



Modeling the Risk of Renal Impairment using Current Status Chronic Kidney Disease Data: A Simulation-based Analysis

KEYWORDS

CKD, Current Status, NPMLE, Proportional Hazards

Gurprit Grover

Associate Professor, Department of Statistics, University of Delhi, Delhi-110007, India.

Barnali Deka

PhD Scholar, Department of Statistics, University of Delhi, Delhi-110007, India.

ABSTRACT

Chronic kidney disease (CKD) is an emerging public health concern of India. We investigate hazards of renal impairment along with various risk factors under survival modeling setup, using a simulated data prepared based on reported prevalences of CKD by a number of population-based cross-sectional surveys of CKD prevalence in India. Since cross-sectional surveys only record, instead of exact time, the current status of kidney function at the time of examination, we encounter current status survival data. The survival analysis of our simulated data has provided evidence of increased risk of developing CKD for female, diabetic and hypertensive patients.

1. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive illness which is silently treading its way into the trajectories of major public health concerns of India. CKD is associated with increasing risk of cardiovascular morbidity, all-cause mortality and other complications of reduced renal function[1,2]. It has a prolonged asymptomatic period, followed by the progression to end stage renal disease (ESRD) or kidney failure. Substantial loss of kidney function might occur before clinical symptoms become apparent, if not detected and managed timely. The 'Indian CKD Registry' reported that about 70% of the patients presented late for treatment[3]. Patients diagnosed with ESRD require dialysis or transplantation for survival, which eventually results in increased competing risk of cardiovascular morbidities and mortality; and decline in quality of life. However patients with CKD are more likely to die, mainly from cardiovascular complications, than to develop ESRD[1,4]. Progress to kidney failure or other adverse outcomes could be prevented or delayed through early detection and treatment[5,6]. Therefore the time elapsed before severe damage to kidney is a failed opportunity to prevent or delay the progression to ESRD.

Thus assessment of risk of developing chronic renal impairment would contribute in great measure towards patient care and management. Here, we have attempted a survival analysis to estimate hazards of developing renal impairment and identify its risk factors. Survival modeling usually requires the onset time of the event of interest, which may be gathered through a longitudinal study of patients with individual follow-up. However, our effort is constrained by the lack of maintained registry data with regular follow-ups in hospitals in India. So far in India, most of the studies have been conducted to conjure up prevalence and other epidemiological parameters of CKD in India. These studies are either population-based observational, cross-sectional studies [7-12] or hospital-based [13-16].

Therefore we attempt to estimate the risk of developing renal impairment using survival modeling within the constraint of the reported cross-sectional survey data in India. We have generated simulated CKD data using the prevalence estimates reported in these population-based studies. Survival analyses have been performed on this simulated data to report an estimate of risk of renal impairment and its risk factors. We begin with a description of the simulated data in Section-2. Section-3 illustrates the survival analyses techniques. Results of the survival analyses on simulated CKD data are presented in Section-4 followed by discussions in Section-5.

2. DATA

Stages of CKD are defined according to the kidney function marker- glomerular filtration rate (GFR). Estimate of GFR (eGFR) is obtained by various formulas that use serum creatinine values and other anthropometric parameters. The eGFR (in ml/min per 1.73m²) is then used to classify subjects into K/DOQI stages of CKD- 1. ≥ 90 (Normal); 2. 60-89 (Mild); 3. 30-59 (Moderate); 4. 15-29 (Severe) and 5. ≤ 15 (ESRD) [17].

Our event of interest is the *renal impairment*, which is defined as CKD stage-3 onwards or *low eGFR* ($eGFR \leq 60$ ml/min/1.73m²). We generated a hypothetical data set, comprising of 1500 cases, simulated on the basis of information reported in some of the studies. Reported prevalences (in %) of low GFR in India are- 0.86, 1.39, 0.79, 13.3, 3.02 and 17.4 [7-12]. We calculated the average of these estimates to derive an empirical estimate of prevalence of low GFR in adult population at 6%.

In our simulated population, every subject is examined once to determine the status of renal function at the time of survey. Since renal impairment is a chronic non-fatal condition, the onset time of damage occurrence cannot be observed exactly; only the current status of renal function is known at the time of examination. So age (in years) of a subject at examination was considered as the observation time (say, O). We generated the observation times from a truncated Normal (54, 12.73) within the interval (20, 85). Time to renal impairment (say T) was generated from an exponential distribution on the interval (20, 100) with an appropriate parameter so that rate of the event onset is 6%. If $T \leq O$, time to renal impairment is left censored; else the observation is right censored. Since information on Sex, Diabetes Mellitus (DM) and Hypertension (HTN) were only available; we generated these covariates only through Bernoulli distribution. We generated several data sets sequentially. The finally selected data set was the one where descriptive characteristics were closer to the reported studies.

