Modeling cumulative incidence function of a competing risk with partially observed cause of failure

Giorgos Bakoyannis and Giota Touloumi

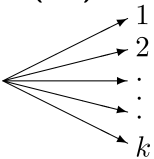
Department of Hygiene, Epidemiology and Medical Statistics
Athens University, Medical School
Greece

Outline

- **Introduction**
  - Competing risks data
  - Example of competing risks data
  - Modeling competing risks data
  - Partially observed cause of failure
  - Objective
- **Methods**
  - The proportional subdistribution hazards model
  - Multiple Imputation of the cause of failure
  - Combination of the results
- **Simulation Study**
  - Description of the simulations
  - Simulation results
- **Conclusions**

Competing risks data

- **Failure time data**
- **k (>1) mutually exclusive causes of failure**



- **Identifiable quantities**
  - Cause-specific hazard (CSH) function for a cause j:

$$\lambda_j(t) = \lim_{h \to 0} \frac{Pr\left(t \leq T < t + h, C = j | T \geq t\right)}{h}$$

  - Functions of CSH like the cumulative incidence (CI) of a cause j:

$$F_j(t) = Pr(T \leq t, C = j) = \int_0^t \lambda_j(u) \exp\left[-\int_0^u \sum_{c=1}^{k} \lambda_c(w)dw\right] du$$

Modeling of competing risks data

CSH of a cause j

Proportional hazards model

$$\lambda_j(t;\mathbf{z}) = \lambda_{0j}(t)\exp(\mathbf{z}'\boldsymbol{\beta}_j)$$

CI of a cause j[1]

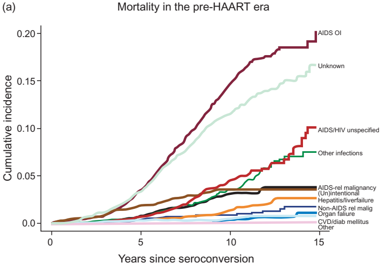
Proportional subdistribution hazards model

$$\lambda_j^{sub}(t;\mathbf{z}) = \lambda_{0j}^{sub}(t)\exp(\mathbf{z}'\boldsymbol{\beta}_j^{sub})$$

[1] Fine JP and Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 1999; **94**:496–509

Example of competing risks data

Cumulative incidence of death from specific causes in HIV-1 (+) patients[2]



Missing cause of death: 25.9%

[2] CASCADE Collaboration. Effective therapy has altered the spectrum of cause-specific mortality following HIV seroconversion. *AIDS*, 2006; **20**:741–749

Partially observed cause of failure

- **Setting**
  - Known failure status
  - Unknown failure cause for some subjects
  - Missing at random (MAR) cause of failure

- **Methodology for CSH modeling**[3]
  - Multiple imputations of the missing causes of failure
    - Logistic regression imputation model
    - Type B (improper) multiple imputation procedure
  - Proportional CSH model fitted in each imputed dataset
  - Combination of the results, accounting for all sources of variability (using the analytic variance expression)

[3] Lu K and Tsiatis A. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, 2001; **57**:1191–1197

Objective

- **Modeling the CI of a competing risk under partially observed cause of failure**
  - Propose a multiple imputation method
  - Testing proposed method's validity through simulations
  - Evaluate the extent of bias in the effect estimate from a complete case analysis under **MAR**

The proportional subdistribution hazards model

Subdistribution hazard function for cause 1 (2 possible causes)

$$\lambda_1^{sub}(t) = \lim_{h \to 0} \frac{Pr\left[t \le T < t+h, C=1 | T \ge t \cup (T \le t \cap C=2)\right]}{h}$$

Proportional subdistribution hazards model

$$\lambda_1^{sub}(t;\mathbf{z}) = \lambda_{01}^{sub}(t)\exp(\mathbf{z}'\boldsymbol{\beta}_1^{sub})$$

Model for the CI of cause 1 as a function of $\lambda_1^{sub}(t;\mathbf{z})$

$$F_1(t;\mathbf{z}) = 1 - \exp\left[-\int_0^t \lambda_1^{sub}(u;\mathbf{z})du\right]$$

$1-1$ relationship between $\lambda_1^{sub}(t;\mathbf{z})$ and $F_1(t;\mathbf{z})$

- Internal time-dependent covariates **cannot** be incorporated in the model

The multiple imputation procedure – 1

- $\mathbf{z}$: Covariates of the main model
- $\mathbf{x}$: Covariates not in the main model, but related to the probability of missingness
- Fitting on the complete cases (known failure) the model:

$$\pi(\mathbf{w}) = logit\left[Pr(C = 1|\mathbf{w})\right] = \mathbf{w}'\boldsymbol{\theta}$$

where $\mathbf{w} = (1, t, \mathbf{z}, \mathbf{x})'$
Bootstrap SE estimates can be used under a possibly misspecified model
The model can include interaction and/or polynomial terms to improve the fit of it
Under MAR

$$Pr(C = 1|\mathbf{w}, r = 1) = Pr(C = 1|\mathbf{w}, r = 0)$$

where r is the missingness indicator (0 indicating missing cause)

