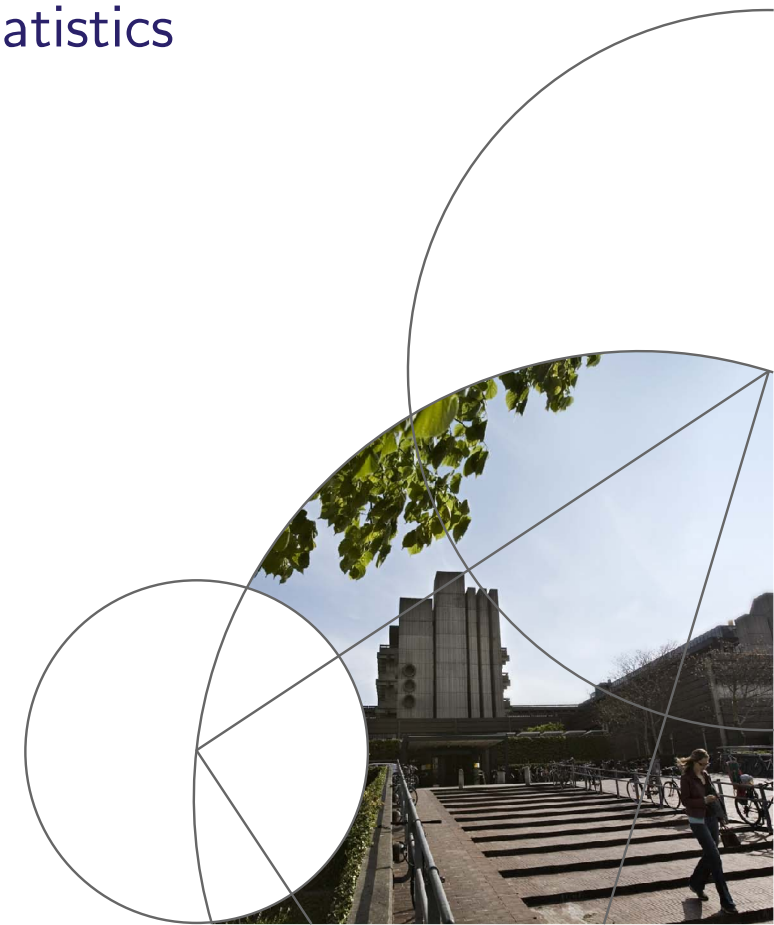
Survival Analysis for PhD students in Statistics

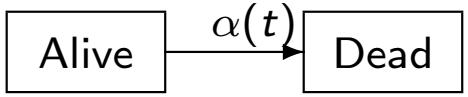
Day 3 2021: Competing risks

Section of Biostatistics



Recap: Survival data

Two-state life-death model: one type of event that eventually will happen to everyone



Let T^* be a continuous time to event

- Cumulative distribution function $F(t) = \text{pr}(T^* \leq t)$
- Survival function $S(t) = 1 - F(t) = \text{pr}(T^* > t)$
- Density $f(t) = \partial F(t)/\partial t$

The hazard rate $\alpha(t)$ is the conditional event rate at time t for those still alive at time t,

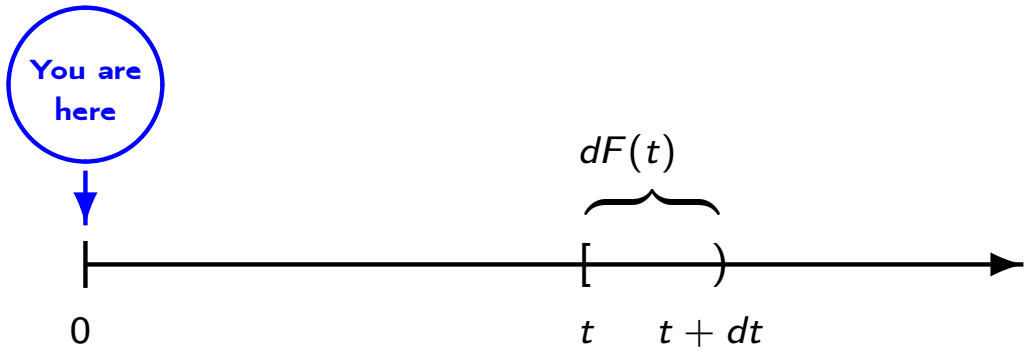
$$\begin{aligned}\alpha(t) &= \lim_{h \to \infty} \frac{\text{pr}(t \leq T^* < t + h | T^* \geq t)}{h} \\ &= \frac{\lim_{h \to \infty} \text{pr}(t \leq T^* < t + h)/h}{\text{pr}(T \geq t)} \\ &= \frac{f(t)}{S(t-)}\end{aligned}$$

Using Leibniz notation, for infinitely small dt, we write

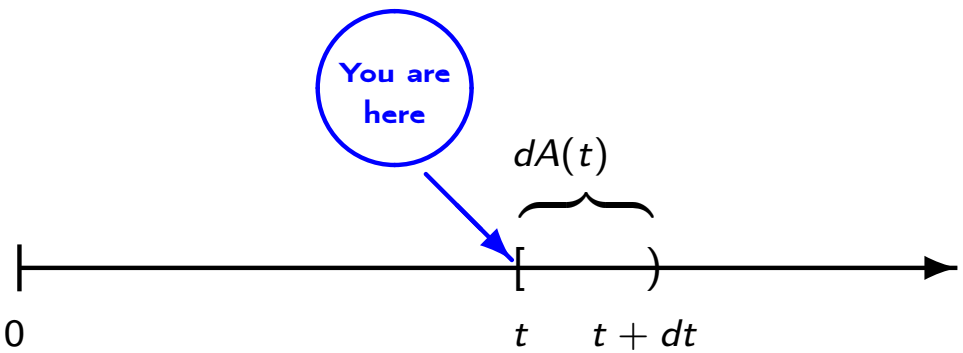
$$\begin{aligned}\alpha(t)dt = dA(t) &= \text{pr}(t \leq T^* < t + dt | T \geq t) \\ dF(t) &= \text{pr}(t \leq T^* < t + dt)\end{aligned}$$

Density vs. hazard

- The density is a marginal rate: $dF(t) = \text{pr}(t \leq T^* < t + dt)$



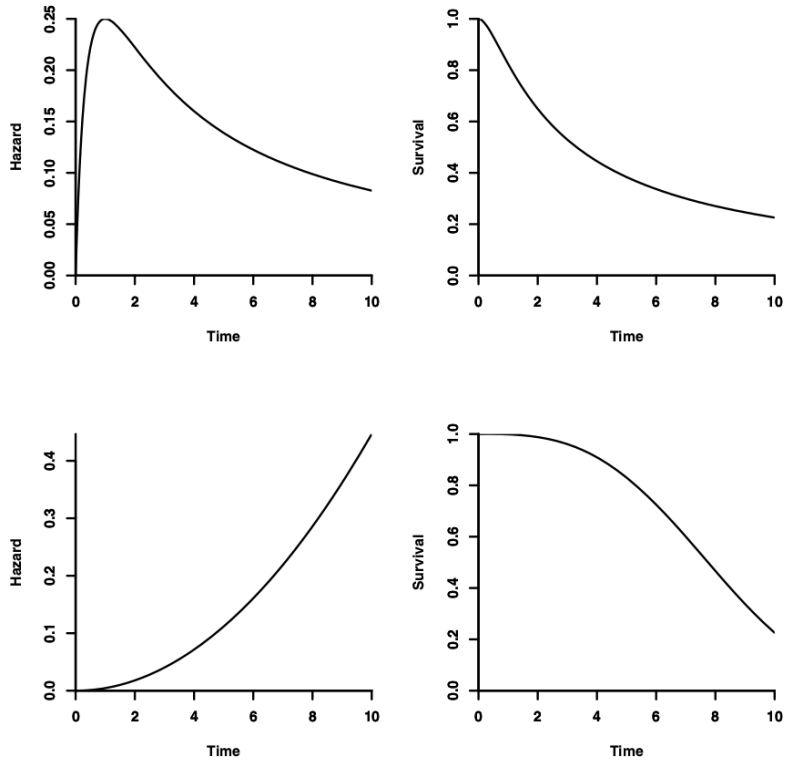
- The hazard is a conditional (on survival) rate: $dA(t) = \text{pr}(t \leq T < t + dt | T \geq t)$



Two state model: Hazard and survival

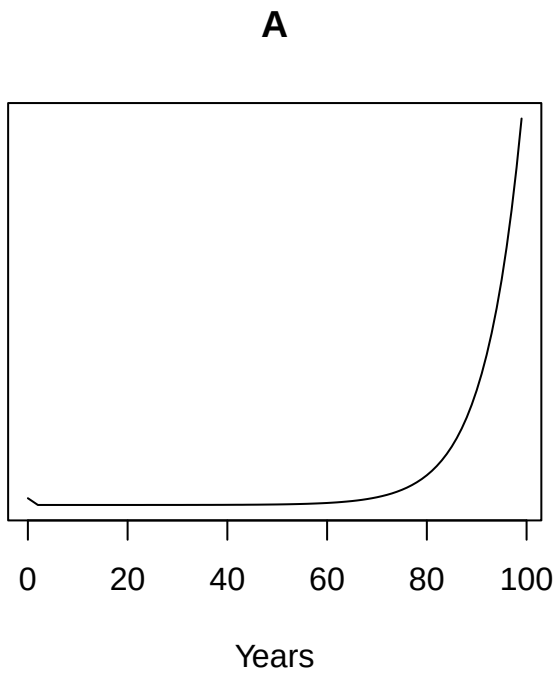
$$\alpha(s) = \frac{f(s)}{S(s)} = -\frac{\partial}{\partial s}\log S(s) \Leftrightarrow \int_0^t \alpha(s)ds = -\log S(t)$$

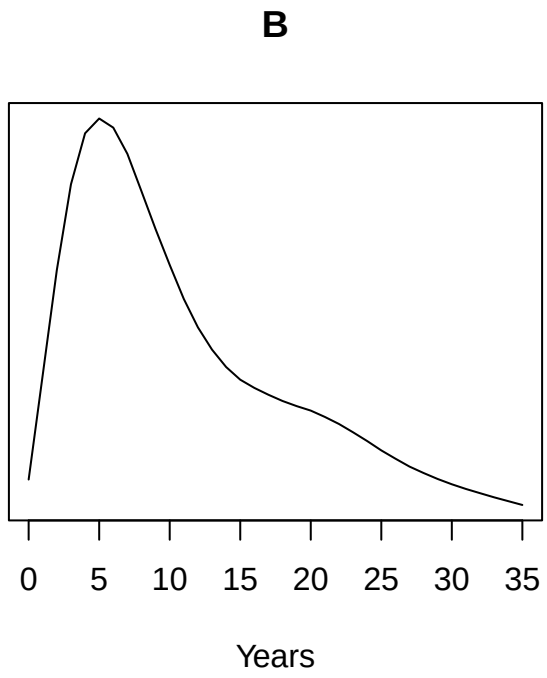
$$\Leftrightarrow S(t) = \exp\left(-\int_0^t \alpha(s)ds\right)$$



There is a one-to-one correspondence between the hazard, density and survival function.

