Regression models for failure time data in the presence of competing risks

Bernhard Haller

Institut für Medizinische Statistik und Epidemiologie

24.11.2011

Outline

Outline

Introduction

- Primary interest: Time to event
- Problem: Event of interest not observed for all individuals
  $\rightarrow$ censored observations (end of study, loss to follow-up)
- Special methods for the analysis of event time data
- Observed: Pair (T_i^*, d_i) with
  - $T_i^* = min(T_i, C_i)$
  - $d_i = I(T_i < C_i)$
- T_i = event time
- C_i = censoring time
- T_i^* = observed time
- d_i = status (0 = censored, 1 = event)

Relevant measures

- Density function: $f(t)$
- Survivor function: $S(t) = P(T > t) = \int_t^\infty f(s)ds$
- Hazard function: $\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$
- $S(t) = exp\Big(- \int_0^t \lambda(s)ds \Big) = exp\Big(- \Lambda(t) \Big)$

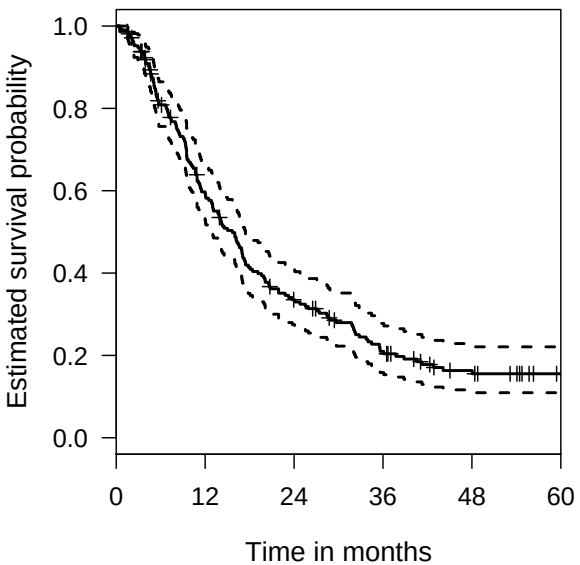
Event-time analysis

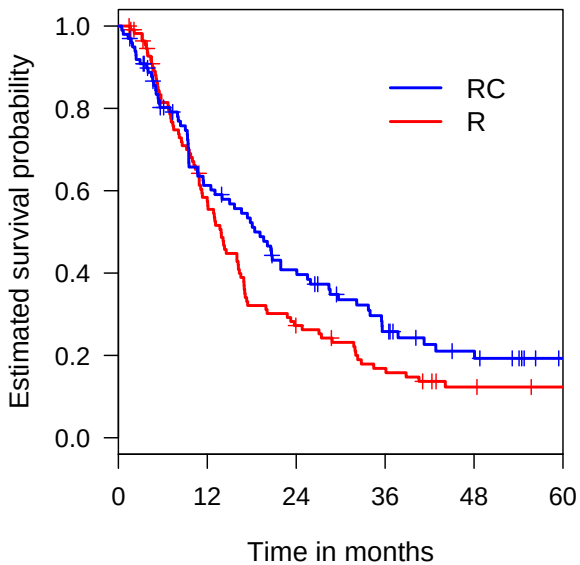
- Kaplan-Meier estimator:

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

- Cox regression:

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

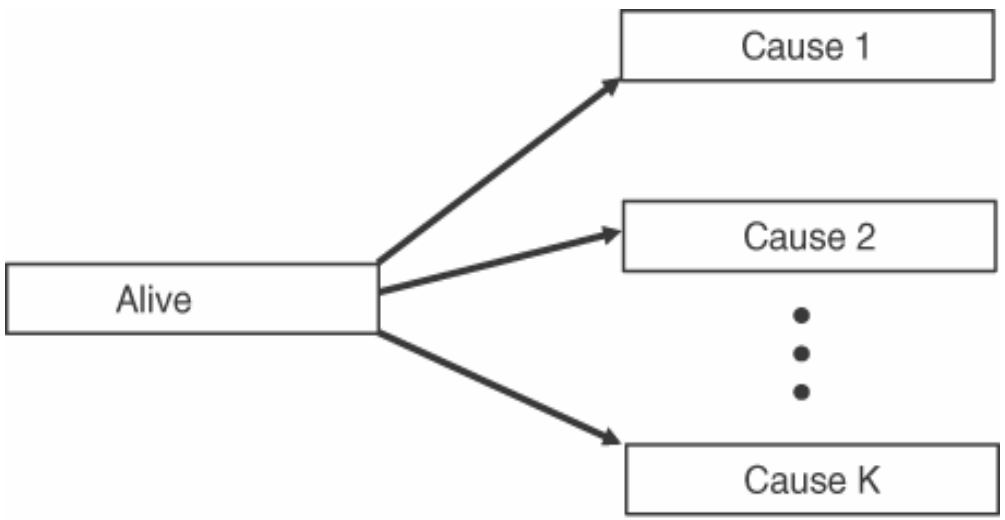




Outline

Competing Risks

- Two or more mutually exclusive types of event
- Not all standard methods for event-time analysis are appropriate
- Different approaches for analysis of competing risks data proposed
- Observed:
  - $T_i^* = min\{T_1, T_2, ..., T_K, C\}$
  - d_i: Indicator for type of event ($d_i \in \{1, ..., K\}$) or censoring ($d_i = 0$)



"A nonidentifiability aspect of the problem of competing risks"

Tsiatis (1975):

- Problem: Only time to first event can be observed
- Correlation structure between times to different events cannot be assessed
- For each set of marginal distributions joint distributions with different correlation structures can be found
- Only under the assumption of independence marginal distributions do uniquely define the joint distribution
- Assumption of independence cannot be tested from observed data

Motivation

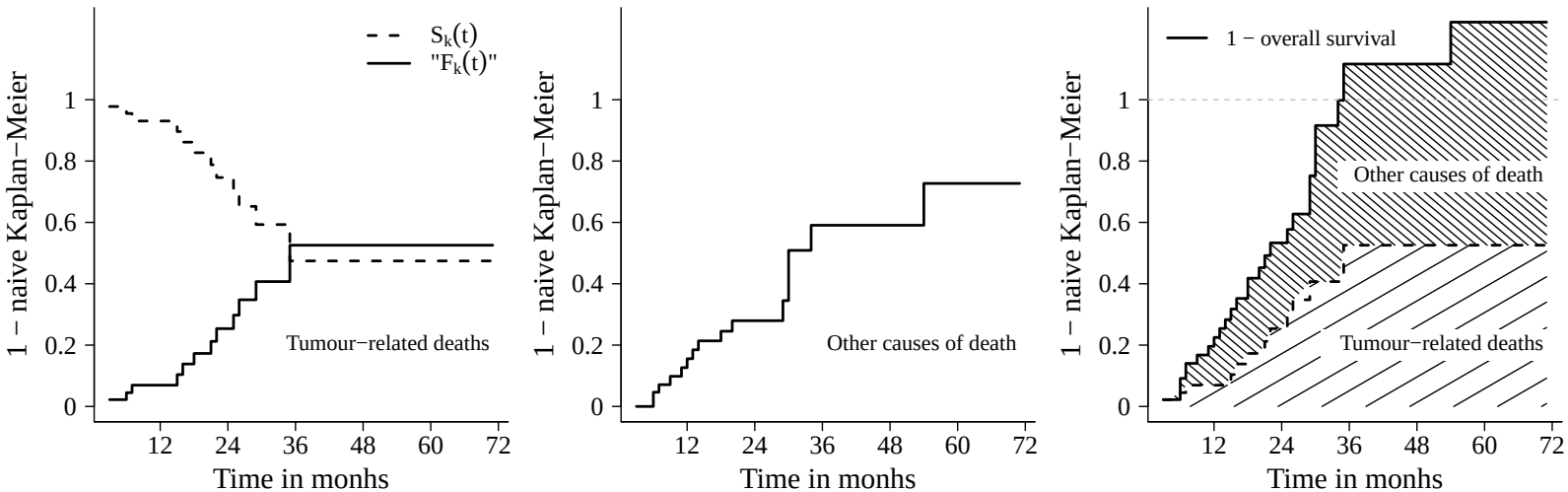
Koller et al. (*Statistics in Medicine*, 2011):
Competing risks and the clinical community: irrelevance or ignorance?

- Analysis of research articles from 2000 - 2010 from leading medical and biostatistical journals with a special focus on the last 50 publications relevant for competing risks analyses
- *"Large developments in competing risks methodology have been achieved over the last decades, but we assume that recognition of competing risks in the clinical community is still marginal."*
- Results:
  Competing risks ...
  - ... were often discussed in statistical journals.
  - ... were present in many studies.
  - ... were not considered or not analysed adequately in most of the studies investigated (n=50).
- Conclusion:
  "A better recognition of competing risks in the clinical community is needed."

