Modeling the cumulative incidence function of multivariate competing risks data allowing for within-cluster dependence of risk and timing

paper presentation

Henrique Laureano
http://leg.ufpr.br/~henrique

Last modification on 2020-03-27 22:25:39

"Modeling the **cumulative incidence function** of **multivariate competing risks data** allowing for **within-cluster dependence** of risk and timing"

what?

"Modeling the **cumulative incidence function** of **multivariate competing risks data** allowing for **within-cluster dependence** of risk and timing"

what?

- **cumulative incidence function**: CIF.

$$\text{for a type } j \text{ failure,} \quad \bar{F}_j(t; X) = \mathbb{P}[T \leq t, J = j; X] = \int_0^t f_j(u; X) du, \quad t > 0,$$

  where $f_j(t; X) = \lambda_j(t; X) F(t; X)$ is the (sub)density for the time to a type j failure.

- ...

"Modeling the **cumulative incidence function** of **multivariate competing risks data** allowing for **within-cluster dependence** of risk and timing"

what?

- ...
- **multivariate competing risks**,

  we have more than one, that is why it is ***multivariate****, cause of interest* ***competing*** *to be responsible by the failure (if not censor).*

"Modeling the **cumulative incidence function** of **multivariate competing risks data** allowing for **within-cluster dependence** of risk and timing"

what?

- . . .
- **multivariate competing risks**,

  *we have more than one, that is why it is **multivariate**, cause of interest **competing** to be responsible by the failure (if not censor).*

- multivariate competing risks **data**, i.e.,

  we'll not do a multivariate competing risks model,
  we'll do a model for multivariate competing risks data!

"Modeling the **cumulative incidence function** of **multivariate competing risks data** allowing for **within-cluster dependence** of risk and timing"

what?

- . . .
- **within-cluster dependence**, i.e., a random/latent effect structure for
  - risk: how a failure occurrence relates to other;
  - timing: some failures aren't likely to happen equally all time.

paper pres. structure

$F_j(t; X)$ 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 $x = 0, 1$.

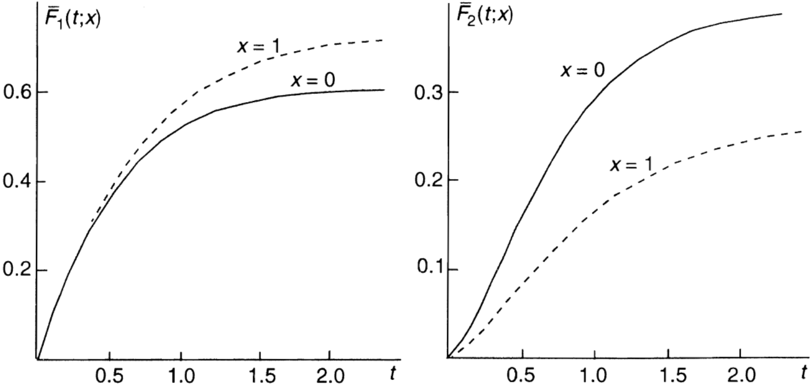

Figure 8.1 Cumulative incidence functions for Example 8.1.

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

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

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

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

Let $t_1 < t_2 < \cdots < t_k$ 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 $1, \ldots, k$, so the partially maximized likelihood can be rewritten using expressions for discrete models, to obtain

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

Likelihoods

Maximization of the multinomial likelihood gives the MLE $\hat{\lambda}_{ji} = d_{ji}/n_i$.

The cumulative hazard function is then estimated by
$\hat{\Lambda}_j(t) = \sum_{i=1}^{k} \mathbf{1}(t_i \leq t) d_{ji}/n_i,\ t \geq 0.$

- 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

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

Likelihoods

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

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

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

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

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

$$\prod_{i=1}^{k} \frac{\exp\{\gamma_{j_i} + Z_i(t_i)^\top \beta_{j_i}\}}{\sum_{j=1}^{m} \sum_{l=1}^{n} Y_l(t_i) \exp\{\gamma_j + Z_l(t_i)^\top \beta_j\}}.$$

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 proportial risk model has some attractive properties. For instance, the probability that an individual with fixed covariate Z has failure type j is

$$\mathbb{P}[J = j; Z] = \frac{\exp\{\gamma_j + Z^\top \beta_j\}}{\sum_{h=1}^{m} \exp\{\gamma_h + Z^\top \beta_h\}}, \quad j = 1, \ldots, m,$$

regardless of $\lambda_0(\cdot)$.

