Survival Analysis I (CHL5209H)

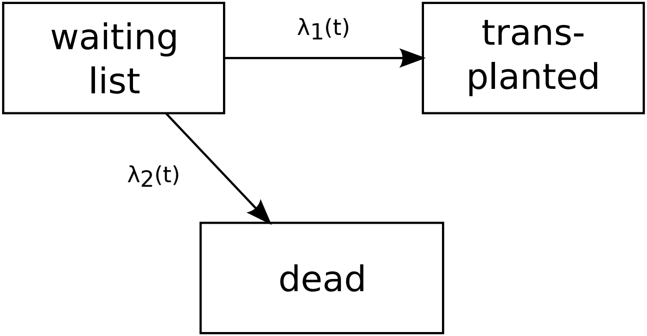
Olli Saarela

Dalla Lana School of Public Health
University of Toronto

olli.saarela@utoronto.ca

March 19, 2019

Example of competing risks



Competing risks

- As seen earlier, the hazard modeling framework generalizes to more than one mutually exclusive event type of interest.
- The time $\tilde{T}_i$ refers to the time of the first event (of any type), and the event type indicator values $\tilde{E}_i \in \{1, \dots, J\}$ refer to J mutually exclusive event types.
- Equivalently, we could introduce the cause-specific counting processes $\tilde{N}_{ij}(t)$, $j = 1, \dots, J$.
- The cause-specific hazard function for each event type $j \in \{1, \dots, J\}$ is defined as

$$\lambda_j(t) \equiv \lim_{h \to 0} \frac{P(t \leq \tilde{T}_i < t + h, \tilde{E}_i = j \mid \tilde{T}_i \geq t)}{h}.$$

Interpretation

- This can be interpreted through the instantaneous probability of an event of a particular type occurring, given that individual i is still at risk at time t (has not experienced event of any type).
- Each $\lambda_j(t)$ could be modeled through a Cox model of the form

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp\{\beta_j' x_i\},$$

  estimated separately for each $j \in \{1, \ldots, J\}$.
- If only one event type is of interest, the other cause-specific hazards need not be estimated.
- The above interpretation does not require assuming that the competing causes are independent (given x_i).

Caveats

- The cause-specific cumulative baseline hazards $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(u)\,\mathrm{d}u$ could be estimated through the Breslow estimator.
- However, the cause-specific cumulative hazards $\Lambda_{ij}(t) = \Lambda_{0j}(t)\exp\{\beta_j' x_i\}$ do not have a direct relationship to an event time distribution.
- If we are willing to assume that the competing causes are independent (given x_i), taking $1 - \exp\{-\Lambda_{ij}(t)\}$ would correspond to a risk of event type j occurring by time t in the absence of the competing causes.
- However, considering the other competing causes to be absent rarely makes sense.

Cause-specific cumulative incidence

- The risk relevant to real life applications is the cause-specific cumulative incidence function

$$\begin{aligned}\pi_{ij}(s) &= P(\tilde{T}_i \leq s, \tilde{E}_i = j \mid x_i) \\ &= \int_0^s \lambda_{ij}(t) \exp\left\{-\sum_{j=1}^J \Lambda_{ij}(t)\right\} \mathrm{d}t.\end{aligned}$$

- Note that $1 - \pi_{ij}(s)$ is not a survival probability. (Why?)
- Plugging in the Cox model based estimators, this could be estimated as

$$\hat{\pi}_{ij}(s) = \sum_{k:t_k \leq s} \mathrm{d}\hat{\Lambda}_{ij}(t_k) \exp\left\{-\sum_{j=1}^J \hat{\Lambda}_{ij}(t_{k-1})\right\}.$$

- Note on terminology: although cumulative incidence and cumulative hazard sound similar, they refer to different concepts.

Non-parametric estimators

- Using Kaplan-Meier estimator in the presence of competing risks would also involve the assumption that the compering causes are independent.
- It would estimate the survival probability in the absence of the competing causes (which again rarely makes sense).
- Usually it is more appropriate to estimate cumulative incidence instead, where we can again consider estimators of the form

$$\hat{\pi}_j(s) = \int_0^s \mathrm{d}\hat{\Lambda}_j(t)\hat{S}(t),$$

now without covariate information.

