Marginal analysis of recurrent events

(Sections 4.2.3, 4.3, 5.5.4)

Per Kragh Andersen

Section of Biostatistics, University of Copenhagen

DSBS Course
Survival Analysis in Clinical Trials
2025

Overview

- Marginal parameters
- The expected number of events, no gaps, no mortality
- The expected number of events, no gaps, with mortality
- The expected number of events with gaps

Marginal parameters

RCTs

Both transition *intensities* and transition *probabilities* are conditional on the past development.

In a randomized trial the past may be influenced by treatment and may, therefore, be *intermediate* between treatment and an ultimate outcome like death or relapse.

In epidemiology, it is well-known that one should avoid conditioning on intermediate variables since this may mask exposure effects, and the same holds true in RCTs.

An example is the *PROVA* trial where the primary outcomes were bleeding and death without bleeding, while death after bleeding was not a primary outcome. This is because experiencing bleeding may have been affected by treatment, and randomization does no longer ensure balance between treatment groups for those who enter the intermediate bleeding state.

For the same reasons, (internal/endogeneous) time-dependent covariates should not be adjusted for when analyzing a hazard model in a RCT

Marginal parameters for recurrent events

For analysis of *recurrent events*, it has therefore been argued (e.g., by Cook and Lawless, Springer book, 2007) that intensity-based models that condition on previous events are not optimal when analyzing RCTs, and *marginal* parameters have been put forward instead.

Marginal parameters in multi-state models include:

- State occupation probabilities $Q_h(t) = P(V(t) = h)$
- Average time spent in a given state
- Times (from 0) to entry into a given state
  In the special case of recurrent events, one could (also) focus on
- Expected number of events in $[0, t]$: $\mu(t) = E(N(t))$

As we shall see, such parameters may some times be analyzed directly, i.e., without going via the intensities.

Times to entry into given states

For recurrent events (at least when there are no terminal events), one could study the distribution of the event times $T_1, T_2, ...$ For this purpose, the model of Wei, Lin and Weisfeld (*JASA*, 1989) (the 'WLW' model) may be used. This is a series of Cox models for $T_1, T_2, ...$, e.g.,

$$\alpha_h(t \mid Z) = \alpha_{h0}(t) \exp(\beta_h^{\mathsf{T}} Z)$$

for the hazard function for the *marginal distribution* of time to event number h. Technically, the model is fitted as separate Cox models for each event time pretending that, at any time where a subject is alive and uncensored, he or she is *at risk for all event numbers*, e.g., at risk for the second event before the occurrence of the first.

Intra-individual dependence is treated by using the robust sandwich variance estimator for the parameters.

A remark on the marginal Cox model

While the WLW model (the 'marginal Cox model') may not be the best model choice for a recurrent events analysis (we will introduce more useful methods below), it is useful for other kinds of 'clustered survival data'.

This could be survival times for related individuals (families, litters, ..) and in *cluster-randomized studies*.

By using a marginal model in such cases it is possible to draw inference on the (marginal) distribution of survival times in presence of an intra-cluster correlation and without having to specify the form of the dependence.

This is the general idea behind 'GEE' methods (generalized estimating equations).

Expected number of events, no gaps, no mortality

Plug-in

The expectation of a discrete random variable $Y = 0, 1, 2...$ is

$$E(Y) = \sum_h h \cdot P(Y = h).$$

For the random variable $N(t)$, this is:

$$\mu(t) = E(N(t)) = \sum_h h \cdot Q_h(t)$$

which may be estimated whenever state occupation probabilities are estimable.

However, how many terms should be included in the sum?

Parametric models

The counting process likelihood may also form the basis for *parametric* models, e.g., models with a *constant* intensity. This gives a *homogeneous Poisson process* and the MLE is a simple occurrence/exposure rate:

$$\widehat{\lambda} = \frac{\sum_i N_i(\infty)}{\sum_i \int_0^\infty Y_i(t)dt}.$$

This is the simple descriptive number that is often used in RCTs with a time-to-event outcome (e.g., LEADER) and may also be used for recurrent events.

If one believes in the model, the estimate $\widehat{\lambda} \cdot t$ of $\mu(t) = E(N(t))$ follows.

Frailty models

The constant intensity model is very simple and often too simple for most practical purposes. However, with a frailty:

$$\alpha_i(t \mid A_i = a_i) = a_i \cdot \alpha_0$$

a more realistic model is obtained. Assuming the frailty to be gamma distributed, the number of events in $[0, t]$ follows a *negative binomial* distribution. This is a model that has been used in practice.

Also the *negative binomial process* obtained by multiplying a gamma frailty to an inhomogeneous Poisson process:

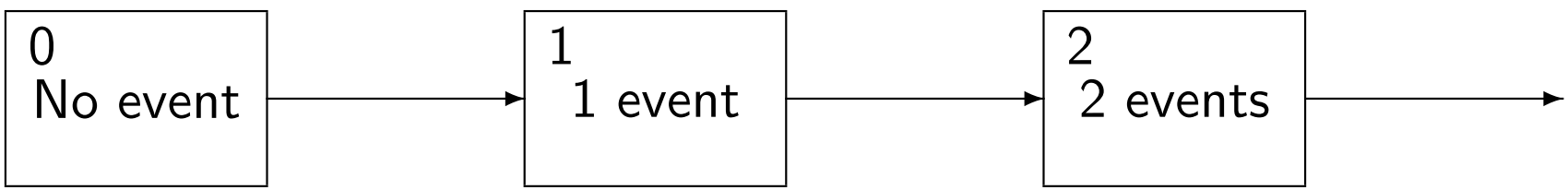
$$\alpha_i(t \mid A_i = a_i) = a_i \cdot \alpha_0(t)$$

has been analyzed (e.g., Cook and Lawless, Springer Book, 2007, Ch. 3).

Model with no mortality and no gaps

Plug-in methods are, however, not what is being used most in practice. This is probably due to the fact that, like in 'standard survival analyses' simple *non-parametric* (and semi-parametric) methods are available.