Examples of hazards





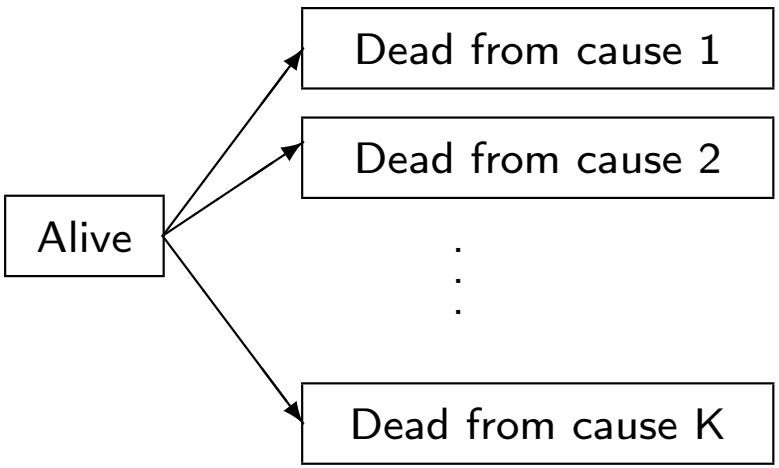
Competing risks

"All-cause mortality" or "event-free survival" is not always the relevant concept.

- In the malignant melanoma study, the composite event "death from any cause" yields a two-state dead or alive model. Instead of overall survival, researchers may be interested in assessing factors that impact the rate and risk of death from malignant melanoma, while death from other causes is of secondary interest.
- In the PBC liver trial, the endpoint "transplant-free survival" (time to failure of medical treatment, defined as either transplant or death) leads to a two-state model. Here researchers may also want to investigate factors that relate to the rate and/or probability of having a liver transplant instead of the composite outcome including also death.

Competing risks

Subjects can experience one and only one of K distinct failure types. The occurrence of one type of event precludes the occurrences of others.



Examples

- Different causes of death. Death from malignant melanoma and non-cancer related mortality
- Liver transplant and death
- Relapse or treatment related mortality in leukemia patients

Competing events and right-censoring

The occurrence of a competing event may change the rate for or preclude events of the type of interest

- Our aim is to estimate quantities such as the survival function $S(t)$, risk $F(t)$ or hazard $\alpha(t)$ from the uncensored population correctly based on a data sample where there are independent censorings
- In the survival model we deal with censorings and estimate the population survival $S(t)$ (the probability of still being alive in the uncensored target population)
- In the competing risks model, we want to estimate the rate and probability of seeing a cause / event in the presence of other events in an uncensored population

Censoring and target population

We wish to estimate population parameters based on censored data. We need to carefully consider

- The target population. The complete non-censored population must be well-defined.
- Censoring. The censoring should not leave us with a biased sample.

In the medical literature, a lot of confusion seems to arise from not clearly defining the target population (distinguishing between censoring and competing risks) and not distinguishing between "rates" and "risks".

Population and right-censoring

The complete population (without censoring) must be well-defined

- In the target (complete uncensored) population, the event will happen to everyone if we wait long enough
  - True for all-cause mortality and composite end-points such as "failure of medical treatment" (transplant or death)
  - Violated if the event of interest is a specific disease or "liver transplant" and we censor for death. Then, in the target population, even with complete follow-up, every one will not experience the event: some subjects will die without ever getting the disease or transplant.
- When censoring for death, a population "without censoring" (i.e., one where subjects are not allowed to die without the disease) is often completely hypothetical and typically inconsistent with the definition of the target population of interest.
- We must acknowledge that individuals may die without the disease and inference for the disease should be made "in the presence of the competing risk of dying"

Censoring and target population

Censoring must not leave us with a biased sample

- Independent censoring then means that the extra information that the subject is not only alive $\{T^* \geq t\}$, but also uncensored at time $\{T^* \geq t\} \cap \{C \geq t\}$, does not change the hazard
- Independent censoring cannot be tested from the available data - it is a matter of discussion
  - Violated if the most severely ill patients are censored and only the healthiest are observed
  - Censoring by being alive when the study ends can typically be taken to be independent
- "Independent censoring" should be thought of as for given covariates; censoring may depend on covariates as long as these are accounted for in the hazard model (e.g., in the Cox model)
- Some non-hazard based methods require that the censoring is not affected by covariates or that the conditional distribution of censoring is explicitly modelled. An example is the Fine and Gray model we will consider later today.

Cause-specific hazard

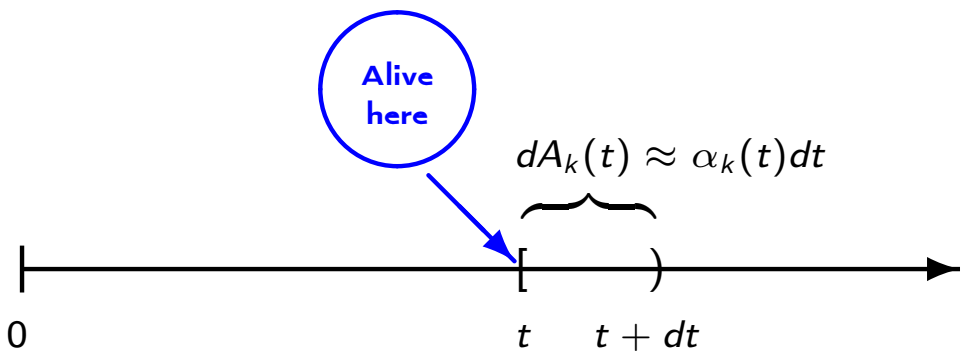
Let T^* be a survival time and let $\varepsilon \in \{1, \ldots, K\}$ denote the type of failure. The cause-specific hazard

$$\alpha_k(t) = \lim_{h \to 0} \frac{\text{pr}(t \leq T^* < t + h, \varepsilon = k | T^* \geq t)}{h}$$

is the type k failure rate among survivors.

- Represents the instantaneous rate for failures of type k in the presence of all other failure types
- Gives a local description of the mechanisms by which subjects may fail

The cause-specific hazard is a conditional (on survival from all risks) rate: $dA_k(t) = \text{pr}(t \leq T^* < t + dt, \epsilon = k | T^* \geq t)$



Cause-specific hazard

The total hazard of T^* (all cause mortality hazard) is

$$\alpha_\bullet(t) = = \lim_{h \to 0} \frac{\text{pr}(t \leq T^* < t + h | T^* \geq t)}{h} = \alpha_1(t) + \ldots + \alpha_K(t).$$

and the marginal distribution of T^* (the overall survival probability) is

$$S(t) = \text{pr}(T^* > t) = \exp\left(-\int_0^t \alpha_\bullet(s)ds\right) = \exp\left(-\sum_{k=1}^{K} A_k(t)\right).$$

Absolute risk

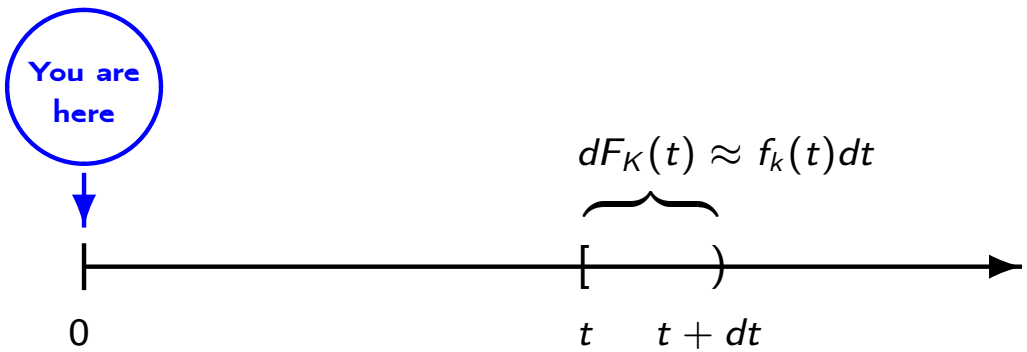
The absolute risk is the marginal probability of dying from cause k in the presence of other risks,

$$F_k(t) = \text{pr}(T^* \leq t, \varepsilon = k).$$

- F_k is a sub-distribution function. $F_k(\infty) = \text{pr}(\epsilon = k) < 1$ when there is a positive probability of experiencing a competing event.
- At any time, the probability of still being alive or having failed from any of causes sums to one

$$S(t) + \sum_{k=1}^{K} F_k(t) = 1$$

The sub-distribution density $f_k(t) = \partial F_k(t)/\partial t$ is the marginal cause k incidence in the presence of other risks: $dF_k(t) = \text{pr}(t \leq T^* < t + dt, \epsilon = k)$



Relation between α_k and F_k

$$\begin{aligned}\alpha_k(t) &= \lim_{h\to 0} \frac{\text{pr}(t \le T^* < t+h, \epsilon = k | T^* \ge t)}{h} \\ &= \lim_{h\to 0} \frac{\text{pr}(t \le T^* < t+h, \epsilon = k)/h}{\text{pr}(T^* \ge t)} \\ &= \frac{f_k(t)}{S(t-)} = \frac{f_k(t)}{1 - \sum_{l=1}^{K} F_l(t-)}\end{aligned}$$

so that

$$f_k(t) = S(t-)\alpha_k(t) = \exp\left(-\sum_{l=1}^{K} A_l(t-)\right)\alpha_k(t)$$

and

$$F_k(t) = \int_0^t S(u-)\alpha_k(u)du = \int_0^t \exp\left(-\sum_{l=1}^{K} A_l(u-)\right)\alpha_k(u)du.$$

The cumulative incidence for cause k is a function of all the cause-specific hazards. And the cause-specific hazard is a function of all the sub-distributions.

