Competing risks, when and how to incorporate them in the analysis

Workshop
Statistical modelling of multivariate longitudinal and survival data in medical research,
University of Cape Town
January 29-31, 2019

Ronald Geskus
Oxford University Clinical Research Unit (OUCRU)
Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam

Part I

Introduction to competing risks

Outline

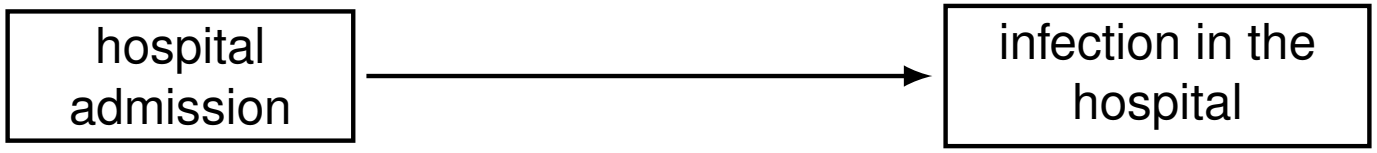
Research Questions
- Examples

Rates and Risks
- Two approaches to competing risks analysis

Estimation
- Multi-state approach
- Subdistribution approach
- Regression

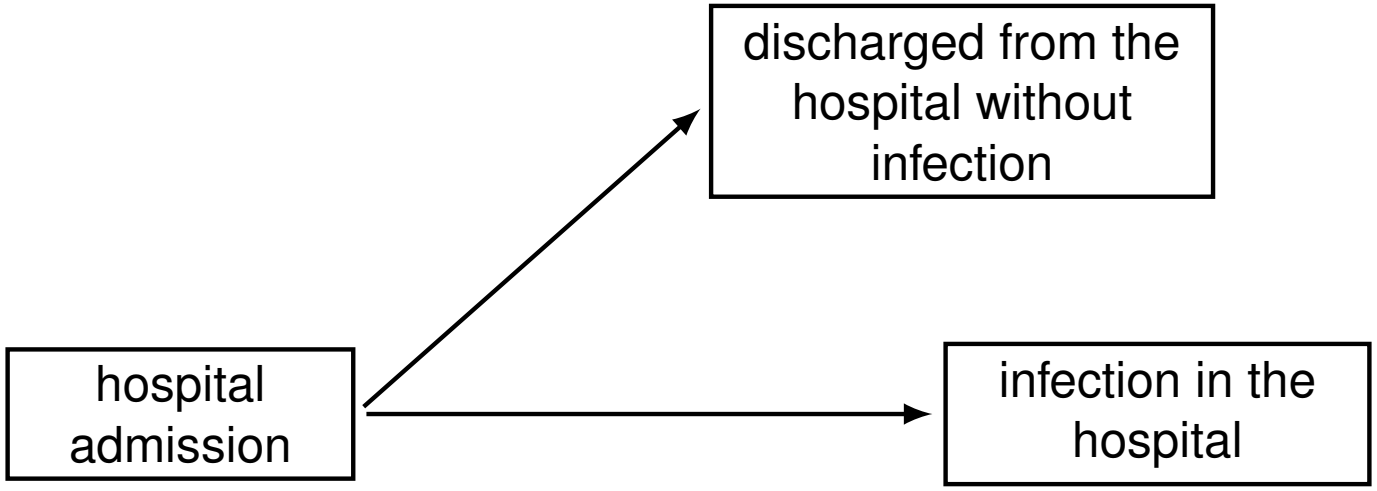
Summary
- Marginal versus competing risks
- Which appoach to choose?

I: Time to staphylococcus infection during hospital stay



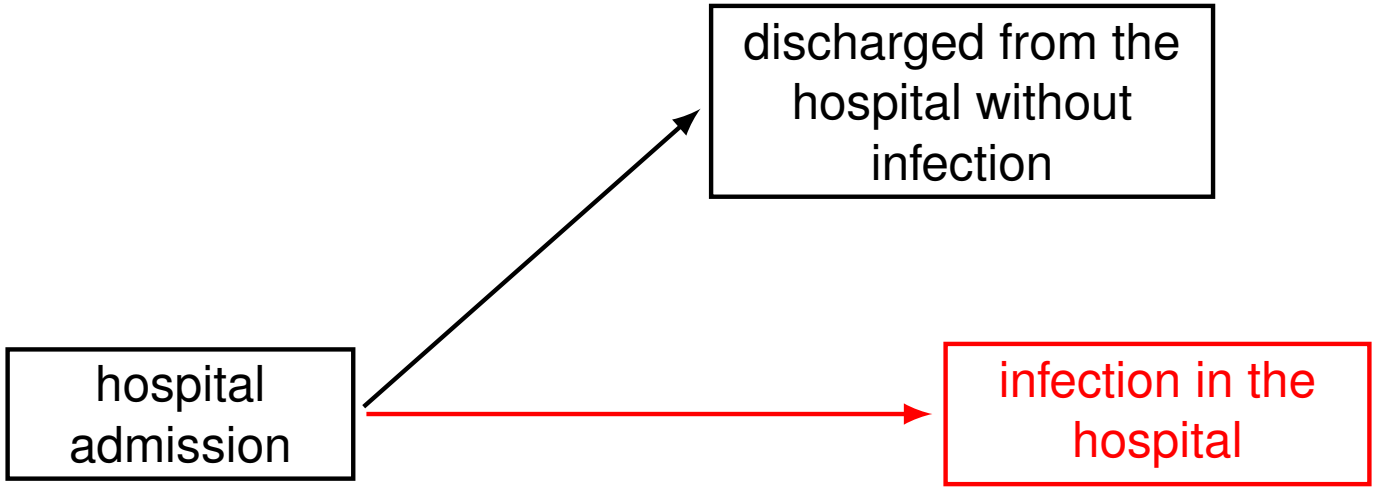
- Etiology (biological question): infection risk in hospital. What would happen if everyone stayed in hospital?

I: Time to staphylococcus infection during hospital stay



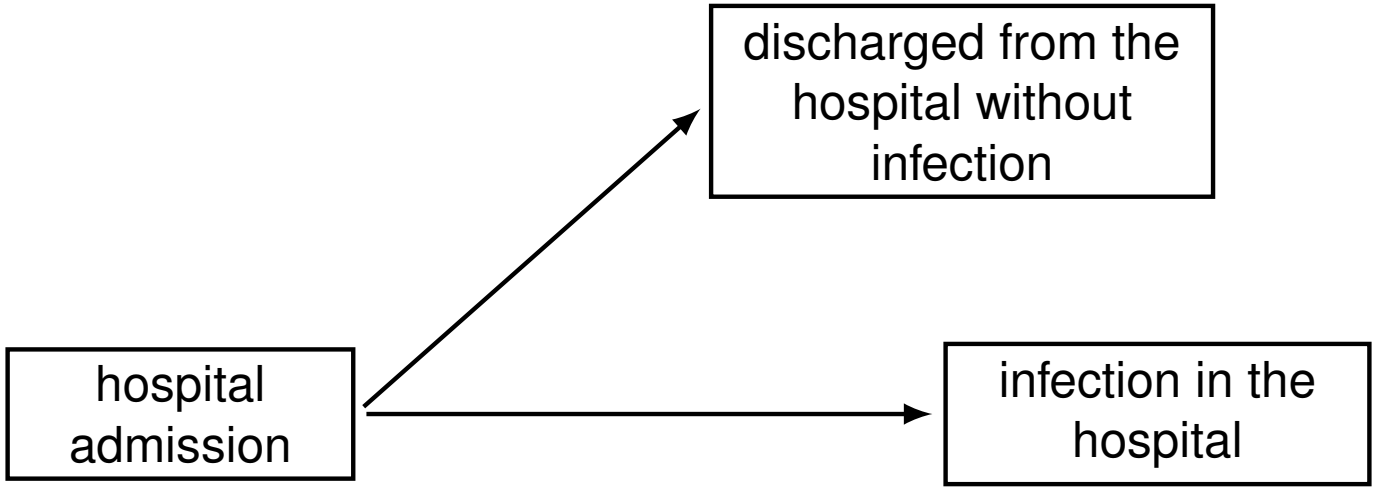
- Etiology (biological question): infection risk in hospital. What would happen if everyone stayed in hospital?

I: Time to staphylococcus infection during hospital stay



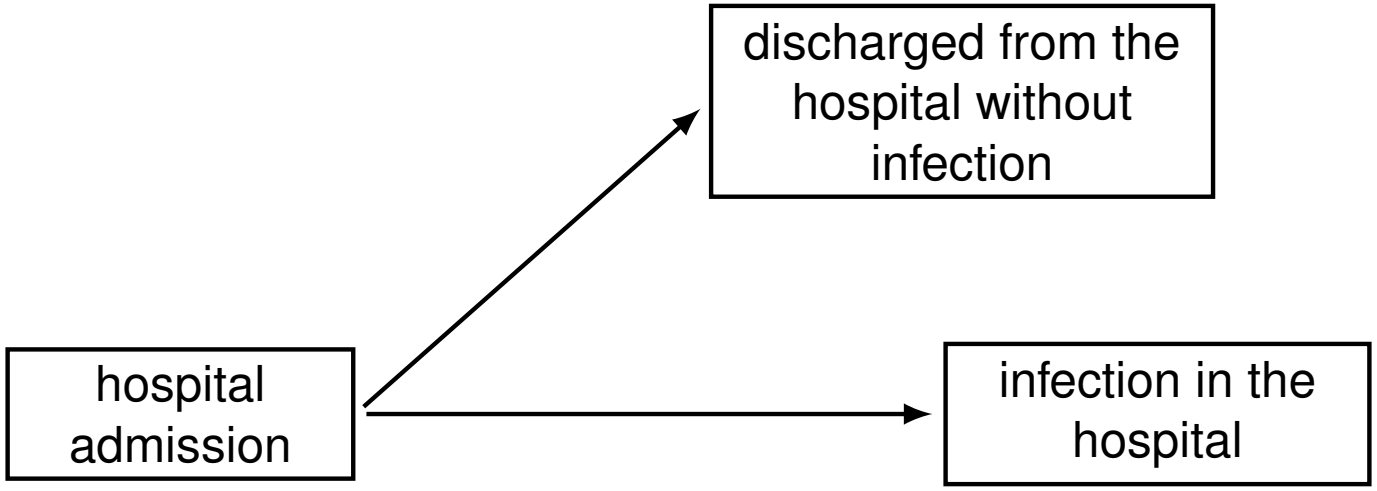
- Etiology (biological question): infection risk in hospital. What would happen if everyone stayed in hospital?
  - Marginal distribution/net risk

I: Time to staphylococcus infection during hospital stay



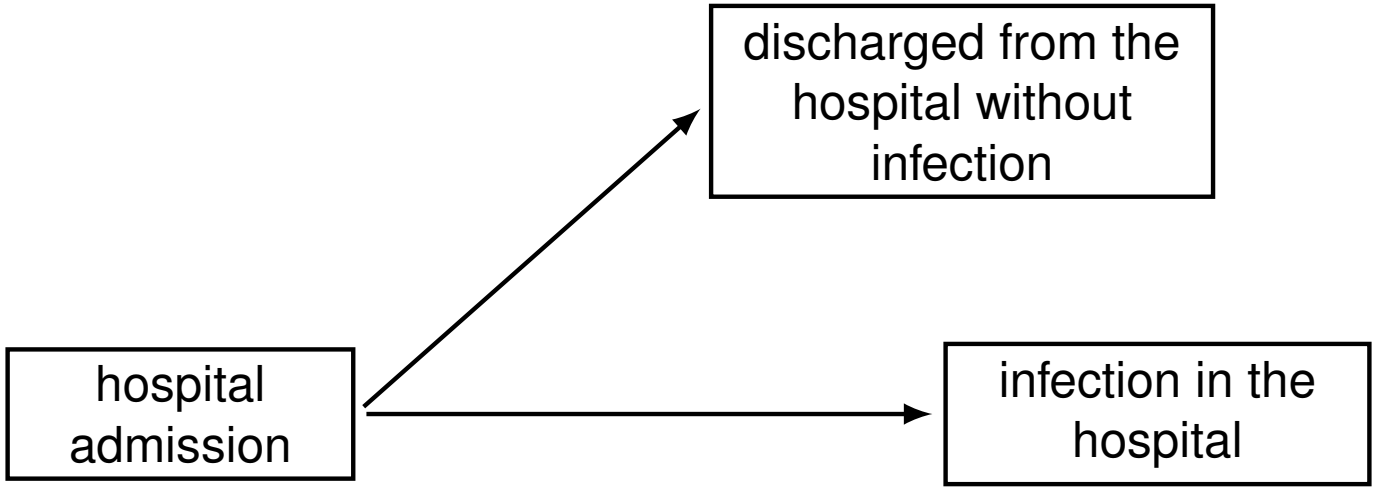
- Etiology (biological question): infection risk in hospital. What would happen if everyone stayed in hospital?
  - Marginal distribution/net risk
- Predict (clinical question): disease burden due to infection while in hospital; discharge prevents event to occur
  - Cause-specific cumulative incidence/crude risk

I: Time to staphylococcus infection during hospital stay



- Etiology (biological question): infection risk in hospital. What would happen if everyone stayed in hospital?
  - Marginal distribution/net risk
- Predict (clinical question): disease burden due to infection while in hospital; discharge prevents event to occur
  - Cause-specific cumulative incidence/crude risk

I: Time to staphylococcus infection during hospital stay



- Etiology (biological question): infection risk in hospital. What would happen if everyone stayed in hospital?
  - Marginal distribution/net risk
- Predict (clinical question): disease burden due to infection while in hospital; discharge prevents event to occur
  - Cause-specific cumulative incidence/crude risk
- Performance when comparing two hospitals may depend on type of question

Estimation with complete follow-up (artificial data)

week	0-1	1-2	2-3	3-4	4-5	5-6	6-7	> 7
infection	1	2	6	11	9	11	2	18
cumulative	1	3	9	20	29	40	42	60
discharge	5	9	6	6	9	12	4	35
cumulative	5	14	20	26	35	47	51	86

Estimation with complete follow-up (artificial data)

week	0-1	1-2	2-3	3-4	4-5	5-6	6-7	> 7
infection	1	2	6	11	9	11	2	18
cumulative	1	3	9	20	29	40	42	60
discharge	5	9	6	6	9	12	4	35
cumulative	5	14	20	26	35	47	51	86

- Marginal: discharged individuals interpreted as censored. Kaplan-Meier: represented by the ones that remain in hospital

Estimation with complete follow-up (artificial data)

week	0-1	1-2	2-3	3-4	4-5	5-6	6-7	> 7
infection	1	2	6	11	9	11	2	18
cumulative	1	3	9	20	29	40	42	60
discharge	5	9	6	6	9	12	4	35
cumulative	5	14	20	26	35	47	51	86

- Marginal: discharged individuals interpreted as censored. Kaplan-Meier: represented by the ones that remain in hospital
- Discharge competing risk: Crude risk estimated as frequency of events:

  $\widehat{P}(\text{infection} \leq 6\text{weeks}) = 40/146$

  $\widehat{P}(\text{discharge} \leq 6\text{weeks}) = 47/146$

Estimation with complete follow-up (artificial data)

week	0-1	1-2	2-3	3-4	4-5	5-6	6-7	> 7
infection	1	2	6	11	9	11	2	18
cumulative	1	3	9	20	29	40	42	60
discharge	5	9	6	6	9	12	4	35
cumulative	5	14	20	26	35	47	51	86

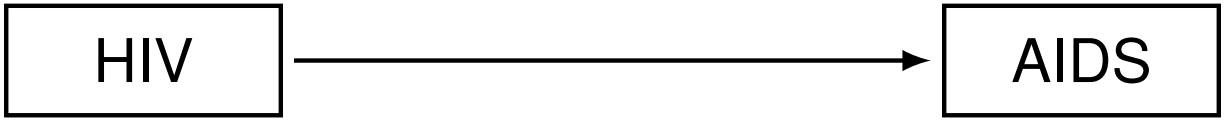
- Marginal: discharged individuals interpreted as censored. Kaplan-Meier: represented by the ones that remain in hospital
- Discharge competing risk: Crude risk estimated as frequency of events:

  $\widehat{P}(\text{infection} \leq 6\text{weeks}) = 40/146$

  $\widehat{P}(\text{discharge} \leq 6\text{weeks}) = 47/146$

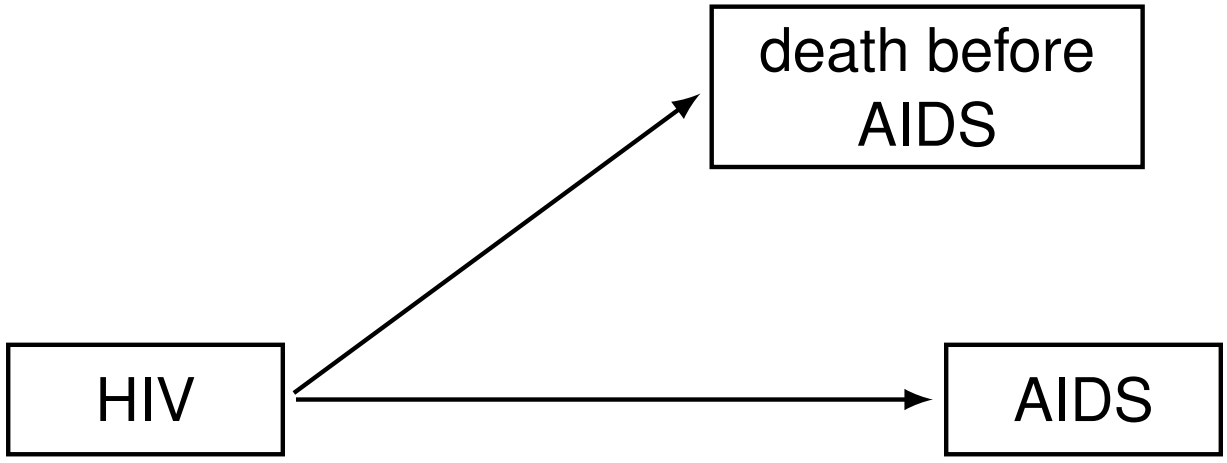
  Individuals with competing event remain in denominator, competing event ignored in estimation

II: Time from HIV infection to AIDS



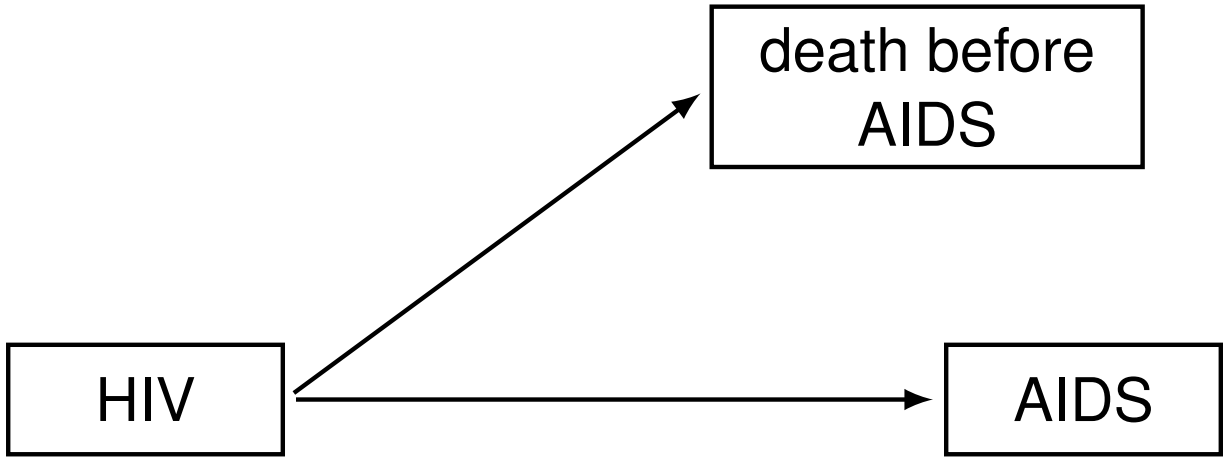
- Compare men who have sex with men (MSM) and injecting drug users (IDU)
- IDUs expected to have faster progression to AIDS

II: Time from HIV infection to AIDS



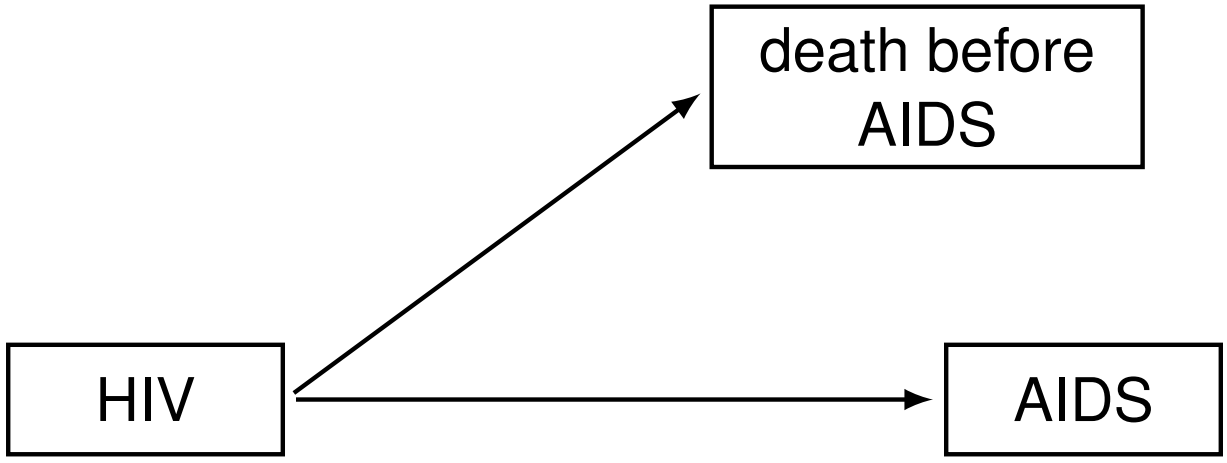
- Compare men who have sex with men (MSM) and injecting drug users (IDU)
- IDUs expected to have faster progression to AIDS
- Data from Amsterdam Cohort Studies: 99 IDU; 127 MSM

II: Time from HIV infection to AIDS



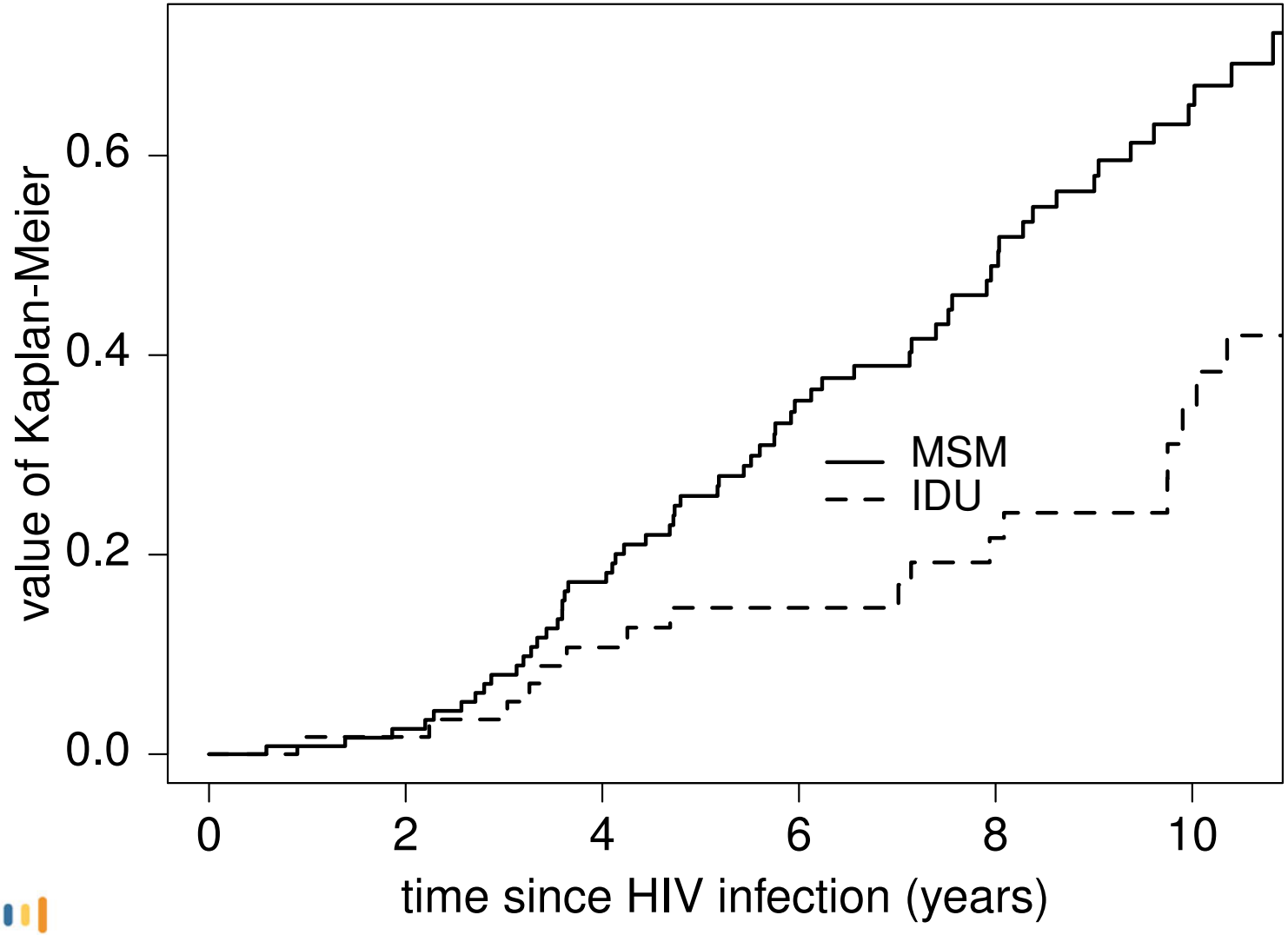
- Compare men who have sex with men (MSM) and injecting drug users (IDU)
- IDUs expected to have faster progression to AIDS
- Data from Amsterdam Cohort Studies: 99 IDU; 127 MSM
- Interest in time to AIDS if there were no pre-AIDS death. Interest in etiology and marginal distribution

