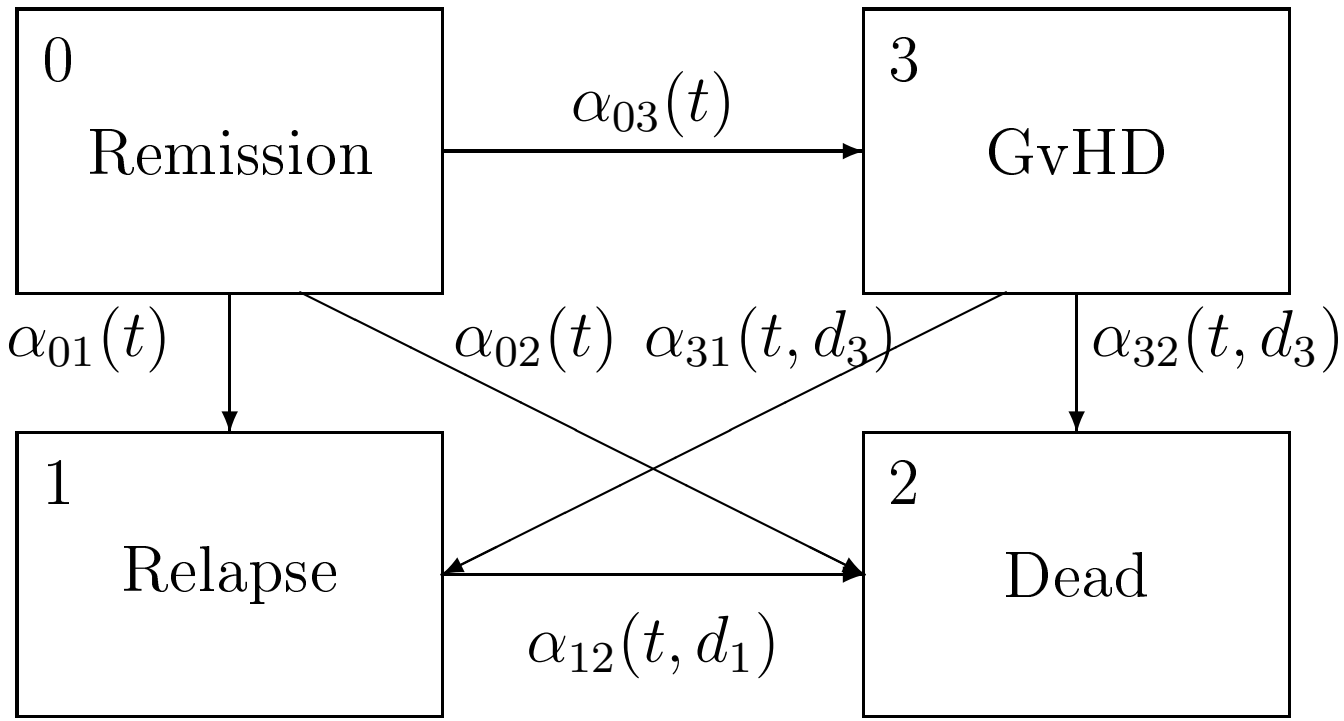
Multi-state models - an overview.

Ljubljana, June 2008

Per Kragh Andersen, Dept. Biostatistics, Univ. Copenhagen
Joint work with Maja Pohar Perme
Paper to appear (?) in *Lifetime Data Analysis.*

Motivation: Bone marrow transplantation.

- BMT is an effective therapy in, e.g. acute and chronic leukemia
- High-dose chemotherapy brings the patient in remission
- Bone marrow from a donor is infused while the patient is in remission
- In the disease course after transplantation, a number of events may occur, including:
  - Graft versus host disease (GvHD) - an immune reaction to the infused bone marrow cells
  - Relapse
  - Death



Figur 1: An extended model for BMT.

Tabel 1: Description of data from 2009 patients who underwent bone marrow transplantation.

Age (mean, SD)	31.9	15.4
Female sex	896	44.6%
Disease type AML	1406	70.0%
Graft type BM only	1153	57.4%
Karnofsky score (mean, SD)	91.3	9.1 (25 missing values)
Relapse	259	12.9%
Death	737	36.7%
Death and relapse	232	89.6% of patients with relapse
AGvHD	590	29.4%
CGvHD	630	31.4%
A or C GvHD	989	49.2%
A or C GvHD and relapse	104	10.9% of patients with GvHD
A or C GvHD and death	400	40.4% of patients with GvHD

Notation

A multi-state model is a model for a stochastic process, $X(t)$, with finite state space, S;

$X_i(t), i = 1, \ldots, n$ are studied (some times assumed to be i.i.d. copies);

$X_i(\cdot)$ is observed when the "censoring process", $C_i(t) = I(C_i \geq t)$, takes the value 1; (more general censoring is possible);

i.i.d. covariate processes $Z_i(t)$ may also be available;

The observed data $(X_i(t)C_i(t), Z_i(t)C_i(t), t \geq 0, i = 1, \ldots, n)$ generate a filtration $(\mathcal{F}_t)$.

Basic parameters.

- Transition probabilities:
  $P_{hj}(s,t \mid \mathcal{F}_{s-}) = \mathrm{pr}(X(t) = j \mid X(s-) = h, \mathcal{F}_{s-})$, $h, j \in S, s < t$
- Transition intensities $\alpha_{hj}(t \mid \mathcal{F}_{t-})$, $h, j \in S$ are given by:

$$\alpha_{hj}(t \mid \mathcal{F}_{t-}) = \lim_{\Delta t \to 0} P_{hj}(t, t + \Delta t \mid \mathcal{F}_{t-})/\Delta t,$$

- State occupation probabilities: $Q_h(t) = \mathrm{pr}(X(t) = h), h \in S$.

Models describe how these parameters depend on time, t, and the past, $\mathcal{F}_{t-}$, possibly via regression models including the covariates.

Statistical models for intensities.

