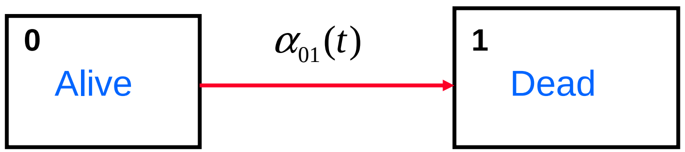
Markov chains

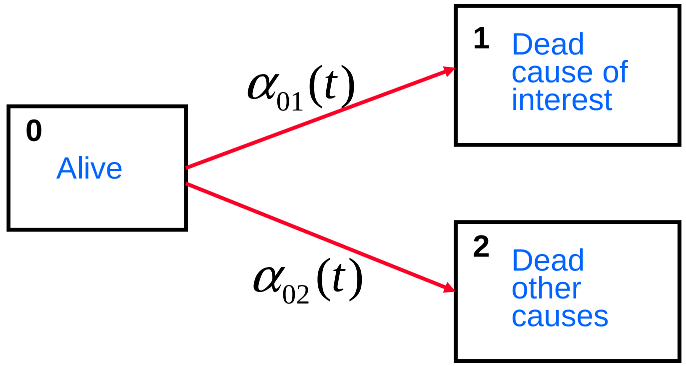
We may use Markov chains to model situations where more than one type of event may occur

We first note that the survival analysis situation may be modelled by a Markov process with two states:



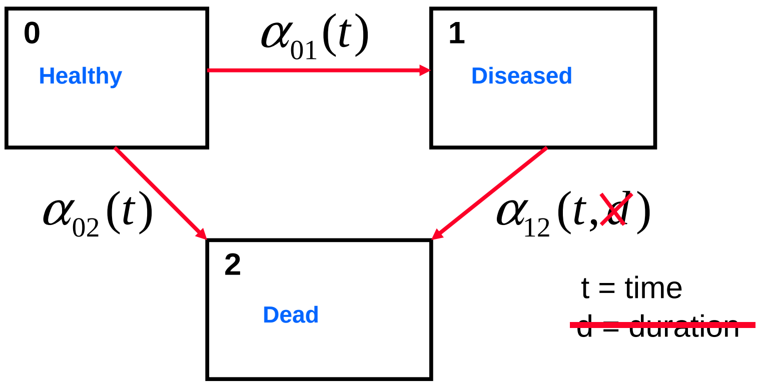
$\alpha_{01}(t)$ is the hazard rate or transition intensity

With two or more causes of failure we get a model for competing risks:



$\alpha_{01}(t)$ and $\alpha_{02}(t)$ are the cause specific hazards or transition intensities (i.e. instantaneous probabilities of a transition per unit of time).

An illness-death model without recovery:



We have a Markov process if the transition intensities do not depend on duration in a state

In general we consider a stochastic process *X*(*t*) with state space $\mathscr{S} = \{0, 1, 2, \ldots, k\}$

The process is a Markov chain (i.e. a Markov process with discrete state space) if future transitions only depend on the current state

May define transition probabilities

$$P_{gh}(s,t) = P(X(t) = h \mid X(s) = g) \qquad s < t, \quad g,h \in \mathscr{S}$$

and transition intensities

$$\alpha_{gh}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(X(t+\Delta t) = h \mid X(t-) = g)$$

for $g \neq h$

Estimating cumulative transition intensities

Consider the transition from state g to state h in a Markov chain, and let $\alpha_{gh}(t)$ denote the corresponding transition intensity

Let $N_{gh}(t)$ count the number of observed g --> h transitions in a group of individuals, and let $Y_g(t)$ be the number of individuals observed in state g just before time t

The intensity process of $N_{gh}(t)$ has the multiplicative form:

$$\lambda_{gh}(t) = Y_g(t) \cdot \alpha_{gh}(t)$$

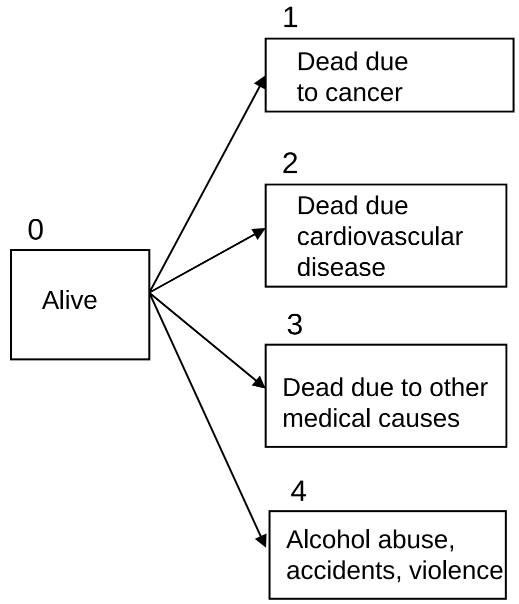
The cumulative transition intensity $A_{gh}(t) = \int_0^t \alpha_{gh}(u)du$ may be estimated by the Nelson-Aalen estimator

