Survival Analysis Workshop

Jooyoung Lee

Department of Applied Statistics

Chung-Ang University

Seoul Medical Center

January 20, 2023

Week 3: Introduction to Competing Risks Model

Week 3 Materials

- An introduction to competing risks models
- Cumulative incidence function (CIF)
- Comparison of CIF: log-rank test, Gray's test
- Regression models: the cause-specific Cox models, the Fine-Gray's model

Recap: Univariate Survival Data

In the univariate survival analysis, $F(t) = 1 - S(t)$, and

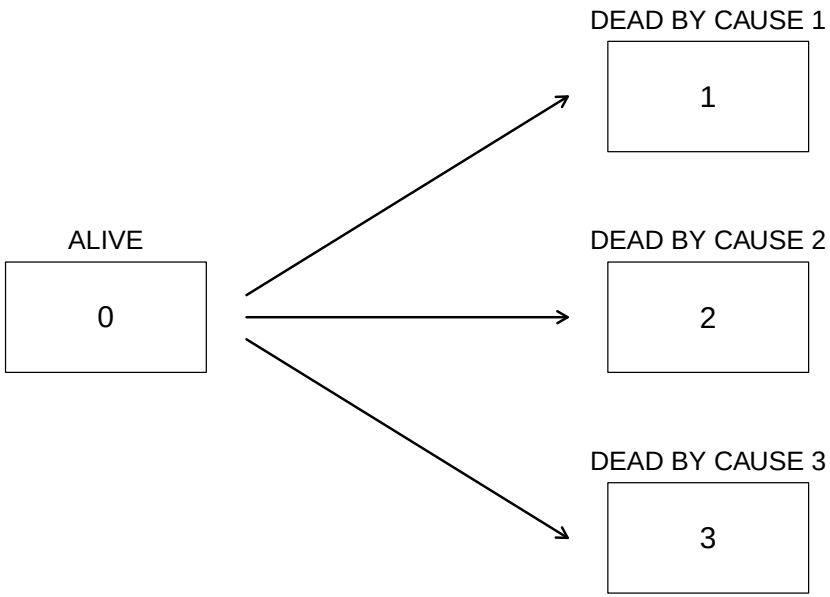
$$F(t) = 1 - \exp\left(\int_0^t \lambda(s)ds\right)$$

, which describes a one-to-one relationship between the risk and hazard.

Competing Risks data

Competing risks data arise when an individual is at risk of more than one mutually exclusive event. Therefore, only event time along with the cause of failure is observable.

Example: Difference causes of death




```
library(survival); data(pbc)
pbc[1:5, c(1,2,3,4)]
```

	id	time	status	trt
1	1	400	2	1
2	2	4500	0	1
3	3	1012	2	1
4	4	1925	2	1
5	5	1504	1	2

Competing Risks data

Competing risks models are interested in

- Model the cause-specific hazard function
- Estimate cumulative incidence function
- Evaluate the effect of risk factors on each cause of failure

Notation

- T: Event time (survival/failure time) of an individual from a population
- C: Censoring time
- X: Observed time
- δ: Censoring indicator
- D: Type of cause of the failure

Cause-Specific Hazard

Given that we have a total of K types of failures, the cause-specific hazard function

$$\lambda_k(t) = \lim_{h \to 0^+} \frac{P(t \leq T < t + h, D = k | T \geq t)}{h}, \quad k = 1, \ldots, K.$$

According to the relationship between the cumulative hazard functions and survival functions, we have

$$S(t) = \exp(-\sum_{k=1}^{K} \Lambda_k(t)),$$

where the cumulative hazard function is defined as

$$\Lambda_k(t) = \int_0^t \lambda_k(s) ds.$$

Cumulative Incidence Function (CIF, Sub-distribution function)

Cumulative incidence function is a probability of the k*th* cause of failure in the presence of competing risks,

$$F_k(t) = P(T \leq t, D = k) = \int_0^t S(s)\lambda_k(s)ds$$
$$= \int_0^t \exp(-\sum_{l=1}^{K} \Lambda_l(t))\lambda_k(s)ds$$

We have the following relationship:

$$S(t) + F_1(t) + \cdots + F_K(t) = 1.$$

Therefore, there is no one-to-one relationship between risk and hazard in competing risks settings.

