mvna, an R-package for the Multivariate Nelson-Aalen Estimator in Multistate Models

A. Allignol[1,2,*] J. Beyersmann[1,2] M. Schumacher[2]

[1]Freiburg Center for Data Analysis and Modelling, Freiburg University
[2]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg
[*]arthur.allignol@fdm.uni-freiburg.de

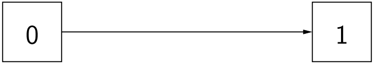
DFG Forschergruppe FOR 534

Multistate Model Framework

- Time-inhomogeneous Markovian multistate model
- Possible right-censoring and left-truncation
  - $(X_t)_{t \in [0, +\infty)}$ a stochastic process with state space $\{0, \ldots, K\}$, and right-continuous sample paths

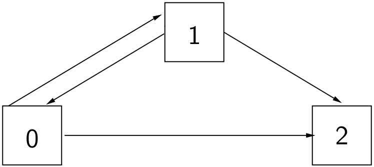
Multistate Model Framework

- Time-inhomogeneous Markovian multistate model
- Possible right-censoring and left-truncation
  - $(X_t)_{t \in [0, +\infty)}$ a stochastic process with state space $\{0, \ldots, K\}$, and right-continuous sample paths



Multistate Model Framework

- Time-inhomogeneous Markovian multistate model
- Possible right-censoring and left-truncation
  - $(X_t)_{t \in [0, +\infty)}$ a stochastic process with state space $\{0, \ldots, K\}$, and right-continuous sample paths



Multistate Model Framework

- Transition hazards:

$$\alpha_{ij}(t)dt = \mathrm{P}(X_{t+dt} = j \,|\, X_t = i)$$

Multistate Model Framework

- Transition hazards:

$$\alpha_{ij}(t)dt = \mathrm{P}(X_{t+dt} = j \mid X_t = i)$$

- Cumulative transition hazards:

$$A_{ij}(t) = \int_0^t \alpha_{ij}(u)du$$

The Nelson-Aalen Estimator

- Nelson-Aalen estimator of the cumulative transition hazards

$$\hat{A}_{ij}(t) = \sum_{t_k \leq t} \frac{\Delta N_{ij}(t_k)}{Y_i(t_k)}$$

  - $N_{ij}(t)$ number of transitions from state i to j by time t
  - $Y_i(t)$ number of individuals in state i just before time t

The Nelson-Aalen Estimator

- Nelson-Aalen estimator of the cumulative transition hazards

$$\hat{A}_{ij}(t) = \sum_{t_k \le t} \frac{\Delta N_{ij}(t_k)}{Y_i(t_k)}$$

  - $N_{ij}(t)$ number of transitions from state i to j by time t
  - $Y_i(t)$ number of individuals in state i just before time t
- The Nelson–Aalen estimator is simply a sum over empirical hazards/empirical conditional transition probabilities

Variance Estimation

- Aalen variance estimator

$$\hat{\sigma}_{ij}^2 = \sum_{t_k \leq t} \frac{dN_{ij}(t_k)}{Y_i(t_k)^2}$$

Variance Estimation

- Aalen variance estimator

$$\hat{\sigma}_{ij}^2 = \sum_{t_k \leq t} \frac{dN_{ij}(t_k)}{Y_i(t_k)^2}$$

- Greenwood variance estimator

$$\check{\sigma}_{ij}^2 = \sum_{t_k \leq t} \left\{ \frac{Y_i(t_k) - \Delta N_{ij}(t_k)}{Y_i(t_k)} \right\} \left\{ \frac{dN_{ij}(t_k)}{Y_i(t_k)^2} \right\}$$

Rationale

- The Nelson-Aalen estimator is the fundamental nonparametric estimator in event history analysis.
- No package available to compute it in multistate models
  - Univariate software tempts people to use objects that are meaningless in multistate framework, *e.g.*, Kaplan-Meier of single hazard estimates
- Cumulative hazard estimates give useful insights *e.g.*,
  - In competing risks analysis
  - With time-dependent covariates

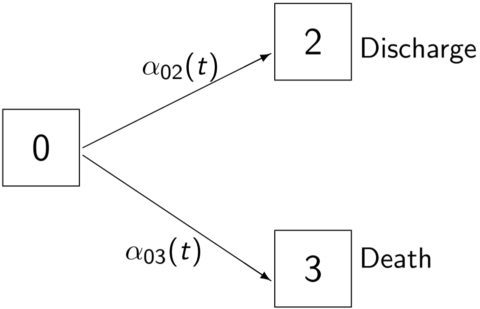
Package Description

- `mvna(data, state.numbers, tra, cens.name)`
- `xyplot.mvna(x, ...)`
- `plot.mvna(x, ...)`
- `print(x, ...)`
- `predict(object, times, ...)`