Kaplan-Meier method in the presence of competing risks

- "Naïve" Kaplan-Meier estimator is applied to competing risks data
- Competing events are treated as censored observations
- A major assumption (independence between event and censoring times) of the Kaplan-Meier estimator is violated

Example: Patients with non-small cell lung cancer (NSCLC)



$$S(t) = 1 - \int_0^t f(s)ds = 1 - \int_0^t \lambda(s)S(s)ds = 1 - \int_0^t \lambda(s)exp\big(-\Lambda(s)\big)ds$$

Cause-specific hazard

- Cause-specific hazard rate:

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}$$

- Cumulative cause-specific hazard rate:

$$\Lambda_k(t) = \int_0^t \lambda_k(s) ds$$

- Estimating λ_k:

$$\hat{\lambda}_k(t_i) = \frac{d_{ki}}{n_i}$$

n_i: Number of subjects under risk (no event until t_i, still under observation)

d_{ki}: Number of events of type k at time t_i

Cumulative incidence function

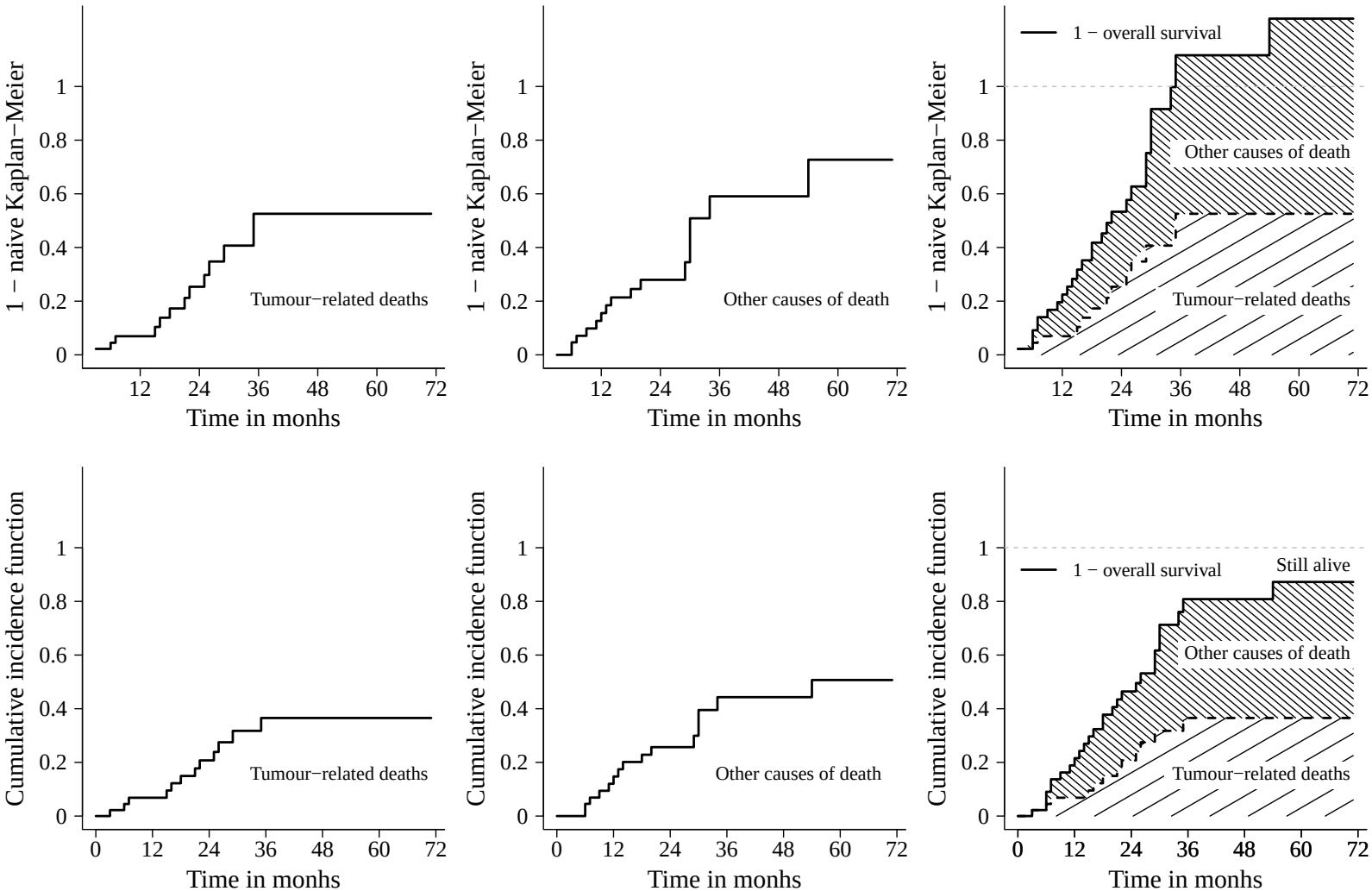
- Cumulative incidence function:

$$F_k(t) = P(T \leq t, D = k) = \int_0^t \lambda_k(s) S(s-) ds$$

- "Naïve" Kaplan-Meier estimator vs. cumulative incidence function:

$$1 - S_k(t) = \int_0^t \lambda_k(s) exp\big(-\Lambda_k(s)\big) ds$$
$$\geq \int_0^t \lambda_k(s) exp\big(-\sum_{j=1}^{K} \Lambda_j(s)\big) ds = F_k(t)$$

"Naïve" Kaplan-Meier estimator vs. cumulative incidence function



Relationship between $\lambda_k(t)$ and $F_k(t)$

- Cause-specific hazard:

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}$$

- Cumulative incidence function:

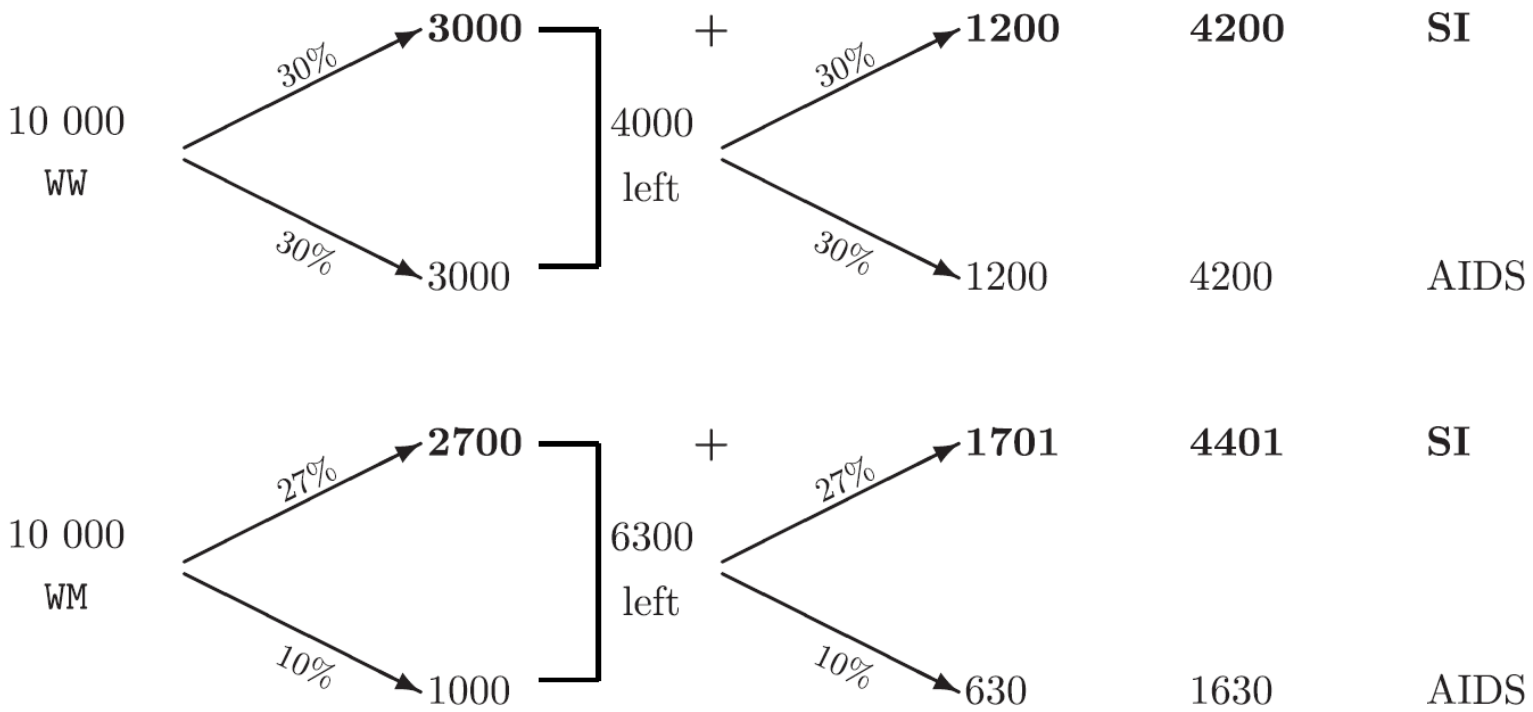
$$F_k(t) = \int_0^t \lambda_k(s) exp\Big(- \sum_{j=1}^{K} \Lambda_j(s)\Big) ds$$

$F_k(t)$ depends on all λ_j, $j = \{1, ..., K\}$

$\rightarrow$ no direct link between $\lambda_k(t)$ and $F_k(t)$

Risk and incidence - an example

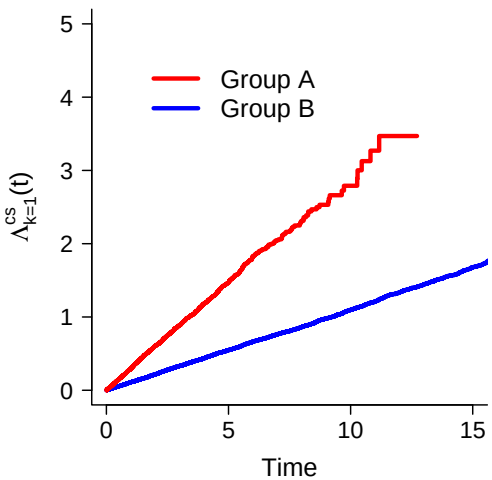
Illustration from Putter et al. (*Statistics in Medicine*, 2007):

