Intensity models for illness-death and recurrent events

(Sections 2.5, 3.9, 5.1, 5.2, 5.4)

Per Kragh Andersen

Section of Biostatistics, University of Copenhagen

DSBS Course
Survival Analysis in Clinical Trials
2025

Overview

- Examples of multi-state models; recurrent events
- Models for intensities
- Estimating transition and state occupation probabilities

Textbook:

'Models for Multi-State Survival Data: Rates, Risks, and Pseudo-Values'
PKA & HR (2023), Chapman & Hall/CRC.

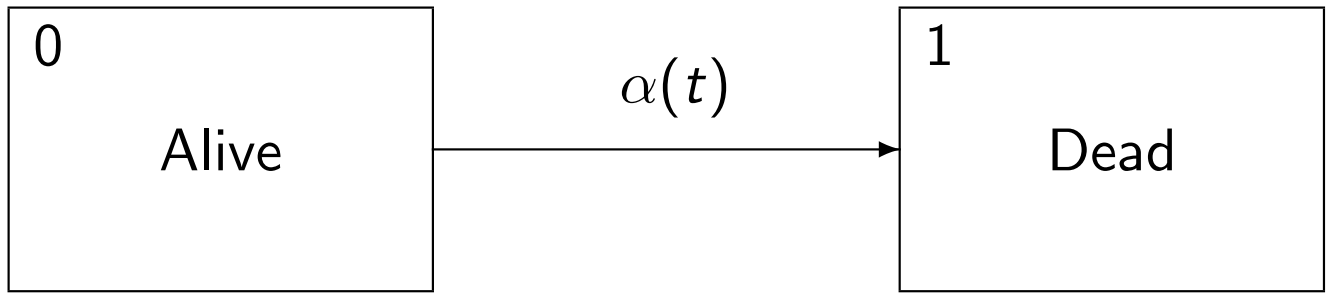
Companion web pages:

https://multi-state-book.github.io/companion

include data sets for download and code for examples in both `R` and `SAS`.

Examples of multi-state models

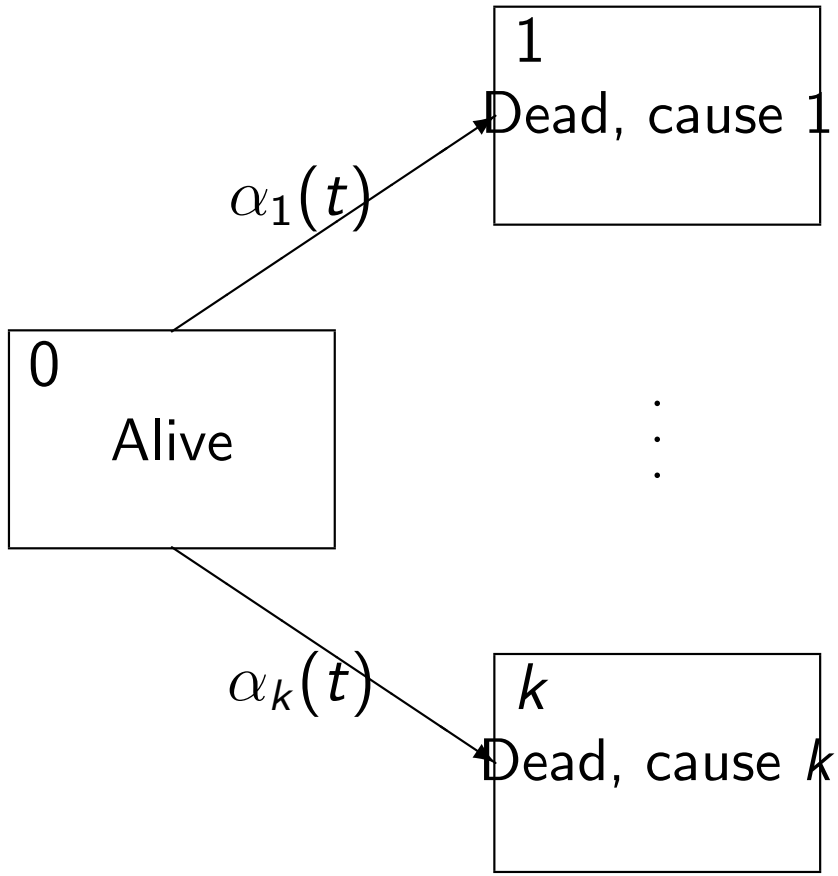
The two-state model for survival data



$$
\begin{aligned}
\alpha(t) &\approx P(\text{state 1 time } t+dt \mid \text{state 0 time } t)/dt \\
S(t) &= P(\text{state 0 time } t) \quad (= Q_0(t)) \\
F(t) &= 1 - S(t) = P(\text{state 1 time } t) \text{ is the cumulative} \\
&\quad \text{probability ('risk') of death over the interval from 0 to } t \\
&= 1 - \exp(-\int_0^t \alpha(u)du) \quad (= Q_1(t))
\end{aligned}
$$

Example: PBC3 trial, composite end-point – death or transplantation.