II: Time from HIV infection to AIDS

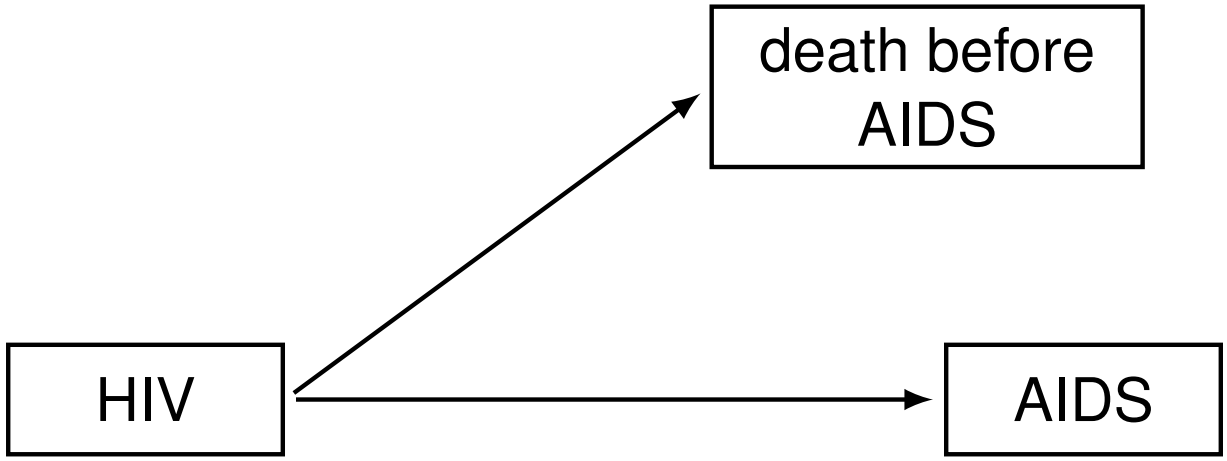


- Compare men who have sex with men (MSM) and injecting drug users (IDU)
- IDUs expected to have faster progression to AIDS
- Data from Amsterdam Cohort Studies: 99 IDU; 127 MSM
- Interest in time to AIDS if there were no pre-AIDS death. Interest in etiology and marginal distribution
- Kaplan-Meier: leave risk set at death before AIDS

Kaplan-Meier: IDU much slower progression ($p = 0.001$)



II: Time from HIV infection to AIDS



- Compare men who have sex with men (MSM) and injecting drug users (IDU)
- IDUs expected to have faster progression to AIDS
- Data from Amsterdam Cohort Studies: 99 IDU; 127 MSM
- Interest in time to AIDS if there were no pre-AIDS death. Interest in etiology and marginal distribution
- Kaplan-Meier: leave risk set at death before AIDS
  Assumption: deaths represented by those that do not die

Explanation: dependent censoring

- Extra information on cause of death before AIDS

	IDU	MSM
Reason of death	Number	
HIV related infections	3	0
overdose/suicide	6	0
violence/accident	2	0
liver cirrhosis	2	0
cancer	0	1
heart attack	0	1
unknown	4	3

- Some causes of pre-AIDS death in IDU related to AIDS progression. Censoring close to AIDS, hence net risk estimate for IDU biased downwards

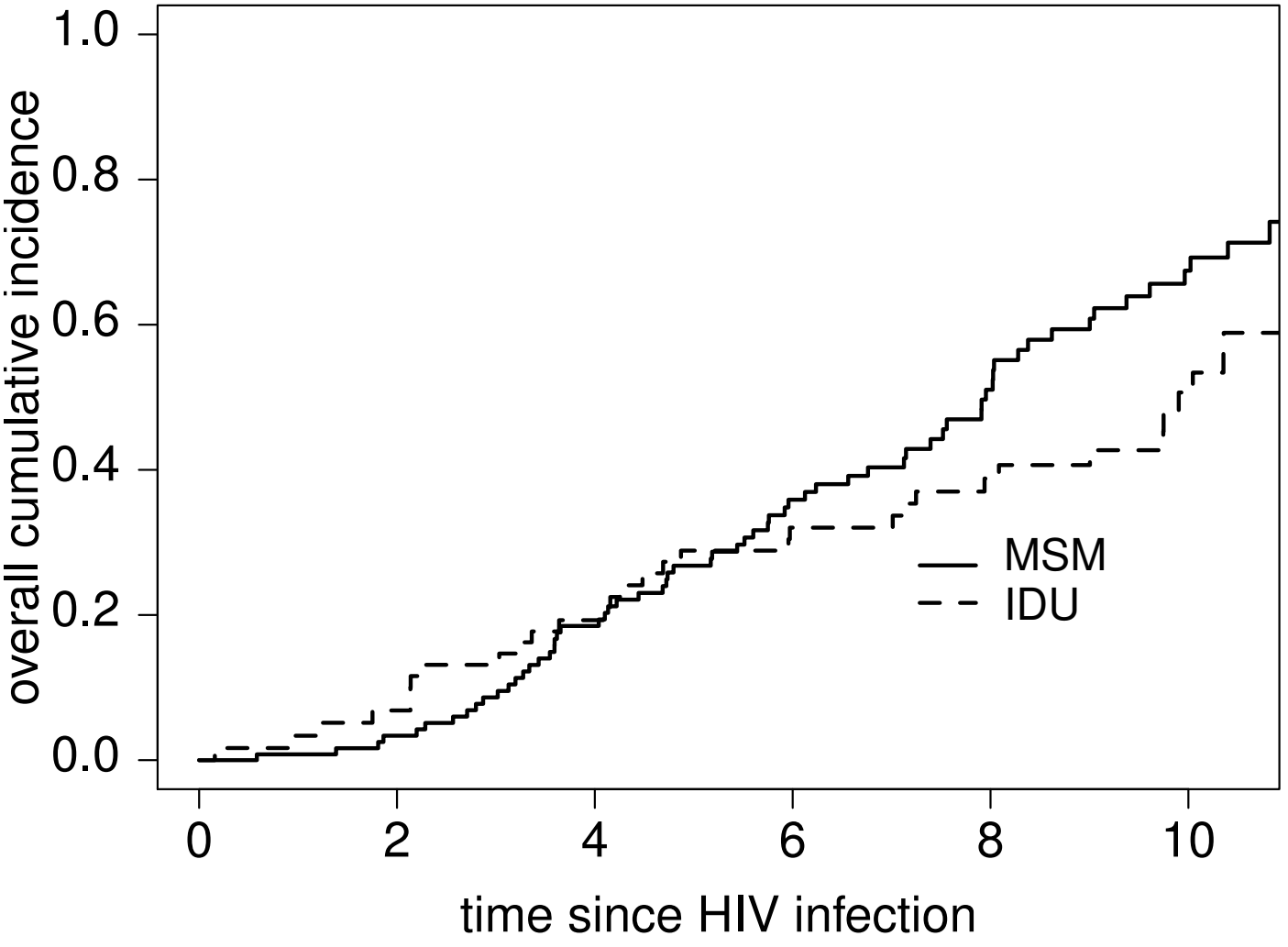
Explanation: dependent censoring

- Extra information on cause of death before AIDS

	IDU	MSM
Reason of death	Number	
HIV related infections	3	0
overdose/suicide	6	0
violence/accident	2	0
liver cirrhosis	2	0
cancer	0	1
heart attack	0	1
unknown	4	3

- Some causes of pre-AIDS death in IDU related to AIDS progression. Censoring close to AIDS, hence net risk estimate for IDU biased downwards
- What if: i) deaths would have developed AIDS right after

i) Combine AIDS and pre-AIDS death



Overall time-to-event distribution (both event types combined)

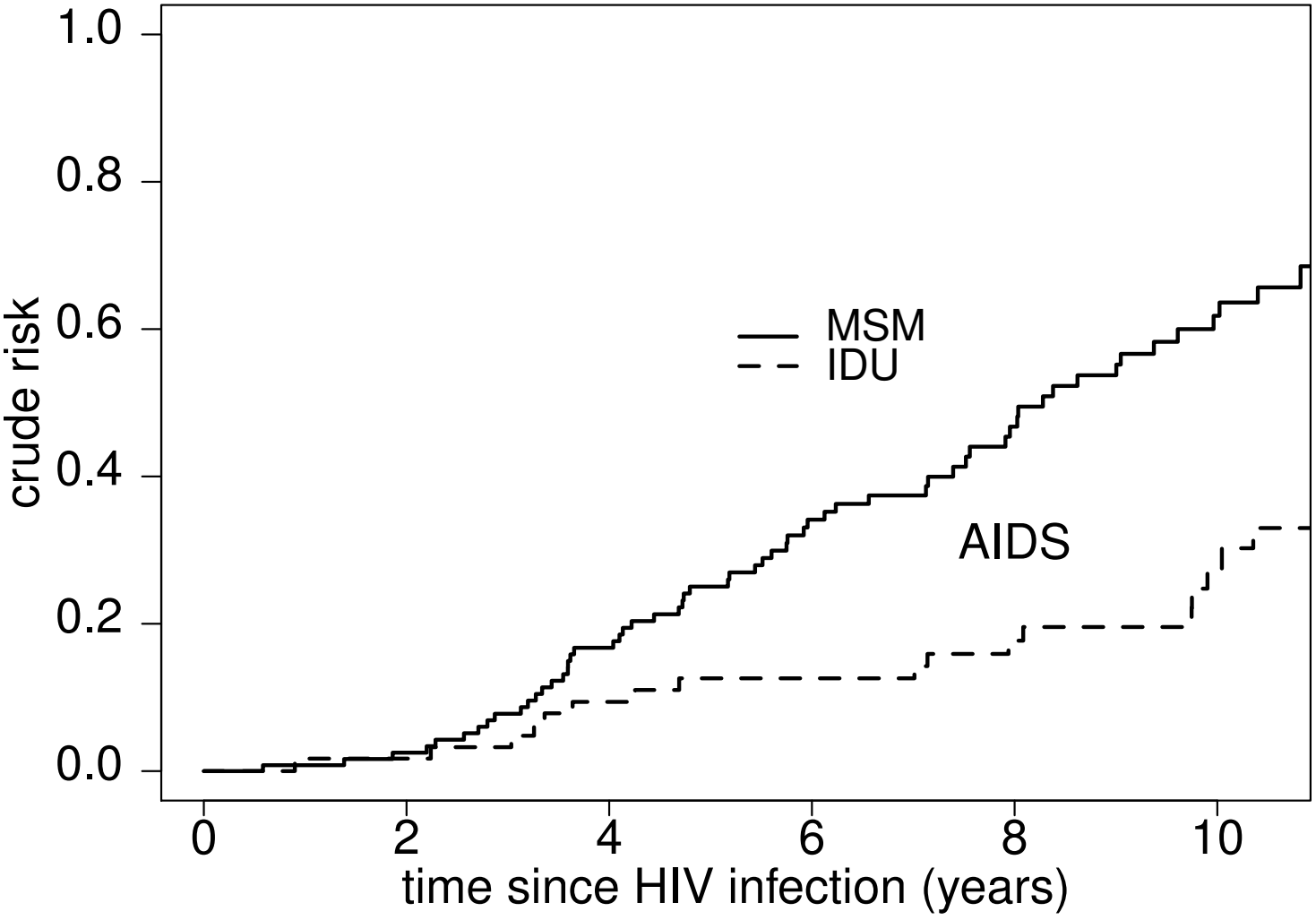
Explanation: dependent censoring

- Extra information on cause of death before AIDS

	IDU	MSM
Reason of death	Number	
HIV related infections	3	0
overdose/suicide	6	0
violence/accident	2	0
liver cirrhosis	2	0
cancer	0	1
heart attack	0	1
unknown	4	3

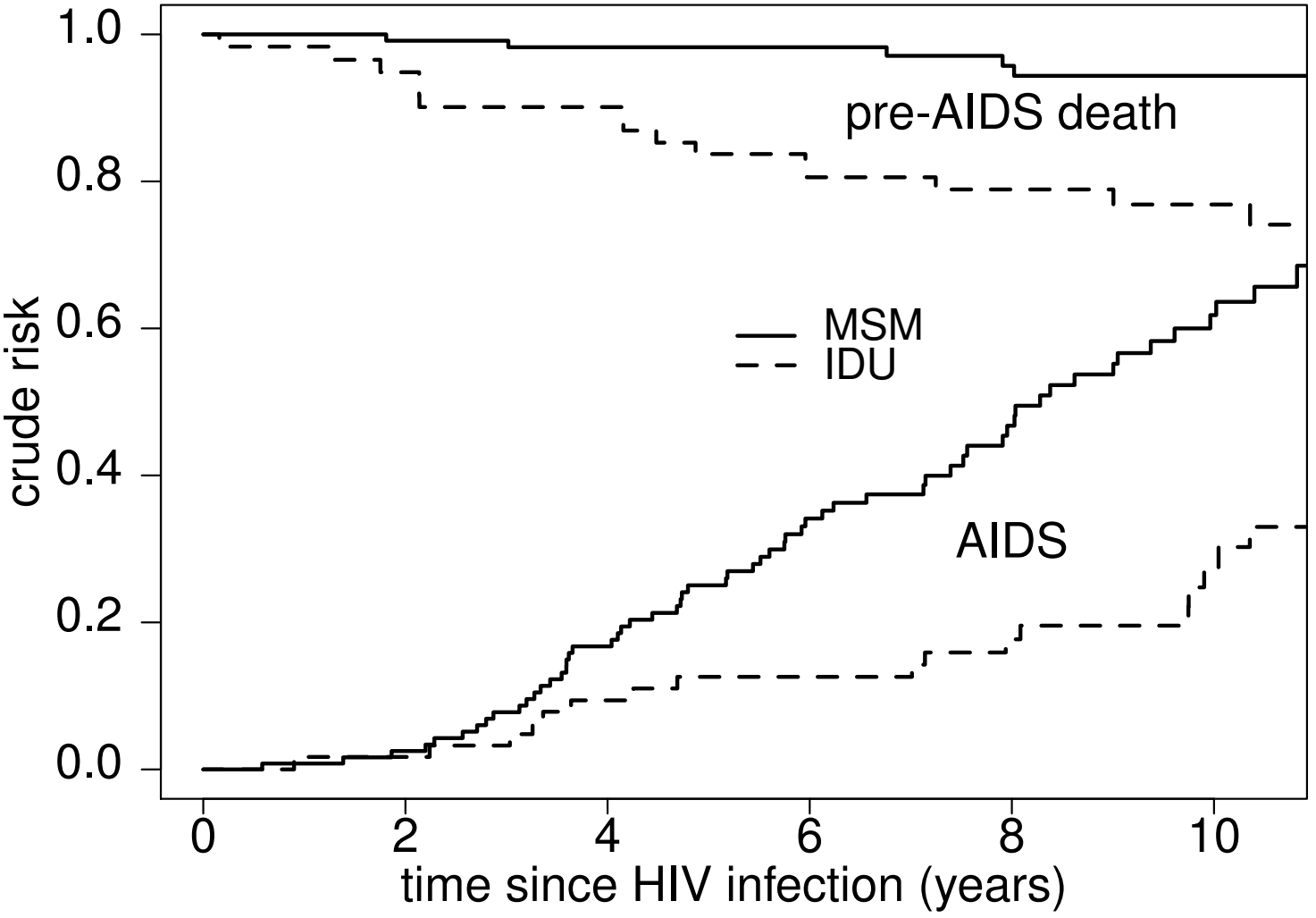
- Some causes of pre-AIDS death in IDU related to AIDS progression. Censoring close to AIDS, hence net risk estimate for IDU biased downwards
- What if: i) deaths would have developed AIDS right after
- What if: ii) deaths would never have developed AIDS

ii) AIDS-specific cumulative incidence



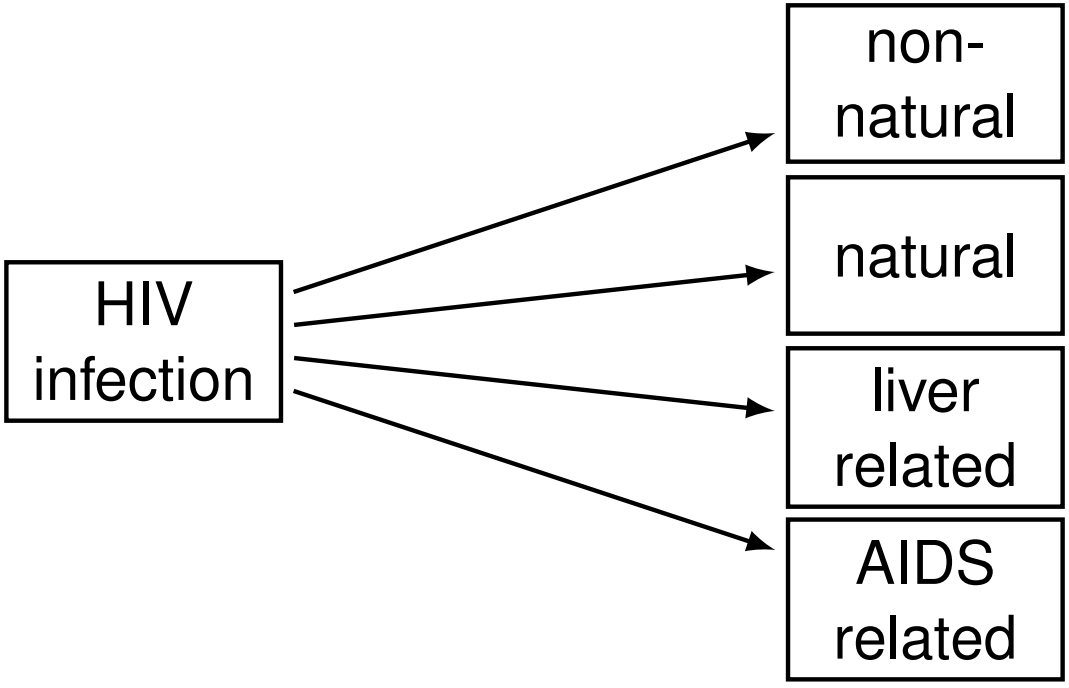
AIDS-specific cumulative incidence (pre-AIDS death prevents AIDS)

ii) AIDS-specific cumulative incidence



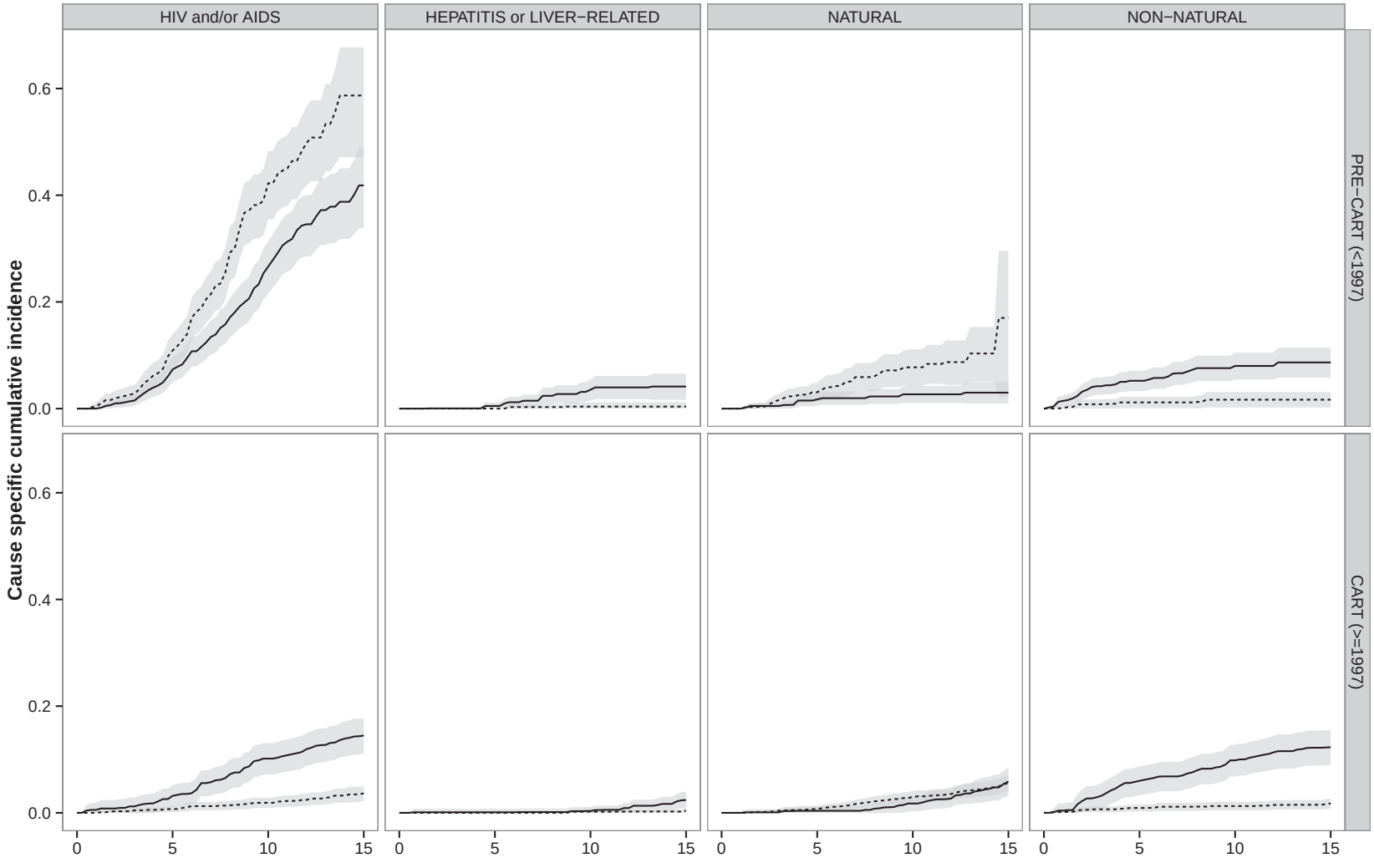
AIDS-specific cumulative incidence (pre-AIDS death prevents AIDS)

III: Causes of death (COD) after HIV infection



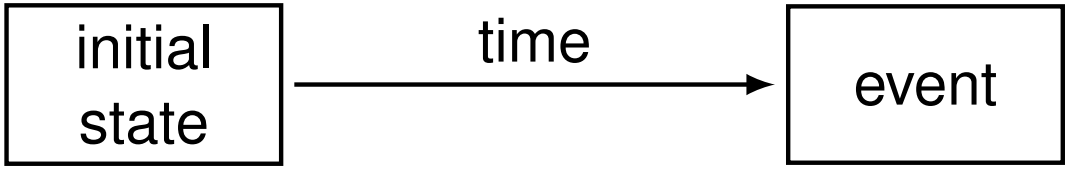
- Has the spectrum in causes of death changed after the introduction of cART (combination anti-retroviral therapy)
- Competing risks analysis most interesting
  No interest in change in AIDS-related death in world in which other COD's do not exist

Cause-specific mortality by calendar period and hepatitis C status



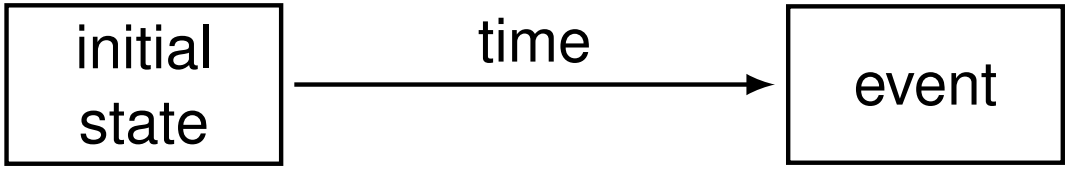
Beyond classical survival analysis

- Classical: transition between two states, one event type.
  "We all die, but not all at the same age"



Beyond classical survival analysis

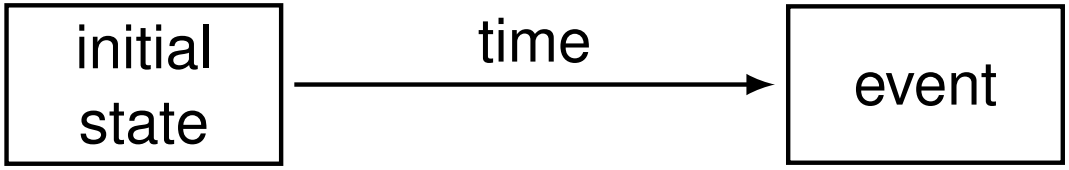
- Classical: transition between two states, one event type.
  "We all die, but not all at the same age"



- Life and death are richer than that
  1. Multiple causes of death. Competing risks:
     "we all die, but not all at the same age and from the same cause"

Beyond classical survival analysis

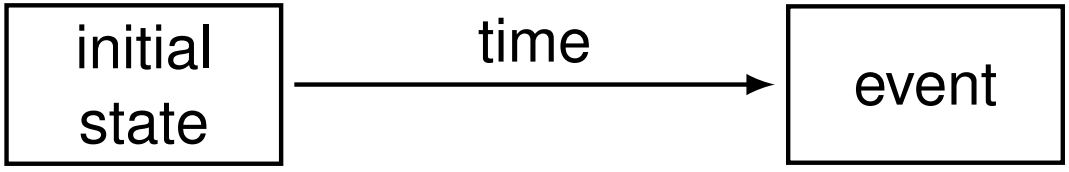
- Classical: transition between two states, one event type.
  "We all die, but not all at the same age"



- Life and death are richer than that
  1. Multiple causes of death. Competing risks:
     "we all die, but not all at the same age and from the same cause"
  2. Intermediate events. Multi-state model:
     "we all die, but not all at the same age, not from the same cause and with different life histories"

Beyond classical survival analysis

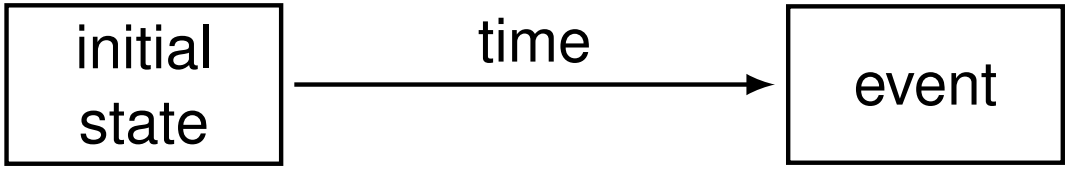
- Classical: transition between two states, one event type.
  "We all die, but not all at the same age"



- Life and death are richer than that
  1. Multiple causes of death. Competing risks:
     "we all die, but not all at the same age and from the same cause"
  2. Intermediate events. Multi-state model:
     "we all die, but not all at the same age, not from the same cause and with different life histories"
- Two components
  - Events/Transitions. Initial, intermediate and final states

Beyond classical survival analysis

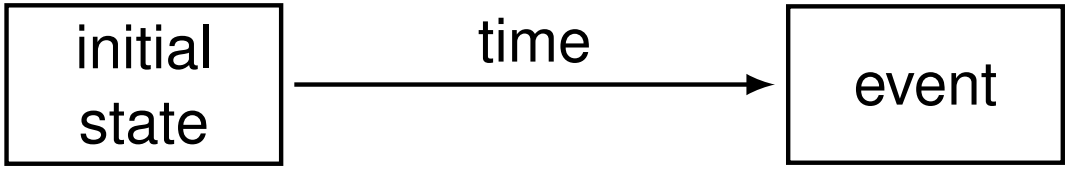
- Classical: transition between two states, one event type.
  "We all die, but not all at the same age"



- Life and death are richer than that
  1. Multiple causes of death. Competing risks:
     "we all die, but not all at the same age and from the same cause"
  2. Intermediate events. Multi-state model:
     "we all die, but not all at the same age, not from the same cause and with different life histories"
- Two components
  - Events/Transitions. Initial, intermediate and final states
  - Time. What is the time origin? Multiple time scales?

