

Competing Risks and Multistate Models



Henrique Laureano

<http://leg.ufpr.br/~henrique>

Last modification on 2020-03-10 14:06:11

Competing Risks and Multistate Models,
8th chapter of *The Statistical Analysis of Failure Time Data*
Kalbfleisch and Prentice, 2002

Outline

- » Competing Risks
 - » Context & definitions
 - » Different **likelihood** functions
 - » **Multiple** decrement functions and **identifiability**
- » Multistate Models



Notation & preliminaries,

- » an underlying **failure time** T that may be subject to (independent) censoring
- » a vector of possibly **time-dependent** covariates,
 $X(t) = \{x(u) : 0 \leq u < t\}$
- » when failure occurs, it may be of any one of **m distinct types or causes** denoted by $J \in \{1, 2, \dots, m\}$

As before, the **overall** rate or hazard function at time t is

$$\lambda[t; X(t)] = \lim_{h \rightarrow 0} h^{-1} \mathbb{P}[t \leq T < t + h \mid T \geq t, X(t)].$$

To model **competing risks**, we consider a **type-specific** or cause-specific hazard function or process

$$\lambda_j[t; X(t)] = \lim_{h \rightarrow 0} h^{-1} \mathbb{P}[t \leq T < t + h, J = j \mid T \geq t, X(t)].$$

for $j = 1, \dots, m$ and $t > 0$.



type-specific hazard function, $\lambda_j[t; X(t)]$, in words,

$\lambda_j[t; X(t)]$ represents the **instantaneous** rate for failures of type j at time t given $X(t)$ and in the presence of all other failure types.

Without ties of the failure types,

$$\lambda[t; X(t)] = \sum_{j=1}^m \lambda_j[t; X(t)].$$

The overall survivor function is

$$F(t; X) = \mathbb{P}[T > t \mid X] = \exp \left\{ - \int_0^t \lambda(u; X) du \right\},$$

and the (sub)density function for the **time to a type j failure** is

$$\begin{aligned} f_j(t : X) &= \lim_{h \rightarrow 0} h^{-1} \mathbb{P}[t \leq T < t + h, J = j \mid X] \\ &= \lambda_j(t; X) F(t; X), \quad j = 1, \dots, m. \end{aligned}$$



When the covariates are of the **fixed** or **external** type,

the likelihood on a sample subject to **independent right censorship** is written as a product of the **survivor functions** for the censored data and the **subdensities** for the observed failure times, i.e., the likelihood function can be written entirely in terms of the **type-specific hazard functions**.

The **cumulative incidence function** for type j failures corresponding to the **external** covariate is

$$\begin{aligned}\bar{F}_j(t; X) &= \mathbb{P}[T \leq t, J = j; X] \\ &= \int_0^t f_j(u; X) du, \quad t > 0,\end{aligned}$$

for $j = 1, \dots, m$. Note that

$$p_j = \mathbb{P}[J = j] = \lim_{t \rightarrow \infty} \bar{F}_j(t; X), \quad j = 1, \dots, m$$

and $\sum_{j=1}^m p_j = 1$.



$F_j(t; X)$ has **no** simple probability interpretation within the **competing risks** model, at least not without introducing strong additional assumptions.

Example 8.1. Suppose that $m = 2$ and that the covariate is a treatment indicator $x = 0, 1$.

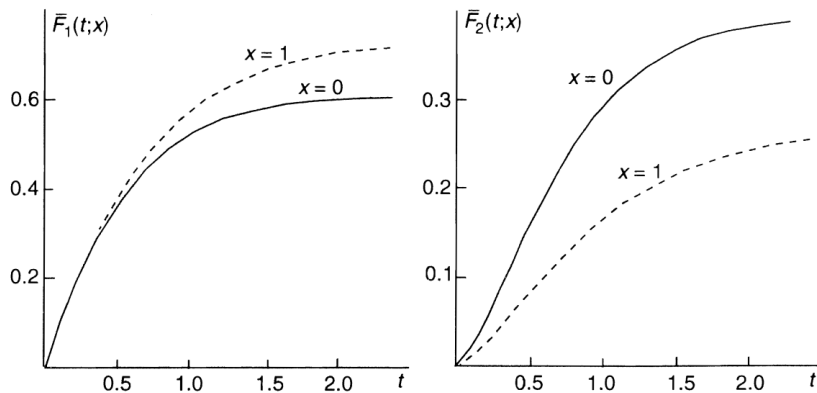


Figure 8.1 Cumulative incidence functions for Example 8.1.



Likelihoods

Consider $\{t_i, \delta_i, j_i, X_i(t_i)\}_{i=1}^n$.