The competing risks multi-state model



Example: PBC3 trial transplantation and death without transplantation.

Basic parameters

Transition intensities $h = 1, ..., k$ ('cause-specific hazards'):

$$\alpha_h(t) \approx P(\text{state } h \text{ time } t + dt \mid \text{ state 0 time } t)/dt.$$

State occupation probabilities (marginal parameters):

1. Overall survival function:

$$\begin{aligned} S(t) &= P(\text{alive time } t) \quad (= Q_0(t)) \\ &= \exp\Big(-\int_0^t \sum_h \alpha_h(u)du\Big). \end{aligned}$$

2. Cumulative incidences $h = 1, ..., k$:

$$\begin{aligned} F_h(t) &= P(\text{dead from cause } h \text{ before time } t) \\ &= \int_0^t S(u)\alpha_h(u)du \quad (= Q_h(t)). \end{aligned}$$

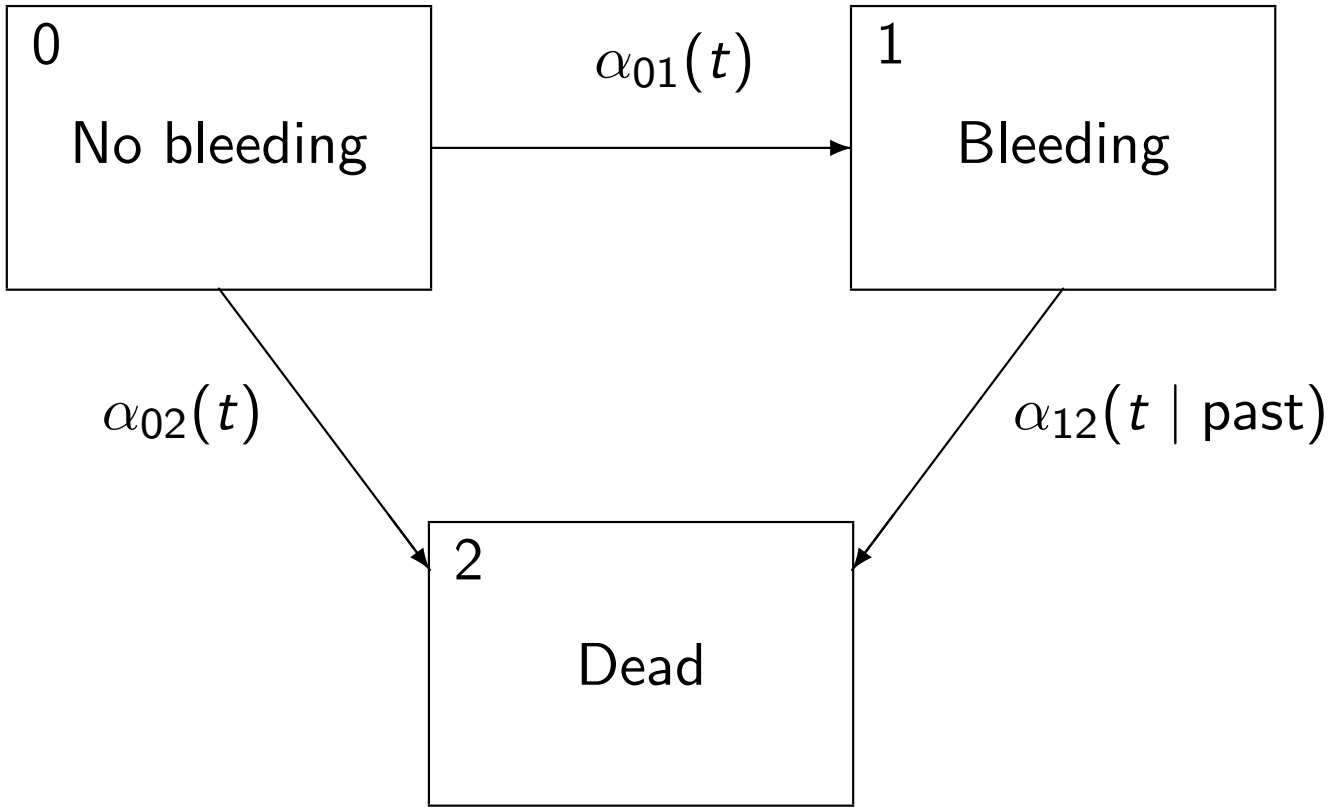
Example 1.1.4: the PROVA trial

(PROVA study group, *Hepatology*, 1991; Andersen, Esbjerg and Sørensen, *Stat. in Med.*, 2000).

