Notes for Biostatistics 140.641.01 'Survival Analysis'

First Term, 2022

Chapter 7. Competing risks models

"Competing Risks" refer to the situation where the population is at risk for more than one type of failures but its eventual failure is attributed to precisely one of the causes.

Example.

- Medicine: T is time to death; different causes of death

  Medicine: To study SARS-CoV-2 (Covid-19) infection for hospitalized patients, T is time from hospital admission to either death or discharge (i.e., being cured).
- Reliability: T is time to system breakdown; numerous reasons for breakdown
- Sociology: T is time until job exit, numerous reasons for job exit.

Standard setting in Survival Analysis

- time to failure, T
- only one type of failure
- potential censoring time, C
- observed failure time, $X = min(T, C)$
- censoring indicator, $\Delta = I(T < C)$
- covariates, Z

Basics of Competing Risks Models

- latent failure times $U_1, \ldots, U_m$
- time to failure, $T = min(U_1, , \ldots, U_m)$
- type of failure, $\pi = 1, ..., m$
- potential censoring time, C
- observed failure time, $X = min(T, C)$
- censoring indicator, $\Delta = I(T < C)$
- covariates, Z

Basic functions

For ease of discussion, we consider continuous failure times.

Cumulative distribution function: $F(t) = Prob(T \leq t)$

Survival function: $S(t) = Prob(T \geq t)$

Density function: $f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$

Hazard function: $\lambda(t) = \lim_{\Delta \to 0^+} \frac{\Pr(t \leq T < t+\Delta \mid T \geq t)}{\Delta}$

Cumulative hazard function: $\Lambda(t) = \int_0^t \lambda(u)du$

Observed Data

- Without censoring, the competing risks failure time data for a subject can be represented by

$$T_i = min(U_{i1}, , \ldots, U_{im}) \text{ and } \Pi_i$$

where T_i is the failure time and $\Pi_i = 1, \ldots, m$ is the cause of failure. The failure is due to the jth cause if $\Pi_i = j$, $j = 1, \ldots, m$.

- With censoring, the observed data can be represented by

$$X_i = min(T_i, C_i) \text{ and } \Pi_i \Delta_i$$

With additional covariate information, the observed data include

$$(x_i, \delta_i, \pi_i \delta_i, z_i), \ i = 1, \ldots, n$$

Remark: $\pi_i \delta_i$ equals 0 if failure event is censored, it equals π_i if failure event is uncensored.

The Cause-Specific Hazard Function

The cause specific hazard function due to cause j at time t, in the presence of other causes of failure, is defined as:

$$\lambda_j(t) = \lim_{dt \to 0^+} \frac{P[t \leq T < t + dt, \Pi = j | T \geq t]}{dt}$$

Relationship to the overall hazard function and survival function of T:

$$\lambda(t) = \sum_{j=1}^{m} \lambda_j(t)$$

$$S(t) = exp\left\{ - \int_0^t \lambda(u) du \right\} = exp\left\{ - \int_0^t \sum_{j=1}^{m} \lambda_j(u) du \right\}$$

Fundamental Relationship

Define the pseudo-survival function

$$S_j(t) = e^{-\int_0^t \lambda_j(u)du}, \quad j = 1, \ldots, m$$

Watch out! In general, $S_j(t)$ does not have a probability interpretation unless $U_1, ..., U_m$ are independent!

Then $S_j(t)$ is decreasing in t, $0 \leq S_j(t) \leq 1$, and $S(t) = \prod_{j=1}^{m} S_j(t)$. Let $x_{(1)} < x_{(2)} < ... < x_{(k)}$ be the ordered and uncensored failure times. Let $R(t) = \#\{j : x_j \geq t\}$ and $d_j(t) = \#\{k : x_k = t, \pi_k \delta_k = j\}, j = 1, ..., m$.
For simplicity of discussion, suppose there are no ties in uncensored data. Denote by n_i the number of subjects at risk at x_i, δ_{ji} the censoring indicator for type-j risk, and d_{ji} the number of failures due to cause j at an uncensored time x_i. Then, using d_{ji}/n_i to estimate $\lambda_j(x_i)$, the same arguments used to derive the usual K-M estimator lead to

$$\hat{S}_j(t) = \prod_{i:x_i<t} \left(1 - \frac{d_{ji}}{n_i}\right)^{\delta_{ji}}$$

and the following finite-sample equality holds: $\hat{S}_{KM}(t) = \prod_{j=1}^{m} \hat{S}_j(t)$.

