to 59 years of age: Third National Health and Nutrition Examination Survey (NHANES III). J Am Soc Nephrol. 2007;18(7):2205-13. [Crossref] [PubMed]
3. Tsai WC, Wu HY, Peng YS, Ko MJ, Wu MS, Hung KY, et al. Risk factors for development and progression of chronic kidney disease: a systematic review and exploratory meta-analysis. Medicine (Baltimore). 2016;95(11):e3013. [Crossref] [PubMed] [PMC]
4. Varma PP. Prevalence of chronic kidney disease in India-where are we heading? Indian J Nephrol. 2015;25(3):133-5. [PubMed] [PMC]
5. Agarwal SK, Srivastava RK. Chronic kidney disease in India: challenges and solutions. Nephron Clin Pract. 2009;111(3):c197-203. [Crossref] [PubMed]
6. Varughese S, Abraham G. Chronic kidney disease in India: a clarion call for change. Clin J Am Soc Nephrol. 2018;13(5):802-4. [Crossref] [PubMed]
7. Anupama YJ, Uma G. Prevalence of chronic kidney disease among adults in a rural community in South India: results from the kidney disease screening (KIDS) project. Indian J Nephrol. 2014;24(4):214-21. [Crossref] [PubMed]
8. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, et al. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Ann Intern Med. 2003;139(2):137-47. [Crossref] [PubMed]
9. Kazancioğlu R. Risk factors for chronic kidney disease: an update. Kidney Int Suppl. 2013;3(4):368-71. [Crossref] [PubMed]
10. Amin N, Mahmood RT, Asad MJ, Zafar M, Raja AM. Evaluating urea and creatinine levels in chronic renal failure pre and post dialysis: a prospective study. J Cardiovas Dis. 2014;2(2):1-4.
11. Hsu CY, Ordoñez JD, Chertow GM, Fan D, McCulloch CE, Go AS. The risk of acute renal failure in patients with chronic kidney disease. Kidney Int. 2008;74(1):101-7. [Crossref] [PubMed]
12. Jackson CH. Multi-state models for panel data: the msm package for R. J Statist Softw. 2011;38(8):1-29. [Crossref]
13. Grover G, Gadpayle AK, Swain PK. A multistate Markov model based on CD4 cell count for HIV/AIDS patients on antiretroviral therapy (ART). Int J Stat Med Res. 2013;2(2):144.
14. Chiang CL. The life table and its construction. Introduction to Stochastic Processes in Biostatistics. 1st ed. Hoboken: Wiley; 1968. p.198-214.
15. Cassandras CG, Lafortune S. Introduction to Discrete Event Systems. 2nd ed. New York, N.Y: Springer Science & Business Media; 2009. p.772.
16. Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. J Am Stat Assoc. 1985;80(392):863-71. [Crossref]

17. Kay R. A Markov model for analysing cancer markers and disease states in survival studies. Biometrics. 1986;42(4):855-65. [Crossref] [PubMed]
18. Cox DR, Miller HD. The Theory of Stochastic Processes. 1st ed. London: Methue; 1965.
19. Moler C, Van Loan C. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Review. 2003;45(1):3-49. [Crossref]
20. Marshall G, Jones RH. Multi-state models and diabetic retinopathy. Stat Med. 1995;14(18):1975-83. [Crossref] [PubMed]
21. Begun A, Icks A, Waldeyer R, Landwehr S, Koch M, Giani G. Identification of a multistate continuous-time non-homogeneous Markov chain model for patients with decreased renal function. Med Decis Making. 2013;33(2):298-306. [Crossref] [PubMed]
22. Anwar N, Mahmoud MR. A stochastic model for the progression of chronic kidney disease. IJERA. 2014;4(11):8-19.