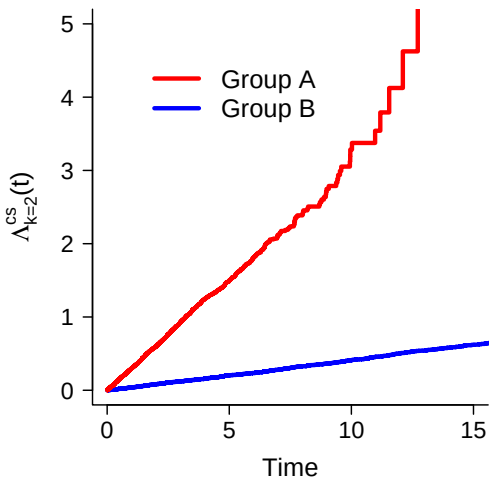


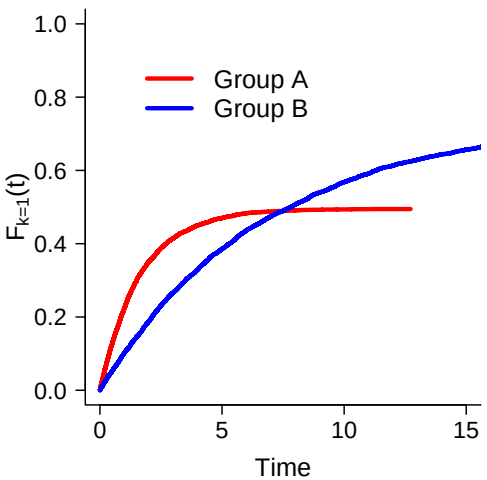
SI: Syncytium inducing HIV phenotype

Risk and incidence - graphical illustration

Similar situation - simulated data:







Subdistribution hazard

- Subdistribution hazard (Fine and Gray, 1999):

$$\lambda_k^*(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = k | T \geq t \cup \{T < t, D \neq k\})}{\Delta t}$$

- Subjects failing from a competing event remain in the risk set
- Estimation of $\lambda_k^{*(1)}$

$$\hat{\lambda}_k^*(t_i) = \frac{d_{ik}}{n_i^*}$$

n_i^*: Number of subjects at risk at time t_i
(no event before t_i OR an event $\neq k$ before t_i)

d_{ki}: Number of events of type k at time t_i

- $F_k(t) = 1 - exp\Big(- \int_0^t \lambda_k^*(s) ds \Big) = 1 - exp\Big(- \Lambda_k^*(t) \Big)$

(1) For the case that no or just administrative censoring is present. In the case that individuals are lost to follow-up, a censoring distribution is estimated from the data.

Illustration of different hazard rates

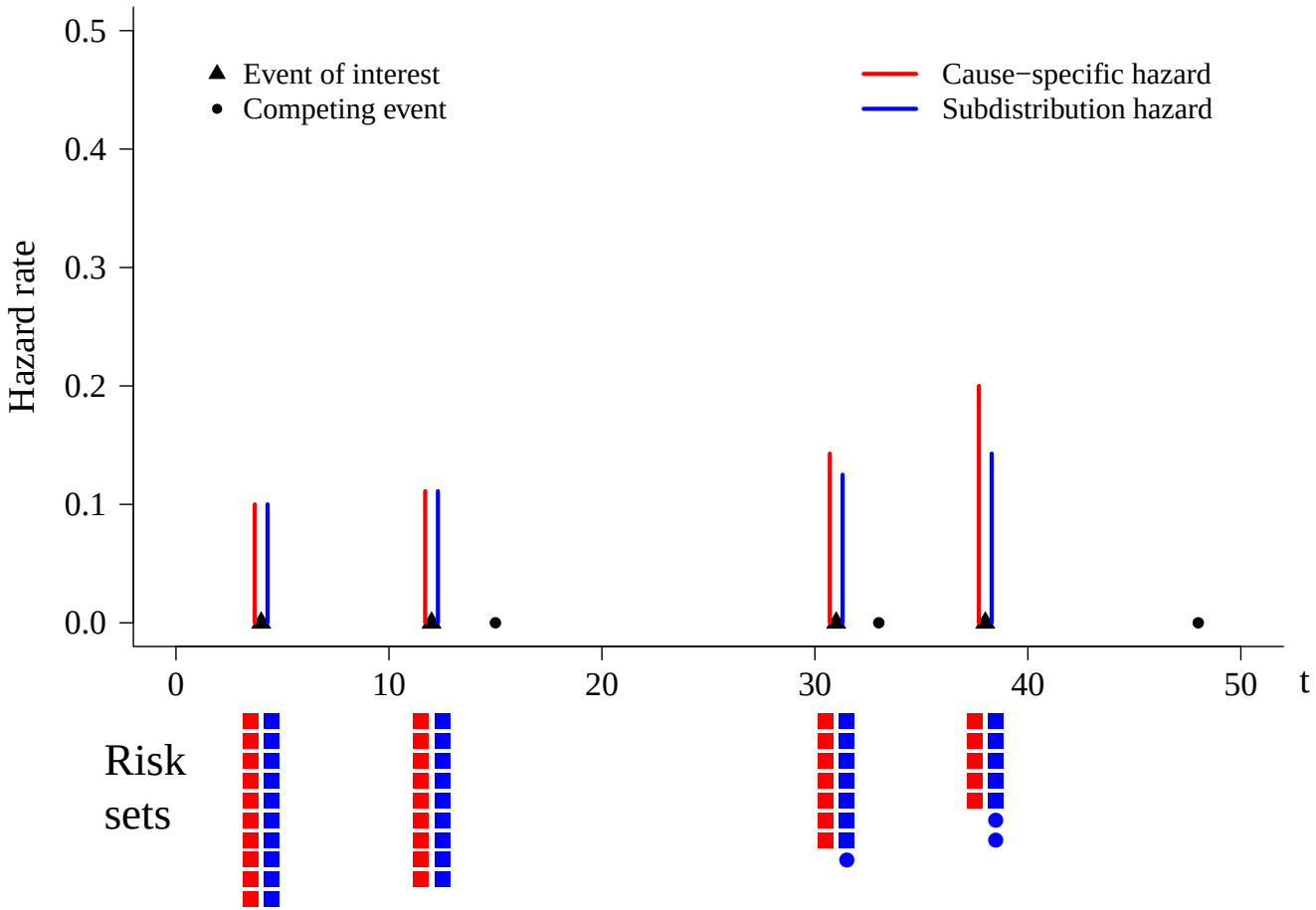
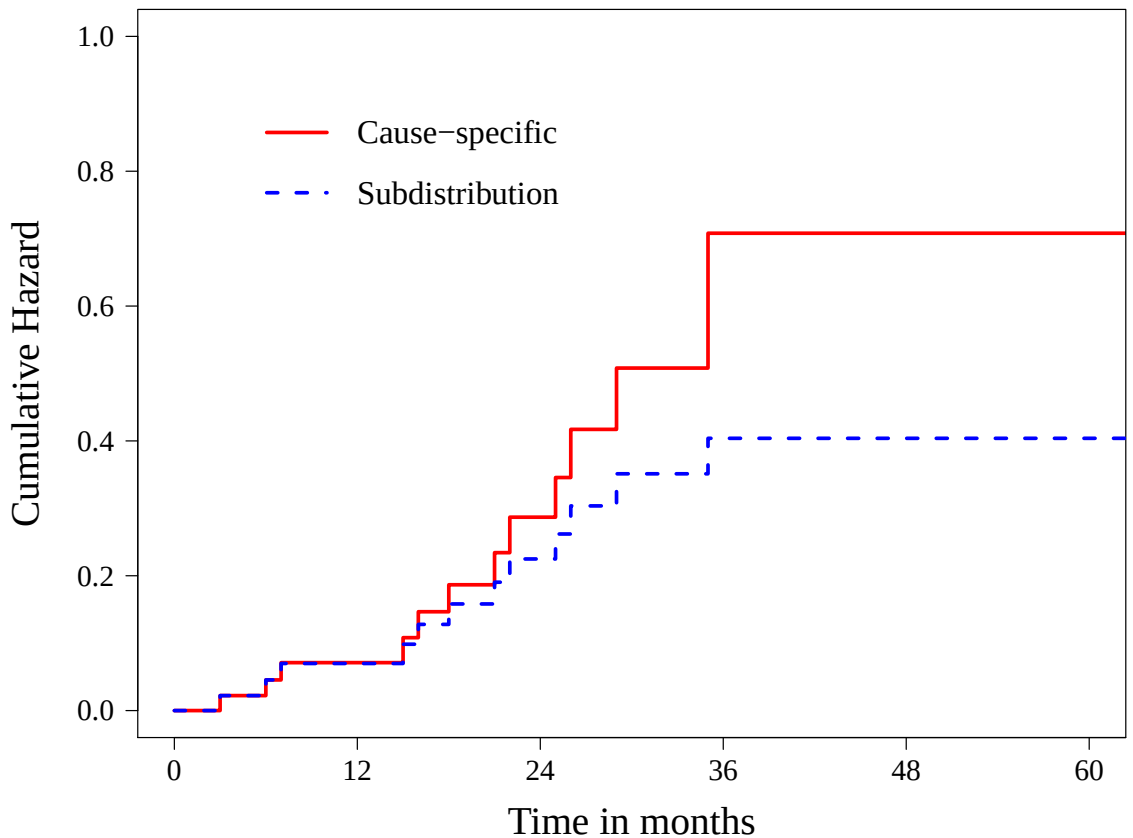


