

Modeling the cumulative incidence function  
of clustered competing risk data:  
computational and numerical aspects  
of a multinomial GLMM approach



Henrique Ap. Laureano

Wagner H. Bonat

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

December/2020

Context: clustered competing risk data

Idea: causes competing by the occurrence of an event such the

confiability analysis

failure of an industrial or  
electronic component

survival analysis

failure or progress of a patient  
or some biological process



## Context: clustered competing risk data

Idea: causes competing by the occurrence of an event such the

### confiability analysis

failure of an industrial or electronic component

### survival analysis

failure or progress of a patient or some biological process

A typical data set consists of

Group	ID	Cause 1	Cause 2	Censorship	Time	Feature
1	1	1	0	0	10	A
1	2	0	0	1	8	A
2	1	0	0	1	7	B
2	2	0	1	0	5	A

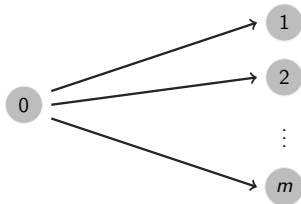


# Survival data designs

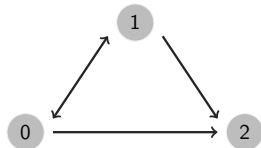
Failure time process



Competing risk process



Multistate process

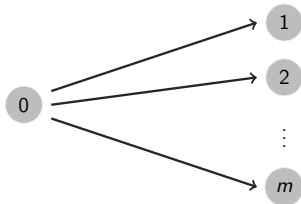


## Survival data designs

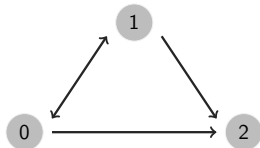
Failure time process



Competing risk process



Multistate process



## Survival modeling framework

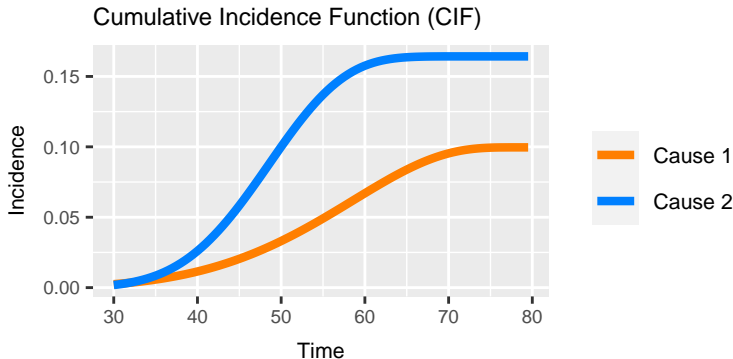
We have to choose which **scale** we model the **survival experience**.  
Usually, is the

**hazard** (failure rate) **scale** :  $\lambda(t | x) = \lambda_0(t) \times c(x, \beta)$



In the competing risk setting ...

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



i.e.

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



Main focus application: cancer incidence in twins



Clustered competing risks data

↳ Clusters? Families

↳ Family studies

↳ **Twins data**



## Main focus application: cancer incidence in twins



Clustered competing risks data

↳ Clusters? Families

↳ Family studies

↳ **Twins data**

- » Taking into account the within-family dependence may reflect both disease **heritability** and the impact of shared **environmental effects**
- » A complication is that we have little information to track that **dependence** since each 'family' consists of only a pair of twins





## Challenges

Besides the small size groups, the data is very simple ...

- » we just know if the event occurred (1 or 0) and the time
  - » with this, we have to be able to construct the **cumulative incidence curves**
- » and we have to accommodate the **within-family dependency**
  - » that can happen in different ways and with different intensities

---

to accomplish all this a powerful modeling framework is made necessary  
... with this,  
computational and numerical challenges has also to be overcome



Thank you



---

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

@hap\_laureano

