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**

$$
\begin{align*}
1 \\
2 \\
\vdots \\
k
\end{align*}
$$

- **Identifiable quantities**
  - Cause-specific hazard (CSH) function for a cause $j$:
    $$
    \lambda_j(t) = \lim_{h \to 0} \frac{Pr(t \leq T < t + h, C = j | T \geq t)}{h}
    $$
  - 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
\[
\lambda_j(t; z) = \lambda_{0j}(t) \exp(z' \beta_j)
\]

CI of a cause j

Proportional subdistribution hazards model
\[
\lambda_{j}^{\text{sub}}(t; z) = \lambda_{0j}^{\text{sub}}(t) \exp(z' \beta_{j}^{\text{sub}})
\]

---

细水JP和Gray RJ. 一个比例风险模型用于竞争风险的子分布。《美国统计学会杂志》，1999年；94:496–509
Example of competing risks data

**Cumulative incidence** of death from specific causes in HIV-1 (+) patients

---

![Graph (a): Mortality in the pre-HAART era](image)

- AIDS OI
- Unknown
- AIDS/HIV unspecified
- Other infections
- AIDS-rel malignancy
- (Un)intentional
- Hepatitis/liver failure
- Non-AIDS rel malign
- Organ failure
- CVD/diab mellitus
- Other

![Graph (b): Mortality in the HAART era](image)

- AIDS OI
- Unknown
- AIDS/HIV unspecified
- Other infections
- (Un)intentional
- Hepatitis/liver failure
- Non-AIDS rel malign
- Organ failure
- Other

---

**Missing cause of death: 25.9%**

---


---

**Giorgos Bakoyannis and Giota Touloumi**

**Modeling CI with partially observed cause of failure**
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)**

\[
\lambda_{1\text{sub}}(t) = \lim_{h \to 0} \frac{Pr [t \leq T < t + h, C = 1|T \geq t \cup (T \leq t \cap C = 2)]}{h}
\]

**Proportional subdistribution hazards model**

\[
\lambda_{1\text{sub}}(t; z) = \lambda_{01\text{sub}}(t) \exp(z' \beta_{1\text{sub}})
\]

**Model for the CI of cause 1 as a function of \( \lambda_{1\text{sub}}(t; z) \)**

\[
F_1(t; z) = 1 - \exp \left[ - \int_0^t \lambda_{1\text{sub}}(u; z) \, du \right]
\]

1 – 1 relationship between \( \lambda_{1\text{sub}}(t; z) \) and \( F_1(t; z) \)

- Internal time-dependent covariates **cannot** be incorporated in the model
The multiple imputation procedure – 1

- \( z \): Covariates of the main model
- \( x \): Covariates not in the main model, but related to the probability of missingness

Fitting on the complete cases (known failure) the model:

\[
\pi(w) = \logit[Pr(C = 1|w)] = w'\theta
\]

where \( w = (1, t, z, x)' \)

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

\[
Pr(C = 1|w, r = 1) = Pr(C = 1|w, r = 0)
\]

where \( r \) is the missingness indicator (0 indicating missing cause)
The multiple imputation procedure – 2

- Generation of \( m \) completed datasets
  - Impute \( m \) times the missing causes with the cause of interest with probability:

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

where \( \theta^*_q \sim N(\hat{\theta}, Var(\hat{\theta})) \)

- This imputation procedure is proper (type A) since the parameters \( \theta^*_q \) 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 ($\hat{\beta}_{1}^{sub}$)

$$\hat{\beta}_{1}^{sub} = \frac{1}{m} \sum_{q=1}^{m} \hat{\beta}_{1q}^{sub}$$

Estimator of the variance of $\hat{\beta}_{1}^{sub}$

$$\hat{Var}(\hat{\beta}_{1}^{sub}) = \frac{1}{m} \sum_{q=1}^{m} \hat{Var}(\hat{\beta}_{1q}^{sub}) + (1 + m^{-1}) \sum_{q=1}^{m} \frac{(\hat{\beta}_{1q}^{sub} - \hat{\beta}_{1}^{sub})(\hat{\beta}_{1q}^{sub} - \hat{\beta}_{1}^{sub})'}{m - 1}$$
The multiple imputation procedure – 4

- **Statistical inference**
  - **Scalar** $\beta_{1}^{sub}$
    \[
    \frac{(\hat{\beta}_{1}^{sub} - \beta_{1}^{sub})^2}{\text{Var}(\hat{\beta}_{1}^{sub})} \sim F_{1,\nu}
    \]
    where $\nu = (m - 1) \left[1 + \frac{W}{(1+m^{-1})B}\right]^2$
  - **Multi-component** $\beta_{1}^{sub}$ (of length $g$)
    \[
    (1+r)^{-1}(\hat{\beta}_{1}^{sub} - \beta_{1}^{sub})[\text{Var}(\hat{\beta}_{1}^{sub})]^{-1}(\hat{\beta}_{1}^{sub} - \beta_{1}^{sub})'k^{-1} \sim F_{g,(g+1)\nu/2}
    \]
    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:
  - \( z \sim \text{bernoulli}(0.4) \)
  - \( x \sim \text{bernoulli}(0.6) \)

- **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:

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

where $w = (1, t, z, x)'$ and

- **MCAR** $b = (-1, 0, 0, 0)'$
- **MAR** $b = (b_0, 1, -2.5, 2)'$ depending on the scenario and $b_0 \in \{-2, -1\}$

- The **imputation model** used was:

$$\text{logit} [Pr(C = 1|w)] = w'\theta$$

Bootstrap SEs used for $\hat{\theta}$
Under the present simulation, the probability for the cause of interest \textit{conditional on} \( t \), \( z \) and \( x \) was:

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

- 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**
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 covariate on the cause of interest
- Known phenomenon in classical survival analysis\(^4\)