Nelson-Aalen estimator

For ordinary survival data, the Nelson-Aalen estimator is used to estimate the cumulative hazard functions:

$$\widehat{\Lambda}(t) = \sum_{j:t_j \leq t} \frac{d_j}{n_j}.$$

If we apply the Nelson-Aalen estimator to competing risks setting,

$$\widehat{\Lambda}_k(t) = \sum_{j:t_j \leq t} \frac{d_j I(D = k)}{n_j}.$$

Estimate cumulative incidence function: Aalen-Johansen Estimator

The Aalen-Johanseon estimator is used to estimate the cumulative incidence function

$$\widehat{F}_k(t) = \sum_{j:t_j \leq t} \widehat{S}(t_{j-1}) \frac{d_j I(D = k)}{n_j},$$

where $\widehat{S}(t_{j-1})$ can be estimated by Kaplan-Meier estimator using all K types of failures.

Why do we care the competing risks model?

If we consider all-cause mortality, we assess the risk ($F(t) = 1 - S(t)$) by $1 - \widehat{S}^{KM}(t)$, the Kaplan-Meier estimates.

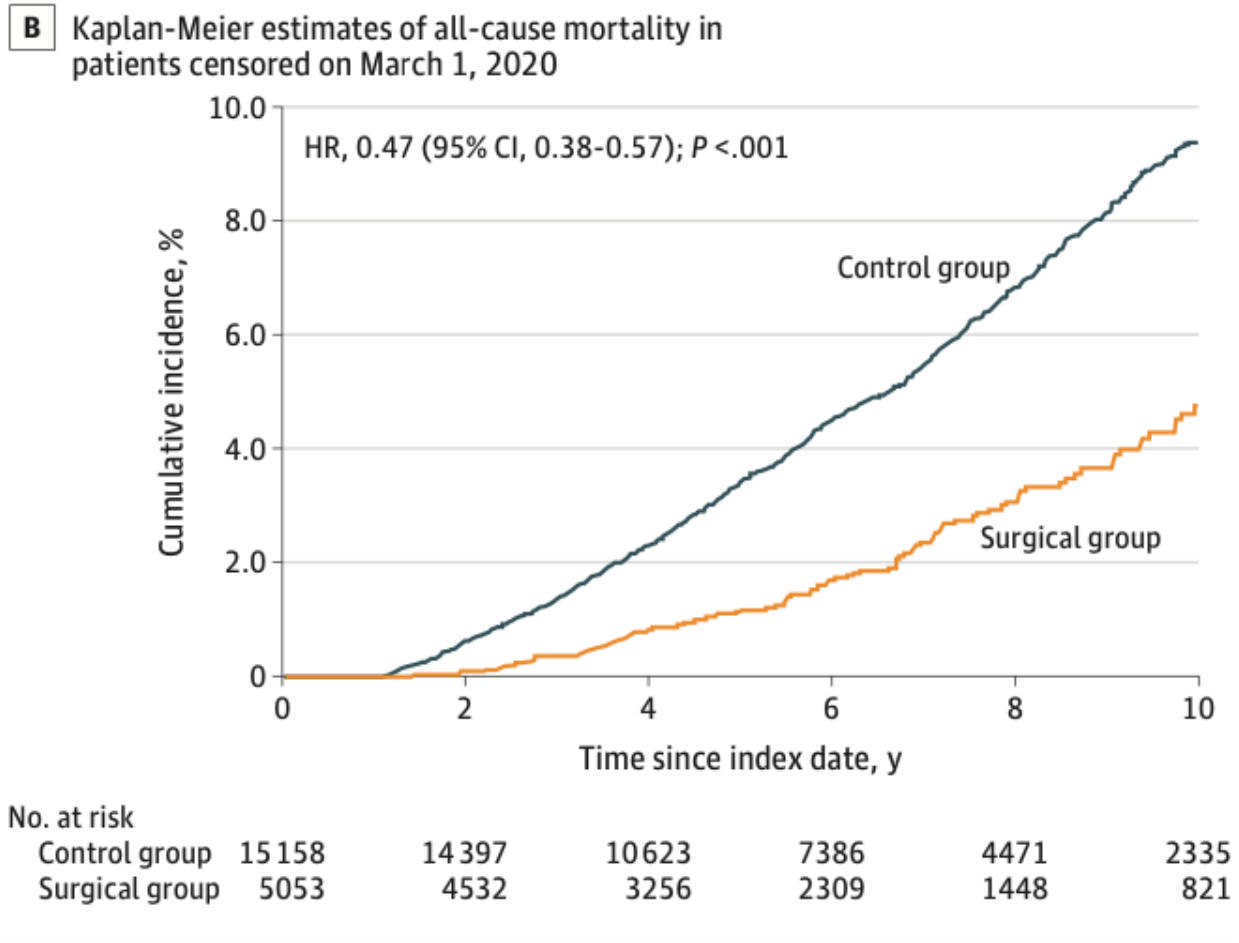


Figure 1: Figure from Aminian et al., *JAMA Surgery* 2022

Why do we care the competing risks model?