The simplest case is when there are no 'gaps' between events and when mortality is negligible:



Model with no mortality and no gaps

Instead of focusing on the transition intensities, we go directly for $\mu(t) = E(N(t))$ or its derivative $\mu'(t)$ which is some times denoted the *marginal rate function*. It has the interpretation:

$$\mu'(t) \approx E(dN(t))/dt$$

in contrast to the intensity that is conditional on the past

$$E(dN(t) \mid \text{past})/dt.$$

Let $Y_i(t) = I(C_i \geq t)$ be the at-risk indicator for subject i and assume first that censoring C is independent of the recurrent events process $N(t)$. Then the estimating equation:

$$\sum_i Y_i(t)\big(dN_i(t) - d\mu(t)\big) = 0$$

is *unbiased*.

The Nelson-Aalen estimator

The equation leads to the consistent estimator

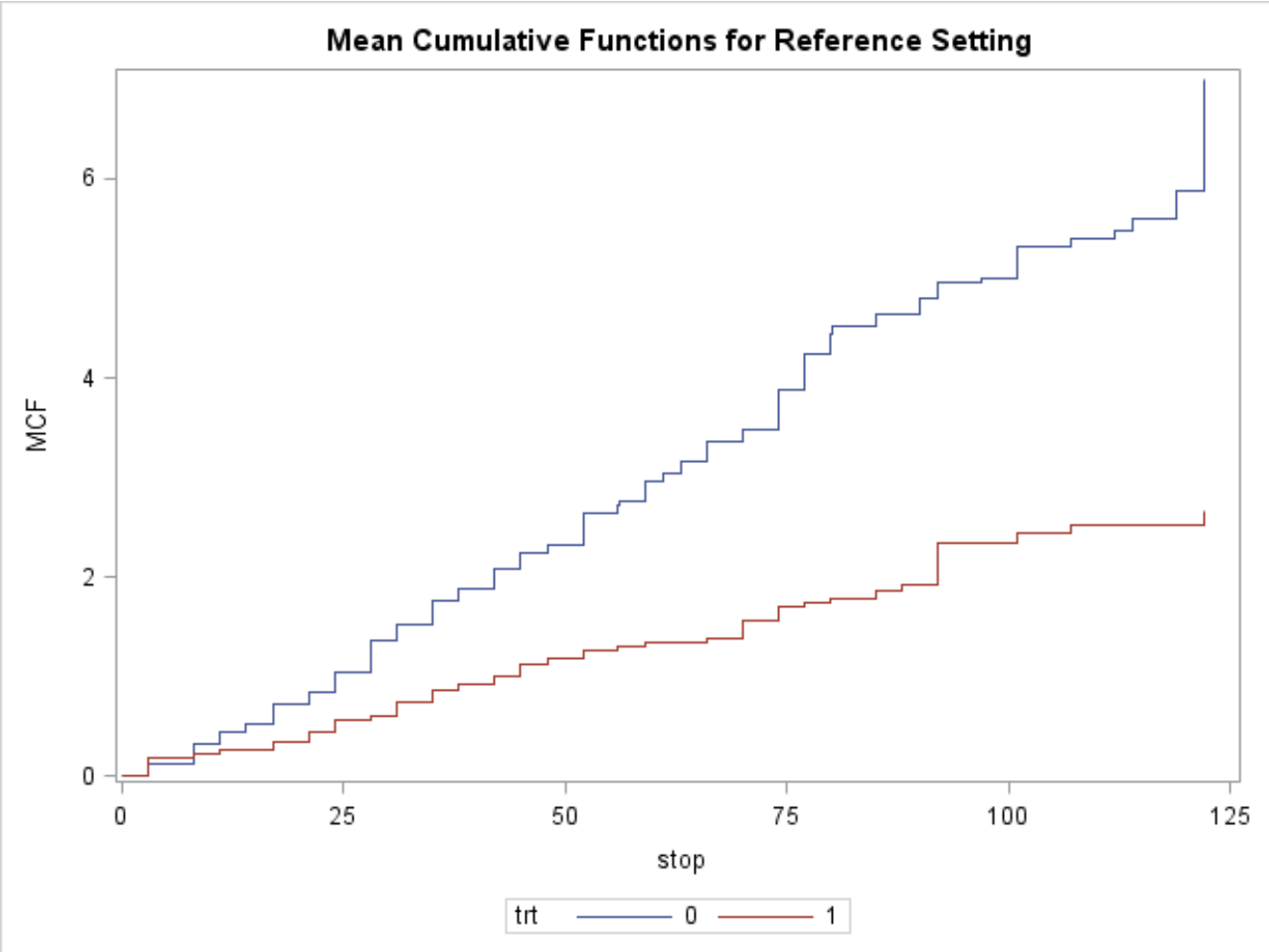
$$d\widehat{\mu}(t) = \frac{\sum_i dN_i(t)Y_i(t)}{\sum_i Y_i(t)}$$

which is simply the *Nelson-Aalen estimator*

$$\widehat{\mu}(t) = \int_0^t \frac{dN(u)}{Y(u)}$$

with $N(t) = \sum_i N_i(t)$ and $Y(t) = \sum_i Y_i(t)$. The estimator was discussed by Lawless and Nadeau (*Technometrics*, 1995) building on earlier work by Nelson.
For the special case where $N(t)$ is an inhomogeneous Poisson process we have $P(dN(t) = 1 \mid \text{past}) = P(dN(t) = 1)$ and the mean then equals the cumulative intensity (which now has a nice interpretation!) but it is important to notice that the estimator works without the Poisson assumption.

The Nelson-Aalen estimator for rats data



A two-sample test

Lawless and Nadeau (1995) also suggested a two-sample test in the spirit of the logrank test:

$$U(t) = \int_0^t \frac{Y_1(u)Y_2(u)}{Y_1(u) + Y_2(u)} \left(d\widehat{\mu}_1(u) - d\widehat{\mu}_2(u) \right)$$

together with a consistent estimator of the variance of $U(\tau)$.

For the rats data, this gives the value 11.20, $P < 0.001$.

The test is obtainable as a *score* test in the Cox-type model ('LWYY') to be described below.