For a particular risk there is no simple one-to-one correspondence between α_k and F_k as for (overall) survival

Example

Consider $K = 2$ causes, two treatments $x \in \{0, 1\}$ and hazards

$$\begin{aligned}
\alpha_1(t; x) &= 1, \\
A_1(t; x) &= t \\
\alpha_2(t; x) &= \frac{t^{1/2}}{2^x}, \\
A_2(t; x) &= \frac{2^{1-x} t^{3/2}}{3}
\end{aligned}$$

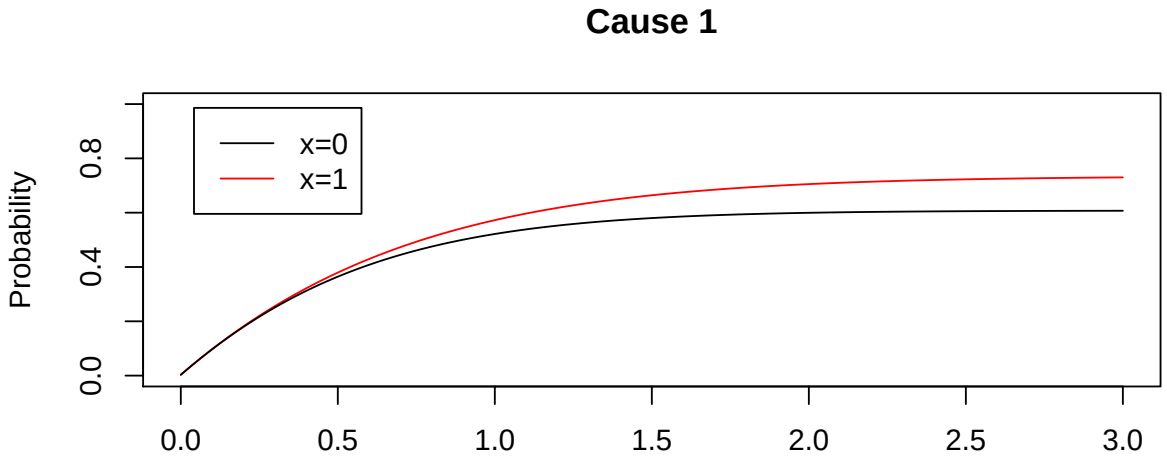
The cause 1 hazard is not affected by the treatment.

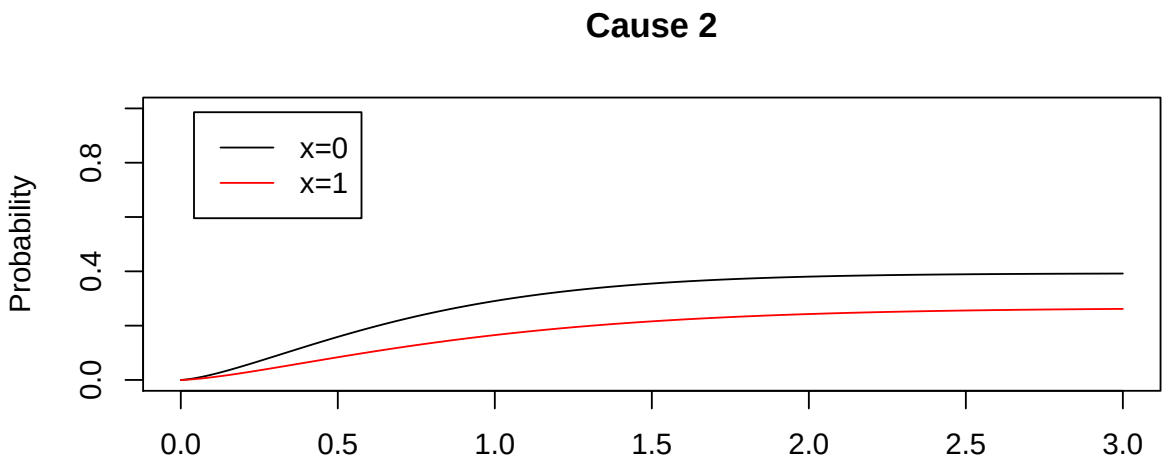
The absolute risks are

$$\begin{aligned}
F_1(t; x) &= \int_0^t \exp\left(-s - \frac{2^{1-x} s^{3/2}}{3}\right) ds \\
F_2(t; x) &= \int_0^t \exp\left(-s - \frac{2^{1-x} s^{3/2}}{3}\right) \frac{s^{1/2}}{2^x} ds
\end{aligned}$$

The treatment has an effect on the probability of type 1 failures, although this is solely as a result of the change in the type 2 rates.

Example


Cause 1
Probability
0.0
0.4
0.8
x=0
x=1
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Cause 2
Probability
0.0
0.4
0.8
x=0
x=1
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Inference for cause-specific hazards

Let ϵ_i, X_i, T_i^* and C_i be the failure type, covariates, and event and censoring time. We observe the censored data $(T_i, \Delta_i \varepsilon_i, X_i)$, $i = 1, \ldots, n$, where $T_i = T^* \wedge C$, $\Delta_i = I(T_i \geq C_i)$. Assume that the censoring is independent and that there are no common parameters among the failure types. The observed data likelihood is proportional to

$$\begin{aligned}
&\prod_{i=1}^{n} \left(S(T_i; X_i) \prod_{k=1}^{K} (\alpha_k(T_i; X_i))^{I(\Delta_i \varepsilon_i = k)} \right) \\
&\quad = \prod_{i=1}^{n} \exp\left(- \sum_{k=1}^{K} A_k(T_i; X_i) \right) \prod_{k=1}^{K} (\alpha_k(T_i; X_i))^{I(\Delta_i \varepsilon_i = k)} \\
&\quad = \prod_{k=1}^{K} \left(\prod_{i=1}^{n} e^{-A_k(t; X_i)} (\alpha_k(T_i; X_i))^{I(\Delta_i \varepsilon_i = k)} \right) \\
&\quad = \prod_{k=1}^{K} \left(\prod_{i=1}^{n} \exp\left(- \int_0^{\tau} Y_i(s) \alpha_k(s; X_i) ds \right) (\alpha_k(T_i; X_i))^{I(\Delta_i \varepsilon_i = k)} \right),
\end{aligned}$$

where $Y_i(t) = I(T_i \geq t)$ and $A_j(t) = \int_0^t \alpha_j(s) ds$.

The kth factor corresponds to what we would have if only cause k events were studied and events of the other types were (independent) right-censorings

Cause-specific hazards

The likelihood factorizes according to causes $k = 1, \ldots, K$.

- Each factor k has the form it would have if only cause k events were events, and events of the other type were treated as (independent) censorings
- This has nothing to do with "independence" of causes. It is solely a consequence of the definition of cause-specific hazards as conditional rates of exclusive events.

Cause-specific estimators

Let $N_{ik}(t) = I(T_i \leq t, \Delta_i \varepsilon_i = k)$ and $Y_i(t) = I(T_i \geq t)$. Then $N_{ik}(t)$ has intensity process

$$\begin{aligned} & E\left(dN_{ik}(t) | N_{i1}(t-), \ldots, N_{iK}(t-), Y_i(t)\right) \\ &= Y_i(t) E\left(dN_{ik}(t) | Y_i(t) = 1\right) \\ &= Y_i(t) E\left(dN_{ik}(t) | T_i^* \geq t, C_i \geq t\right) \\ &= Y_i(t) E\left(dN_{ik}(t) | T_i^* \geq t\right) \\ &= Y_i(t) dA_k(t) \end{aligned}$$

such that

$$\begin{pmatrix} N_{i1}(t) - Y_i(t) dA_1(t) \\ \vdots \\ N_{iK}(t) - Y_i(t) dA_K(t) \end{pmatrix}$$

is a K-dimensional martingale with respect to the observed filtration.

Asymptotics for cause-specific hazard based methods follows by the same martingale arguments as in the survival two-state situation (see day 1 and 2).

Cause-specific hazards

All standard hazard-based models for survival data apply for inference on the cause-specific hazards. We can just pretend that the events of other types are censorings. For example,

- Log-rank test for cause-specific hazards
- Cause-specific Cox regression
- Cause-specific Nelson-Aalen
- Parametric regression models; Poisson, Weibull, ...
- Model checks by cumulative martingale residuals
- ...

But be careful when considering probabilities/risks (more on this)

Risk estimation

When there are no covariates, A_k can be estimated by the cause-specific Nelson-Aalen estimator

$$\hat{A}_k(t) = \int_0^t \frac{dN_{\bullet k}(s)}{Y_\bullet(s)}$$

where $N_{\bullet k}(t) = \sum_{i=1}^n N_{ik}(t)$ and $Y_\bullet(t) = \sum_{i=1}^n Y_i(t)$.

