

Modeling the cumulative incidence function of clustered competing risk data: a multinomial GLMM approach



Henrique Laureano

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

Wagner Bonat

Cause-specific **cumulative incidence function** (CIF), $F(\cdot)$

for a type k failure,
$$F_k(t \mid \{X, R\}) = \mathbb{P}[T \leq t, K = k \mid \{X, R\}]$$
$$= \underbrace{\pi_k(\{X, R\})}_{\text{risk level}} \times \underbrace{\Phi[t \mid \{X, R_k\}]}_{\text{failure time trajectory}}.$$

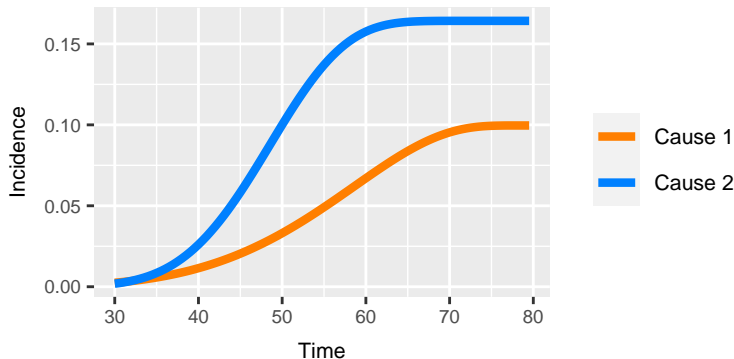
As proposed by [Cederkvist et al., 2019](#).



Cause-specific **cumulative incidence function** (CIF), $F(\cdot)$

$$\begin{aligned} \text{for a type } k \text{ failure, } F_k(t \mid \{X, R\}) &= \mathbb{P}[T \leq t, K = k \mid \{X, R\}] \\ &= \underbrace{\pi_k(\{X, R\})}_{\text{risk level}} \times \underbrace{\Phi[t \mid \{X, R_k\}]}_{\text{failure time trajectory}}. \end{aligned}$$

As proposed by [Cederkvist et al., 2019](#).

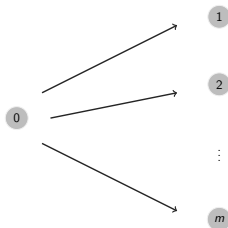


Survival data designs

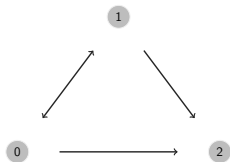
Failure time process



Competing risks process



Multistate process

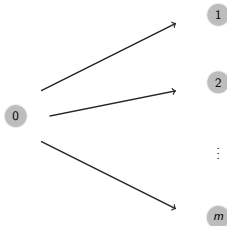


Survival data designs

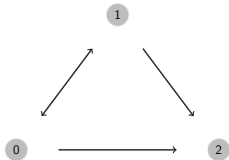
Failure time process



Competing risks process



Multistate process



Survival models

for a type k failure,

$$F_k(t \mid \{X, R\}) = \mathbb{P}[T \leq t, K = k \mid \{X, R\}]$$
$$= \underbrace{\pi_k(\{X, R\})}_{\text{risk level}} \times \underbrace{\Phi[t \mid \{X, R_k\}]}_{\text{failure time}}.$$



- » We work on the probability scale, modeling the

Cause-specific **cumulative incidence function** (CIF)

$$\begin{aligned}\text{for a type } k \text{ failure, } F_k(t | X) &= \mathbb{P}[T \leq t, K = k | X] \\ &= \int_0^t f_k(a | X) da, \quad t > 0,\end{aligned}$$

where $f_k(t | X) = \lambda_k(t | X) \times S(t | X)$ is the (sub)density for the time to a type k failure

- » We go beyond the independence supposition, having clusters of subjects with a possible correlation
- » Our latent structure to capture this possible correlation has audacity, allowing for **within-cluster dependence** of both risk and timing



Considering a competing risks setup of three competing causes. . .



Considering a competing risks setup of three competing causes...

... we want to model the CIF in the following specific manner

$$F_k(t \mid X, U) = \mathbb{P}[T \leq t, K = k \mid X, U]$$
$$= \underbrace{\pi_k(X, u_1, u_2)}_{\text{cluster-specific risk level}} \times \underbrace{\Phi[w_k g(t) - X^\top \gamma_k - \eta_k]}_{\text{cluster-specific failure time trajectory}}, \quad k = 1, 2.$$

i.e., a product of two functions

- » the first models the absolute risk of a competing cause in the presence of another
- » the second models the failure time distribution,

both varying between clusters



Considering a competing risks setup of three competing causes...