The multiple imputation procedure – 2

- Generation of m completed datasets
  - Imputate m times the missing causes with the cause of interest with probability:

    $$Pr(C=1|\mathbf{w}) = \frac{\exp(\mathbf{w}'\boldsymbol{\theta}_q^*)}{1+\exp(\mathbf{w}'\boldsymbol{\theta}_q^*)}\ ,\ q=1,2,...,m$$

    where $\boldsymbol{\theta}_q^* \sim N(\widehat{\theta}, \widehat{Var}(\widehat{\theta}))$
  - This imputation procedure is proper (type A) since the parameters $\boldsymbol{\theta}_q^*$ are not fixed across imputations.
- Fit the proportional subdistribution hazards model on each imputed dataset

The multiple imputation procedure – 3

- Combine results from the m pseudo-complete datasets as in any proper multiple imputation procedure

Estimator of the regression coefficient ($\widehat{\boldsymbol{\beta}}_1^{sub}$)

$$\widehat{\boldsymbol{\beta}}_1^{sub} = \frac{1}{m}\sum_{q=1}^{m}\widehat{\boldsymbol{\beta}}_{1q}^{sub}$$

Estimator of the variance of $\widehat{\boldsymbol{\beta}}_1^{sub}$

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}_1^{sub}) = \underbrace{\frac{1}{m}\sum_{q=1}^{m}\widehat{Var}(\widehat{\boldsymbol{\beta}}_{1q}^{sub})}_{W} + (1+m^{-1})\underbrace{\sum_{q=1}^{m}\frac{(\widehat{\boldsymbol{\beta}}_{1q}^{sub}-\widehat{\boldsymbol{\beta}}_{1}^{sub})(\widehat{\boldsymbol{\beta}}_{1q}^{sub}-\widehat{\boldsymbol{\beta}}_{1}^{sub})'}{m-1}}_{B}$$

The multiple imputation procedure – 4

- Statistical inference
  - Scalar β_1^{sub}

$$\frac{(\widehat{\beta}_1^{sub} - \beta_1^{sub})^2}{\widehat{Var}(\widehat{\beta}_1^{sub})} \sim F_{1,\nu}$$

where $\nu = (m-1)\left[1 + \frac{W}{(1+m^{-1})B}\right]^2$

  - Multi-component $\boldsymbol{\beta}_1^{sub}$ (of lengh g)

$$(1+r)^{-1}(\widehat{\beta}_1^{sub} - \beta_1^{sub})[\widehat{Var}(\widehat{\beta}_1^{sub})]^{-1}(\widehat{\beta}_1^{sub} - \beta_1^{sub})'k^{-1} \sim F_{g,(g+1)\nu/2}$$

where $r = (1+m^{-1})Tr(BW^{-1})g^{-1}$
and $\nu = (m-1)(1+r^{-1})^2$

Simulation setting

- 2 **causes of failure** considered:
  - Type 1 (cause of interest)
  - Type 2 (competing cause)
- 2 **covariates** considered, z and x
- The **cumulative incidence of the cause of interest** was considered to be:

$$F_1(t; z, x) = 1 - \{1 - 0.5[1 - \exp(-t)]\}^{\exp(0.5z + b_x x)}$$

  (where $b_x \in \{0, 0.5, 1\}$) resulting in the **proportional subdistribution hazards** model:

$$\lambda_1^{sub}(t; z, x) = \lambda_{01}^{sub}(t) \exp(0.5z + b_x x)$$

- **MCAR** or **MAR** cause of failure depending on the scenario

Simulation study details – 1

- **Covariates** were simulated from:
  - $z \sim bernoulli(0.4)$
  - $x \sim bernoulli(0.6)$
- **Cause of failure** was simulated from:

$$Pr(C = 1|z, x) = 1 - (1 - 0.5)^{\exp(0.5z + b_x x)}$$

  Where $b_x \in \{0, 0.5, 1\}$ depending on the scenario
- Conditional on cause, **failure time** was simulated from:

$$Pr(T \leq t|C = 1, z, x) = \frac{1 - \{1 - 0.5[1 - \exp(-t)]\}^{\exp(0.5z + b_x x)}}{1 - (1 - 0.5)^{\exp(0.5z + b_x x)}}$$

$$Pr(T \leq t|C = 2, z, x) = 1 - \exp[-\exp(-0.5z + 0.5x)t]$$

- **Censoring time** was simulated from:

$$Pr(U \leq u|z, x) = 1 - \exp(-0.25u)$$

Simulation study details – 2

- **Missingness indicator** $I(r=0)$ was simulated from:

$$Pr(r=0|\mathbf{w}) = \frac{\exp(\mathbf{w}'\mathbf{b})}{1+\exp(\mathbf{w}'\mathbf{b})}$$

where $\mathbf{w} = (1, t, z, x)'$ and
$\mathbf{b} = \overbrace{(-1, 0, 0, 0)'}^{MCAR}$ or $\overbrace{(b_0, 1, -2.5, 2)'}^{MAR}$ depending on the scenario and $b_0 \in \{-2, -1\}$

