On some practical issues in the analysis of survival data

On behalf of STRATOS TG8

Maja Pohar Perme

IBMI, University of Ljubljana

STRATOS mini-symposium, ISCB, 2016

The plan of this talk

TG8 members

- **Michal Abrahamowicz**
- **Per Kragh Andersen**
- **Richard Cook**
- **Pierre Joly**
- **Torben Martinussen**
- **Maja Pohar Perme**
- **Jeremy Taylor**
- **Terry Therneau**

Outline

- **a few well known facts of survival analysis**
- **an outline of TG8 plan**
- **competing risks**

Survival analysis

Data evolving in time

Target population and censoring

- **Inference: parameters in the population, estimated on a sample**
- **Survival analysis: parameters in the population - complete data ($S(t)$, $h(t)$)**
- **Sample: censored data (incomplete data)**
- **The goal: drawing inference for population parameters based on incomplete data.**
- **Assumption: independent censoring**

Independent censoring

- **individuals censored at any given time t should not be a biased subsample of those who are at risk at time t.**
- **the extra information that the subject is not only alive, but also uncensored at time t does not change the hazard:**

$$h(t) \approx \frac{P(T^* \leq t + dt | T^* > t)}{dt} = \frac{P(T^* \leq t + dt | T^* > t, C > t)}{dt}$$

Inference with independent censoring

- **For independent observations (T_i, δ_i), where $T_i = min(T_i^*, C_i)$ and $\delta_i = I(T_i^* < C_i)$, the likelihood can be expressed via the hazard (and cumulative hazard H) functions:**

$$L(\theta) = \prod_i h_\theta(T_i)^{\delta_i} e^{-H_\theta(T_i)}$$

- **This is the basis for the Nelson-Aalen estimator for H and Cox partial likelihood**
- **Using the relations between S and H leads to Kaplan-Meier estimator**

TG8 plan

Level 2 papers

- **Single endpoint**
- **Multiple endpoints**

Level 2 papers

- **Avoiding pitfalls**
- **Checking assumptions**
- **Using state-of-the-art methods**

TG8 plan

Single endpoint

- **The censoring assumption**
- **Cox model - check PH, functional form**
- **time-varying covariates, time-dependent coefficients**
- **Alternatives to Cox models (AFT, cure models ...)**

Immortal time bias

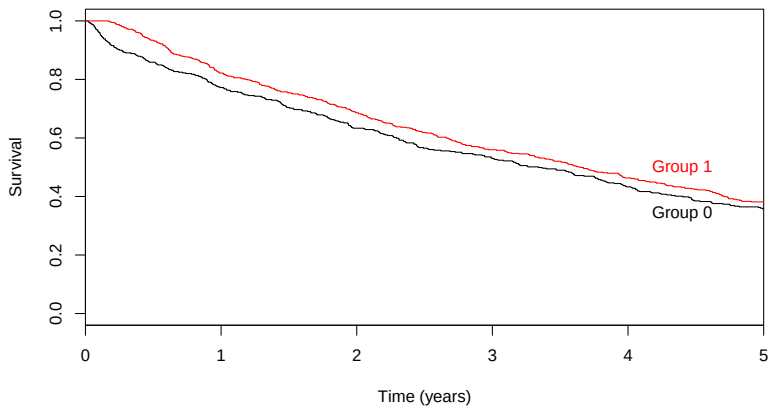
Patients on chemotherapy. Do treatment side effects improve the prognosis?

- **Time 0: start of chemotherapy. With time, some patients develop side effects.**
- **Available data: patients followed for 5 years, some developed side effects, some did not (0/1 variable)**

Immortal time bias

Patients on chemotherapy. Do treatment side effects improve the prognosis?

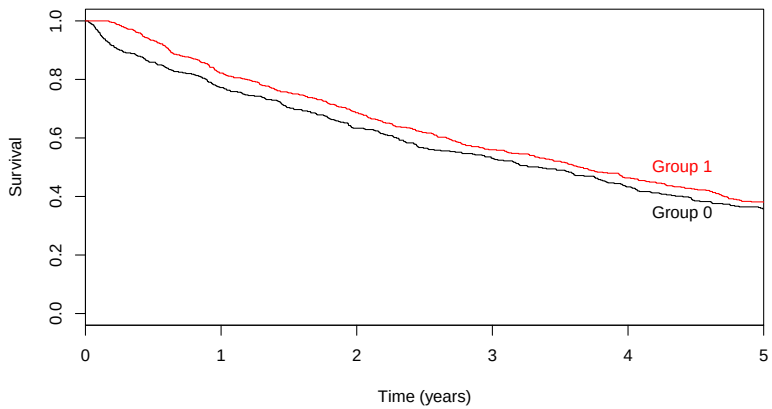
- **Time 0: start of chemotherapy. With time, some patients develop side effects.**
- **Available data: patients followed for 5 years, some developed side effects, some did not (0/1 variable)**



Immortal time bias

Patients on chemotherapy. Do treatment side effects improve the prognosis?