A regression model for $\mu(t)$

One may consider a Cox-type regression model for the mean function $\mu(t)$ (Lawless and Nadeau, 1995; asymptotic results by Lin, Wei, Yang, Ying, *JRSS(B)*, 2000 – the 'LWYY' model):

$$\mu(t \mid Z) = \mu_0(t) \exp(\beta^{\mathsf{T}} Z)$$

with the baseline mean function $\mu_0(t)$ completely unspecified. A set of unbiased estimating equations is:

$$\begin{aligned} \sum_i \int_0^\infty Y_i(t)\big(dN_i(t) - d\mu_0(t)\exp(\beta^{\mathsf{T}} Z_i)\big) &= 0 \\ \sum_i \int_0^\infty Y_i(t) Z_{ij}\big(dN_i(t) - d\mu_0(t)\exp(\beta^{\mathsf{T}} Z_i)\big) &= 0, \quad j = 1, ..., p, \end{aligned}$$

(if there are p covariates). The equations come from a *working Poisson model* but apply more generally.

A regression model for $\mu(t)$

Solving the first equation for fixed β gives the 'Breslow-type' estimator:

$$d\widehat{\mu}_0(t) = \frac{\sum_i dN_i(t) Y_i(t)}{\sum_i Y_i(t) \exp(\beta^{\mathsf{T}} Z_i)}$$

which may then be inserted into the second equation(s) to give:

$$U_j(\beta) = \sum_i \int_0^\infty Y_i(t)\big(Z_{ij} - \frac{\sum_\ell Y_\ell(t) X_{\ell j} \exp(\beta^{\mathsf{T}} Z_\ell)}{\sum_\ell Y_\ell(t) \exp(\beta^{\mathsf{T}} Z_\ell)}\big) dN_i(t) = 0$$

which is just the score equation from the Cox-model.

This means that software for the Cox model can do the job - just use robust variances.
Rats data $\exp(\widehat{\beta}) = 0.442$, 95% c.i. $(0.300, 0.652)$ (model-based: $(0.328, 0.595)$).

R code: rats data

```
plot(survfit(Surv(start,stop,status==1)~trt+cluster(id),
data=rats),cumhaz=TRUE)

coxph(Surv(start,stop,status==1)~trt+cluster(id),
data=rats)
```

Non-independent censoring

When censoring is not completely independent of $N(t)$, the estimating equation (without covariates):

$$\sum_i Y_i(t)\big(dN_i(t) - d\mu(t)\big) = 0$$

is no longer necessarily unbiased and to fix this problem, IPCW may be used.

We need an estimator, $\widehat{G}_i(t)$, of the censoring distribution (distribution of C_i).

This may depend on the past of $N_i(t)$, but not on the future.

Non-independent censoring

It can then be shown that (for the situation without covariates) the estimating equation

$$\sum_i \frac{Y_i(t)}{\widehat{G}_i(t)}\big(dN_i(t) - d\mu(t)\big) = 0$$

is approximately unbiased and leads to the estimator

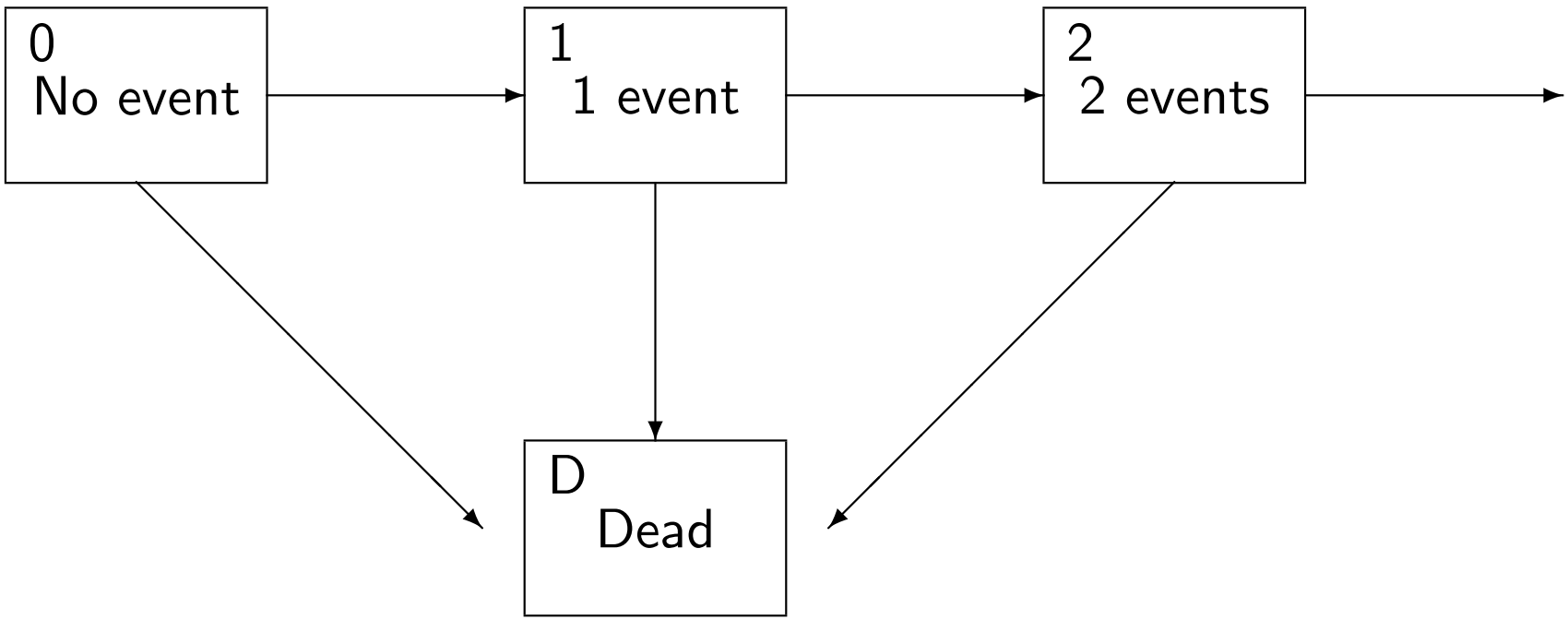
$$d\widehat{\mu}(t) = \frac{\sum_i dN_i(t)Y_i(t)/\widehat{G}_i(t)}{\sum_i Y_i(t)/\widehat{G}_i(t)}.$$

The equations for the Cox-type model may be fixed in a similar way.

In both situations, robust variances are used.