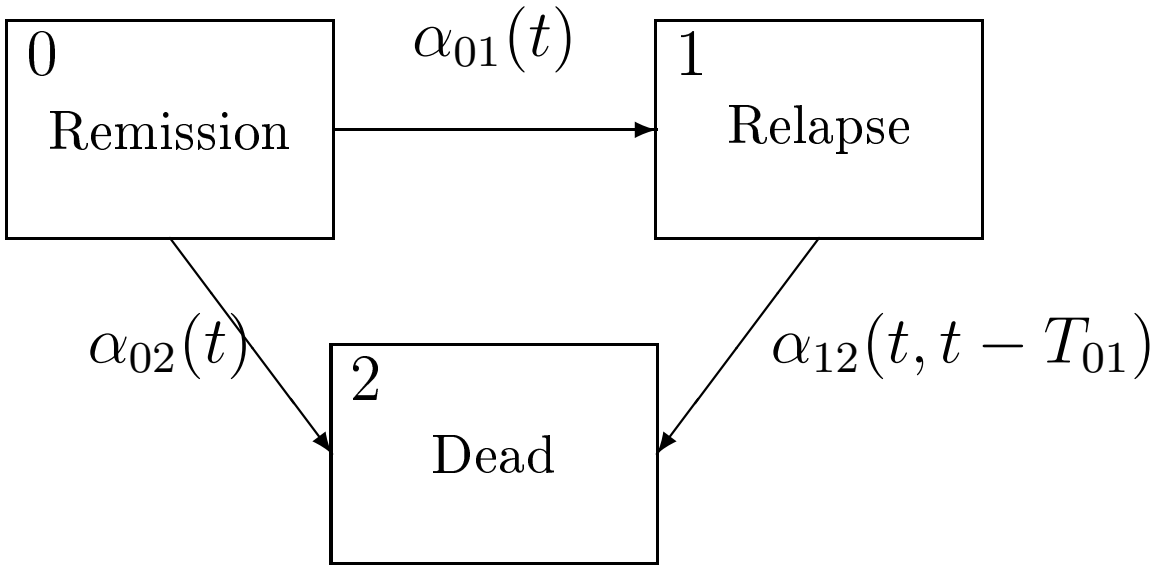
This is a very well developed area. Several hazard models, known from survival analysis, are available, including:

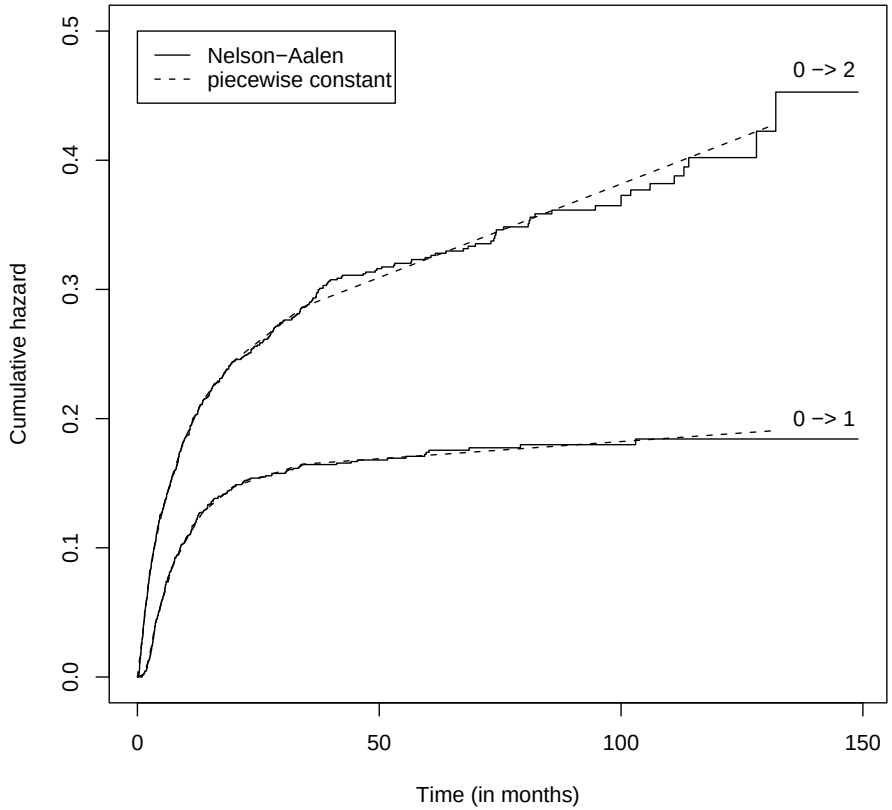
- Non-parametric model - the Nelson-Aalen estimator for the cumulative intensity: $\widehat{A}_{hj}(t) = \int_0^t \mathrm{d}N_{hj}(u)/Y_h(u)$.
- Various parametric models, e.g. $\alpha_{hj}(t \mid \mathcal{F}_{t-})$ constant or piecewise constant,
- Cox regression models: $\alpha_{hji}(t) = \alpha_{hj0}(t) \exp(\beta^\mathsf{T} Z_{hji}(t))$,
- Additive hazard models, flexible combinations of the two.

Markov models where $\alpha_{hj}(s, t \mid \mathcal{F}_{s-})$ only depends on $\mathcal{F}_{s-}$ via the state, $X(s-)$, occupied at s are particularly well developed.