- 286 patients with liver cirrhosis and endoscope-verified oesophageal varices
- randomized in a 2 by 2 design to combinations of sclerotherapy (yes/no) and propranolol (yes/no)
- Primary outcomes: bleeding or death without bleeding, but death after bleeding was also of interest

Treatment	Patients	Bleedings	Deaths without bleeding	Deaths total
Sclerotherapy only	73	13	13	18
Propranolol only	68	12	5	11
Both	73	12	20	30
Neither	72	13	8	16
Total	286	50	46	75

An illness-death model for PROVA



Same type of model would be applicable in the PBC3 trial, if data on mortality after transplantation were available.

Illness-death model: parameters

Transition intensities:

$$\alpha_{01}(t) \approx P(\text{state 1 time } t + dt \mid \text{state 0 time } t)/dt,$$

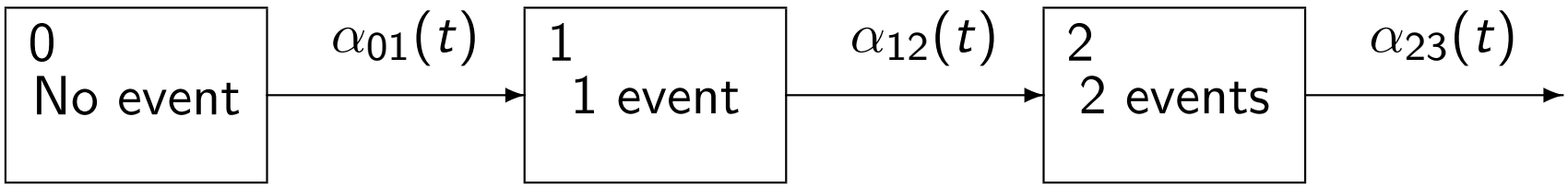
$$\alpha_{02}(t) \approx P(\text{state 2 time } t + dt \mid \text{state 0 time } t)/dt,$$

$$\alpha_{12}(t) \approx P(\text{state 2 time } t+dt \mid \text{state 1 time } t, \text{ past at time } t)/dt.$$

Note that the interpretation of these is identical to those in the simple (two-state or competing risks) multi-state models but that the intensity *out of the 'transient' state* 1 involves 'the past' at time t.
This could include

- Time spent in state 1 at time t (time-dependent)
- Time of entry into state 1 (time-fixed)

Recurrent events: mortality negligible, 'no duration'



Again, transition intensities could depend on the past. This leads to considering a single transition intensity function:

$$\alpha(t) \approx P(\text{event in } (t, t+dt) \mid \text{past})/dt,$$

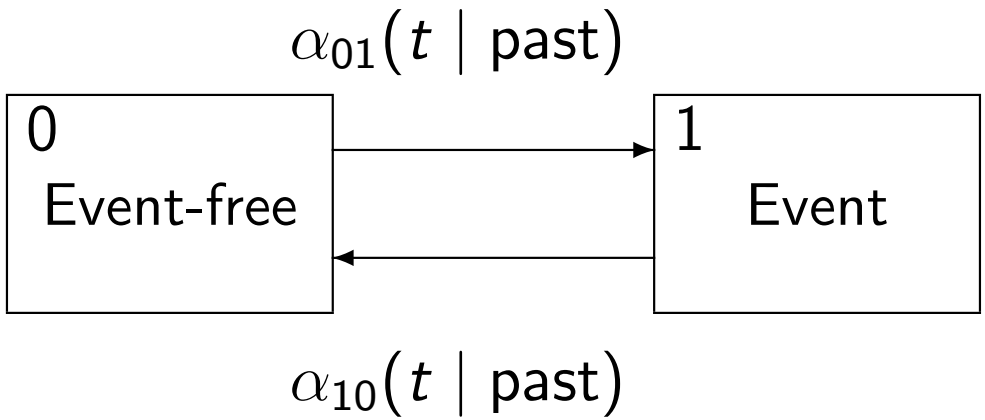
where the past would include information on previous events (i.e., in $(0, t)$), e.g., number and/or times of events.