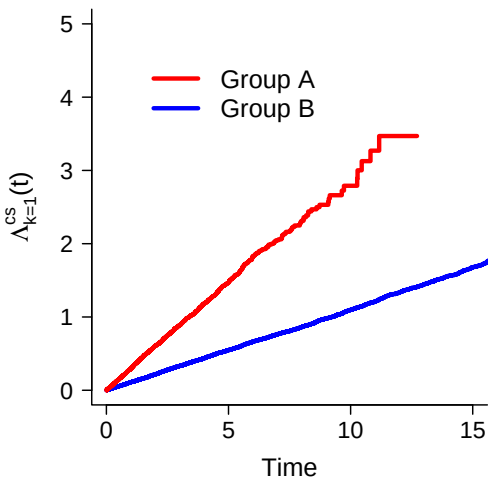
Illustration of different hazard rates

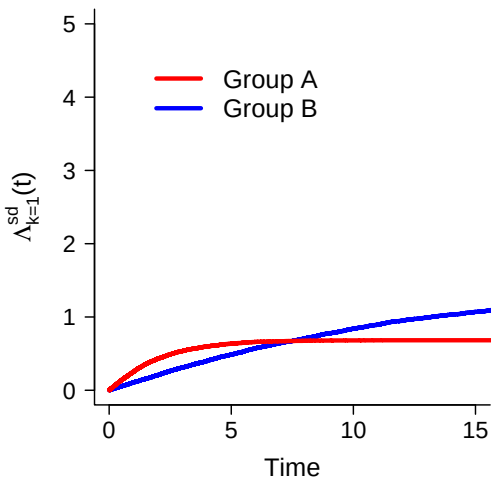
NSCLC patients: Cause-specific and subdistribution hazard for tumour-related death

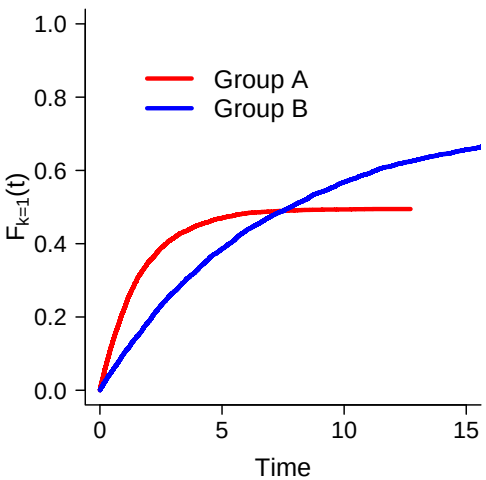


Risk and incidence - graphical illustration

Situation described above - simulated data:







k-sample tests

- Logrank test: Compares cause-specific hazard rates
- Gray's k-samples test: Compares subdistribution hazard rates
- Methods for comparison of cumulative incidence functions:
  - Pepe's Test (integrated weighted difference between two cumulative incidence functions)
  - Lin's Test (Kolmogorov-Smirnov-type test - compares maximum difference between two cumulative incidence functions)

Outline

Cause-specific hazard regression

Prentice et al. (1978):

- $\lambda_k(t)$ as dependent variable
- Prentice et al. proposed Cox-type regression

$$\lambda_k(t|\mathbf{X}) = \lambda_{k,0}(t)exp(\beta_k^\top \mathbf{X})$$

- Competing events are treated as censored observations
- Assumptions and extension as known from classical Cox regression
- Can be performed using standard software

Cause-specific hazard regression

- Cumulative incidences depend on covariate effects on all event types
- Two possible types of event:

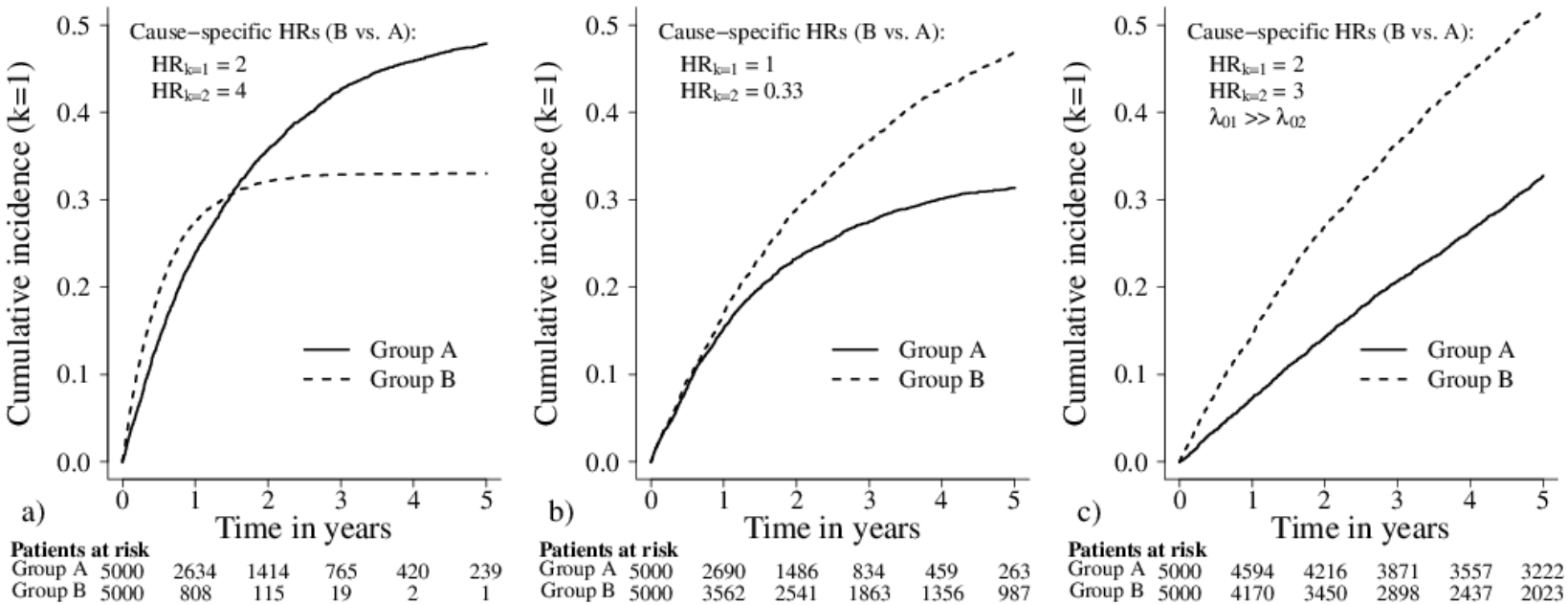
$$F_1(t \mid X) = \int_0^t \lambda_1(s \mid X) exp\Big(- \big(\Lambda_1(s \mid X) + \Lambda_2(s \mid X)\big)\Big) ds$$

- "Naïve" calculation of event probabilities gives biased estimates
- No direct link between regression coefficients and cumulative incidence function

Cause-specific hazard ratios and cumulative incidences

Incidences depend on the effects on all types of event

$\rightarrow$ Simulations (n=10 000); HR: $\frac{Group\ B}{Group\ A}$



$exp(\hat{\beta}_1^{cs}) = 1.93$ $exp(\hat{\beta}_1^{cs}) = 0.97$ $exp(\hat{\beta}_1^{cs}) = 1.96$

Subdistribution hazard regression (Fine&Gray-Model)

Fine and Gray (1999):

- Cox-type regression on the subdistribution hazard proposed

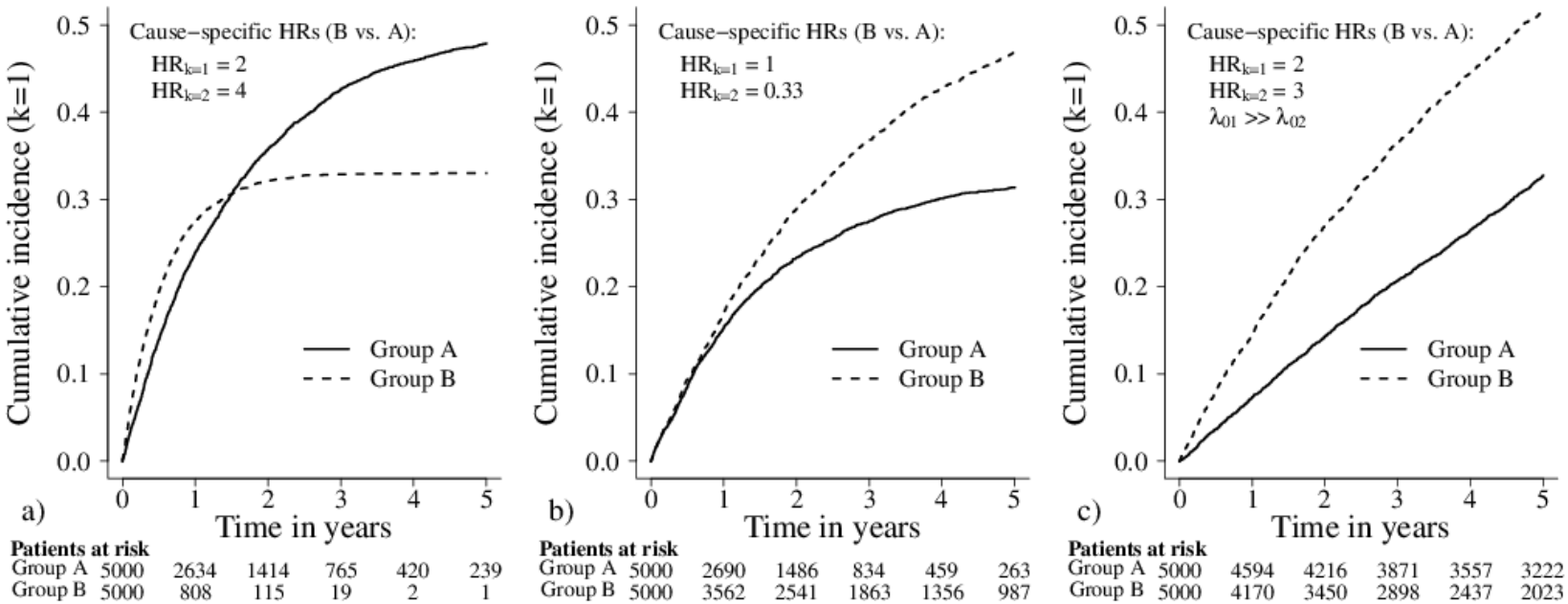
$$\lambda_k^*(t|\mathbf{X}) = \lambda_{k,0}^*(t) exp(\beta_k^{*\top}\mathbf{X})$$