The overall survival can be estimated by the Kaplan-Meier estimator ignoring the failure type

$$\hat{S}(t) = \prod_{s \leq t} \left(1 - \frac{dN_{\bullet\bullet}(s)}{Y_\bullet(s)}\right)$$

where $N_{\bullet\bullet}(t) = \sum_{i=1}^n \sum_{k=1}^K N_{ik}(t)$ and $\prod$ denotes the product integral.

The absolute risk from cause k,

$$F_k(t) = \text{pr}(T^* \leq t, \varepsilon = k) = \int_0^t S(u) dA_k(u).$$

can be estimated by plugging in $d\hat{A}_k$ and $\hat{S}$

$$\hat{F}_k(t) = \int_0^t \hat{S}(u) d\hat{A}_k(u).$$

The Aalen-Johansen estimator

- The product integral is Hadamard differentiable (Gill and Johansen, 1990). The asymptotic distribution of $\sqrt{n}(\hat{F}_1(t) - F_1(t))$ can be obtained from the asymptotic properties of $\hat{A}_k(t)$, $k = 1, \ldots, K$, by the functional delta method applied to the functional

$$(\Lambda_1, \ldots, \Lambda_2) \mapsto \int_0^t \prod_{u \leq s} (1 - dA_\bullet(u)) dA_k(s)$$

  where $A_\bullet = \sum_{j=1}^K A_j$.

- The asymptotic variance of $\hat{F}_1(t)$ can be estimated by

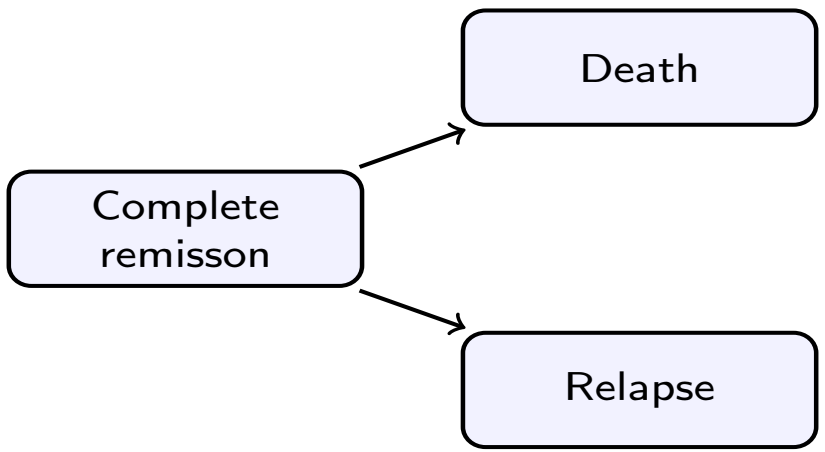
$$\sum_{k=1}^K \int_0^t \frac{\hat{S}(u)^2}{Y_\bullet(u)} \left(I(k = 1) - \frac{\hat{F}_1(t) - \hat{F}_1(u)}{\hat{S}(u)} \right)^2 d\hat{A}_k(u)$$

Alternatively, the properties of the estimator can be deduced from the theory for more general multistate models. The estimator $\hat{F}_1$ is the Aalen-Johansen estimator, a non-parametric estimator of transition probabilites in non-homogeneous Markov processes, see Section 10.1 of Martinussen and Scheike.

Survival after bone marrow transplant

1715 leukemia patients received a bone marrow transplant (International Bone Marrow Transplant Registry). At the end of the follow-up period (median 37.2 months)

- 311 patients relapsed
- 557 patients died in remission
- 847 patients still alive and in remission (right-censored)



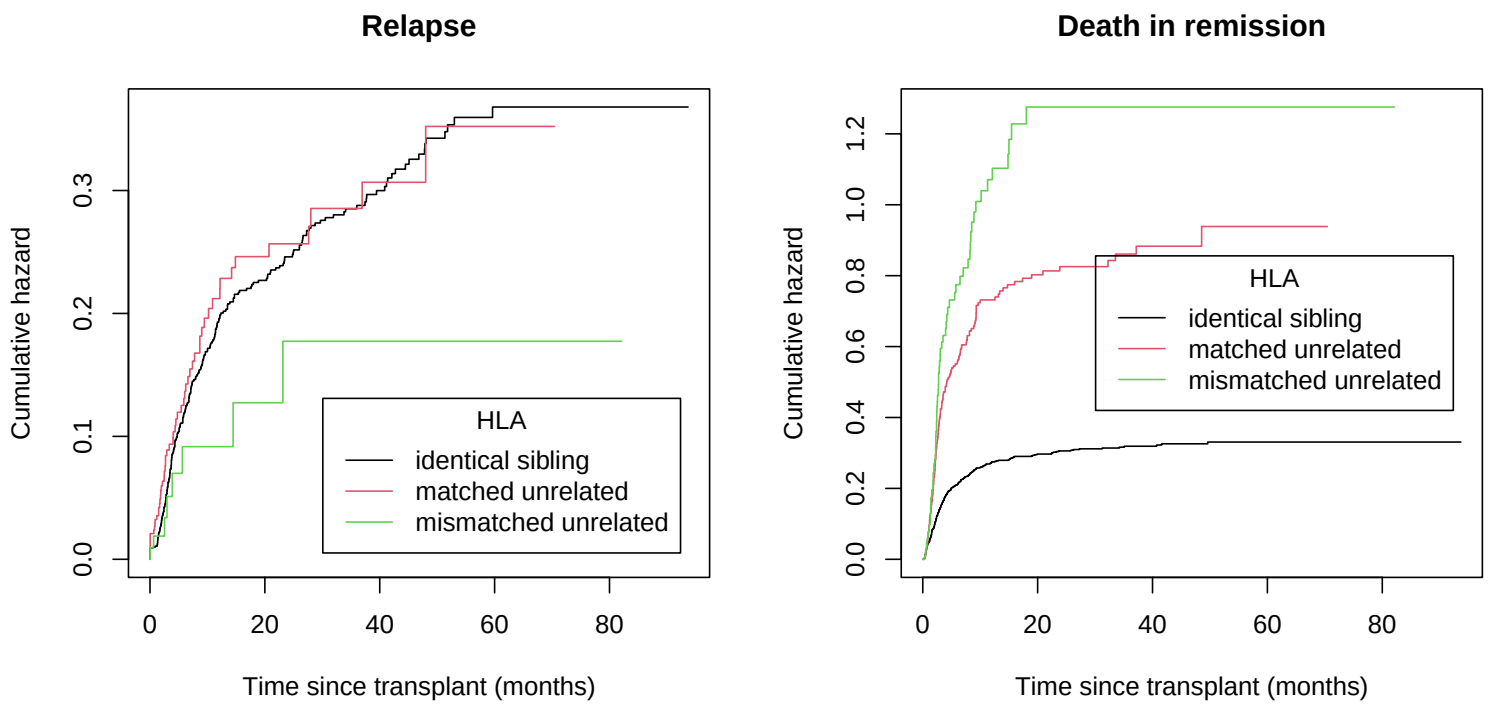
Survival after bone marrow transplant

Information on

- time from transplant to event or censoring (months)
- event : 0=censoring; 1=relapse; 2=death in remission
- donor : 1=HLA identical sibling (n=1224); 2=HLA matched unrelated (n=383); 3=HLA mismatched unrelated (n=108)

Purpose: comparison of relapse and death in remission between patients with different donor types.

```
library(timereg)
bmt <- read.csv2("http://publicifsv.sund.ku.dk/~frank/data/bmt1715.csv")
par(mfrow=c(1,2)) # Plot 1 row, 2 columns
na1 <- survfit(Surv(time, event==1)~donor, data=bmt)
kmplot(na1, fun="cumhaz", main="Relapse", lty=1, add.legend=FALSE, conf.int=FALSE,
       xlab="Time since transplant (months)", ylab="Cumulative hazard")
legend("bottomright", inset=.05,lty=1, col=1:3, title="HLA",
       legend=c("identical sibling","matched unrelated", "mismatched unrelated"))
na2 <- survfit(Surv(time, event==2)~donor, data=bmt)
kmplot(na2, fun="cumhaz", main="Death in remission", lty=1, add.legend=FALSE, conf.int=FALSE,
     xlab="Time since transplant (months)", ylab="Cumulative hazard")
legend("right", inset=.05,lty=1, col=1:3, title="HLA",
       legend=c("identical sibling","matched unrelated", "mismatched unrelated"))
```



