Max Planck Odense Center on the Biodemography of Aging

University of Southern Denmark

Competing risks

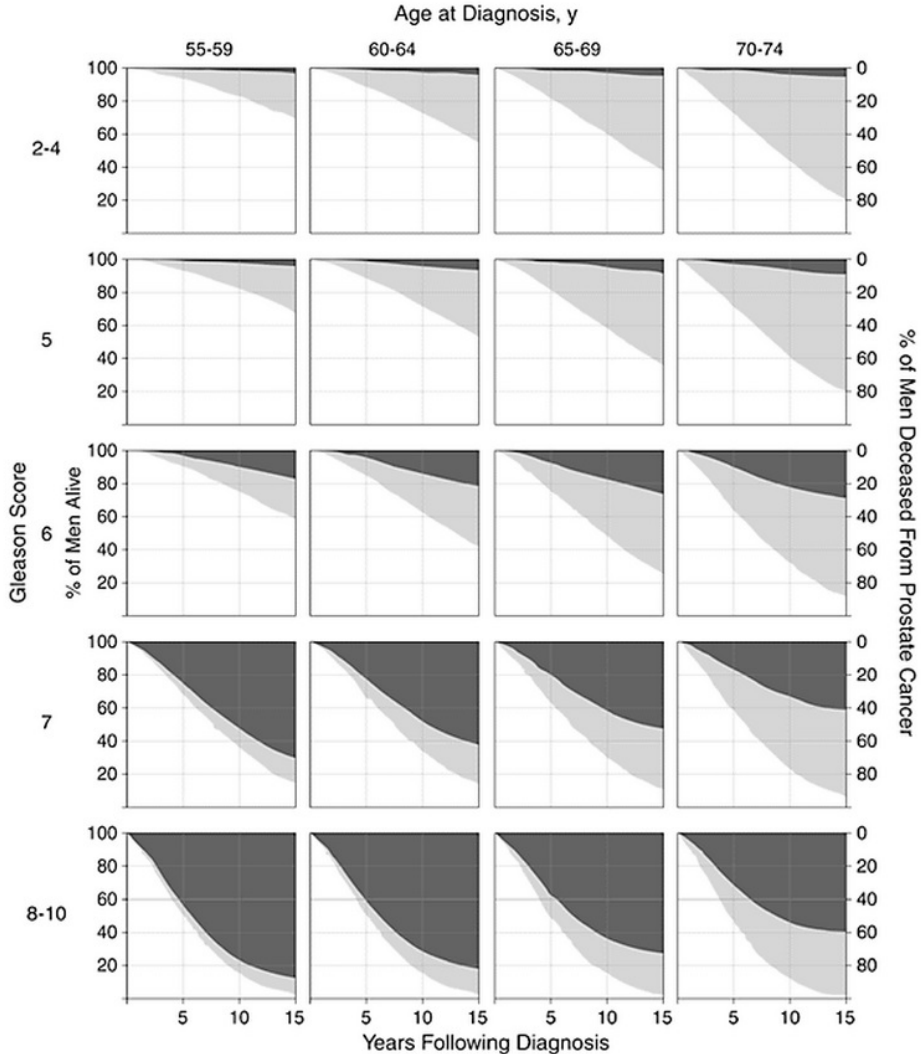
Virginia Zarulli, Adam Lenart

MaxO - SDU

Competing risks

- Analysis of multiples causes of failure;
- Study the relationship between a vector of covariates x and the rate of occurrence of specific types of failure;
- Estimate the risk of one type of failure after removing the others (only under the strong assumption that the competing risks are independent).

Competing risks



Survival (white lower band) and cumulative mortality from prostate cancer (dark gray upper band) and other causes (light gray middle band) up to 15 years after diagnosis stratified by age at diagnosis and Gleason score.
(Albertsen PC, Hanley JA, Gleason DF, Barry MJ. Competing Risk Analysis of Men Aged 55 to 74 Years at Diagnosis Managed Conservatively for Clinically Localized Prostate Cancer. JAMA. 1998;280(11):975-980. doi:10.1001/jama.280.11.975.)

Competing risks

Let T be a continuous random variable representing the survival time; failure can occur from one of m types of failure (causes), indexed by $j_1, j_2, \ldots, j_m$; Let J be a random variable representing the type of failure; x is a vector of covariates.

Cause specific hazard:

$$h_j(t,x) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, J = j | T \geq t, x)}{\Delta t}$$

Overall hazard:

$$h(t,x) = \sum_{j=1}^{m} h_j(t,x)$$

Competing risks

The overall survival and cumulative hazard functions are defined as usual:

$$S(t,x) = e^{-\int_0^t h(u,x)du} = e^{-H(t,x)}$$

When we have multiple causes:

$$S_j(t,x) = e^{-\int_0^t h_j(u,x)du} = e^{-H_j(t,x)}$$

$S(t,x)$ represents the probability of surviving all types of failure up to time t:

$$S(t,x) = \prod_{j=1}^{m} S_j(t,x)$$

Competing risks

$S_j(t,x)$, however, will not have a classical survival function interpretation if $m > 1$, because one individual cannot die more than once.

