Georgia State University
ScholarWorks @ Georgia State University

Mathematics and Statistics Faculty Publications | Department of Mathematics and Statistics

2010

Methods of Competing Risks Analysis of End-stage Renal Disease and Mortality Among People with Diabetes

Hyun J. Lim
University of Saskatchewan College of Medicine, hyun.lim@usask.ca

Xu Zhang
Georgia State University, xzhang19@gsu.edu

Roland Dyck

Nathaniel Osgood

Follow this and additional works at: http://scholarworks.gsu.edu/math_facpub

Part of the Diseases Commons, Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation
Lim *et al.*: Methods of competing risks analysis of end-stage renal disease and mortality among people with diabetes. *BMC Medical Research Methodology* 2010 10:97. doi: 10.1186/1471-2288-10-97

This Article is brought to you for free and open access by the Department of Mathematics and Statistics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Mathematics and Statistics Faculty Publications by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

Lim et al. BMC Medical Research Methodology 2010, 10:97
http://www.biomedcentral.com/1471-2288/10/97

BMC
Medical Research Methodology

RESEARCH ARTICLE Open Access

Methods of competing risks analysis of end-stage renal disease and mortality among people with diabetes

Hyun J Lim[1*], Xu Zhang[2], Roland Dyck[3], Nathaniel Osgood[4]

Abstract

Background: When a patient experiences an event other than the one of interest in the study, usually the probability of experiencing the event of interest is altered. By contrast, disease-free survival time analysis by standard methods, such as the Kaplan-Meier method and the standard Cox model, does not distinguish different causes in the presence of competing risks. Alternative approaches use the cumulative incidence estimator by the Cox models on cause-specific and on subdistribution hazards models. We applied cause-specific and subdistribution hazards models to a diabetes dataset with two competing risks (end-stage renal disease (ESRD) or death without ESRD) to measure the relative effects of covariates and cumulative incidence functions.

Results: In this study, the cumulative incidence curve of the risk of ESRD by the cause-specific hazards model was revealed to be higher than the curves generated by the subdistribution hazards model. However, the cumulative incidence curves of risk of death without ESRD based on those three models were very similar.

Conclusions: In analysis of competing risk data, it is important to present both the results of the event of interest and the results of competing risks. We recommend using either the cause-specific hazards model or the subdistribution hazards model for a dominant risk. However, for a minor risk, we do not recommend the subdistribution hazards model and a cause-specific hazards model is more appropriate. Focusing the interpretation on one or a few causes and ignoring the other causes is always associated with a risk of overlooking important features which may influence our interpretation.

1. Background

In medical research, each person studied can experience one of several different types of events over the follow-up period and survival times are subject to competing risks if the occurrence of one event type prevents other event types from occurring. For example, in a study of bone marrow transplantation, leukemia relapse and death in remission are competing risks [1,2]. Leukemia relapse will not be observed once patients have died. Similarly, in a study of people with diabetes, end-stage renal disease (ESRD) and death compete for the life of the person, and each influence the risk of the other [3,4]. When a person experiences an event other than the one of interest in the study, the probability of experiencing the event of interest is frequently altered. Thus, caution is needed when we estimate survival probability of the event of interest in competing risks analysis [5]. Accordingly, if a person reaches the primary event of interest (e.g, ESRD), the other event (e.g, death without ESRD) is censored. The competing risk model can be described by specifying the cause-specific hazards as visualized as in Figure 1.

With competing risks data, the nonparametric Kaplan-Meier [6] cumulative hazard function, $\{1 - S_{KM}(t)\}$, has been used in some research. However, studies have demonstrated that $\{1 - S_{KM}(t)\}$ is inappropriate because it overestimates the probability of occurrence of the event of interest [7-12]. The bias is especially great when the hazard of the competing events is large [13]. An alternative method to the inappropriate cumulative hazard function is Cox cause-specific hazard [14] and

* Correspondence: hyun.lim@usask.ca
[1]Department of Community Health & Epidemiology College of Medicine, University of Saskatchewan 107 Wiggins Road Saskatoon, SK S7N 5E5, Canada
Full list of author information is available at the end of the article

BioMed Central
© 2010 Lim et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

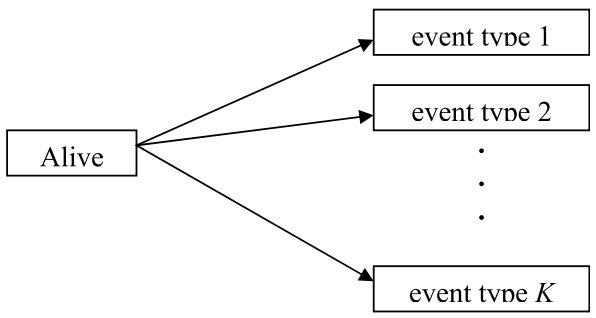

Figure 1 Competing risk models with *K* different event types.

cumulative incidence functions (CIF), which are the most important approaches to analyse competing risks data [12]. The cause-specific hazard measures the instantaneous failure rate due to one risk at a time. It is routinely estimated by constructing the Cox models on cause-specific hazards and treating time to event from the other competing risks as censored [11,12]. For each risk, the effects of prognostic factors are assessed as constant hazards ratios on the instantaneous failure rate of this risk. The CIF is an important quantity related to one risk in the context of competing risks. The CIF curve provides a better incidence curve associated with one risk than $\{1 - S_{KM}(t)\}$. It also provides a meaningful interpretation in terms of failure due to one risk regardless of whether competing risks are independent. Comparing the CIF curves is analogous to the log-rank test and is identical to the log-rank test in the absence of competing risks [15]. Gray considered a class of *K*-sample tests for the cumulative incidence based on weighted averages of subdistribution hazard functions [15]. Such tests do not require the independence assumption and does not adjust for other covariates.

