Statistical methods in public health
Time scales and competing risks

Tommi Härkänen

National Institute for Health and Welfare (THL)
Department of Health (TERO)

September 15, 2015

Contents

Assumptions in survival models of week 2 lectures

- One time variable
- All individuals enter the study at time zero (baseline)
- Event of interest can occur only after baseline

Challenges: Last week only a subset of the Framingham data was used in the excercises. Why?

Young individuals have generally much lower risk of death than old individuals.

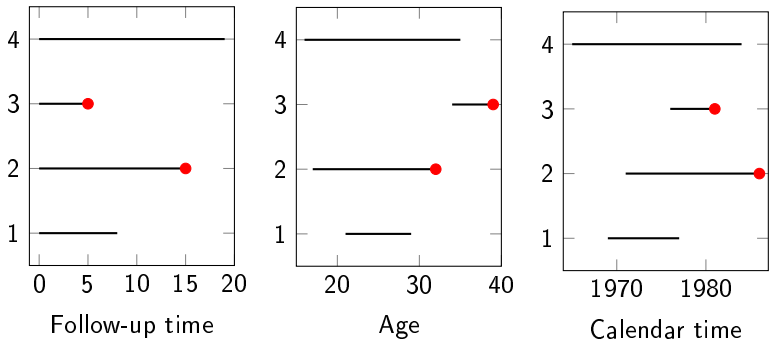
Time origins and time scales

Some examples:

Time origin	Time variable
Birth	Age
Onset of exposure	Exposure time
Entry to study	Follow-up time
Disease onset time	Time since disease onset
Diagnosis	Time since diagnosis
Start of treatment	Time since start of treatment
End of treatment	Time since end of treatment
⋮	⋮

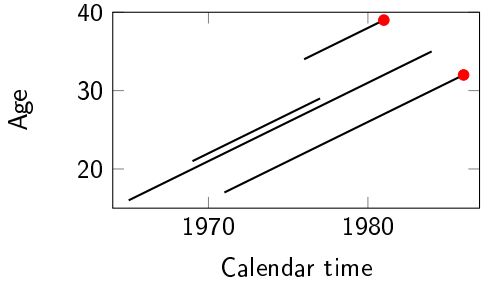
Survival data illustrated using different time scales

Four individuals: two observed failures and two right censored.

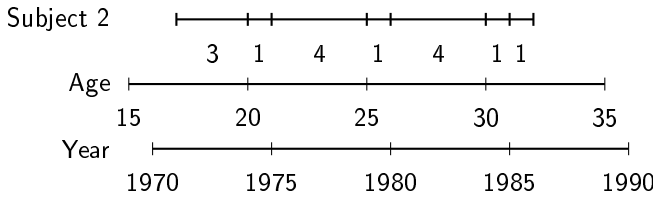


Lexis diagram

Two time scales simultaneously



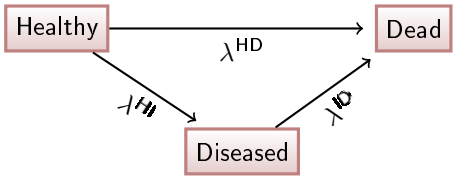
Data of subject 2 for five-year time bands:



Multi-state model

A common way to describe different events is by assuming a (small) number of different *states*, in one of which an individual i is at time t.

One of the simplest multi-state model is the *illness-death model*:



Modeling: For each *transition* we (can) have a different model.
After moving to state diseased, the model of death changes.

Risk sets: When an individual is in state healthy, he/she is *at risk* to move to state diseased or state dead.
After moving to state diseased, he/she is no longer at risk to move from state healthy to state dead.

Proportional hazards model

The hazard rate parameter λ_i for individual i can be parameterized as

$$\lambda_i = \exp\{X_i\beta\}, \tag{1}$$

where X_i is a vector of *covariate values* for individual i and β is a vector of corresponding *regression coefficients*. The inner product $X_i\beta$ is called the *linear predictor* (LP).

Interpretation of β_j (an element of β):

- If the corresponding covariate value in X_i increases by one ($x^*_{ij} := x_{ij} + 1$), then LP changes by $x_{ij}\beta_j$.
- The hazard rate changes by

$$\frac{\lambda^*_i}{\lambda_i} = \frac{\exp\{X^*_i\beta\}}{\exp\{X_i\beta\}} = \exp\{\beta_j\}$$

and $\exp\{\beta_j\}$ is called the *risk ratio* RR or *hazard ratio* HR.

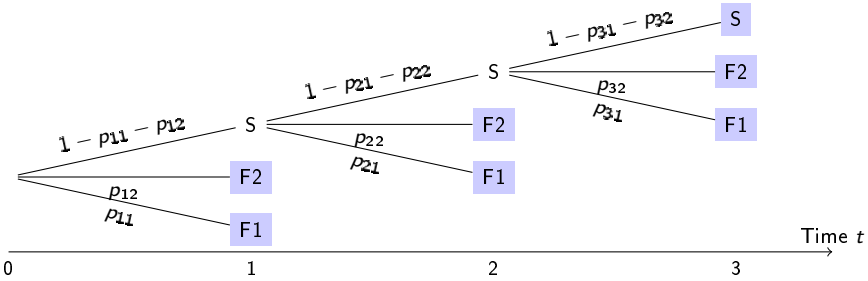
Two categories of censoring

Causes of failure There are other causes of failure than the one we are interested in. E.g. we are studying death from heart diseases, but individuals might die from cancer.
The individual is not in the risk set after the time of the event.