The corresponding MLEs of the proportionality factors $\exp\{\gamma_j\}$, subject to $\gamma_1 = 0$, are $\exp\{\hat{\gamma}_j\} = k_j / k_1$, $j = 2, \ldots, m$.

Example 8.2, $m = 3$.

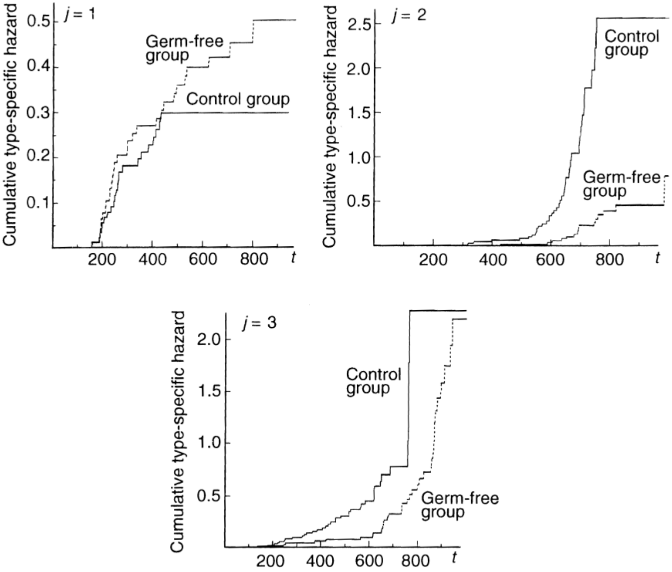

Figure 8.2 Estimates of the cumulative type-specific hazard functions for the data of Example 8.2.

Example 8.2, $m = 3$.

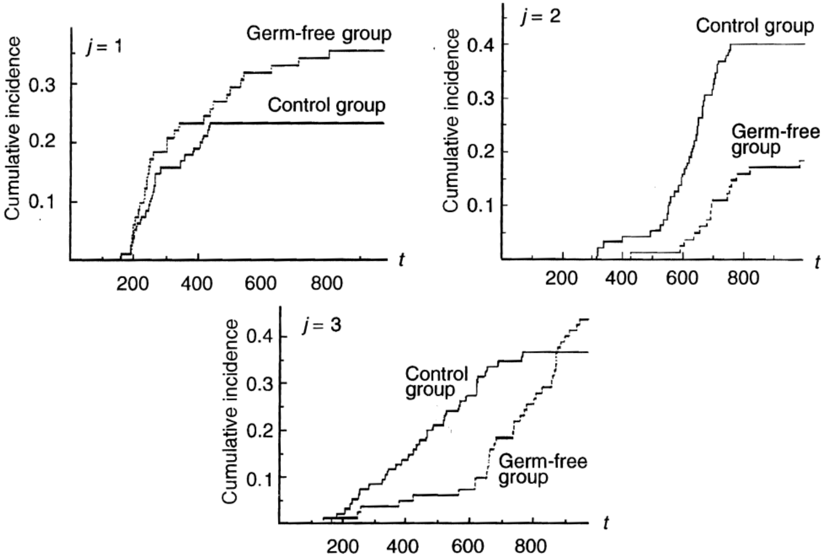

Figure 8.3 Estimates of the cumulative incidence functions (8.11) for the data of Example 8.2.

Multiple Decrement Function

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

The multiple decrement function or joint survivor function,

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

- 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

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

and the type-specific hazard functions

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

$j = 1, \ldots, m$.

Nonidentificability

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.

Counting Processes & Asymptotic Results

Under independent censoring and the other conditions that we already know,

$$\mathbb{P}[dN_{jl}(t) = 1 \mid \mathcal{F}_{t^-}] = Y_l(t)\lambda_j[t; X_l(t)]dt$$

for $0 < t$ and all l, j. It follows that

$$M_{jl}(t) = N_{jl}(t) - \int_0^t Y_l(u)\lambda_j[u; X_l(u)]du, \quad j = 1, \dots, m, \quad l = 1, \dots, n$$

are orthogonal martingales wrt the filtration $\mathcal{F}_t$.