Survival after bone marrow transplant

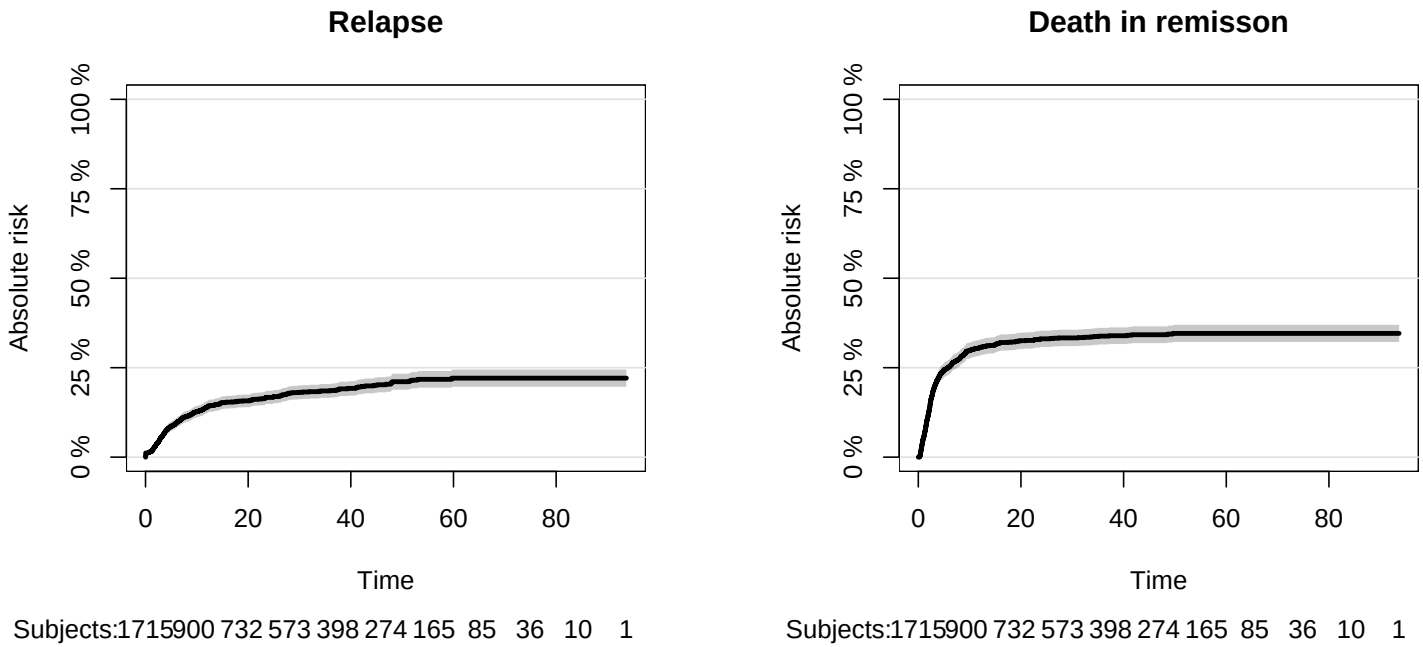
The "local slope" of the cumulative hazard approximates the hazard.

The plots suggest that

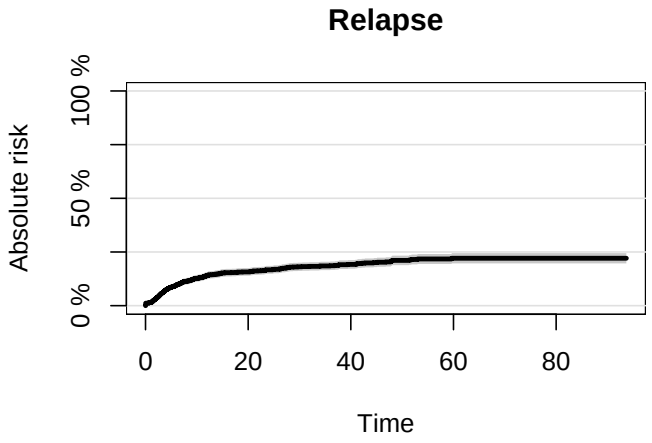
- Patients with HLA mismatched unrelated donors have lower relapse rate than patients with HLA identical sibling donors.
- Little sign of a difference in relapse rates between HLA matched unrelated donors and HLA identical sibling donors
- HLA identical sibling donors have lower treatment related mortality rate than HLA matched unrelated donors
- HLA matched unrelated donor patients have lower treatment related mortality rate than patients with HLA mismatched unrelated donors

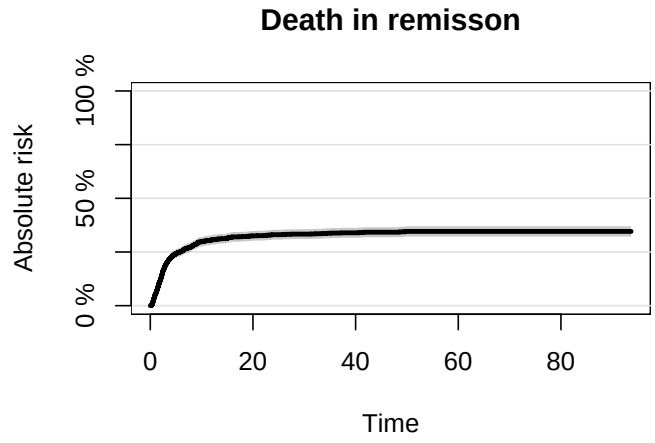
The Aalen-Johansen estimator can be estimated by `prodlim` with `Hist()` instead of `Surv()`

```
library(prodlim)
aj1 <- prodlim(Hist(time, event)~1, data=bmt) # Hist() instead of Surv()
par(mfrow=c(1,2))
plot(aj1, cause=1, plot.main="Relapse") # Plot cause 1
plot(aj1, cause=2, plot.main="Death in remisson") # Plot cause 2
```



Independent censoring should be labelled 0.



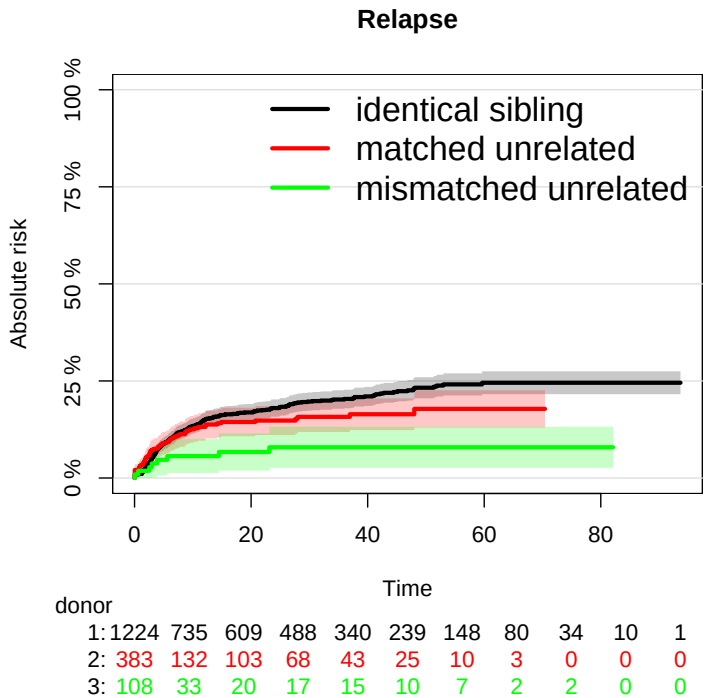


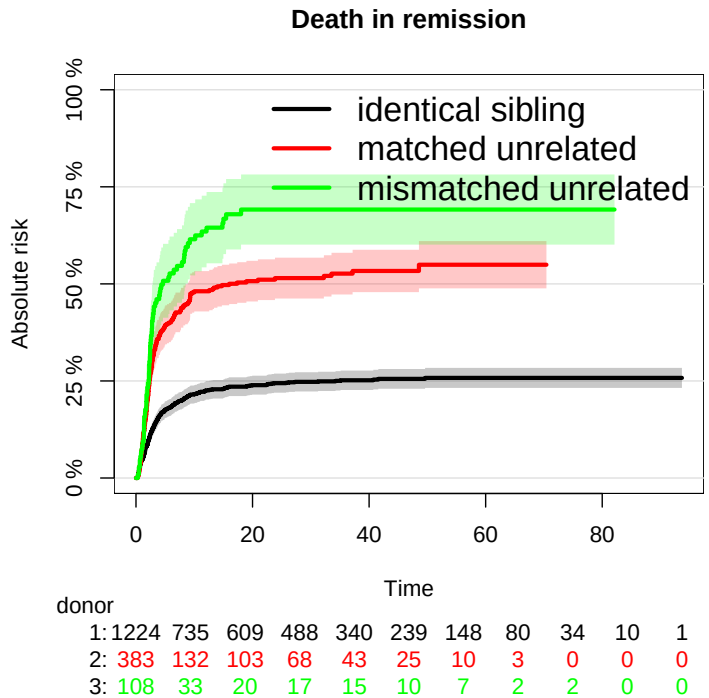

```r
summary(aj1, times=c(20,80)) # CIF and 95%CI at specific times
```

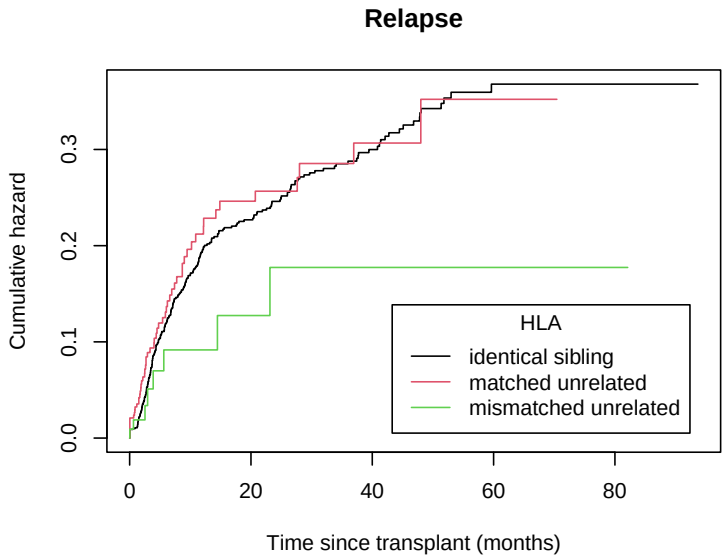
```
##
##
## ----------> Cause:  1
##
##   time n.risk n.event n.lost cuminc se.cuminc lower upper
## 1   20    714       0      0  0.157   0.00907 0.139 0.175
## 2   80     18       0      0  0.221   0.01226 0.197 0.245
##
##
## ----------> Cause:  2
##
##   time n.risk n.event n.lost cuminc se.cuminc lower upper
## 1   20    714       0      0  0.325    0.0116 0.303 0.348
## 2   80     18       0      0  0.346    0.0122 0.322 0.370
```

Probability (risk) of relapse (left) and treatment related death (right) after bone marrow transplant according to HLA match