An individual with renal damage might have developed the condition (realized the event) before the examination, and the individual with the initial kidney impairment might develop severe dysfunction afterwards. Hence the collected data reflected the current status of the renal impairment of a subject rather than the exact time. The data collected from this simulation study can be categorized primarily as Current Status or Case-1 Interval Censored data[18].

3. STATISTICAL ANALYSIS OF CURRENT STATUS SURVIVAL DATA

Suppose Y_i be the time to renal impairment, T_i the examination (censoring) time, independent of Y_i , and Z_i a $p \times 1$ vector of covariates for i^{th} subject. Denote $\delta_i = 1$ when i^{th} subject showed renal impairment during examination ($Y_i \leq T_i$) and $\delta_i = 0$ otherwise. The observed data are $\{(t_i, \delta_i, Z_i), i=1, 2, \dots, n\}$. Denote the survival function and cumulative hazards function (CHF) of i^{th} subject by $S(\cdot; Z_i)$ and $\Lambda(\cdot; Z_i)$ respectively.

Nonparametric Maximum Likelihood Estimates (NPMLE) of Surviving Renal Impairment

The probability of developing (failure distribution) or surviving (survival distribution) renal impairment was estimated using nonparametric maximum likelihood method for current status data. Overall survival curve was compared with Kaplan-Meier (KM) plot taking age at the time of examination as the exact failure time. Also survival curves were compared between different levels of various risk factors considered in our study. We used the Iterative Convex Minorant (ICM) algorithm for computing NPMLE of current status data because of its efficient convergence for NPMLEs[19-21]. Steps of this algorithm are

1. Order the examination times: $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}$ and relabel δ_i accordingly as $\delta_{(1)}, \delta_{(2)}, \dots, \delta_{(n)}$.
2. Plot $(t_i, \delta_i), i=1, 2, \dots, n$.
3. Form the greatest convex minorant G^* of the points in step (ii).
4. Then NPMLE of $F_n(t_i)$ is the left-derivative of G^* at t_i , which can further be expressed as .

Then the estimated survival probabilities were compared between levels of covariates using a log-rank type test[18]. Any p -value less than 0.05 were considered significant.

Semiparametric Cox Proportional Hazards (PH) Regression

Under PH model, the hazards of renal impairment by the examination time, t , given covariate Z is

$$\ddot{E}(t; z_i) = \ddot{E}_0(t) \exp(z_i \ddot{\alpha}) \quad (1)$$

Where $\Lambda_0(t)$ is the unknown base-line CHF. The likelihood function is

$$L(\beta, S_0) = \prod_{i=1}^n [S(t_i)]^{1-\delta_i} \{1 - [S(t_i)]\}^{\delta_i}, \quad \text{where } S(t_i) = \exp(-\Lambda(t_i)) \quad (2)$$

Now (2) can be expressed as

$$L(\beta, \Lambda_0) = \prod_{i=1}^n \exp[-(1-\delta_i)\Lambda_0(t_i)e^{z_i\beta}] [1 - \exp(-\Lambda_0(t_i)e^{z_i\beta})]^{\delta_i} \quad (3)$$

The log-likelihood can be written as

$$l(\beta, \Lambda_0) = \sum_{i=1}^n \{ \delta_i \log[1 - \exp(-\Lambda_0(t_i)e^{z_i\beta})] - (1-\delta_i)\Lambda_0(t_i)e^{z_i\beta} \} \quad (4)$$

The estimates of β and Λ_0 can be derived using maximum likelihood approach, described by Huang (1996) and known as profile likelihood method[22] –first, maximizing (4) over Λ_0 for all fixed values of β to obtain $\Lambda_0(\beta)$; and then maximizing the profile log-likelihood function $l(\beta, \hat{\Lambda}_0(\beta))$ over β and thus $\hat{\Lambda}_0 = \hat{\Lambda}_0(\hat{\beta})$. Let $\beta^{(k)}$ be an initial value and set $k=0$. Then steps to compute the estimates are – (i). Maximize $l(\beta^{(k)}, \Lambda_0)$ over Λ_0 to obtain $\Lambda_0^{(k)}$, using ICM algorithm. (ii). Maximize $l(\beta, \Lambda_0^{(k)})$ with respect to β using Newton-Raphson method. Set $k = k+1$, and let $\beta^{(k)}$ be the maximizer. Repeat the steps until convergence.