Example: tumors in rats

Data from Cook and Lawless 2007 Springer-book "The Statistical Analysis of Recurrent Events" (Gail et al., 1980, *Biometrics*; Thompson et al., 1978, *Proc. Ann. Meet. Amer. Ass. Cancer Res.* **19**, 74).
76 female rats were exposed to a carcinogen and then given retinyl acetate to prevent cancer for 60 days. 48 rats, still tumor-free, were randomized to either continued treatment (23) or control (25) and followed for another 122 days. They were examined for tumors twice weekly and times of tumors were noted. The data set includes the variables:

- `id`
- `start`, `stop`, `status` (tumor or not)
- `num` (record no.)
- `trt` (treatment indicator)

Recurrent events 'with duration', mortality negligible



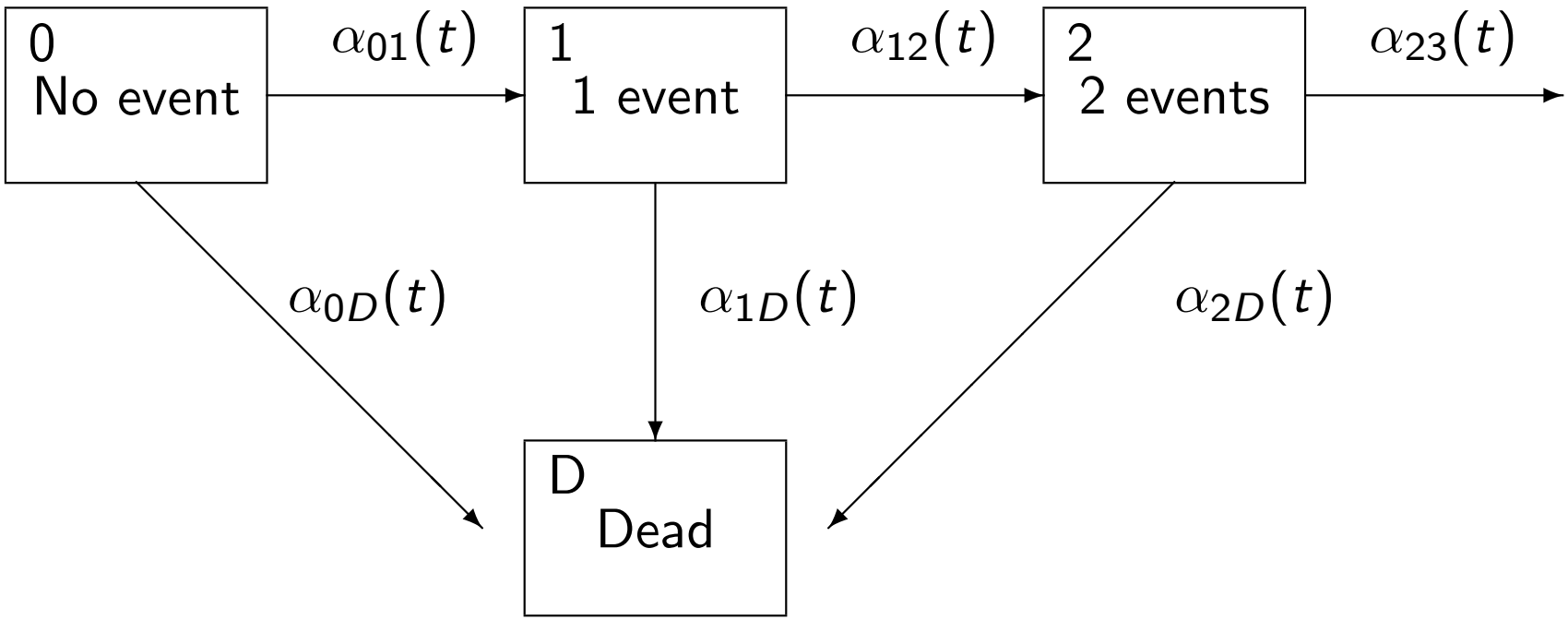
Here, there are intervals after occurrence of the event where the subject is not at risk for a new event. In both states there may be a past to consider when modeling the intensities.

Example: Pulmonary exacerbations

Data from Cook and Lawless (2007) book
(Therneau and Hamilton, 1997, *Statist. in Med.*; Fuchs et al., *NEJM*, 1994).
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.)