- Individuals failing from a competing event remain in the risk set
- Assumptions known from standard models are translated to subdistribution hazards (e.g. proportionality)
- Implemented in the R library *cmprsk*
- Direct link between regression coefficients and cumulative incidence:

$$F_k(t|\mathbf{X}) = 1 - exp\Big(- \int_0^t \lambda_{0k}^*(s) exp(\beta_k^{*\top}\mathbf{X}) ds \Big) = 1 - exp\Big(- \Lambda_k^*(t|\mathbf{X}) \Big)$$

Cause-specific and subdistribution hazard ratios

SDH-models misspecified: Can be interpreted as summary analysis



$exp(\hat{\beta}_1^{cs}) = 1.93$
$exp(\hat{\beta}_1^{sd}) = 0.70$

$exp(\hat{\beta}_1^{cs}) = 0.97$
$exp(\hat{\beta}_1^{sd}) = 1.59$

$exp(\hat{\beta}_1^{cs}) = 1.96$
$exp(\hat{\beta}_1^{sd}) = 1.86$

Mixture models

Larson and Dinse (1985):

- Joint distribution of event time and type of event as product of conditional and marginal distribution
- P(D,T) = P(T|D)P(D)
- Logistic regression model to estimate covariate effects on the probability of an event of interest
- Separate estimation of event time distributions for given event type
- Proposed conditional distributions for event times:
  - piecewise exponential distribution (Larson and Dinse, 1985)
  - generalized gamma distribution (Lau et al, 2008)
- Likelihood contribution of subject i:

$$L_i = [\pi_i f_1(t_i)]^{I(d_i=1)} \times [(1-\pi_i) f_2(t_i)]^{I(d_i=2)} \times [\pi_i S_1(t_i) + (1-\pi_i) S_2(t_i)]^{I(d_i=0)}$$

- Likelihood has to be maximized

Mixture models

Alternative model for survival times:

- Semi-parametric estimation (Ng and McLachlan (2003), Escarela and Bowater (2008))
- No specification for survival time distribution for given type of event
- Hazard rates for given event type are assumed to be proportional
- Estimation via EM algorithm
- Standard errors via bootstrap samples
- Computationally very intensive

Vertical Modelling

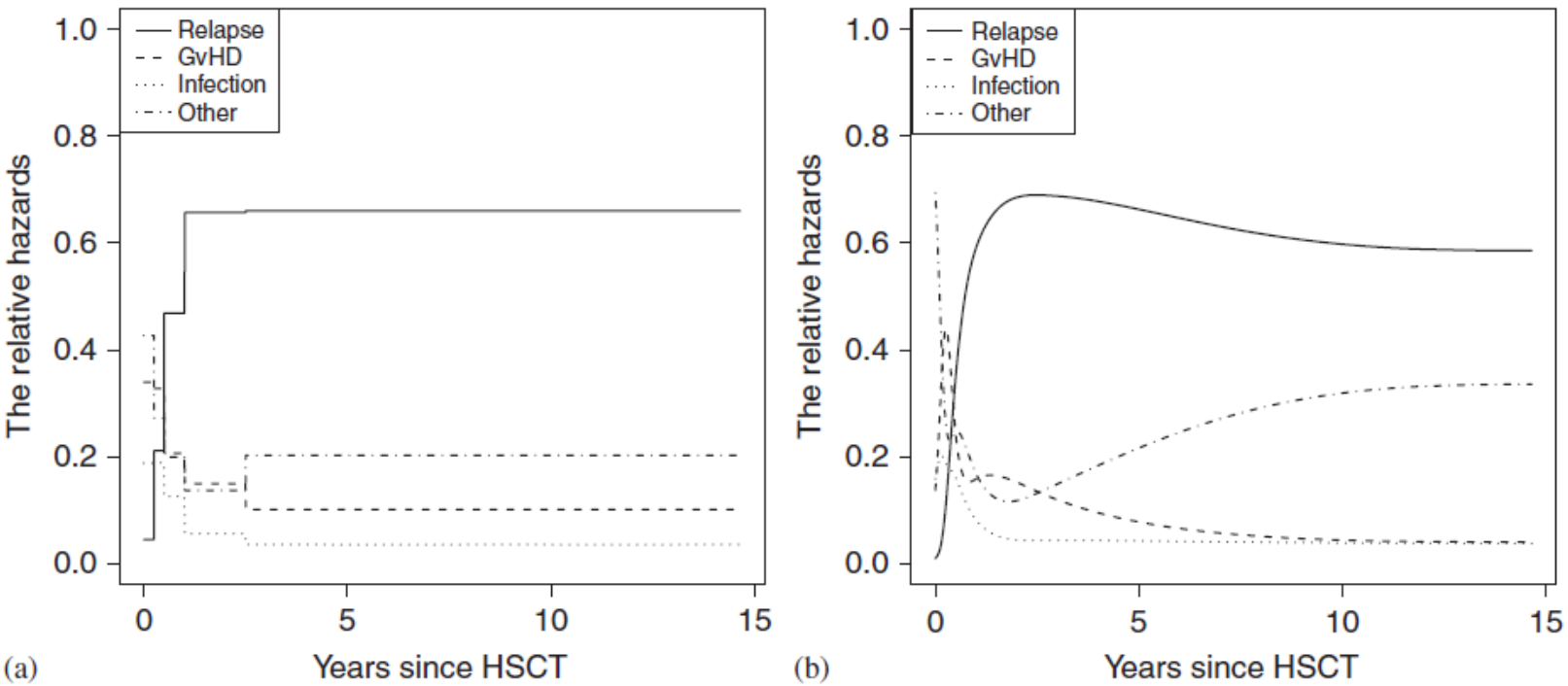
Nicolaie et al. (2009)

- Joint distribution of event time and type of event as product of conditional and marginal distribution
- P(D,T)=P(D|T)P(T)
- Marginal distribution for event times (e.g. parametric model)
- Conditional event probabilities for given event times
  $\rightarrow$ e.g. multinomial logistic regression including time as covariate
- Only individuals with an observed event can be used for estimation of "relative hazards"
- Regression coefficients hard to interpret
  $\rightarrow$ graphical presentation

Vertical Modelling

Nicolaie et al. (*Statistics in Medicine*, 2009)

- Data: 8966 leukaemia patients from the European Group for Blood and Marrow Transplantation
- Proportion of events in pre-specified time intervals or B-Splines to estimate "relative hazards" over time



Event time analysis based on pseudo observations

Andersen, Klein et al. (2003, 2005)

- Calculate H pseudo-observations for each individual i
- Leave-one-out estimates $\hat{F}_k^{(i)}(\tau_h)$ for the cumulative incidence function (Jackknife) at pre-specified time points $\boldsymbol{\tau} = (\tau_1, ..., \tau_H)$

  $\rightarrow$ calculation of pseudo-observations

  $$\hat{\theta}_{ih} = n\hat{F}_k(\tau_h) - (n-1)\hat{F}_k^{(i)}(\tau_h)$$

- Notation:
  - $\boldsymbol{\tau}$: vector of pre-specified time points
  - $\hat{F}_k(\tau_h)$: estimated cumulative incidence function for event k from all observations at time τ_h
  - $\hat{F}_k^{(i)}(\tau_h)$: estimated cumulative incidence function for event k at time τ_h leaving out subject i
- If no censoring is present:
  - $n\hat{F}_k(\tau_h)$ = number of events of interest until time τ_h
  - $\hat{\theta}_{ih}$ is an indicator function $I(T_i \leq \tau_h, d_i = k)$.