In recent years, research methods centered on directly assessing covariate effects on a CIF have been developed [16,17]. One important work is the proportional subdistribution hazards model proposed by Fine and Gray [16]. This approach directly measures the covariate effects on the cumulative failure probability due to one risk, in the presence of other risks. As in any other regression analysis, modeling CIF for competing risks can be used to identify potential prognostic factors for a particular event in the presence of competing risks, or to assess a prognostic factor of interest after adjusting for other potential risk factors in the model.

The primary aim of this paper is to apply regression models on cause-specific hazards and subdistribution hazards to people with diabetes and to examine the estimates obtained by such models. We have identified the competing risk of "ESRD" versus "death without ESRD" in a diabetes population and evaluate the risk factors that are associated with these two outcomes. In the next section, we introduce a description of the diabetes study. In Section 3, models on cause-specific and subdistribution hazards for analysis are reviewed. In Section 4, the models are applied to the diabetes dataset to measure the hazard ratios of covariates and the cumulative incidence function. Section 5 contains a discussion about results and conclusions.

2. Study Description

2.1. Clinical Background

Diabetes is one of the most common chronic diseases occurring globally. In developed countries, population aging, inactivity, growing prevalence of obesity, and improved management of chronic complications have contributed to an epidemic of type 2 diabetes (T2DM) [4,18-23]. T2DM is particularly common among African Americans, Latinos, Native Americans, and Asian Americans/Pacific Islanders [24]. Increasing rates of T2DM among African-American and Canada's First Nations peoples reflect a parallel epidemic of overweight/obesity caused in large part by disruption of traditional cultures and lifestyles [25-27]. In addition to the above, risk factors for diabetes include obesity, physical inactivity, advanced aging, high blood pressure and/or high cholesterol, and family history of diabetes [24,26,28]. It has been reported that T2DM contributes to and is associated with increased mortality in end-stage renal disease (ESRD) populations [18,28,30]. Diabetic nephropathy affects about 10-20% of people with diabetes [31,32] and is a leading cause of ESRD [21-23,31]. Furthermore, studies have shown that individuals with both diabetes and ESRD have higher morbidity and mortality rates than individuals with only one of these conditions [4,33,34]. Finally, it has also been reported that women with ESRD have worse outcomes compared to men [4,28,29].

2.2 Study Population

We conducted a population study of diabetes, utilizing data drawn from the Saskatchewan Ministry of Health administrative databases. Descriptions of the overall study design and profile have been published elsewhere [27]. Briefly, Saskatchewan is a mid-western province of Canada with a population of approximately one million people through the years of study. Approximately 99% of the provincial population are beneficiaries of a universal health care system and recorded in the Ministry of Health's insurance registry. For this sub-study, 8274 First Nations people age 20 years or older with a diabetes incident year between 1980 and 2005 were identified as registered Indians in the databases.

In this study we excluded diabetes records related to a gestational record to ensure that gestational diabetes

cases were not counted as diabetes cases. The ESRD case definition was based on physician fee-for-service codes for chronic dialysis and renal transplantation. To qualify as a chronic dialysis patient in the study, a person was required to have received dialysis for at least 90 days, and to have received that treatment without any break of 21 days or greater. We excluded 20 patients who were classified as having reached ESRD prior to diabetes diagnosis. For all patients in the study, we obtained the following information: birth year, sex, and diabetes incident year. Where applicable, the ESRD incident year, year of death, and any period of health care coverage loss were also provided. A competing risk model was used to analyze the risk of two event types - ESRD or death without ESRD. Censoring time was set at December 31, 2005. In this study we explored and determined the effect of diabetes on ESRD and death when demographic characteristics were taken into account in the competing risks analysis.

3. Models

3.1. Standard single event time model

Let T be a random variable representing survival time that has a density function, $f(t)$, and the distribution function, $F(t)$. The survival function at time t, $S(t)$, is defined to be the probability that the survival time is greater than t, where $S(t) = P(T > t) = 1 - F(t)$. The survival function, therefore, represents the probability that an individual survives from the time origin (for example, time of the study enrollment or disease diagnosis) to sometime beyond t. The hazards function or hazard rate, $h(t)$, is the probability that an individual dies at time t, conditional on having survived to that time, which is defined as:

$$h(t) = \lim_{\Delta t \to 0} \left\{ \frac{P\left(t \le T < t + \Delta t \mid T \ge t \right)}{\Delta t} \right\}.$$

The hazard function, therefore, represents the instantaneous death rate for an individual surviving up to time t and provides a full characterization of the distribution of T. [35].

The main concern with this approach is how to study the impact of important covariates on the distribution of T. To do this, we assume the variation in the distribution of event and censoring times can be characterized by a vector of observed explanatory covariates, $\boldsymbol{z}$, which can be either time-invariant or time-dependent covariates. Under the Cox proportional hazards model, the hazard function for the event time T associated with the covariates $\boldsymbol{z}$ is defined as:

$$h(t) = h_0(t)\ e^{\beta' Z}.$$

Here, the function $h_0(t)$ is an unspecified baseline hazard function and gives the shape of the hazard function. If all explanatory covariates are zero, the hazard function will be the baseline hazard $h_0(t)$. If two individuals have identical values of the measured covariates, they will have identical hazard functions. The cumulative hazard function given $\boldsymbol{z}$ is defined by $\Lambda(t;\boldsymbol{z}) = \Lambda_0(t)\ e^{\beta' Z}$, where $\Lambda_0(t)$ is the cumulative baseline hazard and $\Lambda_0(t) = \int_0^t h_0(\mathrm{u})du$. The survival function is then obtained from the cumulative hazard function such that $S(t) = exp\{- \Lambda(t; \boldsymbol{z})\}$.

