



Introduction to Competing Risk Model in the Epidemiological Research

Abbas Alipour¹, Abolghasem Shokri^{2*}, Fatemeh Yasari³, Soheila Khodakarim⁴

¹Assistant Professor of Epidemiology, Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

²MSc Student of Epidemiology, Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

³Assistant Professor of Nephrology, Department of Internal Medicine, Messiah Daneshvari Medical Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁴Assistant Professor of Biostatistics, Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Abstract

Background and aims: Chronic kidney disease (CKD) is a public health challenge worldwide, with adverse consequences of kidney failure, cardiovascular disease (CVD), and premature death. The CKD leads to the end-stage of renal disease (ESRD) if late/not diagnosed. Competing risk modeling is a major issue in epidemiology research. In epidemiological study, sometimes, inappropriate methods (i.e. Kaplan-Meier method) have been used to estimate probabilities for an event of interest in the presence of competing risks. In these situations, competing risk analysis is preferred to other models in survival analysis studies. The purpose of this study was to describe the bias resulting from the use of standard survival analysis to estimate the survival of a patient with ESRD and to provide alternate statistical methods considering the competing risk.

Methods: In this retrospective study, 359 patients referred to the hemodialysis department of Shahid Ayatollah Ashrafi Esfahani hospital in Tehran, and underwent continuous hemodialysis for at least three months. Data were collected through patient's medical history contained in the records (during 2011-2017). To evaluate the effects of research factors on the outcome, cause-specific hazard model and competing risk models were fitted. The data were analyzed using Stata (a general-purpose statistical software package) software, version 14 and SPSS software, version 21, through descriptive and analytical statistics.

Results: The median duration of follow-up was 3.12 years and mean age at ESRD diagnosis was 66.47 years old. Each year increase in age was associated with a 98% increase in hazard of death. In this study, statistical analysis based on the competing risk model showed that age, age of diagnosis, level of education (under diploma), and body mass index (BMI) were significantly associated with death (hazard ratio [HR]=0.98, $P<0.001$, HR=0.99, $P<0.001$, HR=2.66, $P=0.008$, and HR=0.98, $P<0.020$, respectively).

Conclusion: In analysis of competing risk data, it was found that providing both the results of the event of interest and those of competing risks were of importance. The Cox model, which ignored the competing risks, presented the different estimates and results as compared to the proportional sub-distribution hazards model. Thus, it was revealed that in the analysis of competing risks data, the sub-distribution proportion hazards model was more appropriate than the Cox model.

Keywords: Competing risk, Survival analysis, Epidemiological research

*Corresponding Author:

Abolghasem Shokri,
Tel: +98 912 81 61097,
Email:
epidemiology.shokri@sbmu.
ac.ir

Received: 21 January 2018
Accepted: 25 April 2018
ePublished: 15 September 2018



Introduction

The chronic kidney disease (CKD) is known as a global health problem. This disorder gradually leads to end stage of renal disease (ESRD). The ESRD is defined as reduction of the amount of glomerular filtration rate (GFR) to less than 15 ml/m²/per 1.73 m².¹ In the absence of national registries, no reliable data are available on the incidence and prevalence of ESRD in Iran. Due

to an increase in life expectancy in Iranian context, the contribution of ESRD related to the death factors is increasing. The prevalence and incidence of patients with ESRD based on the latest available statistics for the year 2006 were as follows: 434.83 and 63.8 in each million/year as well as the annual rate of the disease at the end of that year which was estimated between 4% and 5%.²

The survival of ESRD patients is lower than that

of the general population.³ The classic example is to consider the various causes of death. If a subject dies of one particular cause he/she is no longer at risk of death owing to any other cause. Two specific examples in the nephrology world that require competing risk analysis are comparisons of mortality risk in patients with CKD compared with the mortality risks of those patients who have developed ESRD, since the risk of mortality is different for patients who reach ESRD as compared to those who do not progress to it.⁴ In the traditional analysis of this data, most researchers were interested in knowing the distribution of observed survival times in a particular refractive index and all other factors as censored data.

In cohort studies and clinical trials, the incidence of an event is often investigated. Competing risk data occurs when the subjects under study have more than one single event, such as death, from various causes. The term competing risks also refer to data in which different events may not be mutually exclusive but interest in the first event occurs⁵ the most common of which are the competing risks of relapse and death in remission. These two risks are the primary reasons that patients fail treatment. In most epidemiological papers the effects of covariates on three outcomes (relapse, death in remission, and treatment failure) are modeled by distinct proportional hazards regression models. Standard survival analyses might overestimate the rate of the event, especially when the rate of the competing risk is high. The presence of competitive risks is time-to-time analysis and standard survival analysis is not always appropriate and thus should be interpreted with care.⁶ In addition, competing risks refer to situations in which different types of events might occur. For example, when studying death on dialysis, receiving a kidney transplant is an event that competes with the event of interest. Thus, a competing risk is an event that either prevents the event from being viewed or changes the chance of the event.⁷ The aim of this study was to demonstrate the practicality of competing risk models, to estimate the survival of ESRD patients using competing risk analysis, and also to compare its results with other commonly used approaches.