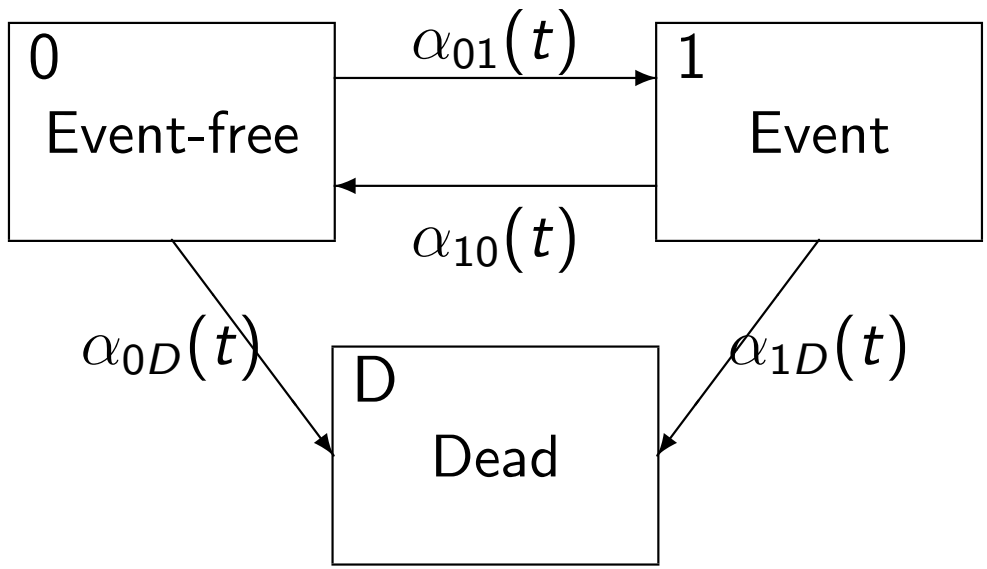
Recurrent events and mortality, 'no duration'



The LEADER trial

- Recurrent MI and all-cause mortality
- Recurrent MACE and non-CV mortality

Recurrent events 'with duration' and mortality



Again, all intensities may depend on the past.

Example 1.1.5: Psychiatric admissions

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

- 119 patients with unipolar ('depression') or bipolar ('manic') disorder; 78 died before 1985
- dates on admission to and discharge from psychiatric hospital
- covariates include
  - sex
  - first diagnosis (unipolar or bipolar)
  - age at first diagnosis
  - year at first diagnosis

Some times, times from admission to re-admission, i.e., disregarding that some time is spent in hospital are considered instead of times from discharge to re-admission. That brings the example into the framework of recurrent events with 'no duration' ('recurrent episodes').

A small sidetrack

It should be mentioned that one simple standard way out of a recurrent events problem is to ignore it (!) by restricting attention to the *first* occurrence of the event (maybe in competition with death), e.g., LEADER trial.

This is an *inefficient* solution since it disregards parts of the data that could contain important and useful information.

Simulation studies have been conducted that illustrate the loss of efficiency.

Models for intensities

Likelihood

A multi-state model, say $V(t)$ involves different *states*: $h = 1, ..., k$ and *types of direct transition*, $h \to j, h, j = 1, ..., k; h \neq j$.
The *counting process* $N_{hj}(t)$ counts the number of direct $h \to j$ transitions in $[0, t]$. Let $Y_h(t)$ be the number of subjects observed in state h at time $t-$. Then the *intensity process* for $N_{hj}(t)$ is:

$$E(dN_{hj}(t) \mid \text{past}) \approx \alpha_{hj}(t) Y_h(t) dt,$$

for some function $\alpha_{hj}(\cdot)$ of time and past.
According to *Jacod's formula*, the likelihood for the α's based on observation of $(N_{hj}(t), Y_h(t), 0 \leq t \leq \tau)$ (censoring allowed) is

$$L = \prod_t \prod_{h,j;h\neq j} \left(\alpha_{hj}(t) Y_h(t) dt\right)^{dN_{hj}(t)} \times \exp(-\sum_{h,j} \int_0^\tau \alpha_{hj}(u) Y_h(u) du)$$

$$= \prod_{h,j;h\neq j} \left(\prod_t \left(\alpha_{hj}(t) Y_h(t) dt\right)^{dN_{hj}(t)} \exp(-\int_0^\tau \alpha_{hj}(u) Y_h(u) du) \right).$$

Models for intensities

As we saw it for the special case of competing risks, this likelihood *factorizes* into a product over transitions.

As a consequence, if no α's have parameters in common then transition intensities may be modelled separately.
Models with shared parameters may be relevant, e.g., the mortality rates in the illness-death model may be *proportional* and/or risk factors may have same effects on $0 \to 2$ and $1 \to 2$ transitions.

We may therefore conclude that:

1. hazard models from simple survival analysis (and competing risks) are applicable, e.g., Cox models
2. modeling transitions out of non-initial transient states poses new challenges

Recurrent events, the PWP model

In all of the recurrent events scenarios studied (i.e., with or without duration and with or without mortality) intensity models may be used.
One option is to use separate models (possibly with shared covariate effects) for $\alpha_{h,h+1}(t), h = 0, 1, ...$

This is known as the *PWP model* (Prentice, Williams and Peterson, *Biometrika*, 1981):

$$\alpha_{h,h+1}(t \mid Z) = \alpha_{h0}(t) \exp(\beta^{\mathsf{T}} Z), h = 0, 1, ...$$

This is a Cox model with *time-dependent strata*.
The effect of Z might vary with the number of previous events, i.e., β_h instead of β.