$$\hat{A}_{gh}(t) = \int_0^t \frac{J_g(u)}{Y_g(u)} dN_{gh}(u)$$

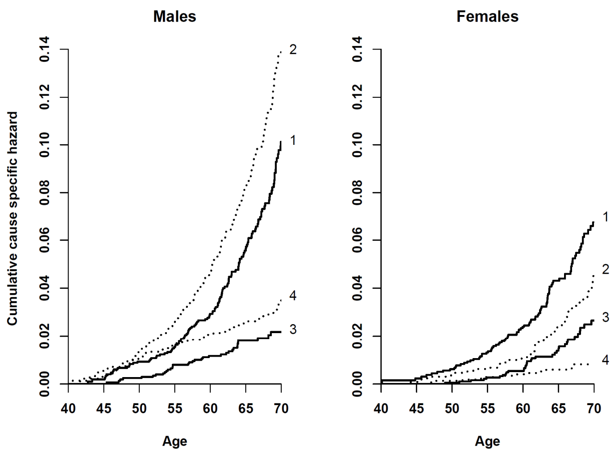
Examples 3.3 and 3.15: Competing causes of death

Data from the health screenings in three Norwegian counties 1974-78.

Followed-up to the end of 2000 by record linking to the cause of death registry at Statistics Norway.



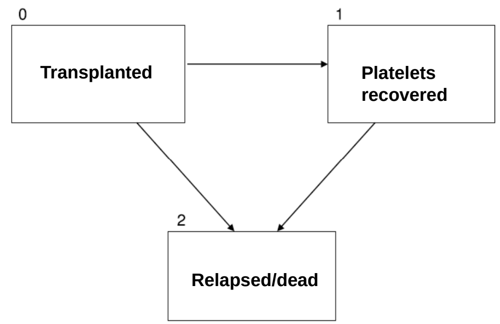
Nelson-Aalen estimates for the cause-specific mortality according to cause of death and sex (data from health screenings in three Norwegian counties):



1) Cancer
2) Cardiovascular disease
3) Other medical
4) Alcohol abuse, violence, accidents

Example 3.16: Platelet recovery, relapse and death for bone marrow transplant patients

137 patients with acute leukemia have had a bone marrow transplantation. Record the time of the events “platelet recovery” and “death/relapse”



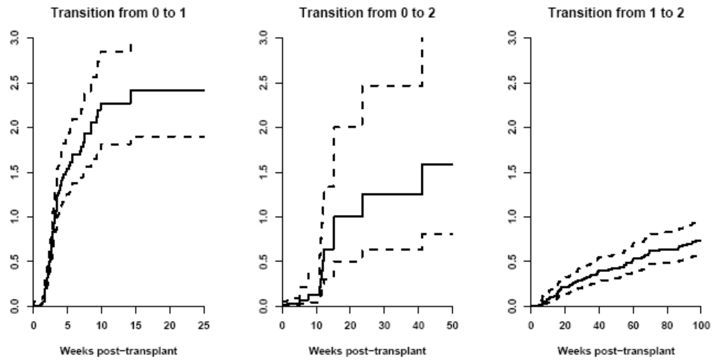


Fig. 3.20 *Nelson-Aalen estimates with log-transformed 95% confidence intervals of the cumulative transition intensities for the bone marrow transplant patients. The states are 0: "transplanted," 1: "platelet recovered," and 2: "relapsed or dead." Note that the time scale is not the same for the three estimates.*

Estimating transition probabilities

For simple Markov chains we have explicit expressions for the transition probabilities $P_{gh}(s,t)$ and these expressions may be used to suggest estimators

For a competing risks model with k causes of death :

$$P_{00}(s,t) = \exp\left(-\int_s^t \sum_{h=1}^{k} \alpha_{0h}(u)du\right)$$

$$P_{0h}(s,t) = \int_s^t P_{00}(s,u)\alpha_{0h}(u)du \qquad \text{(cumulative incidence function)}$$

This suggests the estimators:

$$\widehat{P}_{00}(s,t) = \prod_{s<T_j\le t}\left(1 - \frac{\triangle N_{0\bullet}(T_j)}{Y_0(T_j)}\right) \qquad \text{where } N_{0\bullet}(t) = \Sigma_{h=1}^{k} N_{0h}(t)$$