Fundamental Relationship

The sub-pdf of T due to cause $j, j = 1, ..., m$, is defined as

$$f_j^{\text{sub}}(t) \quad = \quad \lim_{dt \to 0^+} \frac{P[t \leq T < t + dt, \Pi = j]}{dt}$$

Note that $f_j^{\text{sub}}(t) = \lambda_j(t)S(t)$, or equivalently, $\lambda_j(t) = \frac{f_j^{\text{sub}}(t)}{S(t)}$.

Thus, for the cumulative incidence function (CIF) $F_j^{\text{sub}}(t) = P(T \leq t, \Pi = j)$, then

$$f_j^{\text{sub}}(t) = \frac{dF_j^{\text{sub}}(t)}{dt}, \text{ and}$$

$$F_j^{\text{sub}}(t) = \int_0^t f_j^{\text{sub}}(u)du = \int_0^t \lambda_j(u)S(u)du$$

A property for latent variables $U_1, ..., U_m$

In general, $S_j(\cdot)$ is not the survival function of U_j. However, when $U_1, \ldots, U_m$ are independent, $\lambda_j(t)$ and $S_j(t)$ coincide with the hazard and survival function of U_j respectively. (why?)

If $U_1, \ldots, U_m$ are independent, then the cause specific hazard function becomes

$$\begin{aligned}
\lambda_j(t) &= \lim_{d \to 0^+} \frac{P[t \le T < t+d, \Pi = j | T \ge t]}{d} \\
&= \lim_{d \to 0^+} \frac{P[t \le U_j < t+d \,|U_1 \ge t, \ldots, U_m \ge t]}{d} \\
&= \lim_{d \to 0^+} \frac{P[t \le U_j < t+d \,|U_j \ge t]}{d} \\
&= h_j(t) \quad \text{(the hazard function of } U_j\text{)}
\end{aligned}$$

Thus, $S_j(t) = e^{-\int_0^t \lambda_j(u)du} = e^{-\int_0^t h_j(u)du}$, $j = 1, \ldots, m$, is the survival function of U_j.

Fundamental Relationship

Define the joint survival function of $(U_1, \ldots, U_m)$:
$S^*(t_1, \ldots, t_m) = Prob(U_1 \geq t_1, \ldots, U_m \geq t_m)$

Then,

1. $S(t) = S^*(t, t, \ldots, t)$
2. the sub-density $f_j^{\text{sub}}(t) = -\frac{\partial S^*(t_1,\ldots,t_m)}{\partial t_j} \mid_{t_1=\ldots=t_m=t}$
3. the cause-specific hazard is $\lambda_j(t) = -\frac{\partial \log S^*(t_1,\ldots,t_m)}{\partial t_j} \mid_{t_1=\ldots=t_m=t}$

Example: Bivariate exponential distribution (Gumbel, 1960)

Assume $S^*(t_1, t_2) = \exp(-\theta_1 t_1 - \theta_2 t_2 - \nu t_1 t_2)$,
where $\theta_1 > 0, \theta_2 > 0$ and $0 \leq \nu \leq \theta_1 \theta_2$.
The parameter ν specifies the dependence; $\nu = 0$ implies independence between U_1 and U_2. The cause-specific hazard is

$$\begin{aligned} \lambda_j(t) &= -\frac{\partial \log S^*(t_1, t_2)}{\partial t_j} \mid_{t_1=t_2=t} \\ &= (\theta_j + \nu t_{3-j}) \mid_{t_1=t_2=t} \\ &= \theta_j + \nu t \end{aligned}$$

The overall hazard for T is

$$\lambda(t) = \lambda_1(t) + \lambda_2(t) = \theta_1 + \theta_2 + 2\nu t$$

Exercise. Suppose (U_1, U_2) follows the Bivariate exponential distribution with the survival function $S^*(t_1, t_2) = \exp(-\theta_1 t_1 - \theta_2 t_2 - \nu t_1 t_2)$, where $\theta_1 > 0, \theta_2 > 0$ and $0 \leq \nu \leq \theta_1 \theta_2$.

(1) Show that $S^*(t_1, t_2)$ is a valid bivariate survival function. (Hint: prove the monotonicity and boundary properties)
(2) Find the pseudo-survival function $S_j(t)$ and the marginal survival function $S_j^*(t) = P(U_j \geq t)$, $j = 1, 2$.
(3) Show that $S_j(t) = S_j^*(t)$ if and only if $\nu = 0$.

Net and Crude Risks

Net hazard/survival/risk Assume everyone has potential to fail due to cause j. Net hazard/survival refers to risks related to the marginal distribution of latent failure time U_j, $j = 1, \ldots, m$. For example, the marginal survival function of U_j, $S_j^*(t_j) = S^*(0, ..0, t_j, 0..., 0)$.

Crude hazard/survival refers to the risk due to cause j in the presence of all the other risks: These are the cause-specific $\lambda_j(t)$ and the CIF $F_j^{\text{sub}}(t)$ defined earlier.

