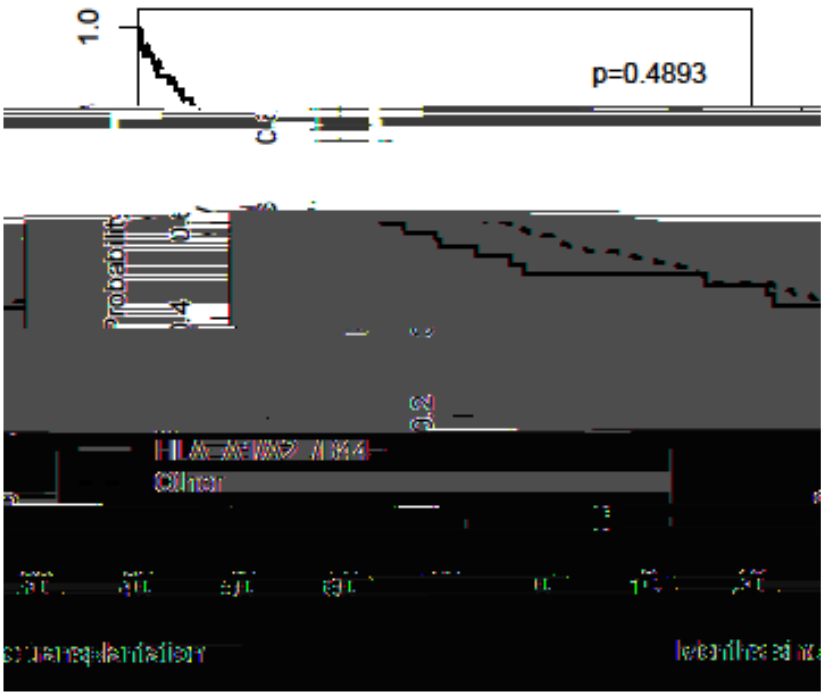
A Review of Competing Risks Data Analysis

Yizeng He[1], Kwang Woo Ahn[1], Ruta Brazauskas[1]

1.

Table 1 Patient characteristics by HLA combination

	HLA Combination	
Variable	HLA-A1, non-A2 and non-B44 (n=68)	Other combinations (n=355)



[Figure 1 Kaplan-Meier curves for overall survival in HCT cohort]

In many studies, a set of explanatory variables or covariates is available for every subject. The covariates may contain information about patients' age, gender, disease characteristics, and treatment. The goal is to identify covariates associated with higher risk of the events of interest. Cox proportional hazards model is the most commonly used regression model in survival analysis for assessing the relationship between the covariates and time to event of interest (4). The Cox model is concerned with the hazard rate which, at each time point, represents the instantaneous rate of failure among individuals who are still at risk at that time. For example, if the event is death, then the hazard rate for death at any particular time is the chance that a patient dies tomorrow given that he or she is alive today. A proportional hazards model assumes that the effect of a covariate is to multiply the baseline hazard by a function of the covariate. Traditionally, results are presented in terms of the hazard ratio or, equivalently, the relative risk quantifying the risk of experiencing the event if the individual was in one group relative to the risk of having the event among individuals from a different group. The theory for inference based on this model has been long established (5) and can be carried out by numerous software packages including SAS and R. Table 2 shows the analysis results of the Cox proportional hazards model. The analysis results can be interpreted via the hazards ratios. For example, the risk of death is 2 times higher among patients who have stable or progressive disease at the time of transplantation as compared to those who are in remission after adjusting for the other covariates.

Table 2 Multivariable analysis for HCT study.

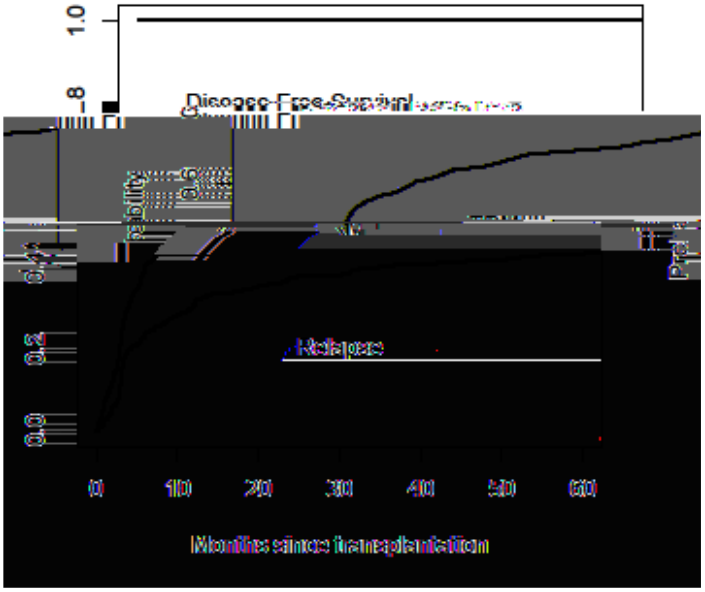
Variable	Hazard Ratio (95% CI)	P-value

HLA-A1, non-A2, non-B44	1	
Other HLA combinations	0.84 (0.60-1.19)	0.3232
Standard risk	1	
High risk	1.12 (0.84-1.49)	0.4513
Remission	1	
Stable/progressive	2.00 (1.55-2.60)	<.0001
		0.3128
Myeloablative	1	
RIC	0.83 (0.59-1.17)	0.2805
NMA	0.78 (0.56-1.08)	0.1313

