Analysis of time-to-event for observational studies: Guidance to the use of intensity models

On behalf of STRATOS TG8

Maja Pohar Perme

IBMI, University of Ljubljana

Applied statistics, Koper, 2023

The plan of this talk

Received: 30 March 2020 | Revised: 4 September 2020 | Accepted: 4 September 2020
DOI: 10.1002/sim.8757

TUTORIAL IN BIOSTATISTICS

Statistics in Medicine WILEY

Analysis of time-to-event for observational studies: Guidance to the use of intensity models

Per Kragh Andersen[1] | Maja Pohar Perme[2] | Hans C. van Houwelingen[3] | Richard J. Cook[4] | Pierre Joly[5] | Torben Martinussen[1] | Jeremy M. G. Taylor[6] | Michal Abrahamowicz[7] | Terry M. Therneau[8]

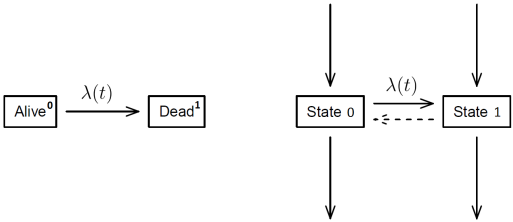
Overview paper:

- Basic ideas and pitfalls of survival analysis, organized as checklists
- Hazard models and beyond
- Illustrative example - patients with peripheral arterial disease

Survival analysis

Occurrence of a particular event in time

- $\lambda(t)$**: intensity (hazard)**
- **incomplete information: censoring or competing risk**



Intensity or hazard function

$$\lambda_i(t) \approx P(\textbf{event in } (t, t+dt) \mid \textbf{past at time } t-)/dt;$$

$$\lambda(t) = -\frac{d \log S(t)}{dt}$$

- **dynamic description of how events occur in time**
- **can be estimated directly (assuming independent censoring assumption)**
- **inclusion of time-dependent covariates**
- **taking account of delayed entry**
- **conditionally dependent censoring**

May be of interest in its own right, insufficient for some questions - absolute risk

Survival analysis - notation

Standard notation

- T_i: **follow-up time**
- δ_i: **censoring indicator**
- V_i: **entry time**
- $Z_i(t)$: **covariate vecor**

Counting process notation

For each individual i

- $Y_i(t)$: **at risk indicator. Drops from 1 to 0 in case of event or censoring. In case of delayed entry: can be 0 at** $t = 0$
- $N_i(t)$: **counting events. Jumps from 0 to 1 in case of event occurrence.**
- $Z_i(t)$: **covariate vector**

Our data example

Peripheral arterial disease

- **Common circulatory problem, narrowed arteries, sign of atherosclerosis, increased risk for CV (cardio-vascular) events**
- **742 PAD patients and 713 controls, Slovenia, 5 years of follow up**
- **Baseline data, measurements at each visit, endpoints**
- **Goal: survival of patients with PAD (in comparison to controls) despite optimal treatment**

Preliminary concepts and issues

In general:

- **Time origin:** unambigously defined, comparable, clinically relevant. Defines time axis, multiple time axes may be relevant
- **Inclusion criteria:** must be met by the time the patient enters the study ($Y(t)$ first becomes 1) - danger of immortal time bias
- **Event definition:** Clearly defined, the definition should be clear at the time of event (when $N(t)$ switches to 1) - danger of immortal time bias
- **Censoring:** We wish to estimate a complete, uncensored, population. Independent censoring assumption. Why was a patient censored?

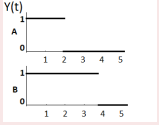
PAD example:

- **Time origin:** enrollment or birth, conditional survival in case of age as time axis.
- **Inclusion criteria:** PAD (and age-matched controls) at the time of enrollment. Ever or never PAD cannot be a criterium, time-varying covariate PAD could be
- **Event definition:** death (CV or non CV), major CV events(stroke, infarction), minor events (revascularization)
- **Censoring:** 5th visit after 5 years. Censored at 5 years. Non CV death as a competing risk.

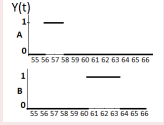
Preliminary concepts and issues - time origin

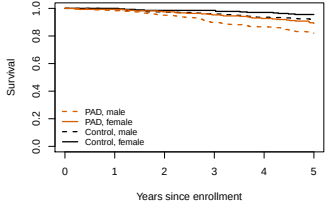
Unambigously defined, comparable, clinically relevant, depends on our scientific questions

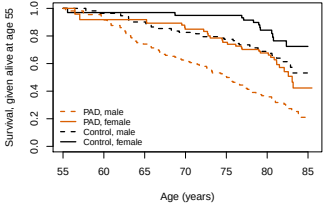
- **Enrollment. Time axis: time since enrollment**