Note that

- Only crude/cause-specific hazard is observable if nothing is observed after the death/failure event.
- $U_1, \ldots, U_m$ are independent $\Longrightarrow$ $\lambda_j(t)$ and $S_j(t)$ coincide with the hazard and survival function of U_j (net = crude).

Question: Is crude hazard or net hazard of interest? Does it makes sense to consider net risks?

Remarks

1. When latent failure times $\{U_j\}_{j=1}^m$ are correlated, the naive Kaplan-Meier estimator for the survival function S_j^* with competing risk events treated as independent censoring is biased. (Warning: This KM estimator does not have a proper interpretation if independent censoring assumption does not hold.)

2. The independence of the latent failure times $\{U_j\}_{j=1}^m$ cannot be tested on the basis of the competing risks data, and must be assumed as a priori on the basis of the physical or biological process leading to the failure of the system.

3. In many practical situations it is **not** realistic to assume independence of latent failure times $\{U_j\}_{j=1}^m$ even if you believe latent failure time model. Check by prior knowledge or experience to see if independent censoring is violated (sicker patients tend to die earlier due to diseases other than the focused disease; so shorter failure time tends to imply shorter censoring time).
4. Many statisticians don't even think that the latent failure time model makes sense in reality!
5. Latent failure time model leads to marginal distributional results which might be welcome in causal inference, but many do not believe this model. Cause-specific risk models produce statistical results connected to real populations, but the models may not offer causal interpretations.

Reference: see independent censoring cannot be tested:
Tsiatis (1975, Proceedings of the National Academy of Science)

Likelihood function for cause-specific hazards model

Recall $F_j^{\text{sub}}(t) = P(T \le t, \Pi = j)$ and $f_j^{\text{sub}}(t) = dF_j^{\text{sub}}(t)/dt$. Let $\boldsymbol{z}_i$ denote the vector of covariates. Assume the conditional independent censoring: (T, Π) is independent of C conditioning on $\boldsymbol{Z} = \boldsymbol{z}$. Then, the likelihood based on the observed data $(x_i, \pi_i\delta_i, \boldsymbol{z}_i), i = 1, \ldots, n$, can be expressed entirely in terms of the cause specific hazard functions. With abused notation: let $f_{\pi_i}^{\text{sub}}(x_i; \boldsymbol{z}_i) = \prod_{j=1}^m f_j^{\text{sub}}(x_i; \boldsymbol{z}_i)^{I(\Pi_i = j)}$, the likelihood function is

$$
\begin{aligned}
L &= \prod_{i=1}^n f_{\pi_i}^{\text{sub}}(x_i; \boldsymbol{z}_i)^{\delta_i} S(x_i; \boldsymbol{z}_i)^{(1-\delta_i)} = \prod_{i=1}^n \lambda_{\pi_i}(x_i; \boldsymbol{z}_i)^{\delta_i} S(x_i; \boldsymbol{z}_i) \\
&= \prod_{i=1}^n \left[\lambda_{\pi_i}(x_i; \boldsymbol{z}_i)^{\delta_i} \prod_{j=1}^m exp\left\{ - \int_0^{x_i} \lambda_j(u; \boldsymbol{z}_i) du \right\} \right] \\
&= \prod_{i=1}^n \left[\prod_{j=1}^m \left\{ \lambda_j(x_i; \boldsymbol{z}_i)^{I(\pi_i = j)\delta_i} \right\} exp\left\{ - \int_0^{x_i} \lambda_j(u; \boldsymbol{z}_i) du \right\} \right] \\
&= \prod_{j=1}^m \left[\prod_{i=1}^n \lambda_j(x_i; \boldsymbol{z}_i)^{I(\pi_i = j)\delta_i} exp\left\{ - \int_0^{x_i} \lambda_j(u; \boldsymbol{z}_i) du \right\} \right]
\end{aligned}
$$

Cause-Specific Proportional Hazards Model

Cause-Specific Proportional Hazards Model models the effects of covariates on the cause specific hazard function:

$$\lambda_j(t;\boldsymbol{z}) = \lambda_{j0}(t) exp\left\{\boldsymbol{z}\beta_j\right\},\ j = 1,\ldots,m$$

Let $\gamma_{i,j} = I(\pi_i = j)\delta_i$. Estimation of β_j can be done by maximization of the partial likelihood (Holt 1978, Biometrika):

$$\prod_{j=1}^{m}\prod_{i=1}^{n}\left\{\frac{\lambda_{j0}(t)exp[\boldsymbol{z}_i\beta_j]}{\lambda_{j0}(t)\sum_{l\in R(x_i)} exp(\boldsymbol{z}_l\beta_j)}\right\}^{\gamma_{i,j}} = \prod_{j=1}^{m}\prod_{i=1}^{n}\left\{\frac{exp[\boldsymbol{z}_i\beta_j]}{\sum_{l\in R(x_i)} exp(\boldsymbol{z}_l\beta_j)}\right\}^{\gamma_{i,j}}$$

Remark: Standard partial likelihood and log-rank test (i.e., as the partial score test) can be extended to the cause-specific (j-specific) proportional hazard model based on $\lambda_j(t;\boldsymbol{z})$, by treating other risk types (those $k \neq j$) as censored data.