Package Description

- `mvna(data, state.numbers, tra, cens.name)`
- `xyplot.mvna(x, ...)`
- `plot.mvna(x, ...)`
- `print(x, ...)`
- `predict(object, times, ...)`
- 2 data sets:
  - Random samples from intensive care unit cohort data on hospital infections, with a minimum length of stay of 2 days
  - `sir.adm`:
    - Effect of pneumonia status on admission on the hazard of discharge and death, respectively
  - `sir.continuation`:
    - Effect of ventilation (time-dependent) on the hazard of end-of-stay, a combined discharge/death endpoint

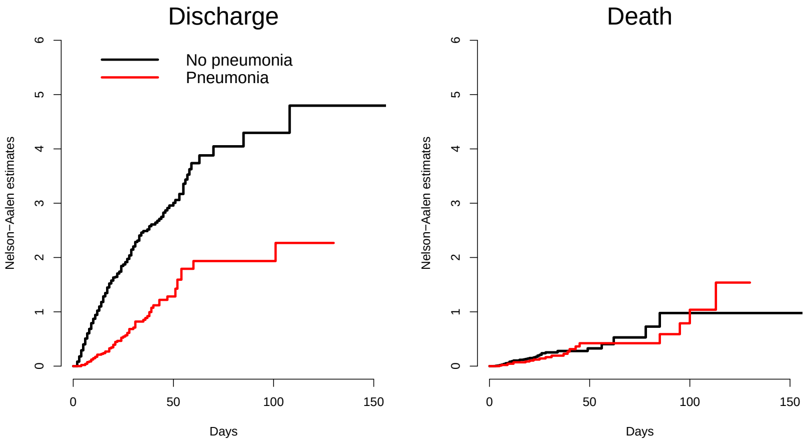
Competing Risks example



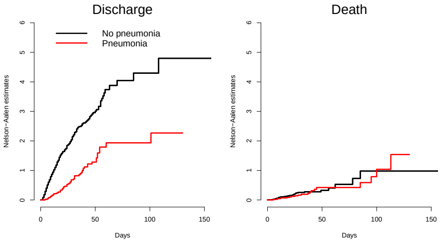
The Data Set

- 765 patients from medical and surgical ICUs
- 14 (2%) censored observations
- 97 (13%) patients with pneumonia on admission
  - 21 (22%) died
- 668 (87%) patients free of pneumonia
  - 56 (8%) died

Nelson-Aalen Estimates

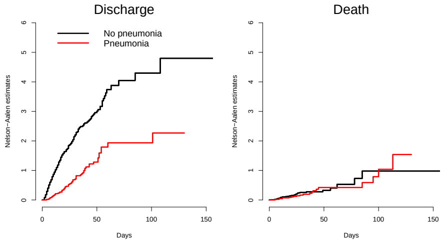


Nelson-Aalen Estimates



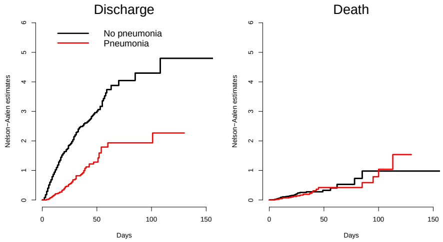
- More patients die after pneumonia on admission

Nelson-Aalen Estimates



- More patients die after pneumonia on admission
  - Pneumonia prolongs hospital stay, as the all-cause hazard is reduced.
  - Patients with pneumonia stay longer in hospital, exposed to an unchanged death hazard

Nelson-Aalen Estimates



- More patients die after pneumonia on admission
  - Pneumonia prolongs hospital stay, as the all-cause hazard is reduced.
  - Patients with pneumonia stay longer in hospital, exposed to an unchanged death hazard

$\Rightarrow$ Pneumonia increases mortality

How This Package Supplements What Already Exists

Competing Risks

Cox Model for the Cause-Specific Hazards

- Pneumonia status as a baseline binary covariate

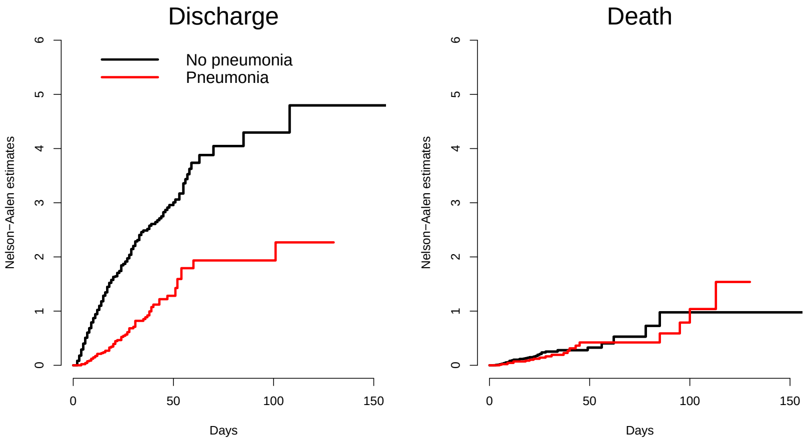
Competing Risks

Cox Model for the Cause-Specific Hazards

- Pneumonia status as a baseline binary covariate
- Results:

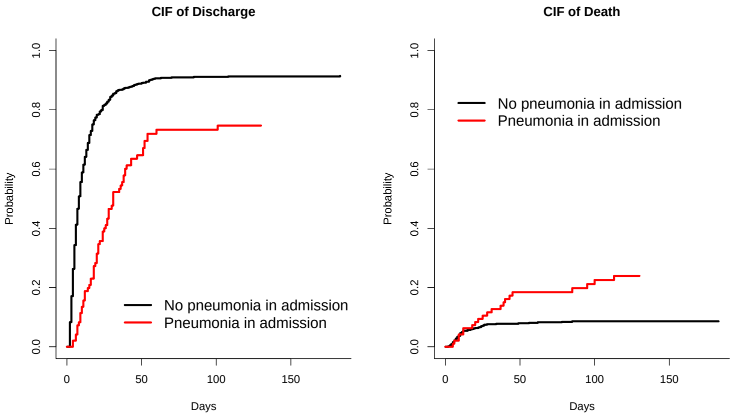
	CSHR	95% CI
Discharge	0.367	[0.258; 0.473]
Death	0.906	[0.537; 1.53]

Nelson-Aalen Estimates



Cumulative Incidence Functions

- Proportion of patients failing due to one risk as time progresses



Note on the Variance Estimators

- Variance estimation is a more concerning problem with multistates

Note on the Variance Estimators

- Variance estimation is a more concerning problem with multistates
- For standard survival data, Klein (1991) found that:
  - The Aalen estimator overestimates the true variance for risk sets ≤ 5
  - The Greenwood estimator underestimates the true variance, but has a smaller MSE
  - The 2 estimators coincide for risk sets ≥ 10
- Preliminary simulations in the multistate framework:
  - Comparable findings
- Recommendations:
  - Use of the Aalen estimator

Summary

- The **mvna** package provides a way to easily estimate and display the cumulative transition hazards from multistate models
- Extremely useful in illustrating and understanding complex event history processes, *e.g.*,
  - with competing risks
  - with a time-dependent covariate
- Outlook:
  - **etm** package for computing the empirical transition matrix (transition probabilities)

Bibliography

Aalen, O. (1978).
Nonparametric Inference for a family of Counting Processes.
The Annals of Statistics, 6:701–726.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993).
Statistical Models Based on Counting Processes.
Springer-Verlag, New-York.

Klein, J. P. (1991).
Small Sample Moments of Some Estimators of the Variance of the Kaplan-Meier and Nelson-Aalen Estimators.
Scandinavian Journal of Statistics, 18:333–340.

Data & Model

- $(X_t)_{t \geq 0}$ the competing risks process
  - $X_t \in \{0, 2, 3\}$
- The failure time T at which patients leave the initial state 0 is
  - $T = \inf\{t \in [0, \infty) | X_t \neq 0\}$
  - X_T denotes the failure cause
- Cause-specific hazard:

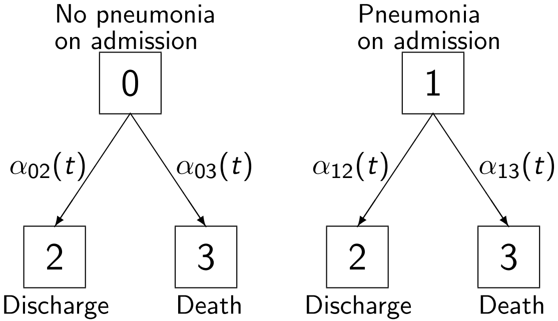
$$\alpha_{0i}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, X_T = i | T \geq t)}{\Delta t}, i = 2, 3$$

How to?

New Definition of the Competing Risks Process

How to?