Note that a PWP model may not be obvious for randomized studies because 'randomization is lost' after first event.

Recurrent events, the AG model

Assuming proportionality between the different $\alpha_{h0}(t)$, a special case of the *AG* model (Andersen and Gill, *Ann. Statist*, 1982) is obtained.
In this model there is a single event intensity:

$$\alpha(t \mid Z) = \alpha_0(t) \exp(\beta^{\mathsf{T}} Z(t))$$

that is allowed to depend on the past (e.g., number of previous events $N(t-)$) via *time-dependent covariates* which may also interact with other covariates.
The model may be fitted using delayed entry for later events.
If there are only time-fixed covariates, this is an *inhomogeneous Poisson process* which is *Markov*.

Recurrent events, gap time models

Gap time models are models where the baseline intensity depends, not on t, but on time $t - T_{N(t-)}$ since last event, e.g.,

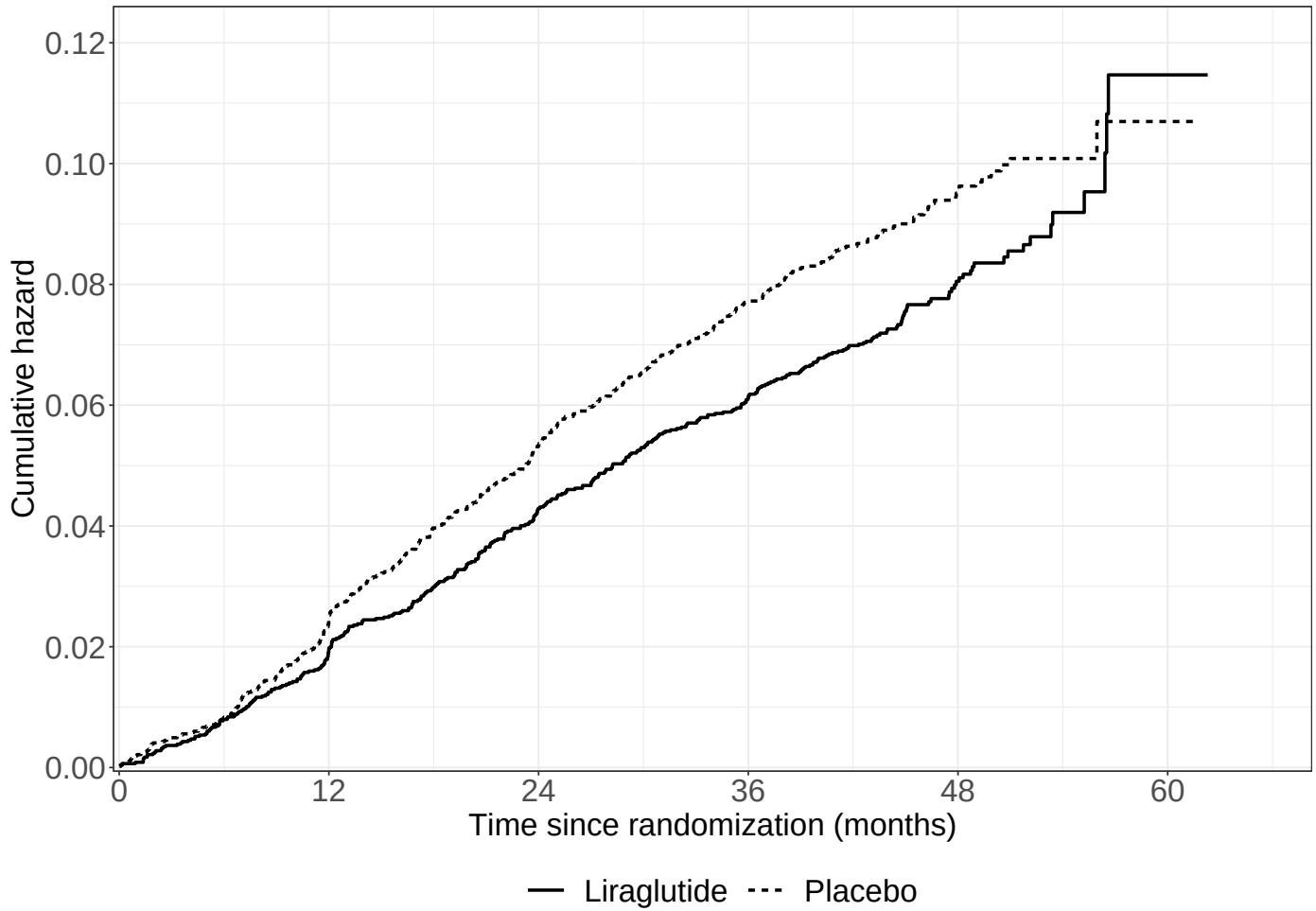
$$\alpha_{h,h+1}(t \mid Z) = \alpha_{h0}(t - T_h) \exp(\beta_h^\mathsf{T} Z).$$