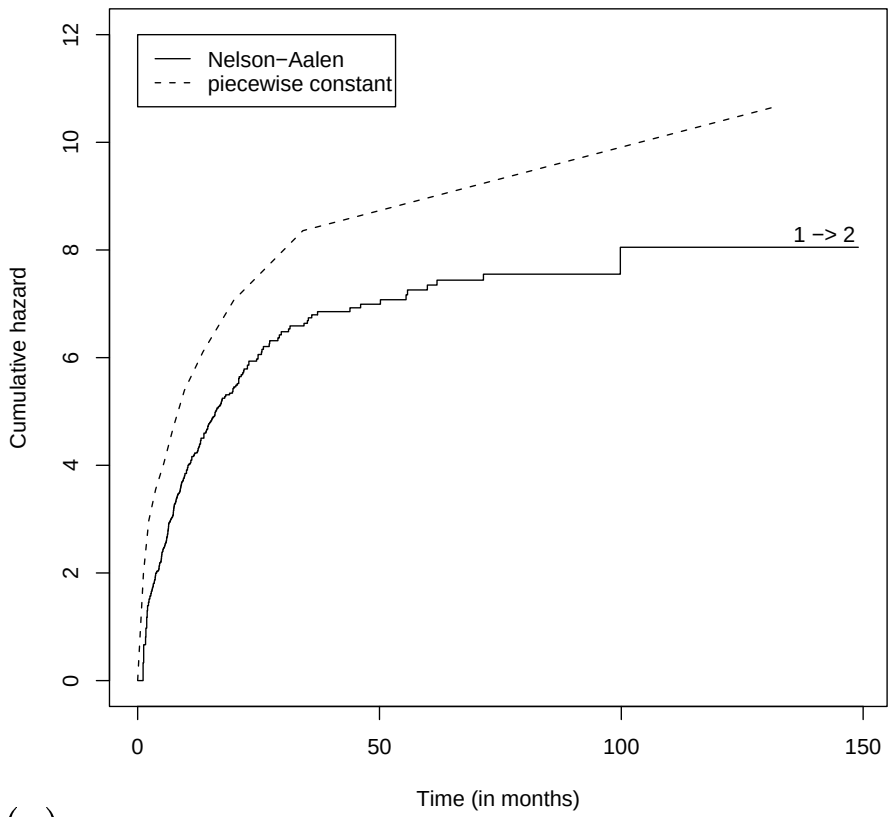
More general semi-Markov models may be studied using time-dependent covariates.

Illness-death model

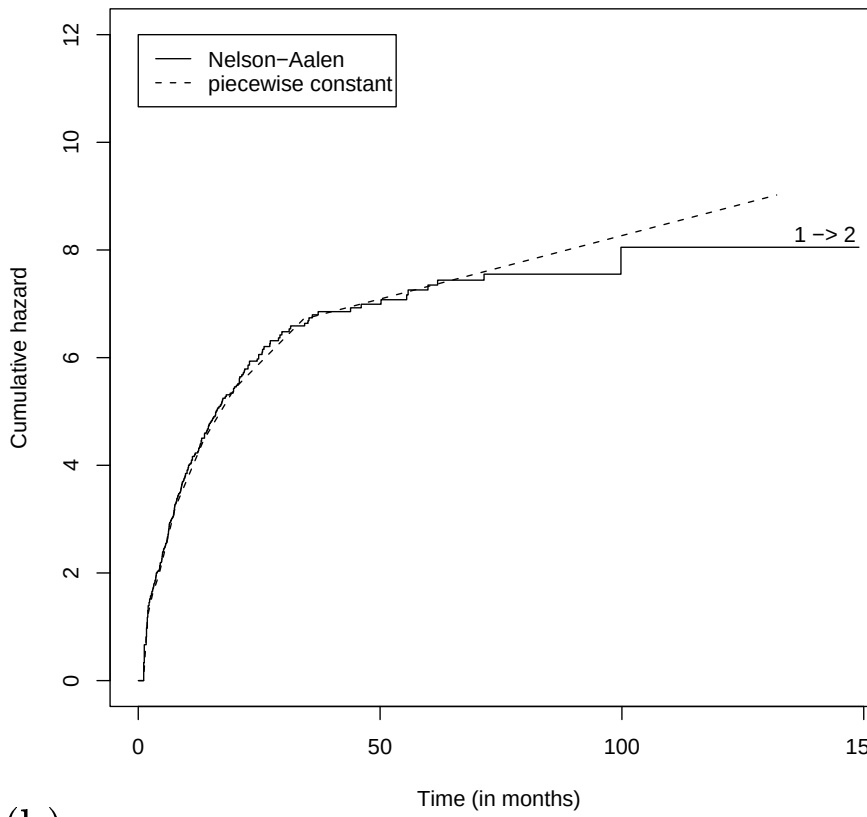




Figur 2: Nelson-Aalen estimates and estimates based on piecewise constant intensities for transitions out of state 0.



(a)



(b)

Figur 3: Nelson-Aalen estimates and estimates based on piecewise constant intensities for transitions out of state 1.Both figures use 10 intervals for the piecewise constant model, with case (b) having more cut-points at the very beginning of the follow-up time.

Tabel 2: Cox models for transition intensities in the illness-death model taking into account graft versus host disease as a time-dependent variable (combined acute and chronic - whichever happens first).

	$0 \rightarrow 1$			$0 \rightarrow 2$			$1 \rightarrow 2$		
	β	SE	p	β	SE	p	β	SE	p
disease (AML vs. ALL)	0.563	0.130	<0.001	0.334	0.098	<0.001	-0.271	0.142	0.055
age (per 10 years)	-0.040	0.045	0.370	0.263	0.033	<0.001	-0.002	0.050	0.970
graft type (BM only vs. other)	0.126	0.135	0.350	0.085	0.096	0.370	-0.046	0.172	0.790
GvHD	-0.184	0.134	0.170	1.040	0.098	<0.001	0.103	0.141	0.470

Tabel 3: Poisson models for transition intensities in the illness-death model taking into account graft versus host disease (combined). The follow-up time is split into 10 intervals.