- However, if we use $1 - S^{KM}(t)$, for example, to estimate the risk of the death from cardiovascular disease with censoring for other causes of death, the risk of the death from cardiovascular disease is overestimated.

- This is because the violation of one of the assumptions underlying the Kaplan–Meier estimator, which is an independent censoring assumption.

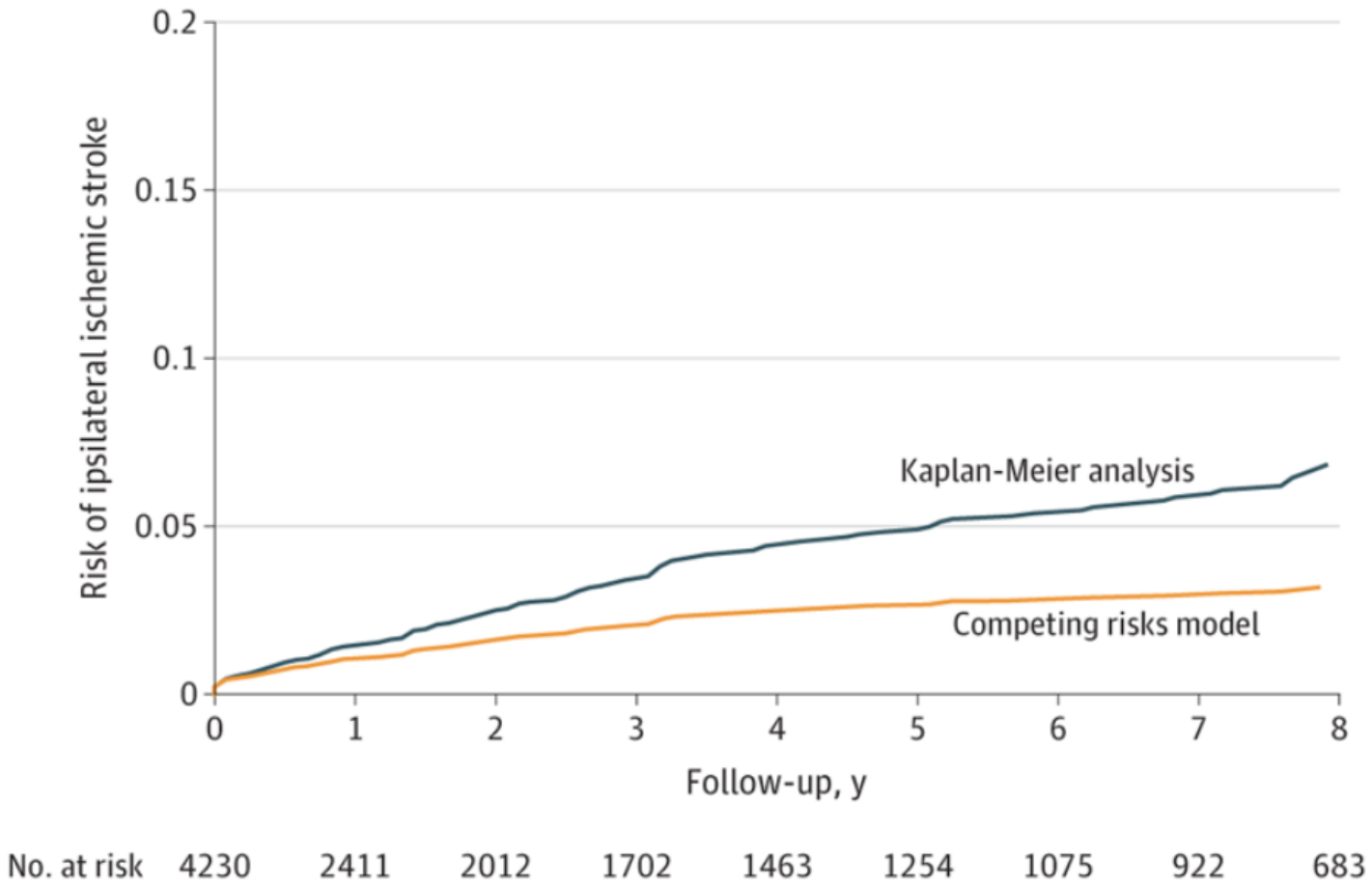
- If the competing event time distributions were independent of the distribution of time to the event of interest, this would imply that the hazard of the event of interest is the same for subjects that have not yet failed and are still under observation as for subjects that have experienced a competing event by that time.