3.2. Models on cause-specific hazards

Competing risks in survival analysis refer to a situation where subjects under investigation are exposed to more than one possible type of events. Thus, each subject is associated with a pair (T, D) where T is the time-to-event (event time or failure time) and D is the type of the event for that subject. Here we assume that the possible causes are numbered from 1, ..., K. The cause-specific hazard function in the competing risks model is the hazard of failing from a given cause k in the presence of the competing events

$$h_k(t) = \lim_{\Delta t \to 0} \left\{ \frac{P\left(t \le T < t + \Delta t, D = k \mid T \ge t \right)}{\Delta t} \right\} \text{with } D = 1, \ldots, K.$$

With covariates, the regression model on cause-specific hazards is $h_k(t; \boldsymbol{z}) = h_{0k}(t)\ e^{\beta Z}$.

The total hazard, $h(t;\boldsymbol{z})$, equals the value of its corresponding hazards function summed up to time t. It is then

$$h(t;\boldsymbol{z}) = \sum_{k=1}^{K} h_k(t).$$

This equation means that the all-cause hazard rate is the sum of K hazards.

Define $\Lambda_k(t;\boldsymbol{z}) = \Lambda_{0k}(t)\ e^{\beta' Z}$, where $\Lambda_{0k}(t) = \int_0^t h_{0k}(\mathrm{u})du$ and $S_k(t; \boldsymbol{z}) = exp\{- \Lambda_k (t; \boldsymbol{z})\}$. Although we can estimate $S_k(t; \boldsymbol{z})$ from the cause-k specific cumulative hazard, $exp\{- \Lambda_k (t; \boldsymbol{z})\}$ is not interpretable as the marginal survival function for cause-k specific alone [12]. Instead $S_k(t; \boldsymbol{z})$ is the survival probability for the kth risk if all other risks were hypothetically removed.

With competing risks data, the cumulative incidence curve derived from cause-specific hazard functions provides important event information for a specific cause. The cause-specific cumulative incidence function (CIF)

of cause k at time t, $I_k(t)$, is defined by the probability of failing from cause k,

$$I_k(t) = P\left(T \le t, D = k\right) \quad k = 1, \ldots, K.$$

Given the covariate value $\boldsymbol{z}$, the CIF for cause k is also defined as

$$I_k(t;\boldsymbol{z}) = \int_0^t S(u;\mathrm{z})\, d\Lambda_k(u;\mathrm{z}) = \int_0^t S(u;\mathrm{z})\, h_k(\mathrm{u}) du$$

where $S(t;\boldsymbol{z})$ and $\Lambda_k(t;\boldsymbol{z})$ are the adjusted overall survival and cumulative hazard based on certain types of cause-specific hazard regression models [12]. This expression shows that the cumulative incidence of a specific cause k is a function of both the probability of not having the event prior to another event first ($S(u)$) up to time t and the cause-specific hazard ($h_k(u)$) for the event of interest at that time [7,8,12]. Estimation of the CIF can be obtained by using the cause-specific hazard.

Lunn-McNeil [36] extends to only one Cox model on cause-specific hazards rather than separate cause-specific models for each competing risk. Their method is an adaptation of Cox regression requiring event type indicator variables, which corresponds to different event types.The Lunn-McNeil approach stratified by event type gives identical results to those obtained from separate Cox models. The unstratified Lunn-McNeil model is an unstratified Cox proportional model, which can be used when constant hazard ratios between risk types is assumed. The unstratified Lunn-McNeil method assumes that different risk types have proportional baseline hazard functions. By contrast, the stratified method permits distinct baseline hazards for each event type [17]. If the proportionality assumption is not satisfied, then the stratified Lunn-McNeil model should be used.

3.3. Model on a subdistribution hazards

Fine and Gray [16] proposed a regression modeling applied directly on a cumulative incidence function for particular use in competing risks analysis. For any event type, this approach focuses on the hazard associated with the CIF, $I_k(t;\boldsymbol{z})$, which expresses the effect of covariates directly on the CIF. This is done via the subdistribution hazard function $h^*_k(t;\boldsymbol{z})$:

$$h^*_k(t;\boldsymbol{z}) = h^*_{0k}(t)\, e^{\beta Z}.$$

The CIF on the subdistribution is the function such that

$$I_k(t;\boldsymbol{z}) = 1 - \exp\{-\int_0^t h^*_k(\mathrm{u}, \mathrm{z})\, du\}.$$

Here $h^*_k(t;\boldsymbol{z})$ is not the cause-specific hazard. The CIF for cause k not only depends on the hazard of cause k, but also on the hazards of all other causes. For this approach, the subdistribution hazard is also defined as

$$h^*_k(t;\boldsymbol{z}) = \lim_{\Delta t \to 0}\left\{\frac{P\left[t \le T < t + \Delta t, D = k \mid T \ge t \cup (T < t \cap D \ne k)\right]}{\Delta t}\right\} = -\frac{d\log\{1 - I_k(t;\mathrm{z})\}}{dt}$$

so that the covariate effect directly relates to the cumulative incidence function [12]. Fine and Gray imposed a proportional hazards assumption on the subdistribution hazards and gave estimators and large sample properties [12]. This method takes into account other events and does not make any assumptions about their independence between the event time and censoring distribution, i.e., the censoring mechanism is independent of disease progression.

Estimation of the covariates coefficients for the models on cause-specific and subdistribution hazards follows the partial likelihood approach used in the standard Cox model. However, the difference between cause-specific and subdistribution hazards lies in the risk set. For the cause-specific hazard, $h_k(t;\boldsymbol{z})$, the risk set decreases at each time point at which there is an event of another cause. For the subdistribution hazard, $h^*_k(t;\boldsymbol{z})$, a person who has an event from another cause remains in the risk set [16]. In our study, we have applied the Cox models on the cumulative incidences of ESRD and death without ESRD, and have determined the subdistribution hazards ratios.