Event time analysis based on pseudo observations

- Estimation of covariate effects on the pseudo-observations using a GEE model with complementary log-log (clogclog) link
- $-log(-log(1-\theta_{ih})) = \alpha_h + \boldsymbol{\beta}^\top \mathbf{X}_{ih}$
- $exp(\beta)$ can be interpreted as subdistribution hazard ratio
- Simulation studies: large standard errors for regression coefficients

  $\rightarrow$ Recommendation (Pohar Perme and Andersen, 2008):
  - Do not use the method for proportional subdistribution hazard regression, since more efficient estimators are available
  - Pseudo observations can be useful for more complex models
  - Use pseudo observations to check model assumptions

Example

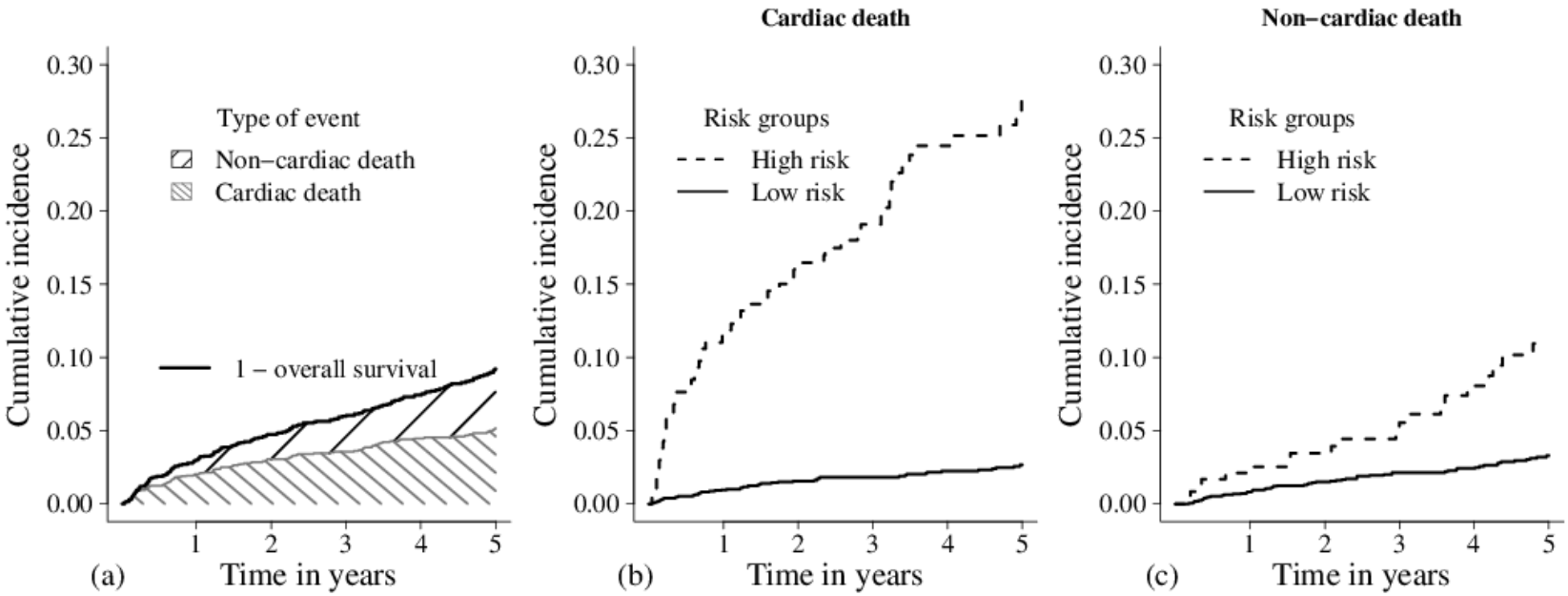
Bauer et al. (2009)

- Observation of 2341 patients after myocardial infarction
- Within 5 years 181 patients died
  - Cardiac deaths: n=104
  - Other types of death: n=77
- Aim: To establish a risk score for cardiac death
- Diabetes (yes/no) and age ($<65/\geq 65$) as further covariates
- Application of presented methods for competing risk regression

Cumulative incidence function

Incidences five years after myocardial infarction :

	Cardiac death		Non-cardiac death	
	$\hat{F}_{card.}(5\ years)$	95% ci	$\hat{F}_{non\text{-}card.}(5\ years)$	95% ci
Overall	5.1%	4.2% to 6.1%	4.1%	3.1% to 5.0%
Low risk	2.7%	1.9% to 3.4%	3.3%	2.4% to 4.2%
High risk	27.5%	21.1% to 33.9%	10.9%	6.3% to 15.7%



Cause-specific hazard regression

```
library(survival)
coxph(Surv(ev.time,ev.type==1)~group + diabetes + age,data=dat)
coxph(Surv(ev.time,ev.type==2)~group + diabetes + age,data=dat)
```

Cardiac death

	$\hat{\beta}$	$\exp(\hat{\beta})$	Std. error	p-value
Risk group	2.36	10.53	0.20	$<$0.001
Diabetes	0.72	2.06	0.21	0.001
Age$\geq$65	0.48	1.60	0.20	0.016

Non-cardiac death

	$\hat{\beta}$	$\exp(\hat{\beta})$	Std. error	p-value
Risk Group	1.06	2.89	0.26	$<$0.001
Diabetes	0.70	2.01	0.25	0.005
Age$\geq$65	1.28	3.69	0.24	$<$0.001

Sudistribution hazard regression (Fine&Gray-Model)

```
library(cmprsk)

summary(crr(dat$ev.time,dat$ev.type,
  cbind(dat$group,dat$diabetes,dat$age),failcode=1,cencode=0))

summary(crr(dat$ev.time,dat$ev.type,
  cbind(dat$group,dat$diabetes,dat$age),failcode=2,cencode=0))
```

	Cardiac death			
	$\hat{\beta}$	$\exp(\hat{\beta})$	Std. error	p-value
Risk group	2.32	10.21	0.20	$<$0.001
Diabetes	0.68	1.98	0.21	0.001
Age$\geq$65	0.47	1.60	0.20	0.017

	Non-cardiac death			
	$\hat{\beta}$	$\exp(\hat{\beta})$	Std. error	p-value
Risk Group	0.84	2.31	0.28	0.002
Diabetes	0.62	1.85	0.24	0.011
Age$\geq$65	1.28	3.58	0.25	$<$0.001

Mixture model

- $P(D,T|\mathbf{X}) = P(T|D,\mathbf{X})P(D|\mathbf{X})$
- Semi-parametric approach proposed by Ng and McLachlan (2003)
- No distribution assumption for survival times given type of event
- Hazard rates are assumed to be proportional
- Parameter estimation via EM algorithm
  - E-step: Determine expected failure type for censored observations given the observed data and current parameter estimates
  - M-step: Maximize the log-likelihood given the observed data and the expected failure types for censored observations
- Estimation of standard errors using 500 bootstrap samples

Mixture model results

- Higher probability of cardiac death for high risk patients
- Increased risk for both types of events for high risk patients

	Event types		Event times			
	Cardiac		Cardiac		Non-cardiac	
	$\hat{\beta}$	95% ci	$\hat{\beta}$	95% ci	$\hat{\beta}$	95% ci
Constant	-2.30	-3.92 to 1.65	—	—	—	—
Risk group	2.22	-1.44 to 3.97	0.88	-0.92 to 3.19	1.76	-0.21 to 2.76
Diabetes	-0.43	-2.00 to 1.82	1.17	-0.44 to 2.02	0.52	-0.49 to 2.07
Age $\geq$ 65	0.96	-1.45 to 2.66	-0.25	-1.60 to 1.02	1.54	0.71 to 2.54

Vertical modelling

P(D=1|T,**X**) $\rightarrow$ Logistic regression

- Smooth influence of time (B-Spline)
- Interaction between time and group
- Only patients with observed event considered

```
glm(ev.type==1~bs(ev.time,degree=3) * group + diabetes + age,
     family=binomial(link="logit"),data=dd)
```

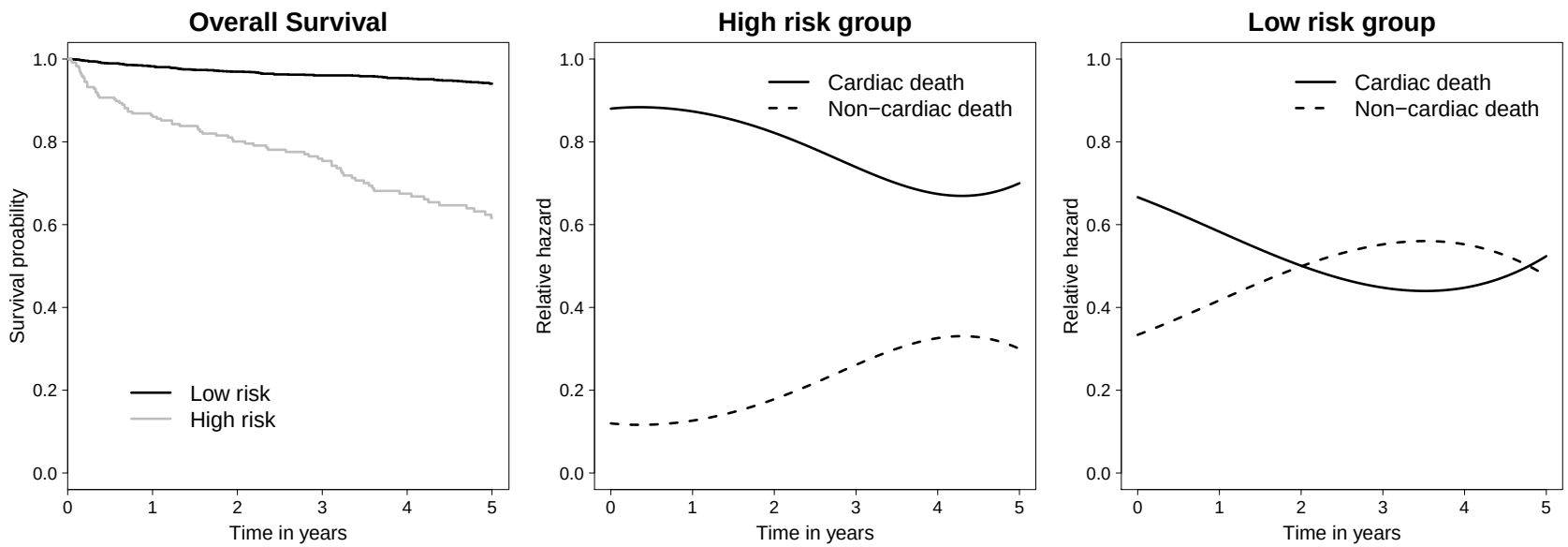
	$\hat{\beta}$	Std. error	p-value
Constant	0.75	0.61	0.218
Risk group	1.27	0.91	0.165
Diabetes	-0.08	0.36	0.825
Age $\geq$ 65	-0.65	0.34	0.053
bs1(time)	-0.76	1.96	0.698
bs2(time)	-1.37	1.69	0.419
bs3(time)	-0.69	1.14	0.545
bs1(time)$\times$risk group	1.08	3.25	0.740
bs2(time)$\times$risk group	-0.64	2.56	0.802
bs3(time)$\times$risk group	-0.46	1.82	0.800

bs: B-Spline components

bs.(time)$\times$ Risk group: Interaction terms

Vertical modelling - graphical presentation

- left: Survivor function adjusted for age and diabetes
- middle, right: Estimated relative hazards for events types given event time