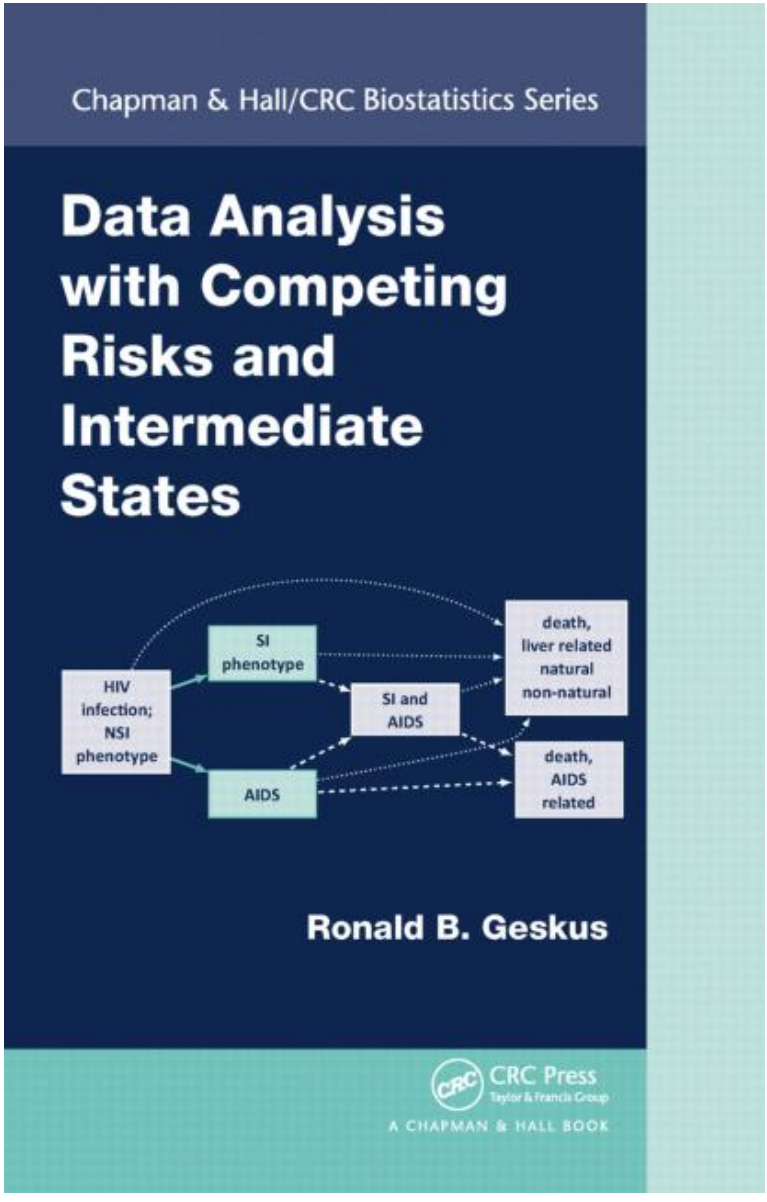
Beyond classical survival analysis

- Classical: transition between two states, one event type.
  "We all die, but not all at the same age"



- Life and death are richer than that
  1. Multiple causes of death. Competing risks:
     "we all die, but not all at the same age and from the same cause"
  2. Intermediate events. Multi-state model:
     "we all die, but not all at the same age, not from the same cause and with different life histories"
- Two components
  - Events/Transitions. Initial, intermediate and final states
  - Time. What is the time origin? Multiple time scales?

Published by CRC Press, 2015



Outline

Research Questions
- Examples

Rates and Risks
- Two approaches to competing risks analysis

Estimation
- Multi-state approach
- Subdistribution approach
- Regression

Summary
- Marginal versus competing risks
- Which appoach to choose?

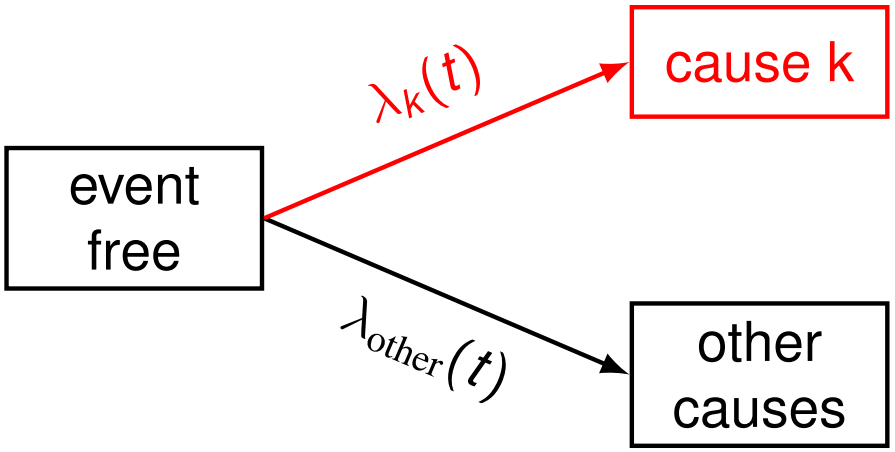
Setup and notation

- Competing risks: $E \in \{1, \ldots, K\}$
- $T \sim F$ time to event (of any type); $F(t) = \mathrm{P}(T \leq t)$
- Overall hazard h: $\mathrm{P}(T > t) = \exp\{-\int_0^t h(s)ds\}$
- Notation: $\overline{F}(t) = 1 - F(t) = \mathrm{P}(T > t)$

Setup and notation

- Competing risks: $E \in \{1, \ldots, K\}$
- $T \sim F$ time to event (of any type); $F(t) = \mathrm{P}(T \leq t)$
- Overall hazard h: $\mathrm{P}(T > t) = \exp\{-\int_0^t h(s)ds\}$
- Notation: $\overline{F}(t) = 1 - F(t) = \mathrm{P}(T > t)$
- Cause-specific cumulative incidence: $F_k(t) = \mathrm{P}(T \leq t, E = k)$

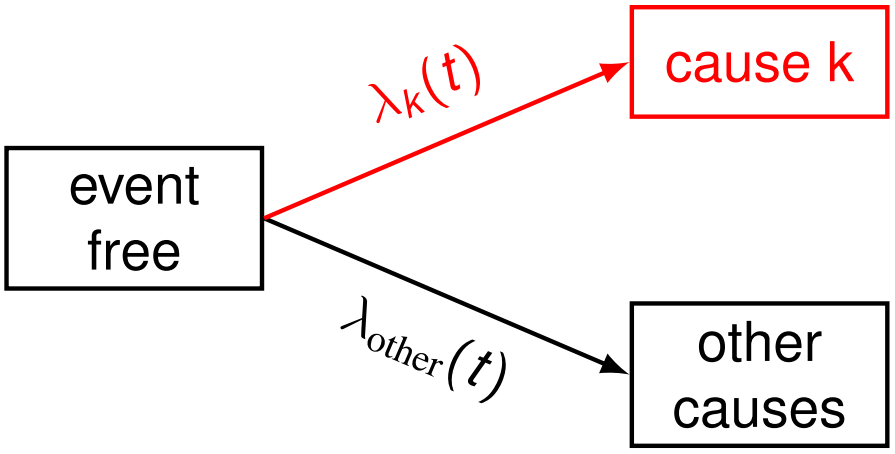
I: The multi-state approach: cause-specific hazard



- Transition rate to cause k. For continuous distribution:

$$\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{P}(t \leq T < t + \Delta t, E = k \mid T \geq t)}{\Delta t}$$

I: The multi-state approach: cause-specific hazard

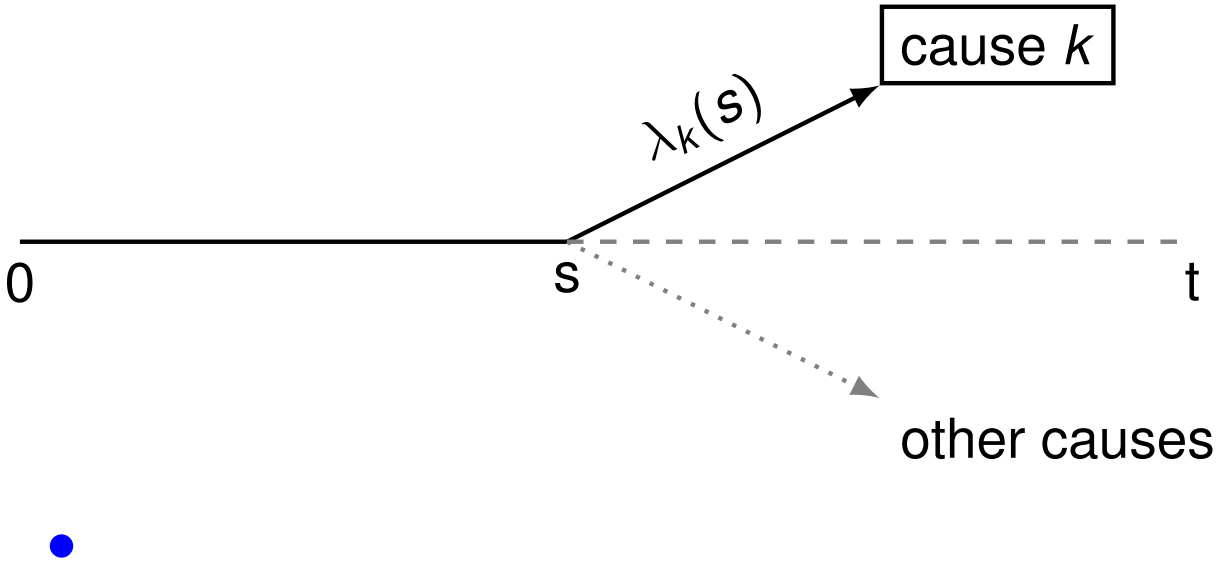


- Transition rate to cause k. For continuous distribution:

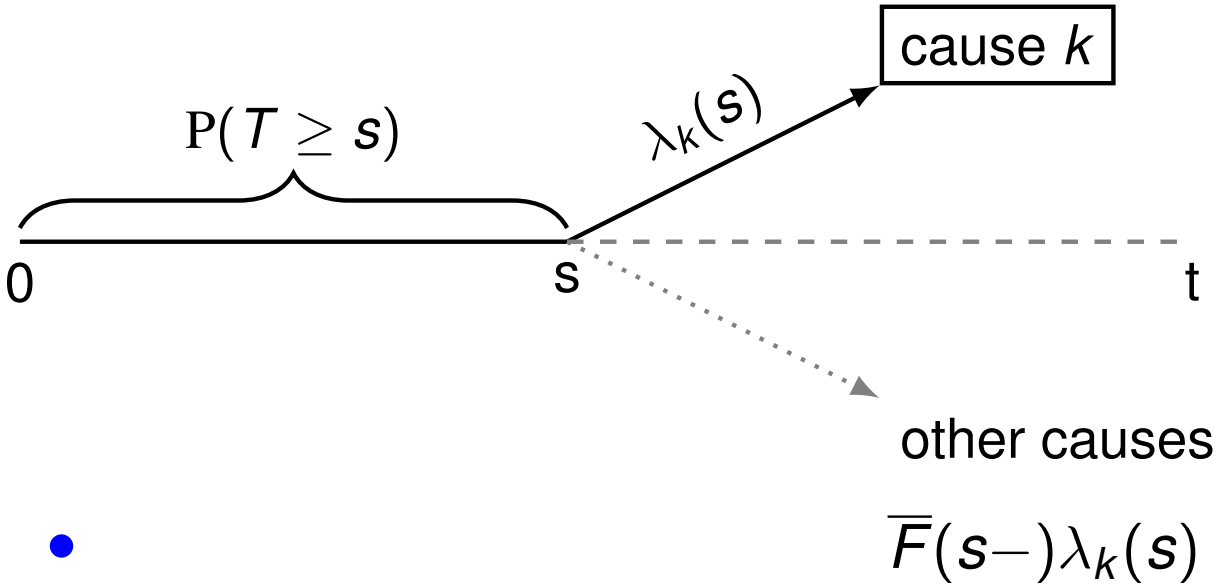
$$\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{P}(t \leq T < t + \Delta t, E = k \mid T \geq t)}{\Delta t}$$

- Sum over causes is overall hazard: $\sum_{e=1}^{K} \lambda_e(t) = h(t)$
- Cause-specific hazard directly generalizes to multi-state setting (called transition hazard)

From hazard to cumulative scale $\mathrm{P}(T \leq t, E = k)$

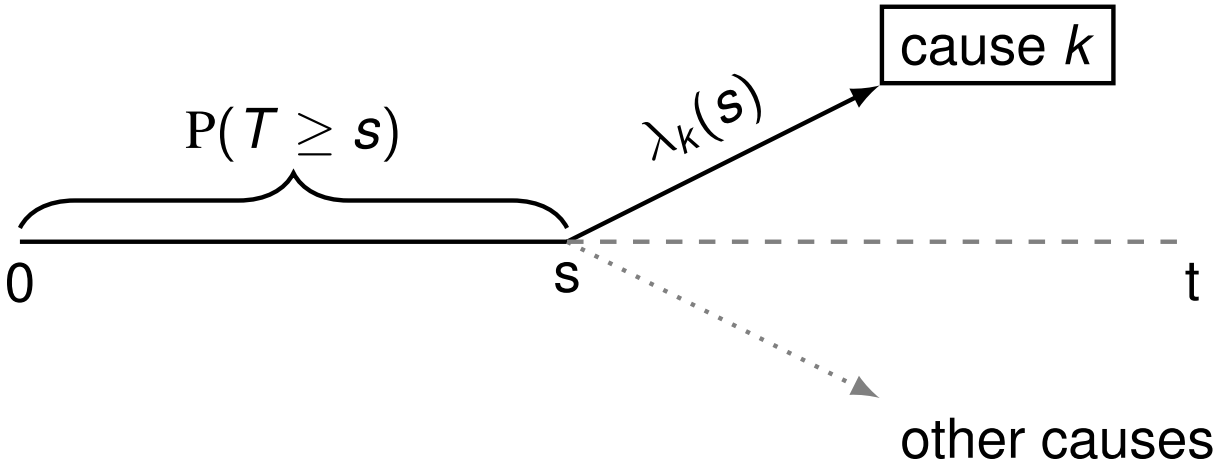


From hazard to cumulative scale $\mathrm{P}(T \leq t, E = k)$



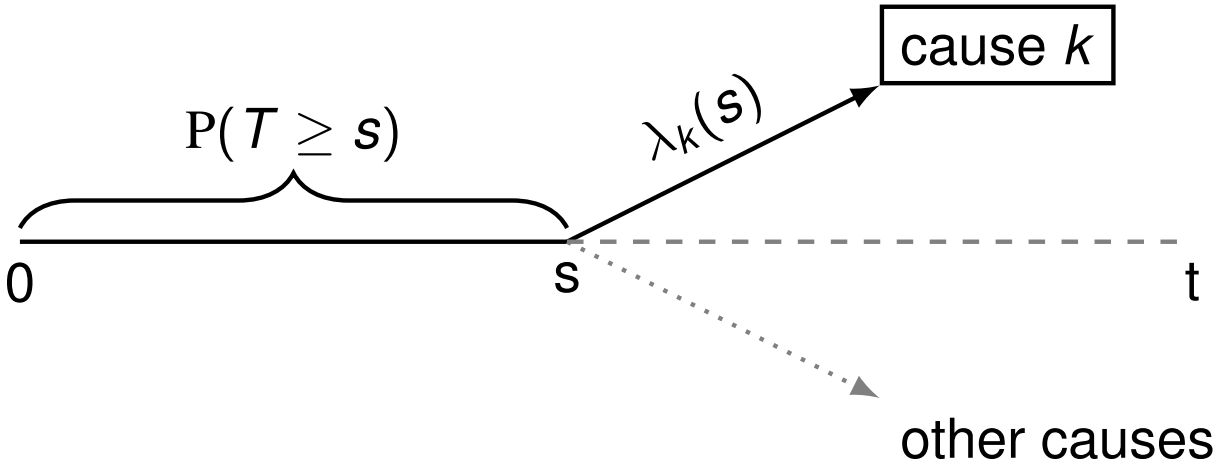
- $\overline{F}(s-)\lambda_k(s)$

From hazard to cumulative scale $\mathrm{P}(T \leq t, E = k)$



- $F_k(t) = \mathrm{P}(T \leq t, E = k) = \int_0^t \overline{F}(s-)\lambda_k(s)ds$

From hazard to cumulative scale $\mathrm{P}(T \leq t, E = k)$



- $F_k(t) = \mathrm{P}(T \leq t, E = k) = \int_0^t \overline{F}(s-)\lambda_k(s)ds$
- Depends on all cause-specific hazards via overall "survival"

$$\overline{F}(s) = \exp\left\{-\int_0^s h(u)du\right\} = \exp\left\{-\sum_{e=1}^{K}\int_0^s \lambda_e(u)du\right\}$$

II: The subdistribution approach

Setup and notation

- Competing risks: $E \in \{1, \ldots, K\}$

- Cause-specific cumulative incidence:
  $F_k(t) = \mathrm{P}(T \leq t, E = k)$
- Subdistribution random variable $T_k \sim F_k$:
  $T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\}$

Setup and notation

- Competing risks: $E \in \{1, \ldots, K\}$

- Cause-specific cumulative incidence:
  $F_k(t) = \mathrm{P}(T \leq t, E = k)$
- Subdistribution random variable $T_k \sim F_k$:
  $T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\}$
- $\mathrm{P}(T_k \leq t) = \mathrm{P}(T \leq t, E = k), \qquad \overline{F_k} = 1 - F_k = \mathrm{P}(T_k > t)$

Setup and notation

- Competing risks: $E \in \{1, \ldots, K\}$

- Cause-specific cumulative incidence:
  $F_k(t) = \mathrm{P}(T \leq t, E = k)$
- Subdistribution random variable $T_k \sim F_k$:
  $T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\}$
- $\mathrm{P}(T_k \leq t) = \mathrm{P}(T \leq t, E = k), \qquad \overline{F_k} = 1 - F_k = \mathrm{P}(T_k > t)$
- Subdistribution hazard h_k:

$$\mathrm{P}(T_k > t) = \exp\left\{ - \int_0^t h_k(s) ds \right\}$$

Setup and notation

- Competing risks: $E \in \{1, \ldots, K\}$
- Overall hazard h: $\mathrm{P}(T > t) = \exp\{-\int_0^t h(s)ds\}$
- Cause-specific cumulative incidence: $F_k(t) = \mathrm{P}(T \leq t, E = k)$
- Subdistribution random variable $T_k \sim F_k$: $T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\}$
- $\mathrm{P}(T_k \leq t) = \mathrm{P}(T \leq t, E = k), \qquad \overline{F_k} = 1 - F_k = \mathrm{P}(T_k > t)$
- Subdistribution hazard h_k:

$$\mathrm{P}(T_k > t) = \exp\left\{-\int_0^t h_k(s)ds\right\}$$

II: The subdistribution approach

- Subdistribution hazard
  $(T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\})$:

$$h_k(s) = \lim_{\Delta s \downarrow 0} \frac{1}{\Delta s} \times \mathrm{P}\{s \leq T_k < s + \Delta s \,|\, T_k \geq s\}$$

II: The subdistribution approach

- Subdistribution hazard
  $(T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\})$:

$$h_k(s) = \lim_{\Delta s \downarrow 0} \frac{1}{\Delta s} \times \mathrm{P}\{s \leq T_k < s + \Delta s \,|\, T_k \geq s\}$$

$$= \lim_{\Delta s \downarrow 0} \frac{\frac{1}{\Delta s}\mathrm{P}\{s \leq T < s + \Delta s, E = k\}}{\mathrm{P}\{T \geq s \text{ or } (T < s, E \neq k)\}}$$

- Denominator: event free or with earlier competing event

II: The subdistribution approach

- Subdistribution hazard
  $(T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\})$:

$$h_k(s) = \lim_{\Delta s \downarrow 0} \frac{1}{\Delta s} \times \mathrm{P}\{s \leq T_k < s + \Delta s \,|\, T_k \geq s\}$$

$$= \lim_{\Delta s \downarrow 0} \frac{\frac{1}{\Delta s}\mathrm{P}\{s \leq T < s + \Delta s, E = k\}}{\mathrm{P}\{T \geq s \text{ or } (T < s, E \neq k)\}}$$

- Denominator: event free or with earlier competing event
- Interpretation controversial
  - Not a rate in epidemiological sense,

II: The subdistribution approach

- Subdistribution hazard
  $(T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\})$:

$$h_k(s) = \lim_{\Delta s \downarrow 0} \frac{1}{\Delta s} \times \mathrm{P}\{s \leq T_k < s + \Delta s \,|\, T_k \geq s\}$$

$$= \lim_{\Delta s \downarrow 0} \frac{\frac{1}{\Delta s}\mathrm{P}\{s \leq T < s + \Delta s, E = k\}}{\mathrm{P}\{T \geq s \text{ or } (T < s, E \neq k)\}}$$

- Denominator: event free or with earlier competing event
- Interpretation controversial
  - Not a rate in epidemiological sense,
  - unless we can assume that those with the competing event were immune for the event of interest (cure model)

II: The subdistribution approach

- Subdistribution hazard
  $(T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\})$:

$$h_k(s) = \lim_{\Delta s \downarrow 0} \frac{1}{\Delta s} \times \mathrm{P}\{s \leq T_k < s + \Delta s \,|\, T_k \geq s\}$$

$$= \lim_{\Delta s \downarrow 0} \frac{\frac{1}{\Delta s}\mathrm{P}\{s \leq T < s + \Delta s, E = k\}}{\mathrm{P}\{T \geq s \text{ or } (T < s, E \neq k)\}}$$

- Denominator: event free or with earlier competing event
- Interpretation controversial
  - Not a rate in epidemiological sense,
  - unless we can assume that those with the competing event were immune for the event of interest (cure model)
- One-to-one relation with crude risk

$$\overline{F_k}(t) = \prod_{t_l \leq t} \left\{1 - h_k(t_l)\right\} \quad \text{or} \quad \overline{F_k}(t) = \exp\left\{-\int_0^t h_k(u)du\right\}$$

oucru

Rates and risks in competing risks setting

	hazard		cumulative	
competing risks	marginal	*	net risk	*
			marginal survival function	
			marginal cumulative incidence	
	cause-specific	λ_k	no corresponding quantity	
	subdistribution	h_k	crude risk	$F_k(t)$
			cause-specific cumulative incidence	
combined	overall	h	overall risk	$F(t)$
			overall survival function	
			overall cumulative incidence	

* Doesn't play a role in competing risks analyses; therefore, no notation is introduced

Rates and risks in competing risks setting

	hazard		cumulative	
competing risks	marginal	*	net risk	*
			marginal survival function	
			marginal cumulative incidence	
	cause-specific	λ_k	no corresponding quantity	
	subdistribution	h_k	crude risk	$F_k(t)$
			cause-specific cumulative incidence	
combined	overall	h	overall risk	$F(t)$
			overall survival function	
			overall cumulative incidence	

* Doesn't play a role in competing risks analyses; therefore, no notation is introduced

Observed data

$\{(x_1, e_1\delta_1), \ldots, (x_N, e_N\delta_N)\}$

- $x_i = \min\{t_i, c_i\}$, $\delta_i = \{t_i \leq c_i\}$, $e_i \in \{1, \ldots, K\}$
- $t_{(i)}$ ordered unique event times of any type

Outline

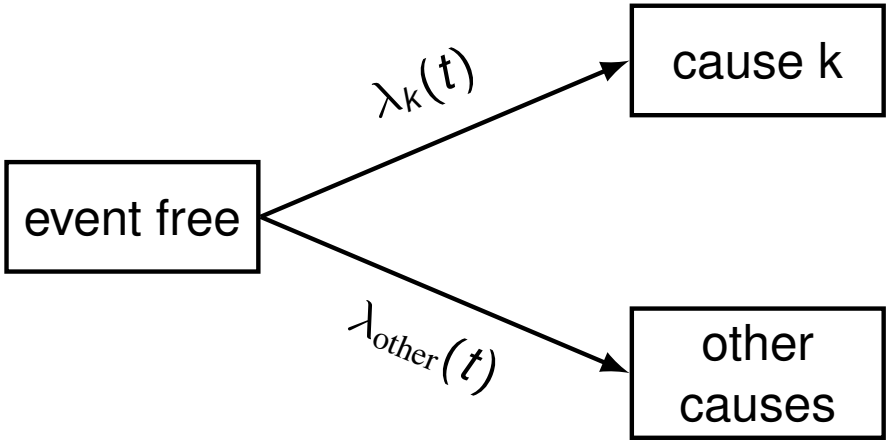
Research Questions
- Examples

Rates and Risks
- Two approaches to competing risks analysis

Estimation
- Multi-state approach
- Subdistribution approach
- Regression

Summary
- Marginal versus competing risks
- Which appoach to choose?

