Survival Analysis

4. Competing Risks

Germán Rodríguez

Princeton University

February 26, 2018

Introduction

We now turn to multiple causes of failure in the framework of competing risks models. IUD users, for example, could become pregnant, expel the device, or request its removal for personal or medical reasons.

Competing risks pose three main analytic questions of interest

1. How covariates relate to the risk of specific causes of failure, such as IUD expulsion
2. Whether people at high risk of one type of failure are also at high risk of another, such as accidental pregnancy
3. What would survival look like if a cause of failure could be removed, for example if we could eliminate expulsion

It turns out we can answer question 1, but question 2 is essentially intractable with single failures, and 3 can only be answered under strong and wholly untestable assumptions.

Cause-Specific Risks

Let T denote survival time and J represent the type of failure, which can be one of $1, 2, \ldots, m$.

We define a cause-specific hazard rate as

$$\lambda_j(t) = \lim_{dt \downarrow 0} \frac{\Pr\{T \in [t, t+dt), J = j | T \geq t\}}{dt}$$

the instantaneous conditional risk of failing at time t due to cause j among those surviving to t.

With mutually exclusive and collectively exhaustive causes the overall hazard is the sum of the cause-specific risks

$$\lambda(t) = \sum_{j=1}^{m} \lambda_j(t)$$

This result follows directly from the law of total probability and requires no additional assumptions.

Cumulative Hazard and Survival

We can also define a cause-specific cumulative hazard

$$\Lambda_j(t) = \int_0^t \lambda_j(u)du$$

which obviously adds up to the total cumulative hazard $\Lambda(t)$.

It may also seem natural to define the function

$$S_j(t) = e^{-\Lambda_j(t)}$$

but $S_j(t)$ does not have a survival function interpretation in a competing risks framework without strong additional assumptions.

Obviously $\prod S_j(t) = S(t)$, the total survival. This suggests interpreting $S_j(t)$ as a survival function when the causes are independent, but as we'll see this assumption is not testable.

Demographers call $S_j(t)$ the associated single-decrement life table.

Cause-Specific Densities

Finally, we consider a cause-specific density function which combines overall survival with a cause specific hazard:

$$f_j(t) = \lim_{dt \downarrow 0} \frac{\Pr\{T \in [t, t+dt), J = j\}}{dt} = \lambda_j(t) S(t)$$

the unconditional rate of type-j failures at time t. By the law of total probability these densities add up to the total density $f(t)$

In order to fail due to cause j at time t one must survive *all* causes up to time t. That's why we multiply the cause-specific hazard $\lambda_j(t)$ by the overall survival $S(t)$.

Our notation so far has omitted covariates for simplicity, but extension to covariates is straightforward. With time-varying covariates, however, a trajectory must be specified to obtain the cumulative hazard or survival.

The Incidence Function

Another quantity of interest is the cumulative incidence function (CIF), defined as the integral of the density

$$I_j(t) = \Pr\{T \le t, J = j\} = \int_0^t f_j(u) du$$

In words, the probability of having failed due to cause j by time t.

A nice feature of the cause-specific CIFs is that they add up to the complement of the survival function. Specifically

$$1 - S(t) = \sum_{j=1}^{m} I_j(t)$$

which provides a decomposition of failures up to time t by cause.

The CIF is preferred to $S_j(t)$ because it is observable, while the latter "has no simple probability interpretation without strong additional assumptions" (K-P, 2002, p. 252.)

Non-Parametric Estimation

Let t_i denote the failure or censoring time for observation i and let $d_{ij} = 1$ if individual i fails due to cause j at time t_i. A censored individual has $d_{ij} = 0$ for all j.

The Kaplan-Meier estimate of overall survival is obtained as usual

$$\hat{S}(t) = \prod_{i:t_j \le t} (1 - \frac{d_i}{n_i})$$

where $d_i = \sum_j d_{ij}$ is the total number of failures at t_i and n_i is the number of individuals at risk just before t_i.