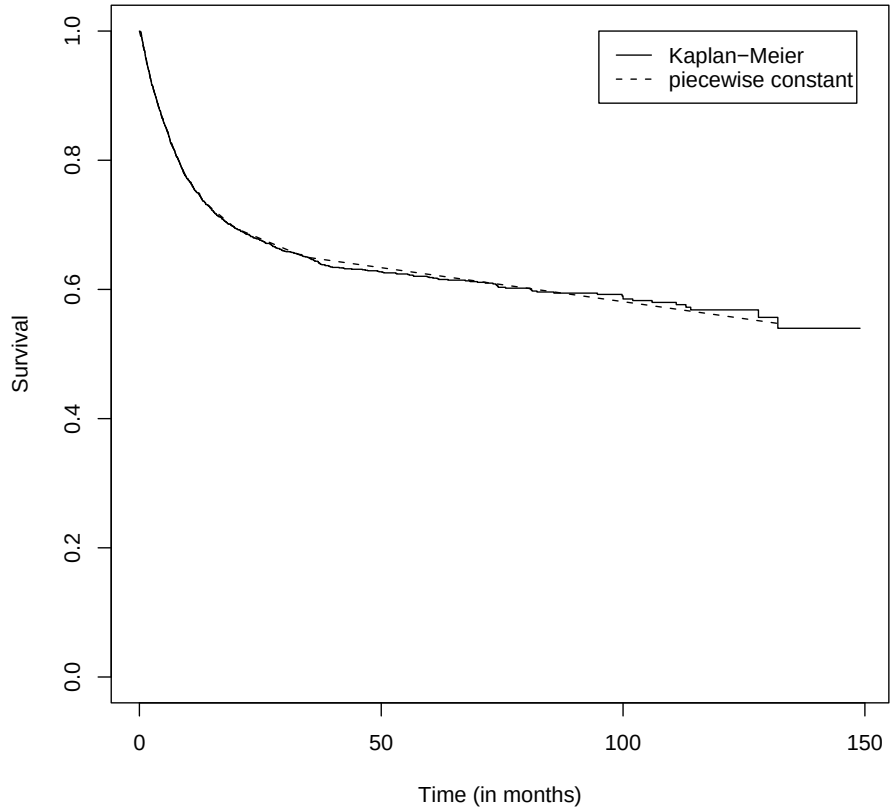
	$0 \rightarrow 1$			$0 \rightarrow 2$			$1 \rightarrow 2$		
	β	SE	p	β	SE	p	β	SE	p
disease (AML vs. ALL)	0.571	0.130	< 0.001	0.310	0.106	0.003	-0.264	0.142	0.0
age (per 10 years)	-0.035	0.045	0.431	0.257	0.035	< 0.001	-0.010	0.049	0.8
graft type (BM only vs. other)	0.131	0.136	0.335	0.191	0.102	0.061	-0.010	0.173	0.9
GvHD	-0.188	0.134	0.160	1.138	0.106	< 0.001	0.083	0.141	0.4

Statistical models for probabilities: 1 - No covariates, Markov - plug-in.

- Non-parametric models: Aalen-Johansen both for $P_{hj}(s,t)$ and for $Q_h(t)$: $\hat{\mathbf{P}}(s,t) = \prod_s^t \left(\mathbf{I} + \mathrm{d}\hat{\mathbf{A}}(u)\right)$.

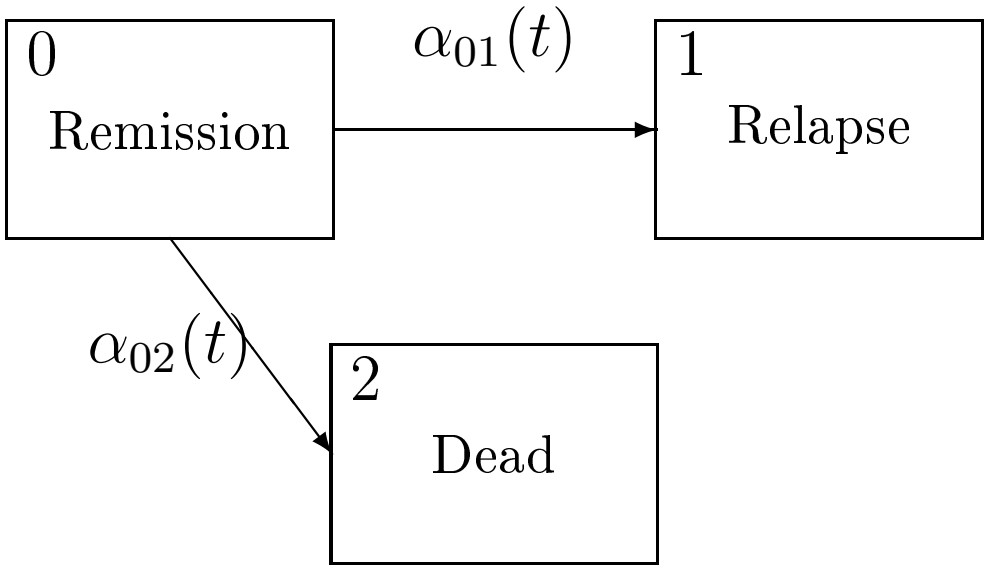
- Constant intensities: exponential formula for $P_{hj}(s,t)$ and thereby $Q_h(t)$, i.e. for $d = t - s$:
  $\mathbf{P}(d) = \exp(d\mathbf{A}) = \mathbf{V}\mathrm{diag}(e^{\rho d})\mathbf{V}^{-1}$ where ρ are the eigenvalues for the intensity matrix, $\mathbf{A}$, and $\mathbf{V}$ the matrix of eigenvectors