4. Results

Of the 8254 subjects in the study, 3718 (45%) were male and mean age at diabetes diagnosis was 47.2 (s.d = ± 14) years old. During the study period (median follow-up time = 8.2 years), 1482 (17.9%) subjects died without ESRD and 200 (2.4%) developed ESRD. Of 200 ESRD patients, mean age of ESRD was 56.5 (s.d.= ± 11.2) years old and 110 (55%) died. The events of interest were the times to ESRD, or death without ESRD. The study censoring time was set at December 31, 2005. In contrast to other contributions that focused on describing the study population [27], we used the data below to illustrate the application of the competing risk hazards models and provide comparisons between techniques.

In our study, we applied the Cox models on cause-specific and subdistribution hazards to obtain the cumulative incidences of ESRD and death without ESRD: the hazards ratios are given in Table 1. The results show that the effect sizes from the cause-specific and subdistribution hazards models are quite close for death events but are different for ESRD events. This indicates that

Table 1 Estimation of hazard ratio (H.R), 95% confidence interval (C.I), and p-value from the Cox cause-specific and subdistribution hazards models of time from the diabetic diagnosis to ESRD and to death without ESRD

Competing Risk	Model	Covariate	H.R	95% C.I	p-value
ESRD		Male	1.513	1.146 - 1.998	0.004
	Cox cause-specific hazards model	Age < 40	-	-	-
		40 < Age < 6	1.149	0.849 - 1.554	0.368
		60 < Age	1.405	0.891 - 2.215	0.144
	Cox subdistribution hazards model	Male	1.323	1.006 - 1.742	0.045
		Age < 40	-	-	-
		40 < Age < 60	0.923	0.685 - 1.243	0.6
		60 < Age	0.53	0.335 - 0.838	0.007
Death without ESRD					
	Cox cause-specific hazards model	Male	1.377	1.243 - 1.525	< 0.0001
		Age < 40	-	-	-
		40 < Age < 60	2.68	2.267 - 3.158	< 0.0001
		60 < Age	10.23	8.653 - 12.09	< 0.0001
	Cox subdistribution hazards model	Male	1.36	1.226 - 1.498	< 0.0001
		Age < 40	-	-	-
		40 < Age < 60	2.65	2.248 - 3.126	< 0.0001
		60 < Age	9.96	8.453 - 11.74	< 0.0001

the covariates interacted with the two event types. Males were at 51% and 32% higher hazard risks of ESRD compared to females in the cause-specific and subdistribution models, respectively (Table 1). Age at diabetes diagnosis had a significant effect on development of ESRD. Data also showed that age greater than 60 years old was not different compared to age less than 40 years old on the risk of ESRD in the cause-specific model (p-value = 0.144). However, it was significant in the subdistribution model (p-value = 0.0066).

Male sex and increasing age were significant predictors for death without ESRD (Table 1). Even when the competing risk of ESRD occurrence was taken into account, males and older age groups had a higher probability of death without ESRD than females and young age groups, respectively. The cause-specific and subdistribution hazards models showed that males faced a 1.37 times higher risk of death without ESRD than females. Risk of death without ESRD also increased with age. People aged 40 to 60 years had 2.65 times higher risk of death compared to those aged younger than 40 years in the subdistribution model (95% C.I: 2.267 - 3.158, p-value < 0.0001). After adjusting for sex, this is interpretable as the risk of death without ESRD for people aged 40 to 60 increasing by 165% compared to people younger than 40.

Table 2 shows the risk of ESRD and death without ESRD when the data were analyzed by the unstratified Lunn-McNeil model under the assumption that the baseline cause specific hazards are proportional. Here the hazard of ESRD and death without ESRD were both increased by 40% for males compared to females. The risk type hazard ratio of 2.44 in the Table 2 indicates that for females younger than 40 years, the hazard of death without ESRD is 2.44 times higher than that of ESRD (95% C.I: 1.788 - 3.328, p-value < 0.0001). As clinically expected, older patients had a higher risk of death, but age did not show an effect on ESRD risk. Note that because the stratified Lunn-McNeil model is identical to the Cox cause-specific model, it is not discussed further.

Figures 2a-c and Figures 3a-c show the estimates of the CIF curves of risk of ESRD and death without ESRD by sex for subjects younger than 40 based on the Cox cause-specific, subdistribution hazards models, and the unstratified Lunn-McNeil model. Estimates for the cause-specific hazards model provided a slightly higher CIF curve than for the subdistribution hazards model and the unstratified Lunn-McNeil model (Figures 2a-c). The cumulative incidence probability of ESRD approached 6.6% in females and 9.4% in males in the cause-specific model (Figure 2a), 6.8% in females and 8.9% in males in the subdistribution

Table 2 Estimation of hazard ratio (H.R), 95% confidence interval (C.I), and p-value from the Lunn-McNeil unstratified models assuming constant ratios between ESRD and death without ESRD

Competing Risk	Covariate	H.R	95% C.I	p-value
ESRD	$Male_{ESRD}$	1.395	1.057 - 1.84	0.0186
	$Age_{(ESRD)} < 40$	-	-	-
	$40 < Age_{(ESRD)} < 60$	1.078	0.797 - 1.457	0.627
	$60 < Age_{(ESRD)}$	1.004	0.641 - 1.574	0.985
Death without ESRD	Risk type *	2.44	1.788 - 3.328	< 0.0001
	$Male_{(death)}$	1.40	1.264 - 1.55	< 0.0001
	$Age_{(death)} < 40$	-	-	-
	$40 < Age_{(death)} < 60$	2.708	2.291 - 3.202	< 0.0001
	$60 < Age_{(death)}$	10.73	9.078 - 12.68	< 0.0001

* The reference competing risk type is ESRD.

model (Figure 2b), and 6.2% in females and 8.2% in males in the unstratified Lunn-McNeil model (Figure 2c) at 20 year after diabetes diagnosis.