- **Birth. Time axis: age**
- **Delayed entry, conditional survival**







Proportional hazards models

Cox PH model

$$\lambda(t|Z_i(t)) = \lambda_0(t)\exp(Z_i(t)^\top \beta)$$

- **Estimation: maximum partial likelihood**
- **Std. errors, tests as in classical likelihood**
- **Valid in simple and more general situation (factorization)**

Alternatives

- **Other PH models: parametric (constant, piecewise constant, Weibull, splines)**
- **Cox extensions: time-varying effects, stratified Cox**
- **Alternative models: additive hazards (Aalen), accelerated failure time (AFT) model**

Cox PH model

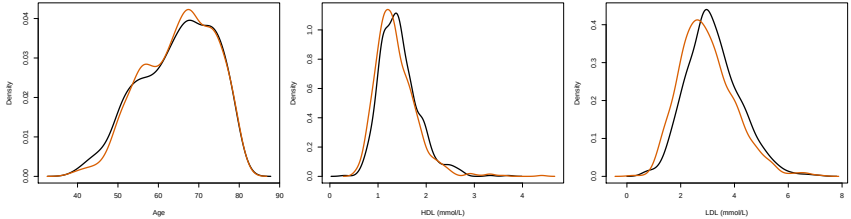
Before fitting the model

In general:

- **Check the covariates, check the dates**
- **Investigate covariate dependent censoring (Cox with censoring as the event): include such variables in the model**
- **Time-dependent covariates (extrapolation, external covariates, reverse causality bias)**

PAD example:

- **Covariates: PAD, Sex, Age, LDL, HDL**
- **Time-dependent covariates: carry last value forward**

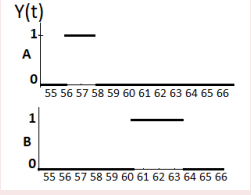


Immortal time bias

The values of $Z(t)$, $N(t)$ and $Y(t)$ should be defined so that they do not depend on $N(s)$, $Y(s)$ or $Z(s)$ for $s > t$

Examples in PAD

- **Age axis: do not forget about delayed entry. Otherwise Y depending on N at a higher age.**



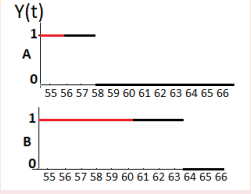
- **Some controls are diagnosed with PAD at later visits. Do not exclude them from the control group. Options:**
  - **PAD status can be time-fixed (value at enrollment)**
  - **Time-dependent (current value)**
  - **but NOT time-fixed at the value at the end of follow-up (ever PAD vs never-PAD). Example of Z depending on later values of itself**

Immortal time bias

The values of $Z(t)$, $N(t)$ and $Y(t)$ should be defined so that they do not depend on $N(s)$, $Y(s)$ or $Z(s)$ for $s > t$

Examples in PAD

- **Age axis: do not forget about delayed entry. Otherwise Y depending on N at a higher age.**



- **Some controls are diagnosed with PAD at later visits. Do not exclude them from the control group. Options:**
  - **PAD status can be time-fixed (value at enrollment)**
  - **Time-dependent (current value)**
  - **but NOT time-fixed at the value at the end of follow-up (ever PAD vs never-PAD). Example of Z depending on later values of itself**

Fitting the Cox PH model - PAD example, part I

Event - death of any cause

The effect of PAD and sex (m vs f) - which time axis?

- **Time since enrollment: add age (per 10 years, assume linearity)**
- **Age axis: add time since enrollment (FU, per year, assume linearity)**
- **Multiple axes: Poisson**

Time since enroll			Age axis			Both axes		
Cov	HR	95% CI	Cov	HR	95% CI	Cov	HR	95% CI
PAD	2.40	(1.71, 3.37)	PAD	2.40	(1.70, 3.37)	PAD	2.38	(1.70, 3.35)
Sex	2.00	(1.40, 2.86)	Sex	2.02	(1.42, 2.90)	Sex	2.01	(1.41, 2.88)
Age	1.93	(1.57, 2.37)	FU	1.18	(1.05, 1.33)			

Fitting the Cox PH model - PAD example, part I

Event - death of any cause

The effect of PAD and sex (m vs f) - which time axis?

- **Time since enrollment: add age (per 10 years, assume linearity)**
- **Age axis: add time since enrollment (FU, per year, assume linearity)**
- **Multiple axes: Poisson**

Time since enroll			Age axis			Both axes		
Cov	HR	95% CI	Cov	HR	95% CI	Cov	HR	95% CI
PAD	2.40	(1.71, 3.37)	PAD	2.40	(1.70, 3.37)	PAD	2.38	(1.70, 3.35)
Sex	2.00	(1.40, 2.86)	Sex	2.02	(1.42, 2.90)	Sex	2.01	(1.41, 2.88)
Age	1.93	(1.57, 2.37)	FU	1.18	(1.05, 1.33)			

Fitting the Cox PH model - PAD example, part I

Event - death of any cause

The effect of PAD and sex (m vs f) - which time axis?