Expected number of events, no gaps, with mortality

Model with no gaps, mortality allowed



Remarks

- The presence of a non-negligible mortality rate complicates the situation
- The resulting model has several similarities with *competing risks*, and we shall see that inference methods for the new model are inspired by those for competing risks
- We shall also see that the same types of bias arise when disregarding mortality in the new model, as those that appear when using '1 - K-M' instead of the correct cumulative incidence estimator with competing risks

The Cook-Lawless estimator

When mortality plays a role, the Nelson-Aalen estimator for $\mu(t) = E(N(t))$ will be *upwards biased*. This is because events can only happen as long as the subject is alive and by 'treating death as censoring' we pretend that we have an immortal population where events can happen all the time.

So, we need an estimator for $\mu(t)$ that accounts for mortality and it turns out that a very simple one exists.

Let

$$\alpha_i^*(t) \approx E(dN_i(t) \mid T_i \geq t)/dt$$

be the *marginal rate function given alive* (T_i is the survival time for subject i).

The Cook-Lawless estimator

The corresponding cumulative function $A^*(t) = \int_0^t \alpha^*(u)du$ (assuming that all subjects have the same $\alpha^*(t)$) can be estimated by the Nelson-Aalen estimator. The function $A^*(t)$, however, has no nice interpretation but the marginal mean is now simply

$$\mu(t) = E(N(t)) = \int_0^t S(u)dA^*(u)$$

where $S(\cdot)$ is the survival function (distribution of T) and can be estimated simply be Kaplan-Meier. This leads to the estimator

$$\widehat{\mu}(t) = \int_0^t \widehat{S}(u-)\frac{\sum_i Y_i(u)dN_i(u)}{\sum_i Y_i(u)}.$$

suggested by Cook and Lawless (*Statist in Med.*,1997) and studied in more details (including asymptotic results) by Ghosh and Lin (*Biometrics*, 2000).

Math

$$
\begin{aligned}
E(N(t)) &= E\left(\int_0^t dN(u)\right) \\
&= \int_0^t E(dN(u)) \\
&= \int_0^t P(dN(u) = 1) \\
&= \int_0^t P(T > u) P(dN(u) = 1 \mid T > u) \\
&+ \int_0^t P(T \leq u) P(dN(u) = 1 \mid T \leq u) \\
&= \int_0^t S(u) dA^*(u) + 0
\end{aligned}
$$

A two-sample test

Ghosh and Lin (2000) suggested a two-sample test in the spirit of the logrank test (and the Lawless-Nadeau test for cumulative means in the case of no mortality):

$$U(t) = \int_0^t \frac{Y_1(u)Y_2(u)}{Y_1(u) + Y_2(u)} \left(d\widehat{\mu}_1(u) - d\widehat{\mu}_2(u) \right)$$

together with a consistent estimator of the variance of $U(\tau)$.

Regression analysis: 'Ghosh-Lin model'

In a follow-up paper, Ghosh and Lin (*Statistica Sinica*, 2002) developed a regression model for $\mu(t)$:

$$\mu(t \mid Z) = \mu_0(t) \exp(\beta^{\mathsf{T}} Z)$$

with asymptotic theory. Estimation follows to a large extent what we saw for the Fine-Gray model in the sense that they first studied the case where censoring times C_i are known for everybody - even for those who died and, next, an IPCW method was derived. For the first case the estimating equation is:

$$U(\beta) = \sum_i \int_0^\infty \left(Z_i - \bar{Z}^C(t)\right) I(C_i \geq t) dN_i(t)$$

where $\bar{Z}^C(t)$ is the average

$$\bar{Z}^C(t) = \frac{\sum_\ell I(C_\ell \geq t) Z_\ell \exp(\beta^{\mathsf{T}} Z_\ell)}{\sum_\ell I(C_\ell \geq t) \exp(\beta^{\mathsf{T}} Z_\ell)}.$$

Regression analysis

In the more common case where censoring times are not known for all, a model for $P(C > t \mid Z)$ (possibly without Z) is needed. Let $\widehat{G}(t)$ be the estimator for the censoring distribution, e.g., Kaplan-Meier if C is independent of Z or a Cox-model-based estimator otherwise.
Ghosh and Lin introduced the weights

$$w_i(t) = I(C_i \geq T_i \wedge t)\widehat{G}(t)/\widehat{G}(\widetilde{T}_i \wedge t)$$

and showed that:

$$E(w_i(t)) \approx G(t)$$

which is the expectation of $I(C_i \geq t)$ and, therefore, this indicator in the previous estimating equation can be replaced by $w_i(t)$.

Regression analysis

The resulting estimating equation is

$$U(\beta) = \sum_i \int_0^\infty \left(Z_i - \bar{Z}^G(t)\right) w_i(t) dN_i(t)$$

with

$$\bar{Z}^G(t) = \frac{\sum_\ell w_\ell(t) Z_\ell \exp(\beta^\mathsf{T} Z_\ell)}{\sum_\ell w_\ell(t) \exp(\beta^\mathsf{T} Z_\ell)}.$$

Finally, variances and the baseline mean function can be estimated, the latter by a Breslow-type estimator:

$$\widehat{\mu}_0(t) = \sum_i \int_0^t \frac{w_i(u) dN_i(u)}{\sum_\ell w_\ell(u) \exp(\widehat{\beta}^\mathsf{T} Z_\ell)}.$$

Regression analysis - studying both end-points

The model for $\mu(t \mid Z)$ may be combined with, e.g., a Cox model for T_j.

However, summarizing results from the two models is not obvious - just as summarizing results from Fine-Gray models for two competing causes is not obvious.

In general, one must study both $N(t)$ and T because 'one way to get few recurrent events would be to kill the patient' – a fact that, obviously, may lead to incorrect interpretation of treatment effects.

Regression analysis

Cook et al. (*JASA*, 2009) discussed an alternative ('plug-in') approach to regression analysis of $\mu(t)$ - very much similar to estimating the cumulative incidence with competing risks by plugging-in models for the cause-specific hazards.

Recall that $\mu(t) = E(N(t)) = \int_0^t S(u)dA^*(u)$ which means that, given covariates Z, we have:

$$\mu(t \mid Z) = E(N(t) \mid Z) = \int_0^t S(u \mid Z)dA^*(u \mid Z)$$

and a regression model for $\mu(t)$ may be obtained by combining regression models for $S(\cdot)$, e.g., a simple Cox model, and for the marginal rate, given survival $\alpha^*(\cdot) = dA^*(\cdot)$, e.g., an AG-type model as discussed above.