Cox PH regression model, as described above, was applied to assess the effect of various covariates on the hazard of developing renal impairment. Goodness-of-fit of our model was assessed by plotting Cox-Snell residuals against cumulative hazard functions[23]. An approximately straight line would indicate a good fit. The proportionality assumption in our PH model was assessed by plotting log estimated cumulative hazard for each group of a covariate together against time; and approximately parallel plots would indicate if our PH model is correct[18].

4. RESULTS

Average age of the study population was 45 years (male-47 years, female-40 years). The prevalence of DM and Hypertension were found to be 16%, and 20.4% respectively.

Plots of survival functions of renal impairment estimated by NPMLE methods for all subjects and for various covariate groups were shown in Figure 1. The log-rank type statistics were significant for all the three covariates. Fig-1a depicted a comparison between NPMLE and Kaplan-Meier (KM) estimates of cumulative survival function. It showed clearly that KM overestimated probability of surviving kidney dysfunction. Fig-1b indicated a fair drop in survival probabilities for females after 40 years of age compare with their male counterparts (p -value < 0.005). Individuals with DM (Fig-1d) and HTN (Fig-1c) showed significantly higher progression towards renal impairment (p -value < 0.005).

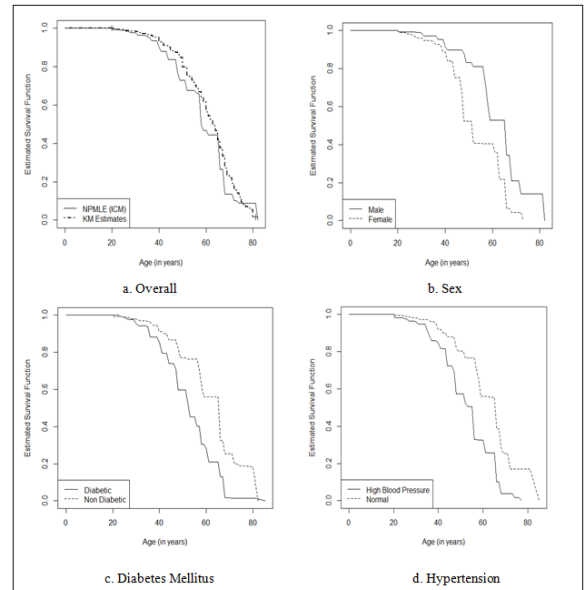


Figure 1: Plots of NPMLEs of Cumulative Survival Probabilities of Renal Impairment.

Table-1 reported the results of Cox PH regression model. Only the final model with significant covariates had been presented. Similar to NPMLEs, females are twice as likely to experience renal impairment as their male counterparts (HR=2.15; 95% CI: 1.83-2.52). A diabetic person was at significantly higher risk of developing kidney dysfunction (HR=1.8; 95% CI: 1.21-2.10) and observation conjoined in case of a hypertensive was akin to that of diabetics (HR=2.07; 95% CI: 1.51-2.21).

Table -1: Cox PH Regression model for Renal Impairment.

Risk Factors	β	Hazard Ratio (e^β)	95% Confidence Interval (CI)
Sex (Female)	0.765	2.15	1.832 – 2.523
DM (present)	0.73	2.075	1.507 – 2.209
HTN (present)	0.589	1.802	1.209 – 2.101

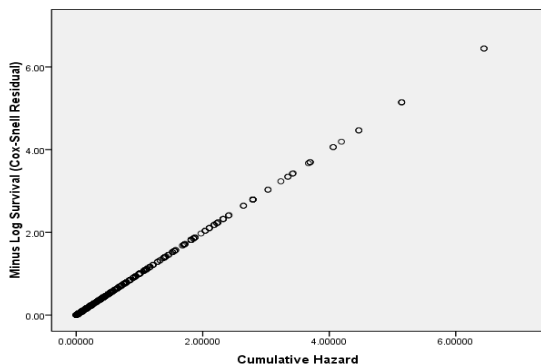


Figure-2 : Plot of Cox-Snell Residuals against Cumulative Hazards.

Our model was a good fit to the observed sample as evident from the straight line plot of Cox-Snell residuals against cumulative hazards (Fig-2). Also the key assumption of proportionality of hazards at different covariate levels had been validated in our model as apparent from the Fig-3.

