Multi-state models: Rates, Risks, and Pseudo-Values

Overview of course

I Introduction to multi-state models

II Non-parametric estimation and regression models for intensities (Cox)

III Estimation of marginal parameters using plug-in

IV Direct regression models for marginal parameters (Cox, Fine-Gray, Ghosh-Lin)

V Pseudo-values

III: Estimation of marginal parameters using plug-in

- Kaplan-Meier estimator
- Aalen-Johansen estimator for competing risks cumulative incidence
- Markov processes
- ELOS
- Recurrent events
- Regression models

The intensities are the basic parameters in multi-state models and marginal parameters, such as state occupation probabilities $Q_h(t)$ may be estimated by *plug-in* (or by simulation).

Two-state model

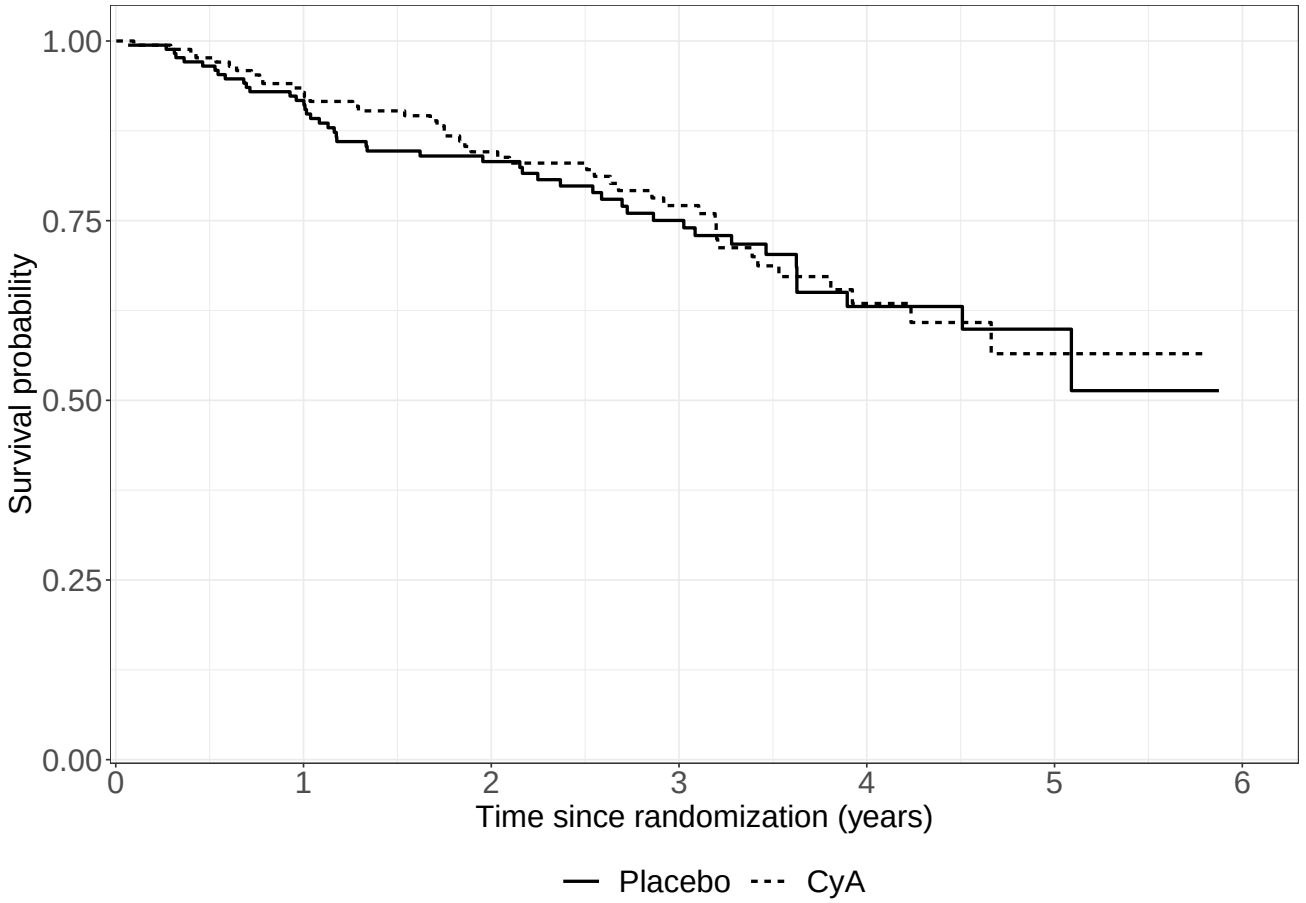
For the two-state model, plug-in leads to the Kaplan-Meier (1958) estimator, Eq. (4.3), for $S(t) = Q_0(t)$:

$$\widehat{S}(t) = \prod_{u \leq t} \left(1 - \frac{\sum_i dN_{01i}(u)}{\sum_i Y_{0i}(u)} \right).$$

Note that we do not use $\exp(-\widehat{A}_{01}(t))$. This is because the 'exponential formula' $S(t) = \exp(-A_{01}(t))$ only holds for *continuous distributions* while, in general, the relationship is given by a 'product-integral'.

The uncertainty of $\widehat{S}(t)$ may be evaluated using the *Greenwood formula.* This may be derived from the uncertainty of the Nelson-Aalen estimator via the *delta-method*.

The PBC-3 trial, composite end-point



Competing risks

The cause-h cumulative incidence is $Q_h(t) = \int_0^t S(u)\alpha_{0h}(u)du$, and the plug-in estimator is the *Aalen-Johansen* estimator, Eq. (4.9):

$$\widehat{Q}_h(t) = \int_0^t \widehat{S}(u-)d\widehat{A}_{0h}(u),$$

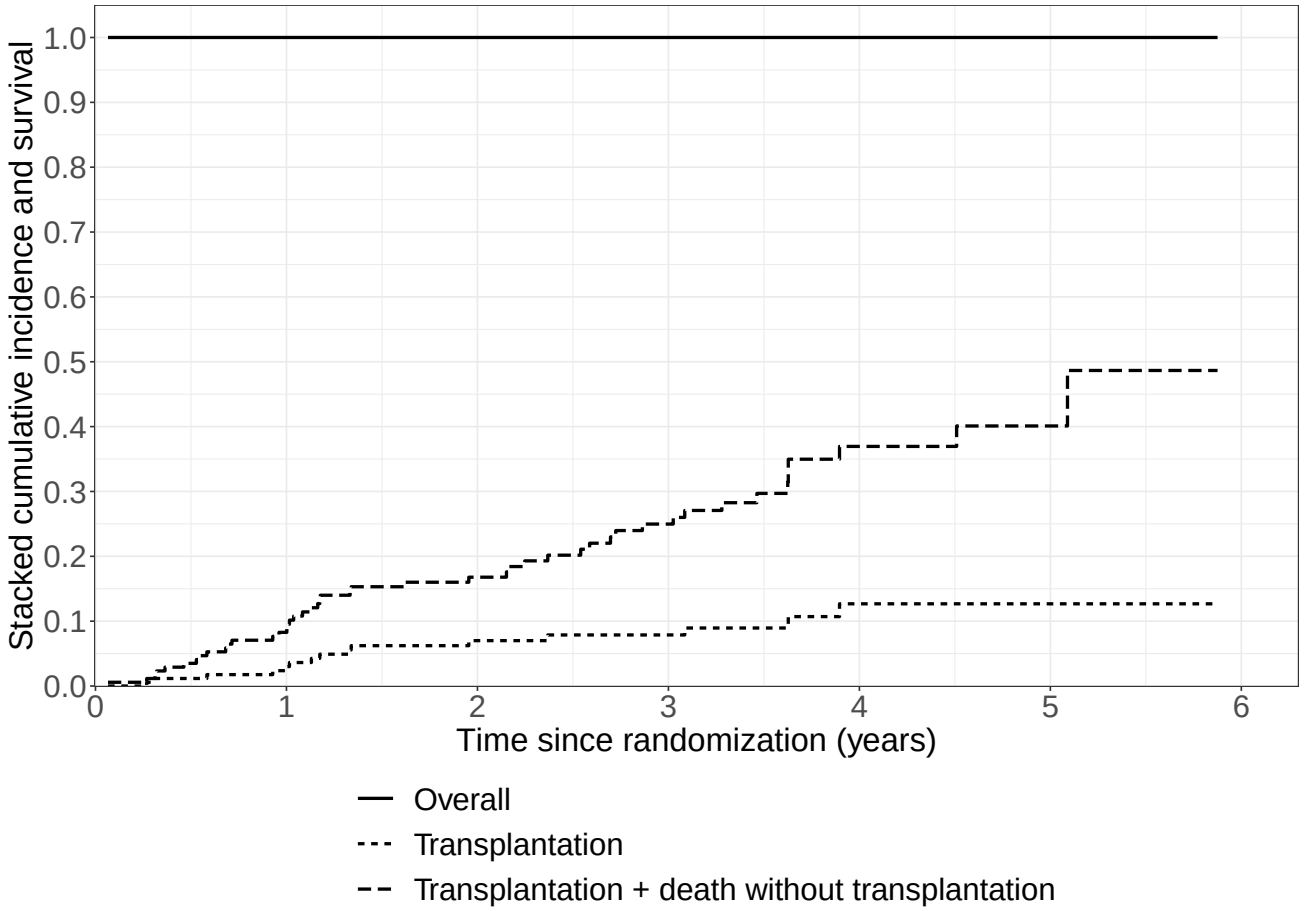
with $\widehat{S}$, the Kaplan-Meier estimator for overall survival (i.e., counting all deaths) and $\widehat{A}_{0h}$ the Nelson-Aalen estimator for the cumulative cause-h-specific hazard. The uncertainty may be evaluated via the delta-method.