Methods

We performed a retrospective study among all ESRD patients older than 20 years who were registered (during 2011 to 2017) in the hemodialysis department of Shahid Ayatollah Ashrafi Esfahani hospital, Tehran, Iran. Patient were included in the study when renal replacement therapy such as hemodialysis was initiated. Those patients who died within three months following the onset of dialysis were excluded from the study. The collected data included

information and characteristics, clinical characteristics, transplantation history, hemodialysis, and also the cause of the end-stage renal disease. The event of interest were death and transplantation. Myocardial infarction, stroke, cancer, hepatitis C, accidents, sepsis, cardiac and respiratory arrest, seizures, diabetic foot, and embolism were considered to be as competing risks. Gender, age, body mass index (BMI), blood markers, blood pressure, diabetes, smoking, education level, marital status, and blood group were considered to be independent variables. Two different models were used to fit the data. In the first model, data were analyzed applying the competing risk approach as explained by Fine and Gray. They proposed the direct use of a regression model on a cumulative incidence function. The second model that was employed to fit the data was a specific hazard model. In the following sections, a brief introduction is provided regarding the concepts of these models.

Statistical Methods for the Analysis of Survival Data in the Presence of Competing Risks

Gray and Fine Models

The model proposed by Fine and Gray is based on the hazard of the sub-distribution and provides a simple relationship between covariates and cumulative incidence.⁸ In this model, the effect of independent variables, in contrast to the direct cause-specific model, is examined on the cumulative incidence function. Therefore, the effect of an independent variable in this model would be very different from its effect in cause-specific model.⁹ In other words, this model seeks to model the cumulative incidence function. Given the definition of the cumulative incidence function as the probability of the given incident occurrence before the specified time T and prior to the occurrence of the competing incidents, by defining the cumulative incidence function for each incident, for example, for the first incident, this model will be as follows:

$$F_1(t; Z) = \Pr(T \leq t, \varepsilon = 1 | Z)$$

where F_1 denotes the cumulative incidence function of the first occurrence, Z represents the independent variables vector, and $\varepsilon = 1$ indicates the first occurrence. In addition, the sub-hazard model of this occurrence is provided as:

$$\lambda_1(t; Z) = -d \log\{1 - F_1(t; Z)\} / dt$$

It will be considered as hazards in the following form:

$$\lambda_1(t; Z) = \lambda_{10}(t) \exp\{Z^T \beta\}$$

where β is the vector of regression coefficients and

$\lambda_{10}(t)$ displays a nonparametric estimation of the baseline hazard. Combining the two recent equations, the cumulative incidence function for the first incident will be as follows:

$$F_1(t; Z) = 1 - \exp \left[- \int_0^t \lambda_{10}(s) \exp\{Z^T \beta\} ds \right]$$

In this model, the nonparametric estimation of the cumulative incidence functions, of course, is calculated with some changes relative to the cause-specific model. The difference between this model and the cause-specific model is the number of individuals at risk for the given incident. In the cause-specific model, the number of individuals at risk is reduced each time a person experiences the rest causes. But in this model, in which the person is experiencing other causes, it is assumed at the risk remained that the time for these individuals has been censored which is longer than all the times of the given incident. This leads to the cumulative incidence function's association with only one's own hazard function for a specific cause and not the rest of the hazards.

Cumulative Incidence Function

A second statistical method in competing risks is the cumulative incidence function. This function is directly estimable from the data without a need for making any distributional assumptions. Standard survival analysis methods have been commonly used to analyze the competing risks data. However, sometimes, inappropriate methods such as the complement of Kaplan-Meier estimate (1-KM) has been applied to estimate the probabilities of the occurrence of an event of interest in a competing risks setting.^{10,11} The 1-KM cannot be interpreted as the actual probability of the occurrence of an event by time t .⁶ In this classic analysis, there is a favorite event and all other events are censored. The assumption of this method is that of non-informative censoring which is based on the idea that censored patients are more likely to experience the event as follow-up patients. However, this assumption has not been confirmed in the presence of numerous competing results.¹² These estimates have been interpreted as the probability of an ideal event in the ideal world in which there are no other types of events.^{6,8} Although, in the presence of competing risks, each event has a hazard. Therefore, the number of failures from the competing risks will reduce the actual number of failures from the event of interest and consequently, influence the estimate of the probability of failure from this event. In these situations, the cumulative incidence function is the appropriate tool for analyzing such data. Cumulative incidence function for a specific event, also known as the