Interpretation suffers from the same problems as those faced in competing risks.

A ‘composite end-point’

In the competing risks model, the situation could be simplified by considering a *composite end-point*.

In the model for recurrent events with mortality one can do the same by defining the ‘recurrent’ event:

event or death

That is, in the simplest case, death is considered as an event ‘on an equal footing’ as the recurrent event, e.g., if death occurs after the first recurrent event then the death event is regarded as the second ‘composite event’.

Adding (utility) *weights* has been suggested to signify that death is worse than a new event (Mao and Lin, 2016, *Biostatistics*).

A 'composite end-point': inference

Mao and Lin (2016, *Biostatistics*) showed that inference for this (weighted) composite end-point may follow that for the Ghosh-Lin model.

Thus, if $N_i(t)$ is the weighted composite event counting process for subject i, with an assumed multiplicative marginal mean

$$E(N_i(t) \mid Z_i) = \mu_0(t) \exp(\beta^{\mathsf{T}} Z_i),$$

then, defining the IPC weights

$w_i(t) = I(C_i \geq T_i \wedge t)\widehat{G}(t)/\widehat{G}(\widetilde{T}_i \wedge t)$

the estimating equation

$$U(\beta) = \sum_i \int_0^\infty \left(Z_i - \bar{Z}^G(t)\right) w_i(t) dN_i(t)$$

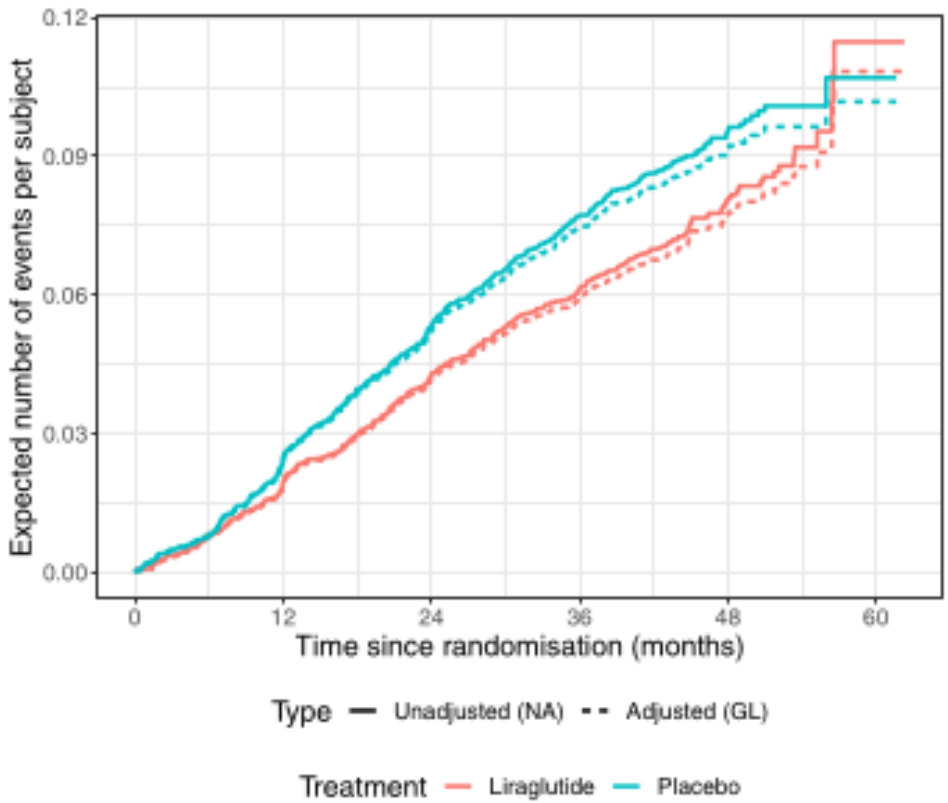
with

$$\bar{Z}^G(t) = \frac{\sum_\ell w_\ell(t) Z_\ell \exp(\beta^{\mathsf{T}} Z_\ell)}{\sum_\ell w_\ell(t) \exp(\beta^{\mathsf{T}} Z_\ell)}.$$

is unbiased.

LEADER trial

Estimated average numbers of myocardial infarctions: One curve for each treatment group where mortality is treated as censoring and one for each group where mortality is treated as a competing risk (CL: Cook-Lawless estimates, NA: Nelson-Aalen estimates).



LEADER trial: liraglutide vs. placebo

Recurrent MI	$\exp(\widehat{\beta})$	95% c.i.
'LWYY' model	0.849	$(0.714, 1.009)$
Ghosh-Lin model	0.853	$(0.718, 1.013)$
Cox model for mortality	0.847	$(0.738, 0.972)$
Mao-Lin model	0.853	$(0.763, 0.954)$

Furberg et al. (2022) generalized the Ghosh-Lin/Mao-Lin models to a situation where some cause-specific mortality (CV death) was part of the composite event (MACE), and other cause-specific mortality (non-CV death) was a competing risk.

R code: LEADER (1)