If the **censoring is independent**, the likelihood (or partial likelihood) is proportional to

$$\begin{aligned} L &= \prod_{i=1}^n \left(\{\lambda_{j_i}[t_i; X_i(t_i)]\}^{\delta_i} \prod_{j=1}^m \exp \left\{ - \int_0^{t_i} \lambda_j[u; X_i(u)] du \right\} \right) \\ &= \prod_{j=1}^m \left(\{\lambda_j[t_i; X_i(t_i)]\}^{\delta_{ji}} \exp \left\{ - \int_0^\infty \sum_{i=1}^n Y_i(t) \lambda_j[t; X_i(t)] dt \right\} \right). \end{aligned}$$

Any of the methods of preceding chapters can be used for inference about the $\lambda_j[t; X(t)]$'s.



Likelihoods

We can also generalize simple explanatory methods such as [Kaplan-Meier](#) and [Nelson-Aalen](#) estimators to [competing risks data](#).

Let $t_1 < t_2 < \dots < t_k$ denote the k distinct failure times for all failure types combined. Then, the likelihood function can be written

$$L = \prod_{i=1}^k \left(\prod_{j=1}^m \{ [F_j(t_i^-) - F_j(t_i)] F(t_i^-) \}^{d_{ji}} \prod_{l=1}^{C_i} [F(t_{il})]^{c_{il}} \right).$$

Its nonparametric MLE places mass only at the observed failure times $1, \dots, k$, so the partially maximized likelihood can be rewritten using expressions for discrete models, to obtain

$$\text{Multinomial likelihood : } \hat{L} = \prod_{i=1}^k \left[\prod_{j=1}^m \lambda_{ji}^{d_{ji}} (1 - \lambda_i)^{n_i - d_i} \right].$$

Maximizing it gives the MLE $\hat{\lambda}_{ji} = d_{ji}/n_i$.



The cumulative hazard function is then estimated by

$$\hat{\Lambda}_j(t) = \sum_{i=1}^k \mathbf{1}(t_i \leq t) d_{ji} / n_i, \quad t \geq 0.$$

- » This yields the **Nelson-Aalen** estimate of the total cumulative hazard and the **Kaplan-Meier** estimate of the overall survivor function $F(t)$.

The estimated **cumulative incidence function** is also discrete, and is given by

$$\hat{\hat{F}}_j(t) = \sum_{\{i | t_i \leq t\}} d_{ji} n_i^{-1} \hat{F}(t_i^-), \quad j = 1, \dots, m.$$



Likelihoods

Consider now a relative risk or Cox model for the **cause-specific hazard functions**

$$\lambda_j[t; X(t)] = \lambda_{0j}(t) \exp\{Z(t)^\top \beta_j\}, \quad j = 1, \dots, m.$$

The corresponding partial likelihood is

$$L(\beta) = \prod_{j=1}^m \prod_{i=1}^{k_j} \frac{\exp\{Z_{ji}(t_{ji})^\top \beta_j\}}{\sum_{l \in R(t_{ji})} \exp\{Z_l(t_{ji})^\top \beta_j\}}.$$

If applicable, a **proportional** risks model

$$\lambda_j[t; X(t)] = \lambda_0(t) \exp\{\gamma_j + Z(t)^\top \beta_j\}, \quad j = 1, \dots, m,$$

would yield more efficient β_j estimators, in which the **cause-specific hazards** are assumed to be **proportional** to each other (for uniqueness set $\gamma_1 = 0$).



Likelihoods

The partial likelihood of the **proportional** risk model can then be written

$$\prod_{i=1}^k \frac{\exp\{\gamma_{\mathbf{j}_i} + \mathbf{Z}_i(t_i)^\top \boldsymbol{\beta}_{\mathbf{j}_i}\}}{\sum_{\mathbf{j}=1}^m \sum_{l=1}^n Y_l(t_i) \exp\{\gamma_{\mathbf{j}} + \mathbf{Z}_l(t_i)^\top \boldsymbol{\beta}_{\mathbf{j}}\}}.$$

As is the general relative risk model,
an adjustment is needed to handle **tied** failure times.

Although it would often be more restrictive than is desirable, the **proportional risk model** has some attractive properties. For instance, the **probability** that an individual with fixed covariate \mathbf{Z} has failure type \mathbf{j} is

$$\mathbb{P}[\mathbf{J} = \mathbf{j}; \mathbf{Z}] = \frac{\exp\{\gamma_{\mathbf{j}} + \mathbf{Z}^\top \boldsymbol{\beta}_{\mathbf{j}}\}}{\sum_{\mathbf{h}=1}^m \exp\{\gamma_{\mathbf{h}} + \mathbf{Z}^\top \boldsymbol{\beta}_{\mathbf{h}}\}}, \quad \mathbf{j} = 1, \dots, m,$$

regardless of $\lambda_0(\cdot)$.