... we want to model the CIF in the following specific manner

$$F_k(t \mid X, U) = \mathbb{P}[T \leq t, K = k \mid X, U]$$
$$= \underbrace{\pi_k(X, u_1, u_2)}_{\text{cluster-specific risk level}} \times \underbrace{\Phi[w_k g(t) - X^\top \gamma_k - \eta_k]}_{\text{cluster-specific failure time trajectory}}, \quad k = 1, 2.$$

i.e., a product of two functions

- » the first models the absolute risk of a competing cause in the presence of another
- » the second models the failure time distribution,

both varying between clusters

-
- » The third competing cause is left to accommodate the censorship
 - » Which probability distribution fits this setup? Multinomial



Scheike works out this idea via a pairwise composite likelihood approach,

```
@article{Scheike,  
  author = {L. Cederkvist and K. K. Holst and K. K. Andersen and  
            T. H. Scheike},  
  title   = {Modeling the cumulative incidence function of  
            multivariate competing risks data allowing for  
            within-cluster dependence of risk and timing},  
  year    = {2019},  
  journal = {Biostatistics},  
  volume  = {20},  
  number  = {2},  
  pages   = {199--217},  
}
```

And we? We try a hierarchical approach via a GLMM



Our approach: what we do



Our approach: what we do

For an observed response y_{ijt} of a subject i , with cluster j , in the time t , we have

$$y_{ijt} \mid \{u_{1j}, u_{2j}, \eta_{1j}, \eta_{2j}\} \sim \text{Multinomial}(p_{1ijt}, p_{2ijt}, p_{3ijt}),$$
$$\begin{bmatrix} u_{1j} \\ u_{2j} \\ \eta_{1j} \\ \eta_{2j} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_1}^2 & \varrho_{u_1, u_2} & \varrho_{u_1, \eta_1} & \varrho_{u_1, \eta_2} \\ & \sigma_{u_2}^2 & \varrho_{u_2, \eta_1} & \varrho_{u_2, \eta_2} \\ & & \sigma_{\eta_1}^2 & \varrho_{\eta_1, \eta_2} \\ & & & \sigma_{\eta_2}^2 \end{bmatrix} \right).$$



Our approach: what we do

For an observed response y_{ijt} of a subject i , with cluster j , in the time t , we have

$$y_{ijt} \mid \{u_{1j}, u_{2j}, \eta_{1j}, \eta_{2j}\} \sim \text{Multinomial}(p_{1ijt}, p_{2ijt}, p_{3ijt}),$$

$$\begin{bmatrix} u_{1j} \\ u_{2j} \\ \eta_{1j} \\ \eta_{2j} \end{bmatrix} \sim \mathcal{N} \left[\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_1}^2 & \varrho_{u_1, u_2} & \varrho_{u_1, \eta_1} & \varrho_{u_1, \eta_2} \\ & \sigma_{u_2}^2 & \varrho_{u_2, \eta_1} & \varrho_{u_2, \eta_2} \\ & & \sigma_{\eta_1}^2 & \varrho_{\eta_1, \eta_2} \\ & & & \sigma_{\eta_2}^2 \end{bmatrix} \right].$$

This model structure is connected via a [link function](#), i.e., a functional form for the multinomial probabilities.

Here we use the CIF derivative w.r.t. time t ,

$$p_{kijt} = \frac{\exp\{\mathbf{x}_{kij}\beta_{ki} + u_{kj}\}}{1 + \sum_{m=1}^{K-1} \exp\{\mathbf{x}_{mij}\beta_{mi} + u_{mj}\}} \quad \{k = 1, 2\}$$

$$\times w_k \frac{\delta}{2\delta t - 2t^2} \phi\left[w_k \operatorname{arctanh}\left(\frac{t - \delta/2}{\delta/2}\right) - \mathbf{x}_{kij}\gamma_{ki} - \eta_{kj}\right].$$



Our approach: how we do



TMB: Template Model Builder

Quickly implement complex random effect models through simple C++ templates. The R package combines

- » CppAD: C++ automatic differentiation
- » Eigen: templated matrix-vector library
- » CHOLMOD: sparse matrix routines available from R

to obtain an efficient implementation of the applied Laplace approximation with exact derivatives.

Also, key features are

- » automatic sparseness detection
- » parallelism through BLAS
- » parallel user templates.



Main application: cancer incidence in twins



Clustered competing risks data

↳ Clusters? Families

↳ Family studies

↳ **Twins data**

within-cluster dependence \Rightarrow within-family dependence

within-family dependence may reflect both disease **heritability** and the impact of shared **environmental effects**



References

» [[Cederkvist et al., 2019](#)]

L. Cederkvist, K. K. Holst, K. K. Andersen and T. H. Scheike (2019). Modeling the cumulative incidence function of multivariate competing risks data allowing for within-cluster dependence of risk and timing. *Biostatistics*, v. 20, n. 2, p. 199-217

