

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

master thesis defense



**Henrique Laureano ([.github.io](https://github.io))**  
LEG @ UFPR

April 3, 2021

- 1 Data
- 2 Application
- 3 Model
- 4 Contributions & challenges
- 5 Template Model Builder

# Clustered competing risk data



Key terms:

- **Clustered**: groups with a dependence structure (e.g. families);
- Causes **competing** by *something*.

Something?

- **Failure** of an industrial or electronic component;
- **Occurrence** or **cure** of a disease or some biological process;
- **Progress** of a patient clinic state.

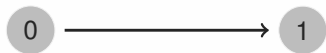
Independent of the application, always the same framework

Cluster	ID	Cause 1	Cause 2	Censorship	Time	Feature
1	1	Yes	No	No	10	A
1	2	No	No	Yes	8	A
2	1	No	No	Yes	7	B
2	2	No	Yes	No	5	A

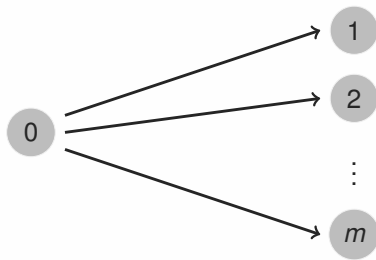
# Data designs



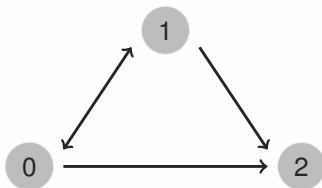
Failure time process



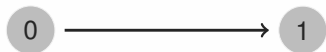
Competing risk process



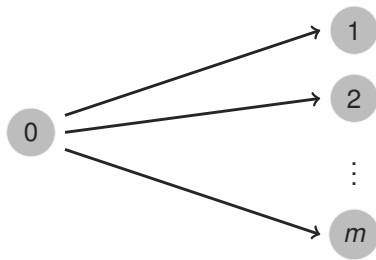
Multistate process



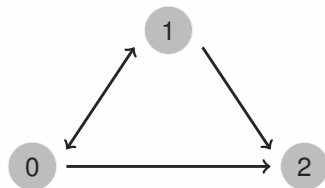
Failure time process



Competing risk process



Multistate process



## Modeling framework

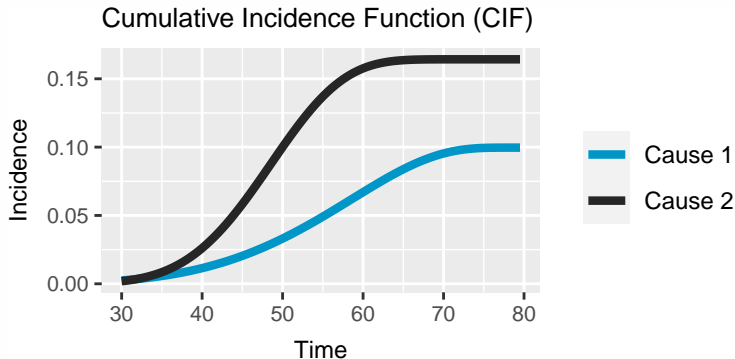
We have to choose which **scale** we model the **survival experience**.  
Usually, the

**hazard** (failure rate) **scale** :  $\lambda(t \mid \text{features}) = \lambda_0(t) \times c(\text{features})$ .

## In the competing risk setting ...



a more attractive possibility is to work on the **probability scale**, focusing on the cause-specific



i.e.

$$\text{CIF} = \mathbb{P}[\text{failure time} \leq t, \text{ a given cause} \mid \text{features}]$$

① Data

② Application

③ Model

④ Contributions & challenges

⑤ Template Model Builder

# Cancer incidence in twins



Clustered competing risks data

↳ Clusters? Families

↳ Family studies

↳ **Twins data**

Family studies  $\Rightarrow$  **within-family dependence**

That may reflect

- Disease **heritability**;
- The impact of shared **environmental effects**;
  - **Parental effects**: continuity of the phenotype across generations.



- 1 Data
- 2 Application
- 3 Model**
- 4 Contributions & challenges
- 5 Template Model Builder

# Our contribution: a hierarchical approach



Thinking on two competing causes

... for the outcome  $y_{ijt}$  of a subject  $i$ , family  $j$ , in the time  $t$ , we have

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

$$\begin{bmatrix} u_{1j} \\ u_{2j} \\ \eta_{1j} \\ \eta_{2j} \end{bmatrix} \sim \text{Multivariate Normal} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1, u_2} & \sigma_{u_1, \eta_1} & \sigma_{u_1, \eta_2} \\ & \sigma_{u_2}^2 & \sigma_{u_2, \eta_1} & \sigma_{u_2, \eta_2} \\ & & \sigma_{\eta_1}^2 & \sigma_{\eta_1, \eta_2} \\ & & & \sigma_{\eta_2}^2 \end{bmatrix} \right)$$

$$\begin{aligned} p_{kijt} &= \frac{\partial \text{CIF}}{\partial t} \\ &= \frac{\partial}{\partial t} \underbrace{\pi_k(X, u_1, u_2 \mid \beta)}_{\text{cluster-specific risk level}} \underbrace{\Phi[w_k g(t) - X^\top \gamma_k - \eta_k]}_{\text{cluster-specific failure time trajectory}}, \end{aligned}$$

$k = 1, 2.$

- 1 Data
- 2 Application
- 3 Model
- 4 Contributions & challenges**
- 5 Template Model Builder

- A clear and simpler modeling structure;
- **There is no free lunch**  
Computational challenges overcome via an efficient implementation and estimation routines, the **TMB**;
- The data is very simple,  
we just know the outcome (**yes** or **no**);
- We have to be able to build the **CIF** curves;
- And accommodate the **within-family dependence** properly,  
that can happen in different manners;
- ...

- ① Data
- ② Application
- ③ Model
- ④ Contributions & challenges
- ⑤ 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.

# Thanks for watching and have a great day



Special thanks to



**PPGMNE**

Programa de Pós-Graduação em  
Métodos Numéricos em Engenharia



Joint work with

Wagner H. Bonat

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

Paulo Justiniano Ribeiro Jr.

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

@hap\_laureano