The corresponding MLEs of the proportionality factors $\exp\{\gamma_{\mathbf{j}}\}$, subject to $\gamma_1 = 0$, are $\exp\{\hat{\gamma}_{\mathbf{j}}\} = k_{\mathbf{j}}/k_1$, $\mathbf{j} = 2, \dots, m$.



Example 8.2, $m = 3$.

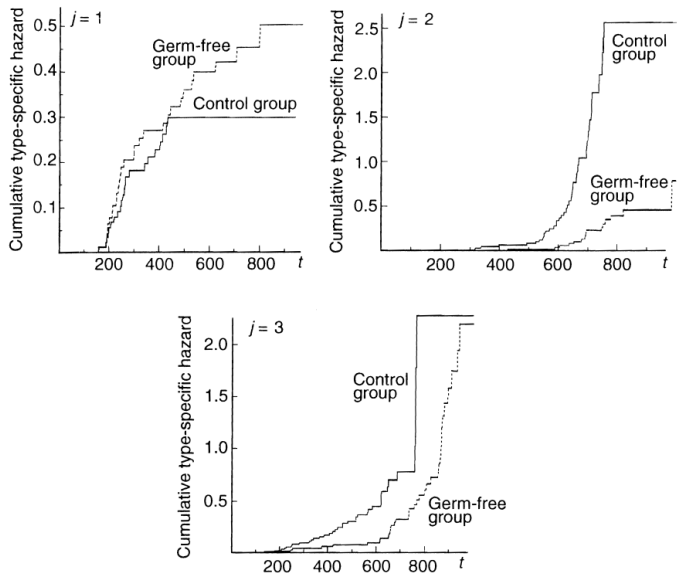


Figure 8.2 Estimates of the cumulative type-specific hazard functions for the data of Example 8.2.



Example 8.2, $m = 3$.

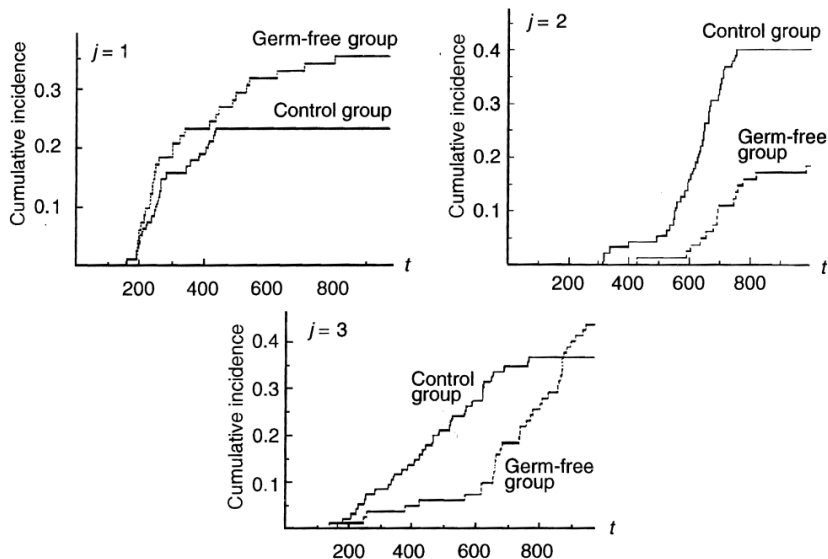


Figure 8.3 Estimates of the cumulative incidence functions (8.11) for the data of Example 8.2.



Multiple Decrement Function

Idea: A joint distribution for the latent failure times, $\bar{T}_1, \dots, \bar{T}_m$.

The **multiple decrement function** or **joint survivor function**,

$$Q(t_1, \dots, t_m; x) = \mathbb{P}[\bar{T}_1 > t_1, \dots, \bar{T}_m > t_m; x].$$

- » This model gives a complete specification of the probability laws for the m variate failure time model.
- » Thus, quantities introduced earlier can be expressed in terms of Q , such as the **overall survivor function**

$$F(t; x) = \mathbb{P}[T > t; x] = Q(t, t, \dots, t; x),$$

and the **type-specific hazard functions**

$$\begin{aligned} \lambda_j(t; x) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq T_j < t + \Delta t \mid T \geq t; x]}{\Delta t} \\ &= \left. \frac{-\partial \log Q(t_1, \dots, t_m; x)}{\partial t_j} \right|_{t_1 = \dots = t_m = t}, \end{aligned}$$

$$j = 1, \dots, m.$$



Nonidentifiability

Quantities that cannot be expressed as functions of the **type-specific hazard functions**, are **nonidentifiable**, and so, cannot be estimated without introducing additional model assumptions.

e.g.,

- » the **marginal survivor functions** are generally **nonidentifiable**.