Cause-specific hazard



- Individuals with a competing event are no longer at risk
  $\Longrightarrow$ leave the risk set

$$\widehat{\lambda_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r(t_{(i)})} \ .$$

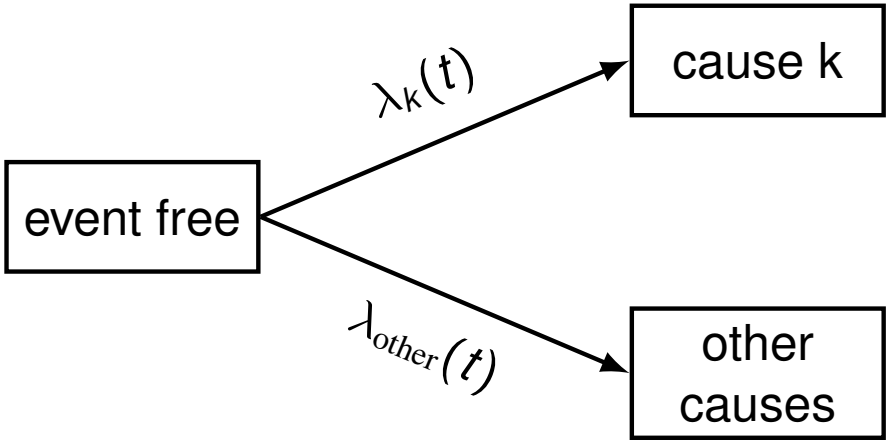
Observed data

$\{(x_1, e_1\delta_1), \ldots, (x_N, e_N\delta_N)\}$

- $x_i = \min\{t_i, c_i\}$, $\delta_i = \{t_i \leq c_i\}$, $e_i \in \{1, \ldots, K\}$
- $t_{(i)}$ ordered unique event times of any type
- $r(t_{(i)})$ number observed at risk

- $d_k(t_{(i)})$ number of events at $t_{(i)}$ of type k

Cause-specific hazard



- Individuals with a competing event are no longer at risk $\Longrightarrow$ leave the risk set

$$\widehat{\lambda_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r(t_{(i)})} \ .$$

- Standard rate estimation. Same estimator as marginal hazard, but different interpretation, unless censoring due to competing risks is non-informative

Aalen-Johansen estimator of F_k

- Plug-in estimator based on $F_k(t) = \int_0^t \mathrm{P}\{T \geq s\}\lambda_k(s)ds$:

$$\widehat{F_k}^{\mathrm{AJ}}(t) = \sum_{i:t_{(i)} \leq t} \widehat{\overline{F}}^{\mathrm{PL}}(t_{(i)}-) \times \widehat{\lambda_k}(t_{(i)}) \text{ with}$$

$$\widehat{\lambda_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r(t_{(i)})} \qquad \text{cause specific hazard}$$

Aalen-Johansen estimator of F_k

- Plug-in estimator based on $F_k(t) = \int_0^t \mathrm{P}\{T \geq s\}\lambda_k(s)ds$:

$$\widehat{F_k}^{\mathrm{AJ}}(t) = \sum_{i:t_{(i)} \leq t} \widehat{\overline{F}}^{\mathrm{PL}}(t_{(i)}-) \times \widehat{\lambda_k}(t_{(i)}) \text{ with}$$

$$\widehat{\lambda_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r(t_{(i)})} \qquad \text{cause specific hazard}$$

Observed data

$\{(x_1, e_1\delta_1), \ldots, (x_N, e_N\delta_N)\}$

- $x_i = \min\{t_i, c_i\}$, $\delta_i = \{t_i \leq c_i\}$, $e_i \in \{1, \ldots, K\}$
- $t_{(i)}$ ordered unique event times of any type
- $r(t_{(i)})$ number observed at risk

- $d_k(t_{(i)})$ number of events at $t_{(i)}$ of type k
- $d(t_{(i)})$ total number of events at $t_{(i)}$

Aalen-Johansen estimator of F_k

- Plug-in estimator based on $F_k(t) = \int_0^t \mathrm{P}\{T \geq s\}\lambda_k(s)ds$:

$$\widehat{F_k}^{\mathrm{AJ}}(t) = \sum_{i:t_{(i)} \leq t} \widehat{\overline{F}}^{\mathrm{PL}}(t_{(i)}-) \times \widehat{\lambda_k}(t_{(i)}) \text{ with}$$

$$\widehat{\lambda_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r(t_{(i)})} \qquad \text{cause specific hazard}$$

$$\widehat{\overline{F}}^{\mathrm{PL}}(t_{(i)}-) = \prod_{j:t_{(j)} < t_{(i)}} \left(1 - \frac{d(t_{(j)})}{r(t_{(j)})}\right) \text{ Kaplan-Meier}$$

Aalen-Johansen estimator of F_k

- Plug-in estimator based on $F_k(t) = \int_0^t \mathrm{P}\{T \geq s\}\lambda_k(s)ds$:

$$\widehat{F_k}^{\mathrm{AJ}}(t) = \sum_{i:t_{(i)} \leq t} \widehat{\overline{F}}^{\mathrm{PL}}(t_{(i)}-) \times \widehat{\lambda_k}(t_{(i)}) \text{ with}$$

$$\widehat{\lambda_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r(t_{(i)})} \qquad \text{cause specific hazard}$$

$$\widehat{\overline{F}}^{\mathrm{PL}}(t_{(i)}-) = \prod_{j:t_{(j)} < t_{(i)}} \left(1 - \frac{d(t_{(j)})}{r(t_{(j)})}\right) \text{ Kaplan-Meier}$$

- With single event type equal to Kaplan-Meier

Outline

Research Questions
- Examples

Rates and Risks
- Two approaches to competing risks analysis

Estimation
- Multi-state approach
- Subdistribution approach
- Regression

Summary
- Marginal versus competing risks
- Which appoach to choose?

II: The subdistribution approach

- Subdistribution hazard
  $(T_k = T \times I\{E = k\} + \infty \times I\{E \neq k\})$:

$$h_k(s) = \lim_{\Delta s \downarrow 0} \frac{1}{\Delta s} \times \mathrm{P}\{s \leq T_k < s + \Delta s \,|\, T_k \geq s\}$$

$$= \lim_{\Delta s \downarrow 0} \frac{\frac{1}{\Delta s}\mathrm{P}\{s \leq T < s + \Delta s, E = k\}}{\mathrm{P}\{T \geq s \text{ or } (T < s, E \neq k)\}}$$

- Denominator: event free or with earlier competing event
- Interpretation controversial
  - Not a rate in epidemiological sense,
  - unless we can assume that those with the competing event were immune for the event of interest (cure model)
- One-to-one relation with crude risk

$$\overline{F_k}(t) = \prod_{t_l \leq t} \left\{1 - h_k(t_l)\right\} \quad \text{or} \quad \overline{F_k}(t) = \exp\left\{-\int_0^t h_k(u)du\right\}$$

oucru

Estimation with complete follow-up (artificial data)

week	0-1	1-2	2-3	3-4	4-5	5-6	6-7	> 7
infection	1	2	6	11	9	11	2	18
cumulative	1	3	9	20	29	40	42	60
discharge	5	9	6	6	9	12	4	35
cumulative	5	14	20	26	35	47	51	86

- Discharge competing risk: Crude risk estimated as frequency of events:

  $\widehat{P}$(infection $\leq$6weeks)=40/146

  $\widehat{P}$(discharge $\leq$6weeks)=47/146

  Individuals with competing event remain in denominator, competing event ignored in estimation

Observed data

$\{(x_1, e_1\delta_1), \ldots, (x_N, e_N\delta_N)\}$

- $x_i = \min\{t_i, c_i\}$, $\delta_i = \{t_i \leq c_i\}$, $e_i \in \{1, \ldots, K\}$
- $t_{(i)}$ ordered unique event times of any type
- $r(t_{(i)})$ number observed at risk
- $r^*(t_{(i)})$ number in risk set (for subdistribution hazard)
- $d_k(t_{(i)})$ number of events at $t_{(i)}$ of type k
- $d(t_{(i)})$ total number of events at $t_{(i)}$

Subdistribution $\widehat{F}_k$: product-limit estimator

$$\widehat{\overline{F}}_k^{\mathrm{PL}}(t) = \prod_{i:t_{(j)} \leq t} \left\{ 1 - \widehat{h}_k(t_{(j)}) \right\} \text{ with } \widehat{h}_k(t_{(j)}) = \frac{d_k(t_{(j)})}{r^*(t_{(j)})}$$

No censoring: individuals with competing event remain in risk set forever. Small change in data

Subdistribution $\widehat{F_k}$: product-limit estimator

$$\widehat{\overline{F_k}}^{\mathrm{PL}}(t) = \prod_{i:t_{(j)} \leq t} \left\{ 1 - \widehat{h_k}(t_{(j)}) \right\} \text{ with } \widehat{h_k}(t_{(j)}) = \frac{d_k(t_{(j)})}{r^*(t_{(j)})}$$

No censoring: individuals with competing event remain in risk set forever. Small change in data

Administrative censoring: individuals with competing event leave risk set at date of administrative censoring.

Subdistribution $\widehat{F_k}$: product-limit estimator

$$\widehat{\overline{F_k}}^{\mathrm{PL}}(t) = \prod_{i:t_{(j)} \le t} \left\{ 1 - \widehat{h_k}(t_{(j)}) \right\} \text{ with } \widehat{h_k}(t_{(j)}) = \frac{d_k(t_{(j)})}{r^*(t_{(j)})}$$

No censoring: individuals with competing event remain in risk set forever. Small change in data

Administrative censoring: individuals with competing event leave risk set at date of administrative censoring.

General censoring: Estimate time-to-censoring distribution. Then for those with competing event:

- multiply impute censoring times
- reweight them by probability to remain uncensored

Right Censored Data

$$\widehat{h_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r^*(t_{(i)})}$$

Contribution $\omega_l(t_{(i)})$ of individual l to the risk set $r^*(t_{(i)})$ is:

- censored or event of type k before $t_{(i)}$: 0

Right Censored Data

$$\widehat{h_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r^*(t_{(i)})}$$

Contribution $\omega_l(t_{(i)})$ of individual l to the risk set $r^*(t_{(i)})$ is:

- censored or event of type k before $t_{(i)}$: 0
- still at risk at $t_{(i)}$: 1

Right Censored Data

$$\widehat{h_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r^*(t_{(i)})}$$

Contribution $\omega_l(t_{(i)})$ of individual l to the risk set $r^*(t_{(i)})$ is:

- censored or event of type k before $t_{(i)}$: 0
- still at risk at $t_{(i)}$: 1
- competing event at $t_{(j)}$ before $t_{(i)}$:

estimate of $\mathrm{P}\{C \geq t_{(i)} | C \geq t_{(j)}\}$:

Right Censored Data

$$\widehat{h_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r^*(t_{(i)})}$$

Contribution $\omega_l(t_{(i)})$ of individual l to the risk set $r^*(t_{(i)})$ is:

- censored or event of type k before $t_{(i)}$: 0
- still at risk at $t_{(i)}$: 1
- competing event at $t_{(j)}$ before $t_{(i)}$:

$$\text{estimate of } \mathrm{P}\{C \geq t_{(i)} | C \geq t_{(j)}\} : \widehat{\bar{\Gamma}}(t_{(i)}-)/\widehat{\bar{\Gamma}}(t_{(j)}-)$$

Right Censored Data

$$\widehat{h_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r^*(t_{(i)})}$$

Contribution $\omega_l(t_{(i)})$ of individual l to the risk set $r^*(t_{(i)})$ is:

- censored or event of type k before $t_{(i)}$: 0
- still at risk at $t_{(i)}$: 1
- competing event at $t_{(j)}$ before $t_{(i)}$:

$$\text{estimate of } \mathrm{P}\{C \geq t_{(i)} | C \geq t_{(j)}\} : \widehat{\bar{\Gamma}}(t_{(i)}-)/\widehat{\bar{\Gamma}}(t_{(j)}-)$$

- $\widehat{\bar{\Gamma}}$: reverse role of event time T_i and censoring C_i:

$$\widehat{\bar{\Gamma}}(t) = \prod_{j:c_{(j)} \leq t} \left\{ 1 - \frac{m_j}{r(c_{(j)})} \right\}$$

m_j: number of censorings at $c_{(j)}$

Equivalence

If weights in r^* based on the PL-form of $\widehat{\widehat{\Gamma}}$, then we have

$$\widehat{F_k}^{\mathrm{AJ}} \equiv \widehat{F_k}^{\mathrm{PL}}$$

(Geskus 2011, Biometrics)

Outline

Research Questions
- Examples

Rates and Risks
- Two approaches to competing risks analysis

Estimation
- Multi-state approach
- Subdistribution approach
- Regression

Summary
- Marginal versus competing risks
- Which appoach to choose?

Regression

- Cause-specific hazards: standard Cox model
  - Does account for competing risks
  - Interpretation is different: cause-specific event rate among event-free individuals
  - Not a marginal hazard, unless competing risks independent

Regression

- Cause-specific hazards: standard Cox model
  - Does account for competing risks
  - Interpretation is different: cause-specific event rate among event-free individuals
  - Not a marginal hazard, unless competing risks independent
- Proportional subdistribution hazards model (Fine and Gray)
  - Interpretation: direct relation with cause-specific cumulative incidence
  - Estimation: those with competing event remain in risk set (with a decreasing censoring weight)

Regression

- Cause-specific hazards: standard Cox model
  - Does account for competing risks
  - Interpretation is different: cause-specific event rate among event-free individuals
  - Not a marginal hazard, unless competing risks independent
- Proportional subdistribution hazards model (Fine and Gray)
  - Interpretation: direct relation with cause-specific cumulative incidence
  - Estimation: those with competing event remain in risk set (with a decreasing censoring weight)
- Example
  - AIDS-specific mortality reduced by cART
  - Other COD's: more frequent, even if cART has no side effects. No change in cause-specific hazard, but subdistribution hazard increases ("in the end we all die")
  - Subdistribution hazard includes impact on other event types

Outline

Research Questions
- Examples

Rates and Risks
- Two approaches to competing risks analysis

Estimation
- Multi-state approach
- Subdistribution approach
- Regression

Summary
- Marginal versus competing risks
- Which appoach to choose?

Estimators in competing risks setting

hazard	estimate	cumulative
marginal	$d_k(t)/r(t)\{=\widehat{\lambda_k}(t)\}$	$\prod_{t_{(j)}\le t}\left\{1-\widehat{\lambda_k}(t_{(j)})\right\}$
cause-specific	$\widehat{\lambda_k}(t)=d_k(t)/r(t)$	$\widehat{F_k}^{\mathrm{AJ}}(t)=\sum_{t_{(j)}\le t}\widehat{\overline{F}}^{\mathrm{PL}}(t_{(j)}-)\widehat{\lambda_k}(t_{(j)})$
subdistribution	$\widehat{h_k}(t)=d_k(t)/r^*(t)$	$\widehat{\overline{F_k}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h_k}(t_{(j)})\right\}$
overall	$\widehat{h}(t)=d(t)/r(t)$	$\widehat{\overline{F}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h}(t_{(j)})\right\}$

Estimators in competing risks setting

hazard	estimate	cumulative
marginal	$d_k(t)/r(t)\{=\widehat{\lambda_k}(t)\}$	$\prod_{t_{(j)}\le t}\left\{1-\widehat{\lambda_k}(t_{(j)})\right\}$
cause-specific	$\widehat{\lambda_k}(t)=d_k(t)/r(t)$	$\widehat{F_k}^{\mathrm{AJ}}(t)=\sum_{t_{(j)}\le t}\widehat{\overline{F}}^{\mathrm{PL}}(t_{(j)}-)\widehat{\lambda_k}(t_{(j)})$
subdistribution	$\widehat{h_k}(t)=d_k(t)/r^*(t)$	$\widehat{\overline{F_k}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h_k}(t_{(j)})\right\}$
overall	$\widehat{h}(t)=d(t)/r(t)$	$\widehat{\overline{F}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h}(t_{(j)})\right\}$

Marginal distribution

- Estimated via (marginal) hazard, basis for Kaplan-Meier estimate of cumulative incidence/net risk
- Assumption: Censored individuals can be represented by the ones that remain at risk. Reason for censoring should give no information on residual time-to-event

Marginal distribution

- Estimated via (marginal) hazard, basis for Kaplan-Meier estimate of cumulative incidence/net risk
- Assumption: Censored individuals can be represented by the ones that remain at risk. Reason for censoring should give no information on residual time-to-event
- Otherwise Kaplan-Meier has no meaning.
  Does not describe survival in (hypothetical) world with competing event removed, unless we know that censoring is independent

Marginal distribution

- Estimated via (marginal) hazard, basis for Kaplan-Meier estimate of cumulative incidence/net risk
- Assumption: Censored individuals can be represented by the ones that remain at risk. Reason for censoring should give no information on residual time-to-event
- Otherwise Kaplan-Meier has no meaning.
  Does not describe survival in (hypothetical) world with competing event removed, unless we know that censoring is independent
- Extra information may allow to show informative/dependent censoring (IDU and pre-AIDS death), but independence can never be tested for

Competing risks

- Competing risk is a separate event
  - Individuals censored by competing event don't have to be represented by the ones that remain at risk.
    Other censoring (administrative/loss to follow-up) must be independent

Competing risks

- Competing risk is a separate event
  - Individuals censored by competing event don't have to be represented by the ones that remain at risk.
    Other censoring (administrative/loss to follow-up) must be independent
- Cause-specific hazard
  - Basis for Aalen-Johansen estimator of cause-specific cumulative incidence/crude risk

Estimators in competing risks setting

hazard	estimate	cumulative
marginal	$d_k(t)/r(t)\{=\widehat{\lambda_k}(t)\}$	$\prod_{t_{(j)}\le t}\left\{1-\widehat{\lambda_k}(t_{(j)})\right\}$
cause-specific	$\widehat{\lambda_k}(t)=d_k(t)/r(t)$	$\widehat{F_k}^{\mathrm{AJ}}(t)=\sum_{t_{(j)}\le t}\widehat{\overline{F}}^{\mathrm{PL}}(t_{(j)}-)\widehat{\lambda_k}(t_{(j)})$
subdistribution	$\widehat{h_k}(t)=d_k(t)/r^*(t)$	$\widehat{\overline{F_k}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h_k}(t_{(j)})\right\}$
overall	$\widehat{h}(t)=d(t)/r(t)$	$\widehat{\overline{F}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h}(t_{(j)})\right\}$

Competing risks

- Competing risk is a separate event
  - Individuals censored by competing event don't have to be represented by the ones that remain at risk.
    Other censoring (administrative/loss to follow-up) must be independent
- Cause-specific hazard
  - Basis for Aalen-Johansen estimator of cause-specific cumulative incidence/crude risk
  - If censoring due to competing event is independent, then marginal and cause-specific hazard are equal. Cumulative quantities different: Kaplan-Meier versus Aalen-Johansen

Estimators in competing risks setting

hazard	estimate	cumulative
marginal	$d_k(t)/r(t)\{=\widehat{\lambda_k}(t)\}$	$\prod_{t_{(j)}\le t}\left\{1-\widehat{\lambda_k}(t_{(j)})\right\}$
cause-specific	$\widehat{\lambda_k}(t)=d_k(t)/r(t)$	$\widehat{F_k}^{\mathrm{AJ}}(t)=\sum_{t_{(j)}\le t}\widehat{\overline{F}}^{\mathrm{PL}}(t_{(j)}-)\widehat{\lambda_k}(t_{(j)})$
subdistribution	$\widehat{h_k}(t)=d_k(t)/r^*(t)$	$\widehat{\overline{F}_k}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h_k}(t_{(j)})\right\}$
overall	$\widehat{h}(t)=d(t)/r(t)$	$\widehat{\overline{F}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h}(t_{(j)})\right\}$

Estimators in competing risks setting

hazard	estimate	cumulative
marginal	$d_k(t)/r(t)\{=\widehat{\lambda_k}(t)\}$	$\prod_{t_{(j)}\le t}\left\{1-\widehat{\lambda_k}(t_{(j)})\right\}$
cause-specific	$\widehat{\lambda_k}(t)=d_k(t)/r(t)$	$\widehat{F_k}^{\mathrm{AJ}}(t)=\sum_{t_{(j)}\le t}\widehat{\overline{F}}^{\mathrm{PL}}(t_{(j)}-)\widehat{\lambda_k}(t_{(j)})$
subdistribution	$\widehat{h_k}(t)=d_k(t)/r^*(t)$	$\widehat{\overline{F_k}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h_k}(t_{(j)})\right\}$
overall	$\widehat{h}(t)=d(t)/r(t)$	$\widehat{\overline{F}}^{\mathrm{PL}}(t)=\prod_{t_{(j)}\le t}\left\{1-\widehat{h}(t_{(j)})\right\}$

Competing risks

- Competing risk is a separate event
  - Individuals censored by competing event don't have to be represented by the ones that remain at risk.
    Other censoring (administrative/loss to follow-up) must be independent
- Cause-specific hazard
  - Basis for Aalen-Johansen estimator of cause-specific cumulative incidence/crude risk
  - If censoring due to competing event is independent, then marginal and cause-specific hazard are equal. Cumulative quantities different: Kaplan-Meier versus Aalen-Johansen
- Subdistribution hazard: one-to-one relation with crude risk

Outline

Research Questions
- Examples

Rates and Risks
- Two approaches to competing risks analysis

Estimation
- Multi-state approach
- Subdistribution approach
- Regression

Summary
- Marginal versus competing risks
- Which appoach to choose?

Marginal or competing risks?

- Example I: staphylococcus infection in hospital
  - Marginal: what if everyone would stay in hospital
  - Competing risks: how many infections are observed in hospital
- Example II: difference in natural history between IDU en MSM
  Marginal analysis
- Example III: spectrum in COD
  Competing risks; marginal analysis completely hypothetical

Etiology or prediction?

- Which hazard quantifies etiology?

Etiology or prediction?

- Which hazard quantifies etiology? Competing risk is:
  - intervention: marginal (hospital discharge; IDU/MSM)
  - biological: cause-specific (spectrum in COD's)

Etiology or prediction?

- Which hazard quantifies etiology? Competing risk is:
  - intervention: marginal (hospital discharge; IDU/MSM)
  - biological: cause-specific (spectrum in COD's)
- When can we interpret results as effects on marginal hazards?

Etiology or prediction?

- Which hazard quantifies etiology? Competing risk is:
  - intervention: marginal (hospital discharge; IDU/MSM)
  - biological: cause-specific (spectrum in COD's)
- When can we interpret results as effects on marginal hazards?
  - event types independent: cause-specific hazard
  - high positive correlation: overall hazard
  - cure: subdistribution

Etiology or prediction?