Non-parametric estimator for cumulative incidence

- Substitute in Nelson-Aalen estimator for the increment of the cause-specific cumulative hazard $\mathrm{d}\hat{\Lambda}_j(t)$ and Kaplan-Meier estimator for the overall survival function $\hat{S}(t)$.
- This gives a non-parametric estimator of the form

$$\hat{\pi}_j(s) = \sum_{k:t_k \leq s} \frac{d_{jk}}{n_k} \hat{S}_{\mathrm{KM}}(t_{k-1}),$$

where d_{jk} is the number of events of type j that occurred at time t_k, and where

$$\hat{S}_{\mathrm{KM}}(t) = \prod_{k:t_k \leq t} \left(1 - \frac{\sum_{j=1}^{J} d_{jk}}{n_k} \right).$$

Subdistribution hazard

- Although the cause-specific hazard is interpretable in itself, it does not directly tell how the covariates are related to the cumulative incidence.
- One can define an alternative hazard function

$$\psi_j(t) \equiv \lim_{h \to 0} \frac{P(t \leq \tilde{T}_i < t + h, \tilde{E}_i = j \mid \tilde{T}_i \geq t \text{ or } (\tilde{T}_i < t, \tilde{E}_i \neq j))}{h}$$

  called the subdistribution hazard.
- Note that the added condition here does not mean that someone who has experienced an event of type other than j would still be at risk for j (since the first event has already taken place).
- However, such individuals would be included in the denominator (riskset) in the estimation.

Connection to cumulative incidence

- The reason we would be interested in subdistribution hazard is that it is related to the cumulative incidence function by

$$\psi_j(t) = -\frac{\mathrm{d} \log(1 - \pi_j(t))}{\mathrm{d}t},$$

or alternatively,

$$\pi_j(t) = 1 - \exp\{-\Psi_j(t)\},$$

where $\Psi_j(t) = \int_0^t \psi_j(u) \, \mathrm{d}u$.

Fine & Gray model

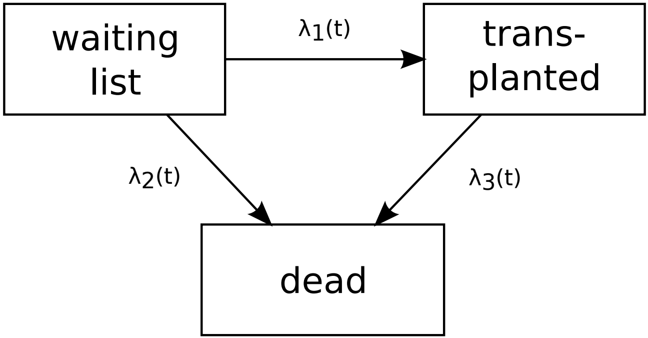
- Similar to Cox model, the subdistribution hazard can be modeled as a function of covariates, and assumed to be proportional:

$$\psi_{ij}(t) = \psi_{0j}(t)\exp\{\gamma_j' x_i\}.$$

- Such a model is known as the Fine & Gray model.
- The interpretation of the regression coefficients here relates to the subdistribution hazard function, whereas the usual Cox model coefficients relate to the cause-specific hazard function.
- The regression estimates combined with a Breslow-type estimator for the cumulative subdistribution baseline hazard can be used to estimate individual specific cumulative incidence through

$$\hat{\pi}_{ij}(s) = 1 - \exp\{-\hat{\Psi}_{0j}(s)\exp\{\hat{\gamma}_j' x_i\}\}.$$

Example of multi-state model



Multi-state models

- The survival analysis setting generalizes further to situations where we want to consider more than one possible event for the same individual.
- The events can be related to transition between states in a multi-state model.
- The possible events in the example are transplantation, death while on the waiting list, and death after transplantation, each characterized by its own transition intensity function.
- Similar to cause-specific hazards, if we are only interested in one type of transition, the others need not be modeled.
- However, the probabilities related to the multi-state model can generally be complicated functions of all intensity functions.
- Certain comparisons related to the multi-state model (here comparison of $\lambda_2(t)$ and $\lambda_3(t)$) could be addressed through time-dependent covariates.