If there are no covariates and successive gap times are i.i.d., then this is a *renewal process*.
These models are some times called *semi-Markov*.
To fit the model, no delayed entry is needed: for each transition $h \to h+1$, subjects are at risk from the 'new time zero'.

Putter et al. (*Statist. in Med.*, 2007), in their tutorial on multi-state models, called such models 'clock reset' (in contrast to Markov models which were called 'clock forward').

Note that gap time models may not be obvious for randomized studies because 'randomization is lost' after first event.

The LEADER trial, recurrent MI



Example: LEADER, recurrent MI

Estimated coefficients (and SD) from models for the hazard of recurrent myocardial infarctions for liraglutide vs. placebo.

Model		$\widehat{\beta}$	SD
Cox model	1^{st} event	-0.159	0.080
AG model	Cox type	-0.164	0.072
	Piece-wise constant	-0.164	0.072
PWP model	2^{nd} event	-0.047	0.197
	3^{rd} event	-0.023	0.400
	4^{th} event	0.629	0.737
	5^{th} event	-0.429	1.230
	All events	-0.130	0.072

R code (examples)

```
library(survival)
NAafit <- survfit(Surv(start, stop, mistatus == 1) ~
treat, data = leader_mi)

plot(NAafit, cumhaz=TRUE)

agfit <- coxph(Surv(start, stop, mistatus == 1) ~ treat,
data = leader_mi)


pwpfit <- coxph(Surv(start, stop, mistatus == 1) ~
treat + strata(eventno),  data = leader_mi)
```

Random effects ('frailty') models

Intra-individual correlation may be modeled using *time-dependent covariates* as in the AG model.
Another model allowing for intra-individual correlation is the *frailty model*:

$$\alpha_i(t \mid Z_i, A_i = a_i) = a_i \cdot \alpha_0(t) \exp(\beta^{\mathsf{T}} Z_i(t))$$

- (Alternatively: gap time model with $\alpha_0(w), \quad w = t - T_{ih}$)
- $Z_i(t)$: explanatory variables including treatment and number of previous episodes
- A_i: random, unobserved *frailty* assumed to follow some distribution with mean 1 and SD σ across the patient population, e.g., the gamma distribution.

Note that the results from a frailty model have *subject-specific* interpretations, i.e., hazard ratios *for given frailty* – this may or may not be what you want!

Joint frailty models

The frailty accounts for unobserved risk factors in the recurrent events process, and inference builds on integrating the frailty out of the likelihood.

If there is a non-negligible mortality and if mortality rates depend on the same frailty then a *joint frailty model* for both the recurrent events process and the mortality rate is needed.

Rondeau et al. (*Biostatistics*, 2007) developed such a model which is implemented in the `R` package `frailtypack`.

See also Liu et al. (*Biometrics*, 2004) and Huang and Wang (*JASA*, 2004).

Joint frailty models

The frailties $A_1, ..., A_n$ are assumed to be i.i.d. gamma variates with mean 1 and SD σ. The model is then:

$$\alpha_i(t \mid Z_i, A_i = a_i) = a_i \cdot \alpha_0(t) \exp(\beta_1^{\mathsf{T}} Z_i(t))$$

for the recurrent events intensity and

$$\alpha_{Di}(t \mid Z_i, A_i = a_i) = a_i^{\gamma} \alpha_{D0}(t) \exp(\beta_2^{\mathsf{T}} Z_i)$$

for the mortality rate. Thus, the frailty is shared but its effect on recurrent events and mortality could be different, modeled by the power γ. The sets of covariates could differ between the two models.

The likelihood is not extremely nice and the authors approximate the baseline intensities by cubic splines and use a penalized likelihood.

Joint frailty models

If there are intervals between successive episodes, then one must either

1. assume that their distribution is independent of the frailty A or
2. model the dependence, e.g., by using $\frac{1}{A}$ as the frailty for that transition

The second option is not available in `frailtypack`.

Example: The LEADER trial, gamma frailty models

Frailty models for recurrent myocardial infarctions (MI).

	Piece-wise constant		Cox-type	
	$\widehat{\beta}$	SD	$\widehat{\beta}$	SD
Liraglutide vs. placebo	-0.177	0.088	-0.177	0.088
Frailty SD	2.38		2.39	

Joint frailty model for recurrent myocardial infarctions (MI) and all-cause mortality – piece-wise constant baseline hazards.