Why do we care the competing risks model?

- However, a subject who have experienced a competing event will never experience the event of interest.
- Since subjects that will never fail are treated as if they could fail (they are censored), the naive Kaplan–Meier overestimates the probability of failure of interest.
- The bias is greater when the hazard of a competing event increases.

Why do we care the competing risks model?

Figure 2. Cumulative Risk of Ipsilateral Ischemic Stroke per Unique Artery After Initial Diagnosis of Asymptomatic Severe Carotid Stenosis



Competing risks model comprises death and carotid intervention. Median

Log-rank Test

Recall the log-rank test is used to test the equality of the survival probability (or the risk) for two groups:

$$H_0 : S_1(t) = S_2(t) \quad \forall t,$$

or we can also write

$$H_0 : F_1(t) = F_2(t) \quad \forall t.$$

Log-rank Test for competing risks

We can perform the log-rank test on time to event type k censoring the other event types at the times they occur.

$$H_0 : F_{k1}(t) = F_{k2}(t) \quad \forall t.$$

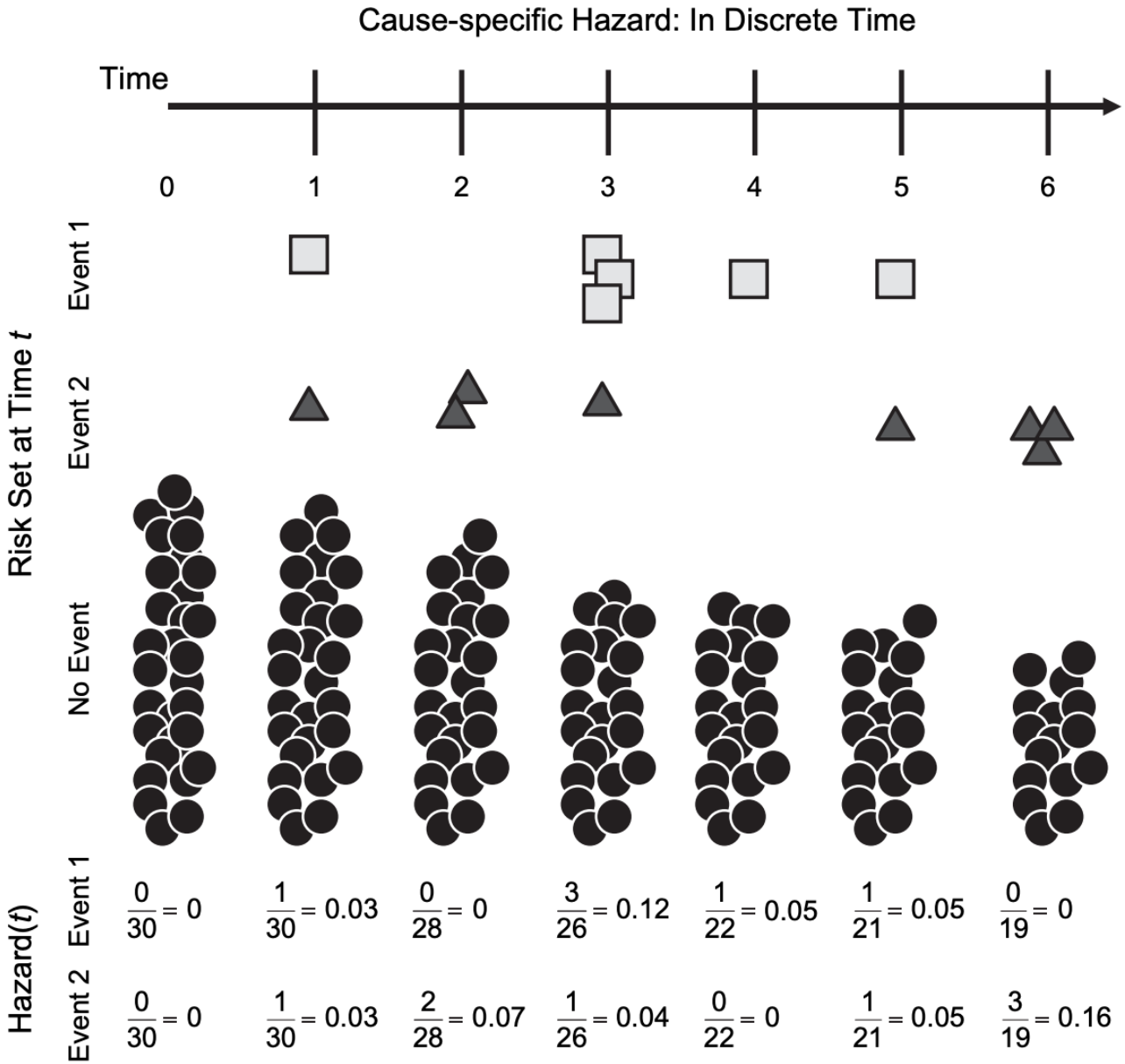


Figure 3: Figure from Lau et al., *American Journal of Epidemiology* 2009

Sub-distributionla hazard function

Define the sub-distributionla hazard function:

$$\tilde{\lambda}_k(t) = \lim_{h \to 0^+} \frac{P(t \leq T < t+h, D = k | T \geq t \bigcup (T \leq t \bigcap D \neq k))}{h}$$

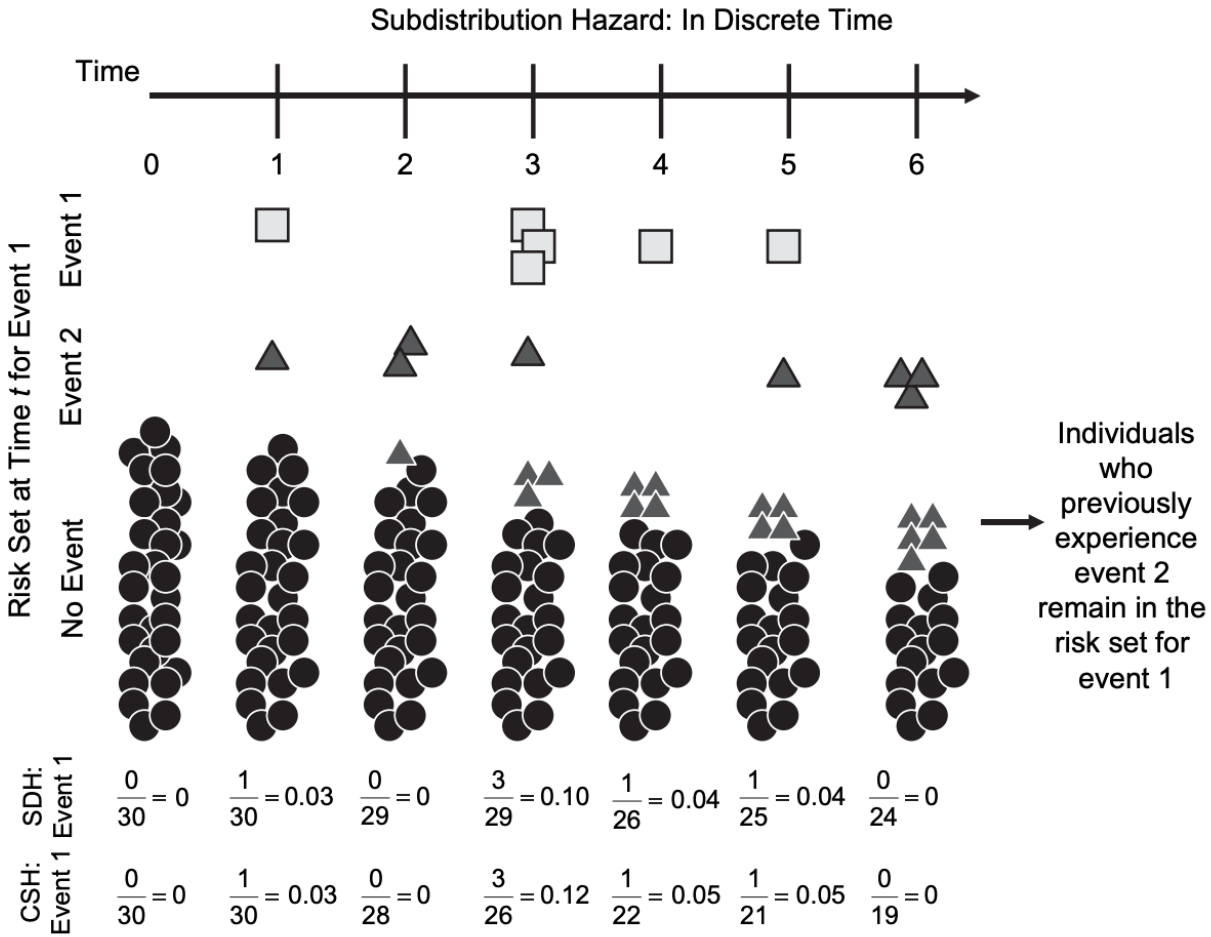
Then, we can define

$$\tilde{\Lambda}_k(t) = -\log(1 - F_k(t)).$$

Gray's Test

Based on the cumulative incidence function (sub-distribution function) , Gray (1988) suggested the test:

$$H_0 : F_{k1}(t) = F_{k2}(t) \quad \forall t.$$



Which test?

