Survival Analysis

Week 7: Competing risks

Section of Biostatistics

Recap: Survival data

Let T be a continuous time to event

- Cumulative distribution function $F(t) = \text{pr}(T \leq t)$
- Survival function $S(t) = \text{pr}(T > t)$
- 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,

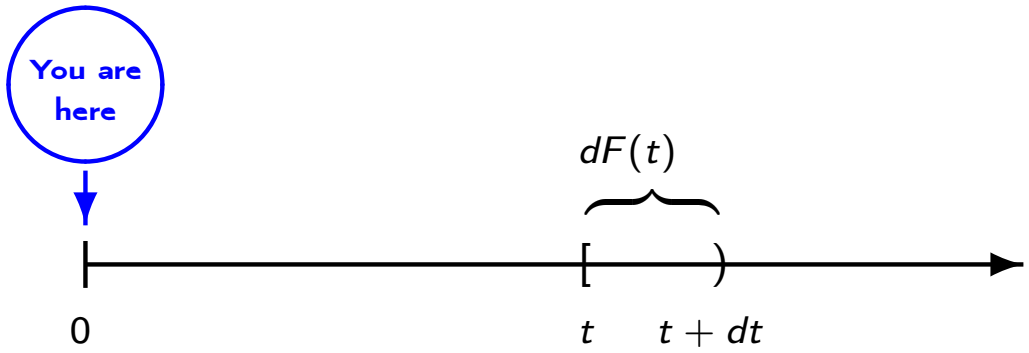
$$\begin{aligned}\alpha(t) &= \lim_{h \to \infty} \frac{\text{pr}(t \leq T < t + h | T \geq t)}{h} \\ &= \frac{\lim_{h \to \infty} \text{pr}(t \leq T < t + h)/h}{\text{pr}(T \geq t)} \\ &= \frac{f(t)}{S(t-)}\end{aligned}$$

Using Leibniz notation, for infinitely small dt, we write

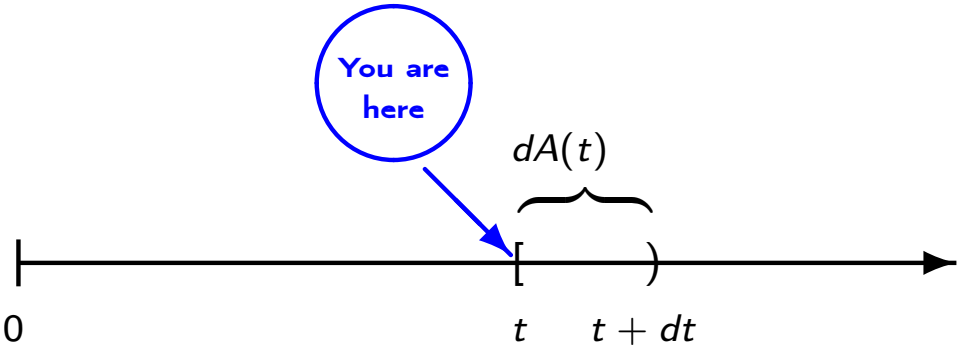
$$\begin{aligned}\alpha(t)dt = dA(t) &= \text{pr}(t \leq T < t + dt | T \geq t) \\ dF(t) &= \text{pr}(t \leq T < t + dt)\end{aligned}$$

Density vs. hazard

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



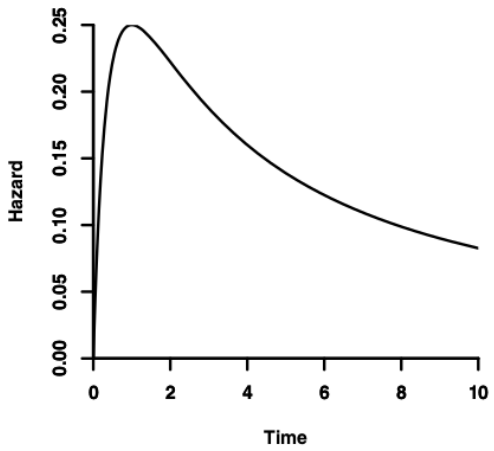
- The hazard is a conditional (on survival) rate: $dA(t) = \text{pr}(t \leq T < t + dt | T \geq t)$