sub-distribution function, is defined as the probability of failing from a given cause in the presence of competing events, given that a subject has survived or has already failed due to different causes.^{6,11} In the present study, the estimate of the cumulative incidence for a specific event was simultaneously calculated based on the estimate of the overall survival function when all types of events are considered and on the hazard estimate of the specific event. Besides, the cumulative incidence function for a specific event depends not only on the number of individuals who have experienced this type of event, but also on the number of those who have not experienced any other event.¹¹ This function is often of interest in epidemiological research and its graphical display over the time is intuitive and appealing.^{11,13,14} To analyze the differences in cumulative incidence between various patient groups, Gray's test was used. Comparing the cumulative incidence functions gives an idea of the probability of occurrence of the event of interest, and therefore can be translated into an actual number of patients with the event of interest.¹¹

Results

Characteristics of Patients

Of the 400 hemodialysis patient participants, 41 (10.25%) of them were excluded from the study because they lacked complete data on all the predictors. Thus, the samples available for the analysis consisted of 359 patients. The participants included in these analyses had a mean age of 58.93 years and 230 (64.07%) of them were males. The sex ratio was 1.78. The median duration of follow-up was 3.12 years and mean age at ESRD diagnosis was 66.47 years. The baseline characteristics of our study are shown in Table 1. A total of 123 (34.26%) patients died. The findings regarding the dead patients of the center showed that of 122 dead patients, 72 (31.30%) of them were males and 50 (38.769%) of them were females. The most common blood type was A. In this study, the majority (76.32%) of people ($n=274$) were married.

Comparison of Different Regression Models

The method of analysis resulted in markedly different estimates. In this study, statistical analysis, based on the competing risk model, showed that age, age of diagnosis, level of education (under diploma), and body mass index were significantly associated with death (hazard ratio (HR) = 0.98, $P < 0.001$, HR = 0.99, $P < 0.001$, HR = 2.66, $P = 0.008$, and HR = 0.98, $P < 0.020$, respectively). Each year increase in age was associated with a 98% increase in hazard of death (Table 2).

Discussion

Competing risks are prevalent most epidemiological

Table 1. Characteristics of the Patients

Variables	No. (%)
Gender	
Male	230 (64.07)
Female	129 (35.93)
Status	
Alive	236 (65.74)
Death	123 (34.26)
Job	
Unemployed	126 (35.10)
Housewife	71 (19.78)
Employed	115 (32.03)
Retirement	47 (13.09)
Smoking	
Yes	140 (39.00)
No	219 (61.00)
Blood group	
A	115 (32.03)
B	71 (19.78)
AB	57 (15.88)
O	116 (32.31)
Underlying cause	
DM	78 (21.73)
HTN	91 (25.35)
Others	190 (52.92)
Marital status	
Single	22 (6.13)
Married	274 (76.32)
Divorced	1 (.28)
Widow	62 (17.27)
Level of education	
Illiterate	193 (53.76)
Diploma	102 (28.41)
Under diploma	10 (2.79)
Upper diploma	54 (15.04)

Table 2. Comparison of Different Regression Models Based on Significant Variables

Variables	HR	CI 95% for HR	P Value
Competing Risk Model			
Age	0.98	(0.97-0.98)	<0.001
Age of diagnosis	0.99	(0.998-0.999)	<0.001
Level of education			
Illiterate (reference)	1	1	1
Under diploma	2.66	(1.23-3.93)	0.008
BMI	0.98	(0.96-0.99)	0.020
Gender female (reference)	1.30	(1.06-1.59)	0.011
Hb	1.05	(1.00-1.11)	0.030
Cox Regression Model			
Age	1.03	(1.02-1.04)	0.001
Age of diagnosis	1.00	(1.000-1.001)	0.005
Level of education			
Illiterate (reference)	1	1	1
Under diploma	9.6	(0.00-0.74)	0.009
BMI	1.04	(1.01-1.07)	0.002
Gender female (reference)	0.72	(0.50-1.03)	0.080
Hb	0.79	(0.72-0.87)	0.001

researches. Failure to correctly account for competing events can result in adverse consequences including probability overestimation of the occurrence of the event and magnitude mis-estimation of relative effects of covariates on the incidence of the outcome. When estimating crude incidence of the outcome of interest, it is inappropriate to use the complement of Kaplan-Meier survival function because this will lead to an overestimation of the incidence of the outcome of interest when competing risks are present. It is important to provide the results for all causes and also for both cause-specific as well as and sub-distribution hazard functions. Our findings showed that the women with ESRD had a better survival rate than men. However, the difference was not statistically significant. Similar to a study by Tabrizi et al.¹ in our study, the majority of patients were hypertension (nearly 26%). In contrast, cumulative incidence was grossly overestimated by standard survival analysis. The Cox proportional regression approach requires a proportionality assumption. Our study showed that the estimates of the covariate coefficients were different on the cause-specific and also sub-distribution hazard models. The cause-specific hazard can be modeled using the Cox model which is broadly used in epidemiological research. These risk factors are often missing from competing risk analyses.¹⁵

Limitations of the Study

This study had some limitations which need to be acknowledged. The most important limitation was the lack of information about incidence and prevalence of ESRD in Iran. As a result, it is suggested that further studies to be conducted, using more complete information, to estimate the incidence and prevalence in Iranian ESRD patients.