- The log-rank test correctly detects differences in cause-specific hazards. Unless there is strong dependence between failure times, between-group differences in hazards for other causes is not largely affected.

- Gray's test correctly detects whether there is a difference in cumulative incidence between groups for a given event. This test may reject the null hypothesis if there is a significant difference in the cause-specific hazards between the groups or a significant difference in hazards for other causes.

- Gray's test will detect the difference in CIF even for the situation in which the treatment has simply increased the incidence of a competing event, not our event of interest.

Two approaches for competing risks regression models

We have two approaches for regression models in competing risks settings.

- The cause-specific Cox proportional hazard model
- The Fine-Gray models (Sub-distribution proportional hazard model)

The Cause-Specific Cox Model

The k*th*-cause hazard function is

$$\lambda_k(t|\mathbf{z}) = \lambda_{k0}(t)\exp(\mathbf{z}'\boldsymbol{\beta}_k),$$

where $\mathbf{z} = (z_1, \ldots, z_p)'$, and $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp})'$.

Interpretation: $exp(\beta_{kj})$: the relative change in the cause-specific hazard for the k*th* event corresponding to a 1-unit increase in the covariate z_j.

The Cause-Specific Cox Model

The cause-specific Cox model can be fitted by considering each event type separately and treating the competing events as censored observations.

However, the cause-specific hazard model does not a simple interpretation to describe the relationship between covariate and cumulative incidences.

The Fine-Gray Model

Fine and Gray (1999) proposed the sub-distributional hazard regression model.

$$\tilde{\lambda}_k(t|\mathbf{z}) = \tilde{\lambda}_{k0}(t)\exp(\mathbf{z}'\tilde{\boldsymbol{\beta}}_k),$$

$$F_k(t|\mathbf{z}) = 1 - \exp\left(-\int_0^t \tilde{\lambda}_{k0}(t)\exp(\mathbf{z}'\tilde{\beta}_k)\right).$$

- The Fine-Gray model provides parameters describing the relationship between the covariates and the risk of the k*th* risk.

- Suppose we consider a covariate z and $\beta_k > 0$. Then, $F_k(t|z=1) > F_k(t|z=0)$, meaning that individuals with $z=1$ have an increased risk for the k*th* event compared to those with $z=0$.

Comparisons between the cause-specific Cox Model and Fine-Gray Model

CS_1 for cause 1	CS_2 for cause 2	FG_1 for cause 1
<1	<1	$FG_1 > CS_1$
<1	>1	$FG_1 < CS_1$
>1	<1	$FG_1 > CS_1$
>1	>1	$FG_1 < CS_1$
$=1$	>1	$FG_1 < CS_1$