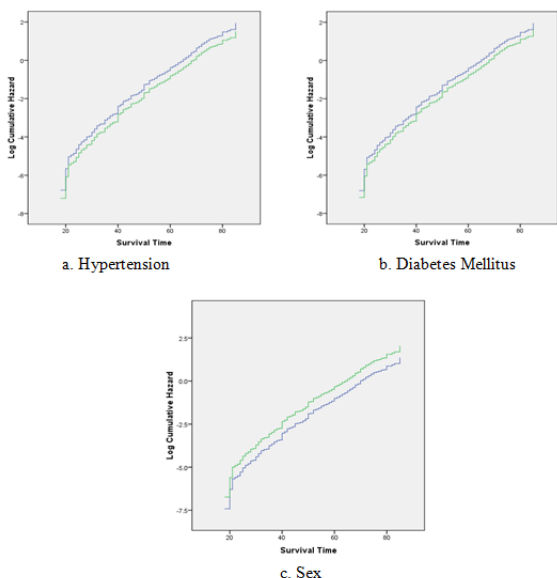


Figure 3: Plots of Log Cumulative Hazard of Renal Impairment against Survival Time. a.Hypertension (—High, —Normal); b. Diabetes Mellitus (—Yes, —No); c. Sex (—Male, —Female).

5. DISCUSSION

We have tried to provide a novel way of exploiting available databases and to initiate a platform for further in-depth

survival research in CKD. Our survival analysis of simulated data has provided evidence of increased risk of developing chronic renal impairment within various risk groups. Renal impairment in our study is defined as eGFR of a person less than 60 ml/min/1.73m³. Our study is first to study the risk of kidney impairment in India under a survival analysis set-up. Therefore direct comparisons of our findings with other studies, based on Indian subcontinent, might not be appropriate from methodological perspectives. However empirical comparisons can always lead to some important implications.

Our study revealed that females were twice as much likely to develop kidney impairment than male. Original study by Singh et al. reported a similar risk estimate in terms of Odds Ratio (OR). Two other studies on Asian population reported fairly comparable estimates for females[1,24]. A person with diabetes mellitus has a significantly higher risk of developing kidney dysfunction similar to other reports[9]. Diabetes mellitus is an independent risk factor for CKD. Epidemiological studies of Indian population revealed an increasing trend in its prevalence from 2.1% in 1979 to nearly 16% in 2009 [25]. Prevalence of diabetes and hypertension in our simulated population were 16% and 20% respectively; and also accounted for nearly 44% and 50% of cases with renal impairment which is fairly comparable with the findings of other studies from India. Our study population had a prevalence of hypertension at 20% and it accounted for nearly 50% cases of kidney impairment. These are fairly comparable with the findings of other studies from India[8,6,11].

Implications of our findings are stark realities that are lurking behind the changing landscape of India's socio-economic and behavioral paradigms. The magnitude of CKD will be of grave concerns in the years to come with the steadily increase in two of its risk populations. International Diabetes Federation (IDF) estimates that India's diabetic population will rise up to 70 million by 2025; and the prevalence of hypertension increased steeply from 6.2% in 1959 to 30.9% in 2005 [26]. However implication of our results should be carefully observed. Our study population has been generated through simulation using information reported in various population-based observational surveys. To conjure up a comprehensive picture of epidemiological progression of CKD in India, a follow-up study of CKD patients coming to nephrologists in hospital set-up would be more informative compared to our study, in terms of evaluating various clinical risk factors of CKD.

With a growing number of ESRD patients and in the absence of any health policy support from the government for treatment, CKD is gradually transforming to a major public health crisis in India which will eventually put a huge economic and disease burden on individual as well as on health planning. Interventions targeted at various risk groups would initiate early interventions to prevent progression to ESRD or eventually dialysis. Inclusion of CKD in the target groups of future government health programs would contribute greatly towards achieving this goal.