- The **imputation model** used was:

$$logit\left[Pr(C=1|\mathbf{w})\right] = \mathbf{w}'\boldsymbol{\theta}$$

Bootstrap SEs used for $\widehat{\boldsymbol{\theta}}$

Simulation study details – 3

- Under the present simulation, the probability for the cause of interest **conditional on** t, z and x was:

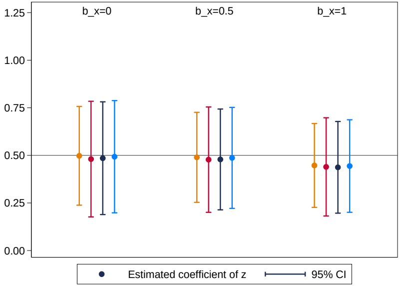
$$Pr(C=1|t,z,x) = \frac{1}{1+\frac{\exp(-0.5z+0.5x)\exp[-\exp(-0.5z+0.5x)t]}{[1+\exp(-t)]^{[\exp(0.5z+b_x x)-1]}\exp(0.5z+b_x x)\exp(-t)}}$$

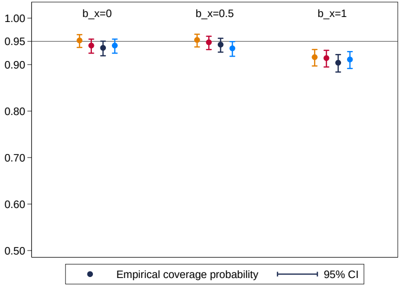
- 1000 simulated datasets consisting of 500 subjects per scenario
- 3 methods applied in each scenario
  - Analysis using full datasets (without missing cause of failure)
  - Analysis based on complete cases only (cases with missing cause of failure deleted)
  - Proposed multiple imputation method (with bootstrap SEs for the parameters of the imputation model)
- In all methods covariate x was not included in the main (for the CI) model

Simulation results – MCAR

Methods

- Analysis using full datasets (orange lines)
- Complete case analysis (red lines)
- Multiple imputation method (dark blue lines)
- Multiple imputation method, with interaction term of z (and x when $b_x = 0$) with time (light blue lines)

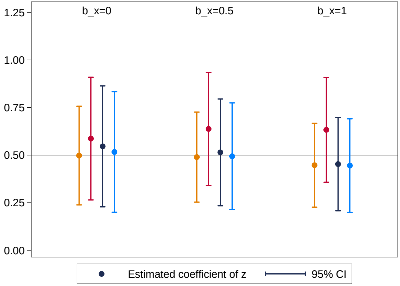


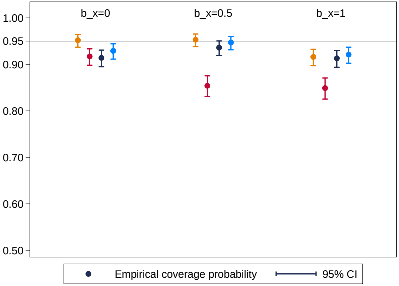


Simulation results – MAR (33.4% missing cause of failure)

Methods

- Analysis using full datasets (orange lines)
- Complete case analysis (red lines)
- Multiple imputation method (dark blue lines)
- Multiple imputation method, with interaction term of z (and x when $b_x = 0$) with time (light blue lines)

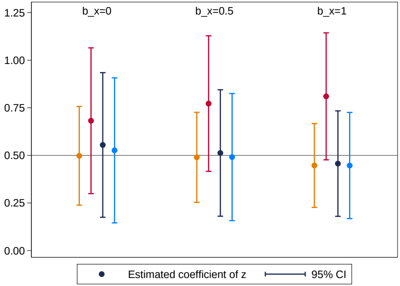


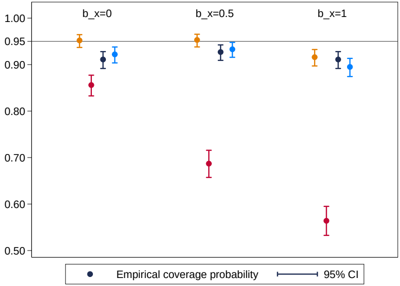


Simulation results – MAR (49.2% missing cause of failure)

Methods

- Analysis using full datasets (orange lines)
- Complete case analysis (red lines)
- Multiple imputation method (dark blue lines)
- Multiple imputation method, with interaction term of z (and x when $b_x = 0$) with time (light blue lines)





Conclusions – 1

- Under **MCAR** cause of failure both methods give similar results
- Under **MAR** cause of failure
  - **Compete case (CC) analysis** gives estimates with serious bias and poor coverage probabilities
  - **Proposed multiple imputation method (MI)** gives better estimates w.r.t bias and coverage probability than the CC analysis **even under a strongly misspecified imputation model**

Conclusions – 2

- **Omitting a significant and independent of the others covariate from the proportional subdistribution hazard model** leads to **attenuated** effect estimate of the included covariates
  - More pronounced as stronger the effect of the omitted covatiate on the cause of interest
  - Known phenomenon in classical survival analysis[4]

[4] Schmoor C and Schumacher M. Effects of covariate omission and categorization when analysing randomized trials with the Cox model. *Statistics in Medicine*, 1997; **16**:225–237