Comparisons between the cause-specific Cox Model and Fine-Gray Model

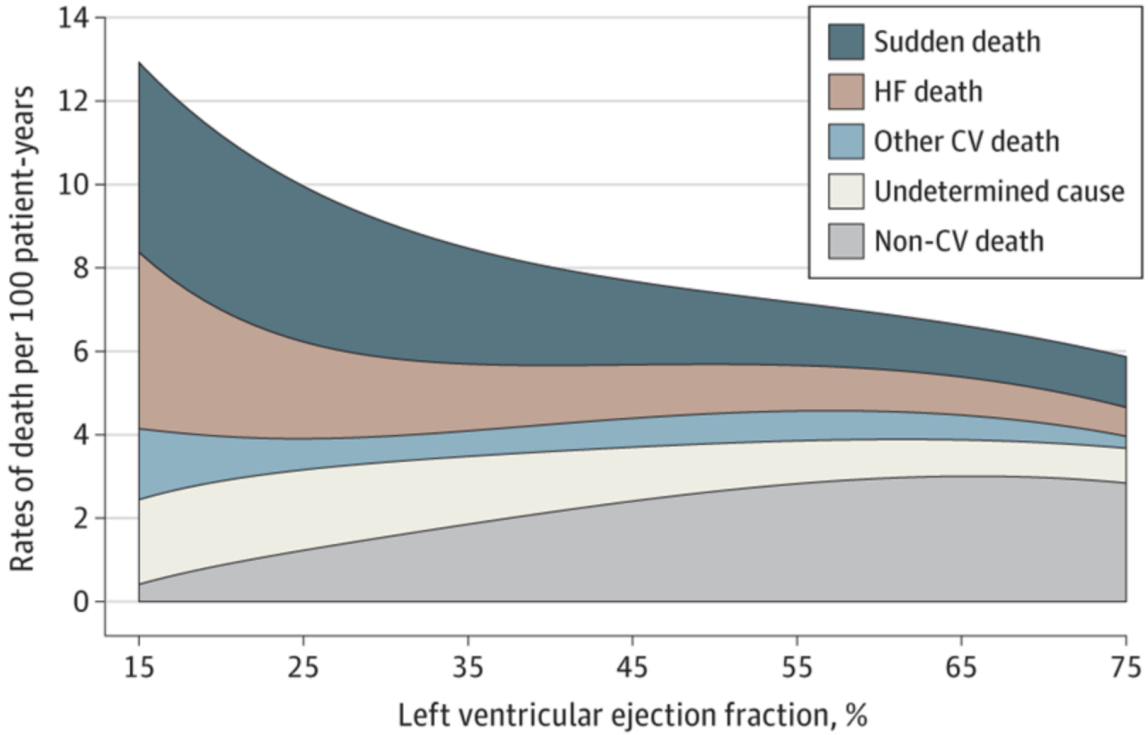
- If the hazard functions for other causes, then the hazard ratios from two models are equivalent.
- Gray's model assess the association of the treatment at the event of interest and the possibly differential treatment effect of competing events.
- The cause-specific Cox models does not evaluate the treatment effect on the risk of the event of interest since the risk of the event of interest also depends on regression parameters associated with the competing events.
- The quantitative interpretation of the sub-distribution hazard ratios is not simple.

Comparisons between the cause-specific Cox Model and Fine-Gray Model

- The cause-specific Cox model is more appropriate for studying the etiology of diseases.
- The Fine-Gray model is more suitable for predicting an individual's risk for an outcome.

Example: Cause-specific Cox Model

Figure 2. Variation in Incidence Rates of Death by Cause and Continuous Left Ventricular Ejection Fraction for Pooled DAPA-HF and DELIVER Populations



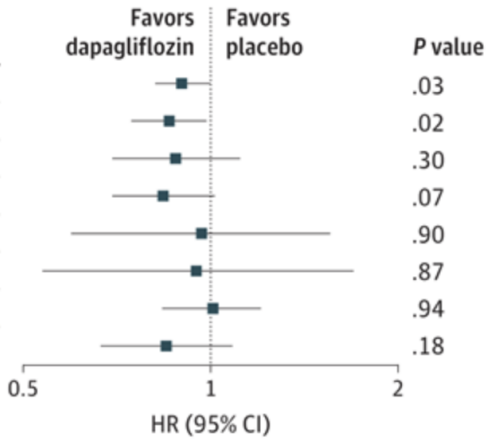
CV indicates cardiovascular; DAPA-HF, Dapagliflozin and Prevention of Adverse Outcomes in Heart Failure; DELIVER, Dapagliflozin Evaluation to Improve the Lives of Patients With Preserved Ejection Fraction Heart Failure; HF, heart failure; MI, myocardial infarction.

Figure 4: Figure from Desai et al., *JAMA Cardiol* 2022

Example: Cause-specific Cox Model

Figure 3. Effect of Dapagliflozin Compared With Placebo on Cause-Specific Mortality for the Dapagliflozin and Prevention of Adverse Outcomes in Heart Failure (DAPA-HF) and Dapagliflozin Evaluation to Improve the Lives of Patients With Preserved Ejection Fraction Heart Failure (DELIVER) Populations

Outcome	Dapagliflozin Events, rate/ 100 patient-year (n = 5504)	Placebo Events, rate/ 100 patient-year (n = 5503)	HR (95% CI)	*P* value
All-cause death	773 (7.4)	855 (8.3)	0.90 (0.82-0.99)	.03
CV death	404 (3.9)	468 (4.5)	0.86 (0.75-0.98)	.02
HF death	136 (1.3)	153 (1.5)	0.88 (0.70-1.11)	.30
Sudden death	202 (1.9)	239 (2.3)	0.84 (0.70-1.01)	.07
Stroke death	34 (0.3)	35 (0.3)	0.97 (0.60-1.55)	.90
MI death	23 (0.2)	24 (0.2)	0.95 (0.54-1.69)	.87
Non-CV death	245 (2.4)	242 (2.3)	1.01 (0.84-1.20)	.94
Unknown death	124 (1.2)	145 (1.4)	0.85 (0.67-1.08)	.18



CV indicates cardiovascular; HR, hazard ratio; MI, myocardial infarction.