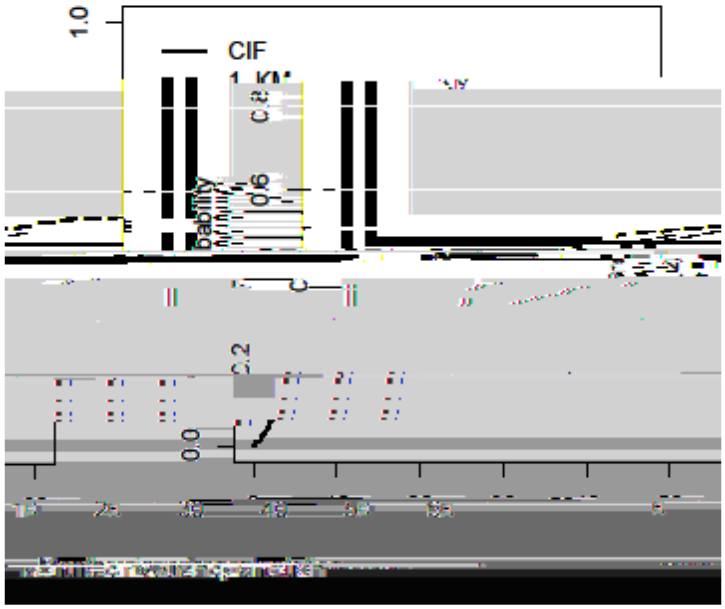
Competing risks data arises when subjects can potentially fail from multiple causes but experiencing failure from one cause precludes the subject from experiencing any other types of events. The most natural example is death from multiple causes such as cancer, cardiovascular disease or accidental death. Another simple example of such a scenario is relapse of leukemia. Relapse is not observed for those who died from treatment related complications before they could experience a relapse. In this case, death prior to relapse (or TRM), is the competing risk for relapse. When more than two competing risks are present in the study, all the failure types that are not of direct interest can be grouped together and considered a singular type. For this reason, we will consider the case where there are two competing risks: the failure type that is of interest and all the other competing failure types in the study. In this section, we will review methods used to summarize competing risks data as well as regression models used to establish the relationship between a set of risk factors and the occurrence of the event of interest.

The cumulative incidence function for competing risks data is a descriptive tool which represents the cumulative probability of the event of interest over time in the presence of other competing events. The calculations for estimating the cumulative incidence for a specific cause account for its dependence on the frequency and timing of other types of failures. Cumulative incidence function starts at 0 (the incidence of the event being evaluated is 0 at the start of the study) and is increasing in a stepwise fashion with a jump up at each time point when an event of interest occurs. Cumulative incidence probabilities should be estimated for all acting competing risks. At each time point, the sum of the cumulative incidence probabilities for all possible causes of failure will not exceed 1. In case there is only one type of failure, cumulative incidence function reduces to the complement of a KM estimate (1-KM). However, the presence of competing risks results in dependency between failure types and 1-KM is no longer correct estimate for the probability of experiencing any event of interest.

Now we will revisit the transplantation example. At each time point, a patient can be in one of the 3 states: dead from treatment related complications (TRM), relapsed, and alive and disease-free. Since these events are mutually exclusive, the probabilities of be