- Which hazard quantifies etiology? Competing risk is:
  - intervention: marginal (hospital discharge; IDU/MSM)
  - biological: cause-specific (spectrum in COD's)
- When can we interpret results as effects on marginal hazards?
  - event types independent: cause-specific hazard
  - high positive correlation: overall hazard
  - cure: subdistribution
- Prediction: subdistribution (based on cause-specific or subdistribution hazard, but only latter has one-to-one relation with cumulative probability)

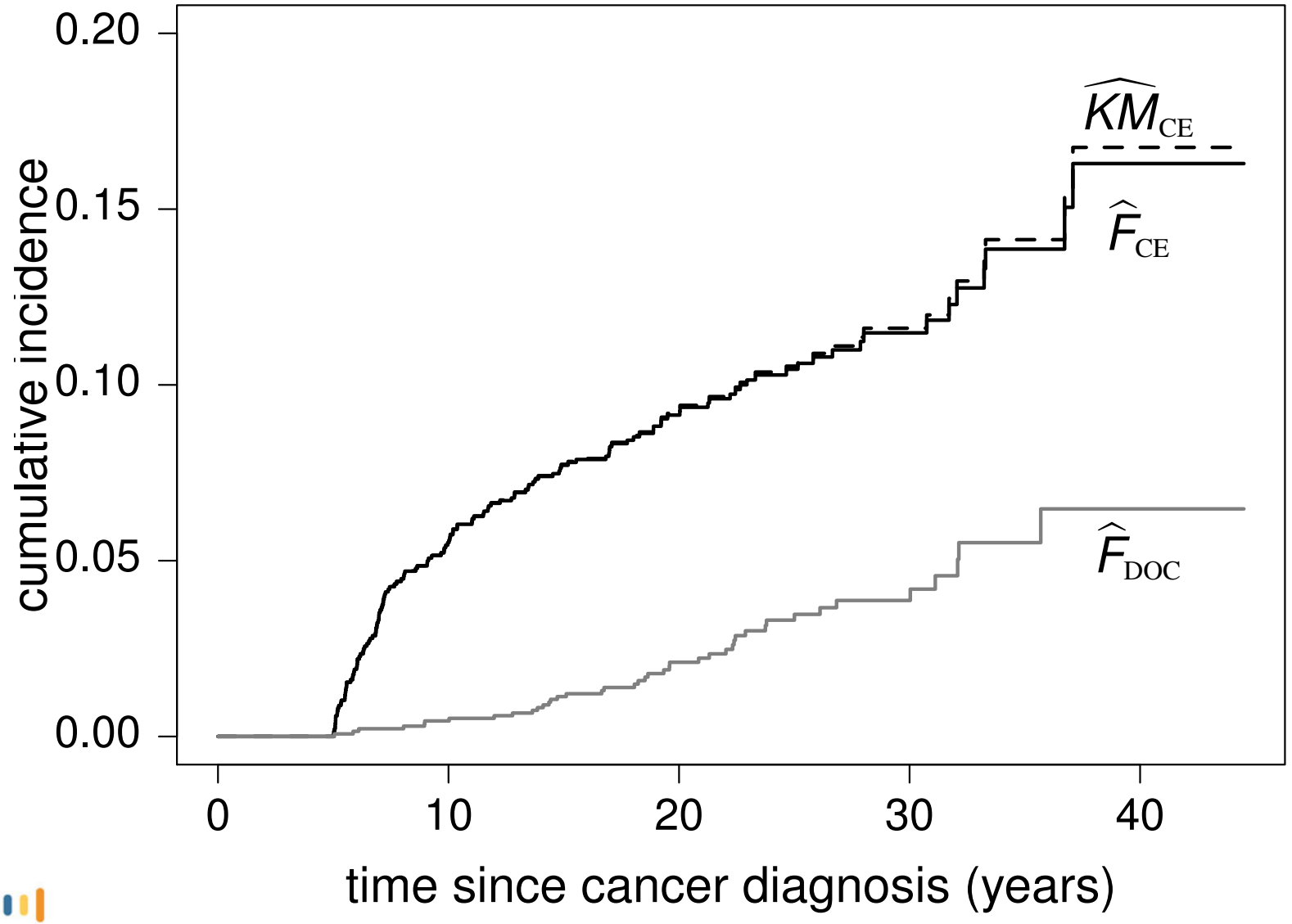
Etiology or prediction?

- Which hazard quantifies etiology? Competing risk is:
  - intervention: marginal (hospital discharge; IDU/MSM)
  - biological: cause-specific (spectrum in COD's)
- When can we interpret results as effects on marginal hazards?
  - event types independent: cause-specific hazard
  - high positive correlation: overall hazard
  - cure: subdistribution
- Prediction: subdistribution (based on cause-specific or subdistribution hazard, but only latter has one-to-one relation with cumulative probability)
- Both Cox and Fine and Gray model make sense in presence of competing risks

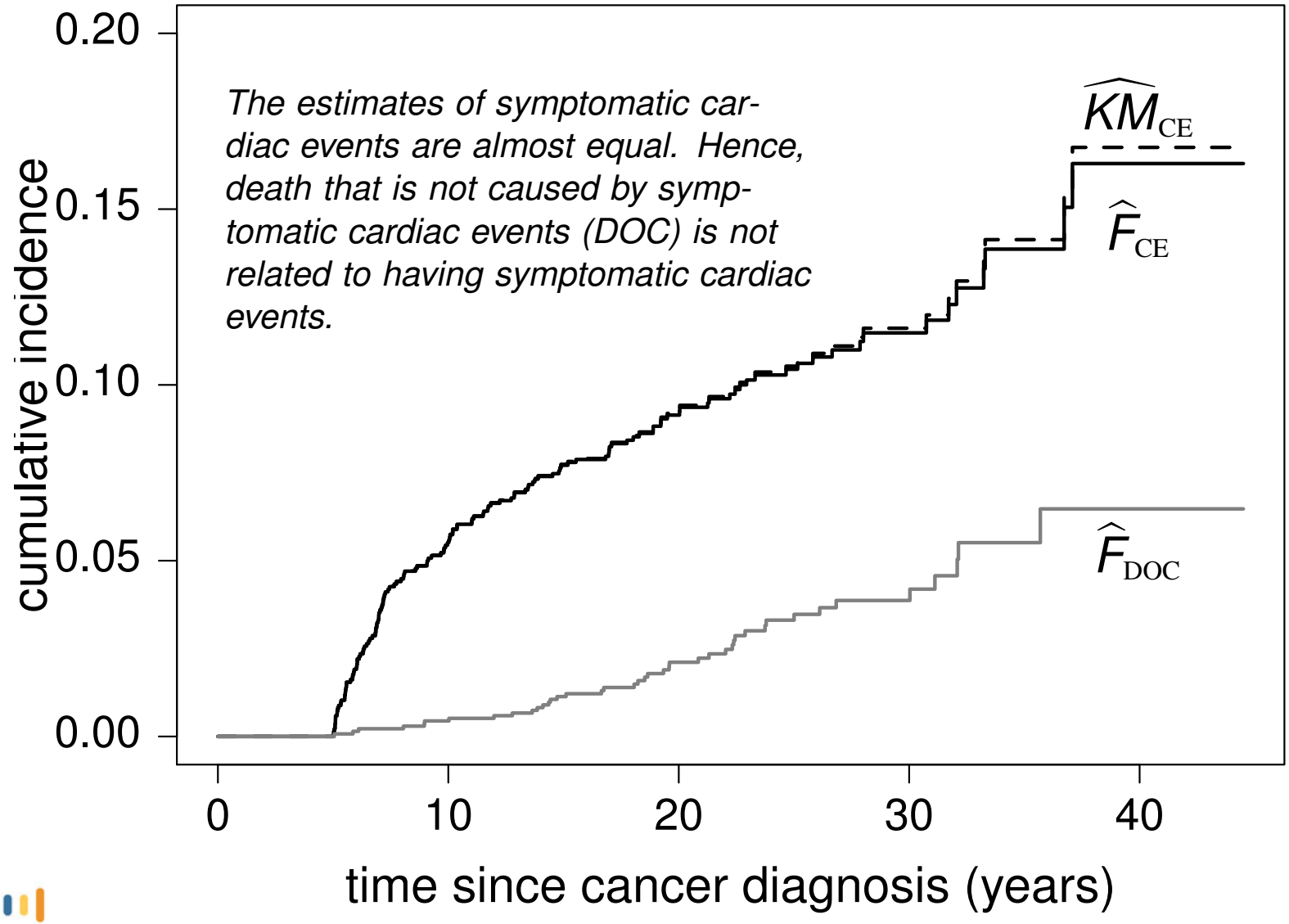
Etiology or prediction?

- Which hazard quantifies etiology? Competing risk is:
  - intervention: marginal (hospital discharge; IDU/MSM)
  - biological: cause-specific (spectrum in COD's)
- When can we interpret results as effects on marginal hazards?
  - event types independent: cause-specific hazard
  - high positive correlation: overall hazard
  - cure: subdistribution
- Prediction: subdistribution (based on cause-specific or subdistribution hazard, but only latter has one-to-one relation with cumulative probability)
- Both Cox and Fine and Gray model make sense in presence of competing risks
- Can we use Fine and Gray with time dependent variables?

Childhood cancer survivors; cardiac event, DOC competing



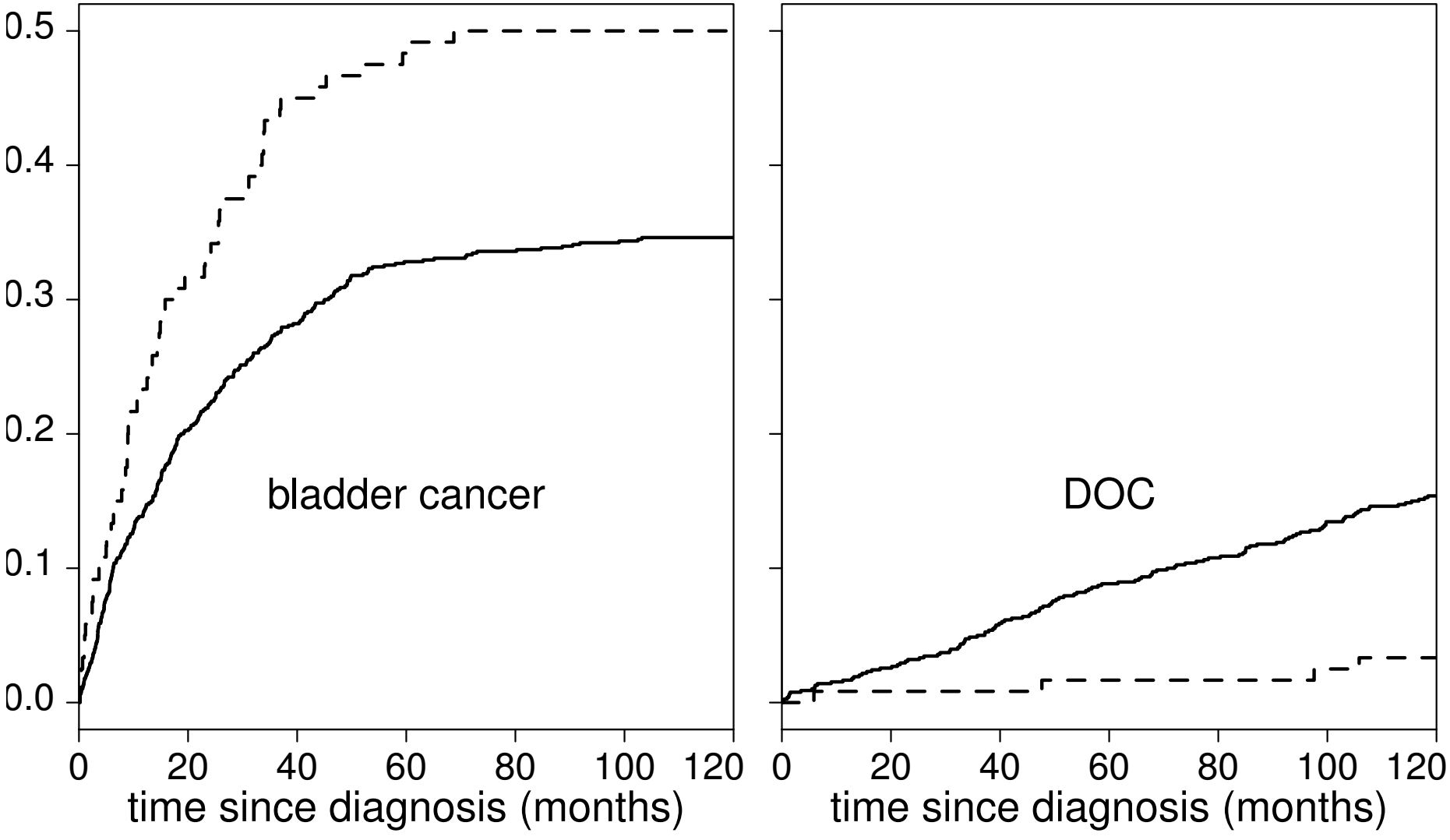
Childhood cancer survivors; cardiac event, DOC competing



Answer

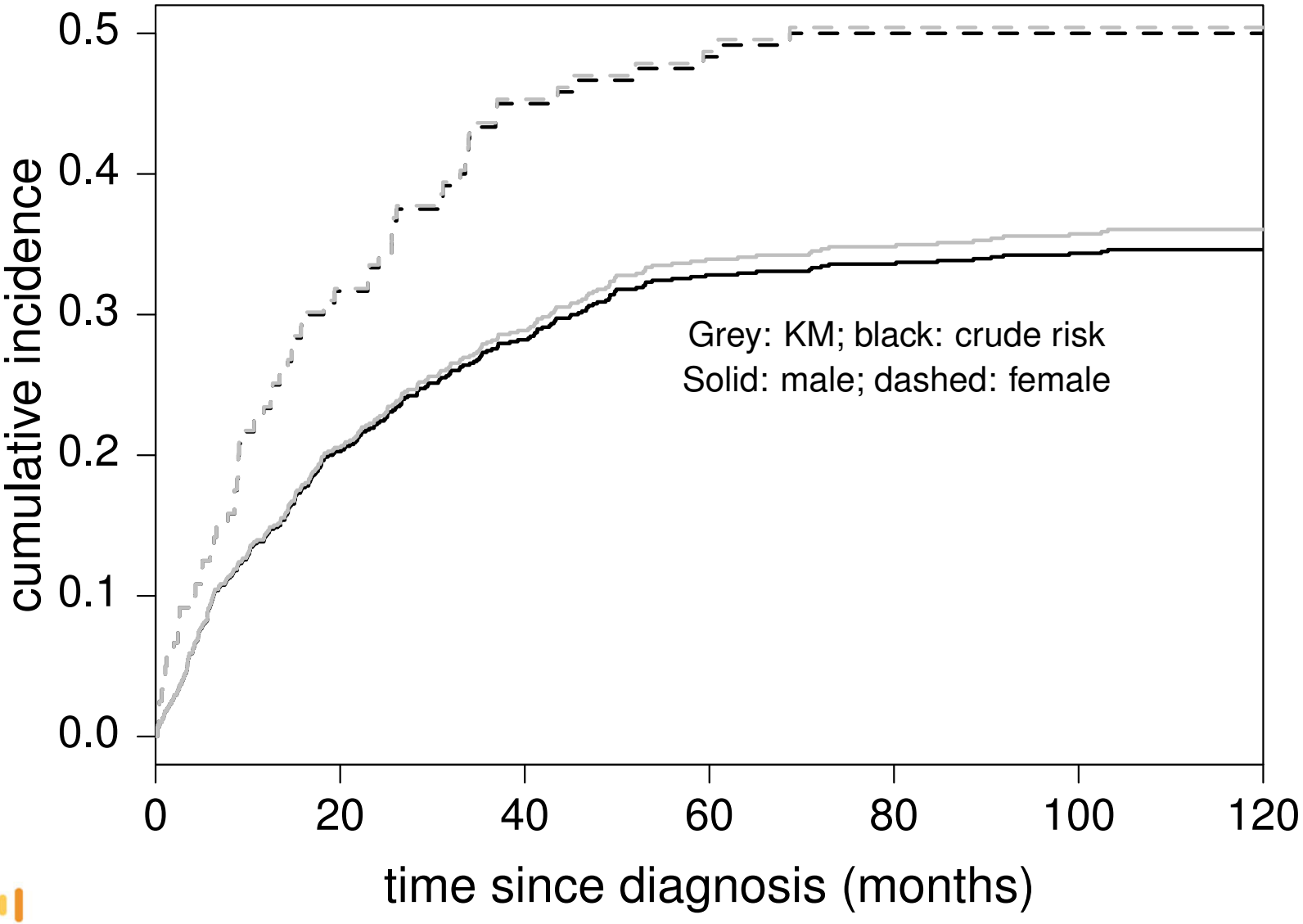
- The Kaplan-Meier and the estimator of the CE-specific cumulative incidence try to estimate different quantities
- Both curves are similar because there is little mortality due to other causes, at least during the first 20 years, when most of the CE's occur.
- Note that on one hand it is said that death due to other causes may not be related to CE events, whereas on the other hand it is called "informative censoring".

Bladder cancer; relapse, DOC competing

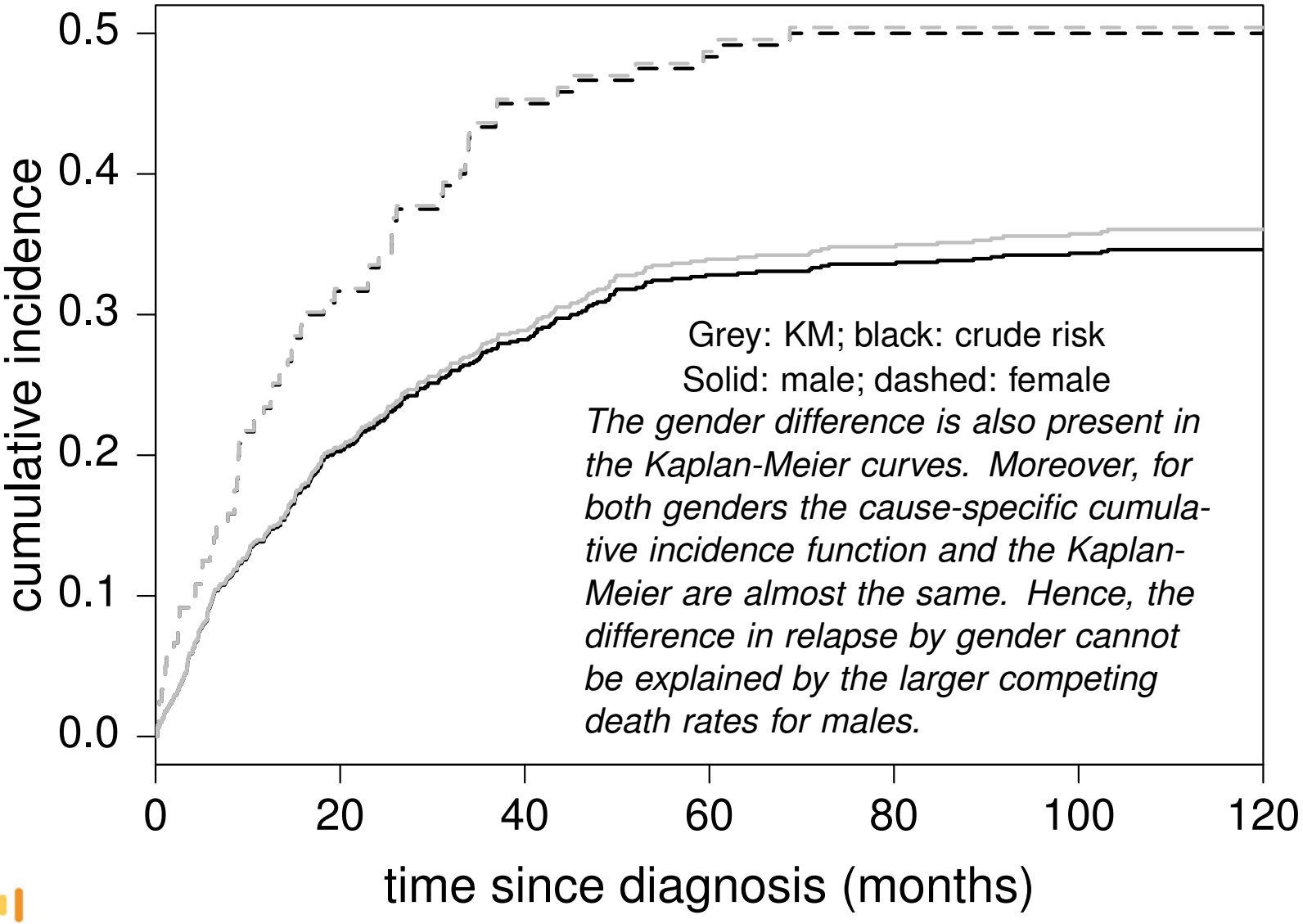


Left: relapse. Solid: male; dashed: female

Bladder cancer; relapse, DOC competing



Bladder cancer; relapse, DOC competing



Answer

- The Kaplan-Meier tries to compare the marginal distribution of time to relapse for males and females. Only valid if DOC is noninformative for relapse.
- Estimates almost equal because there is little mortality due to other causes, at least during the first 40 months.

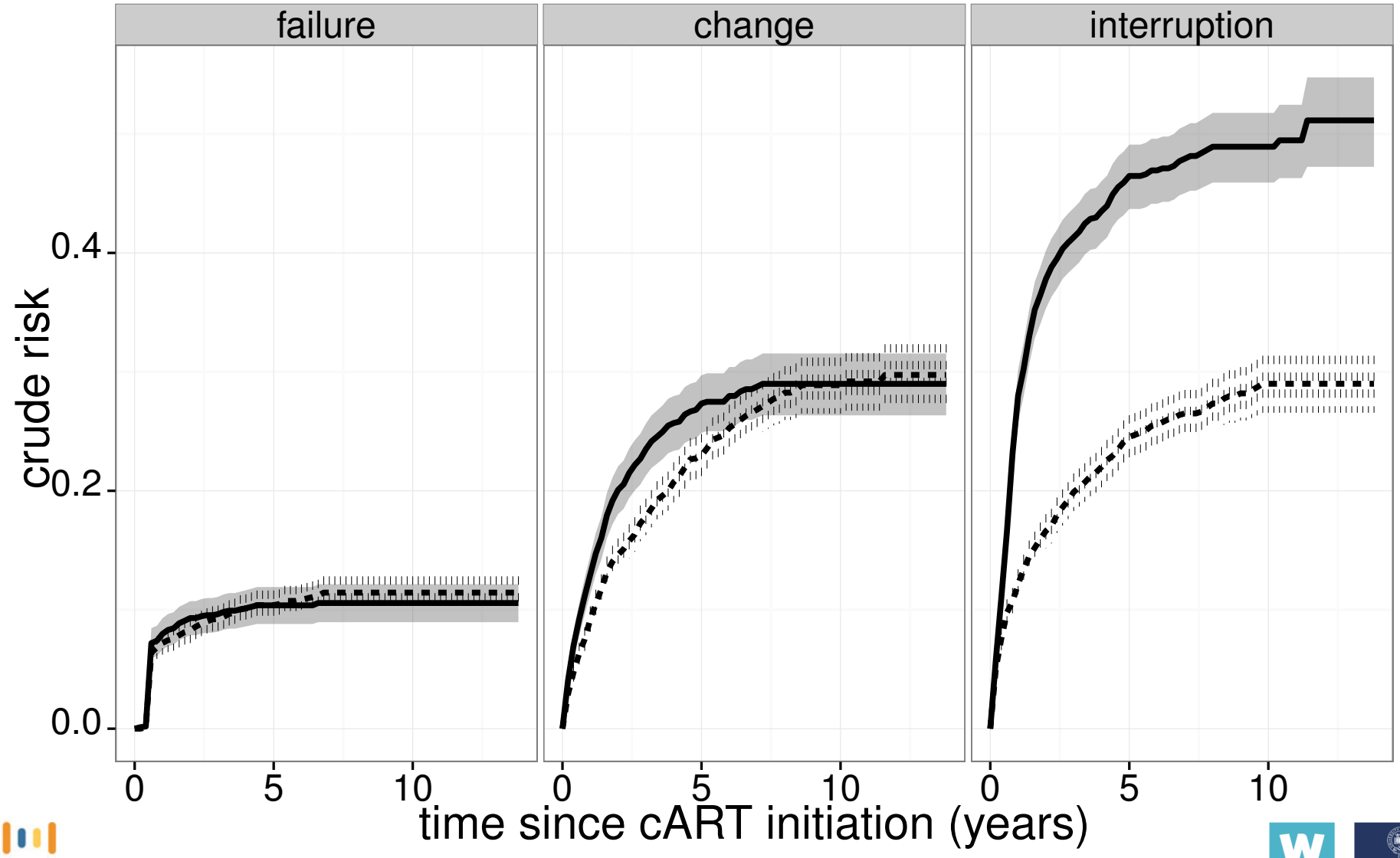
Answer

- The Kaplan-Meier tries to compare the marginal distribution of time to relapse for males and females. Only valid if DOC is noninformative for relapse.
- Estimates almost equal because there is little mortality due to other causes, at least during the first 40 months.
- If we combine both event times, the curves for males and females will become similar. Would estimate marginal hazard if every person that died would have progressed on the next day.

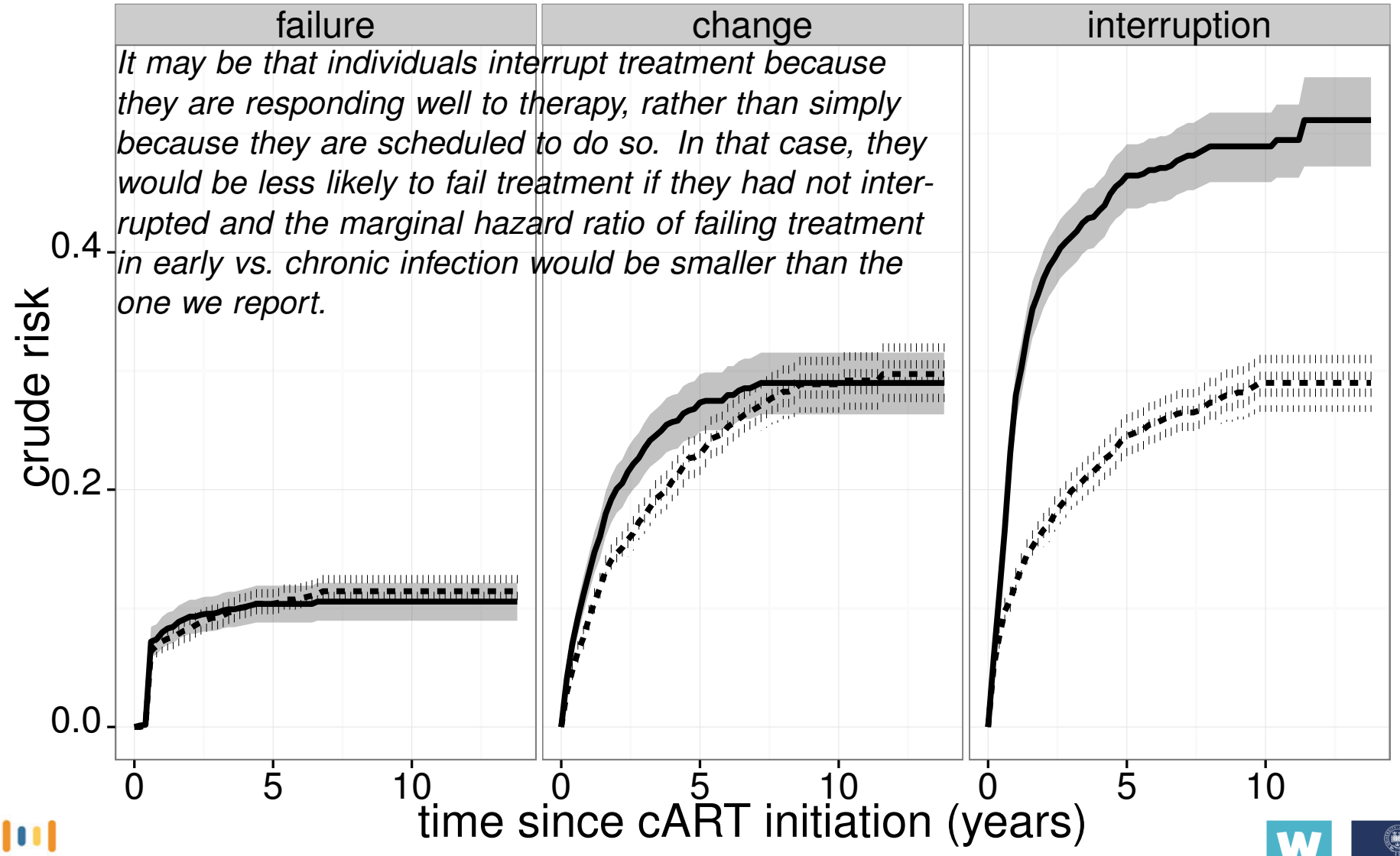
Answer

- The Kaplan-Meier tries to compare the marginal distribution of time to relapse for males and females. Only valid if DOC is noninformative for relapse.
- Estimates almost equal because there is little mortality due to other causes, at least during the first 40 months.
- If we combine both event times, the curves for males and females will become similar. Would estimate marginal hazard if every person that died would have progressed on the next day.
- All we can conclude is that females have a higher relapse-specific hazard than males. And females have a lower DOC-specific hazard than males.

cART; early versus late starters



cART; early versus late starters



Answer

- Assume that treatment failure and treatment interruption are the only two competing events
- Most extreme scenario: those who interrupt treatment will never fail. Marginal hazard same as the subdistribution hazard, i.e. the reported one.
- It may be true if the effect was observed for the cause-specific hazard.

