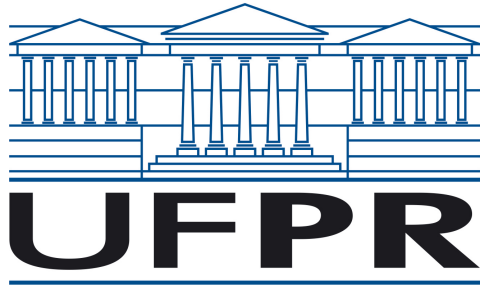
Modeling the cumulative incidence function of clustered competing risk data



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

December/2020

Clustered competing risk data

Idea: causes competing by the occurence 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

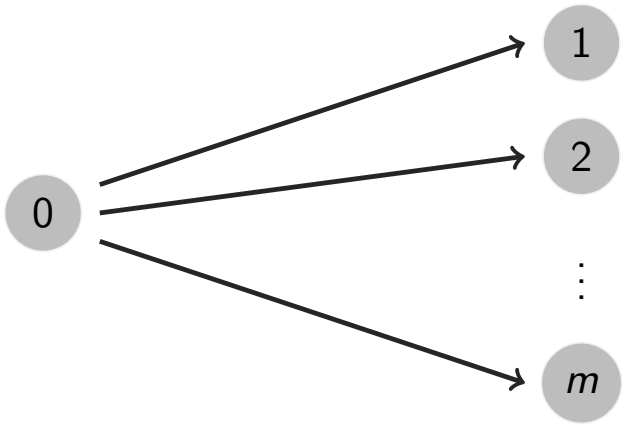
A typical data set consists of

Group	ID	Cause 1	Cause 2	Censorship	Time	Feature
1	1	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