- » **why?**

- We don't know the **dependence structure** of the latent failure times.



how 2 solve

Nonidentifiability of the marginal survivor functions

- » With a **internal time-dependent covariate** $\tilde{x}_j(t)$, that given other variables in the model, are highly predictive of the rate of type j failures
 - » A test for **no** association between $\tilde{x}_j(t)$ and the failure rate for type j' failures

Dependent censoring

- » Insertion of a **time-dependent covariate**
 - » If the censoring mechanism were well explained by the **time-dependent covariate**, the censoring scheme would become **independent**.

Possible problem? Confounding.



Asymptotics: Cox model

In the case of **no ties**, the **score function**, U , based on the partial likelihood, is a **mean 0 vector-valued martingale** wrt \mathcal{F}_t and can be written as

$$U(\beta, t) = \sum_{i=1}^n \int_0^t \left[Z_i(u) - \sum_{l=1}^n Z_l(u) \frac{Y_l(u) \exp\{Z_l(u)^\top \beta\}}{\sum_{j=1}^n Y_j(u) \exp\{Z_j(u)^\top \beta\}} \right] dM_i(u).$$

Under some more boring and too mathly conditions, as $n \rightarrow \infty$,

$$\begin{aligned} n^{-1/2} U(\beta, \tau) &\xrightarrow{\mathcal{D}} N(0, \Sigma(\beta, \tau)), \\ n^{1/2}(\hat{\beta} - \beta) &\xrightarrow{\mathcal{D}} N(0, \Sigma(\beta, \tau)^{-1}), \text{ with} \end{aligned}$$

$$\begin{aligned} \Sigma(\beta, \tau) = \int_0^\tau &\left[\sum_{i=1}^n Y_i(u) Z_i(u) Z_i(u)^\top \exp\{Z_i(u)^\top \beta\} - \right. \\ &\left. \frac{\left[\sum_{i=1}^n Y_i(u) Z_i(u) \exp\{Z_i(u)^\top \beta\} \right]^{\otimes 2}}{\sum_{i=1}^n Y_i(u) \exp\{Z_i(u)^\top \beta\}} \right] \lambda_0(u) du, \end{aligned}$$

$$t \in [0, \tau].$$



Asymptotics: parametric models

- ✓ Asymptotic results apply to independent right censoring and left truncation.
- ✗ They do not cover interval censoring or right truncation.

And again, we have

$$\mathbb{P}[dN_i(t) = 1 | \mathcal{F}_{t-}] = Y_i(t) \lambda_i(t | \theta) dt, \quad i = 1, \dots, n.$$

The log-likelihood function arising from continuous failure time data on the interval $(0, \tau]$ can be written

$$l(\theta) = \sum_{i=1}^n \int_0^{\tau} \log \lambda_i(t | \theta) dN_i(t) - \sum_{i=1}^n \int_0^{\tau} Y_i(t) \lambda_i(t | \theta) dt.$$



Asymptotics: parametric models

The **score process** on data on $(0, \tau]$ is

$$U(\theta, t) = \sum_{i=1}^n \int_0^t \left[\frac{\partial}{\partial \theta} \log \lambda_i(u|\theta) \right] dM_i(u), \quad 0 < t < \tau$$

where $M_i(t)$ are **orthogonal martingales**.

Via **central limit** for the score function

(and some other conditions, of course), as $n \rightarrow \infty$,

$$n^{-1/2} U(\theta, \tau) \xrightarrow{\mathcal{D}} N(0, \Sigma(\theta)),$$

$$n^{1/2}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, \Sigma(\theta)^{-1}), \text{ with } \Sigma(\theta) \text{ est. consistently by } n^{-1}I(\hat{\theta}).$$

$$I(\theta) = \sum_{i=1}^n \int_0^\tau \left[\frac{\partial}{\partial \theta} \log \lambda_i(u|\theta) \right]^{\otimes 2} dN_i(u) - \sum_{i=1}^n \int_0^\tau \frac{\partial^2 \lambda_i(u|\theta)}{\partial \theta \partial \theta^\top} \lambda_i(u|\theta)^{-1} dM_i(u).$$



Efficiency of the Cox model estimator, $\hat{\beta}_{\text{cox}}$

- » With a hazard function specified up to a certain unknown parameter α , e.g., $\lambda_0(t) = \alpha h_0(t)$, which approach would lead to more precise asymptotic estimation of β , $\hat{\beta}_{\text{cox}}$ or a MLE $\hat{\beta}$?

If Z is time independent, $\hat{\beta}_{\text{cox}}$ is then asymptotically fully efficient;
If Z is time dependent, it is not.

Why?

The average Z value over the risk set varies with time in a way that such variations introduce asymptotic correlations between the estimator of β and that of α .