The researchers did not collect follow-up blood pressure, medications as well as proteinuria information and thus, they cannot comment on the time dependence of the outcomes on these risk factors. Therefore, similar studies focusing on the above-mentioned issues are subject to further investigation.

Conclusion

In the presence of competing risk outcomes, Kaplan-Meier estimates are biased as they overestimated the probability of the occurrence of an event of interest. In this paper, 2 common methods have been discussed for handling competing risks and their applications in regression settings.

Ethical Approval

This study was obtained from a master's thesis in epidemiology, which was conducted by a grant (No.

IR.SBMU.PHNS.REC.1396.79) given by Sahid Behesti University of Medical Sciences. The Ethics Committee of Sahid Behesti University of Medical Sciences approved the study.

Conflict of Interest Disclosures

The authors declare that they have no conflict of interests.

Acknowledgements

This article is a part of a Master's thesis in epidemiology. The authors thank all the participants as well as the reviewers and the editor for their helpful comments that substantially improved the manuscript.

References

1. Tabrizi R, Moosazadeh M, Sekhavati E, Jalali M, Afshari M, Akbari M, et al. Competing Risk Analyses of Patients with End-Stage Renal Disease. *Electron Physician*. 2015;7(7):1458-63. doi: 10.19082/1458.
2. Mousavi SS, Soleimani A, Mousavi MB. Epidemiology of end-stage renal disease in Iran: a review article. *Saudi J Kidney Dis Transpl*. 2014;25(3):697-702.
3. Roudbari M, Foruzandeh F, Roudbari S. Survival analysis of dialysis patients and its associated factors in Zahedan, Iran. *Saudi Med J*. 2010;31(1):91-3.
4. Noordzij M, Leffondre K, van Stralen KJ, Zoccali C, Dekker FW, Jager KJ. When do we need competing risks methods for survival analysis in nephrology? *Nephrol Dial Transplant*. 2013;28(11):2670-7. doi: 10.1093/ndt/gft355.
5. Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. A note on competing risks in survival data analysis. *Br J Cancer*. 2004;91(7):1229-35. doi: 10.1038/sj.bjc.6602102.
6. Kleinbaum DG. *Survival Analysis, a Self-Learning Text*. *Biom J*. 1998;40(1):107-8. doi: doi:10.1002/(SICI)1521-4036(199804)40:1<107::AID-BIMJ107>3.0.CO;2-9.
7. Pintilie M. *Competing Risks: A Practical Perspective*. John Wiley & Sons; 2006.
8. Lim HJ, Zhang X, Dyck R, Osgood N. Methods of competing risks analysis of end-stage renal disease and mortality among people with diabetes. *BMC Med Res Methodol*. 2010;10:97. doi: 10.1186/1471-2288-10-97.
9. Pintilie M. Analysing and interpreting competing risk data. *Stat Med*. 2007;26(6):1360-7. doi: 10.1002/sim.2655.
10. Lawless JF. *Statistical models and methods for lifetime data*. John Wiley & Sons; 2011.
11. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med*. 1999;18(6):695-706.
12. Teixeira L, Rodrigues A, Carvalho MJ, Cabrita A, Mendonca D. Modelling competing risks in nephrology research: an example in peritoneal dialysis. *BMC Nephrol*. 2013;14:110. doi: 10.1186/1471-2369-14-110.
13. Gichangi A, Vach W. The analysis of competing risks data: A guided tour. *Stat Med*. 2005;132(4):1-41.
14. Berry SD, Ngo L, Samelson EJ, Kiel DP. Competing risk of death: an important consideration in studies of older adults. *J Am Geriatr Soc*. 2010;58(4):783-7. doi: 10.1111/j.1532-5415.2010.02767.x.
15. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41(3):861-70. doi: 10.1093/ije/dyr213.
16. Eriksen BO, Ingebretsen OC. The progression of chronic kidney disease: a 10-year population-based study of the effects of gender and age. *Kidney Int*. 2006;69(2):375-82. doi: 10.1038/sj.ki.5000058.

How to cite the article: Alipour A, Shokri A, Yasari F, Khodakarim S. Introduction to competing risk model in the epidemiological research. *Int J Epidemiol Res*. 2018;5(3):98-102. doi: 10.15171/ijer.2018.21.