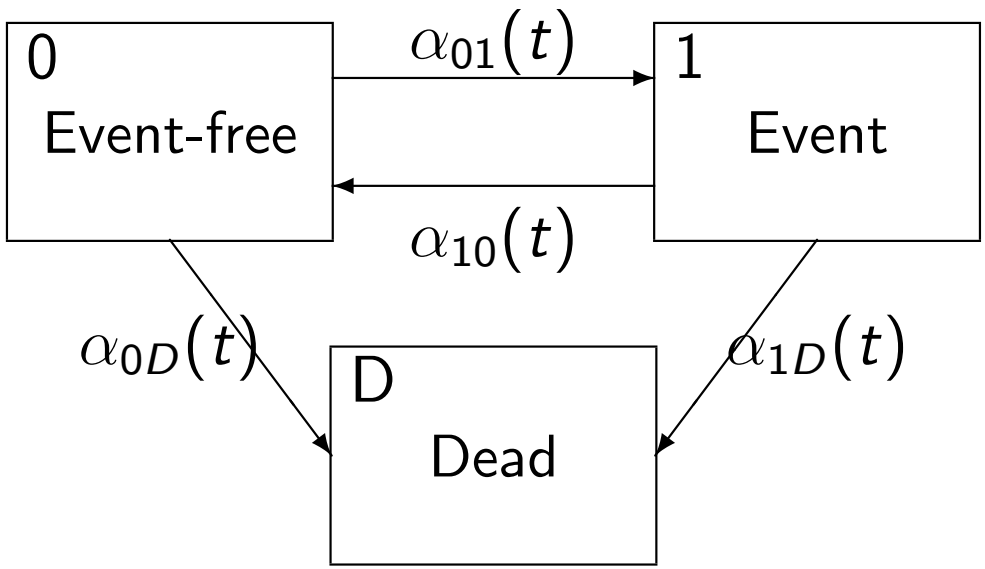
```
library(survival)

#Nelson-Aalen estimators
NAafit <- survfit(Surv(start, stop, mistatus == 1) ~
treat + cluster(id), data = leader_mi)

plot(NAafit, cumhaz=TRUE)

#LWYY model (not correct with competing risks)
LWYYfit <- coxph(Surv(start, stop, mistatus == 1) ~
treat + cluster(id), data = leader_mi)

#Cox model for survival
coxfit <- coxph(Surv(start, stop, mistatus == 2) ~ treat,
data = leader_mi)
```

R code: LEADER (2)

```
library(mets)

# Cook-Lawless estimators
plot(recurrentMarginal(Event(start,stop,mistatus)~
strata(treat) + cluster(id),
data = leader_mi, cause = 1, death.code = 2))

# Ghosh-Lin model
fitGL <- recreg(Event(start, stop, mistatus)~factor(treat)
+ cluster(id), data = leader_mi, cause = 1, cens.code = 0,
        death.code = 2)

# Mao-Lin model
fitML <- recreg(Event(start, stop, mistatus)~factor(treat)
+ cluster(id), data = leader_mi, cause = c(1,2),
cens.code = 0, death.code = 2)
```

Expected number of events with gaps

Intensity-based model



All intensities may depend on the past.

Example: Pulmonary exacerbations

Data from Cook and Lawless (2007) book (Therneau and Hamilton, 1997, *Statist. in Med.*; Fuchs et al. (*NEJM*, 1974)). 645 patients with cystic fibrosis randomized to rhDNase (321) or placebo (324) followed from randomization and about 169 days. The data set includes the variables:

- `id`
- `trt` (treatment indicator), `fev` (baseline value)
- `start`, `stop`, `status`
- `etype` (1 if 'at risk', 2 if 'under treatment')
- `enum` (record no.), `enum1` (gap time no.), `enum2` (treatment period no.)

Here, mortality is negligible.

Example: Psychiatric admissions

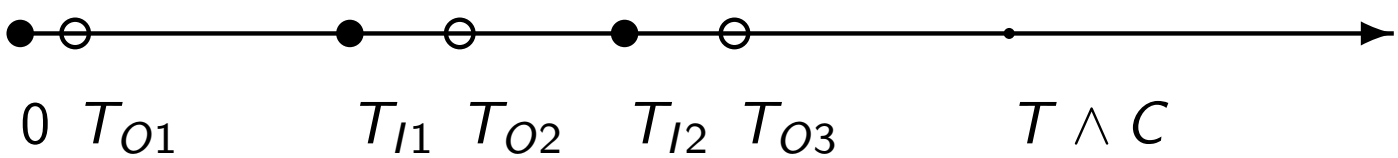
Clinical data collected by Swiss psychiatrist Jules Angst in Zürich. We restrict attention to prospectively collected data on patients with a first diagnosis after 1958:

- 119 patients with unipolar ('depression') or bipolar ('manic-depression') disorder
- dates on admission to and discharge from psychiatric hospital
- 78 patients died during follow-up
- covariates include
  - sex
  - age at first diagnosis
  - year at first diagnosis

Direct estimation of $\mu(t)$

In both examples one could consider 'cycles' instead of times from beginning of the at-risk period to new event, i.e.,

- time from a pulmonary exacerbation to the next, disregarding that there was an 'under treatment' period where, in principle, a new exacerbation was not possible
- time from a psychiatric admission to the next, disregarding that there was an admission period where a new admission was not possible