So the cause specific density function, the unconditional risk that a subject dies at time t from cause j, is:

$$f_j(t,x) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, J = j | x)}{\Delta t} = h_j(t,x) S(t,x)$$

Overall density of deaths:

$$f(t,x) = \sum_{j=1}^{m} f_j(t,x)$$

Competing risks

The likelihood function:

Consider n observations, for which we know (t_i, d_i, j_i, x_i)

$$L = \prod_{i=1}^{n} h_{j,i}(t_i, x_i)^{d_i} S(t_i, x_i) = \prod_{i=1}^{n} h_{j,i}(t_i, x_i)^{d_i} \prod_{j=1}^{m} S_j(t_i, x_i)$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} h_{j,i}(t_i, x_i)^{d_{i,j}} S_j(t_i, x_i)$$

The overall lilkelihood is a product of the m likelihoods for each type of failure;

The likelihood for a specific type of failure is the same likelihood that would be obtained by considering all other types of failures as censored observations.

Competing risks

Estimation in practice:

- Because all that we observe is the minimum failing time of the many potential latent times (one for each cause of failure), we cannot study the relationship among failure models and we cannot test for their independence (the assumption of independence between the competing risks is not testable).
- Under this assumption, then, all we obtain is the survival from cause $j - th$ after having removed all the other causes.

Competing risks

Estimation in practice:

- Using the techniques that we have studied so far (parametric, Cox...), all we can estimate are the $j-th$ cause-specific parameters of the baseline hazards and the $j-th$ cause-specific effect of the covariates, by treating the failures fo causes other than j as censored cases.
- If we want, we can use different covariates for each type of failure.

Cumulative incidence function

When you introduce covariates in the competing risks setting (for example with a Cox model), the focus is on the covariates of the risk of each type of event.

However, a covariate may appear to increase the risk of incidence of a certain type of failure just by lowering the occurrence of the other types of failure, even if, in reality, it has no effect on the event in question.

Another approach to the analysis of covariates is to focus on the probability of each type of event.

Cumulative incidence function

Cumulative incidence function (CIF): the probability that an event of type j occurred by time t, accounting for the fact that failure can occur also from other causes other than j.

$$I_j(t) = \int_0^t f_j(u)du = \int_0^t S(u)h_j(u)du$$

I_j increases with time, as more and more events of type j occur.

Cumulative incidence function

Non-parametric CIF:

With d_{ij} equal to the number of failures of type j at time t_i, n_i equal to the number of observations at risk at time t_i and $\hat{S}(t_i)$ equal to the Kaplan-Meier estimator up to time t_i:

$$\hat{I}_j(t) = \sum_{i:t_i \leq t} \hat{S}(t_i) \frac{d_{ij}}{n_i}$$

The non-parametric estimator of CIF is very convenient, because its complement can be formally treated as a survival function (showed by Fine and Gray) and, therefore, used to calculate the underlying cause-specific hazard (called *sub-hazard*) and estimate a proportional hazard model for it.

Fine and Gray model

Sub-hazard, instantaneous risk of dying from cause j given that the subject has not died from other causes:

$$\bar{h}_j(t) = -\frac{d}{dt} log(1 - \hat{l}_j(t))$$

It is different from a proper hazard because, as risk set, it considers those alive at time t and all those who failed before t but due to causes other than j.

The Fine and Gray model is a way to model the cumulative incidence function via the convenient framework of the proportional hazard models:

$$\bar{h}_j(t, x) = \bar{h}_{j0}(t) e^{x\beta_j}$$

Stata syntax

The function `stcrreg` computes the sub-hazard ratios (SHR) and the predicted CIF.

The function `stcompet` computes nonparametrically the CIF.

Exercise

1. Read the data colon.dta in and stset them without specifying which competing risk you are looking at.
2. Describe the data and look at the coding of the status variable.
3. Overall K-M function and save the plot
4. Generate the variable for the overall KM curve

Exercise

2. For the cumulative incidence function we use the command `stcompet`. To install it type `findit stcompet`.
3. Stset the data specifying death from cancer as event of interest (death from other causes will be the competing risk).
4. Calculate the CIF for the cause of interest and for the each named other competing risk: `stcompet cif=ci, compet(2)`
5. Plot the incidence of death from cancer (our cause of interest) and save it
```
twoway line cif _t if status==1, sort c(J)
export graph cifcancer.png
```
6. Plot and save the incidence of death from other causes
```
twoway line cif _t if status==2, sort c(J)
graph export cifother.png
```

Exercise

10. Stack the three estimates on top of each other using the relationship S(t)+I1(t)+I2(t)=1. How?
    - Sort the variable _t, in order to have ordered cif values.
    - Extract the cif for cancer deaths and fill in the blanks for other types of failures

      ```
      gen cifcancer = cif if status==1
      replace cifcancer=cifcancer[_n-1] if
      missing(cifcancer)
      ```
    - Generate the boundaries of the plot which are 1, 1-I1(t), S(t) and 0

      ```
      gen b1 = 1
      gen b2 = 1-cifcancer
      gen b3 = km
      gen b4 = 0
      sort _t
      ```

Exercise

11. Make the plot. How?

```
 twoway rarea b1 b2 _t, color(blue) c(J)
///
|| rarea b2 b3 _t, color(red) c(J) ///
|| rarea b3 b4 _t, color(green) c(J)
legend(off) ///
text(0.8 150 "Cancer") text(0.3 200 "Other
causes") text(0.1 50 "Alive") ///
xtitle(Months)
```

12. Fit the Fine and Gray model focussing on cancer, treating other causes as competing risk

```
stcrreg i.agegr i.stage, compete(status==2)
```

13. Predict CIF for cancer by stage

```
stcurve, cif at1(stage=1) at2(stage=0)
at3(stage=3)
```