Note that $\widehat{Q}_h(t)$ (via $\widehat{S}$) depends on *all cause-specific hazards*.

A naive '1-Kaplan-Meier estimator', say $1 - \widehat{S}_h(t)$, counting only cause-h events (and censoring for the competing risks) is upwards biased

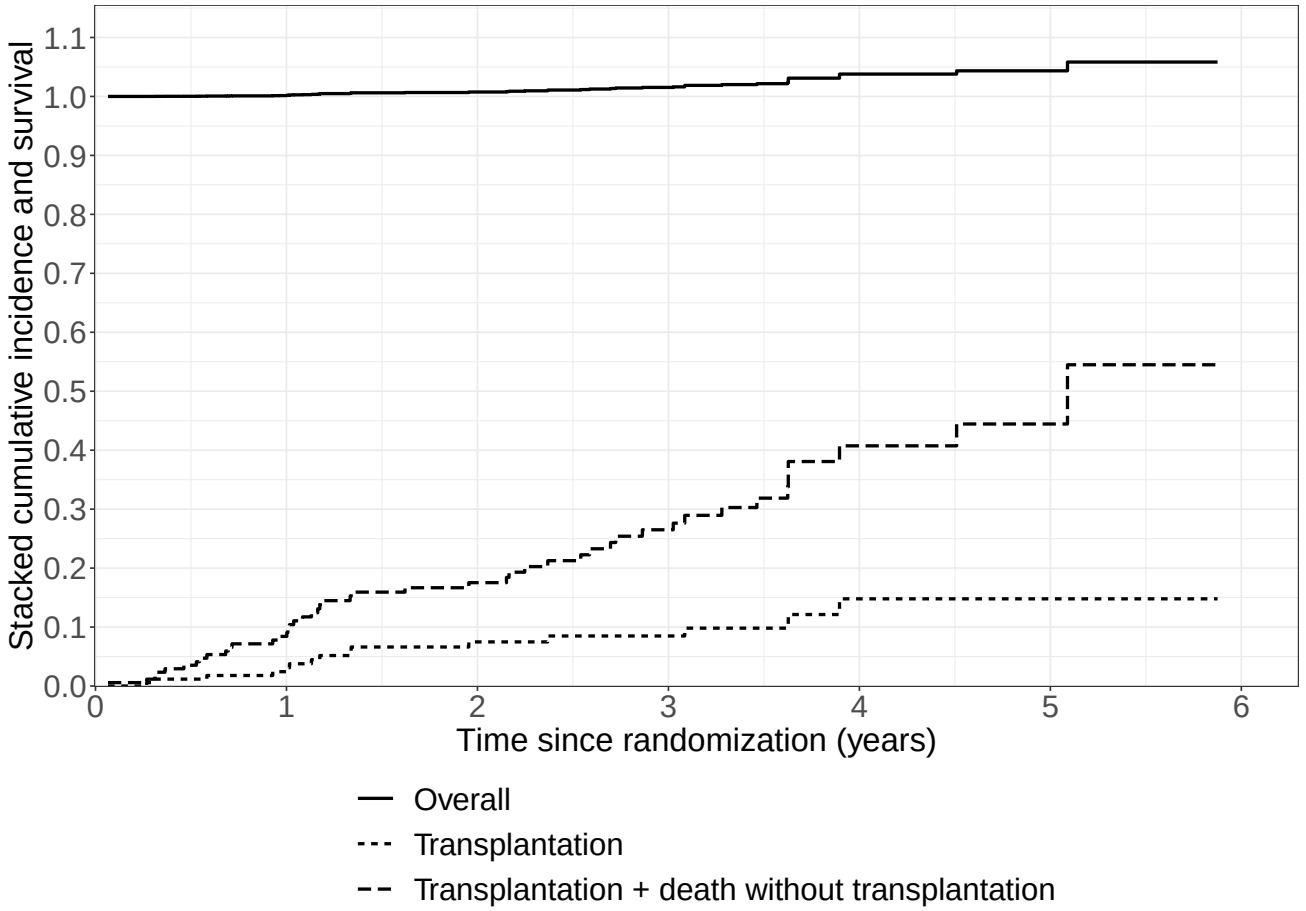
$$\widehat{Q}_h(t) \leq 1 - \widehat{S}_h(t).$$

The PBC-3 trial, death and transplantation



'Stacked' Aalen-Johansen estimators (placebo) – correct!

The PBC-3 trial, death and transplantation



'Stacked' 1 minus Kaplan-Meier estimators (placebo) – biased!

Markov processes

For the survival and competing risks models, $Q_h(t)$ is equal to the *transition probability* $P_{0h}(0,t)$ where

$$P_{hj}(s,t) = P(V(t) = j \mid V(s) = h), h, j = 1, ..., k, s \leq t.$$

For the special case of a *Markov process*, transition probabilities may be estimated non-parametrically using plug-in.

The *Markov property* is:

$$P(V(t+dt) = j \mid V(t-) = h, \text{past}) = P(V(t+dt) = j \mid V(t-) = h),$$

i.e., at any time t, the transition intensity $\alpha_{hj}(t)$ only depends on the past via the current state h (and on time t).

Markov processes

Assume that there are k states and that the transition intensities and cumulative intensities are, respectively, $\alpha_{hj}(t)$ and $A_{hj}(t)$ for $h, j = 1, ...k, h \neq j$. (Some of these may be zero everywhere, e.g., $\alpha_{10}(t)$ in the two-state model.) The matrix $\mathbf{P}(s,t)$ of *transition probabilities* $P_{hj}(s,t) = P(V(t) = j \mid V(s) = h)$ is then given by a *matrix product-integral*:

$$\mathbf{P}(s,t) = \prod_{(s,t]} \big(\mathbf{I} + d\mathbf{A}(u)\big),$$

where $A_{hh}(t) = -\sum_{j \neq h} A_{hj}(t)$.