THANKS!

References I


J. P. Fine and R. J. Gray (1999).
A proportional hazards model for the subdistribution of a competing risk.
J. Am. Statist. Assoc. **94,** 496–509.


R. B. Geskus (2011).
Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring.
Biometrics **67**, 39–49


Ronald B. Geskus (2015).
Data analysis with competing risks and intermediate states.
CRC Press, Boca Raton.


H. Putter, M. Fiocco, and R. B. Geskus (2007).
Tutorial in biostatistics: competing risks and multi-state models.
Stat Med, 26(11):2389–2430.


J. van der Helm, R. B. Geskus, C. Sabin *et al.* (2013).
Effect of HCV infection on cause-specific mortality after HIV seroconversion, before and after 1997.
Gastroenterology, 144:751–760, 2013.

Part II

Software for competing risks analyses

Right Censored Data

Representation Two columns: `time` since origin and `status` variable. `status=1` if event observed and `status=0` if event right censored

```
id time status
 1  7.7      1
 2  4.3      1
 3  5.6      0
 ...
```

In R via `Surv(time=time, event=status)`

Right Censored Data

Representation Two columns: `time` since origin and `status` variable. `status=1` if event observed and `status=0` if event right censored

```
id time status
 1  7.7      1
 2  4.3      1
 3  5.6      0
 ...
```

In R via `Surv(time=time, event=status)`

Kaplan-Meier Main function: `survfit.formula`

```
survfit(Surv(time,status)~1, data=...)
```

Four example individuals

`aidssi` data set (available in `mstate` package and in Stata)

patnr	time	status	cause	ccr5
14	5.054	0	event-free	WW
3	2.234	1	AIDS	WW
15	10.196	1	AIDS	WM
8	8.605	2	SI	WW

Four example individuals

`aidssi` data set (available in `mstate` package and in Stata)

patnr	time	status	cause	ccr5
14	5.054	0	event-free	WW
3	2.234	1	AIDS	WW
15	10.196	1	AIDS	WM
8	8.605	2	SI	WW

Kaplan-Meier both event types combined:

```
KM.curve <- survfit(Surv(time,status!=0)~1,
                                    data=aidssi)
```

Note: `event` argument can be a logical expression (`status!=0`)

Four example individuals

`aidssi` data set (available in `mstate` package and in Stata)

patnr	time	status	cause	ccr5
14	5.054	0	event-free	WW
3	2.234	1	AIDS	WW
15	10.196	1	AIDS	WM
8	8.605	2	SI	WW

Kaplan-Meier both event types combined:

```
KM.curve <- survfit(Surv(time,status!=0)~1,
                                    data=aidssi)
```

Note: `event` argument can be a logical expression (`status!=0`)

Estimate per value of CCR5:

```
survfit(Surv(time,status!=0)~ccr5, data=aidssi)
```

Numerical summary

```
summary(KM.curve)
```

gives:

```
   time n.risk n.event survival std.err lower CI upper CI
  0.112    329       1    0.997 0.00303    0.991    1.000
  0.137    328       1    0.994 0.00429    0.986    1.000
  0.474    325       1    0.991 0.00525    0.981    1.000
  0.824    321       1    0.988 0.00607    0.976    1.000
      .      .       .        .       .        .        .
 12.936     41       1    0.217 0.02604    0.171    0.274
 13.361     22       1    0.207 0.02665    0.161    0.266
 13.936      1       1    0.000     NaN       NA       NA
```

Survival at 12 years obtained via

```
summary(KM.curve, time=12)
```

Plotting Kaplan-Meier survival curves

`plot.survfit` for first plot;
`lines.survfit` adds curves

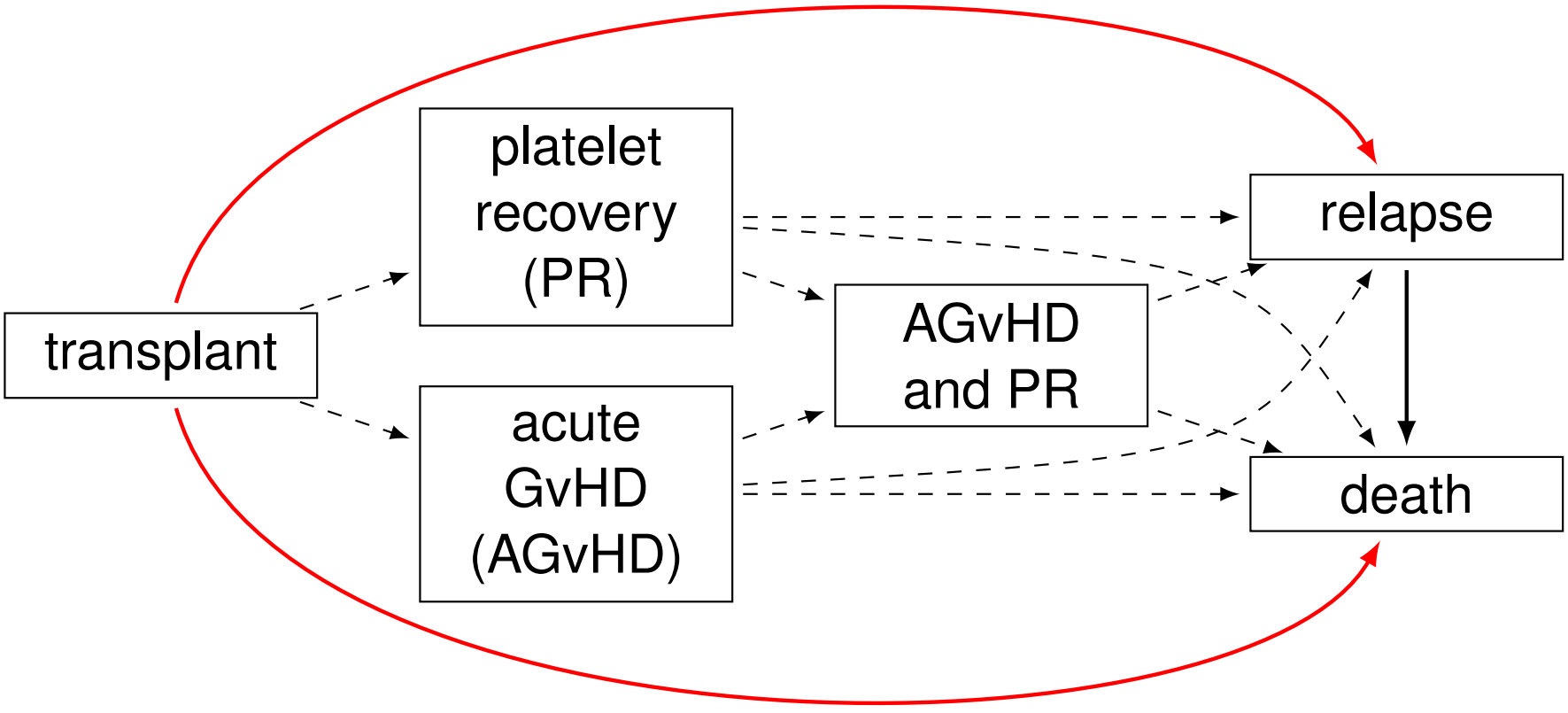
```
plot.survfit
 function(x, conf.int,
    mark.time = FALSE, pch = 3,
    col = 1, lty = 1, lwd = 1, cex = 1,
    log = FALSE, fun,
    xscale = 1, yscale = 1, firstx = 0, firsty = 1,
    xmax, ymin = 0, xlab = "", ylab = "", xaxs = "S",
    conf.times, conf.cap = 0.005, conf.offset = 0.012,
    ...)
```

`fun="event"` plots cumulative incidence (i.e. upwards from 0):

```
plot(KM.curve, mark.time=FALSE, fun="event")
```

`fun="cumhaz"` plots cumulative hazard

Computer practical



1. A first look at the data Have a look at the data explanation:
   `help(ebmt1)`
   Add columns **time** and **stat**.
2. Estimation of overall cumulative incidence Compute the Kaplan-Meier estimator for relapse-free survival. Plot the estimate on the scale of the cumulative incidence.

Software to compute Aalen-Johansen estimator

- Stata: `stcompet` command
- SAS: macros `%CUMINCID` or `%CIF`
- R, some options:
  - standard `survival` package
  - `cmprsk` or `prodlim` package
  - any package for multi-state models, e.g. `etm`, `mstate`, `msSurv`

Software to compute Aalen-Johansen estimator

- Stata: `stcompet` command
- SAS: macros `%CUMINCID` or `%CIF`
- R, some options:
  - standard `survival` package
  - `cmprsk` or `prodlim` package
  - any package for multi-state models, e.g. `etm`, `mstate`, `msSurv`

Four example individuals

`aidssi` data set (available in `mstate` package and in Stata)

patnr	time	status	cause	ccr5
14	5.054	0	event-free	WW
3	2.234	1	AIDS	WW
15	10.196	1	AIDS	WM
8	8.605	2	SI	WW

R: survival package

If event column is numeric:

```
csiSurv <- survfit(Surv(time,status,type="mstate")~1,
                                                data=aidssi)
```

If event column is a factor variable:

```
survfit(Surv(time,relevel(cause,"event-free"))~1,
                                              data=aidssi)
```

Summary of estimate:

```
summary(csiSurv, times=seq(2,8,by=2))

 time n.risk n.event   P(1)    P(2)   P()
    2    300      15 0.00632 0.0404 0.953
    4    240      52 0.10322 0.1112 0.786
    6    170      54 0.17070 0.2259 0.603
    8    122      41 0.25430 0.2916 0.454
```

Confidence intervals: `csiSurv$lower` and `csiSurv$upper`

3. Estimation of cause-specific cumulative incidence Compute the Aalen-Johansen estimator for relapse and relapse-free mortality. What is the probability, with 95% confidence intervals (on the default log scale), to have a relapse within one year and within five years.

4. Some plots (a) Plot the estimated cause-specific cumulative incidence for each end point using the overlaid display format. Plot the 95% confidence intervals as well.
(b) Plot the estimated cause-specific cumulative incidence using the stacked format (without the confidence intervals). First plot the relapse-specific cumulative incidence, and plot the death-specific cumulative incidence on top of this curve.
(c) Plot the cause-specific cumulative incidence estimates for each end point using the alternate display format.

Outline

Single event type

Multi-state: Aalen-Johansen estimator

Subdistribution: product-limit estimator

- Representation of the weights
- Creation of the data set with weights
- The product-limit estimator

Creation of the weights; example data

Three individuals from larger data set, “1” event type of interest

id	event.time	event.type
1	0.63644	0
2	0.64358	1
3	0.25615	2

Right Censored Data

$$\widehat{h_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r^*(t_{(i)})}$$

Contribution $\omega_l(t_{(i)})$ of individual l to the risk set $r^*(t_{(i)})$ is:

- censored or event of type k before $t_{(i)}$: 0
- still at risk at $t_{(i)}$: 1
- competing event at $t_{(j)}$ before $t_{(i)}$:

$$\text{estimate of } \mathrm{P}\{C \geq t_{(i)} | C \geq t_{(j)}\} : \widehat{\bar{\Gamma}}(t_{(i)}-)/\widehat{\bar{\Gamma}}(t_{(j)}-)$$

- $\widehat{\bar{\Gamma}}$: reverse role of event time T_i and censoring C_i:

$$\widehat{\bar{\Gamma}}(t) = \prod_{j:c_{(j)} \leq t} \left\{ 1 - \frac{m_j}{r(c_{(j)})} \right\}$$

m_j: number of censorings at $c_{(j)}$

Creation of the weights; example data

Three individuals from larger data set, "1" event type of interest

id	event.time	event.type
1	0.63644	0
2	0.64358	1
3	0.25615	2

$\Longrightarrow$

id	Tstart	Tstop	status	weight.cens	count	failcode
1	0.00000	0.63644	0	1.00000	1	1
2	0.00000	0.64358	1	1.00000	1	1
3	0.00000	0.25615	2	1.00000	1	1
3	0.25615	0.31778	2	1.00000	2	1
3	0.31778	0.37693	2	1.00000	3	1
3	0.37693	0.38928	2	1.00000	4	1
3	0.38928	0.46029	2	1.00000	5	1
3	0.46029	0.50979	2	1.00000	6	1
3	0.50979	0.64358	2	0.67849	7	1
3	0.64358	0.64724	2	0.67849	8	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Events of type 1 observed at 0.31778 0.37693 0.38928 0.46029 0.50979 0.64358 0.64724

Outline

Single event type

Multi-state: Aalen-Johansen estimator

Subdistribution: product-limit estimator

- Representation of the weights
- Creation of the data set with weights
- The product-limit estimator

Software

Stata `stcrprep` command

SAS `%PSHREG` macro

R `finegray` function in `survival` package
`crprep` function in `mstate` package

R: create data set with weights

patnr	time	status	cause	ccr5
14	5.054	0	event-free	WW
3	2.234	1	AIDS	WW
15	10.196	1	AIDS	WM
8	8.605	2	SI	WW

```
finegray(Surv(time,status,type="mstate")~.,
                                    data=aidssi, etype=1)
finegray(Surv(time,relevel(cause,"event-free"))~.,
                                  data=aidssi, etype="AIDS")
```

```
crprep(Tstop="time", status="status", data=aidssi,
                               trans=1, cens=0, keep="ccr5")
aidssi.w <- crprep(Tstop="time", status="cause",
            data=aidssi, trans="AIDS", cens="event-free",
                                                keep="ccr5")
```

Resulting data set

patnr	time	status	cause	ccr5
14	5.054	0	event-free	WW
3	2.234	1	AIDS	WW
15	10.196	1	AIDS	WM
8	8.605	2	SI	WW

	id	Tstart	Tstop	status	weight.cens	ccr5	count	failcode
3	3	0.000	2.234	AIDS	1.000	WW	1	AIDS
17	8	0.000	8.605	SI	1.000	WW	1	AIDS
18	8	8.605	8.638	SI	0.991	WW	2	AIDS
19	8	8.638	8.755	SI	0.982	WW	3	AIDS
.	.	.	.	.	.	.		
78	14	0.000	5.054	event-free	1.000	WW	1	AIDS
79	15	0.000	10.196	AIDS	1.000	WM	1	AIDS

Outline

Single event type

Multi-state: Aalen-Johansen estimator

Subdistribution: product-limit estimator

- Representation of the weights
- Creation of the data set with weights
- The product-limit estimator

PL-form: Kaplan-Meier with probability weights

Stata Specify `pweights` option in `stset` command; use standard `sts` command

SAS `PROC LIFEREG`

R `weights` argument in `survfit` function (`survival` package)

Example in R

	id	Tstart	Tstop	status	weight.cens	ccr5	count	failcode
3	3	0.000	2.234	AIDS	1.000	WW	1	AIDS
17	8	0.000	8.605	SI	1.000	WW	1	AIDS
18	8	8.605	8.638	SI	0.991	WW	2	AIDS
19	8	8.638	8.755	SI	0.982	WW	3	AIDS
78	14	0.000	5.054	event-free	1.000	WW	1	AIDS
79	15	0.000	10.196	AIDS	1.000	WM	1	AIDS
			⋮					

```
survfit(Surv(Tstart,Tstop,status=="AIDS")~1,
                    data=aidssi.w, weights=weight.cens)
```

Example in R

	id	Tstart	Tstop	status	weight.cens	ccr5	count	failcode
3	3	0.000	2.234	AIDS	1.000	WW	1	AIDS
17	8	0.000	8.605	SI	1.000	WW	1	AIDS
18	8	8.605	8.638	SI	0.991	WW	2	AIDS
19	8	8.638	8.755	SI	0.982	WW	3	AIDS
78	14	0.000	5.054	event-free	1.000	WW	1	AIDS
79	15	0.000	10.196	AIDS	1.000	WM	1	AIDS
			⋮					

```
survfit(Surv(Tstart,Tstop,status=="AIDS")~1,
                    data=aidssi.w, weights=weight.cens)
```

- Both event types at once (via `trans=c("AIDS","SI")`): use

  `Surv(Tstart,Tstop,status==failcode)~failcode`

Summarize results

```
aidssi.w <- crprep(Tstop="time", status="cause",
            data=aidssi, trans=c("AIDS","SI"),
            cens="event-free", keep="ccr5")
csiPL <- survfit(Surv(Tstart,Tstop,status==failcode)~failcode,
                 data=aidssi.w, weights=weight.cens)
csiPL$strata
summary(csiPL["failcode=SI"],times=seq(2,10,by=2))
```

and obtain

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
2	302	13	0.959615	0.0110	0.938344	0.981369
4	272	22	0.888783	0.0177	0.854687	0.924239
6	218	34	0.774112	0.0240	0.728466	0.822619
8	190	18	0.708440	0.0265	0.658364	0.762324
10	161	11	0.665410	0.0279	0.612936	0.722377

or use `csiPL[2]`

5. Estimation of cause-specific cumulative incidence (PL-form)

Use the `crprep` function to create the data set with weights. Include the covariables **score** and **age** and also store the **type** column in the new data set. Compute the weights for both end points.

Compare the estimates and confidence intervals at one and five years with the estimates based on the Aalen-Johansen form.

Part III

Time-varying covariables

Outline

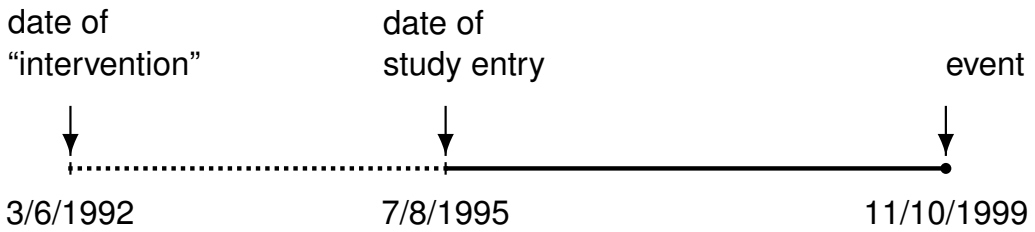
Standard setting
- Left truncated data
- Time-varying covariables

Competing risks
- Aalen-Johansen estimator
- Competing risks: subdistribution

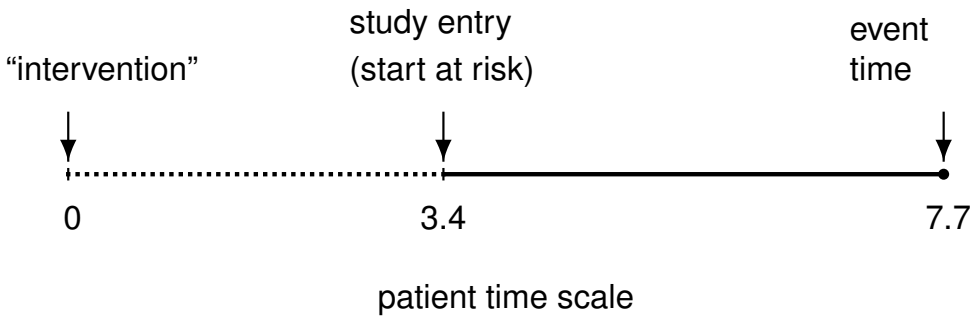
Summary

Left truncated data: structure

- In calendar time scale



- In patient time scale



Left truncated data: data representation and analysis

Individuals only contribute while they are in follow-up

Left truncated data: data representation and analysis

Individuals only contribute while they are in follow-up

- Data: extra column, describing entry time in risk set

id	entry time	event time	status
1	0.0	4.3	1
2	0.0	5.6	0
3	3.4	7.7	1

- In R: `Surv(entry.time, event.time, status)`

Outline

Standard setting
- Left truncated data
- Time-varying covariables

Competing risks
- Aalen-Johansen estimator
- Competing risks: subdistribution

Summary

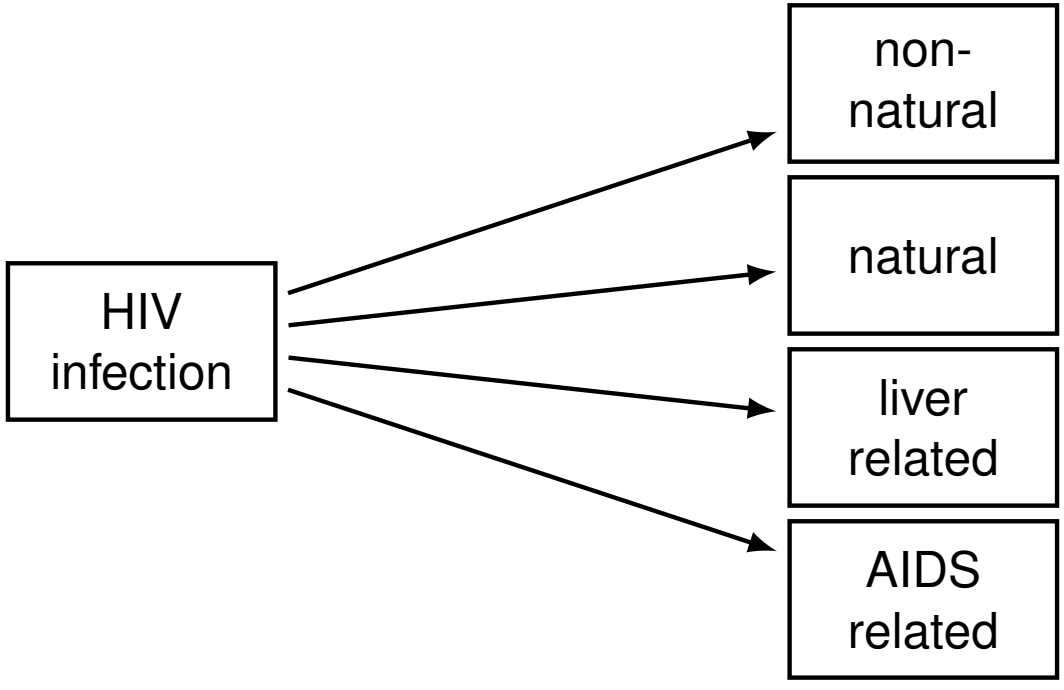
Two types [Kalbfleisch & Prentice, 2002]

- External: develops independently from disease process
  - **Defined:** all values known at time origin
    Examples: calendar period, age
  - **Ancillary:** external process
    Example: air pollution.
- Internal: reflects disease process, markers
  - Random process
  - Exists only as long as the person is alive
  - Direct causal relation with event

Two types [Kalbfleisch & Prentice, 2002]