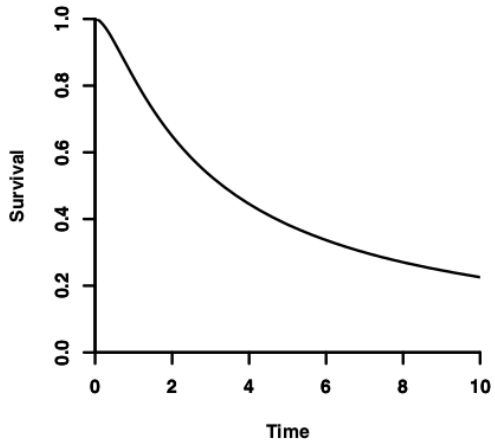


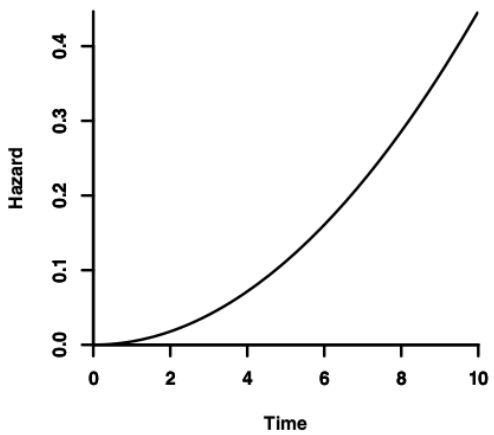
Hazard and survival

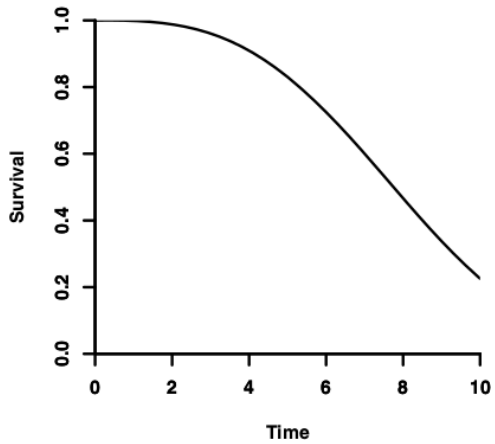
$$\alpha(s) = \frac{f(s)}{S(s)} = -\frac{\partial}{\partial s}\log S(s) \Leftrightarrow \int_0^t \alpha(s)ds = -\log S(t)$$

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



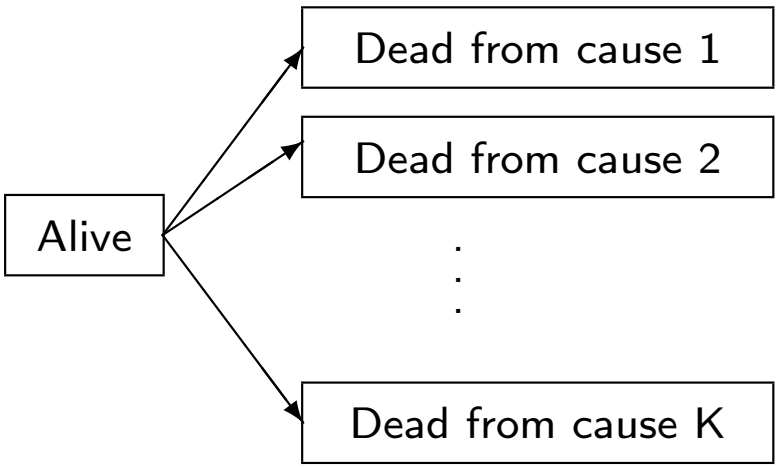






Competing risks

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
- Relapse or treatment related mortality in leukemia patients

Two meaningful ways to extract information from competing risks data

- Cause-specific hazard
- Cumulative incidence (sub-distribution)

Cause-specific hazard

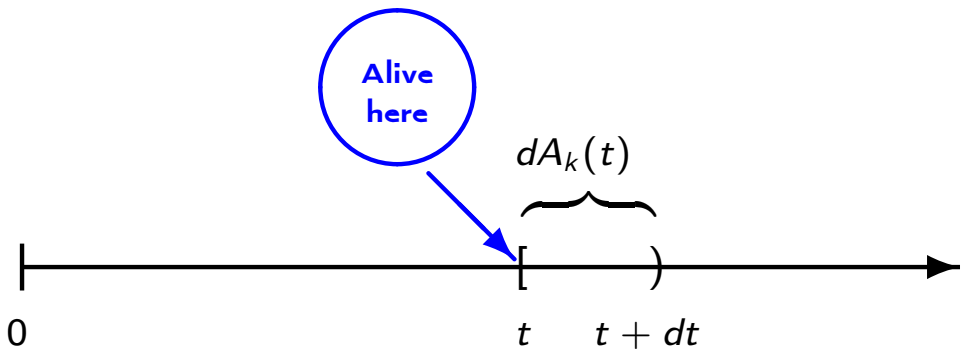
Let T be a survival time and let $\varepsilon \in \{1, \ldots, K\}$ denote the type of failure. The cause-specific hazard