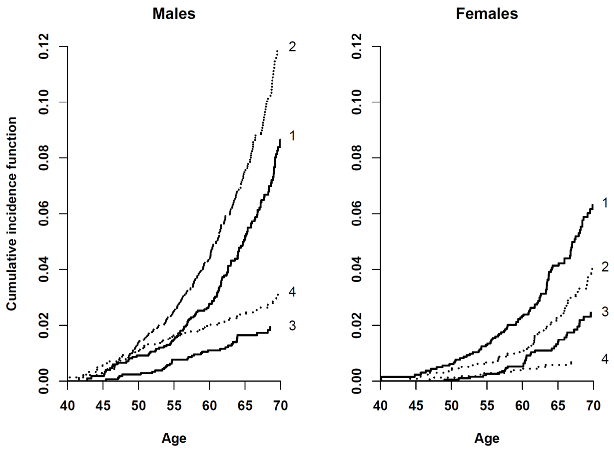
$$\widehat{P}_{0h}(s,t) = \sum_{s<T_j\le t} \widehat{P}_{00}(s,T_{j-1})\triangle\widehat{A}_{0h}(T_j)$$

Example 3.15: Causes of death in Norway

Estimates of

$P_{0h}(40,t)$

1) Cancer
2) Cardiovascular disease
3) Other medical
4) Alcohol abuse, violence, accidents



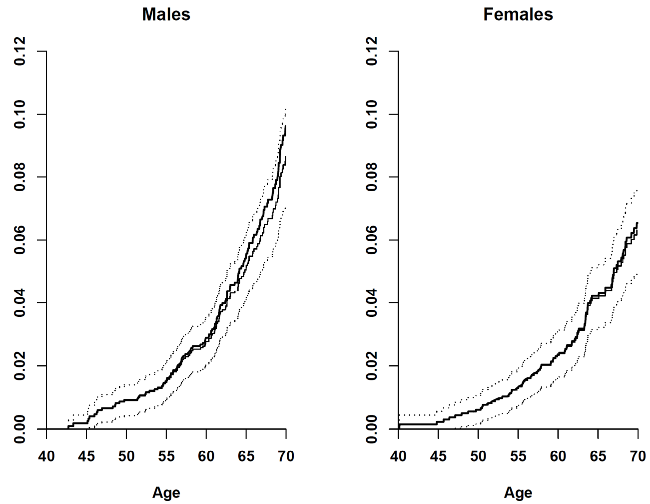
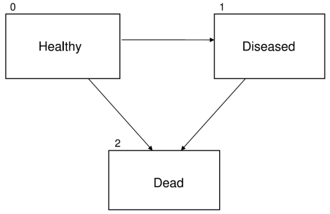


Fig. 3.18 *Empirical cumulative incidence functions for cancer death (thin drawn line) among middle-aged Norwegian males (left) and females (right) with 95% confidence intervals (dotted lines). Absolute risk estimates disregarding other causes of death are also given (thick drawn line).*

For a Markov illness-death model with no recovery we have :



$$P_{00}(s,t) = \exp\left\{-\int_s^t [\alpha_{01}(u) + \alpha_{02}(u)]\,du\right\}$$

$$P_{11}(s,t) = \exp\left(-\int_s^t \alpha_{12}(u)du\right)$$

$$P_{01}(s,t) = \int_s^t P_{00}(s,u)\alpha_{01}(u)P_{11}(u,t)du$$

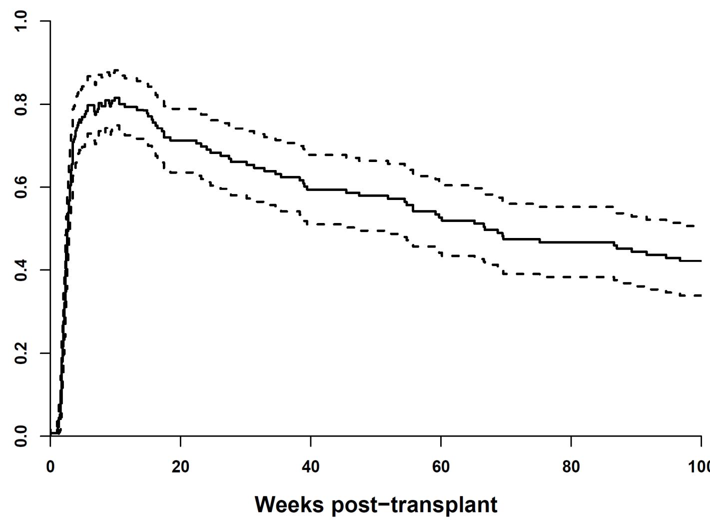
This suggests the estimators:

$$\widehat{P}_{00}(s,t) = \prod_{s<T_j\le t}\left(1 - \frac{\triangle N_{0.}(T_j)}{Y_0(T_j)}\right) \qquad \widehat{P}_{11}(s,t) = \prod_{s<T_j\le t}\left(1 - \frac{\triangle N_{12}(T_j)}{Y_1(T_j)}\right)$$

$$\widehat{P}_{01}(s,t) = \sum_{s<T_j\le t} \widehat{P}_{00}(s,T_{j-1})\triangle\widehat{A}_{01}(T_j)\widehat{P}_{11}(T_j,t)$$