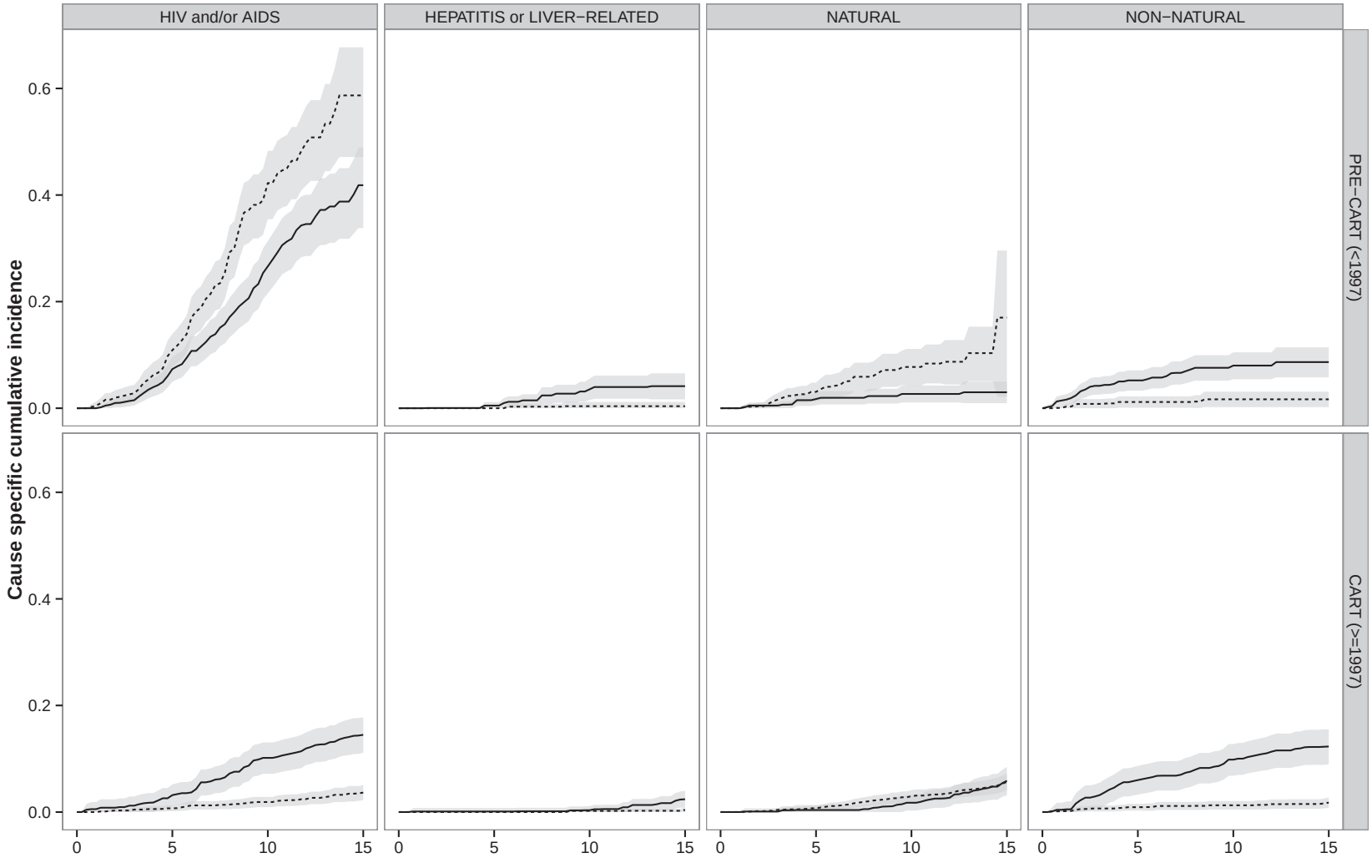
- External: develops independently from disease process
  - **Defined:** all values known at time origin
    Examples: calendar period, age
  - **Ancillary:** external process
    Example: air pollution.
- Internal: reflects disease process, markers
  - Random process
  - Exists only as long as the person is alive
  - Direct causal relation with event
- Hazard: instantaneous event risk $\longleftrightarrow$ instantaneous covariable value

Causes of death after HIV infection



- Has the spectrum in COD changed after introduction of combination anti-retroviral therapy (cART)?
  Two periods: I: $\leq$ 1996; II: $\geq$ 1997
- Some individuals HIV infected $\leq$ 1996, but follow-up in period II

Cause-specific mortality by calendar period and hepatitis C status



Single event (overall mortality)

- Counting process representation: split into pseudo-individuals based on periods with constant covariable value

id	start.time	stop.time	status	cal.period
1	0	6	0	$\leq$ 1996
1	6	8	1	$\geq$ 1997

- administrative censoring in 1996: `stop.time` in 1st row
- enters in risk set "$\geq$ 1997" after 6 years: `start.time` in 2nd row
  similar to late entry/left truncation

Single event (overall mortality)

- Counting process representation: split into pseudo-individuals based on periods with constant covariable value

id	start.time	stop.time	status	cal.period
1	0	6	0	$\leq$ 1996
1	6	8	1	$\geq$ 1997

- administrative censoring in 1996: `stop.time` in 1st row
- enters in risk set "$\geq$ 1997" after 6 years: `start.time` in 2nd row
  similar to late entry/left truncation
  "internal left truncation" [Andersen et al, 1993]
  "administrative late entry"

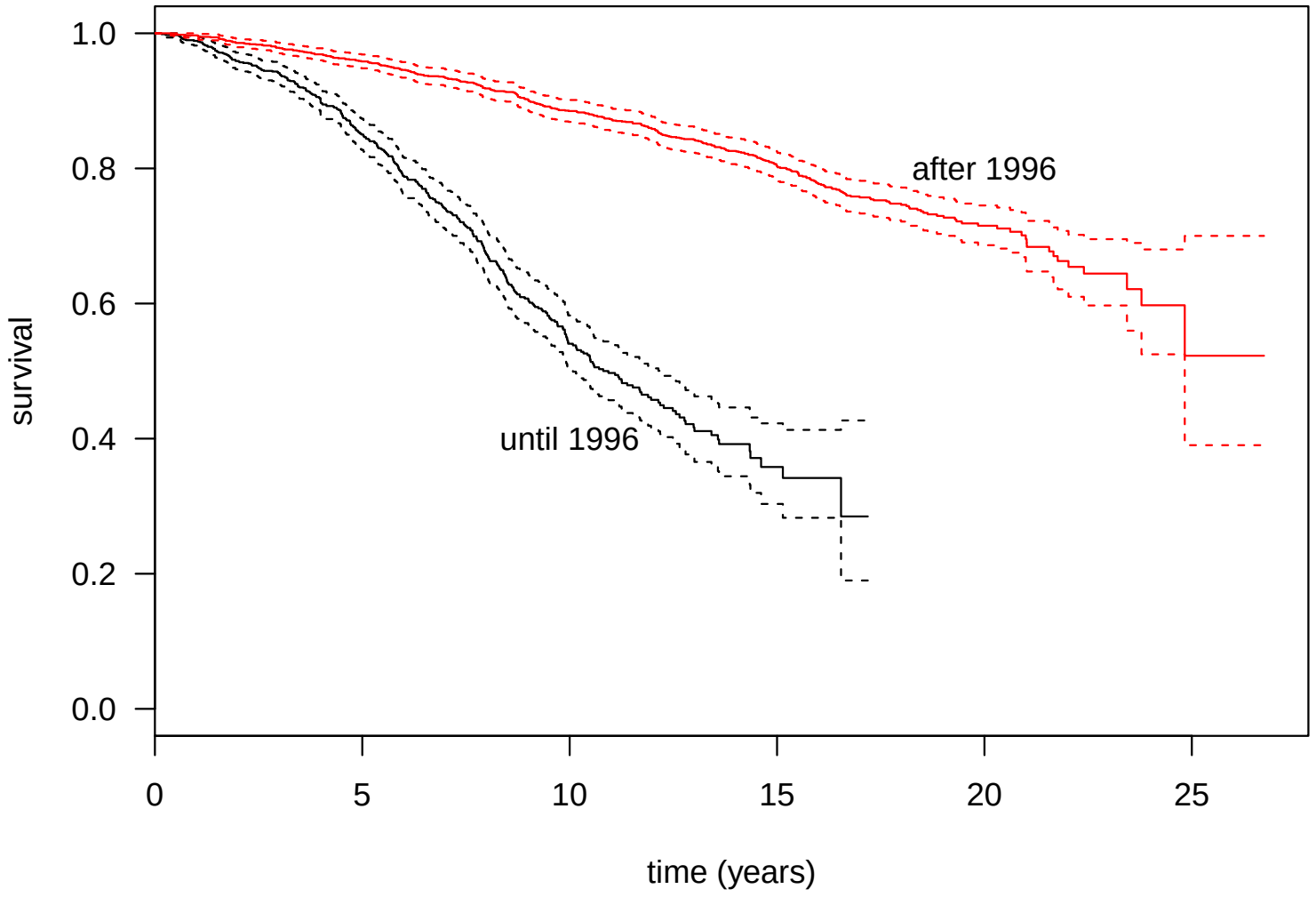
Single event (overall mortality)

- Counting process representation: split into pseudo-individuals based on periods with constant covariable value

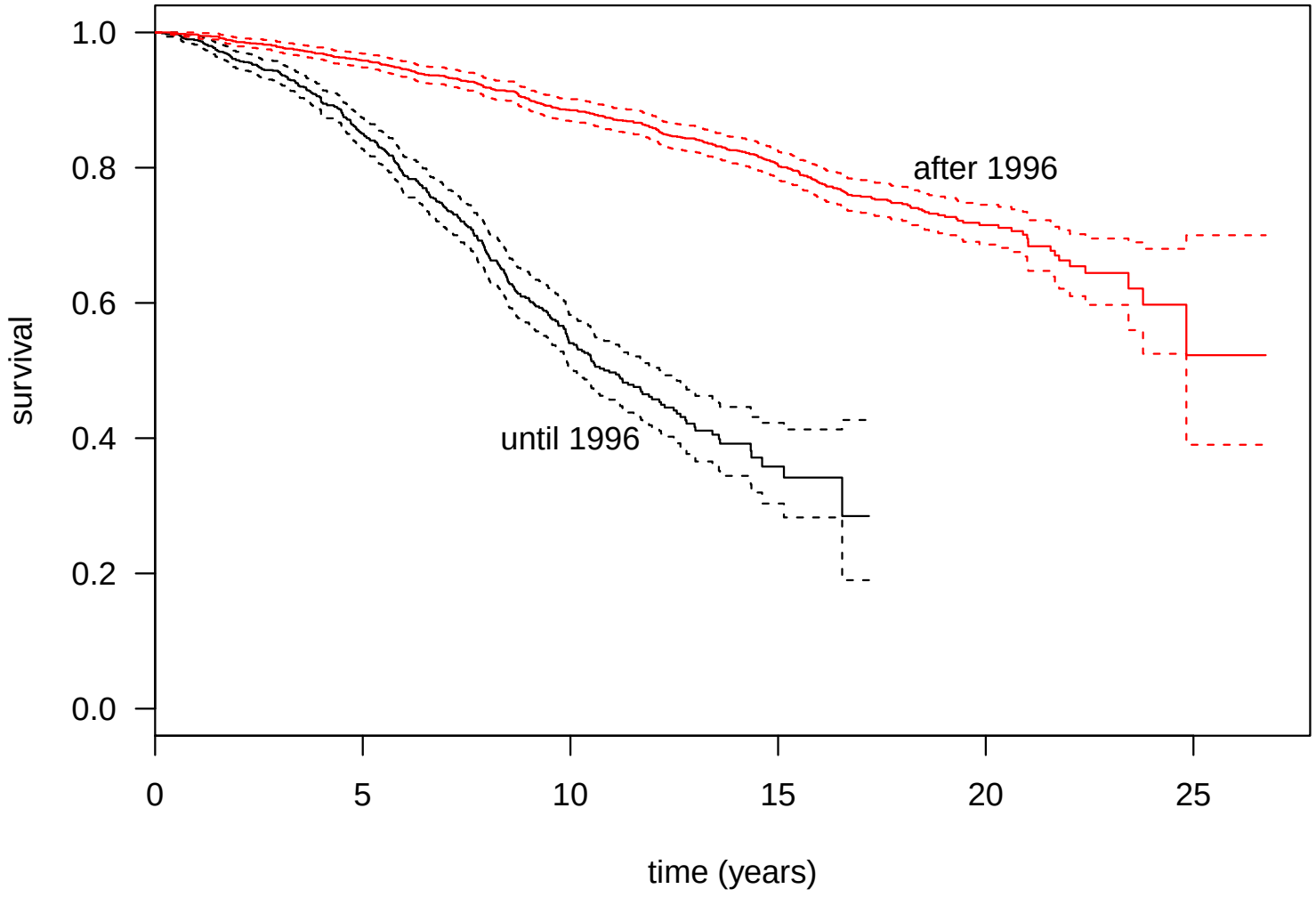
id	start.time	stop.time	status	cal.period
1	0	6	0	$\leq$ 1996
1	6	8	1	$\geq$ 1997

- administrative censoring in 1996: `stop.time` in 1st row
- enters in risk set "$\geq$ 1997" after 6 years: `start.time` in 2nd row
  similar to late entry/left truncation
  "internal left truncation" [Andersen et al, 1993]
  "administrative late entry"
- Kaplan-Meier per period: two separate analyses

Kaplan-Meier per period



Kaplan-Meier per period interpretation?



Interpretation

- We use $\leq$ 1996 to provide additional data for $\geq$ 1997
- Single event type
  - Assumption: all individuals $\geq$ 1997 same death hazard
  - K-M estimates for those that remained in single period

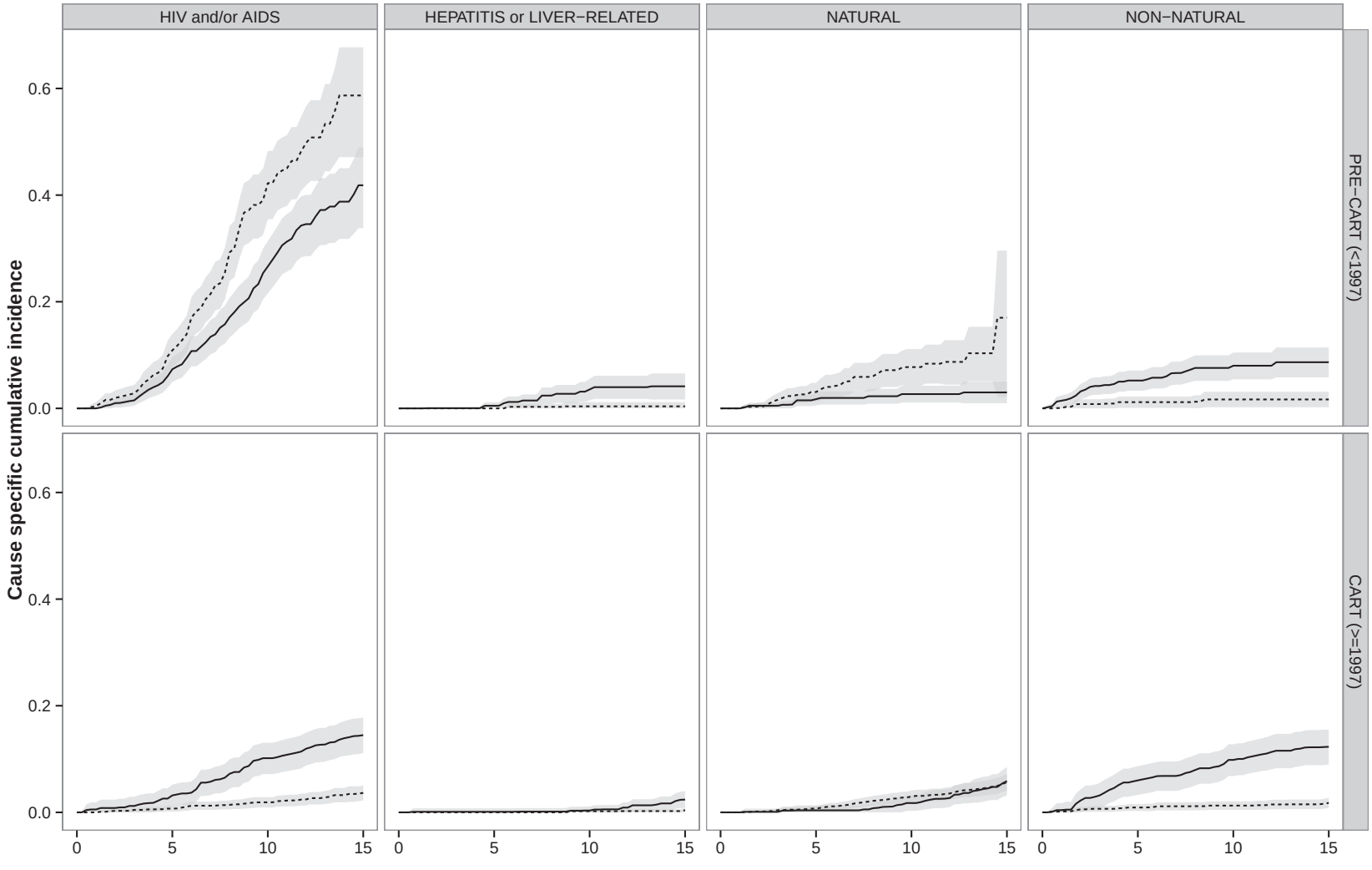
Outline

Standard setting
- Left truncated data
- Time-varying covariables

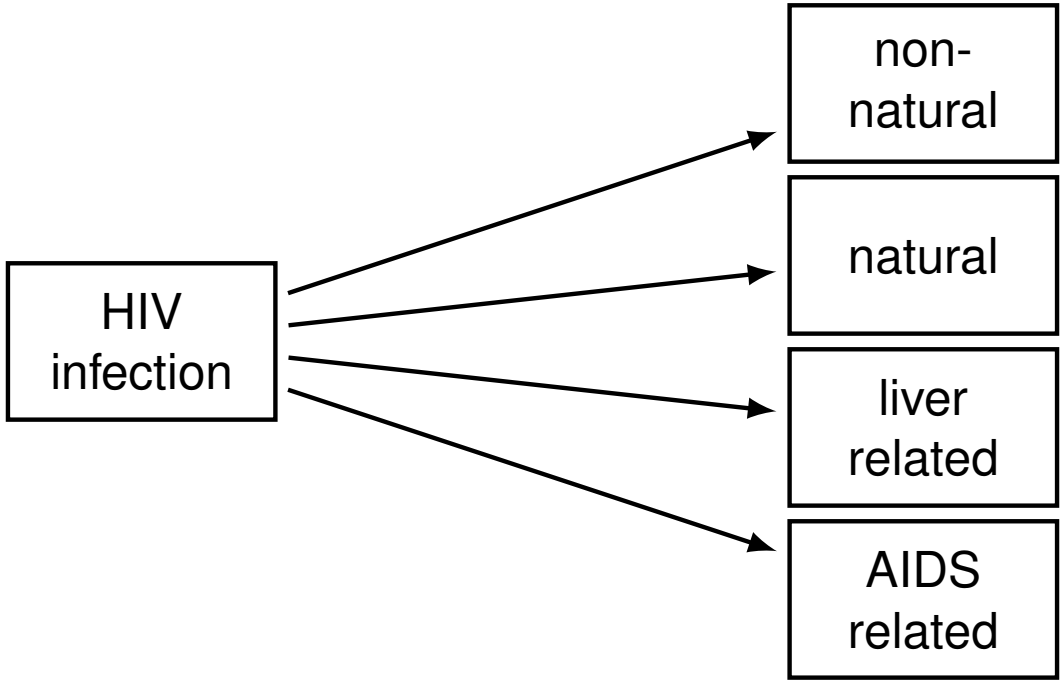
Competing risks
- Aalen-Johansen estimator
- Competing risks: subdistribution

Summary

Cause-specific mortality by calendar period and hepatitis C status



Causes of death after HIV infection



- Has the spectrum in COD changed after introduction of combination anti-retroviral therapy (cART)?
  Two periods: I: $\leq$ 1996; II: $\geq$ 1997
- Some individuals HIV infected $\leq$ 1996, but follow-up in period II
- Two hazards, cause-specific and subdistribution

Observed data

$\{(v_1, x_1, e_1\delta_1), \ldots, (v_N, x_N, e_N\delta_N)\}$

- $x_i = \min\{t_i, c_i\}$, $\delta_i = \{t_i \leq c_i\}$, $e_i \in \{1, \ldots, K\}$
- v_i entry time (or change in time-varying covariable)
- $t_{(1)}, \ldots, t_{(n)}$ ordered unique event times of any type
- $d_k(t_{(i)})$ number of events at $t_{(i)}$ of type k
- $d(t_{(i)})$ total number of events at $t_{(i)}$
- $r(t_{(i)})$ number observed at risk

Covariables $\mathbf{Z}_i(t) = (Z_{i1}(t), \ldots, Z_{ip}(t))^\top$

Aalen-Johansen estimator

$$\widehat{F_k}^{\mathrm{AJ}}(t) = \sum_{i:t_{(i)} \leq t} \mathrm{KM}(t_{(i)}-) \times \widehat{\lambda_k}(t_{(i)})$$

$$\mathrm{KM}(t_{(i)}-) = \prod_{j:t_{(j)} < t_{(i)}} \left(1 - \frac{d(t_{(j)})}{r(t_{(j)})}\right) \quad \text{Kaplan-Meier}$$

$$\widehat{\lambda_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r(t_{(i)})} \quad \text{cause specific hazard}$$

- Standard rate estimation
  - Individual with competing event leaves the risk set
  - Create pseudo-individuals for change in calendar period

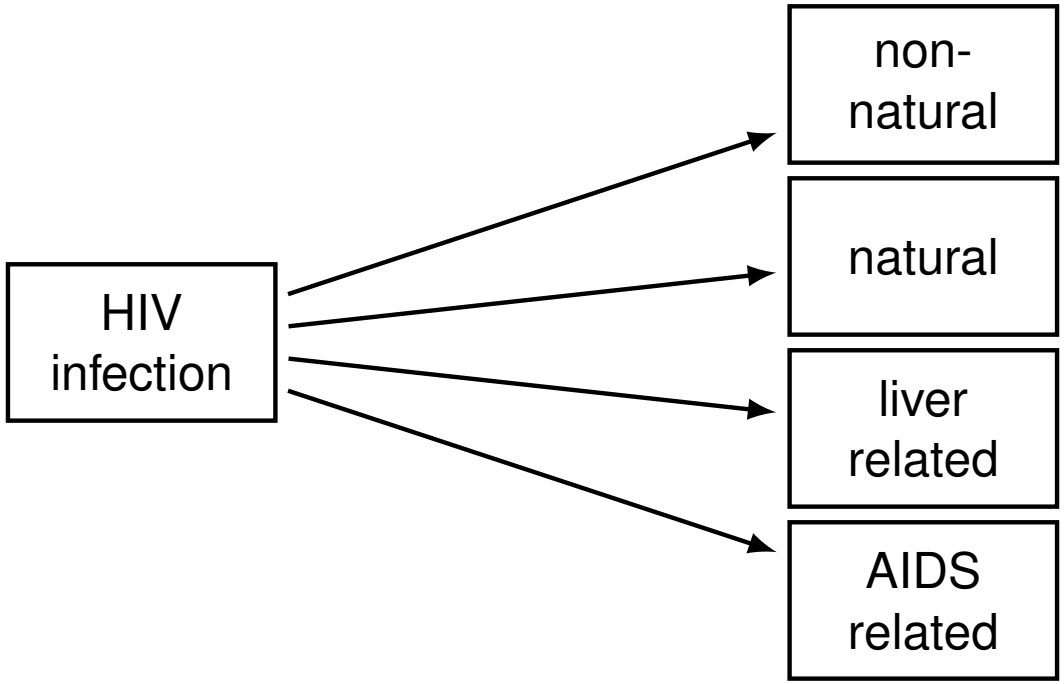
Single event (overall mortality)

- Counting process representation: split into pseudo-individuals based on periods with constant covariable value

id	start.time	stop.time	status	cal.period
1	0	6	0	$\leq$ 1996
1	6	8	1	$\geq$ 1997

- administrative censoring in 1996: `stop.time` in 1st row
- enters in risk set "$\geq$ 1997" after 6 years: `start.time` in 2nd row
  similar to late entry/left truncation
  "internal left truncation" [Andersen et al, 1993]
  "administrative late entry"
- Kaplan-Meier per period: two separate analyses

Causes of death after HIV infection



- Has the spectrum in COD changed after introduction of combination anti-retroviral therapy (cART)?
  Two periods: I: $\leq$ 1996; II: $\geq$ 1997
- Some individuals HIV infected $\leq$ 1996, but follow-up in period II
- Two hazards, cause-specific and subdistribution
- Here: change in spectrum of COD $\rightarrow$ subdistribution

Outline

Standard setting
- Left truncated data
- Time-varying covariables

Competing risks
- Aalen-Johansen estimator
- Competing risks: subdistribution

Summary

Time-varying covariables and the subdistribution hazard

Question on Researchgate.net, May 2015

> *Is there any possibility to add time-dependent covariates in the Fine-Gray model?*
>
> ***R1:*** *Time-dependent variables is not possible using cmprsk/crr. Is there other packages that can do this?*

Time-varying covariables and the subdistribution hazard

Question on Researchgate.net, May 2015

> *Is there any possibility to add time-dependent covariates in the Fine-Gray model?*
>
> ***R1:*** *Time-dependent variables is not possible using cmprsk/crr. Is there other packages that can do this?*
>
> ***R2:*** *It seems that the inclusion of time-dependent covariates in the Fine and Gray model leads to biased results (Latouche A., Porcher R. & Chevret S. (2005) and Putter H., Fiocco M. and Geskus R. (2007)).*

Subdistribution $\widehat{F}_k$: product-limit estimator

$$\widehat{\widehat{F}_k}^{\mathrm{PL}}(t) = \prod_{i:t_{(j)} \leq t} \left\{ 1 - \widehat{h_k}(t_{(j)}) \right\} \text{ with } \widehat{h_k}(t_{(j)}) = \frac{d_k(t_{(j)})}{r^*(t_{(j)})}$$

No censoring: individuals with competing event remain in risk set forever. Small change in data

Administrative censoring: individuals with competing event leave risk set at date of administrative censoring.