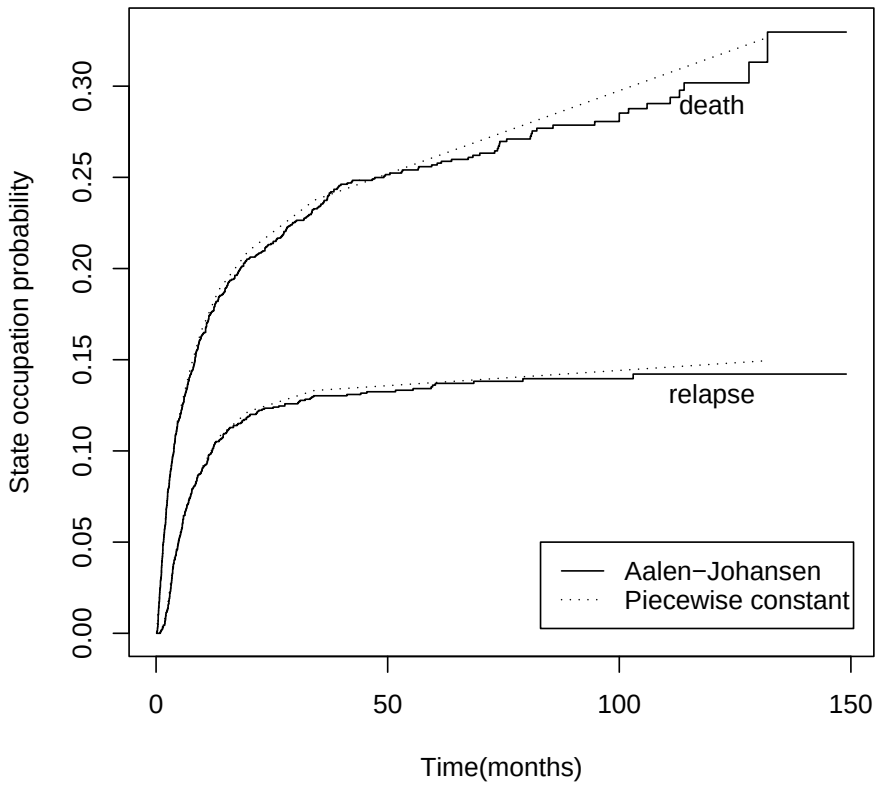
- Piecewise constant intensities: exponential formula for $P_{hj}(s,t)$ and Chapman-Kolmogorov equations:
  $P_{hj}(s,t) = \sum_{l \in S} P_{hl}(s,u) P_{lj}(u,t), s < u < t.$



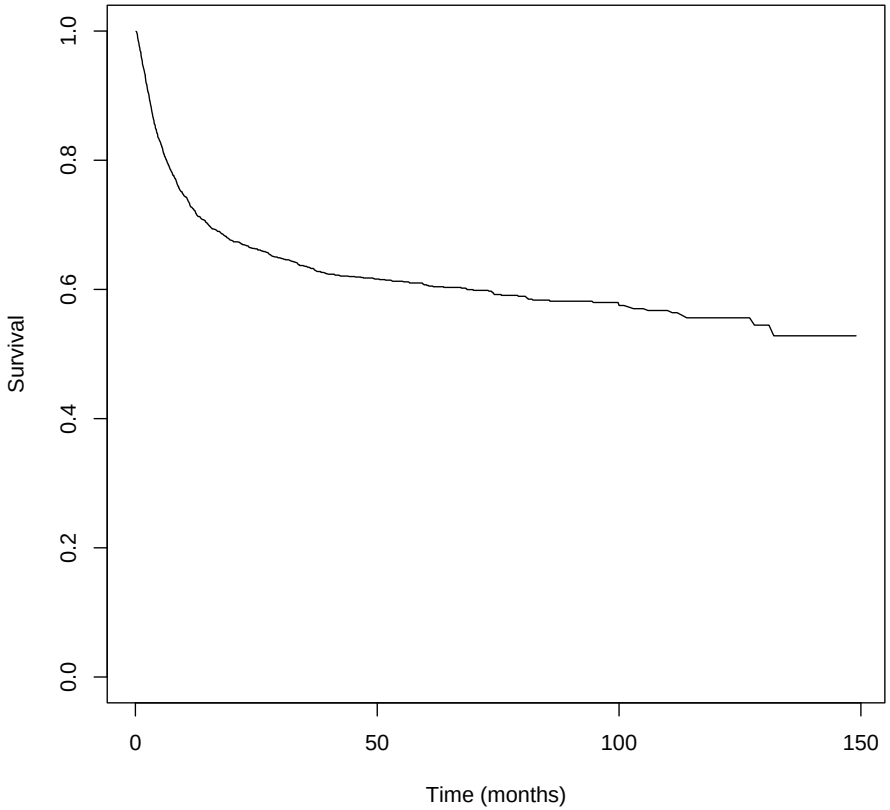
Figur 4: Kaplan-Meier estimate and estimate based on a piecewise constant hazard in the two-state model.

Competing risks model





Figur 5: Cumulative incidence estimates: Aalen-Johansen and piecewise constant hazards.



Figur 6: Aalen-Johansen (Kaplan-Meier) estimate for leukemia-free survival function.

Statistical models for probabilities: 2 - No covariates, non-Markov.