REFERENCE

1. Hosseinpah F, Kasraei F, Nassiri AA, Azizi F. High prevalence of chronic kidney disease in Iran: a large population-based study. *BMC Public Health* 2009; 9: 44 (doi: 10.1186/1471-2458-9-44). | 2. Snively CS, Gutierrez C: Chronic kidney disease: prevention and treatment of common complications. *Am Fam Physician* 2004, 70:1921-1928. | 3. Indian CKD Registry. First Annual Report (<http://www.ckdri.org/1stannualreportckd/ppf>). | 4. Astor BC, Schmid H, Miller ER 3rd, Yeung E and Coresh J: Glomerular filtration rate, albuminuria, and risk of cardiovascular and all-cause mortality in the US population. *Am J Epidemiol* 2008, 167(10):1226-1234. | 5. Kausz AT, Khan SS, Abichandani R, Kazmi WH, Obrador GT, Ruthazer R and Pereira BJ: Management of patients with chronic renal insufficiency in the Northeastern United States. *J Am Soc Nephrol* 2001, 12:1501-1507. | 6. Locatelli F, Vecchio LD, Pozzoni P: The importance of early detection of chronic kidney disease. *Nephrol Dial Transplant* 2002, 17(Suppl 11):2-7. | 7. Mani MK. Prevention of chronic renal failure at the community level. *Kidney Int* 2003; 63 (suppl 83): 86-89. | 8. Mani MK. Experience with a program for prevention of chronic renal failure in India. *Kidney Int* 2005; 67 (suppl 94): 75-78. | 9. Agarwal SK, Dash SC, Irshad M, Raju S, Singh R, Pandey RM. Prevalence of chronic renal failure in adults in Delhi, India. *Nephrol Dial Transplant* 2005; 20: 1638-1642. | 10. Mani MK. Nephrologist sans frontières: preventing chronic kidney disease on a shoestring. *Kidney Int* 2006; 70: 821-823. | 11. Modi GK, Jha V. The incidence of end-stage renal disease in India: A population-based study. *Kidney Int* 2006; 70(12): 2131-2133. | 12. Singh NP, Ingle GK, Saini VK, Jami A, Beniwal P, Lal M, Meena. Prevalence of low glomerular filtration rate, proteinuria and associated risk factors in North India using Cockcroft-Gault and Modification of Diet in Renal Disease equation: an observational, cross-sectional study. *BMC Nephrology* 2009; 10: 4 (doi:10.1186/1471-2369-10-4). | 13. Agarwal SK, Dash SC. Spectrum of renal diseases in India in adults. *J Assoc Physicians India* 2000;48: 594-600. | 14. Dash SC, Agarwal SK. Incidence of chronic kidney disease in India. *Nephrol Dial Transplant* 2005; 21: 232-233. | 15. Mittal S, Kher V, Gulati S, Agarwal LK, Arora P. Chronic renal failure in India. *Ren Fail* 1997; 19(6): 763-770. | 16. Sakhuja V, Jha V, Ghosh AK, Ahmed S, Saha TK. Chronic renal failure in India. *Nephrol Dial Transplant* 1994; 9: 871-872. | 17. Srinivas S, Annigeri RA, Mani MK, Rao BS, Kowdle PC, Seshadri R. Estimation of glomerular filtration rate in South Asian healthy adult kidney donors. *Nephrology (Carlton)* 2008; | 18. Sun J. *The Statistical Analysis of Interval-Censored Failure Time Data*. 1st Ed New York: Springer 2006. | 19. Groeneboom P, Wellner JA. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. New York: Birkhauser 1992. | 20. Jongbloed G. The iterative convex minorant algorithm for nonparametric estimation. *Journal Of Computational and Graphical Statistics* 1998; 7: 310-321. | 21. Huang J, Wellner JA. Asymptotic normality of the NPMLE of linear functional for interval censored data, case I. *Statistics Neerlandica* 1995; 49: 153-163. | 22. Huang J. Efficient estimation for the proportional hazards model with interval censoring. *Ann Statist* 1996; 24:540-568. | 23. Farrington CP. Residuals for proportional hazards model with interval-censored survival data. *Biometrics* 2000; 56: 473-482. | 24. Imai E, Horio M, Iseki K, Yamagata K, Watanabe T, Hara S, Ura N, Kiyohara Y and Hirakata H. Prevalence of chronic kidney disease (CKD) in the Japanese general population predicted by the MDRD equation modified by a Japanese coefficient. *Clin Exp Nephrol* 2007; 11: 156-163. | 25. Mohan V, Pradeepa R, Deepa M. Type 2 diabetes in India : An epidemiological overview. In : *Managing Diabetes: Challenges & Opportunities*. John BM (Ed), MediNEWS.Direct. Ilogy Healthcare Solutions, Bangalore 2010; Q3: pp 17- 21. | 26. Sicree R, Shaw J, Zimmet P: Diabetes and impaired glucose tolerance. In: *Diabetes Atlas*, International Diabetes Federation. 3rd Ed International Diabetes Federation, Belgium 2006; pp 15-103.