Figures 3a-c show the estimates of the CIF curves of risk of death without ESRD by sex for subjects younger than 40 based on the cause-specific and subdistribution hazards models, and the unstratified Lunn-McNeil model. The estimated CIF curves based on the three models are almost identical for the first 22 years. The cumulative incidence probability of death approached 15% in females and 20% in males at 20 years after diabetes diagnosis.

5. Discussion

In this study we used diabetes data to demonstrate analyses for competing risks and showed the differences in estimates obtained by the cause-specific and subdistribution hazards models, and the Lunn-McNeil model. Our analyses showed that the three models yielded different results with regard to the effects of covariates. The CIF of the cause-specific hazards model revealed a higher CIF curve than the subdistribution hazards model: the unstratified Lunn-McNeil model was lower yet. However, the cumulative incidence curves of risk of death without ESRD on those three models were very similar. Our data showed such noticeable phenomenon consistently throughout the other covariate (age, sex) categories.

The Kaplan-Meier survival estimate is not applicable for competing risks analysis. The Cox proportional regression approach requires a proportionality assumption. If the number of competing events is large or some events are rare, the proportionality assumption is

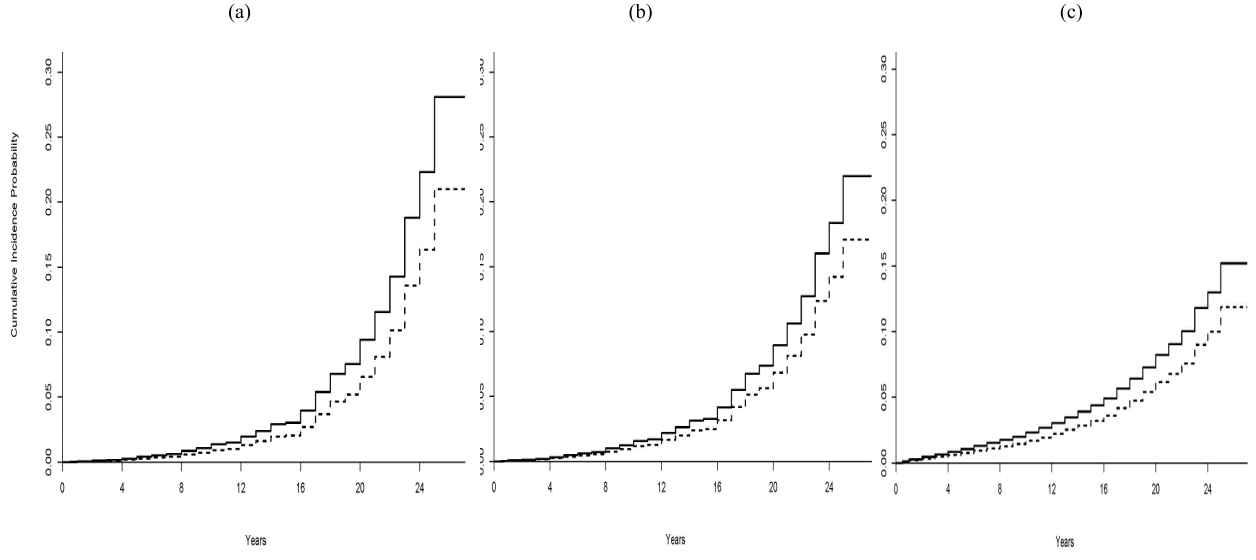

Figure 2 2a-2c: Estimates of the cumulative incidence curves of risk of ESRD. Estimates were by sex for subjects younger than 40 years old patient based on (a) the cause-specific hazards model; (b) the subdistribution hazards model; (c) the unstratified Lunn-McNeil model. Dashed line is for males and dotted line is for females.

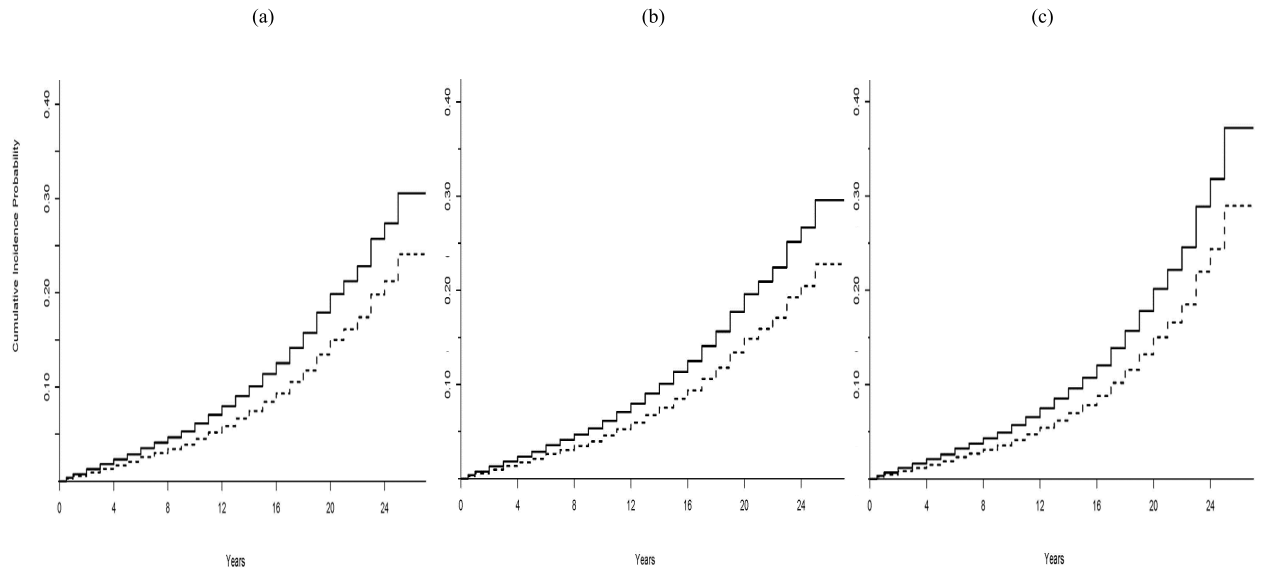

Figure 3 3a-3c: Estimates of the cumulative incidence curves of risk of death without ERSD. Estimates were by sex for subjects younger than 40 years old patient based on (a) the cause-specific hazards model; (b) the subdistribution hazards model; (c) the unstratified Lunn-McNeil model. Dashed line is for males and dotted line is for females.

often not satisfied. When the proportionality assumption is violated, the effects of covariate on the CIF curve can no longer be expressed by a simple number [13]. Gooley showed how the cumulative incidence curve can also be obtained [37]. The cumulative incidence of a specific event k is a function of both the probability of not having the event prior to another event ($S(u)$) up to time t and the cause-specific hazard ($h_k(u)$) for the event of interest at that time [7,8,12].