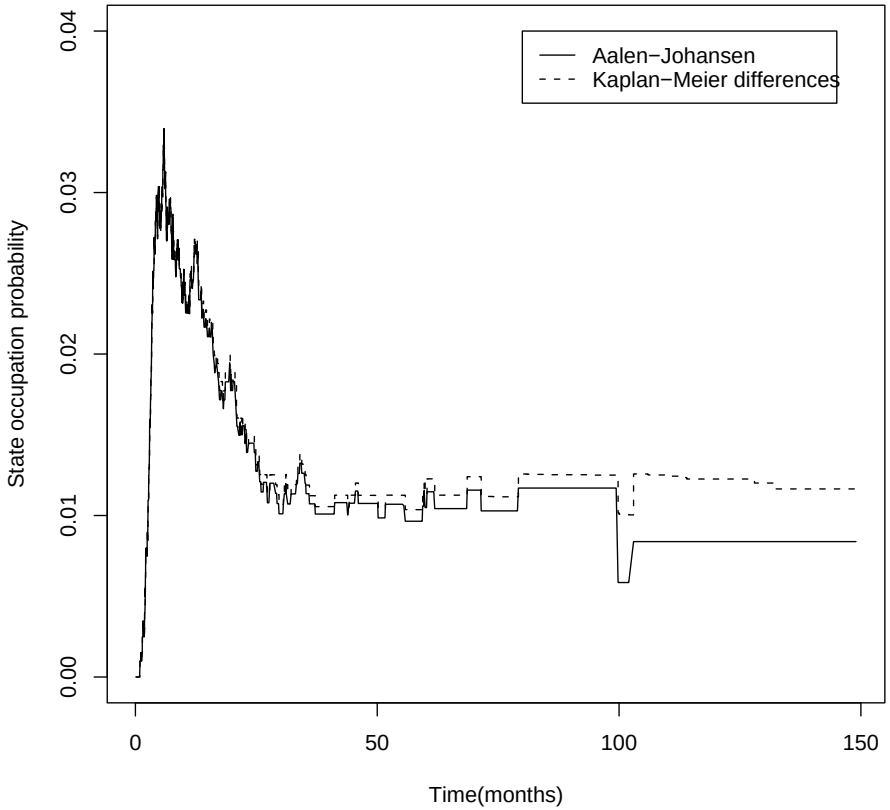
- For $Q_h(t)$, Datta & Satten (2001, *Stat. & Prob. Letters*) showed that Aalen-Johansen works.
  1. Nelson-Aalen: $\int_0^t \frac{\mathrm{d}N_{hj}(u)}{Y_h(u)}$ estimates $\int_0^t \tilde{\alpha}_{hj}(u)\mathrm{d}u$ where $\tilde{\alpha}_{hj}(t) = \lim_{\Delta t \to 0} \mathrm{pr}(X(t+\Delta t) = j \mid X(t-) = h)/\Delta t$,
  2. $\frac{1}{n}\mathbf{Y}(0)^{\mathsf{T}}\hat{\mathbf{P}}(0,t)$ estimates $\mathbf{Q}(t)$.
- For a transient state, Pepe (1991, *JASA*) estimated $Q_h(t)$ by Kaplan-Meier differences, e.g. illness-death model, T = time to death, S = time to illness, $1 \sim$ alive with illness

$$\hat{Q}_1(t) = \hat{S}_T(t) - \hat{S}_{S \wedge T}(t).$$

Example: testing the Markov assumption.

Add time-dependent covariate to model for $\alpha_{12}(\cdot)$:

- Cox model, coefficient for $I(t - T_{i1} > d_0)$ is -1.43, $P < 0.001$,
- Cox model, coefficient for T_{i1} is 0.07, $P < 0.001$,
- Poisson model, coefficient for $t - T_{i1}$ is -0.03, $P < 0.001$,



Figur 7: The Aalen-Johansen estimator and the Pepe Kaplan-Meier difference estimator for the state occupation probability $Q_1(t)$ in a three-state illness-death model.

Meira-Machado et al. estimator

For the illness-death model without recovery, Meira-Machado, de Una-Alvarez & Cadarso-Suarez (2006, *LIDA*), derived non-parametric estimators for $P_{hj}(s,t)$ using Kaplan-Meier integrals:

Let $\tilde{T}_i, \tilde{S}_i$ be the (possibly censored) times to death/illness and let w_i be the "Kaplan-Meier weight" for $\tilde{T}_{(i)}$, i.e.

$$w_i = \hat{P}(T = \tilde{T}_{(i)}) = \frac{D_{(i)}}{n-i+1}\hat{S}(\tilde{T}_{(i-1)}).$$

Then:

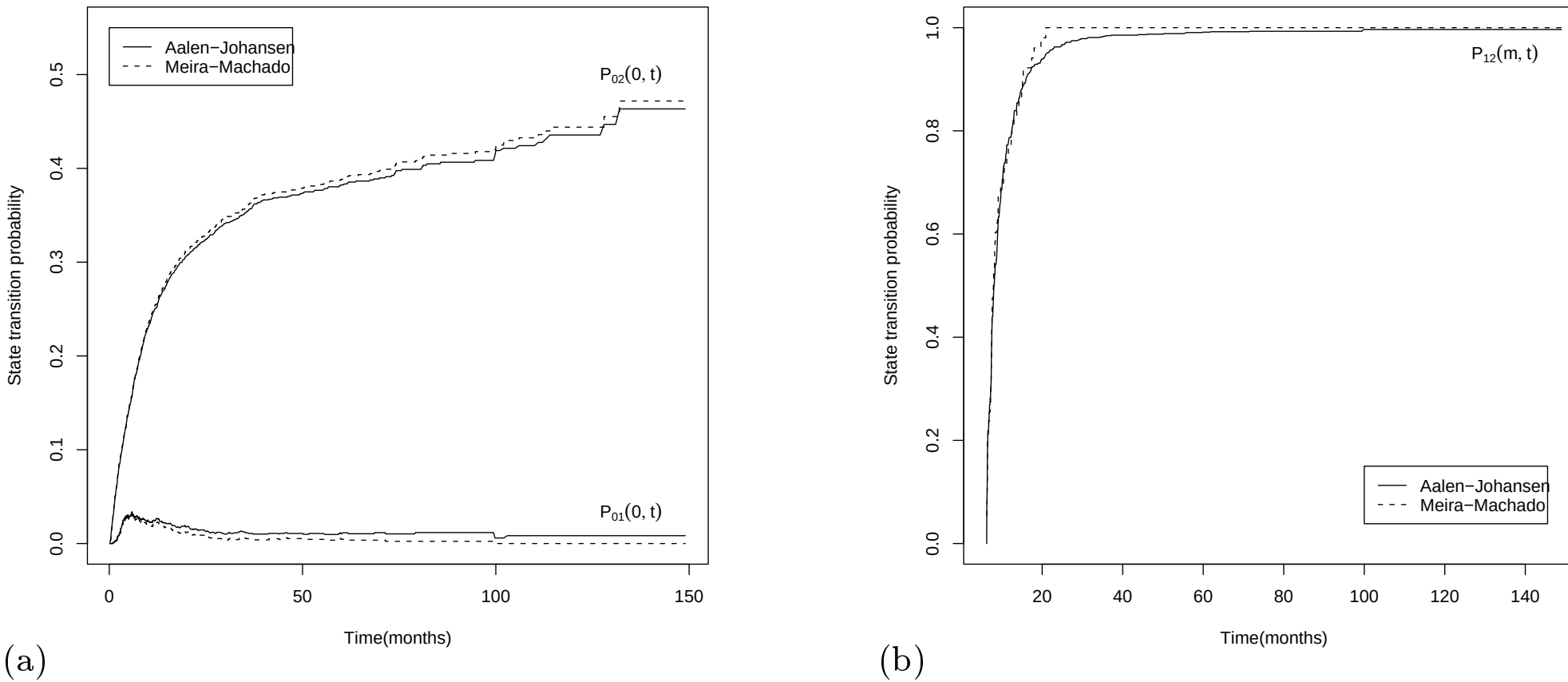
$$\hat{P}_{01}(s,t) = \frac{\sum_i w_i I(s < \tilde{T}_i \wedge \tilde{S}_i \leq t, \tilde{T}_i > t)}{\hat{S}_{T\wedge S}(s)}.$$