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

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

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

Cumulative incidence

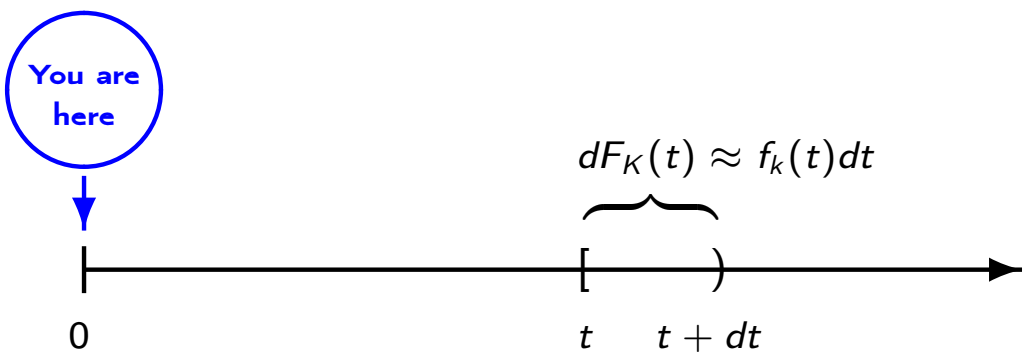
The cumulative incidence is the marginal probability of dying from cause k in the presence of other risks,

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

- 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

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

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

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

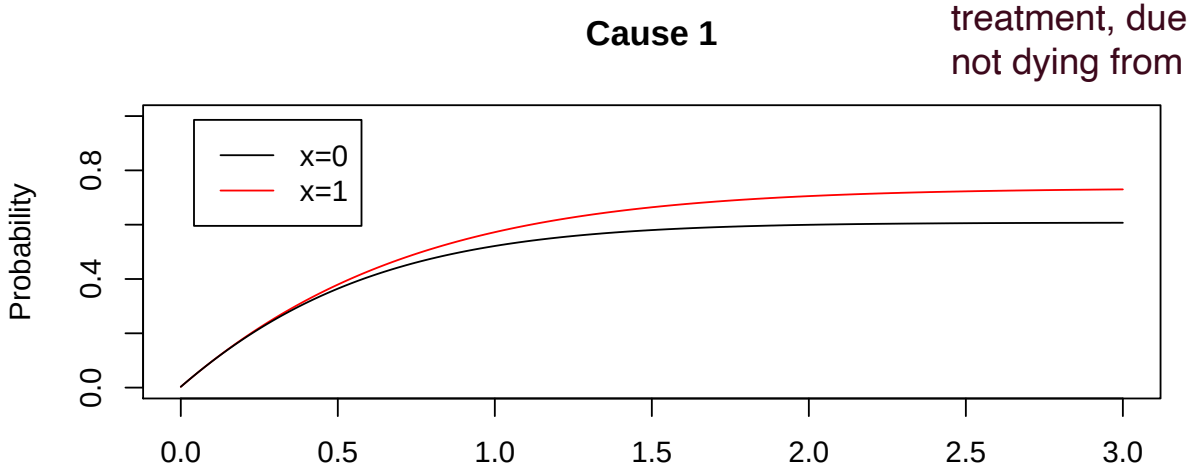
Then

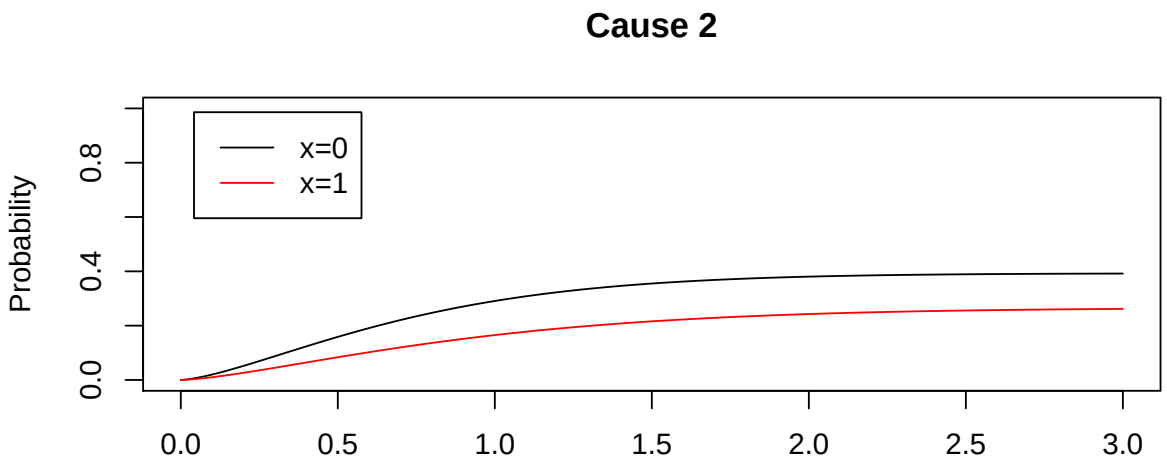
$$\begin{aligned}
F_1(t; x) &= \int_0^t \exp\left(-s - \frac{2^{1-x} s^{3/2}}{3}\right) ds \\
F_2(t; x) &= \int_0^t \exp\left(-s - \frac{2^{1-x} s^{3/2}}{3}\right) \frac{s^{1/2}}{2^x} ds
\end{aligned}$$