The Nelson-Aalen estimate of the cumulative hazard of failure due to cause j is

$$\hat{\Lambda}_j(t) = \sum_{i:t_i \le t} \frac{d_{ij}}{n_i}$$

a sum of cause-specific failure probabilities. This estimate is easily obtained by censoring failures due to any cause other than j

Estimating the CIF

What you should *not* do is calculate a Kaplan-Meier estimate where you censor failures due to all causes other than j. You'll get an estimate, but it is not in general a survival probability.

What you *can* do is estimate the cumulative incidence function

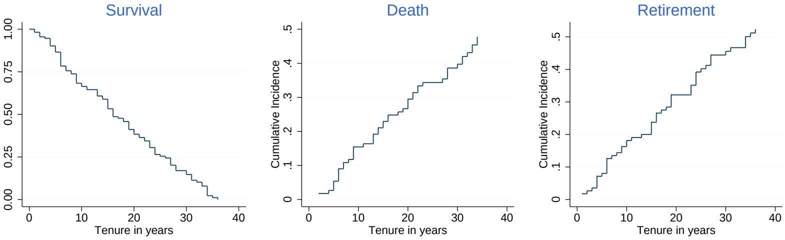
$$\hat{I}_j(t) = \sum_{i:t_i \le t} \hat{S}(t_i) \frac{d_{ij}}{n_i}$$

using KM to estimate the probability of surviving to t_i and d_{ij}/n_i for the conditional probability of failure due to cause j at time t_i.

Pointwise standard errors of the CIF estimate can be obtained using the delta method, but the derivation is more complicated than in the case of Greenwood's formula.

In the computing logs we study how long Supreme Court Justices serve, treating death and retirement as competing risks. The nine current justices are censored at their current (updated) length of service.

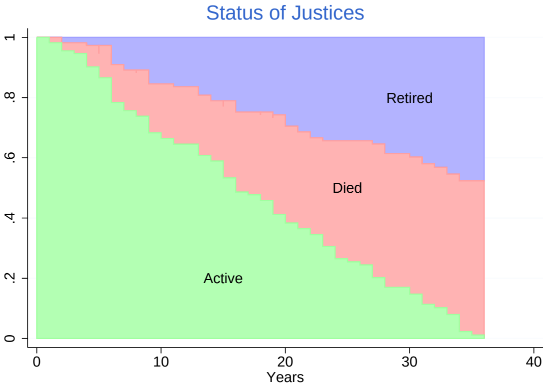
The graphs below show the Kaplan-Meier survival curve and the cumulative incidence functions for death and retirement



The median length of service is 16.5 years. The CIF plots have similar shapes, and indicate that about half the justices leave by death and the other half retire.

Supreme Court Justices (continued)

I like to stack these plots, taking advantage of the fact that $1 - S(t) = \sum_j I_j(t)$, so we can see at a glance the status of the justices by the years since they were appointed.



We now turn to regression models to see how these probabilities vary by age and period.

A natural extension of proportional hazard models to competing risks writes the hazard of type-j failures as

$$\lambda_j(t|x) = \lambda_{0j} e^{x'\beta_j}$$

where λ_{0j} is the baseline hazard and $e^{x'\beta_j}$ the relative risk, both for type-j failures.

The baseline hazard may be specified parametrically, for example using a Weibull or Gompertz hazard, or may be left unspecified, as we did in Cox models, which focus on the relative risks.

The most remarkable result is that these models may be fitted using the techniques we already know! All you do is treat failures of cause j as events and failures due to any other cause as censored observations.

The next two slides justify this remark.

Parametric Likelihoods for Competing Risks

The parametric likelihood for failures of type j in the presence of all other causes has individual contributions given by

$$d_{ij} \log \lambda_j(t_i|x) - \Lambda(t_i|x)$$

where I assumed for simplicity that observation starts at zero.

The cumulative hazard for all causes is a sum of cause-specific hazard, so we can write

$$d_{ij} \log \lambda_j(t_i|x) - \Lambda_j(t_i|x) - \sum_{k \neq j} \Lambda_k(t_i|x)$$

If the hazards for the other causes involve *different* parameters they can be ignored. What's left is exactly the parametric likelihood we would obtain by censoring failures due to causes other than j.