The numerator is a Stute (1996, *SJS*) Kaplan-Meier integral:

$$KMI = \frac{1}{n} \sum_i w_i \phi(\tilde{T}_i, Z_i)$$

.

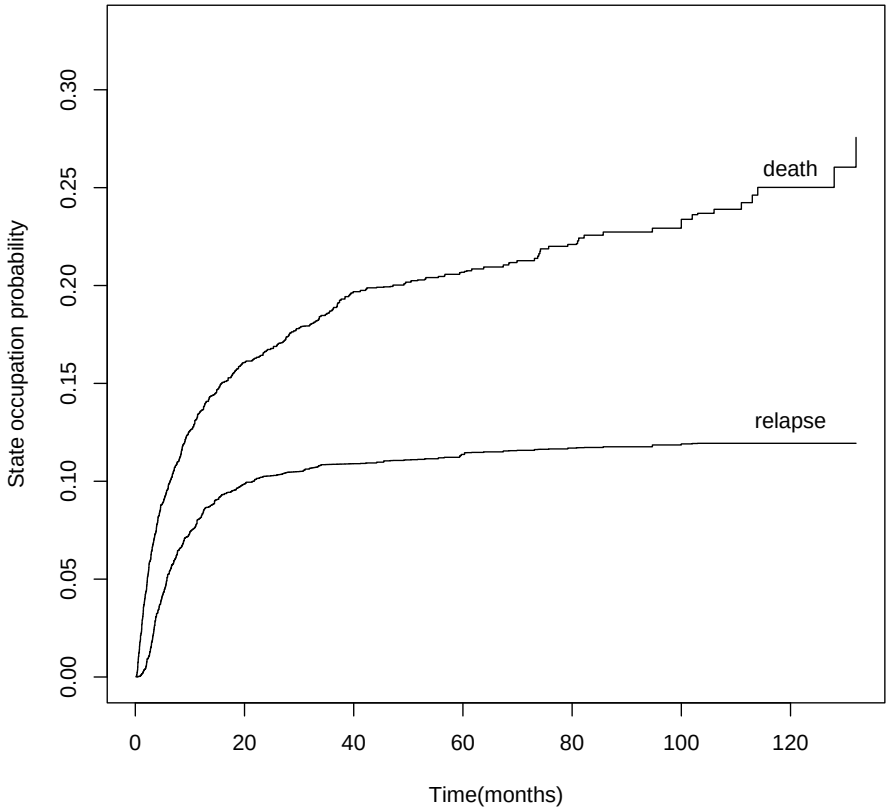
Also expressions for $\hat{P}_{00}(s,t)$ (simple!) and for $\hat{P}_{11}(s,t)$ (less simple) are presented.



Figur 8: Estimated transition probabilities in a three-state illness-death model using the Meira-Machado estimators and the Aalen-Johansen estimators: (a) $P_{01}(0,t)$, $P_{02}(0,t)$; (b) $P_{12}(6.22,t)$.

Statistical models for probabilities: 3 - Regression.

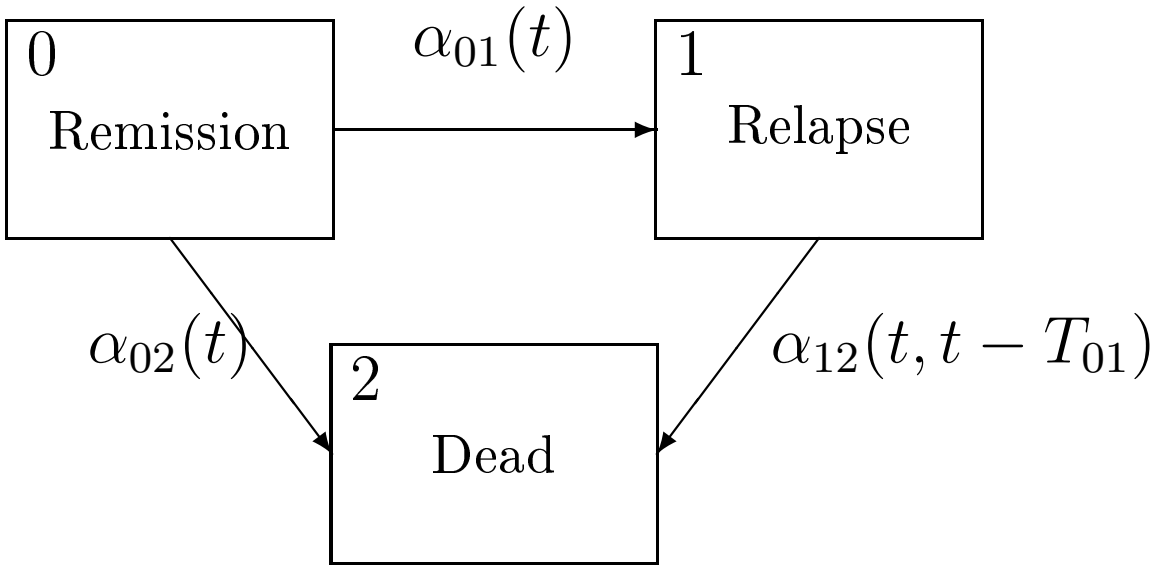
- Markov processes
  - Plug-in, i.e. $P_{hj}(s,t)$ and $Q_h(t)$ may be predicted for given (time-fixed) covariates;
    1. Andersen, Hansen & Keiding (1992, *SJS*): general Markov processes with Cox type intensities
    2. Cheng, Fine & Wei (1998, *Biometrics*): competing risks with Cox hazards
    3. Scheike and Zhang (2003, *Biometrics*): competing risks with flexible "Cox-Aalen" hazards
  - no simple relation between covariates and probabilities.