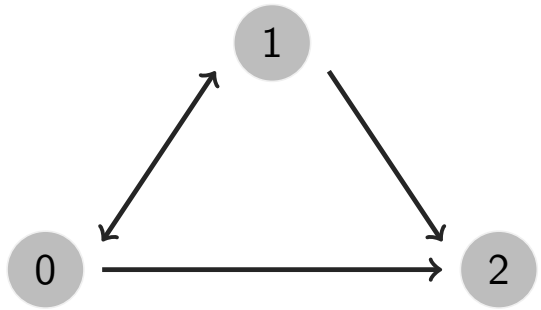
Survival data designs

Failure time process

Competing risk process



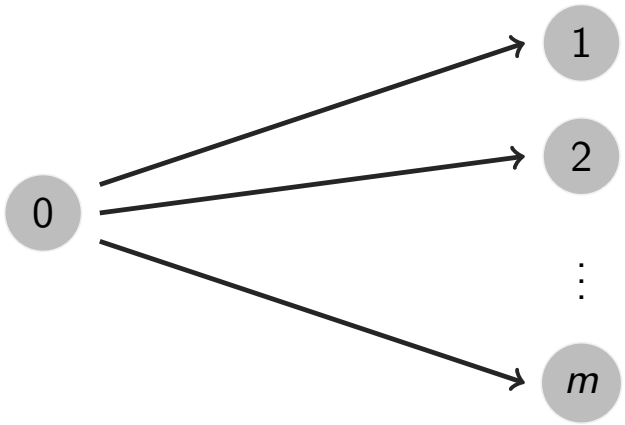
Multistate process



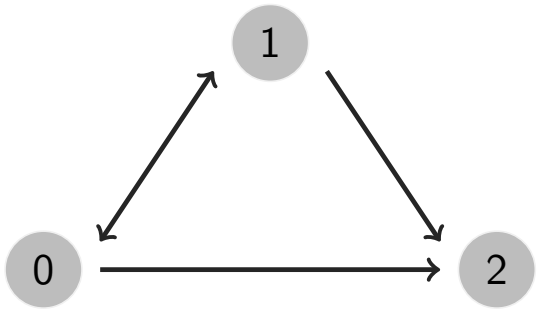
Survival data designs

Failure time process

Competing risk process



Multistate process



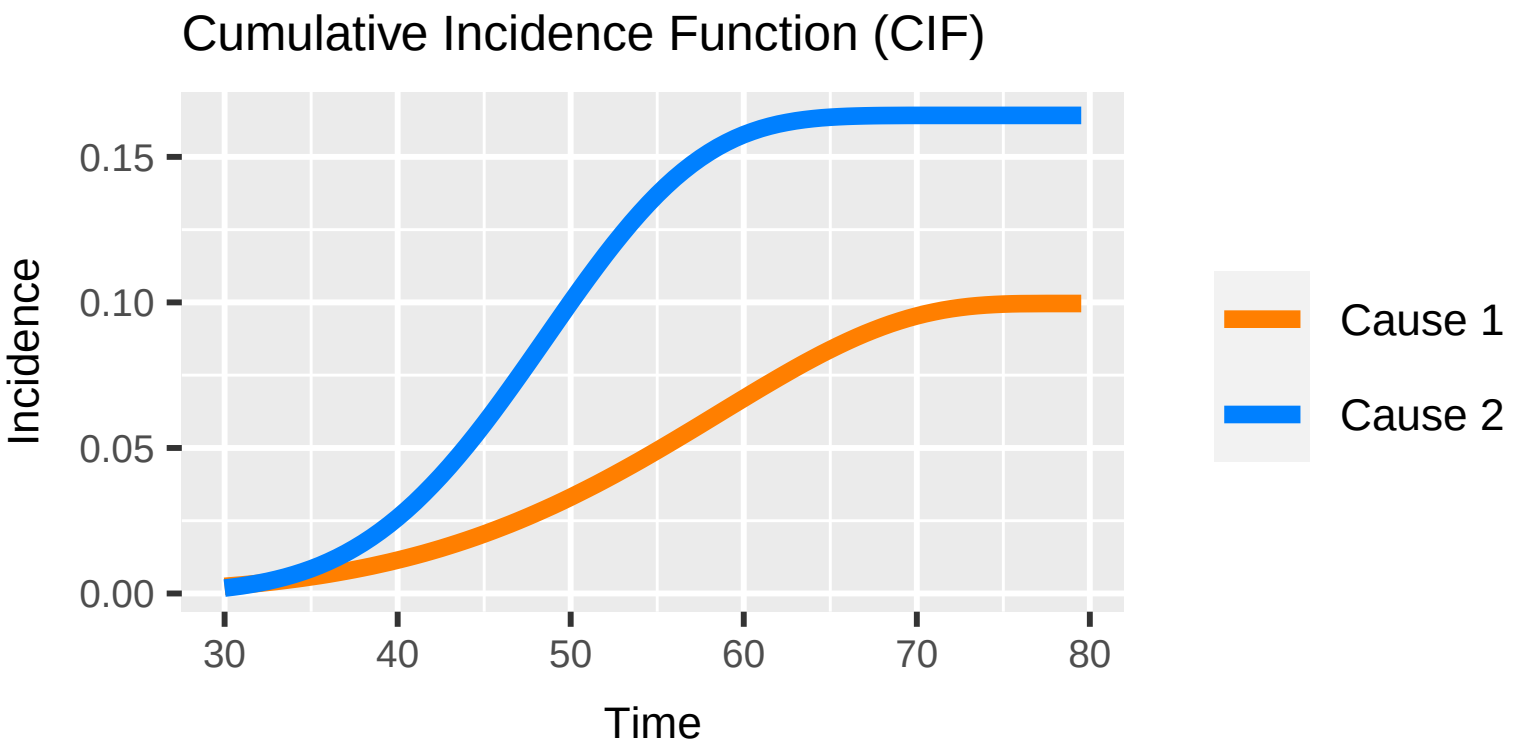
Survival modeling framework

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

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

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

Family studies $\Rightarrow$ within-family dependence

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

And what we do? 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

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

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

$$p_{kijt} = \frac{\partial \text{CIF}}{\partial t}$$

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

$k = 1, 2$

Challenges

Thinking in the twins data application,
the small group/family size is a problem that implies in the following

$$\text{cycle : } \begin{array}{c}\text{small}\\ \text{groups}\end{array} \Rightarrow \begin{array}{c}\text{little}\\ \text{information}\end{array} \Rightarrow \begin{array}{c}\text{complex}\\ \text{model}\end{array} \Rightarrow \begin{array}{c}\text{bigger number}\\ \text{of groups}\end{array}$$

. . . with this,
computational challenges appear and have also to be overcome

Challenges

Thinking in the twins data application,
the small group/family size is a problem that implies in the following

$$\text{cycle : } \begin{array}{c}\text{small}\\ \text{groups}\end{array} \Rightarrow \begin{array}{c}\text{little}\\ \text{information}\end{array} \Rightarrow \begin{array}{c}\text{complex}\\ \text{model}\end{array} \Rightarrow \begin{array}{c}\text{bigger number}\\ \text{of groups}\end{array}$$

. . . with this,
computational challenges appear and have also to be overcome

Besides, the data is very simple . . .

- we just know if the event occured (yes or no) and the time
  - with this, we have to be able to construct the cumulative incidence curves
- and we have to accommodate the within-family dependence
  - that can happen in different manners

Joint work with

Wagner H. Bonat
http://leg.ufpr.br/~wagner

Paulo Justiniano Ribeiro Jr.
http://leg.ufpr.br/~paulojus

Thank you

PPGMNE
PROGRAMA DE PÓS-GRADUAÇÃO EM MÉTODOS NUMÉRICOS EM ENGENHARIA

CAPES

UFPR

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

@hap_laureano

leg