- **Time 0: start of chemotherapy. With time, some patients develop side effects.**
- **Available data: patients followed for 5 years, some developed side effects, some did not (0/1 variable)**



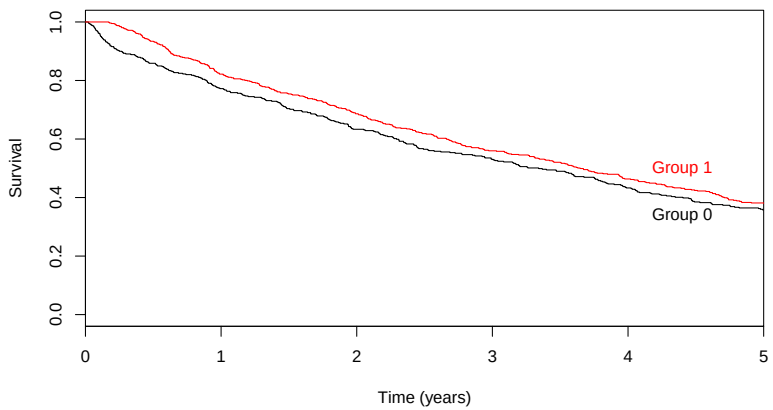
Patients with side effects immortal between time 0 and time of side effects

Immortal time bias

A time-varying covariate

Solution?

- **A hazard regression model with a time-varying covariate**

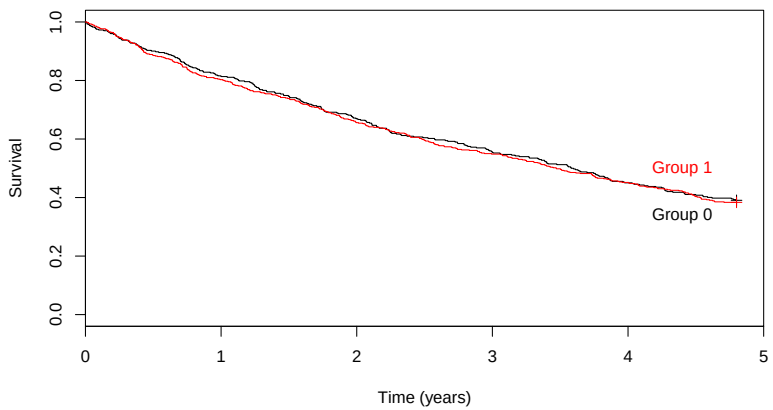


Immortal time bias

A time-varying covariate

Solution?

- **A hazard regression model with a time-varying covariate**
- **Conditional survival**

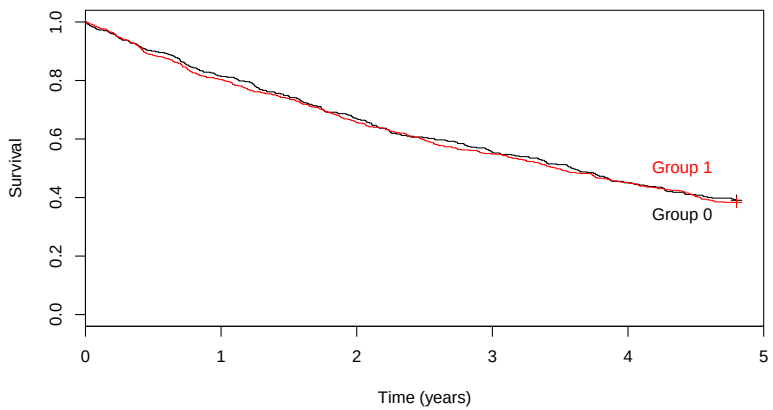


Immortal time bias

A time-varying covariate

Solution?

- **A hazard regression model with a time-varying covariate**
- **Conditional survival**



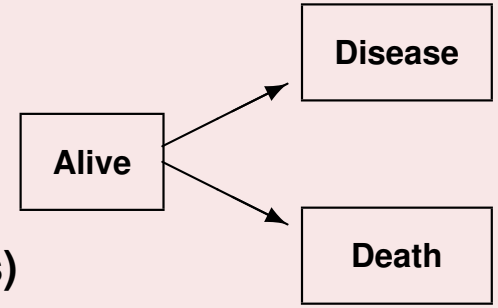
It takes time to measure time ...

... and things can happen in between.