Figur 9: Cumulative incidence estimates for AML patient, 32 years, BM only.

- Non-Markov:
  - Plug-in for Datta-Satten and Pepe estimators of $Q_h(t)$, no simple relation between covariates and probabilities.
  - Generalize Meira-Machado et al.?
  - For some semi-Markov processes modelled using time-dependent covariates for duration effects, e.g. illness-death model without recovery, plug-in may be used.

Illness-death model



$$P_{01}(s,t) = \int_s^t P_{00}(s,u) \exp\left(-\int_u^t \alpha_{12}(x, x-u)\mathrm{d}x\right) \alpha_{01}(u)\mathrm{d}u.$$

Statistical models for probabilities: 4 - Direct regression models.

- Competing risks:
  - Fine & Gray (1999, *JASA*), cloglog model,
  - Fine (2001, *Biostatistics*), general links
  - (Sub-distribution) hazard for
    $T^* = T \cdot I(X(\infty) = h) + \infty \cdot I(X(\infty) \neq h)$ is
    $\tilde{\alpha}_{0h}(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log(1 - P_{0h}(0,t))$, model:
    $\log(\tilde{\alpha}_{0h}(t \mid Z) = \log(\tilde{\alpha}_{0h0}(t)) + \beta^{\mathsf{T}} Z$ where β is estimated by partial likelihood with no or known censoring and by an IPCW score equation with general censoring.

- General approaches:
  - pseudo-observations (Andersen, Klein et al., 2003, *Biometrika*; 2004, *LIDA*; 2005, *Biometrics*; 2007, *SJS*),
  - Direct binomial regression using inverse probability of censoring weights (Scheike & Zhang, 2007, *SJS*; Scheike, Zhang & Gerds, 2008, *Biometrika*).

Direct models for probabilities:

1. Pseudo-observations:

$\hat{\theta}(t)$ estimator based on entire sample, $\hat{\theta}_{-i}(t)$ estimator based on data obtained by deleting i.

Pseudo-observation no. i is then given by

$$\hat{\theta}_i(t) = n \cdot \hat{\theta}(t) - (n-1) \cdot \hat{\theta}_{-i}(t).$$

These are used as outcome variables in standard regression models using GEE and sandwich estimator for variances.

2: Direct binomial regression (first: competing risks model).

Without censoring, $I(X_i(t) = h)$ is observed and can be used as outcome variable.

With censoring, let
$V_i = T_i I(C_i > T_i, X_i(\infty) = h) + \infty(I(X_i(\infty) \neq h) + I(C_i < T_i))$; then

$$\mathrm{E}\left(\frac{I(V_i \leq t)}{S_C(V_i)}\right) = \mathrm{pr}(X_i(t) = h).$$

Then inverse probability of censoring weights can be used in the regression.

1. and 2. shown by Graw, Gerds & Schumacher (2007) to be equivalent (for competing risks model):
$\hat{\theta}_i(t) = \frac{I(X_i(t)=h, T_i \leq C_i)}{S_C(T_i)} + o_P(1)$.

Generalizability beyond competing risks: Scheike & Zhang (2007, *SJS*).

Tabel 4: Estimates in direct regression models for cumulative incidences.

	Fine & Gray			pseudo - cloglog			pseudo - logit		
	β	SE	p	β	SE	p	β	SE	p
relapse									
disease	0.472	0.131	<0.001	0.394	0.147	0.007	0.415	0.155	0.008
age	-0.093	0.045	0.040	-0.089	0.050	0.077	-0.095	0.053	0.074
graft type	0.057	0.125	0.650	0.093	0.140	0.509	0.098	0.148	0.509
death									
disease	0.337	0.097	0.001	0.363	0.121	0.003	0.409	0.134	0.002
age	0.282	0.032	<0.001	0.268	0.040	<0.001	0.297	0.045	<0.001
graft type	-0.085	0.090	0.343	-0.100	0.121	0.408	-0.111	0.134	0.409

Interval-censored data

All of the above methods assume continuous observation, i.e. times of transitions are observed exactly (if observed).

For interval-censored survival data, the survival function may be estimated non-parametrically by the Turnbull estimator.

For multi-state Markov models, except for a few special cases, only parametric inference (in particular for models with constant or piecewise constant intensities) seems to be available.

Perhaps, this is all that we can hope for?

Summary

What is available:

- Transition intensities, continuous observation - all you can fancy!
- State occupation probabilities, continuous observation - not bad!
- Transition probabilities, continuous observation, Markov processes, no covariates - all you can fancy!
- Transition probabilities, continuous observation, non-Markov processes, no covariates - not much, except illness-death model.
- Transition probabilities, continuous observation, Markov processes, regression - plug-in works but no simple parameters. Direct regression techniques are emerging.

- State occupation probabilities, continuous observation, regression - plug-in works but no simple parameters. Direct regression techniques are emerging.
- Transition probabilities, continuous observation, non-Markov processes, regression - not much.
- Interval-censored data, Markov - parametric inference (incl. regression) for intensities, otherwise not much.

Special methods are available for:

1. Competing risks,
2. Illness-death,
3. Recurrent events
4. "Forward" processes (incl. illness-death) using time-dependent covariates.