We can estimate the cumulative transition intensities using the Nelson-Aalen estimator:

$$\widehat{A}_{hj}(t) = \int_0^t \frac{dN_{hj}(u)}{Y_h(u)}.$$

Markov processes: The Aalen-Johansen estimator

This means that $\mathbf{P}(s,t)$ can be estimated by *plug-in*:

$$\widehat{\mathbf{P}}(s,t) = \prod_{(s,t]} \left(\mathbf{I} + d\widehat{\mathbf{A}}(u)\right)$$

where, again, $\widehat{A}_{hh}(t) = -\sum_{j \neq h} \widehat{A}_{hj}(t)$ (Aalen and Johansen, *Scand. J. Statist.*, 1978).

Both the Kaplan-Meier estimator for the two-state model and the Aalen-Johansen estimator for the competing risks cumulative incidence have this form and (slightly confusingly) the term 'Aalen-Johansen' is used for both the general estimator and for the special case of competing risks.

The Aalen-Johansen estimator, simplest cases

For the simple survival model the 2×2 matrix at a death time, T is:

$$\begin{pmatrix} 1 - \frac{dN(T)}{Y(T)} & \frac{dN(T)}{Y(T)} \\ 0 & 1 \end{pmatrix}$$

For the competing risks model with two causes of failure, the 3×3 matrix at time T is:

$$\begin{pmatrix} 1 - \frac{dN_1(T)+dN_2(T)}{Y(T)} & \frac{dN_1(T)}{Y(T)} & \frac{dN_2(T)}{Y(T)} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

For a general Markov process, the $k \times k$ transition matrix at time T can be set up similarly.

ELOS

The state occupation probabilities $Q_h(t) = P(V(t) = h)$ equal the transition probabilities $P_{0h}(0, t)$ if every one begins in state 0 at time 0 (otherwise it is a mixture of such transition probabilities over the *initial state distribution*).

This means that state occupation probabilities can also be estimated in *Markov* processes using the Aalen-Johansen estimator.

The good news is that, as shown by Datta and Satten (*Stat. Prob. Letters*, 2001) and others, even for *non-Markov* processes the Aalen-Johansen estimator consistently estimates $Q_h(t)$.

This means that ELOS:

$$\varepsilon_h(t_0) = \int_0^{t_0} Q_h(t)dt$$

may also be estimated by plug-in.

The PBC-3 trial, ELOS, death and transplantation

	Placebo	CyA
RMST, 3 years	2.606	2.678
Years lost, 3 years		
Transplantation	0.143	0.086
Death without transpl.	0.251	0.236
	3.000	3.000

Estimates are areas under Kaplan-Meier, resp. Aalen-Johansen estimates at t_0 =3 years.

R code, K-M, Aa-J, ELOS

```
#Kaplan-Meier estimator (NB: same as for Nelson-Aalen):
library(survival)
pbcna <- survfit(Surv(years, status != 0) ~ tment, data = pbc3)
plot(pbcna)

#Aalen-Johansen estimator:
ajfit <- survfit(Surv(years, factor(status)) ~ tment,
data = pbc3)
plot(ajfit)

#ELOS:
print(ajfit,rmean=3)
```

The code gives a crude (non-stacked) version of the correct figure with cumulative incidences that may be improved upon.

Estimation of marginal parameters: recurrent events

For recurrent events with competing risks, plug-in leads to the Cook-Lawless (1997) estimator, Eq. (4.13):

$$\widehat{\mu}(t) = \int_0^t \widehat{S}(u-) d\widehat{A}^*(u)$$

($\widehat{S}$: marginal Kaplan-Meier, i.e., ignoring recurrent events; $\widehat{A}^*$: Nelson-Aalen for cumulative marginal rate given survival).