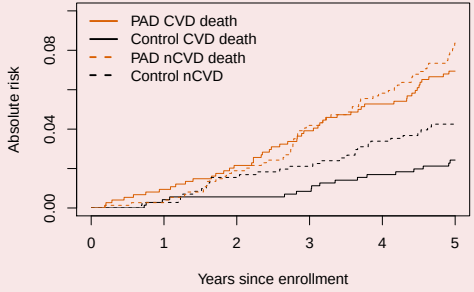
- **Time since enrollment: add age (per 10 years, assume linearity)**
- **Age axis: add time since enrollment (FU, per year, assume linearity)**
- **Multiple axes: Poisson**

Time since enroll			Age axis			Both axes		
Cov	HR	95% CI	Cov	HR	95% CI	Cov	HR	95% CI
PAD	2.40	(1.71, 3.37)	PAD	2.40	(1.70, 3.37)	PAD	2.38	(1.70, 3.35)
Sex	2.00	(1.40, 2.86)	Sex	2.02	(1.42, 2.90)	Sex	2.01	(1.41, 2.88)
Age	1.93	(1.57, 2.37)	FU	1.18	(1.05, 1.33)			

Competing risks analysis - PAD

Death of cardio-vascular reasons

- **Non-CV cause: competing risk, not censoring (present in the complete population, elimination not of interest)**
- **Estimate probabilities: Aalen-Johansen**



Fitting the Cox model - PAD example, part II

Competing risks

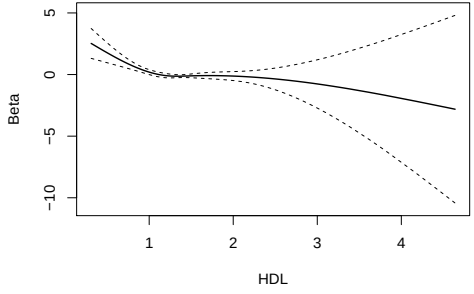
- Non-CV cause: can be treated at censoring in the Cox model (factorization of the likelihood)
- Time fixed or time-dependent covariates
- All CV causes (death + stroke, infarction)

	CV death Time-fixed	
	HR	95% CI
PAD	2.87	(1.65-5)
Sex (m vs. f)	1.67	(0.97-2.88)
Age (per10yrs)	1.93	(1.40-2.66)
HDL (mmol/l)	0.74	(0.39-1.41)
LDL (mmol/l)	0.92	(0.72-1.18)

After fitting the Cox model - PAD example, part III

Check assumptions

- **Proportional hazards, linearity (continuous variables)**
- **Many methods available: Schoenfeld residuals, martingale residuals**
- **What to do if violated: confounder or the variable of interest (omission of strong predictors!)**

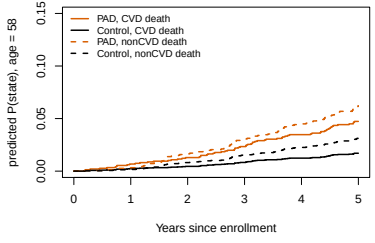


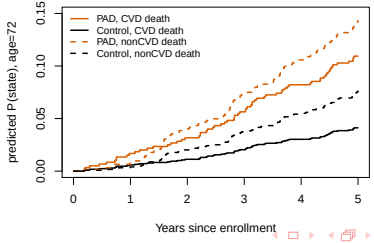
After fitting the Cox model - PAD example, part IV

Reporting and interpretation

- **If only HRs are reported - no absolute risks can be obtained**
- **Competing risks: hazard vs probability**
- **Absolute risks: prediction from $t = 0$ onwards**

	Time-fixed, other cause	
	HR	95% CI
PAD	2.04	(1.31–3.19)
Sex (m vs. f)	2.12	(1.29–3.50)
Age (per 10 yrs)	1.93	(1.45–2.56)
HDL	0.82	(0.43–1.55)
LDL	1.02	(0.83–1.26)





Concluding remarks

The subtitles in the paper

- **Preliminary concepts and issues**
- **The intensity**
- **Proportional hazard models and alternatives**
- **A check-list when fitting the Cox model**
- **Immortal time bias**
- **Prediction in the absence/presence of competing risks**
- **Issues in causal inference**
- **Illustrative applications + supplement with code**

Analysis of time-to-event for observational studies: Guidance to the use of intensity models

Per Kragh Andersen[1] | Maja Pohar Perme*[2] | Hans C. van Houwelingen[3] | Richard J. Cook[4] | Pierre Joly[5] | Torben Martinussen[1] | Jeremy M.G. Taylor[6] | Michal Abrahamowicz[7] | Terry M. Therneau[8] | for the STRATOS TG8 topic group