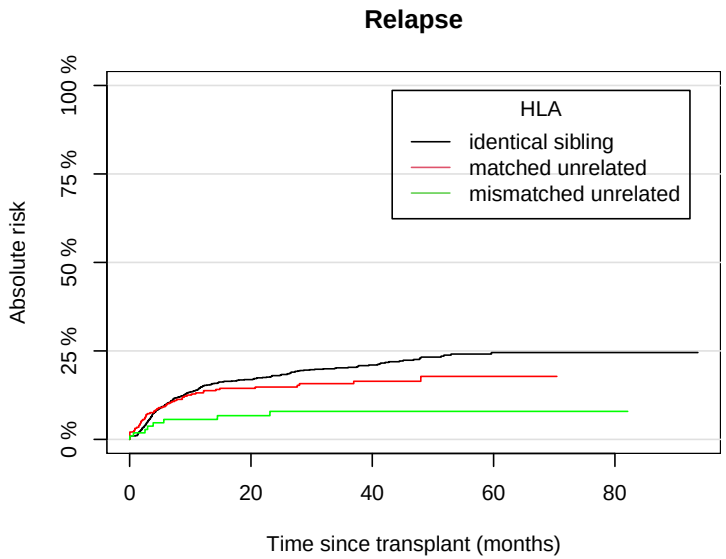
```
par(mfrow=c(1,2))
aj <- prodlim(Hist(time, event)~donor, data=bmt)
plot(aj, cause=1, plot.main="Relapse", col=c("black","red","green"),
     legend.legend=c("identical sibling","matched unrelated","mismatched unrelated"))

plot(aj, cause=2, plot.main="Death in remission", col=c("black","red","green"),
     legend.legend=c("identical sibling","matched unrelated","mismatched unrelated"))
```

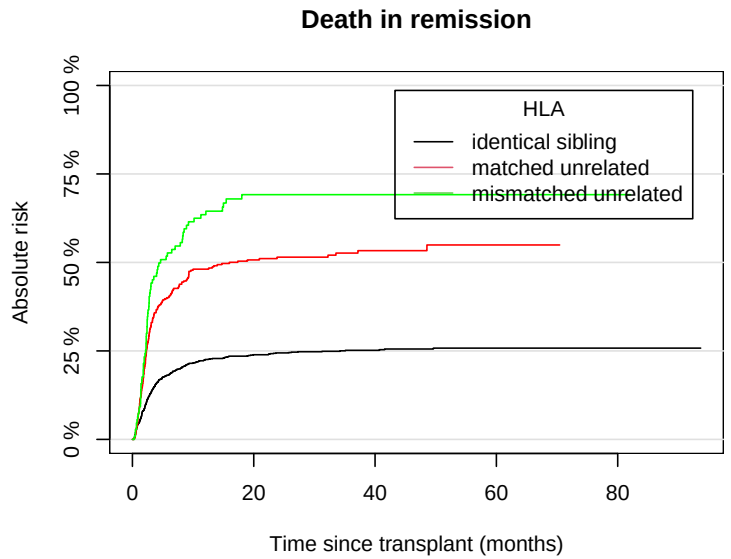





Relapse
Cumulative hazard
0.0
0.1
0.2
0.3
0
20
40
60
80
Time since transplant (months)
HLA
identical sibling
matched unrelated
mismatched unrelated


Death in remission
Cumulative hazard
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
20
40
60
80
Time since transplant (months)
HLA
identical sibling
matched unrelated
mismatched unrelated


Relapse
Absolute risk
0 %
25 %
50 %
75 %
100 %
0
20
40
60
80
Time since transplant (months)
HLA
identical sibling
matched unrelated
mismatched unrelated


Death in remission
Absolute risk
0 %
25 %
50 %
75 %
100 %
0
20
40
60
80
Time since transplant (months)
HLA
identical sibling
matched unrelated
mismatched unrelated

Hazard and risk for relapse and treatment related mortality

The plots on the previous slide show the cumulative hazards (upper row) and absolute risks (lower row) according to HLA match.

1. Does receiving bone marrow from an HLA matched unrelated compared to form a HLA identical sibling donor increase or decrease the hazard for relapse ?
2. Does receiving bone marrow from an HLA matched unrelated compared to from a sibling donor increase or decrease the risk of relapse ?
3. Relate the two answers above to each other and to treatment related mortality
4. Would you choose a HLA mismatched unrelated or a HLA identical sibling donor if you had the choice?

Digression: Latent failure times

Consider (continuous) "latent failure times" $T_1^L, \ldots, T_K^L$ with joint survival distribution

$$Q(t_1, \ldots, t_K) = \text{pr}(T_1^L > t_1, \ldots, T_K^L > t_K).$$

Let $T = \min\{T_1^L, \ldots, T_K^L\}$, and let $\varepsilon = \text{argmin}_k T_k^L$ denote the corresponding failure type.

$$\begin{aligned}
S(t) &= \text{pr}(T > t) = Q(t, \ldots, t) \\
\alpha_k(t) &= \lim_{h \to 0} \frac{\text{pr}(T < t + h, \epsilon = k | T \geq t)}{h} \\
&= \lim_{h \to 0} \frac{\text{pr}(T_k^L < t + h | T_1^L \geq t, \ldots, T_K^L \geq t)}{h} \\
&= \frac{\lim_{h \to 0} \frac{1}{h} \overbrace{\text{pr}(T_1^L \geq t, \ldots, t \leq T_k^L < t + h, \ldots, T_K^L \geq t)}^{\text{pr}(T_1^L \geq t, \ldots, T_k^L \geq t, \ldots, T_K^L \geq t) - \text{pr}(T_1^L \geq t, \ldots, T_k^L \geq t + h, \ldots, T_K^L \geq t)}}{\text{pr}(T_1^L \geq t, \ldots, T_K^L \geq t)} \\
&= -\frac{\lim_{h \to 0}(Q(t, \ldots, t + h, \ldots, t) - Q(t, \ldots, t))/h}{Q(t-, \ldots, t-)} \\
&= -\frac{\frac{\partial}{\partial t_k} Q(t_1, \ldots, t_K)|_{t_1 = \ldots = t_K = t}}{Q(t, \ldots, t)} \\
&= -\frac{\partial}{\partial t_k} \log Q(t_1, \ldots, t_K)|_{t_1 = \ldots = t_K = t}.
\end{aligned}$$

When we only observe (censored) competing risks data $T_i = \min\{T_{1i}^L, \ldots, T_{Ki}^L, C_i\}$, and $\Delta_i = I(T_i < C_i)$, $\Delta_i \varepsilon_i$, what parameters may be identified?

From the likelihood

$$\prod_{k=1}^{K} \left(\prod_{i=1}^{n} \exp\left(-\sum_{k=1}^{K} \int_0^{T_i} dA_k(u)du \right) (dA_k(T_i))^{I(\Delta_i \varepsilon_i = k)} \right),$$

we may identify the cause-specific hazards $dA_k(t)$ but not the whole joint distribution $Q(\cdot)$.

We can for example not identify the "marginal distribution" of T_k^L,

$$\text{pr}(T_k^L > t) = Q(0, \ldots, 0, t_k, 0, \ldots, 0) = S_k(t)$$

nor the corresponding "net" hazard

$$h_k(t) = \lim_{h \to 0} \frac{\text{pr}(T_k^L < t + h | T_k^L \geq t)}{h} = -\frac{\partial}{\partial t} \log S_k(t).$$

without additional untestable assumptions on the joint distribution.

"Independent" competing risks

- Historically, authors considered
  - Independent failure times $T_{\mathbf{1}}^L, \ldots, T_K^L$ such that

$$Q(t_{\mathbf{1}}, \ldots, t_K) = \prod_{k=\mathbf{1}}^{K} S_k(t_k)$$

    for all $t_{\mathbf{1}}, \ldots, t_K$.
  - or the weaker

$$Q(t, \ldots, t) = \prod_{k=\mathbf{1}}^{K} S_k(t)$$

    so that the marginal "net" and cause-specific "crude" hazards are the same $h_k(t) = \alpha_k(t)$
- Because S_j and h_j cannot be identified from the likelihood without further unidentifiable conditions, these assumption are not possible to verify.
- One cannot objectively make inference on the latent multivariate failure times based solely on the competing risks data.
- From a scientific perspective, it is often counter-intuitive or problematic to hypothesize multiple events (such as deaths) occurring to a single subject.
- The question of "what would happen if certain causes were removed" ("partial crude hazards") is quite hypothetical in most biological settings (except possibly for failure of technical systems due to components in "unrelated parts" of the system).

Predicting absolute risk from hazards

With covariates, given estimators $d\hat{A}_k(t|X)$ of the cause-specific hazards, the cumulative incidence can be estimated by

$$\hat{F}_k(t|X) = \int_0^t \hat{S}(s|X) d\hat{A}_k(s|X) = \int_0^t \prod_{u \leq s} \left(1 - d\hat{A}_\bullet(u|X)\right) d\hat{A}_k(s|X).$$

Based on regression models (e.g., Cox models) for all cause-specific hazard $\alpha_k(t, x)$, $k = 1, \ldots, K$, we can predict the absolute risks

- Cox models impose a simple structure between covariates and the cause-specific hazards
- Due to the non-linear realationship between rates and risks in a competing risk model, the simple relationship does not carry over to the absolute risks.
- The way in which a covariate affects a rate can be very different from the way it affects the corresponding risk. This will depend on if and how it affects the rates for the competing causes.

Sub-distribution hazard

Plugging in cause-specific hazard models does not provide parameters that in a simple way describe the relationship between covariates and risks. As we have seen, there is no one-to-one relation between the cause k-specific hazard α_k and the risk of a type k event, F_k.

In order to reestablish such a relation it has been suggested to consider the sub-distribution hazard, the function $\alpha^{\#}(t;X)$ such that

$$F_k(t|X) = \text{pr}(T^* \leq t, \epsilon = k|X) = 1 - \exp\left(-\int_0^t \alpha_k^{\#}(s;X)ds\right)$$

mimicking the relation between survival and the all-cause hazard

$$F(t|X) = \text{pr}(T \leq t|X) = 1 - S(t|X) = 1 - \exp\left(-\int_0^t \alpha_\bullet(s;X)ds\right).$$

That is, the transformation which for all-cause mortality takes us from absolute risk to hazard is used for absolute risk in a competing risk setting.