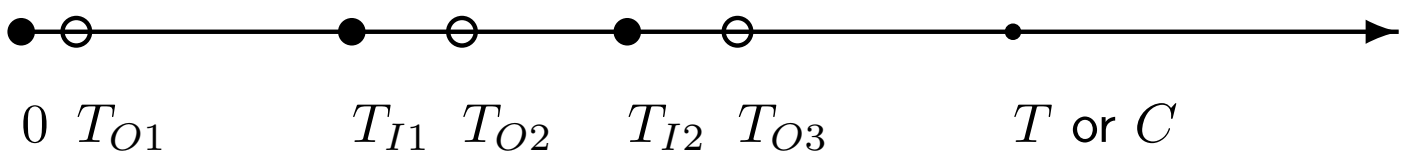
Without competing risks (i.e., $S(t) = 1$), the Nelson-Aalen estimator estimates $\mu(t)$ (which then equals the cumulative marginal rate).

Note that, as it was the case for competing risks, treating deaths as censorings (and estimating $\mu(t)$ by Nelson-Aalen), an upwards biased estimator is obtained.

Recurrent events: periods not at risk ('episodes')

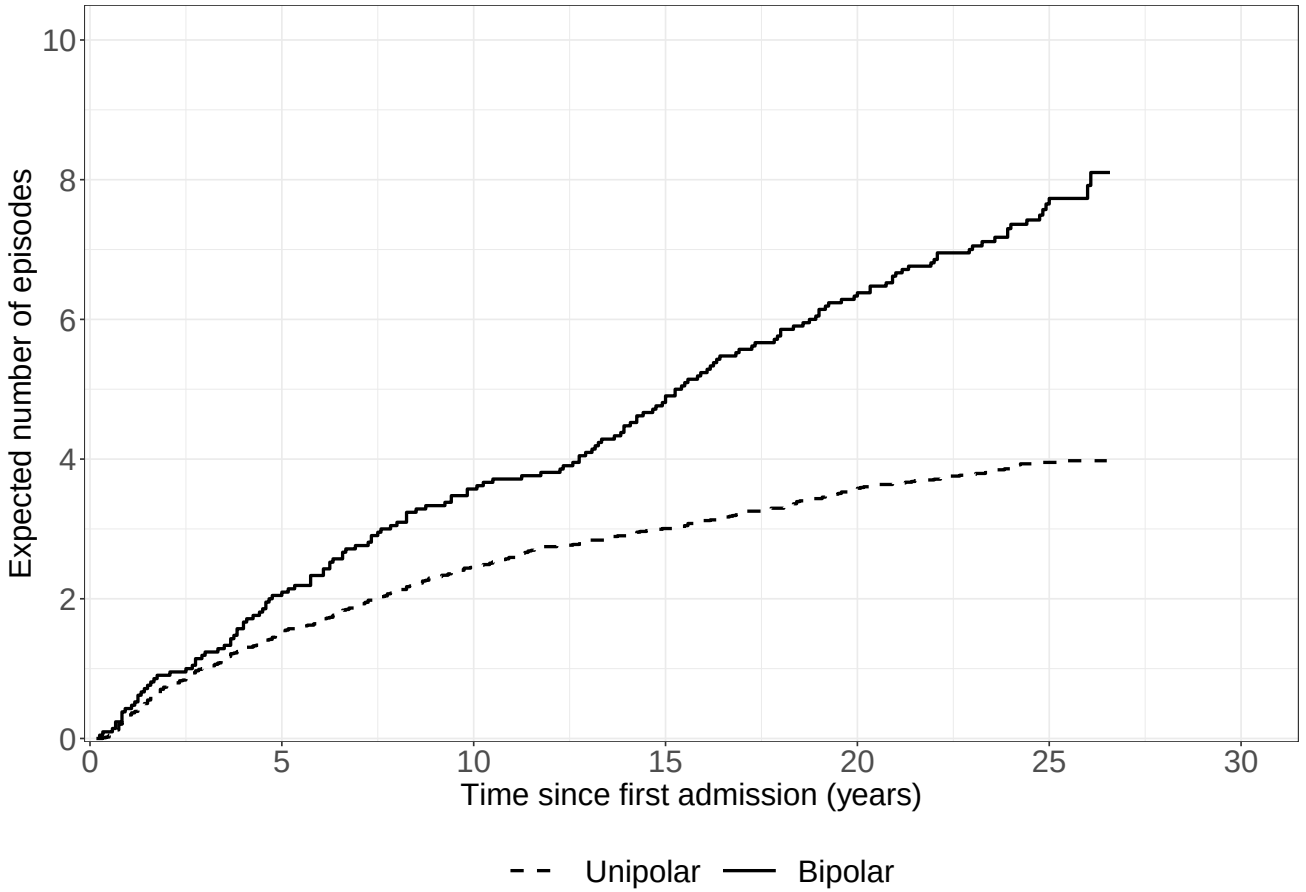
For the data on repeated psychiatric episodes, there are periods of on-going episodes where patients are not at risk for a new episode ('in hospital'), and these periods are accounted for when estimating the cumulative intensity.