1.0
.8
Relapse
0
100
200
300
400
500
600
Months since transplantation

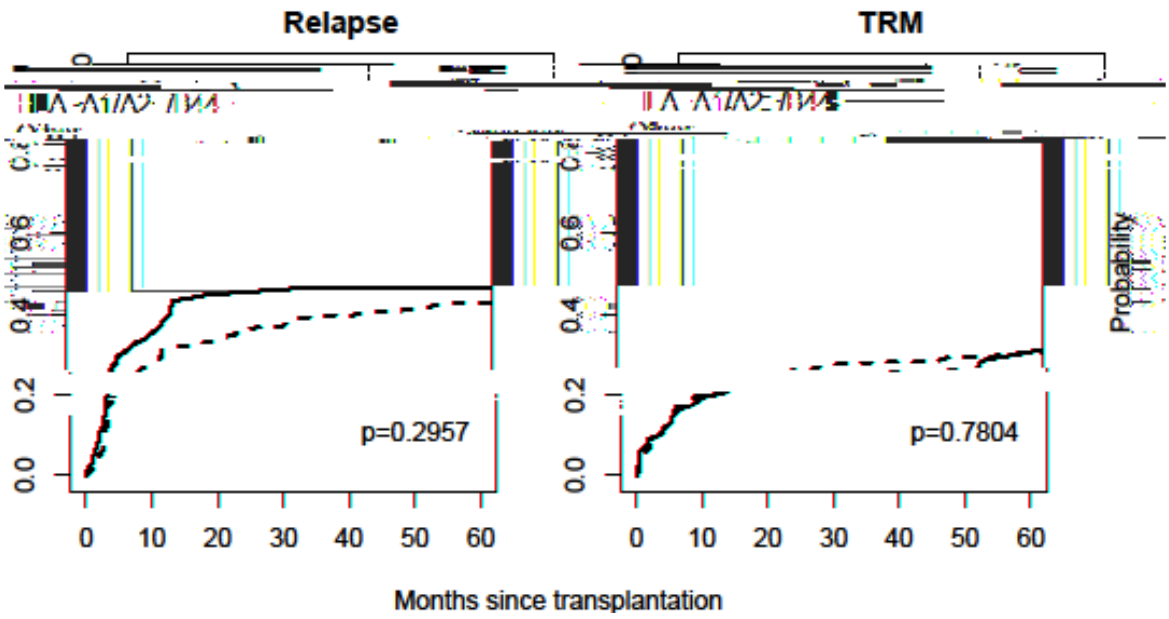
[Figure 2



[Figure 3 Relationship between cumulative incidence curve and 1-KM curve for relapse]

When cumulative incidence probabilities are being compared between two or more groups of patients, a graph depicting their experience consists of several curves representing cumulative incidence functions for the event of interest in each group. A formal comparison of the cumulative incidence LJN2ōI·ōλŧůλŜă 2ŦúǿƧ 2N Y2NŜ ƎN2dzLJă 2Ŧ LJI·ůλŜyűă λă Ř2yŜ dzăλyƎ DNI·è℺ă ŰŜăű ó6ú℗ DNI·è℺ă ŰŜăű λă I·y′I·ŘI·ptation of the log-rank test developed for competing risks data.

In HCT ŘI·ůI· ŜÈI·Y′LJfŜƗ DNI·è℺ă ůŜăű ŐI·y′ōŜ dzăŜŘ ůƧ ŜÈI·YʹλyŜ ǿKŜűKŜN ůKŜNŜ λă I· ăλƎyλŦλŐI·yű ŘλŦŦŜNŜyŐŜ λy′ cumulative incidence probabilities of each outcome between HLA-A1, non-A2 and non-B44 combinations and other HLA combinations. Figure 4 presents the cumulative incidence functions by HLA combinations for relapse and TRM respectively. There is no significant difference in cumulative incidence of relapse (p=0.2957) nor TRM (p=0.7804) between patients with HLA-A1, non-A2 and non-B44 combinations as compared to patients with other HLA combinations.



[Figure 4 Cumulative incidence functions by HLA combinations]

Regression models are employed to assess the effect of various risk factors on the occurrence of a certain type of event. In competing risks setting, this type of analysis is commonly carried out using one of two methods: Cox model or Fine-Gray model (4,7).

Cox model introduced in section 2 can be applied to analyze competing risks data. In the presence of multiple causes of failure, the rate of occurrence of each one of them is quantified by the cause-specific hazard. Cause-specific hazard at each time point for any failure type is defined as the instantaneous rate of occurrence of the event of interest at that time for the subjects who have not yet experienced any type of event (i.e. subjects who have not yet experienced the event of interest or the competing risks).

Since the probability of failure of a certain type depends on the rates of other competing events, there is no longer a direct relationship between cause-specific hazard rate and the probability of a particular type of event. In addition, covariates are not necessarily associated with the cumulative incidence function in the same way as they are associated with the cause-specific hazard. This difficulty motivated regression models which would directly link the covariates and the cumulative incidence function. Fine and Gray proposed a modification of the Cox model based on the transformation of the cumulative incidence function (7).

Fine-Gray regression model is based on an alternative failure rate summary measure, the subdistribution hazard function. The subdistribution hazard for a specific cause is the instantaneous rate of experiencing that particular cause given the individual have not yet experienced failure from that cause. For example, if the subdistribution hazard for relapse is of interest, patients who died before they experienced relapse are considered still at risk for relapse. Note the subtle difference between the cause-specific hazard and subdistribution hazard. For the cause-specific hazard, patients who die from other causes are no longer considered to be in the risk set, that is they are unable to experience the event of interest. With the subdistribution hazard, subjects who fail from another cause remain in the

graft-versus-host disease (aGVHD) has on mortality. These types of covariates are referred to as time-dependent covariates since all patients belong to the non-event group at the time of transplant and only change to the event group at the time of experiencing such an event. In contrast to time-dependent covariates, the variables we have considered so far in this paper, such as disease status or conditioning regimen intensity, are referred to as fixed covariates, meaning the groups are predetermined at the time of the transplantation, and will not change over time. One important feature of Cox cause-specific hazards model is that it allows the inclusion of time-dependent covariates. On the other hand, in most competing risks problems, time-dependent covariates cannot be incorporated into Fine-Gray model (8). Time-

References

1.