New Definition of the Competing Risks Process



How to?

```
> ### Matrix of logical indicating the possible transitions
> tra <- matrix(ncol=4,nrow=4,FALSE)
> tra[1:2,3:4] <- TRUE
> tra
      [,1]  [,2]  [,3]  [,4]
[1,] FALSE FALSE  TRUE  TRUE
[2,] FALSE FALSE  TRUE  TRUE
[3,] FALSE FALSE FALSE FALSE
[4,] FALSE FALSE FALSE FALSE
```

How to?

```
> ### Nelson_Aalen estimates
> na.pneu <- mvna(data=dat.sir,
+                 state.names=c("0","1","2","3"),
+                 tra=tra,cens.name="cens")
```

How to?

```
> ### Nelson_Aalen estimates
> na.pneu <- mvna(data=dat.sir,
+                 state.names=c("0","1","2","3"),
+                 tra=tra,cens.name="cens")

> na.pneu
Estimated cumulative hazard for transition 0 to 2

Time
[1]   1  16  30  46 183

Nelson-Aalen estimates
[1] 0.024 1.336 2.166 2.863 5.613

Variance estimates
[1] 0.000 0.005 0.014 0.033 1.411

Alternative variance estimates
[1] 0.000 0.004 0.013 0.031 0.266
```

How to?

```
> ### Plot
> xyplot(na.pneu,
+ tr.choice=c("0 2","1 2","0 3","1 3"),
+ aspect=1,strip=strip.custom(bg="white",
+ factor.levels=
+   c("No pneumonia on admission - Discharge",
+     "Pneumonia on admission - Discharge",
+     "No pneumonia on admission - Death",
+     "Pneumonia on admission - Death"),
+ par.strip.text=list(cex=0.9)),
+ scales=list(alternating=1),xlab="Days",
+ ylab="Nelson-Aalen estimates")
```

Cumulative Incidence Functions

- CIF of discharge

$$F_2(t) = P(T \leq t, X_T = 2) \\ = \int_0^t P(T > u-)\alpha_{02}(u)du$$

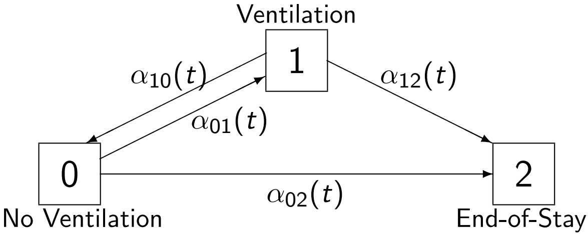
- Depends on both cause-specific hazards

$$P(T > t) = \exp\left(\int_0^t \alpha_{02}(u) + \alpha_{03}(u)du\right)$$

- Loss of the one to one relationship between hazard and probability

Time-dependent Covariate

as a Transient State in a Multistate Model



Model

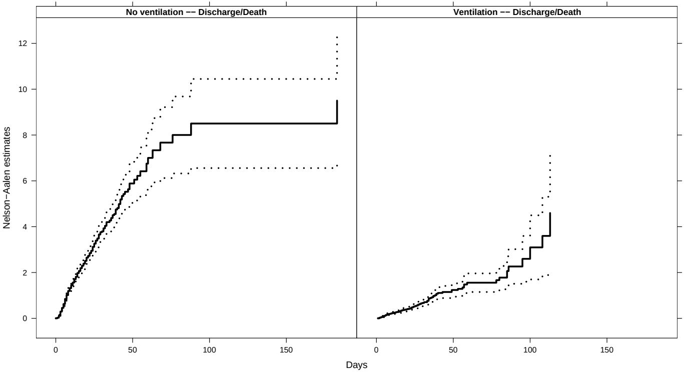
- $(V_t)_{t \geq 0}$ the stochastic process
  - $V_t \in \{0, 1, 2\}$
- Survival time:

$$T = \inf\{t \in [0, \infty) | X_t = 2\}$$

- The transition hazard is

$$\alpha_{ij}(t)dt = P(X_{t+dt} = j | X_t = i)$$

Nelson-Aalen Estimates



Time-dependent Covariate

Cox Model

Time-dependent Covariate

Cox Model

- Cox proportional hazards model:

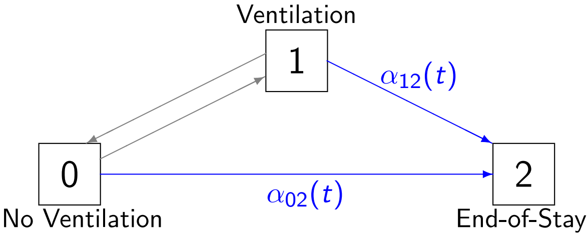
$$\alpha_{12}(t) = \alpha_{02}(t).\exp(\beta)$$

Time-dependent Covariate

Cox Model

- Cox proportional hazards model:

$$\alpha_{12}(t) = \alpha_{02}(t). \exp(\beta)$$



Cox Model

- Cox model:

HR	95% CI
0.179	[0.149; 0.214]

Cox Model

- Cox model:

HR	95% CI
0.179	[0.149; 0.214]