Cause-Specific PHM: time-dependent covariates

Cause-Specific PHM with time-dependent covariates: The Cause-Specific PHM can be extended to handle time-dependent covariates.

$$\lambda_j(t; \boldsymbol{z}(t)) = \lambda_{j0}(t) exp\left\{\boldsymbol{z}(t)\beta_j\right\}, \; j = 1, \ldots, m$$

Let $\gamma_{i,j} = I(\pi_i = j)\delta_i$. Estimation of β_j can be done by maximization of the partial likelihood (Holt 1978, Biometrika):

$$\prod_{j=1}^{m}\prod_{i=1}^{n}\left\{\frac{\lambda_{j0}(t)exp[\boldsymbol{z}_i(x_i)\beta_j]}{\lambda_{j0}(t)\sum_{l\in R(x_i)} exp(\boldsymbol{z}_l(x_i)\beta_j)}\right\}^{\gamma_{i,j}} = \prod_{j=1}^{m}\prod_{i=1}^{n}\left\{\frac{exp[\boldsymbol{z}_i(x_i)\beta_j]}{\sum_{l\in R(x_i)} exp(\boldsymbol{z}_l(x_i)\beta_j)}\right\}^{\gamma_{i,j}}$$

PHM: Pros and Cons

- only need to assume proportionality for cause type-j to get β_j.
- correlation between different failure types is allowed.
- interpretation subject to 'failures happening in real life' - regression effect is not causal if correlation between different failure types exists.
- if one wish to use the PHM for the latent variable U_j (i.e., model the hazard of U_j, not the cause specific hazard of U_j) and treat other failure types as censoring, then one might encounter informative censoring problem due to correlation between different failure types. It is a generally challenging topic to develop statistical inference/methods for U_j.
- stronger interpretation if failure types are independent (i.e., model the hazard of U_j) - in this case, with latent events setting, other failure types would form independent censoring for observing type-j failures.

The Cumulative Incidence Function (CIF)

In the presence of competing risks, the probability of experiencing type-j failure is $F_j^{\text{sub}}(t) = P(T \leq t, \Pi = j)$, which is termed as **the cumulative incidence function (CIF)**.

This is a useful quantity from the patient perspective. In studying the effect of treatment on failure, the benefit derived over a person's lifetime (or any other significant time frame) may be of more interest and more focused than cause-specific hazard.

The cumulative incidence function can be thought of as:

$$F_j^{\text{sub}}(t) = \int_0^t \lambda_j(u) S(u) du$$

A nonparametric estimator can be obtained in a two step process:

- Calculate the KM estimate, $\hat{S}(t)$, of the overall survival function based on $\{(x_1, \delta_1), ..., (x_n, \delta_n)\}$.
- At the observed time x_i, where $\pi_i \delta_i = j > 0$ (uncensored case), estimate the cause-specific hazard probability $d\Lambda_j(x_i) (= \lambda_j(x_i) dx_i)$ by

$$\frac{\text{no. of type-j failures at } x_i}{Y(x_i)}$$

where $Y(x_i) =$ no. of subjects in the risk set defined at x_i.

Nonparametric Estimation of CIF

Then, for $j = 1, \ldots, m$, the cumulative incidence function can be estimated by

$$\hat{F}_j^{\text{sub}}(t) = \int_0^t \hat{S}(u) d\hat{\Lambda}_j(u) = \int_0^t \frac{\hat{S}(u) dN_j(u)}{Y(u)}$$

where

$$\begin{aligned} N_j(t) &= \sum_{i=1}^n I(X_i \leq t,\ \delta_i \Pi_i = j) \\ Y(t) &= \sum_{i=1}^n I(X_i \geq t) = \text{no. of subjects in risk set } R(t) \end{aligned}$$

Asymptotic normality property can be developed using Martingale theory.