TG8 plan

Multiple endpoints

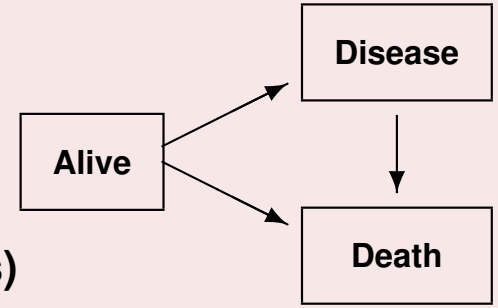
- **Introduction to multistate models:**
  - **competing risks**
  - **illness-death model**
  - **general multistate model**
- **Aalen Johansen estimator (no covariates)**
- **Regression:**
  - **Cause-specific hazard models (modelling hazard)**
  - **Fine-Gray model (modelling probability)**
- **Extensions (recurrent events, ...)**



TG8 plan

Multiple endpoints

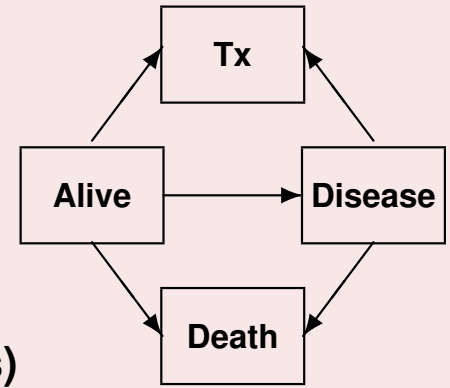
- **Introduction to multistate models:**
  - **competing risks**
  - **illness-death model**
  - **general multistate model**
- **Aalen Johansen estimator (no covariates)**
- **Regression:**
  - **Cause-specific hazard models (modelling hazard)**
  - **Fine-Gray model (modelling probability)**
- **Extensions (recurrent events, ...)**



TG8 plan

Multiple endpoints

- **Introduction to multistate models:**
  - **competing risks**
  - **illness-death model**
  - **general multistate model**
- **Aalen Johansen estimator (no covariates)**
- **Regression:**
  - **Cause-specific hazard models (modelling hazard)**
  - **Fine-Gray model (modelling probability)**
- **Extensions (recurrent events, ...)**



Competing risks, multistate models

Same ideas, but care needed

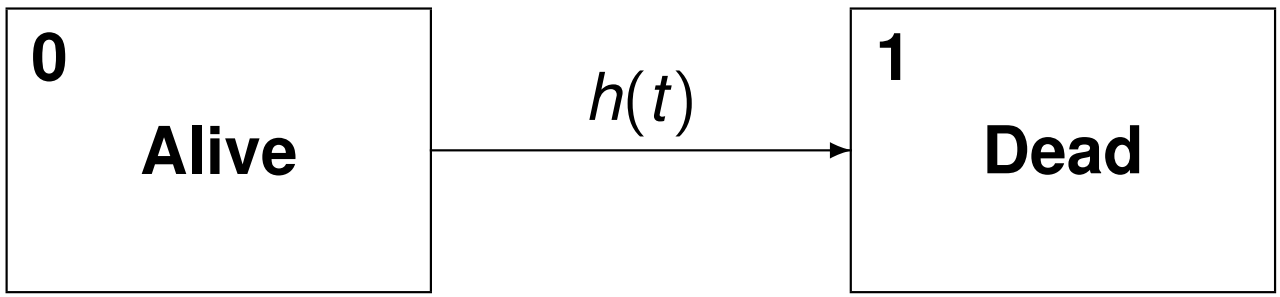
- **Everything can be defined via hazard functions**
- **Cox model can still be used for modelling hazard functions**
- **No one-to-one relationship between hazard and survival**
- **More difficult to state what is of interest**

Competing risks example

Survival of patients on dialysis

- **Time 0: kidney failure, start of dialysis**
- **Events: death (the event of interest), kidney transplant**
- **Not everyone experiences the event of interest in the complete data**
- **Transplants are a competing risk, patients are not randomly transplanted**
- **Patients with best prognosis get transplanted**
- **If these patients are considered as censored, survival gets underestimated**