The cause-specific hazards can then be used to estimate overall survival and cause-specific incidence functions.

Partial Likelihood for Competing Risks

The construction of a partial likelihood follows the same steps as before. We condition on the times at which we observe failures of type j and calculate the conditional probability of observing each failure given the risk set at that time. With no ties this is

$$\frac{\lambda_{0j}(t_i)e^{x_i'\beta_j}}{\sum_{k\in R_i}\lambda_{0j}(t_i)e^{x_k'\beta_j}} = \frac{e^{x_i'\beta_j}}{\sum_{k\in R_i} e^{x_k'\beta_j}}$$

Once again the baseline hazard cancels out and we get an expression that depends only on β_j. Moreover, this is exactly the same partial likelihood we would get by treating failures due to other causes as censored observations.

The hazards in the model reflect risks of failures of one type in the presence of all the other risks, so no assumption of independence is required. It is only if you want to turn them into counterfactual survival probabilities that you need a strong additional assumption.

Cox Models for the Supreme Court

In the computing logs I fit Cox models to estimate age and period effects on Supreme Court tenure, using simple log-linear specifications. Here's a summary of hazard ratios for each cause.

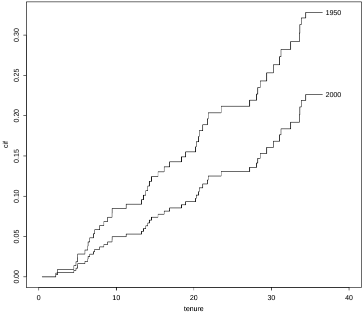
Predictor	All	Death	Retire
Age	1.084	1.071	1.106
Year	0.994	0.989	0.999

- The risk of leaving the court is 8% higher for every year of age and about half a percent lower per calendar year
- The risk of death is about 7% higher per year of age and has declined just over one percent per calendar year
- The risk of retirement is about 10% higher per year of age and shows essentially no trend by year of appointment

Can we turn these estimates into meaningful probabilities? Yes!

Incidence Functions from Cox Regression

In the computing logs I use the hazards of death and retirement to estimate cumulative incidences of death and retirement by tenure. The figure below shows the CIF of death for justices appointed at age 55 in 1950 and 2000.



The probability of dying while serving in the court has declined from 32.8% to 22.6% over the last 50 years, largely as a result of declines in mortality with no trend in retirement.

The Fine-Gray Model

Fine and Gray (1999) proposed a competing risks model that focuses on the incidence function for events of each type.

Let $I_j(t|x)$ denote the incidence function for failures of type j, defined as

$$I_j(t|x) = \Pr\{T \leq t, J = j|x\}$$

the probability of a failure of type j by time t given x.

The complement or probability of not failing due to that cause can be treated formally as a survival function, with hazard

$$\bar{\lambda}_j(t|x) = -\frac{d}{dt}\log(1 - I_j(t|x)) = \frac{f_j(t)}{1 - I_j(t)}$$

We follow Fine-Gray in calling this a *sub-hazard* for cause j, not to be confused with the cause-specific hazard $\lambda_j(t|x)$.
This hazard is a bit weird (the authors say "un-natural") because the denominator reflects all those alive at t or long since dead of other causes.

The Fine-Gray Model (continued)

They then propose a proportional hazards model for the sub-hazard for type j, writing

$$\bar{\lambda}_j(t|x) = \bar{\lambda}_{0j}(t)e^{x'\beta_j}$$

where $\bar{\lambda}_{0j}(t)$ is a baseline sub-hazard and $e^{x'\beta_j}$ a relative risk for events of type j.

The model implies that the incidence function itself follows a glm with complementary log-log link

$$\log(-\log(1 - I_j(t|x))) = \log(-\log(1 - I_{j0}(t))) + x'\beta_j$$

where $I_{j0}(t)$ is a baseline incidence function for type-j failures.

In the end Fine and Gray argue that their formulation is just a convenient way to model the incidence function and I agree. Because the transformation is monotonic, a positive coefficient means higher CIF, but ascertaining how much higher requires additional calculations.