CIF estimates

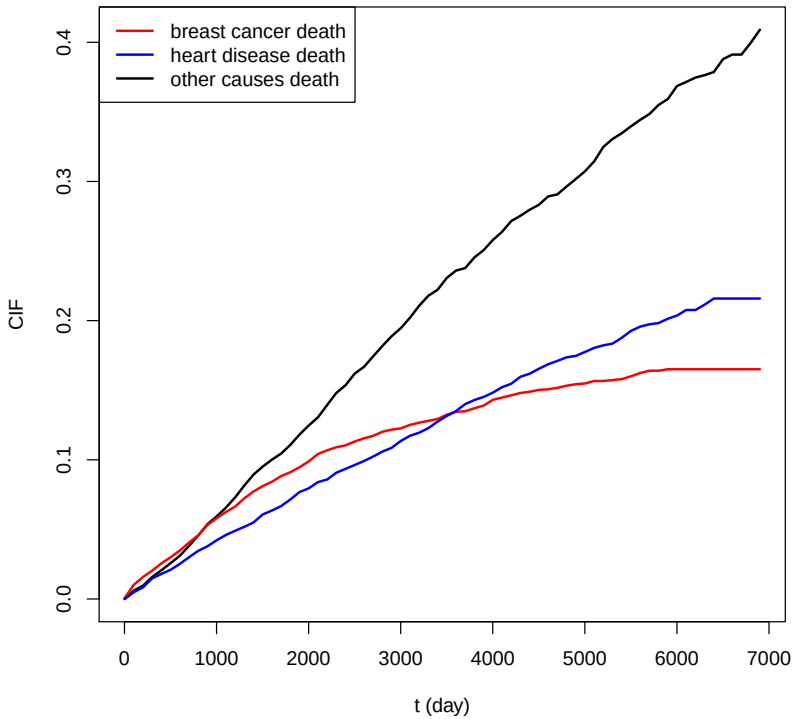


Figure: CIFs for women diagnosed with Breast Cancer with age 65+ (SEER database)

CIF regression model

With covariates $\boldsymbol{Z}$, Fine & Gray (JASA, 1999) proposed regression models based on CIF $F_j^{\text{sub}}(t;\boldsymbol{Z}) = P[T \le t, \Pi = j \mid \boldsymbol{Z}\ ]$.

- Define the sub-distribution hazard function $\lambda_j^{\text{sub}}(t;\boldsymbol{Z}) = \frac{f_j^{\text{sub}}(t;\boldsymbol{Z})}{1-F_j^{\text{sub}}(t;\boldsymbol{Z})}$.

  Note that the interpretation of $\lambda_j^{\text{sub}}(t)$ is not as 'clean' as the hazard function under non-competing risks setting.

- Define $\Lambda_j^{\text{sub}}(t;\boldsymbol{Z}) = \int_0^t \lambda_j^{\text{sub}}(u;\boldsymbol{Z})du$. Then we can show $1 - F_j^{\text{sub}}(t;\boldsymbol{Z}) = e^{-\Lambda_j^{\text{sub}}(t;\boldsymbol{Z})}$.

- Fine-Gray model assumes $\lambda_j^{\text{sub}}(t;\boldsymbol{Z}) = \lambda_j^{\text{sub}}(t) \times e^{\beta'\boldsymbol{Z}}$, where $\lambda_j^{\text{sub}}(t)$ is the baseline hazard function for type-j risk. Equivalently, the model is $\Lambda_j^{\text{sub}}(t;\boldsymbol{Z}) = \Lambda_j^{\text{sub}}(t) \times e^{\beta'\boldsymbol{Z}}$.