Sub-distribution hazard

The relation

$$F_1(t|X) = 1 - \exp\left(-\int_0^t \alpha_1^{\#}(s;X)ds\right)$$

can also be expressed as

$$\alpha_1^{\#}(t;X) = -\frac{\partial}{\partial t}\log(1 - F_1(t|X))$$

$$= \lim_{h\to 0} \frac{\text{pr}(t \leq T^* < t+h, \epsilon = 1 | \overbrace{\{T^* < t, \epsilon \neq 1\} \cup \{T^* \geq t\}}^{\text{"No type 1 event by time } t\text{"}}, X)}{h}$$

The instantaneous type 1 rate, given that the subject has not failed from cause 1. The conditioning set includes the event that the subject already failed, but for another cause.

Sub-distribution hazard

One may also think of

$$\alpha_1^{\#}(t;X) = \lim_{h\to 0} \frac{\text{pr}(t \le T^* < t+h, \epsilon = 1 | \{T^* < t, \epsilon \neq 1\} \cup \{T \ge t\}, X)}{h}$$

as the hazard of the event time of the type 1 risk in the presence of other risks (and ∞ if the type 1 event didn't happen), i.e., the improper random variable

$$T^{\#} = T^* \times I(\epsilon = 1) + \infty \times I(\epsilon \neq 1).$$

From

$$\begin{aligned}\text{pr}(T^{\#} > t|X) &= \text{pr}(T^* > t|X) + \sum_{k=2}^{K} \text{pr}(T^* \le t, \epsilon = k|X) \\ &= S(t|X) + \sum_{k=2}^{K} F_k(t|X) = 1 - F_1(t|X)\end{aligned}$$

we see that $T^{\#}$ has distribution function $F_1(t|X)$ for $t < \infty$ and a point mass at ∞.

The sub-distribution hazard is the usual hazard of $T^{\#}$,

$$\alpha_1^{\#}(t;X) = \lim_{h\to 0} \frac{\text{pr}(t \le T^{\#} < t+h | T^{\#} \ge t, X)}{h}$$

Gray's test

Without covariates (or within groups of covariates)
$A_k^{\#}(t) = \int_0^t (1 - F_k(s))^{-1} dF_k(s)$ can be estimated by $\int_0^t d\hat{A}_k^{\#}(s)$ where

$$d\hat{A}_k^{\#}(s) = \frac{d\hat{F}_k(s)}{1 - \hat{F}_k(s)}$$

where $\hat{F}_k(t)$ is the Aalen-Johansen estimate of $F_k(t)$ from before.

Gray's test with two groups $X \in \{0, 1\}$

$$\int_0^\tau W(s) d(\hat{A}_1^{\#}(s; X = 1) - \hat{A}_1^{\#}(s; X = 0))$$

- W is a weight function (depending on data)
- A log-rank type test for the null hypothesis that that cumulative incidences are equal in group $X = 0$ and $X = 1$.
- Can be stratified and extended to multiple groups

Gray's test and the log-rank test

In a competing risks setting, both Gray's test and the usual log-rank test where we treat competing events are valid and applicable, but the null hypothesis is different.

- Gray's test tests

$$H_0 : \alpha_1^{\#}(t; X = 0) = \alpha_1^{\#}(t; X = 1)$$

  for all $t \in [0, \tau]$, but because of the one-to-one relation between $\alpha^{\#}(t)$ and $F_1(t)$, $\alpha_1^{\#}(t; X) = -\frac{\partial}{\partial t} \log(1 - F_1(t|X))$, this is equivalent to testing

$$H_0 : F_1(t|X = 0) = F_1(t|X = 1)$$

  for all $t \in [0, \tau]$

- The log-rank test tests

$$H_0 : \alpha_1(t|; X = 0) = \alpha_1(t; X = 1)$$

  for all $t \in [0, \tau]$

Gray's test is implemented in the function `cuminc` of the `cmprsk` `R` package

Proportional sub-distribution hazards

The fact that plugging-in cause-specific hazard models does not provide parameters that in a simple way describe the relationship between covariates and cumulative incidences has led to the development of direct regression models for the cumulative incidences.

Fine & Gray proposed the Cox-type proportional sub-distribution hazard model

$$\alpha_k^{\#}(t; X) = \alpha_{k0}^{\#}(t) \exp(\beta^T X)$$

where the covariate parameters β are log-sub-distribution hazard ratios

Fine and Gray's score function with complete data

- Let $N_{ik}^{\#}(t) = I(T_i^* \leq t, \epsilon_i = k)$ denote the complete data type k counting process and

$$Y_{ik}^{\#}(t) = I\left\{\{T_i^* \geq t\} \cup \{T_i^* < t, \epsilon_i \neq k\}\right\} = 1 - N_{ik}^{\#}(t-).$$

  the corresponding "sub-distribution at-risk indicator".
- Anyone who hasn't experienced a type k event is "at risk", including those who failed from other causes.
- With complete data (no censoring), $N_{ik}^{\#}(t)$ and $Y_{ik}^{\#}$ are observable.

Fine and Gray's score function with complete data

- Define the filtration $\mathcal{F}_t^k = \sigma(N_{jk}(u), X_j, u \leq t, j = 1, \ldots, n)$.
- $N_{ik}^{\#}(t)$ has compensator $\int_0^t Y_{ik}^{\#}(u) dA_k^{\#}(u; X_i, \beta_0)$ with respect to $\mathcal{F}_t^k$

$$\begin{aligned} E(dN_{ik}^{\#}(t)|\mathcal{F}_{t-}^k) &= I(N_{ik}^{\#}(t-) = 0) E(dN_{ik}^{\#}(t)|N_{ik}^{\#}(t-) = 0, X_i) \\ &= Y_{ik}^{\#}(t) E(dN_{ik}^{\#}(t)|\{T_i^* \geq t\} \cup \{T_i^* < t, \epsilon_i \neq k\}, X_i) \\ &= Y_{ik}^{\#}(t) dA_k^{\#}(t; X_i, \beta_0) \end{aligned}$$

- With respect to $\mathcal{F}_t^k$,

$$N_{ik}^{\#}(t) - \int_0^t Y_{ik}^{\#}(u) dA_k^{\#}(u; X_i, \beta_0)$$

is a martingale (but not w.r.t. the natural filtration for the cause-specific hazard, $\sigma(N_{jk}(u), X_j, u \leq t, j = 1, \ldots, n, k = 1, \ldots, K)$).

Fine and Gray's score function with known censoring times

- Who is "at risk" (in the sub-distribution world) when there is censoring? Remember that anyone experiencing a competing event is kept "alive". In this hypothetical cohort, anyone who is not censored and has not experienced a type 1 event is "at risk".
- Assume that the censoring times are known for everyone, including those that failed (e.g., administrative censoring at the time when the data is collected for analysis)
- Define the cause k "risk set" indicator at time t as

$$I\{\underbrace{\{T_i^* \wedge C_i \geq t\}}_{\text{Known}} \cup \underbrace{\{T_i^* < t, C_i \geq t, \epsilon_i \neq k\}}_{\text{Generally unknown}}\} = \underbrace{(1 - N_{ik}^{\#}(t-))}_{=Y_{ik}^{\#}(t)} I(C_i \geq t)$$

Individuals who experience an uncensored event from a competing risk are kept at risk until the point in time when they had been censored (if they still had been alive).

Fine and Gray's score function with known censoring times

- If $(T^*, \epsilon) \perp\!\!\!\perp C|X$, the "crude" sub-distribution hazard is equal to the "net" sub-distribution hazard

$$\begin{aligned}
&\lim_{h\to 0} \frac{\text{pr}(t \le T^* < t+h, \epsilon = k | \{T \ge t\} \cup \{T^* < t, C \ge t, \epsilon \ne k\}, X)}{h} \\
&= \frac{\lim_{h\to 0} \text{pr}(t \le T^* < t+h, \epsilon = k, C \ge t | X)/h}{\text{pr}(\{\{T^* \ge t\} \cup \{T^* < t, \epsilon \ne k\}\} \cap \{C \ge t\} | X)} \\
&= \frac{\lim_{h\to 0} \text{pr}(t \le T^* < t+h, \epsilon = k | X)\text{pr}(C \ge t | X)/h}{\text{pr}(\{T^* \ge t\} \cup \{T^* < t, \epsilon \ne k\} | X)\text{pr}(C \ge t | X)} \\
&= \frac{\lim_{h\to 0} \text{pr}(t \le T^* < t+h, \epsilon = k | X)/h}{\text{pr}(\{T^* \ge t\} \cup \{T^* < t, \epsilon \ne k\} | X)} = \frac{f_k(t|X)}{1 - F_k(t-|X)} = \alpha_k^{\#}(t|X)
\end{aligned}$$

- When C_i is always observed,

$$\int_0^t I(C_i \ge u) dN_{ik}^{\#}(u) - \int_0^t I(C_i \ge u) Y_{ik}^{\#}(u) \alpha_k^{\#}(u|X_i, \beta_0) du,$$

is a martingale with respect to the filtration

$$\sigma(I(C_j \ge u), I(C_j \ge u) N_{jk}^{\#}(u), I(C_j \ge u) Y_{jk}^{\#}(u), X_j, u \le t, j = 1, \ldots, n).$$

Fine and Gray's score function

- When there is no censoring ($C_i = \infty$), or when the censoring times are known for everyone, the log-sub-distribution hazard ratios β can be estimated by the partial likelihood score equation

$$\sum_{i=1}^{n} \int_0^{\tau} \left(X_i - \frac{\sum_{j=1}^{n} I(C_j \geq t) Y_{jk}^{\#}(t) X_j \exp(\beta^T X_j)}{\sum_{j=1}^{n} I(C_j \geq t) Y_{jk}^{\#}(t) X_j \exp(\beta^T X_j)} \right) I(C_i \geq t) dN_{ki}^{\#}(t) = 0.$$

- This corresponds to replacing times to failures from other causes than k by ∞ in the usual Cox score function.
- Asymptotic normality of the estimator for β and weak convergence of a Breslow-type estimator for $\Lambda_{k0}^{\#}(t)$ follows by the martingale central limit theorem by the same arguments as for the Cox model.