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

It is therefore seen to be more likely to die from cause 1 prior to time t, given treatment, due to more subjects not dying from cause 2.





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

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

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 get if only cause k was studied and all other causes were right-censorings.

Cause-specific hazards

The likelihood for α_k is the same (up to a constant) as if all other risks were right-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.
- All standard hazard-based models for survival data apply for inference on the cause-specific hazards, e.g,
  - Log-rank test for cause-specific hazards
  - Cox regression
  - Nelson-Aalen
  - Parametric regression
  - ...

  But be careful when considering probabilities/risks (more on this)
- Since the likelihood can be written entirely in terms of the α_k's, the cause-specific hazards and functionals of these (e.g. $\alpha_\bullet$, S and F_k) can be identified from the data without further assumptions.

The product-limit estimator

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

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

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.

Cause-specific hazards

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 $N_{\bullet\bullet}(t) = \sum_{i=1}^n \sum_{k=1}^K N_{ik}(t)$.

The cumulative incidence can be estimated by

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

Cause-specific hazards

- The asymptotic distribution of $\sqrt{n}(\hat{F}_1(t) - F_1(t))$ can be obtained 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

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

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

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

Multistate models

Consider the time-to-event data as a Markov process $X(t)$ with statespace $\{0, 1, \ldots, K\}$.

The transition probabilities, i.e., the probability that a subject is in state h at time t given that it was in state g at time s, are

$$P_{gh}(s,t) = \text{pr}(X(t) = h | X(s) = g)$$

and the transition intensities for $g \neq h$ are

$$a_{gh}(t) = \lim_{r \to 0} \frac{\text{pr}(X(t+r) = h | X(t) = g)}{r} = \lim_{r \to 0} \frac{P_{gh}(t, t+r)}{r}.$$

Collect the intensities in the matrix $\text{a}(t)$ with diagonal $-\sum_{g \neq h} a_{gh}(t)$.

Survival data as a two-state model

Transition from state 0 (*"alive"*) to state 1 (*"dead"*) is governed by the intensity $a_{01}(t) = \alpha(t)$ and

$$\mathsf{a}(t) = \begin{pmatrix} -\alpha(t) & \alpha(t) \\ 0 & 0 \end{pmatrix}.$$

The transition probability matrix is

$$\mathsf{P} = \begin{pmatrix} P_{00} & P_{01} \\ 0 & 1 \end{pmatrix}$$

such that

$$\mathsf{P}(0,t) = \begin{pmatrix} S(t) & 1 - S(t) \\ 0 & 1 \end{pmatrix}.$$

Two-cause competing risks as a multistate model

The transition intensity from state 0 (*"alive"*) to the state k (*"death from cause k"*) is $a_{0k}(t) = \alpha_k(t)$ and the transition intensity matrix is

$$\alpha(t) = \begin{pmatrix} -\alpha_1(t) - \alpha_2(t) & \alpha_1(t) & \alpha_2(t) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The transition probabilities from time 0 to time t are

$$\mathrm{P}(0,t) = \begin{pmatrix} S(t) & F_1(t) & F_2(t) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Multistate models

The instantaneous transition probabilities are

$$\mathsf{P}(u, u + du) \approx \mathsf{I} + \mathsf{a}(u)du$$

By the Markov property, for $0 = t_1 < \ldots < t_M = t$,

$$\mathsf{P}(0, t) = \mathsf{P}(t_0, t_1) \times \cdots \times \mathsf{P}(t_{M-1}, t_M)$$

Letting the partition become infinitely fine, yields the matrix valued product integral

$$\mathsf{P}(s, t) = \prod_{s<u\leq t} (\mathsf{I} + \mathsf{a}(u)du)\,. \tag{1}$$