Example 3.16: Bone marrow transplantation

Estimate of $P_{01}(0,t)$



We then consider a Markov chain with states 0, 1, …, k

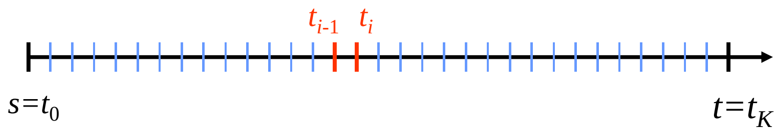
In general we cannot find an expression of the transition probabilities $P_{gh}(s,t)$ by means of the transition intensities $\alpha_{gh}(t)$

Consider the transition probability matrix:

$$\mathrm{P}(s,t) = \begin{pmatrix} P_{00}(s,t) & P_{01}(s,t) & .... & P_{0k}(s,t) \\ P_{10}(s,t) & P_{11}(s,t) & .... & P_{1k}(s,t) \\ .... & .... & .... & .... \\ P_{k0}(s,t) & P_{k1}(s,t) & .... & P_{kk}(s,t) \end{pmatrix}$$

We will see how this may be expressed as a (matrix valued) product-integral of the matrix of transition intensities

We partition $[s,t]$ into K small time intervals:



Then we may write:

$$\mathbf{P}(s,t) = \mathbf{P}(t_0,t_1) \times \mathbf{P}(t_1,t_2) \times \cdots \times \mathbf{P}(t_{K-1},t_K)$$

When K increases and the length of the intervals goes to zero, the product approaches a limit which is a matrix-valued product-integral

The product-integral may be expressed in terms of the matrix of transition intensities:

$$\boldsymbol{\alpha}(u) = \begin{pmatrix} \alpha_{00}(u) & \alpha_{01}(u) & .... & \alpha_{0k}(u) \\ \alpha_{10}(u) & \alpha_{11}(u) & .... & \alpha_{1k}(u) \\ .... & .... & .... & .... \\ \alpha_{k0}(u) & \alpha_{k1}(u) & .... & \alpha_{kk}(u) \end{pmatrix}$$

Here $\alpha_{gg}(u) = -\sum_{h \neq g} \alpha_{gh}(u)$ and $\alpha_{gh}(u) = 0$ if a transition from state *g* to state h may not occur

Now we have $\mathbf{P}(u, u+du) \approx \mathbf{I} + \boldsymbol{\alpha}(u)du$

Therefore the transition probability matrix may be written as a matrix-valued product integral:

$$\mathbf{P}(s,t) = \prod_{s \le u \le t} \{\mathbf{I} + \boldsymbol{\alpha}(u)du\}$$

Alternatively we may express the product integral in terms of the matrix

$$\mathbf{A}(t) = \{A_{gh}(t)\}$$

of cumulative transition intensities (with $A_{gg}(t) = -\sum_{h \neq g} A_{gh}(t)$):

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

From this we obtain the Aalen-Johansen estimator

$$\hat{\mathbf{P}}(s,t) = \prod_{s < u \le t} \left(\mathbf{I} + \Delta\hat{\mathbf{A}}(u)\right)$$

where $\hat{\mathbf{A}}(t) = \{\hat{A}_{gh}(t)\}$ the matrix of Nelson-Aalen estimators with $\hat{A}_{gg}(t) = -\sum_{h \neq g} \hat{A}_{gh}(t)$

The Aalen-Johansen estimator is a finite product over the observed transitions times

If e.g. *k*=2 and a 1–> 2 transition is observed at time *u*, we have:

$$\mathbf{I} + \Delta\hat{\mathbf{A}}(u) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 - \Delta\hat{A}_{12}(u) & \Delta\hat{A}_{12}(u) \\ 0 & 0 & 1 \end{pmatrix}$$

The estimators given earlier for the competing risks and illness-death models are special cases of the Aalen-Johansen estimator

The statistical properties of the estimator may be derived in a similar manner as for Kaplan-Meier, cf. sections 3.4.4 and 3.4.5 in the ABG-book (which are not part of the curriculum)