The simplest way of taking these periods into account when estimating the marginal mean is to ignore them – estimate the expected number of 'bullets' in $[0, t]$ using the Nelson-Aalen/Cook-Lawless estimator and ignore the 'circles':



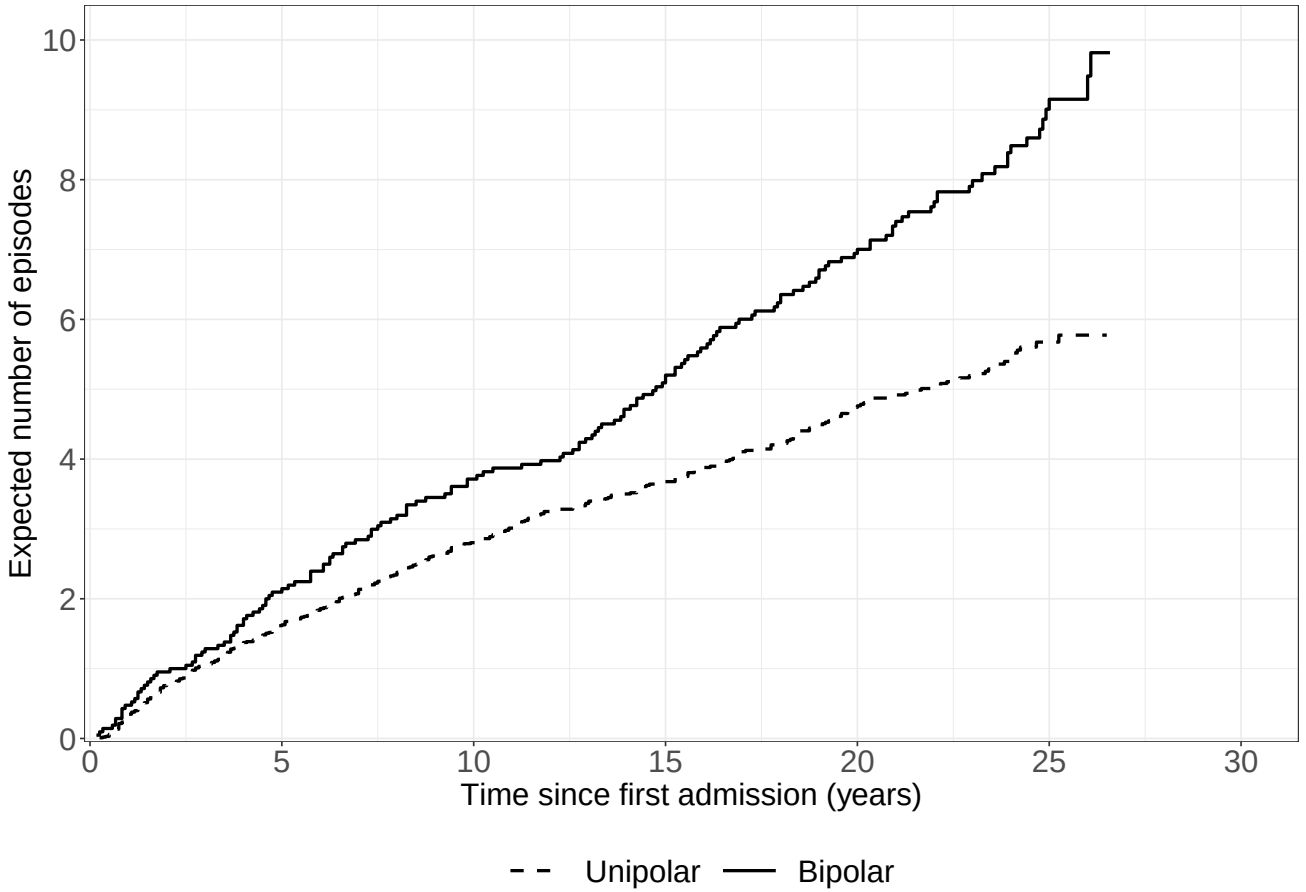
Supplement by analysis of time 'in hospital'/mortality: one way of getting few re-admissions is to keep the patient in hospital/kill him!