Drop-out from follow-up We are not able to continue the follow-up of some patients after some point in time. Causes of drop-out can depend on the risk factors or the outcome, in which case the estimates can be wrong.
The individuals remain in the risk set after the drop-outs.

Competing risks

If there are more than one event types, and any of them terminates the follow-up, the event types are *competing risks*.



Probabilities $p_{\cdot 1}$ and $p_{\cdot 2}$ are *cause-specific* failure probabilities for causes F1 and F2, respectively.

Cause-specific hazards

Likelihood

Data: End of follow-up time t_i, and a cencoring indicator δ_{i1}, δ_{i2}, ... for each cause of failure.
Parameters: Following the same reparameterization (and two causes) for $p_{t1} =: \lambda_{t1} h$ and $p_{t2} =: \lambda_{t2} h$ (as in lectures 2), we obtain cause-specific hazard rates λ_{t1} and λ_{t2}.

Likelihood can be approximated as before:

$$L(\lambda;\, t_i,\, \delta) = (\lambda_{t1} h)^{\delta_{i1}} (\lambda_{t2} h)^{\delta_{i2}} \exp\left\{ -\sum_{s=1}^{t_i} \lambda_{s1} h + \lambda_{s2} h \right\}$$

$$= \underbrace{\left[(\lambda_{t1} h)^{\delta_{i1}} \exp\left\{ -\sum_{s=1}^{t_i} \lambda_{s1} h \right\} \right]}_{\text{Cause 1}} \times \underbrace{\left[(\lambda_{t2} h)^{\delta_{i2}} \exp\left\{ -\sum_{s=1}^{t_i} \lambda_{s2} h \right\} \right]}_{\text{Cause 2}}. \quad (2)$$

Note that the likelihood factorizes into two parts thus the hazard rate parameters $\lambda_{\cdot 1}$ and $\lambda_{\cdot 2}$ can be estimated separately.

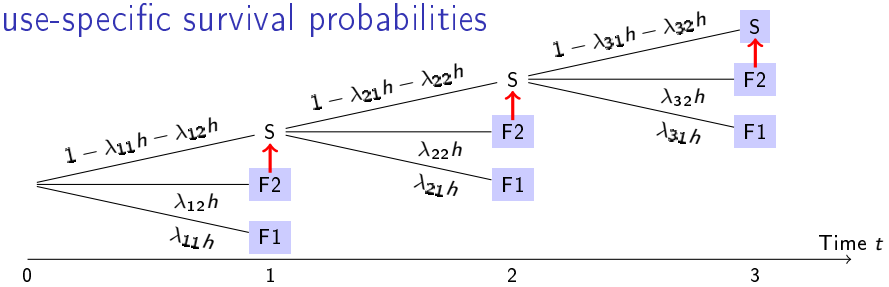
Independence of causes of death?

Likelihood (2) factorized into two parts, and in probability calculus this suggests that events are independent.

But

- If the causes of failure are different causes of death, then death from cancer inhibits death from a heart disease and vice versa.
  Different causes are almost completely dependent (negative association)!
- Individuals with advanced atherosclerosis have high risk for both myocardial infarction (MI) and stroke.
  If stroke incidence (hazard rate) could be reduced by some preventive measures, it is likely that the high risk individuals would face MI a bit later
  ⇒ Lower stroke incidence but higher MI incidence.

Cause-specific survival probabilities



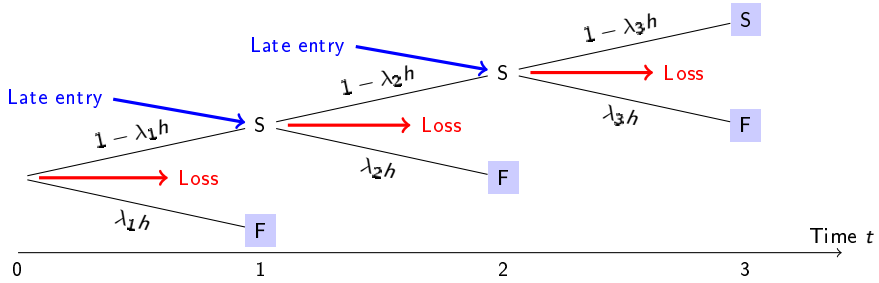
It is common to calculate cause-specific survival probabilities $S_k(t) := \exp\{-\sum_s^t \lambda_s h\}$ for cause k.
Usually $S_k(\cdot)$ is interpreted as "if all other causes could be avoided then the survival probability at time t is $S_k(\cdot)$".

Problem: This interpretation assumes that the causes are independent. In this case elimination of failure F2 at time t should not alter failure rates after time t.
The example on MI and stroke above illustrated that elimination of one cause may well increase incidence later on!

Late entry

Dynamic cohort

There can be an ideal starting point for a follow-up, e.g. onset of a disease or a treatment.
Sometimes part of the individuals are recruited at a later stage, e.g. all cancer patients being treated in a hospital at a certain date and new cases after that date.
Problem: The patients being treated are the survivors. It is possible that patients with severe symptoms have died away before the starting date.
E.g. the *healthy worker* effect.



Assumptions on censoring mechanism

Closed cohort

Some individuals Cannot be followed until the event of interest occurs. Censoring can be completely random (independent from the analysis variables).

Problem: A selection mechanism causing the drop-out can depend on the risk factors.
E.g. individuals with good physical health stay in the follow-up, but individuals with weaker functional capacity drop out, and part of new disease cases are unobserved.