Figure 5: Figure from Desai et al., *JAMA Cardiol* 2022

Example: Fine-Gray Model

Table 2. All-Cause Death Analyzed Using Cox Proportional Hazards Regression

Table 2. All-Cause Death Analyzed Using Cox Proportional Hazards Regression

Antiseizure medication	Hazard ratio (95% CI)	
	Crude (n = 2577)	Adjusted (n = 2577)[a]
Carbamazepine	1 [Reference]	1 [Reference]
Lamotrigine	0.75 (0.63-0.90)	0.72 (0.60-0.86)
Levetiracetam	0.96 (0.80-1.14)	0.96 (0.80-1.15)
Valproic acid	1.65 (1.45-1.87)	1.40 (1.23-1.59)
Phenytoin	1.62 (1.24-2.11)	1.16 (0.88-1.51)
Oxcarbazepine	1.24 (0.87-1.77)	1.16 (0.81-1.66)

[a] Adjusted for age, sex, stroke type, living conditions before and after stroke, dependence in activities of daily living before and after stroke, status epilepticus, hypertension, atrial fibrillation, type 1 or 2 diabetes, statins, antidepressants, and smoking. Missing covariate data have been imputed.

Figure 6: Figure from Desai et al., *JAMA Cardiol* 2022

Example:Fine-Gray Model

Table 3. Cardiovascular Death Analyzed Using Fine-Gray Competing Risk Regression

Table 3. Cardiovascular Death Analyzed Using Fine-Gray Competing Risk Regression

Antiseizure medication	Hazard ratio (95% CI)	
	Crude (n = 2577)	Adjusted (n = 2577)[a]
Carbamazepine	1 [Reference]	1 [Reference]
Lamotrigine	0.78 (0.63-0.97)	0.76 (0.61-0.95)
Levetiracetam	0.81 (0.64-1.03)	0.77 (0.60-0.99)
Valproic acid	1.68 (1.44-1.96)	1.40 (1.19-1.64)
Phenytoin	1.39 (0.98-1.98)	1.02 (0.71-1.47)
Oxcarbazepine	0.76 (0.46-1.26)	0.71 (0.42-1.18)

[a] Adjusted for age, sex, stroke type, living conditions before and after stroke, dependence in activities of daily living before and after stroke, status epilepticus, hypertension, atrial fibrillation, type 1 or 2 diabetes, statins, antidepressants, and smoking. Missing covariate data have been imputed.

Figure 7: Figure from Larssonet al., *JAMA nerology* 2022

Reference

- Aminian, A., Tu, C., Milinovich, A., Wolski, K. E., Kattan, M. W., & Nissen, S. E. (2022). Association of weight loss achieved through metabolic surgery with risk and severity of COVID-19 infection. JAMA surgery, 157(3), 221-230.

- Chang, R. W., Tucker, L. Y., Rothenberg, K. A., Lancaster, E., Faruqi, R. M., Kuang, H. C., ... & Nguyen-Huynh, M. N. (2022). Incidence of Ischemic Stroke in Patients With Asymptomatic Severe Carotid Stenosis Without Surgical Intervention. JAMA, 327(20), 1974-1982

- Lau, B., Cole, S. R., & Gange, S. J. (2009). Competing risk regression models for epidemiologic data. American journal of epidemiology, 170(2), 244-256.

- Desai, A. S., Jhund, P. S., Claggett, B. L., Vaduganathan, M., Miao, Z. M., Kondo, T., ... & Solomon, S. D. (2022). Effect of dapagliflozin on cause-specific mortality in patients with heart failure across the spectrum of ejection fraction: a participant-level pooled analysis of dapa-hf and deliver. JAMA cardiology, 7(12), 1227-1234.

- Larsson, D., Baftiu, A., Landmark, C. J., von Euler, M., Kumlien, E., Åsberg, S., & Zelano, J. (2022). Association between antiseizure drug monotherapy and mortality for patients with poststroke epilepsy. JAMA neurology, 79(2), 169-175.