Single endpoint

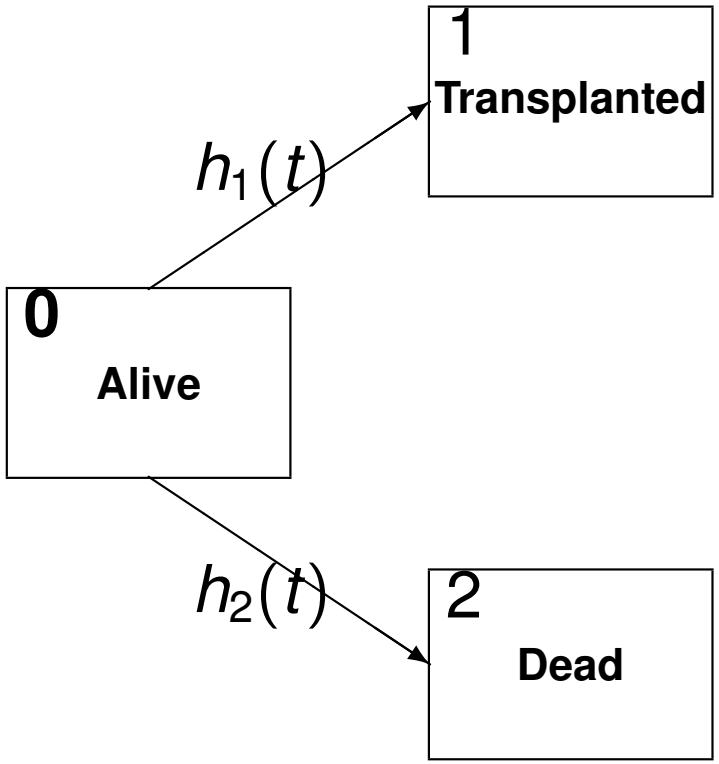


In the population (complete observation):

- **Every one ends up in state 1**
- **The probability of being in state 1 by time t is given uniquely from the hazard:**

$$F(t) = 1 - S(t) = 1 - \exp(-H(t))$$

Competing risks



- **Several events can happen**
- **Even in the population, not every one experiences the same event**
- **Even if only one event is of interest, one cannot see others as 'independent censoring'**

Relationship between rates and risks

Both rates are needed to compute one risk

- **Cause-specific hazards $j = 1, 2$:**

$$h_j(t) \approx \frac{\mathrm{P}(\textbf{state } j \textbf{ by time } t + dt \mid \text{state } 0 \text{ time } t)}{dt}$$

- **Overall survival function:**

$$S(t) = \mathrm{P}(\textbf{alive at time } t) = e^{-[H_1(t)+H_2(t)]}$$

- **Probability of experiencing event j at time u**

$$\mathrm{P}(\textbf{state } j \textbf{ at time } u) \approx S(u-)h_j(u)du$$

- **Cumulative incidence function for event j**

$$F_j(t) = \mathrm{P}(\textbf{state } j \textbf{ by time } t) = \int_0^t S(u-)h_j(u)du$$

Inference

The likelihood function

- **For independent observations $(T_i, \delta_i \cdot D_i)$, where D_i =final state, $T_i = min(T_i^*, C_i)$ $\delta_i = I(T_i^* < C_i)$, the likelihood may again be expressed via the hazard (and cumulative hazard H) functions:**

$$L(\theta) = \prod_i h_{1\theta}(T_i)^{\delta_i I(D_i=1)} h_{2\theta}(T_i)^{\delta_i I(D_i=2)} e^{-H_{1\theta}(T_i) - H_{2\theta}(T_i)}$$

- **This likelihood can be factorized as:**

$$L(\theta) = \prod_i h_{1\theta}(T_i)^{\delta_i I(D_i=1)} e^{-H_{1\theta}(T_i)} \prod_i h_{2\theta}(T_i)^{\delta_i I(D_i=2)} e^{-H_{2\theta}(T_i)}$$

Inference

Consequence of the factorization

- **If the model uses different parameters for different hazards - hazard regression analysis can be performed by censoring the other cause**
- **Hazards can be modeled by censoring the other risk, for probabilities, both are needed**
- **Distinction between hazard rate and probability of an event (equal in single event case)**

Back to dialysis example

What is the interest of the analysis?

- **How many patients are still on dialysis after 5 years: overall survival**
- **Probability of dying in 5 years: cumulative incidence function**
- **Is one type of dialysis (PD, HD) safer than the other: model hazards**

In general

- **To describe the fraction ending in a state: cumulative incidence function**
- **To understand the mechanisms by which subjects may fail: hazards**
- **Both can be useful for a complete description of the competing risks situation**

Concluding remarks

The TG8 plan

- **Single endpoint and multistate models**
- **Avoiding the common pitfalls:**
  - **Mistaking competing risks for censoring**
  - **Not recognizing a time-varying covariate**
- **Distinction between hazard rates and probabilities in multistate models**

TG8 members

- **Michal Abrahamowicz**
- **Per Kragh Andersen**
- **Richard Cook**
- **Pierre Joly**
- **Torben Martinussen**
- **Maja Pohar Perme**
- **Jeremy Taylor**
- **Terry Therneau**