Pseudo observations

- Estimation of cumulative incidence for all data
- Jackknife estimates (n-1 obs.) for cumulative incidence
- $\tau = \{0, 3, 6, 9, 12, 15, ..., 57, 60\}$ months

  $\rightarrow 2341 \times 21$ leave-one-out estimates $\hat{F}^{(i)}_{card.}$
- Pseudo observations:

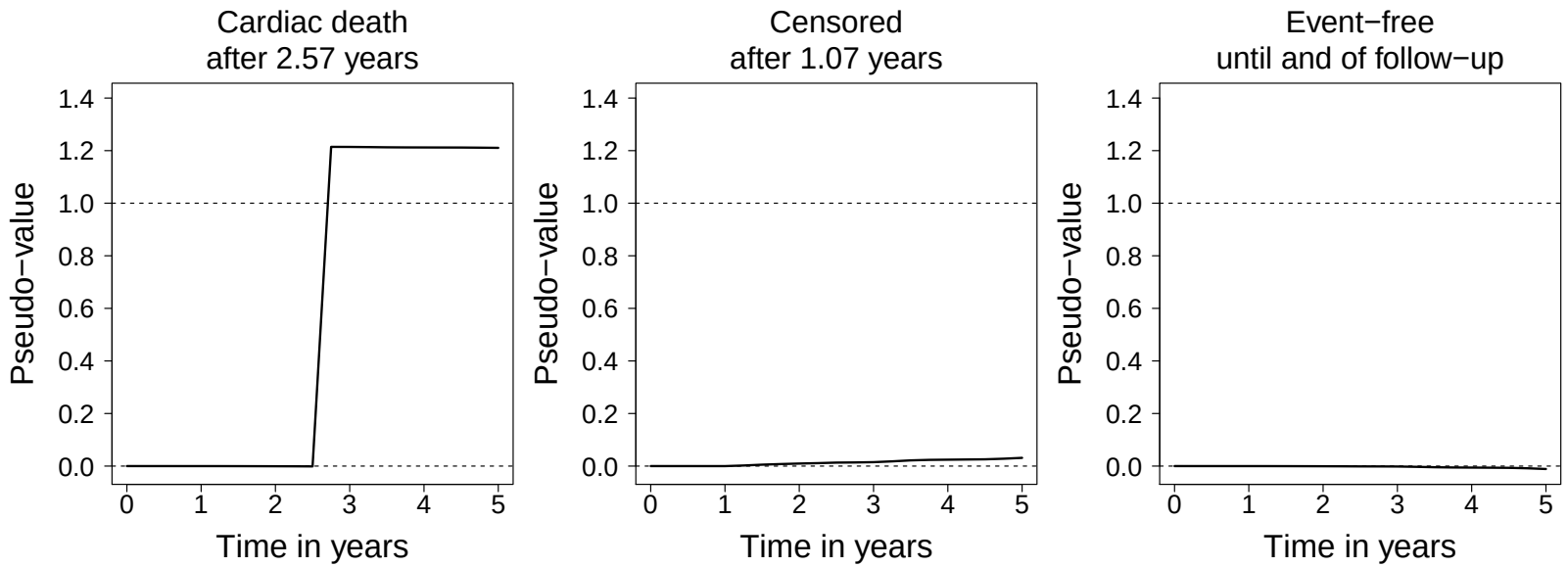
$$\hat{\theta}_{ih} = n\hat{F}_{card.}(\tau_h) - (n-1)\hat{F}^{(i)}_{card.}(\tau_h)$$

```
times <- seq(0,5,by=0.25)
CUM <- timepoints(cuminc(dat$ev.time,dat$ev.type),times)$est[1,]


CUM_i <- matrix(nrow=dim(dat),ncol=length(times))

for(i in 1:dim(dat)[1])
{
Cuminc <- timepoints(cuminc(dat$ev.time[-i],dat$ev.type[-i]),times)
CUM_i[i,] <- Cuminc$est[1,]
}
```

Examples for pseudo observations



Estimation of covariate effects:

- Pseudo observations as dependent variable
- GEE model with cloglog-link

```
library(geepack)
fit <- geese(pseudo ~factor(time) + group + diab + alter, id=id,
          data=na.omit(Dlong), scale.fix=T, family=gaussian,
              mean.link="cloglog",corstr="independence")
```

Pseudo observations - results

	$\hat{\beta}$	$\exp(\hat{\beta})$	Std. error	p-value
Constant	-6.81	0.00	0.35	<0.001
Risk group	2.36	10.59	0.22	<0.001
Diabetes	0.81	2.25	0.25	0.001
Age $\geq$ 65	0.53	1.70	0.26	0.043
Time = 3 months	1.04	2.83	0.30	<0.001
Time = 6 months	1.21	3.35	0.26	<0.001
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Time = 60 months	2.74	15.49	0.18	<0.001

- $\exp(2.36) = 10.59 \rightarrow$ can be interpreted as subdistribution HR
- Similar to Fine&Gray regression
  - F&G: $HR^{sd}_{k=1} = 10.21$
  - F&G: Smaller standard errors

Discussion of regression approaches

- Cause-specific and subdistribution hazard regression mainly used
  - Similar to standard analysis
  - Available in standard software
  - Different measures used as dependent variable
    - → Different assumptions
    - → Different interpretation
    - → May give different results
    - → Be aware of relationships and limitations
- Mixture model and vertical modelling useful for data exploration
  - Many coefficients → hard to interpret
  - No standard software available
  - Computationally intensive
- Model to be used for hypothesis testing?
  - Depending on research question
    - → Effects on risk: Cause-specific hazard regression
    - → Effects on incidence: Subdistribution hazard regression
  - Specify model for primary analysis a priori
  - Consider covariate effects on all event types (graphical presentation)

Outline

Simulation

- Simulation is based on cause-specific hazard rates
- Beyersmann et al. (*Statistics in Medicine*, 2009): Simulating competing risks data in survival analysis
- Cause-specific hazard rates *"completely determine the competing risks process"*
- Evaluation of methods using cause-specific hazards possible
- Methods using subdistribution hazard / cumulative incidence functions: ?

Simulation - cause-specific hazard (k=2)

1. Choose baseline hazards $\lambda_{0;1}(t)$ and $\lambda_{0;2}(t)$ for both types of event
2. Specify cause-specific regression coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$
3. Generate the matrix of covariates $\mathbf{X}$ from a multivariate distribution
4. Simulate event times based on individual overall hazards
   $\lambda_i(t \mid \mathbf{X}_i) = \lambda_{0;1}(t)exp(\mathbf{X}_i^\top \boldsymbol{\beta_1}) + \lambda_{0;2}(t)exp(\mathbf{X}_i^\top \boldsymbol{\beta_2})$
5. Determine the type of event for each individual by drawing a random number with probabilities
   $$P(D_i = 1 \mid \mathbf{X}_i) = \lambda_{1,i}(T_i \mid \mathbf{X}_i)/\big(\lambda_{1,i}(T_i \mid \mathbf{X}_i) + \lambda_{2,i}(T_i \mid \mathbf{X}_i)\big)$$
   $$P(D_i = 2 \mid \mathbf{X}_i) = \lambda_{2,i}(T_i \mid \mathbf{X}_i)/\big(\lambda_{1,i}(T_i \mid \mathbf{X}_i) + \lambda_{2,i}(T_i \mid \mathbf{X}_i)\big)$$
6. Generate independent censoring times

Simulation - subdistribution hazard

- Aim: To evaluate methods using the subdistribution hazard
- Problem: Simulation of event times for given β_1^*
- Possible solution: Use the relationship:

$$exp\big(\beta_1(t)\big) = \frac{1 + \frac{F_2(t|X=1)}{S(t|X=1)}}{1 + \frac{F_2(t|X=0)}{S(t|X=0)}} \times exp\big(\beta_1^*(t)\big)$$