Fine and Gray's score function

- When the censoring times C_i are unknown for individuals with $\Delta_i = 1$, we don't know how long to keep individuals who experience an observed competing event in the risk set. The indicator from before

$$Y_{ik}^{\#}(t) I(C_i \geq t)$$

  may not be computable because C_i is unknown when $T_i^* \leq C_i$.
- Fine and Gray suggest approximating the risk sets when they are unknown.
- The term

$$r_i(t) = I(C_i \geq T_i^* \wedge t)$$

  indicates knowledge of the vital status of individual i just before time t
  - $r_i(t) = 1$ for individuals with no event of any kind, nor censoring, in $[0, t)$ and for individuals with an observed event of type $1, \ldots, K$ in $[0, t)$. For such an indivudal, $N_{ik}^{\#}(t)$ and $Y_{ik}^{\#}(t)$ are computable from the observed data up to time t.
  - Although, $N_{ik}^{\#}(t)$ and $Y_{ik}^{\#}(t)$ are not computable for $r_i(t) = 0$, $r_i(t) N_{ik}^{\#}(t)$ and $r_i(t) Y_{ik}^{\#}(t)$ are computable.

Fine and Gray's score function

- With, $G(t) = \text{pr}(C \geq t)$, the survival function of the censoring, define

$$w_i(t) = \frac{r_i(t)G(t)}{G(T_i \wedge t)}.$$

- For a censored event time, $T_i^* > C_i$,

$$w_i(t) = \frac{r_i(t)G(t)}{G(T_i \wedge t)} = \frac{I(C_i \geq T^* \wedge t)G(t)}{G(T_i^* \wedge C_i \wedge t)} = \frac{I(C_i \geq t)G(t)}{G(C_i \wedge t)} = I(C_i \geq t)$$

- Assume that $C \perp\!\!\!\perp X$, then

$$\begin{aligned} E\left(\frac{r_i(t)G(t)}{G(T_i \wedge t)} \middle| T_i^*, \varepsilon_i, X_i \right) &= G(t)E\left(\frac{I(C_i \geq T_i^* \wedge t)}{G(T_i \wedge t)} \middle| T_i^*, \varepsilon_i, X_i \right) \\ &= G(t)E\left(\frac{I(C_i \geq T_i^* \wedge t)}{G(T_i^* \wedge t)} \middle| T_i^*, \varepsilon_i, X_i \right) \\ &= G(t)\frac{E\left(I(C_i \geq T_i^* \wedge t) \middle| T_i^*, \varepsilon_i, X_i \right)}{G(T_i^* \wedge t)} = G(t) \end{aligned}$$

- Thus, $w_i(t)$ reduces to $I(C_i > t)$ when $T_i^* > C_i$ and it's conditional expectation is the same as that of $I(C_i > t)$.

Fine and Gray's score function

- $G(t)$ is unknown but can be estimated by the Kaplan-Meier estimator of the censoring $\hat{G}(t)$. Let

$$\hat{w}_i(t) = \frac{r_i(t)\hat{G}(t)}{\hat{G}(T_i \wedge t)}.$$

- $\hat{w}_i(t) Y_{ik}^{\#}(t)$ approximates the "sub-distribution risk set"
- We can estimate β by the solution to the estimating equation with estimated risk sets,

$$\sum_{i=1}^{n} \int_0^{\tau} \left(X_i - \frac{\sum_{j=1}^{n} \hat{w}_j(t) Y_{ik}^{\#}(t) X_j \exp(\beta^T X_j)}{\sum_{j=1}^{n} \hat{w}_j(t) Y_{ik}^{\#}(t) X_j \exp(\beta^T X_j)} \right) \hat{w}_i(t) dN_{ki}^{\#}(t) = 0$$

- The asymptotic properties of the model can be studied using empirical process theory (see Martinussen & Scheike p. 365-369). Some of the considered processes are not martingales, and thus the martingale central limit theorem isn't applicable to these.
- The covariance estimator is more complicated than that for the cause-specific Cox models, because estimation of the sub-distribution risk sets also contributes to the covariance.

Sub-distribution hazard ratios

- The Fine and Gray model gives a direct relation between covariates and the risk, but the quantitative meaning of the "sub-distribution hazard ratios" e^{β_j} is not simple.
- The hazard function has the useful "epidemiological rate" interpretation

$$dA(t) = \text{pr}(\text{“die before } t + dt\text{“}|\text{“alive at } t\text{“}),$$

and so has the cause-specific hazard

$$dA_1(t) = \text{pr}(\text{“die from cause 1 before } t + dt\text{“}|\text{“alive at } t\text{“}),$$

but the sub-distribution hazard

$$dA_1^{\#}(t) = \text{pr}(\text{“die from cause 1 before } t + dt\text{“}| \\ \text{“either alive at } t \text{ or dead from a competing cause by } t\text{“}),$$

has not.

Sub-distribution hazrard ratios

The Fine and Gray model,

$$F_1(t) = 1 - \exp\left(-A^{\#}(t)e^{\beta_1 X_1 + \ldots + \beta_p X_p}\right)$$

can be written as

$$\text{cloglog}(F_1(t)) = \log(A_0^{\#}(t)) + \beta_1 X_1 + \ldots + \beta_p X_p$$

where $\text{cloglog}(x) = \log(-\log(1-x))$. The qualitative meaning of the covariates is as expected

- $\beta_1 > 0$ implies that $F_1(t)$ increases with X_1 (for $X_2, \ldots, X_p$ held fixed)
- $\beta_1 < 0$ implies that $F_1(t)$ decreases with X_1
- $\beta_1 = 0$ means no associtation

Thus, the sign and significance test can be interpreted, but the quantitative interpretation of β is not simple as it describes effects on $\text{cloglog}(F_1(t))$.

Proportional odds sub-distribution : odds ratios

Other link functions $h(\cdot)$ are also possible,

$$h(F_1(t, X)) = \log(A_0^{\#}(t)) + \beta^T X.$$

Choosing $h(x) = \text{logit}(x)$ yields a logistic regression for type 1 events

$$\text{logit}\left(F_1(t)\right) = \log\left(\frac{F_1(t)}{1 - F_1(t)}\right) = \log(A_0^{\#}(t)) + \beta_1 X_1 + \ldots \beta_p X_p$$

- The odds for the event for a baseline ($X = (0, \ldots, 0)$) individual is the unspecificed function of time $A_0^{\#}(t)$ which can be estimated nonparametrically ...
- ... while the covariate effect e^{β_1} is the odds ratio for a one unit difference in X_1, that do not depend on time, just as in logistic regression
- The proportional odds assumption (the odds ratios do not depend on time) is a modelling assumption that needs to be checked (just as the proportional (sub-distribution) hazards assumption in the Fine and Gray and Cox models).

Absolute risk regression and censoring

- Absolute risk regression for type 1 events doesn't require modelling the risks of events of other types...
- ...but requires a model for the censoring distribution.
- The standard solution is to use Kaplan-Meier. This requires that the censoring (unconditionally) independent
- One can also use regression model to estimate the censoring distribution if it depends on covariates
- Dealing with left-truncation is possible, but more complicated than for the cause-specifc hazards. Left-truncation is not supported by the `cmprsk` package.

Summary

- Cause-specific hazards are driving the underlying dynamics
  - When assessing the local (in time) strength of a cause in the short interval $[t, t + dt)$, the other causes need not be taken into account (we are conditioning on no death from any cause before time t)
  - Are local in time and calculated conditional on being alive at time t. Therefore we can treat competing events as censorings.
  - Cause-specific hazards are easy to estimate using standard hazard-based methods (Nelson-Aalen, Cox regression, . . .)
- Absolute risks describe the fraction of the population that fails from given causes.
  - Are not local in time but cumulated and we cannot treat competing events as censorings.
  - A rather simple, unbiased estimator for the risk exists - the "Aalen-Johansen" estimator. Direct regression possible; Fine and Gray model, proportional odds.
  - Risks may al be estimated by "plugging in" results from cause-specific hazard models
- Both are useful (and both are needed?) for a complete description of the competing risks situation
- The covariate effects can be qualitatively different on the cause-specific hazards and cause-specific risks.

Summary

- In studies of all-cause mortality, there is a one-to-one correspondence between rates (hazard) and risk (probability of event). One may be obtained from the other.
- Without competing events, e.g. in studies of all-cause mortality, the absolute risk can be calculated as 1 minus the Kaplan-Meier curve
- In studies where the event will not happen to everyone in the population, this is no longer the case.
- The quantity estimated by 1 minus the Kaplan Meier estimator treating censorings from other causes as censorings is

$$1 - \exp(-A_1(t)),$$

  a quantity with no clear interpretation. It is the probability of a cause 1 event in the hypothetical world where all other causes have been removed without changing the hazard for type 1 events (unverifiable assumption of "independent risks").
- With competing risks, using "1-Kaplan-Meier for a single cause" is upwards biased for the absolute risk and should not be used for estimating risks
- The magnitude of the bias depends on the frequency of the competing events
- Questions like independence of events and what would happen if certain causes were removed may be important and interesting, but their answers rely on unverifiable assumptions.

Take-home message

Both cause-specific hazards and absolute risk functions are useful and needed for a complete description in the competing risks setting.

"Rates are good for etiological questions and risks for prediction" (Koller et. al., *Stat.Med.*, 2012)