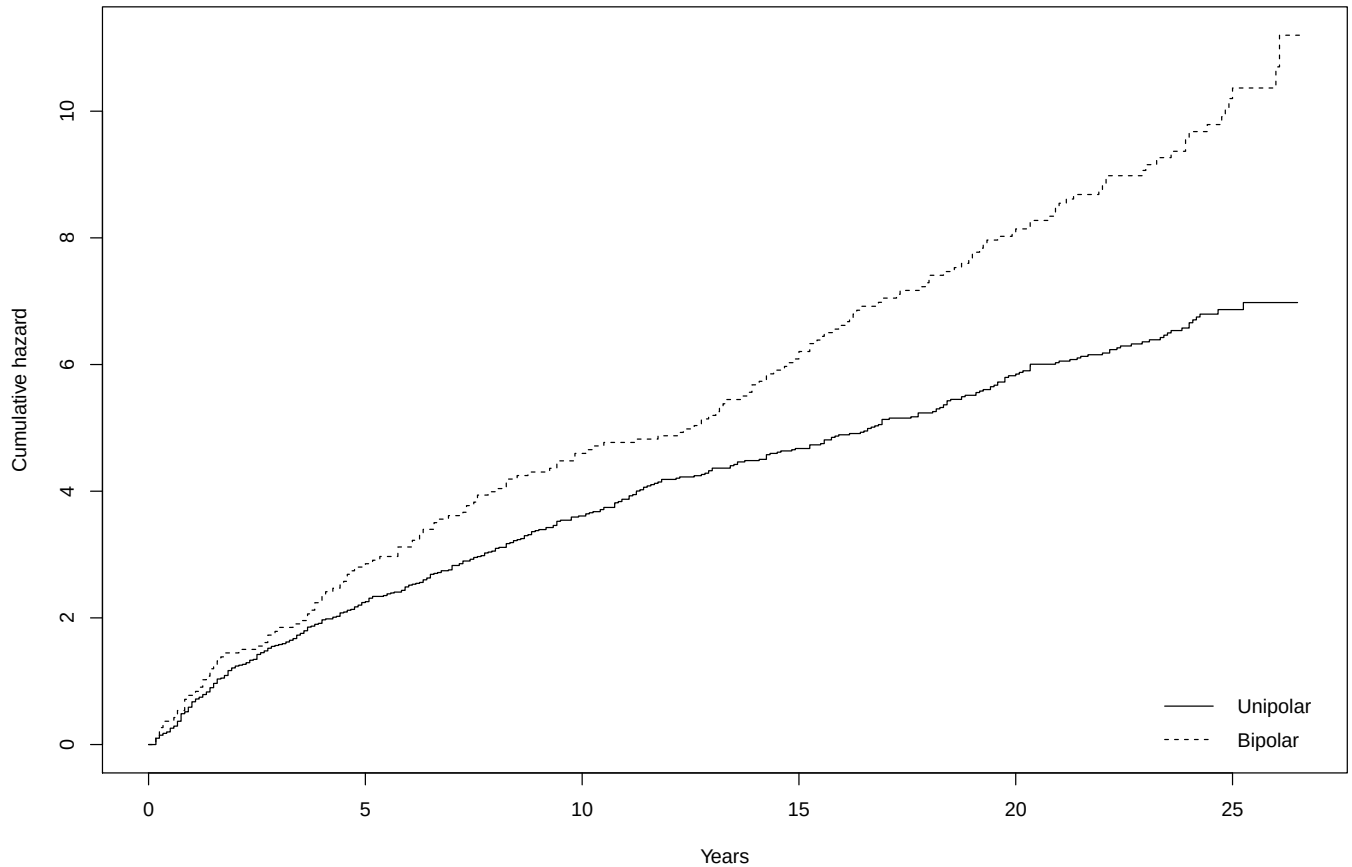
Direct estimation of $\mu(t)$

The Nelson-Aalen/Cook-Lawless estimators estimate the mean number of events in a world where events have a certain duration - a parameter that is of interest even though it would be different in a world where, say, admission intensities were the same but discharge intensities different.

The method works because the estimating equations from earlier are still unbiased for $E(dN(t))$.

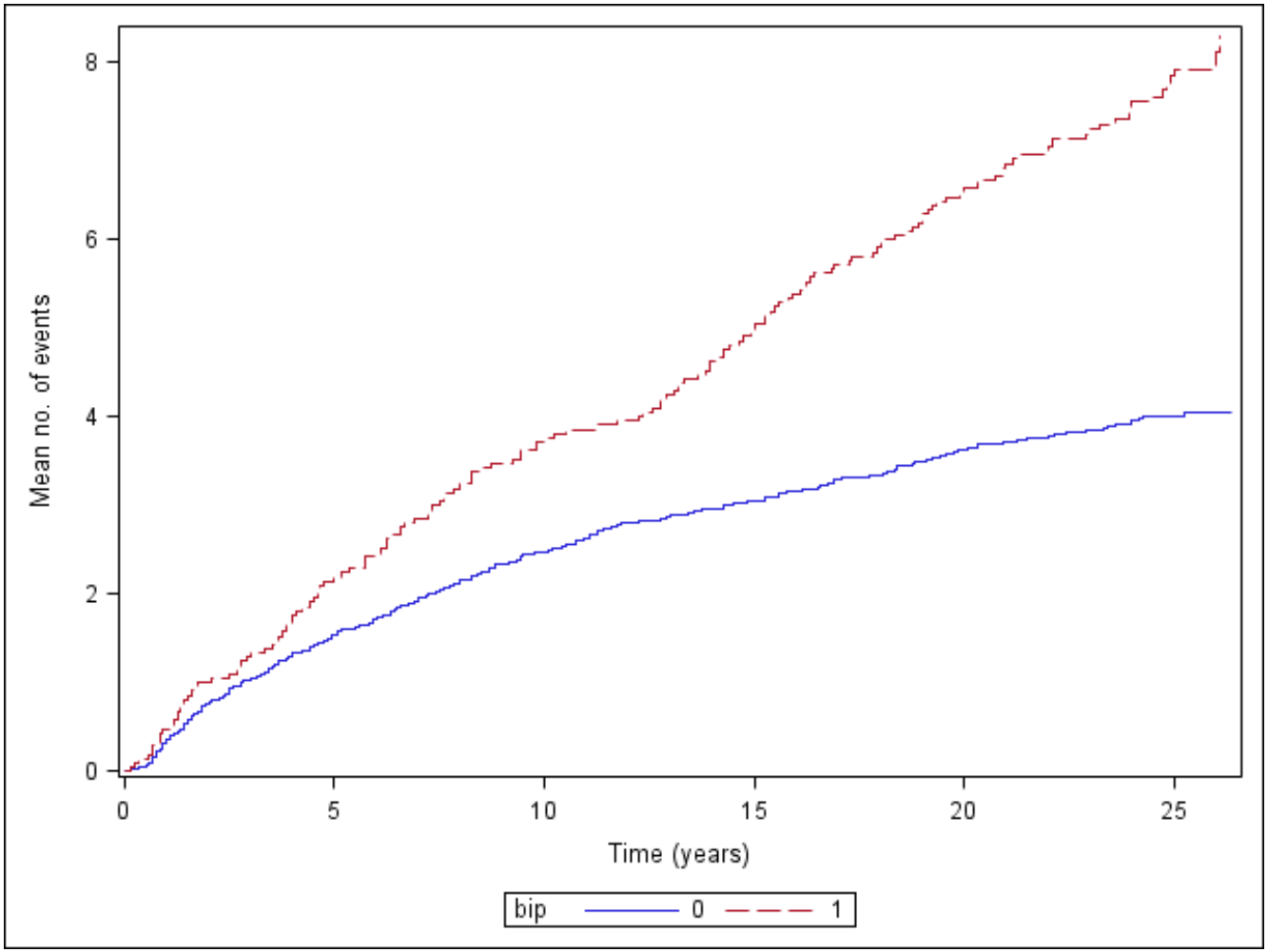
The Nelson-Aalen/Cook-Lawless estimates may be supplemented by estimates of average times in hospital and out of hospital and of the mortality.