Our study showed that the estimates of the covariates coefficients on the cause-specific hazards and on the subdistribution hazards models were different. Latouche et al showed that the effects of covariate on the cause-specific hazard and on the subdistribution hazard were normally different [38]. Their paper addressed the relationship between the Cox cause-specific and subdistribution hazards models using a simulation study. The cumulative incidence of the minor (smaller number of events) risk may be greatly governed by the outcome of the dominant (larger number of events) risk. For example, age <40 that is protective against both risks may yield higher cumulative incidences regarding the minor risk than is observed in those age 40 - 60. The potential for misleading information makes it undesirable to interpret a covariate effect on a minor risk. In our study, there were 200 ESRD patients and 1482 deaths without ESRD, and the two risks are highly unbalanced. This is a good example of how use of the subdistribution hazards model on ESRD for a minor risk is not reliable, and why people should be cautious in interpreting such analysis. Thus, direct assessment of the covariate effect in the subdistribution model should not be conducted on the cumulative incidence of the minor risk. For death without ESRD as the dominant risk, one can use either a cause-specific hazard or subdistribution hazard model. However, for a minor risk, only the Cox cause-specific hazards model appears reliable.

The cause-specific hazard can be modeled using the Cox model, which is broadly used in medical research. The cause-specific hazard model may be more clinically understandable when assessing the prognostic effect of the covariates on a specific cause because we see that the covariate effect would be to reduce or increase the instantaneous probability of the event of interest irrespective of other covariate effect. However, when the study objective is to compare the probability of the event of interest, then the subdistribution hazards model will be appropriate. While the subdistribution hazards model might be limited to populations with similar characteristics and similar competing risk rate, the cause-specific hazard model is applicable for any population with similar characteristics regardless of the rates of competing risk events [39]. A cause-specific hazard can be expressed graphically, but they are not easy to interpret. Additional issues arise with interpretation of covariates on the hazard scale. The Cox subdistribution hazards model provides a methodology for modeling CIF with covariates using a proportional hazards assumption. The CIF are well suited to summarize competing risks data with a graphic display of the probabilities of event causes against time [8]. The CIF curve derived from a cause-specific hazard function provides

the probability of failure due to the event of interest in competing risks analysis [16,37-40].

An advantage of the Lunn and McNeil approach is that it facilitates direct comparisons between different event types. Depending on whether the assumption of proportional baseline cause-specific hazards holds, an unstratified or a stratified Cox regression could be applied. The unstratified method assumes that the baseline hazards for different risk types are proportional, while the stratified one allows for different hazards in each event type [41]. When the assumption of proportional baseline cause-specific hazards is satisfied, interpretation of the estimates from the unstratified Lunn and McNeil model is straightforward and allows assessment of the relative clinical importance of different event types. An advantage of the L-M model compared to the Cox cause-specific model is the flexibility to perform statistical inferences about various features of the competing risks using the information directly provided in the computer output. However, for the unstratified L-M model the constant assumption must be held within strata, otherwise the model is not valid. In most studies, different competing risk types will have substantially different underlying hazard functions and thus the applicability of the unstratified L-M model is restricted. Another limitation is that carrying out the L-M model requires additional data layout coding [42].

The cause-specific and subdistribution models share the same proportional hazards assumptions but normally the covariate effects on the cause-specific hazards and on the subdistribution hazards models are different [37]. This occurs because the effect of a covariate on the cumulative incidence of a particular cause is mediated via its direct effect on the cause specific hazard of that cause and via its indirect effect on the cause specific hazards for other cause [43]. Regarding the covariate effects, the results of the subdistribution hazards model have similar interpretations compared to the Cox model approach for competing risk data analysis: e^{β} represents the increase of the hazard of the subdistribution due to one unit increase of z. However, the cause-specific models

do not allow for a probability interpretation because the cumulative probability depends on other cause-specificevents. Thus, a summarizing the probability of the different effects of the cause-specific hazards is challenging [2,43,44]. Nonparametric inference for general summary measures for differences on the cumulative incidence functions was proposed [43]. It had also been shown that a proportional subdistribution hazards model provides an interpretable summary when the overall effects of covariate on the CIF are of interest [2]. Under the non-proportional subdistribution hazards, the estimated subdistribution hazards ratio for the CIFs is also interpretable as a time-averaged hazard ratio [44,45]. Further, even if the proportional subdistribution hazards model is misspecified, it provides an interpretable summary of separate cause-specific analyses [44].

The cause-specific and subdistribution models both require assumptions of proportional hazards. The proportionality assumption can be checked by evaluating the log{- log $S_{KM}(t)$} or by plotting residuals (Cox-Snell , Martingale, or Deviance residuals) or by adding time-dependent covariates in the model [12,35,46]. For the Cox cause-specific hazard model, the statistical software is available in many commercial statistical software packages and makes it easy to fit the models. However, for the sub distribution hazard models, currently standard procedure is not available in SAS, but SAS macros [47], STATA with ***compet.ado*** or **R**-package ***cmprsk*** are available.