The score vector is expressed as a stochastic integral of a predictable process wrt a martingale, where

$$\begin{aligned} U_j(t) &= \int_0^t \sum_{l=1}^n [Z_l(u) - \mathcal{E}(\beta_j, u)]dN_{jl}(u) \\ &= \int_0^t \sum_{l=1}^n [Z_l(u) - \mathcal{E}(\beta_j, u)]dM_{jl}(u), \quad j = 1, \dots, m, \end{aligned}$$

$t > 0$ and $\mathcal{E}(\beta_j, u)$ is the weighted average of $Z_l(u)$ over the risk set as before.

The score process $U(t)$ is a martingale whose predictable variation process can be seen to be a block diagonal matrix.

The asymptotic arguments and results of Chapter 5 apply directly here.

To finish the competing risks part,

» The Nelson-Aalen estimator for the jth failure type can be written as

$$\hat{\Lambda}_{0j}(t) = \int_0^t \frac{dN_{j\cdot}(u)}{\sum_{l=1}^n Y_l(u) \exp\{Z_l(u)^\top \hat{\beta}_j\}}.$$

» In the case of external (or fixed) covariates, estimators of the baseline cumulative incidence functions can be obtained as

$$\hat{\hat{F}}_{0j}(t) = \int_0^t \exp\left\{-\sum_{j'=1}^m \hat{\Lambda}_{0j'}(u)\right\} d\hat{\Lambda}_{0j}(u).$$

Multistate Models

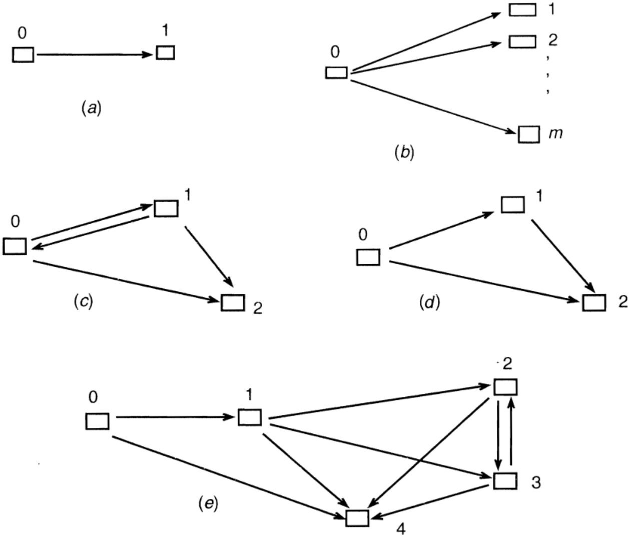

Figure 8.4 Compartment illustrations of multistate models for (*a*) failure time process; (*b*) competing risks; (*c*) illness–death model; (*d*) progressive illness death model; and (*e*) HIV, AIDS, and opportunistic infections.

Markov Processes

First, with no covariates

Let the Markov process $A(t)$ be the state occupied at time t, $t > 0$.

The transtion rate or intensity from i at time t^- to j at time t is given by the memoryless process

$$d\Lambda_{ij}(t) = \mathbb{P}[A(t^- + dt) = j \mid A(u), 0 \leq u < t, A(t^-) = i]$$
$$= \mathbb{P}[A(t^- + dt) = j \mid A(t^-) = i].$$

It is convenient to define

$$d\Lambda_{ii}(t) = -\sum_{j \neq i} d\Lambda_{ij}(t)$$

so that the row sums of the matrix

$$d\Lambda(t) = [d\Lambda_{ij}(t)]_{q \times q}$$

are all 0.

r-step transition probability matrix

Discrete case

$$\boldsymbol{P}^{(r)} = \prod_{k=1}^{r} \boldsymbol{P}_k = \boldsymbol{P}_1 \boldsymbol{P}_2 \dots \boldsymbol{P}_r, \quad r = 0, 1, \dots,$$

$$= \prod_{k=1}^{r} [I + d\Lambda(a_k)] = \prod_{k=1}^{r} (\mathbb{P}[A(a_k) = j \mid A(a_k^-) = i])_{q \times q}$$

$$= ({p_{ij}}^r)_{q \times q} = (\mathbb{P}[A(a_r) = j \mid A(0) = i])_{q \times q}.$$

Continuous case

$$\boldsymbol{P}(t) = (\mathbb{P}_{ij}(t))_{q \times q} = (\mathbb{P}[A(t) = j \mid A(0) = i])_{q \times q}$$

$$= \mathcal{P}_0^t [I + d\Lambda(u)] = \lim \prod_{i=1}^{M} [I + \Lambda(u_i) - \Lambda(u_{i-1})],$$

where the limit is taken as $M \to \infty$ and $\Delta u_i = u_i - u_{i-1} \to 0$.

