

# 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 8, 2021

- 1 Data
- 2 Model
- 3 Contributions & challenges
- 4 Template Model Builder
- 5 Simulation study
- 6 Results
- 7 Final considerations
- 8 References

# 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

# Big picture: Failure time data

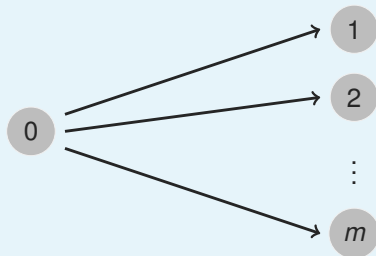


Failure time process



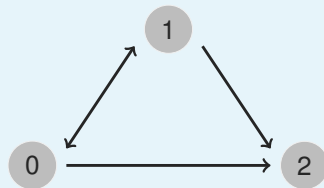
*Same methodologies,  
different names.*

Competing risk process



**Survival analysis** Biomedical studies;  
**Reliability analysis** Industrial life testing.

Multistate process



A comprehensive reference is Kalbfleisch and Prentice (2002)'s book.

- 1 Data
- 2 Model
- 3 Contributions & challenges
- 4 Template Model Builder
- 5 Simulation study
- 6 Results
- 7 Final considerations
- 8 References

# Modeling clustered competing risks data



What?



Why?



How?

# Failure time data → Survival models



First of all, we have to choose which **scale** we model the **survival experience**.

- Usually, is in the

$$\text{hazard (failure rate) scale : } \lambda(t | \text{features}) = \lambda_0(t) \times c(\text{features}). \quad (1)$$

We have a Equation 1 for each competing cause.

The cluster dependence is something actually not measured...

Not measured dependence → **random/latent effects** → Frailty models.

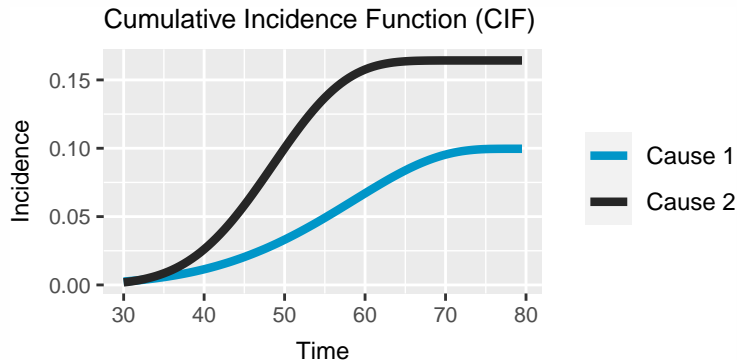
Full likelihood analysis with frailty models for competing risks data is generally complicated, when not impracticable.

- *Not* usually, the **probability scale**.

## Probability scale → Cause-specific CIF



Besides the within-cluster dependence, there is an often interest in describing the time at event onset, directly described by the cause-specific



i.e.,  $\text{CIF} = \mathbb{P}[\text{failure time} \leq t, \text{ a given cause} \mid \text{features \& latent effects}]$ .



for a cause-specific of failure  $k$ , the cumulative incidence function (CIF) is defined as

$$\begin{aligned} F_k(t \mid \mathbf{x}) &= \mathbb{P}[T \leq t, K = k \mid \mathbf{x}] = \int_0^t f_k(z \mid \mathbf{x}) \, dz \\ &= \int_0^t \lambda_k(z \mid \mathbf{x}) S(z \mid \mathbf{x}) \, dz, \quad t > 0, \quad k = 1, \dots, K, \end{aligned}$$

# 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.

# 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 Model
- 3 Contributions & challenges**
- 4 Template Model Builder
- 5 Simulation study
- 6 Results
- 7 Final considerations
- 8 References

- 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 Model
- 3 Contributions & challenges
- 4 Template Model Builder**
- 5 Simulation study
- 6 Results
- 7 Final considerations
- 8 References

# 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.

- 1 Data
- 2 Model
- 3 Contributions & challenges
- 4 Template Model Builder
- 5 Simulation study**
- 6 Results
- 7 Final considerations
- 8 References



- 1 Data
- 2 Model
- 3 Contributions & challenges
- 4 Template Model Builder
- 5 Simulation study
- 6 Results**
- 7 Final considerations
- 8 References

- 1 Data
- 2 Model
- 3 Contributions & challenges
- 4 Template Model Builder
- 5 Simulation study
- 6 Results
- 7 Final considerations**
- 8 References

# 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>



[henriquelaureano.github.io](https://henriquelaureano.github.io)

- 1 Data
- 2 Model
- 3 Contributions & challenges
- 4 Template Model Builder
- 5 Simulation study
- 6 Results
- 7 Final considerations
- 8 References

Kalbfleisch, J. D., and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. Second Edition. Hoboken, New Jersey: John Wiley & Sons, Inc.