Psychiatric admissions: cumulative intensities



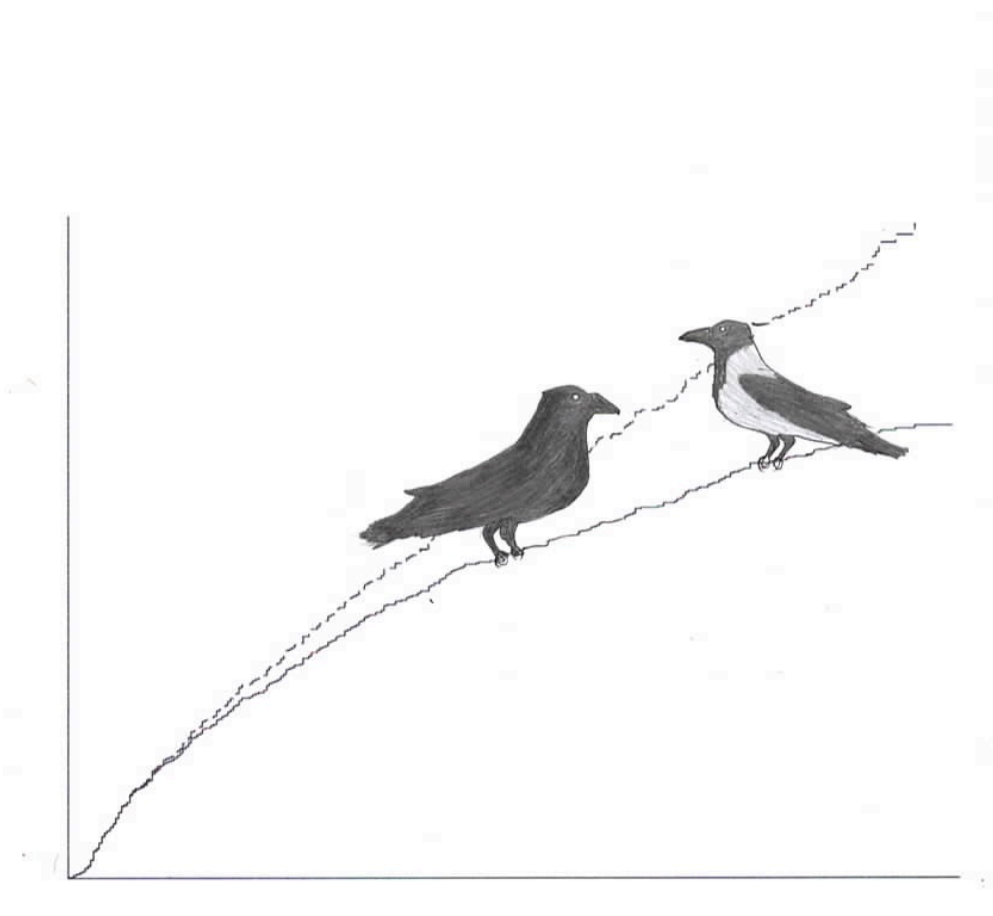
Cox model gives $\exp(\widehat{\beta}) = \exp(0.372) = 1.45 \quad (1.08, 1.94)$.
Adjusted for $N_i(t-)$: $\exp(\widehat{\beta}) = 1.10 \quad (0.83, 1.43)$.

Psychiatric admissions: estimated mean functions

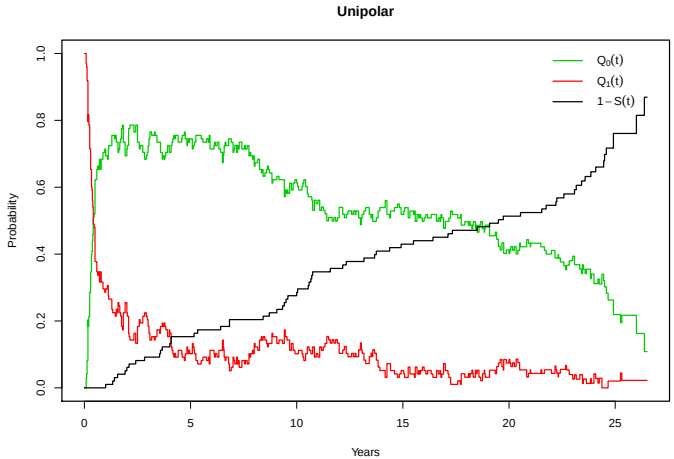


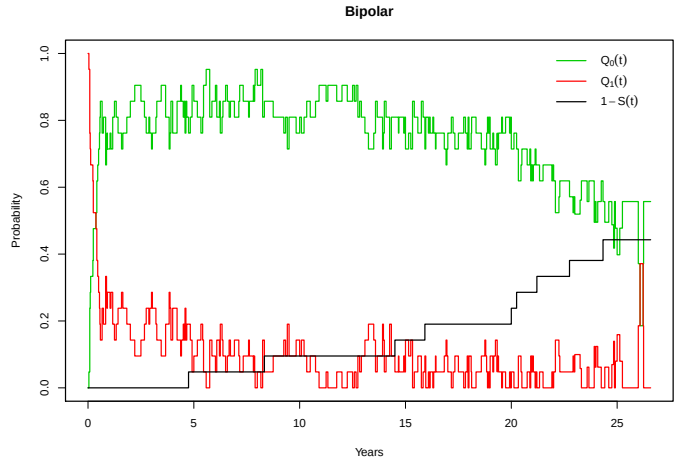
Ghosh-Lin regression model: estimated mean ratio is
$\exp(0.673) = 1.96 \quad (1.49, 2.57)$.

Psychiatric admissions: estimated mean functions, unipolar patients



Psychiatric admissions: state occupation probabilities





Cox model for mortality: estimated hazard ratio is
$\exp(-0.891) = 0.410(0.204, 0.825)$

Furthermore, bipolar patients tend to have shorter hospitalisations (in addition to lower mortality and higher admission intensity).

Results from Andersen, Angst, Ravn *LIDA* (2019).

Summary

- Models for recurrent events can be *intensity-based* or *marginal*, the latter being most relevant for RCTs
- Distinguish between situations with (most relevant!) or without competing risks
- For recurrent events with competing risks, the expected number of events may be estimated non-parametrically using the Cook-Lawless estimator and analyzed using the Ghosh-Lin regression model
- Using the Nelson-Aalen estimator instead of the Cook-Lawless estimator provides an upwards bias (similar to survival data with competing risks)