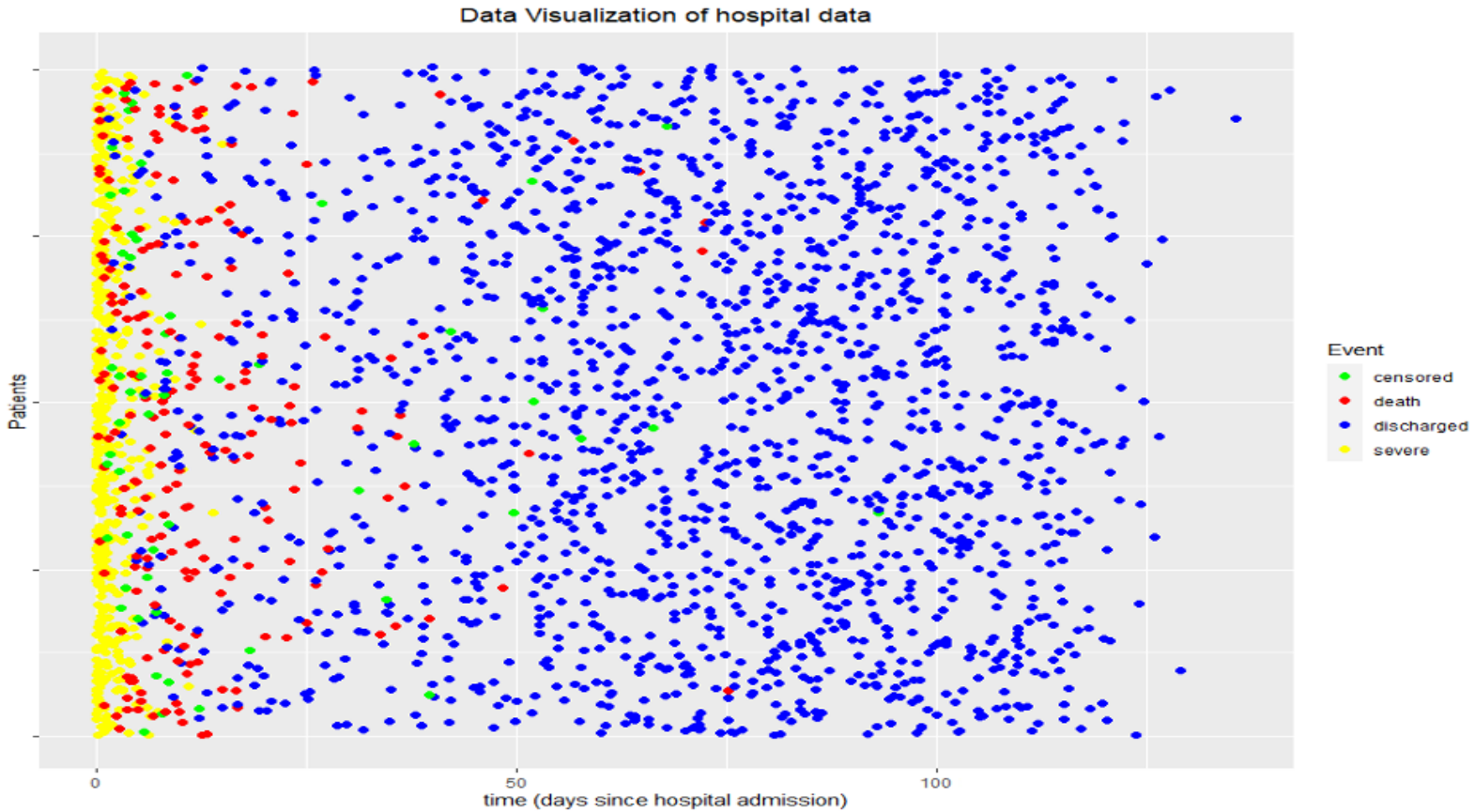
CIF regression model

- The CIF regression model does not possess good interpretation in hazard, but can be interpreted via the sub-distribution function:
  $F_j^{\text{sub}}(t;\boldsymbol{Z}) = 1 - e^{-\Lambda_j^{\text{sub}}(t;\boldsymbol{Z})} = 1 - e^{-\Lambda_j^{\text{sub}}(t)\times e^{\beta'\boldsymbol{Z}}}$.
  That is, $\beta'\boldsymbol{Z}_1 \geq \beta'\boldsymbol{Z}_2$ implies $F_j^{\text{sub}}(t;\boldsymbol{Z}_1) \geq F_j^{\text{sub}}(t;\boldsymbol{Z}_2)$.

- The prediction of CIF based on cause-specific Cox models relies on knowledge of ALL the competing risks models. In contrast, the prediction of CIF based on CIF regression model is self-contained (- a big advantage!) See Appendix for details.

Example 1: Johns Hopkins Covid-19 hospital data

Hospital data collected from 03/05/2020 to 07/17/2020 from 5 hospitals, the Johns Hopkins Medicine System.



Johns Hopkins COVID-19 hospital data

- Data include demographics, medical history, comorbidity conditions, symptoms, vital signs, respiratory measurements, medications, laboratory results, imaging, and WHO disease severity scores.
- n=1512; about 13% of patients died, most occurring within 20 days after admission.
- about 40% of patients developed severe disease (WHO scale 5$\sim$7) or died (scale 8).
- among all the deaths, about 90% with age$>$60
- hospital discharges (recovery) spreading out over the whole follow-up time.
- Two competing event outcomes: time to death vs. time to discharge

Example 2: NIH Covid-19 clinical trial

An NIH adaptive COVID-19 Treatment Trial (ACTT-1) evaluated intravenous remdesivir in adults who were hospitalized with COVID-19 and had evidence of lower respiratory tract infection.

T: time from hospital admission to death or discharge
Two competing events: death ($\Pi = 1$) vs. discharge ($\Pi = 2$).