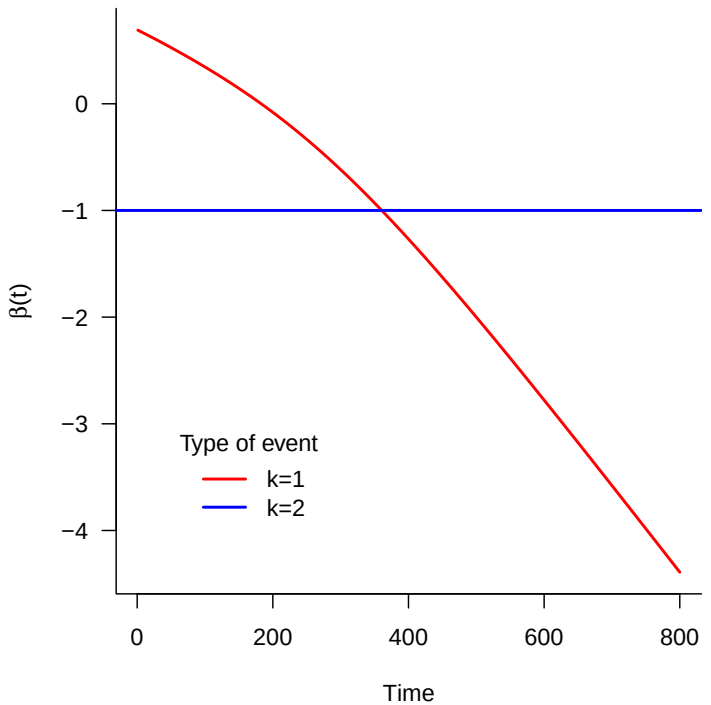
Proposal for k=2, binary covariate

- Choose covariate effects on the subdistribution hazard β_1^*
- Specify $\lambda_{0;1}(t)$, $\lambda_{0;2}(t)$ and $\beta_2(t)$
- Determine $\beta_1(t)$ for all t
- Generate event times following an algorithm for time-varying covariate effects
  $\rightarrow$ Binomial algorithm by Sylvestre and Abrahamowicz (2007)

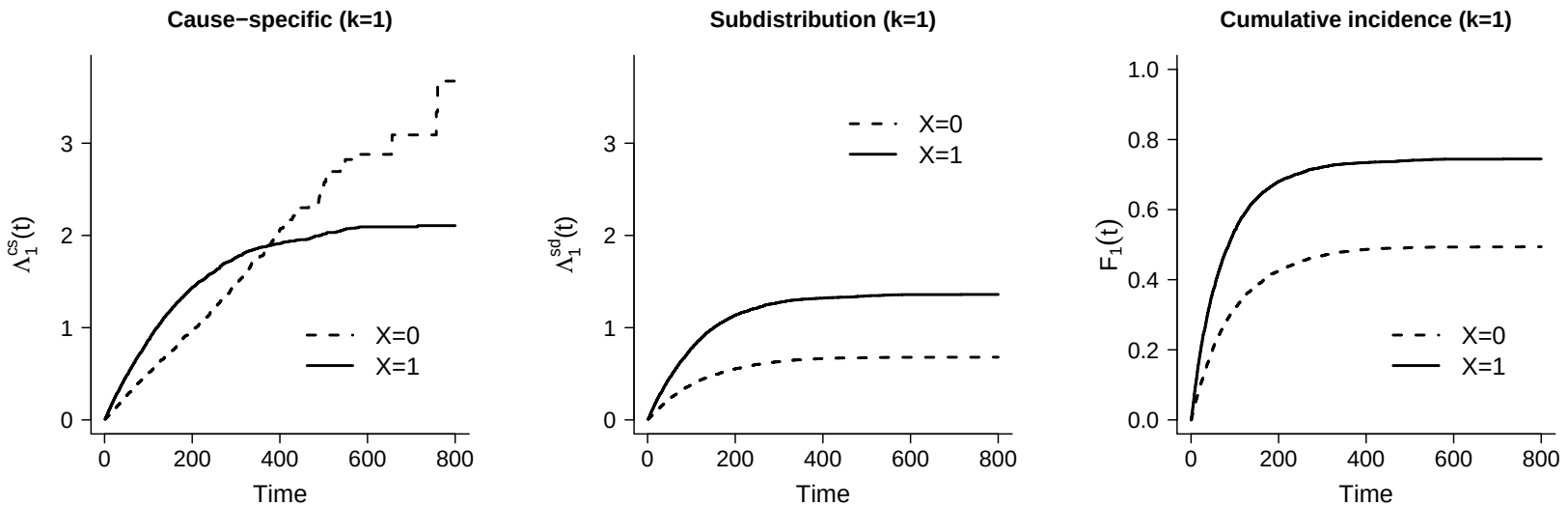
Example for data following a proportional sdh model

Aim: Generate event times for two groups with a $HR^{sd}_{k=1} = 2$.

- Constant baseline hazards
- Constant $HR^{cs}_{k=2}$
- How to choose $\beta^{cs}_{k=1}(t)$?



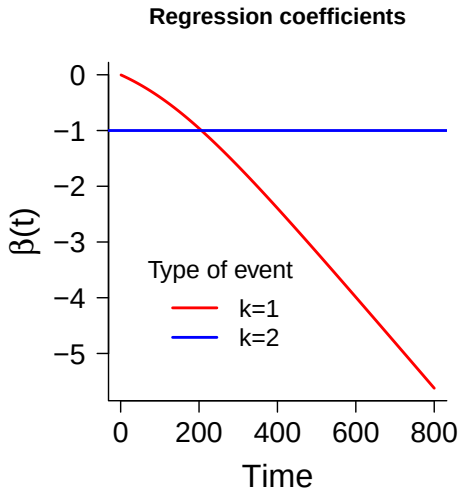
Example for data following a proportional sdh model

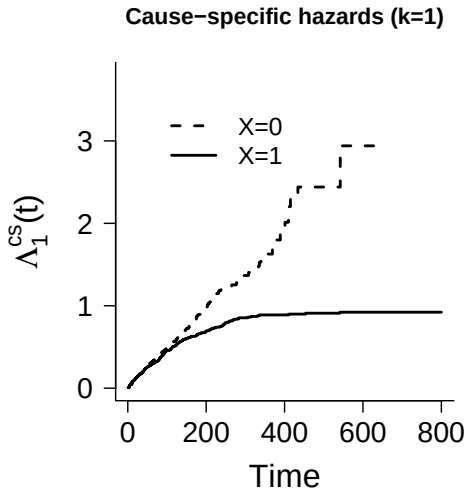


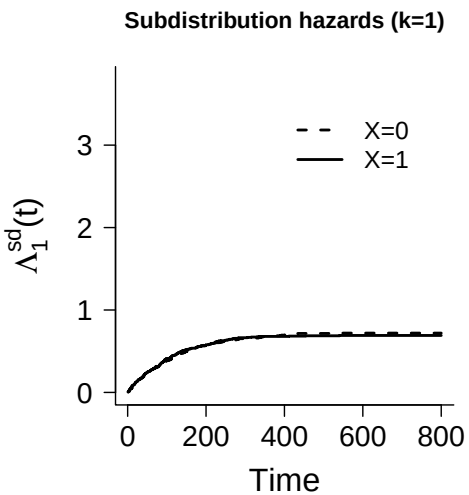
- $HR_{k=1}^{cs}$ is time-dependent
- $HR_{k=1}^{sd}$ is constant over time: $exp(\beta_{k=1}^{sd}) = 2$

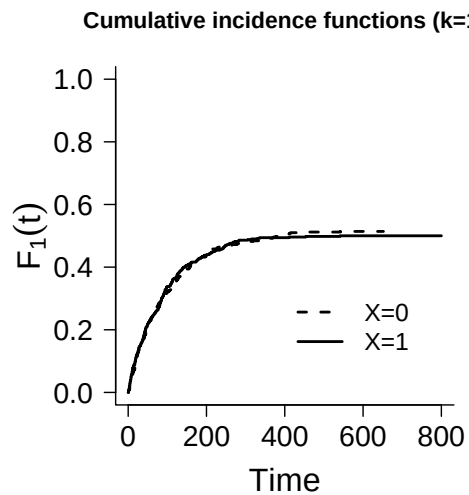
Example II

Aim: $HR^{sd}_{k=1} = 0$









Research questions

- Choice of adequate measures for specific research questions
- When do cause-specific and subdistribution hazard regression give similar / different results?
- Identification of important covariates and high risk groups using different regression approaches
- Assessment of predictive accuracy and goodness of fit
- Small sample properties of proposed models
  $\rightarrow$ alternatives for small sample studies

References

- Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* **72**(1), 187-220.
- Koller, M. T., Raatz, H., Steyerberg, E. W. & Wolbers M. (2011). Competing risks in the clinical community: irrelevance or ignorance? *Statistics in Medicine* DOI: 10.1002/sim.4384.
- Putter, H., Fiocco, M. & Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine* **26**, 2389–2430.
- Beyersmann, J., Latouche, A., Buchholz, A. & Schumacher, M. (2009) Simulating competing risks data in survival analysis. *Statistics in Medicine* **28**:956-971
- Sylvestre, M. & Abrahamowicz, M. (2008) Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine* **27**:2618-2634
- Bauer, A., Barthel, P., Schneider, R., Ulm, K., Müller, A., Joeining, A., Stich, R.,Kiviniemi, A., Hnatkova, K., Kuikuri, H. et al. (2009) Improved Stratification of Autonomic Regulation for risk prediction in post-infarction patients with preserved left ventricular function (ISAR-Risk) *European Heart Journal* **30**:576-583.

References - regression approaches

- Prentice, R. L., Kalbfleisch, J. D., Peterson A., Flournoy, N., Farewell, V. & Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541-554.
- Fine, J. P. & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *JASA* **94**, 496–509.
- Larson, M. G. & Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **34**, 201–211.
- Nicolaie, M. A., van Houwelingen, H. C. & Putter, H. (2010). Vertical modelling: A pattern mixture approach for competing risks modelling. *Statistics in Medicine* **29**, 1190-1205.
- Klein, J. P. & Andersen, P. K. (2005). Regression Modelling of Competing Risks Based on Pseudovalues of the Cumulative Incidence Function *Biometrics* **61**, 223-229.