The Aalen-Johansen estimator

- Let $N_{gh}(t)$ count the number of transitions from g to h, $g \neq h$, in $[0, t]$ and let $Y_g(t)$ count the number of individuals in state g and under observation at time t.
- The Nelson-Aalen estimate of $\int_0^t a_{gh}(s)ds$ is

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

- Expression (1) suggests estimating $P(0, t)$ by

$$\hat{\mathsf{P}}(0, t) = \prod_{u \leq t} \left(\mathsf{I} + d\hat{\mathsf{A}}(u) \right),$$

  where $\hat{\mathsf{A}}$ is the matrix of Nelson-Aalen estimators with diagonal $\hat{A}_{gg}(t) = -\sum_{h \neq g} \hat{A}_{gh}(t)$.
- $\hat{\mathsf{P}}$ is called the Aalen-Johansen estimator.

The Aalen-Johansen estimator

- Define $\mathrm{A}^*(t) = \{\int_0^t I(Y_g(t) > 0) dA_{gh}(u)\}_{gh}$ and let $\mathrm{P}^*(s,t) = \prod_{(s,t]} \{\mathrm{I} + d\mathrm{A}^*(u)\}$.

- By Duhamel's equation (see Gill-Johansen, 1990),

$$\hat{\mathrm{P}}(0,t)\mathrm{P}^*(0,t)^{-1} - \mathrm{I} = \int_0^t \hat{\mathrm{P}}(0,u-) d(\hat{\mathrm{A}} - \mathrm{A}^*)(u)\mathrm{P}^*(0,u)^{-1} \qquad (2)$$

  where the right-hand side is a zero-mean martingale.
- Thus, the empirical transition matrix is almost unbiased,

$$E\left(\hat{\mathrm{P}}(0,t)\mathrm{P}^*(0,t)^{-1}\right) = \mathrm{I}$$

The Aalen-Johansen estimator

We replace P^* by P in

$$\hat{\mathrm{P}}(0,t)\mathrm{P}^*(0,t)^{-1} - \mathrm{I} = \int_0^t \hat{\mathrm{P}}(0,u-)d(\hat{\mathrm{A}} - \mathrm{A}^*)(u)\mathrm{P}^*(0,u)^{-1} \qquad (3)$$

(they are equivalent for large-sample purposes) and multiply by $\mathrm{P}(0,t) = \mathrm{P}(0,u)\mathrm{P}(u,t)$. Asymptotically,

$$\hat{\mathrm{P}}(0,t) - \mathrm{P}(0,t) \approx \int_0^t \mathrm{P}(0,u-)d(\hat{\mathrm{A}} - \mathrm{A}^*)(u)\mathrm{P}(u,t).$$

Asymptotic normality of the elements of $\hat{\mathrm{P}}$ follows from the asymptotic distribution of the Nelson-Aalen estimator.

For a general K-state model, the covariance of the $K \times K$ matrix $\hat{\mathrm{P}}$ can be derived by similar arguments as for the Kaplan-Meier estimator, but the algebra is more involved and quite a bit of bookkeeping is required. The covariance matrix has dimension $K^2 \times K^2$.

Digression: Latent failure times

Consider "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.

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

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

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

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

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

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

nor the corresponding "net" hazard

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

without additional untestable assumptions on the joint distribution.

"Independent" competing risks

- Back in the days 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 times 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).

Sub-distribution hazard

As we have seen, there is no one-to-one relation between the cause-specific hazard α_k and the cumulative incidence 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

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

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

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

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

The instantaneous type 1 rate, given that the subject has not failed from cause 1, including those who have failed from other causes.

Notice the difference between the sub-distribution hazard

$$\alpha_1^{\#}(t) = \frac{f_1(t)}{1 - F_1(t)}$$

and the cause-specific hazard

$$\alpha_1(t) = \frac{f_1(t)}{1 - \sum_{k=1}^{K} F_k(t)} = \frac{f_1(t)}{S(t)}$$

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, 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 $\tilde{T}$,

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

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 estimate of $F_k(t)$ from before (Aalen-Johansen).

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 $N_{ik}^{\#}(t) = I(T^* \leq t, \epsilon_i = k)$ 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.
- 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)
- With $T_i = T_i^* \wedge C_i$, $\Delta_i = I(T_i^* \leq C_i)$, define the cause k "risk set" at time t as

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

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
  - If $r_i(t) = 1$, then $N_{k\mathbf{1}}^{\#}(t)$ 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.

Sub-distribution hazard ratios

- The estimates e^{β_j} are sub-distribution hazard ratios. The quantitative meaning of the regression coefficient 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.