$F_1^{\text{sub}}(t)$: CIF for time to death
$F_2^{\text{sub}}(t)$: CIF for time to discharge

Common problems in data analysis

Many physicians and practitioners have used the **Kaplan-Meier estimator** (Kaplan & Meier, 1958) and **Cox PHM** (Cox, 1972) when conducting medicine or public health related research.

Unfortunately, not too many physicians or practitioners know competing risks data well. In the presence of competing events, they frequently still use the Kaplan-Meier estimator and Cox PHM for data analysis:

- **K-M estimator $\neq$ CIF estimator !**
- Cox PHM is equivalent to cause-specific PHM by treating 'other events' as censoring (physicians and practitioners are so lucky!)

A common mistake when estimating the CIF

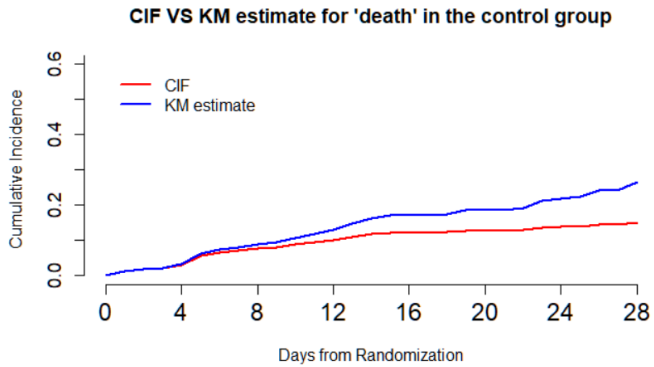
How do you nonparametrically estimate the CIF $F_1^{\text{sub}}(t)$ for time to death?

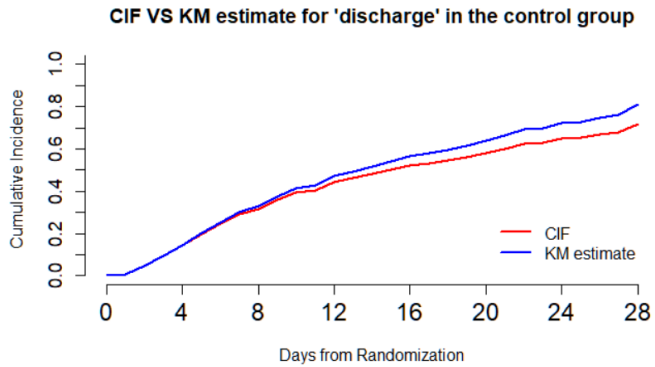
- A common mistake is to treat all the discharges as censoring and use the Kaplan-Meier estimator to estimate the CIF.
- A nonparametric estimator of CIF has been developed for estimation. This nonparametric estimator is NOT the K-M estimator.

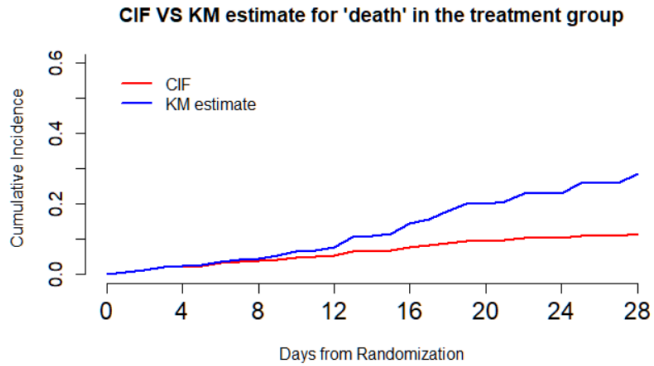
What do you actually estimate if you use the K-M estimator?

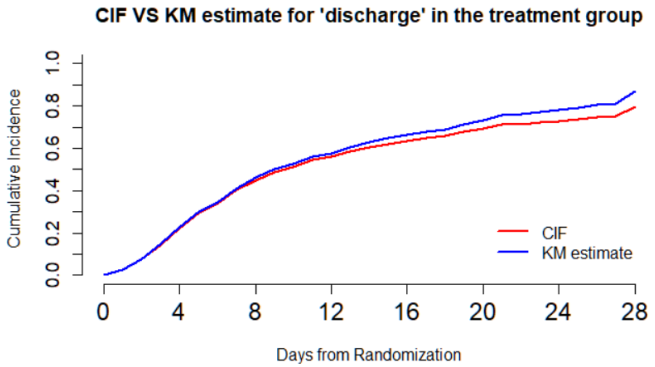
- You are 'trying' to estimate the survival function of latent variable U_1 (but it is usually biased due to informative censoring).
- The K-M estimator typically overestimates the CIF.