	Recurrent MI		All-cause death	
	$\widehat{\beta}$	SD	$\widehat{\beta}$	SD
Liraglutide vs. placebo	-0.186	0.068	-0.211	0.078
Frailty SD ($\widehat{\sigma}$)	0.947			
Association ($\widehat{\gamma}$)	1.86			

Estimating transition and state occupation probabilities

Probabilities in multi-state models

We have seen that hazard models from survival and competing risks may also form the basis for modeling intensities in multi-state models (albeit with some new challenges).

When studying survival analysis and competing risks we also saw that getting *state occupation probabilities* $Q_h(t) = P(V(t) = h)$ depended on the structure of the model (competing risks or not).

This is still the case for general multi-state models and even though a specification of all intensities specifies the likelihood and, thereby, the whole probability distribution for the multi-state process, there are some such models where no simple plug-in techniques are available.

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

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

Markov processes

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 current state h (and on time t).
Assume that there are k states and that the transition intensities and cumulative intensities are, respectively, $\alpha_{hj}(t)$ and $A_{hj}(t)$ for $h, j = 1, \ldots k$. (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 $A_{hh}(t) = -\sum_{j \neq h} A_{hj}(t)$.

Markov processes: The Aalen-Johansen estimator

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

$$\widehat{A}_{hj}(t) \int_0^t \frac{I(Y_h(u) > 0)}{Y_h(u)} dN_{hj}(u).$$

This means that we can also easily estimate $\mathbf{P}(s,t)$ by *plug-in*:

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

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

The Aalen-Johansen estimator

The calculations have been implemented in the `R` package `mstate`.

The `mstate` package can also take results, $\widehat{A}_{hj}(\cdot \mid Z)$, e.g., from fitted Cox models, as the input and compute the product-integrals for given covariates.

That package works, in particular, for the competing risks model for which, alternatively, the `survival` package is applicable as demonstrated both in Part I and later today.

State occupation probabilities

The (marginal) 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), even for *non-Markov* processes the Aalen-Johansen estimator consistently estimates $Q_h(t)$.

State occupation probabilities

For certain *transient* states, an alternative and simple estimator for the state occupation probability is available (Pepe, *JASA*, 1991): 'the difference between Kaplan-Meier's estimator'.
For the illness-death model:

1. Compute the Kaplan-Meier estimator, say $\widehat{S}_0(t)$ for the initial state 0. (It estimates $Q_0(t)$.)
2. Compute the Kaplan-Meier estimator, say $\widehat{S}_{01}(t)$ for survival. (It estimates $1 - Q_2(t) = Q_0(t) + Q_1(t)$.)
3. Compute the difference $\widehat{S}_{01}(t) - \widehat{S}_0(t)$. (It estimates $Q_1(t)$.)

This idea can be used more generally for transient states in other multi-state models and often it gives curves similar to Aalen-Johansen estimates (e.g., BMT data):

Transition probabilities in non-Markov processes

Methods are developing for estimating transition probabilities in non-Markov processes:

- For the illness-death model, Meira-Machado et al. (*LIDA*, 2006) proposed an estimator based on 'weighted Kaplan-Meier integrals'. This is implemented in an `R` package `TPmsm`.
- Titman (*Biometrics*, 2015) proposed estimators for general non-Markov processes based on either the Pepe idea or on an idea of Allignol et al. (*LIDA*, 2014) using 'competing risks methods for the illness-death model'.
- Putter and Spitoni (*Stat. Meth. Med. Res.*, 2018) used the Datta-Satten idea to those observed in state h at time s to estimate $P_{hj}(s,t)$, for $t > s$ ('landmarking').

However, we are now on the edge of what is in general use and implemented.

Micro-simulation

A technique that *is used* in practice, e.g., in demography, is *micro-simulation*.

If all intensities in the multi-state model are specified, then it is possible to generate a large number of realizations of the process and to estimate, e.g., transition probabilities as simple averages taken over the repeated realizations (Sect. 5.4).

Summary

- The transition intensities are the basic building blocks in multi-state models:
  - The likelihood is given via the intensities
  - Micro-simulation is based on the intensities
- In illness-death, recurrent events, and other multi-state models, modeling intensities out of non-initial states poses new challenges
- Intra-individual correlation in recurrent events models may be modeled using time-dependent covariates or random effects (frailties)
- For Markov processes, transition probabilities are (rather) explicit functions of intensities and may be estimated using the Aalen-Johansen estimator
- The Aalen-Johansen estimator also works for estimation of state occupation probabilities in non-Markov multi-state models