In our study, we used data from administrative databases to estimate the competing risks of ESRD and mortality in First Nations people with diabetes. Since it was not a prospective study design and the subjects' clinical characteristics were not available, other important risk factors for ESRD and mortality could not be assessed. If the study had access to patients' demographics beyond age, gender and ethnicity and clinical information such information could be incorporated in the competing events analyses of ESRD and death. Interplay between the competing risks of ESRD and death might give the complete story about the effects of risk factors. If risk factors are different for two competing events then it is necessary to examine the decomposed outcomes on ESRD and mortality since pathways to the two events may be different. Establishing risk factors that cause progression to ESRD and distinguishing such risk factors from those that increase mortality can clearly predict two endpoints of ESRD and death, and can also be used as a decision making instrument.

6. Conclusion

In the analysis of competing risk data it is important to present both the results of the event of interest and the results of competing risks. One can use either the cause-specific hazards model or the subdistribution hazards model for a dominant risk. However, for a minor risk we do not recommend the subdistribution hazards model and a cause-specific hazards model is more appropriate in competing risk data analysis. In interpreting the results of a competing risks analysis, we should always take into account all causes. Focusing the interpretation on one or a few causes and ignoring the other causes is always associated with a risk of overlooking important features which may influence our interpretation. Investigators should take care in setting up

the right models to answer the questions of interest in their research. A graphic illustration of CIF curves will provide important additional insight, although the statistical tests like the log-rank test remains appropriate for testing group differences on each event type. Applying them together as complementary measures of risk clearly can expand a decision-making instrument for many competing risks studies.

Acknowledgements
This study is based in part on non-identifiable data provided by the Saskatchewan Ministry of Health. The interpretations and conclusions herein do not necessarily represent those of the Government of Saskatchewan or the Saskatchewan Ministry of Health. The authors would like to thank the reviewers for their constructive and valuable comments. We wish to thank Dr. Mary Rose Stang from the Saskatchewan Ministry of Health for her invaluable assistance in acquiring the data for this study.

Author details
[1]Department of Community Health & Epidemiology College of Medicine, University of Saskatchewan 107 Wiggins Road Saskatoon, SK S7N 5E5, Canada. [2]Department of Mathematics &Statistics Georgia State University 750 COE, 7th floor, 30 Pryor Street Atlanta, Georgia 30303, USA. [3]Department of Medicine, University of Saskatchewan 103 Hospital Drive Saskatoon, SK S7J 5B6, Canada. [4]Department of Computer Science University of Saskatchewan 110 Science Place Saskatoon, SK S7N 5C9, Canada.

Authors' contributions
HJL conducted the literature review, developed the mathematical framework, derived the results, and prepared the manuscript. XZ wrote computer programs and produced graphs.
RD defined the original study populations, acquired the data and provided advice on how to relate the topic to previous work in the field. NO reviewed the work and gave important editorial suggestions that greatly improved the manuscript. All authors have read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 25 February 2010 Accepted: 21 October 2010
Published: 21 October 2010