Estimation of CIF









Typical analysis for data examples 1 & 2

- **One-sample**: Use the CIF estimator $\hat{F}_j^{\text{sub}}(t)$, $j = 1, 2$, to estimate the sub-distribution of time to death/discharge 'in real world.'
- **One-sample**: Do we want to estimate the survival distribution of latent time U_j?
  - Not of interest.
- **Regression**: Use cause-specific hazards model?
  - As hospitals in different regions/countries might have different discharge policies, the underlying cause-specific hazard population depends on discharge policies, so the data analysis results do not extend easily.
- **Regression**: Use Fine-Gray subdistribution model?
  - The model gives better model interpretation, but the model can handle only baseline covariates (a disadvantage for prediction analysis).
- **Regression**: Latent variable models (i.e., use U_j as outcome variable)?
  - It is not interesting for Covid-19 studies, but it might be interesting in other studies.

Appendix 1: Risk prediction

Risk prediction using baseline covariates under cause-specific PHM

Risk prediction/estimation is an important topic for precision medicine. We estimate the CIF under the competing risks setting.

- The prediction methods that we introduced for standard survival data, in Chapter 5, do **not** work under the competing risks models.
- We want to estimate CIF for type-j risk.

  The CIF is $F_j^{\text{sub}}(t;Z) = P(T \le t, \Pi = j; Z) = \int_0^t \lambda_j(u;Z)S(u;Z)du$. Recall

  $$S(t;Z) = exp\left\{-\int_0^t \lambda(u;Z)du\right\} = exp\left\{-\int_0^t \sum_{j=1}^m \lambda_j(u;Z)du\right\} = exp\left\{-\sum_{j=1}^m \Lambda_j(t;Z)\right\},$$

  where $\Lambda_j(u;Z) = \int_0^t \lambda_j(u;Z)du$, $j = 1, ..., m$.

- Define $\Lambda_{j0}(u) = \int_0^t \lambda_{j0}(u)du$. Extend the Breslow estimator to derive

$$\hat{\Lambda}_{j0}(t) = \sum_{\{i:\, x_{(i)}<t, \Pi_{(i)}=j\}} \frac{1}{\sum_{j\in R(x_{(i)})} e^{\hat{\beta}_j' \boldsymbol{x_j}}}$$

- Then, we can use the estimate $\hat{\Lambda}_j(t;Z) = e^{\hat{\beta}_j' \boldsymbol{Z}} \times \hat{\Lambda}_{j0}(t)$ to derive

$$\hat{S}(t;Z) = exp\left\{-\sum_{j=1}^{m} \hat{\Lambda}_j(t;Z)\right\}$$

- The CIF for type-j risk can be estimated by

$$\hat{F}_j^{\text{sub}}(t;Z) = \int_0^t \hat{\lambda}_j(u;Z)\hat{S}(u;Z)du = \int_0^t \hat{S}(u;Z)d\hat{\Lambda}_j(u;Z)$$

- It is important to understand that, in order to estimate the CIF for risk type-j, one needs to estimate the cause-specific PHM for all risk types (type-j, $j = 1, ..., m$). This is considered a drawback of the cause-specific PHM.

- It is important to understand that, in order to estimate the CIF for cause type-j, one needs to estimate the cause-specific PHM for all cause types (type-j, $j = 1, ..., m$). This is considered a drawback of the cause-specific PHM.
- Risk prediction can be extended to landmark cause-specific PHM so time-dependent covariates can be used dynamically.
- Fine-Gray model can also be used and the risk prediction is self-contained. That is, in order to estimate the CIF for cause type-j, one needs to estimate the Fine-Gray model for cause type-j.

Appendix 2: Semi-competing risks models

X^0: time to incidence of a disease
Y^0: time to death (or a failure event)
C: time to censoring
Z: $1 \times p$ vector of covariates

Example: In a randomized treatment trial on the prophylaxis of pneumocystis carinii pneumonia (pcp), a group of patients with an initial episode of pcp could die without recurrence of pcp or be censored. In this example, X^0: time to recurrence of pcp, Y^0: time to death, and C: time to censoring.

Data: Observe i.i.d. $(X_i, Y_i, \Delta_i, \Pi_i, Z_i),\ i = 1, \ldots, n$

$$X_i = X_i^0 \wedge Y_i^0 \wedge C_i,\ \ \Delta_i = I(X_i^0 < \min\{Y_i^0, C_i\}),\quad Y_i = Y_i^0 \wedge C_i,\ \Pi_i = I(Y_i^0 < C_i).$$

Note: The data under a semi-competing risks model is richer than data under a standard competing risks model.

Latent variable model based on semi-competing risks data: Modeling the latent outcomes X^0 and Y^0. Some details are included in the Appendix.

Ref: Lin, Robins and Wei (1996, Biometrika), Fine et al. (2001, Biometrika), Peng and Fine (2007, Biometrics).