The Fine-Gray Results for Supreme Court

In the computing logs I fit the Fine-Gray model to the Supreme Court data, treating the risk of death and retirement as competing risks.

The table below shows the estimated age and year effects on the sub-hazard ratio (SHR) of death. I show exponentiated coefficients and a Wald test.

Predictor	SHR	z
Age	1.0074	0.42
Year	0.9916	-3.62

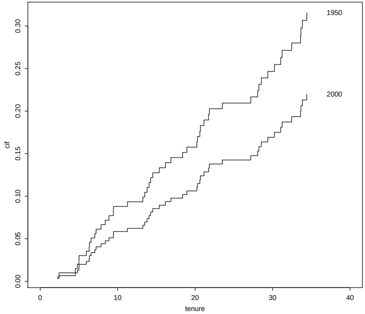
The cumulative incidence of death does not vary with age at appointment beyond what could be expected by chance, but it has declined with year of appointment with a significant linear trend.

To understand the magnitude of these effects we need to translate the sub-hazard ratios into something easier to understand, namely predicted cumulative incidence.

The Fine-Gray CIF for Supreme Court

In the computing logs I show how to obtain predicted CIF curves "by hand", so you can see exactly how it is done.

Here are the estimated CIF for death for justices appointed at age 55 in 1950 and 2000



We estimate that the probability of dying in the court for justices appointed at age 55 has declined from 31.6% to 22.0% over the last 50 years. The results are very similar to the Cox estimates.

The Identification Problem

A useful framework for understanding competing risks introduces latent survival times $T_1, T_2, \ldots, T_m$ representing the times at which failures of each type would occur, with joint distribution

$$S_M(t_1, \ldots, t_m) = \Pr\{T_1 > t_1, \ldots, T_m > t_m\}$$

The problem is that we only observe the *shortest* of these and its type: $T = \min\{T_1, \ldots, T_m\}$ and $J : T = T_j$.

To be alive at t all potential failure times have to exceed t, so the distribution of the observed survival time is

$$S(t) = S_M(t, t, \ldots, t)$$

Taking logs and partial derivatives we obtain the cause-specific hazards

$$\lambda_j(t) = \frac{\partial}{\partial t_j} \log S_M(t, t, \ldots, t)$$

These two functions can be identified from single-failure data, but the joint survival function cannot.

The Marginal Distributions

The marginal distribution of latent time T_j is given by

$$S_j^*(t) = \Pr\{T_j > t\} = S_M(0, \ldots, 0, t, 0, \ldots, 0)$$

and represents how long one would live if only cause j operated.

The hazard underlying this survival function is

$$\lambda_j^*(t) = -\frac{d}{dt}\log S_j^*(t) = -\frac{\partial}{\partial t}\log S_M(0, \ldots, 0, t, 0, \ldots, 0)$$

and represents the risk of failure if j was the only cause operating.

These functions are *not* identified. But if $T_1, T_2, \ldots, T_m$ are independent then

$$S_j^*(t) = S_j(t) \quad \text{and} \quad \lambda_j^*(t) = \lambda_j(t)$$

The assumption of independence, however, cannot be verified!

Illustrating the Identification Problem

In the notes I provide an analytic example involving two bivariate survival functions which produce the same observable consequences, yet the latent times are independent in one and correlated in the other.

An alternative approach uses simulation to illustrate the problem:

- Generate a sample of size 5000 from a bivariate standard log-normal distribution with correlation $\rho = 0.5$. (The underlying normals have means zero and s.d.'s one.) Let's call these variables t_1 and t_2.
- Set the overall survival time to $t = \min(t_1, t_2)$. Censoring is optional. Verify that the Kaplan-Meier estimate tracks $S(t, t)$.
- Compute a Kaplan-Meier estimate treating failures due to cause 2 as censored. Verify that this differs from the Kaplan-Meier estimate based on t_1, which tracks $S(t, 0)$. Unfortunately, t_1 is not observed.

Hint: To generate bivariate normal r.v.'s with correlation ρ make $Y_1 \sim N(0, 1)$ and $Y_2|y_1 \sim N(\rho y_1, 1 - \rho^2)$.