Transition probabilities in different scenarios

Failure time model: $\hat{\mathbb{P}}_{00}(t)$ is the Kaplan-Meier estimate of the survivor function.

Competing risks model: $\hat{\mathbb{P}}_{0j}(t) = \hat{\bar{F}}_j(t)$, the estimates of the jth cumulative incidence function.

Now, with time-dependent covariates, i.e., Continuous-time modulated Markov models

Intensity function,

$$\lambda_{ijl}(t) = \lim_{h \to 0} h^{-1} \mathbb{P}[A_l(t^- + h) = j \mid A(t^-) = i, X_l(u), 0 < u < t].$$

Continuous-time modulated Markov models

Parametric model

With $\mathbb{P}[dN_{ijl}(t) = 1 \mid \mathcal{F}_{t^-}] = Y_{il}(t)\lambda_{ijl}(t)$,
the full log-likelihood on data over the interval $[0, \tau]$ is

$$\log L_M = \sum_{i \neq j} \left\{ \int_0^\tau \sum_{l=1}^n [\log \lambda_{ijl}(t;\theta) dN_{ijl}(t) - Y_{il}(t)\lambda_{ijl}(t;\theta)dt] \right\}.$$

Semiparametric relative risk model

With a contribution term to the partial likelihood given by

$$\mathbb{P}[dN_{ijl}(t) = 1 \mid dN_{ij\cdot}(t) = 1, \mathcal{F}_{t^-}] = \frac{Y_{il}(t)\exp\{Z_l(t)^\top \beta_{ij}\}}{\sum_{u=1}^n Y_{iu}(t)\exp\{Z_u(t)^\top \beta_{ij}\}}.$$

The log partial likelihood is then

$$\sum_{\text{all } i,j} \left\{ \int_0^\tau \sum_{l=1}^{n} Z_l(t)^\top \beta_{ij} dN_{ijl}(t) - \log\left[\sum_{l=1}^{n} Y_{il}(t) \exp\{Z_l(t)^\top \beta_{ij}\} dN_{ij\cdot}(t) \right] \right\}.$$

The procedures that we already know, can be used for estimation here. The asymptotics follows the general results outlined in Chapter 5.

The modulated Markov model uses time since on study as the basic time variable. In some instances, however, there may be a strong dependence on time since entry to a state. Such dependencies can be accommodated by considering

modulated semi-Markov models

key: time since the individual entered the stated occupied at time t^-,
$B(t) = \inf[s \mid A(t-s) \neq A(t^-)]$.

With covariates and under a parametric model,
the likelihood can be written as

$$\log L_{SM} = \sum_{i \neq j} \sum_{l=1}^{n} \left\{ \int_0^\infty \log \lambda_{ijl}[B_l(t);\theta] dN_{ijl}(t) - Y_{il}(t) \lambda_{ijl}[B_l(t);\theta] dt \right\}.$$

Modulated Semi-Markov Models

Now, under a relative risk model and partial likelihood analysis with

- $\nu = B_l(t)$ representing the current sojourn time;
- and with the idea of having the sth visit to state i.

The log partial likelihood can be written

$$\log L = \sum_{l=1}^{n} \int_0^\infty Z_l(a_{ijl} + \nu)^\top \beta_{ijs} dN^*_{ijsl}(\nu) - \int_0^\infty \log \left\{ \sum_{u=1}^{n} Y^*_{isu} \exp\{Z_u(a_{isu} + \nu)^\top \beta_{ijs}\} \right\} dN^*_{ijs}(\nu),$$

where a_{isl} is the time at which individual l enters the state i for the sth time.


THANK YOU
memegenerator.net