Repeated episodes in affective disorder



Cook-Lawless estimator – correct!

Repeated episodes in affective disorder



Nelson-Aalen estimator – biased!

R code: recurrent events

```
library(mets)

# Cook-Lawless estimator
# prev is start of previous episode

plot(recurrentMarginal(
  Event(prev,stop,status) ~ strata(bip) + cluster(id),
   data = angst, cause = 1, death.code = 2))

# Add wrong Nelson-Aalen estimator to plot (censor at death):

lines(survfit(Surv(prev,stop,status==1) ~ bip , data = angst),
      fun = "cumhaz")
```

Regression models

All of the plug-in methods discussed can also be applied when the estimated intensities are based on *regression models (with time-fixed covariates*).

As examples, survival functions and cumulative incidences can be estimated based on Cox models for the (cause-specific) hazards.

For competing risks, however, this technique does not provide parameters that directly quantify the association between the covariate and the cumulative incidence.

More on direct marginal regression models later!

R code: plug-in of Cox models

```
# Survival function for given covariates

mcox<-coxph(Surv(years,status>0) ~ tment + alb + log2bili,
data=pbc3, method="breslow")
predsurv<-data.frame(tment=0:1, alb=38, log2bili=log2(45))
plot(survfit(mcox,newdata=predsurv),lty=1:2)

# Cumulative incidences for given covariates

fitcr<-coxph(Surv(years,factor(status))~ tment + alb + log2bili,
             method = "breslow", data = pbc3, id=id)

predcr<-data.frame(tment=0, alb=38, log2bili=log2(45))
plot(survfit(fitcr,newdata=predcr), lty=1:2)
```

R exercises - Plug-In

Use data set PBC3.

Two-state model

1. Estimate and plot survival curves using the Kaplan-Meier estimator for each treatment arm. Also, plot Kaplan-Meier curves for treatment arms within each of the two bilirubin categories (`bili2`).
2. Based on the Kaplan-Meier estimates for each treatment arm (not by `bili2`), estimate the risk difference at year 3. Use the function `riskdiff()` provided at course web https://multi-state-book.github.io/Salerno25/
3. Based on the Kaplan-Meier estimates for treatment, estimate the RMST difference at year 3 between treatment arms. Use the function `rmstdiff()` at course web.

R exercises - Plug-In

Competing risks

4. For each transition, estimate and plot cumulative incidences for each treatment arm using the Aalen-Johansen estimator.

5. For each transition and based on the Aalen-Johansen estimates, estimate the risk differences at year 3 between treatment arms. Use the function `cidiff()` provided at course web.

6. For each transition and based on the Aalen-Johansen estimates, estimate the ELOS (cause-specific time lost) differences at year 3 between treatment arms. Use the function `yldiff()` provided at course web.

Recurrent events

Use data set `rr.csv` and see variables on
https://multi-state-book.github.io/Salerno25/ and use the `mets` package
https://cran.r-project.org/web/packages/mets/index.html.

7. Plot the Cook-Lawless estimator by treatment group.

8. Add wrong Nelson-Aalen estimator to plot (censor at death).