References

1. Klein JP: **Modelling competing risks in cancer studies.** *Statistics in Medicine* 2006, **25**:1015-1034.
2. Beyersmann J, Dettenkofer M, Bertz H, Schumacher M: **A competing risks analysis of bloodstream infection after stem-cell transplantation using subdistribution hazards and cause-specific hazards.** *Statistics in Medicine* 2007, **26**:5360-5369.
3. Agarwal R, Bunaye Z, Bekele D, Light R: **Competing risk factor analysis of end-stage renal disease and mortality in chronic kidney disease.** *Amer J of Nephrology* 2008, **28**:569-575.
4. Karame A, Labeeuw M, Trolliet P, *et al*: **The impact of type 2 diabetes on mortality in end stage renal disease patients differ between genders.** *Nephro Clin Pract* 2009, **112**:c268-c275.
5. Satagopan JM, Ben-Porat L, Berwick M, *et al*: **A note on competing risks in survival analysis.** *Br J Cancer* 2004, **91**:1229-1235.
6. Kaplan EL, Meier P: **Nonparametric estimation from incomplete observations.** *Journal of the American Statistical Association* 1958, **53**:457-481.
7. Gaynor JJ, Feuer EJ, Tan CC, Wu DH, Little CR, Straus DJ, Clarkson BD, Brennan MF: **On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data.** *J Am Stat Assoc* 2006, **88**:400-409.
8. Pepe MS, Mori M: **Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data.** *Stat in Medicine* 1993, **12**:737-751.
9. Lin DY: **Non-parametric inference for cumulative incidence functions in competing risks studies.** *Stat in Medicine* 1997, **16**:901-910.
10. Southern DA, Faris PD, Brant R, Galbraith PD, *et al*: **Kaplan-Meier yielded misleading results in competing risk scenarios.** *J of Clinical Epid* 2006, **59**:1110-1114.
11. Kalbfleisch JD, Prentice RL, Peterson AV Jr, *et al*: **Analysis of failure times in presence of competing risks.** *Biometrika* 1978, **34**:541-554.
12. Kalbfleisch JD, Prentice RL: *The statistical analysis of failure time data* New York, NY John Wiley & Sons Inc, 2 2002.
13. Putter H, Fiocco M, Geskus B: **Tutorial in Biostatistics: Competing risks and multi-state models.** *Statistics in Medicine* 2007, **26**:2389-2430.
14. Cox DR: **Regression models and life-tables (with discussion).** *Journal of the Royal Statistical Society B* 1972, **34**:187-220.
15. Gray RJ: **A class of K-sample tests for comparing the cumulative incidence of a competing risk.** *Ann Stat* 1988, **16**:1141-1154.
16. Fine JP, Gray RJ: **A proportional hazards model for the subdistribution of competing risks in survival analysis.** *J Am Stat Assoc* 1999, **94**:496-509.
17. Jung JH, Fine JP: **Parametric regression on cumulative incidence function.** *Biostatistics* 2007, **8**:184-196.
18. van Dijk PC, Jager KJ, Stengel B, Gronhagen-Riska C, Feest TG, Briggs JD: **Renal replacement therapy for diabetic end-stage renal disease: data from 10 registries in Europe (1991-2000).** *Kidney Int* 2005, **67**:1489-1499.
19. US Renal Data System: **USDRS 2009 Annual Data Report - Atlas of End-Stage Renal Disease in the United States.** Bethesda, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases 2009.
20. Sorensen VR, Mathiesen ER, Heaf J, Feldt-Rasmussen B: **Improved survival rate in patients with diabetes and end-stage renal disease in Denmark.** *Diabetologia* 2007, **50**:922-929.
21. Nelson RG, Knowler WC, Pettitt DJ, Bennett PH: In *Kidney disease in diabetes. Diabetes in America.* Edited by: Harris MI. Bethesda, MD: National Institutes of Health; , 2 1995:, Publication no 95-1468.
22. Groggel GC: **Diabetic nephropathy.** *Arch Fam Med* 1996, **5**:513-521.
23. Foote EF: **Prevention and treatment of diabetic nephropathy.** *Am J Health-System Pharm* 1995, **52**:1781-1792.
24. Ubink-Veltmaat LJ, Bilo HJG, Groenier KH, Houweling ST, Rischen RO, Meyboom-de Jong B: **Prevalence, incidence and mortality of type 2 diabetes mellitus revisited: a prospective population-based study in the Netherlands (ZODIAC-1).** *Eur J Epidemiol* 2003, **18**:793-800.
25. Tell GS, Hylander B, Craven TE, Burkart J: **Racial differences in the incidence of end-stage renal disease.** *Ethn Health* 1996, **1**:21-31.
26. Public Health Agency of Canada website:[http://www.phac-aspc.gc.ca/cd-mc/diabetes-diabete/index-eng.php], Accessed at Dec. 5, 2009.
27. Dyck R, Osgood N, Lin TH, *et al*: **Epidemiology of diabetes mellitus among first Nations and non-first Nations adults.** *CMAJ* 2010, [published online Jan 19, 2010].
28. Gu K, Cowie CC, Harris MI: **Mortality in adults with and without diabetes in a national cohort of the US population, 1971-1993.** *Diabetes Care* 1998, **21**:1138-1145.
29. Villar E, Chang SH, McDonald SP: **Incidences, treatments, outcomes, and gender effect on survival in end-stage renal disease patients by diabetic status in Australia and New Zealand (1991-2005).** *Diabetes Care* 2007, **30**:3070-3076.
30. Salles GF, Bloch KV, Cardoso CR: **Mortality and predictors of mortality in a cohort of Brazilian type 2 diabetic patients.** *Diabetes Care* 2004, **27**:1299-1305.
31. Levin ME, Pfeiffer MA: *The Uncomplicated Guide to Diabetes Complications. American Diabetes Association* Alexandria Verginia 1998.
32. Krolewski AS, Warram JH: **Natural history of diabetic nephropathy.** *Diabetes Rev* 1995, **3**:446-459.
33. Defronzo RA: **Diabetic nephropathy: etiologic and therapeutic considerations.** *Diabetes Rev* 1995, **3**:510-564.
34. Hellerstedt WL, Hohnson WJ, Ascher N, *et al*: **Survival rates of 2,728 patients with end stage renal disease.** *Mayo Clin Proc* 1984, **59**:776-783.
35. Collett D: *Modelling survival data in medical research* Capman & Hall/CRC. New York NY, 2 2003.
36. Lunn M, McNeil N: **Applying Cox regression to competing risks.** *Biometrics* 1995, **51**:524-532.

37. Gooley TA, Leisenring W, Storere BE: **Estimation of failure probabilities in the presence of competing risks: new representations of old estimators.** *Statistics in Medicine* 1999, **18**:695-706.
38. Latouche A, Boisson V, Chevret S, Porcher R: **Misspecified regression model for the subdistribution hazard of a competing risk.** *Statistics in Medicine* 2007, **26**:965-974.
39. Pintilie M: **Analysing and interpreting competing risk data.** *Statistics in Medicine* 2007, **26**:1360-1367.
40. Lau B, Cole SR, Gange SJ: **Competing risk regression models for epidemiologic data.** *Amer J of Epidemiology* 2009, **26**:3676-3679.
41. Tai BC, Machin D, White I, Gebski V: **Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches.** *Statistics in Medicine* 2001, **20**:661-684.
42. Kleinbaum DG, Klein M: *Survival Analysis: A self learning Text* Springer New York NY, 2 2005.
43. Zhang M, Fine J: **Summarizing differences in cumulative incidence functions.** *Statistics in Medicine* 2008, **27**:4939-4949.
44. Grambauer N, Schumacher M, Beyersmann J: **Proportional subdistribution hazards modeling offers a summary analysis, even if misspecified.** *Statistics in Medicine* 2010, **29**:875-884.
45. Beyersmann J, Schumacher M: **Letter to the editor on 'Misspecified regression model for the subdistribution hazard of a competing risk by Latouche A, Boisson V, Chevret S and Porcher R. *Statistics in Medicine* 2006; DOI: 0.2002/sim.2600'.** *Statistics in Medicine* 2007, **26**:1649-1651.
46. Mau J: **On a graphical method for the detection of time-dependent effects of covariate in survival data.** *Applied Statistics* 1986, **35**:24-255.
47. Zhang X, Zhang M: **SAS macros for estimation of direct adjusted cumulative incidence curves under proportional subdistribution hazards models.** *Computer Methods and Programs in Biomedicine* 2010.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2288/10/97/prepub

doi:10.1186/1471-2288-10-97
Cite this article as: Lim *et al.*: Methods of competing risks analysis of end-stage renal disease and mortality among people with diabetes. *BMC Medical Research Methodology* 2010 **10**:97.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central