General censoring: Estimate time-to-censoring distribution. Then for those with competing event:

- multiply impute censoring times
- reweight them by probability to remain uncensored

Left truncation Weights determined by time-to-entry distribution

Late entry weights

- $\widehat{\widehat{\Gamma}}$: reverse role of event time T_i and censoring C_i:

$$\widehat{\widehat{\Gamma}}(t) = \widehat{\mathrm{P}}(C > t) = \prod_{j:c_{(j)} \leq t} \left\{ 1 - \frac{m_j}{r(c_{(j)})} \right\}$$

  m_j: number of censorings at $c_{(j)}$

- Left truncation. Entry time $V_i \sim \Phi$, $\Phi(t) = \mathrm{P}(V_i \leq t)$
- $\widehat{\Phi}$: reverse role of $X_i = T_i \wedge C_i$ and truncation time V_i, V_i ("event") is right truncated by X_i:

$$\widehat{\Phi}(t) = \widehat{\mathrm{P}}(V \leq t) = \widehat{\mathrm{P}}(-V \geq -t) = \prod_{j:-v_{(j)} < -t} \left\{ 1 - \frac{w_j}{r(v_{(j)})} \right\}$$

$$= \prod_{j:v_{(j)} > t} \left\{ 1 - \frac{w_j}{r(v_{(j)})} \right\}.$$

w_j: number of entries at $v_{(j)}$

General product-limit estimator of F_k

$$\widehat{\overline{F}}_k^{\mathrm{PL}}(t) = \prod_{i:t_{(i)} \leq t} \left\{ 1 - \widehat{h_k}(t_{(i)}) \right\}$$

$$\widehat{h_k}(t_{(i)}) = \frac{d_k(t_{(i)})}{r^*(t_{(i)})}$$

Weights: contribution $\omega_l(t_{(i)})$ of individual l to $r^*(t_{(i)})$ is

- censored or event of type k before $t_{(i)}$: 0
- still at risk at $t_{(i)}$: 1
- competing event at $t_{(j)}$ before $t_{(i)}$: weight

$$\frac{\widehat{\overline{\Gamma}}(t_{(i)}-)}{\widehat{\overline{\Gamma}}(t_{(j)}-)} \times \frac{\widehat{\Phi}(t_{(i)}-)}{\widehat{\Phi}(t_{(j)}-)}$$

$$\approx \widehat{\mathrm{P}}\{C > t_{(i)} | C > t_{(j)}\} \times 1/\widehat{\mathrm{P}}\{V < t_{(j)} | V < t_{(i)}\}$$

Subdistribution $\widehat{F}_k$: product-limit estimator

$$\widehat{\widehat{F}_k}^{\mathrm{PL}}(t) = \prod_{i:t_{(j)} \leq t} \left\{ 1 - \widehat{h_k}(t_{(j)}) \right\} \text{ with } \widehat{h_k}(t_{(j)}) = \frac{d_k(t_{(j)})}{r^*(t_{(j)})}$$

No censoring: individuals with competing event remain in risk set forever. Small change in data

Administrative censoring: individuals with competing event leave risk set at date of administrative censoring.

General censoring: Estimate time-to-censoring distribution. Then for those with competing event:

- multiply impute censoring times
- reweight them by probability to remain uncensored

Left truncation Weights determined by time-to-entry distribution

Internal left truncation $\approx$ administrative censoring

Two approaches [Geskus, 2015]

id	hiv	start.time	stop.time	status	cal.period
1	1991	0	6	0	$\leq$ 1996
1	1991	6	8	3	$\geq$ 1997
2	1991	0	4	2	$\leq$ 1996

Two approaches [Geskus, 2015]

id	hiv	start.time	stop.time	status	cal.period
1	1991	0	6	0	$\leq$ 1996
1	1991	6	8	3	$\geq$ 1997
2	1991	0	4	2	$\leq$ 1996

- Pseudo-individual approach: consider rows as coming from different individuals. Weights also determined by:
  - $\leq$ 1996: censorings at end of first period
  - $\geq$ 1997: late entries into second period

  PL-form equivalent to AJ-form [Geskus, 2011]

Two approaches [Geskus, 2015]

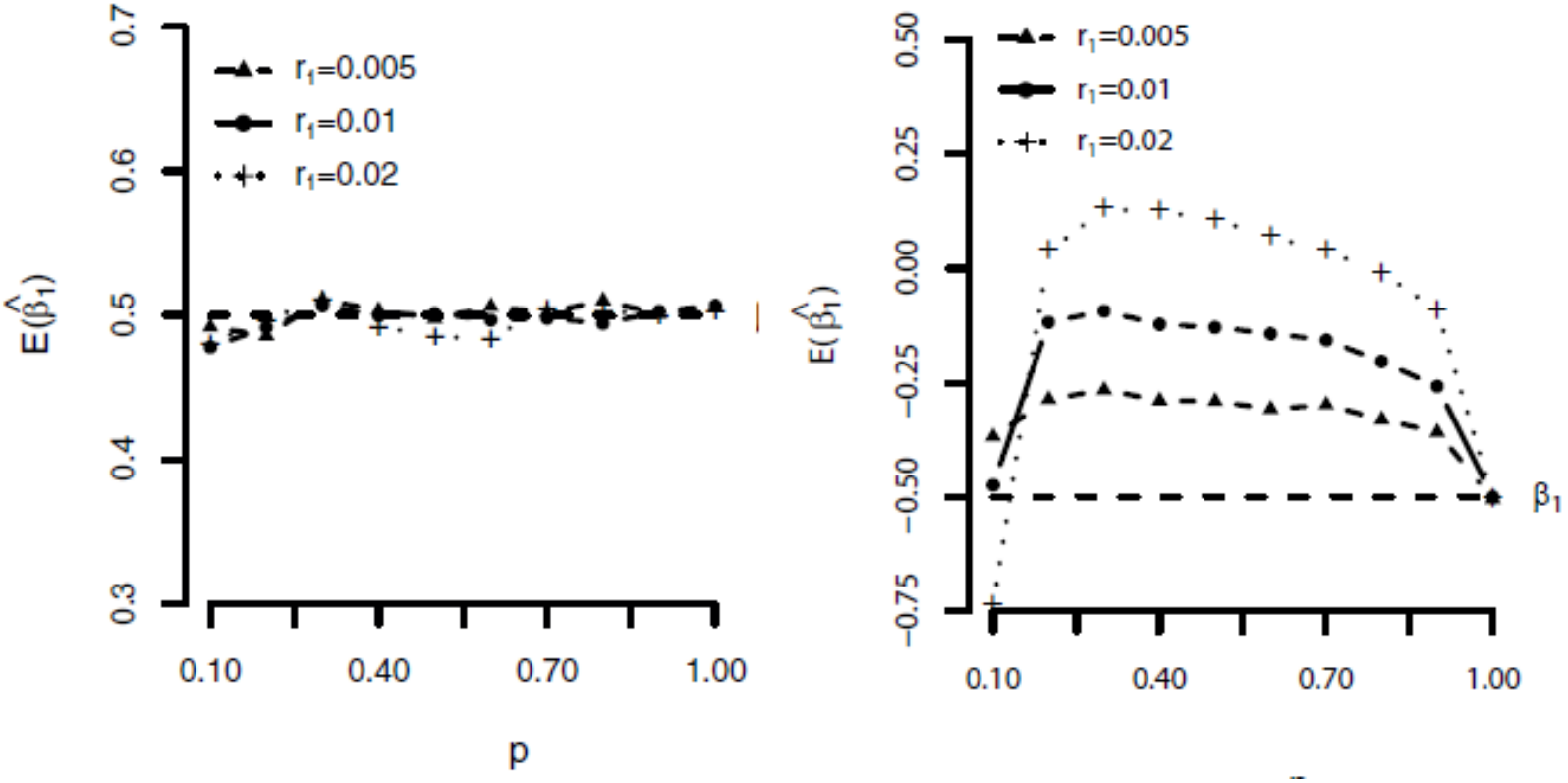
id	hiv	start.time	stop.time	status	cal.period
1	1991	0	6	0	$\leq$ 1996
1	1991	6	8	3	$\geq$ 1997
2	1991	0	4	2	$\leq$ 1996

- Pseudo-individual approach: consider rows as coming from different individuals. Weights also determined by:
  - $\leq$ 1996: censorings at end of first period
  - $\geq$ 1997: late entries into second period

  PL-form equivalent to AJ-form [Geskus, 2011]
- Internal approach: consider rows as continuing follow-up from same individual
  - No time-to-entry weights
    Individual 2 also contributes to period $\geq$ 1997
  - Classical situation: unobserved cure as competing event

Literature

- [Latouche et al, 2005]. Relapse and death after BMT
  aGvHD binary internal covariable.
  Internal approach. Simulation study, no censoring:
  - identifiable path (non-absorbing competing risk): no bias
  - non-identifiable path with LOCF: serious bias

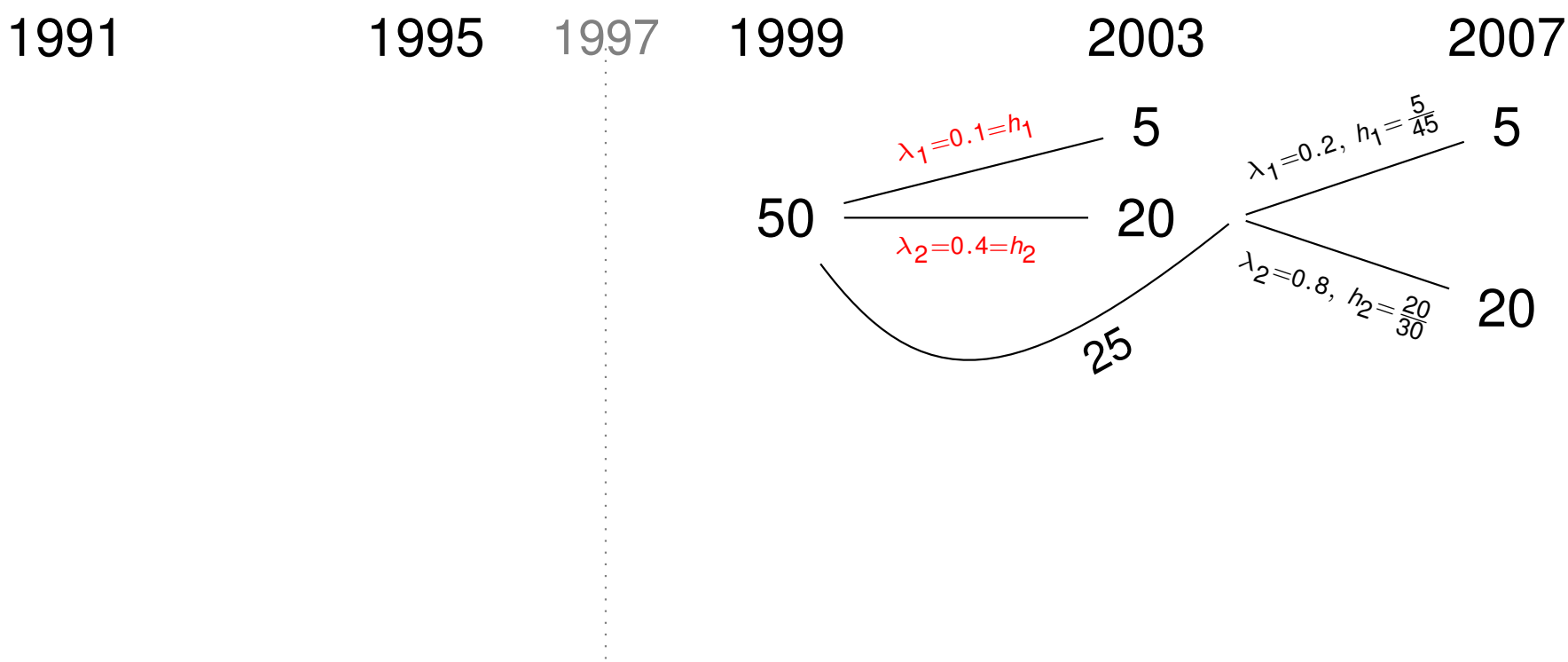

r_1=0.005
r_1=0.01
r_1=0.02
$E(\hat{\beta}_1)$
0.3
0.4
0.5
0.6
0.7
0.10
0.40
0.70
1.00
p
r_1=0.005
r_1=0.01
r_1=0.02
$E(\hat{\beta}_1)$
−0.75
−0.50
−0.25
0.00
0.25
0.50
β_1
0.10
0.40
0.70
1.00

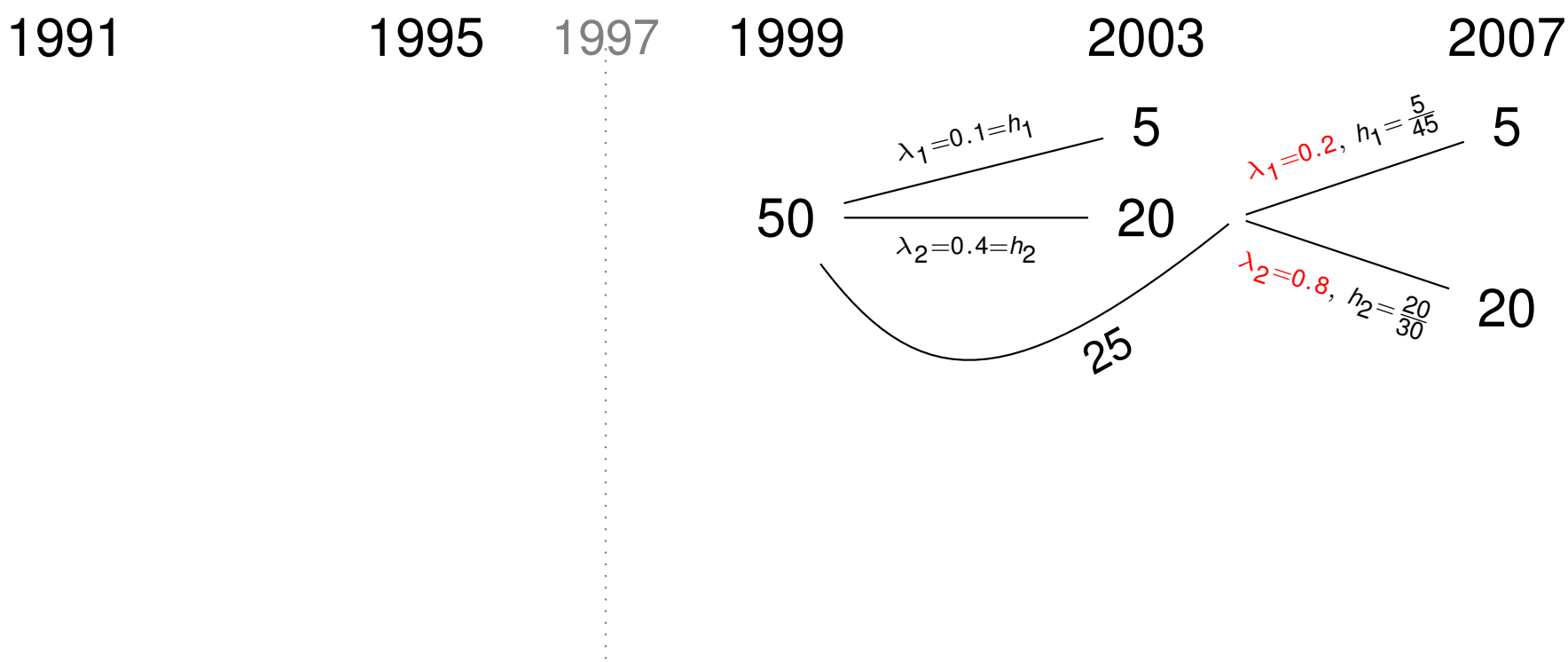
Literature

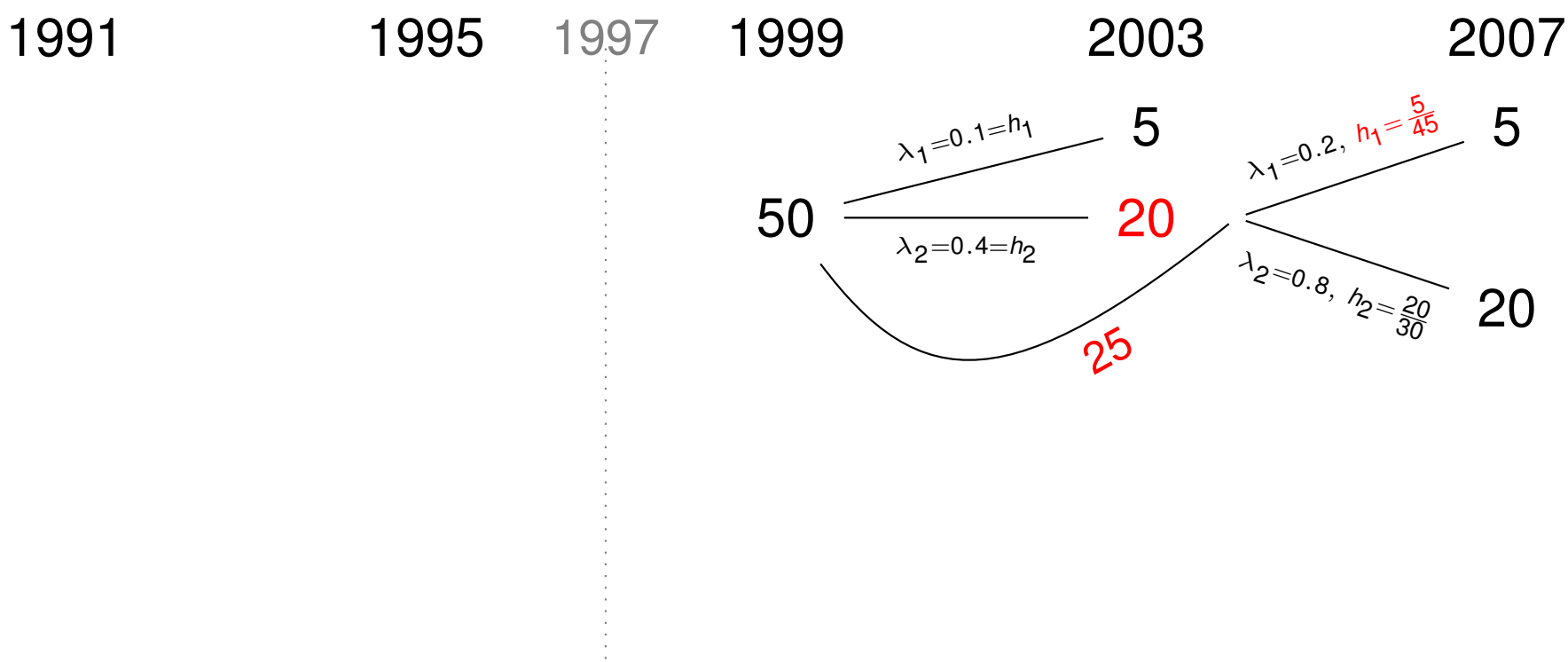
- [Latouche et al, 2005]. Relapse and death after BMT
  aGvHD binary internal covariable.
  Internal approach. Simulation study, no censoring:
  - identifiable path (non-absorbing competing risk): no bias
  - non-identifiable path with LOCF: serious bias
- [Beyersmann & Schumacher, 2008]. Death and discharge in ICU
  pneumonia binary internal covariable.
  Internal approach.
  "stopped covariate process" $Z(t \wedge T)$. Is same as LOCF

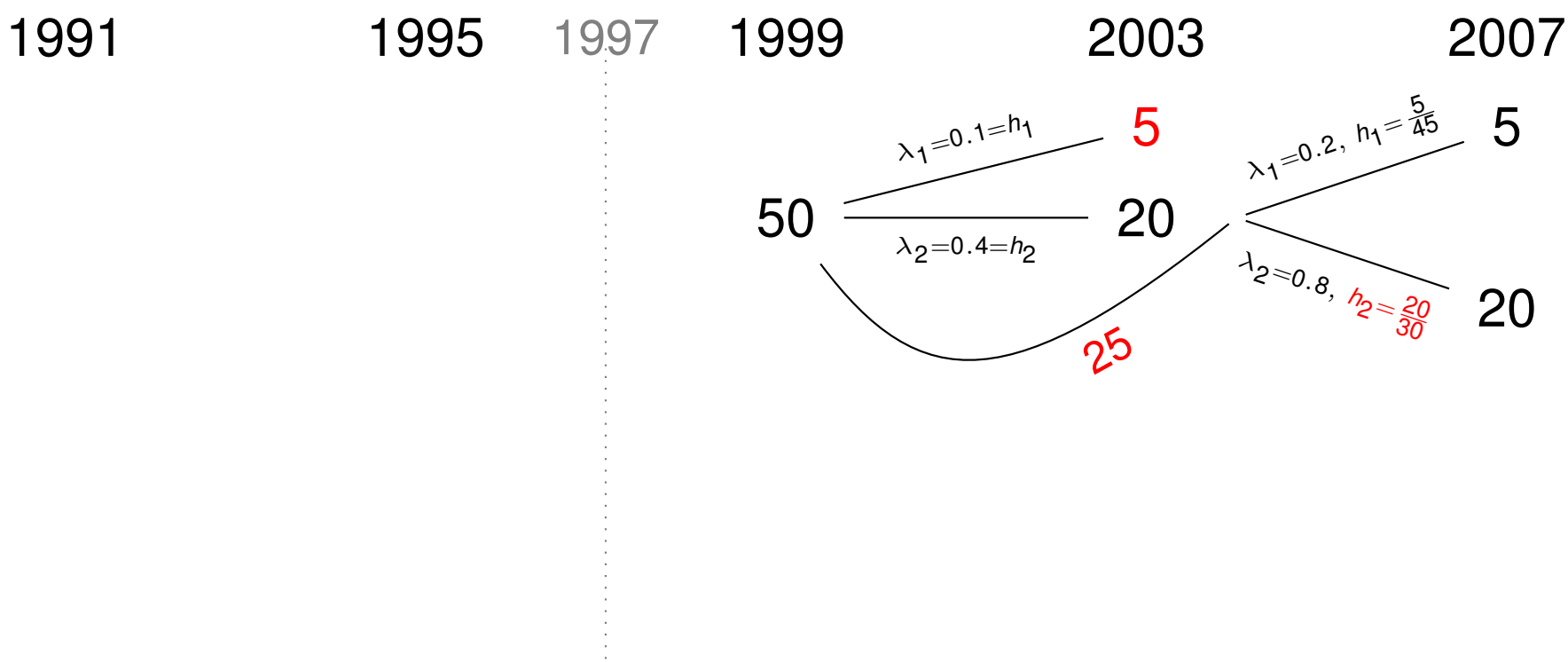
Literature

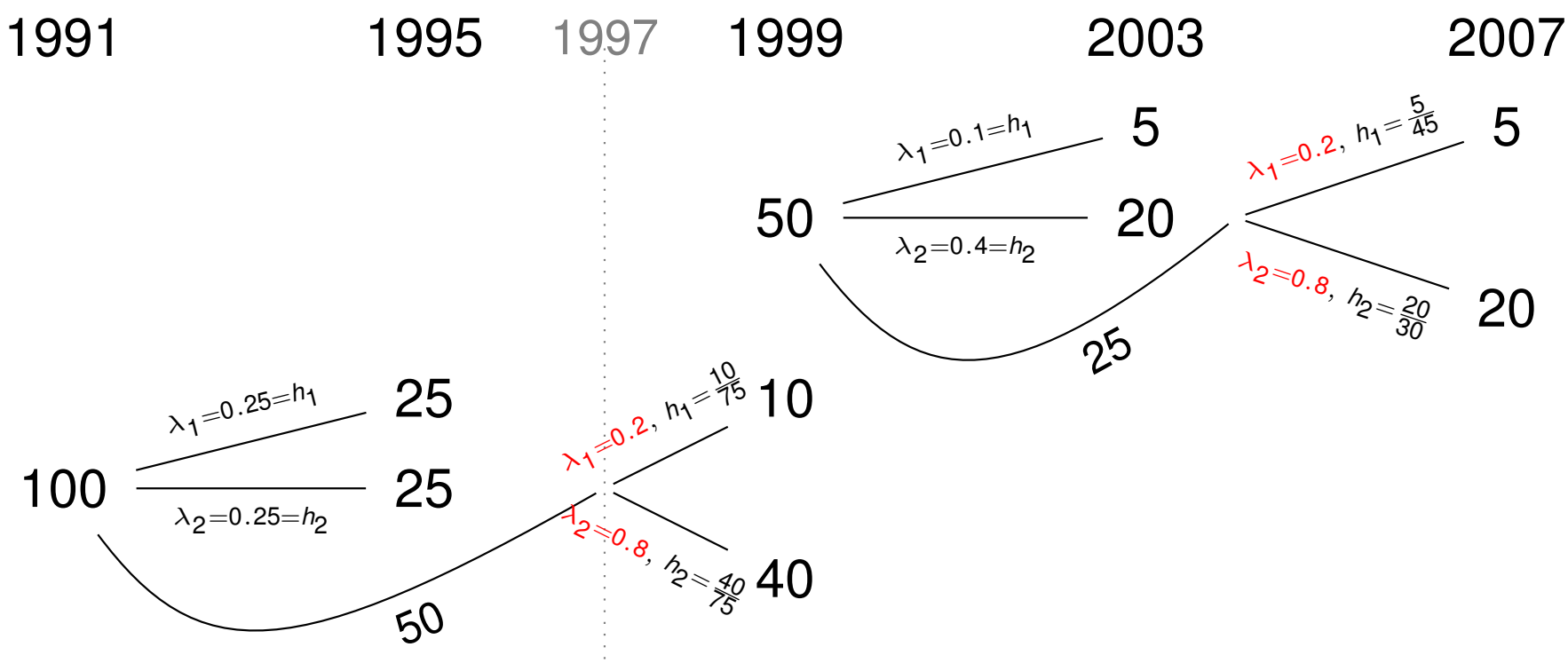
- [Latouche et al, 2005]. Relapse and death after BMT
  aGvHD binary internal covariable.
  Internal approach. Simulation study, no censoring:
  - identifiable path (non-absorbing competing risk): no bias
  - non-identifiable path with LOCF: serious bias
- [Beyersmann & Schumacher, 2008]. Death and discharge in ICU
  pneumonia binary internal covariable.
  Internal approach.
  "stopped covariate process" $Z(t \wedge T)$. Is same as LOCF
- [Deslandes & Chevret, 2010]. Death and discharge in ICU
  SOFA score internal continuous covariable.
  Internal approach.
  Joint model, using predicted value
  Simulation study: good performance.

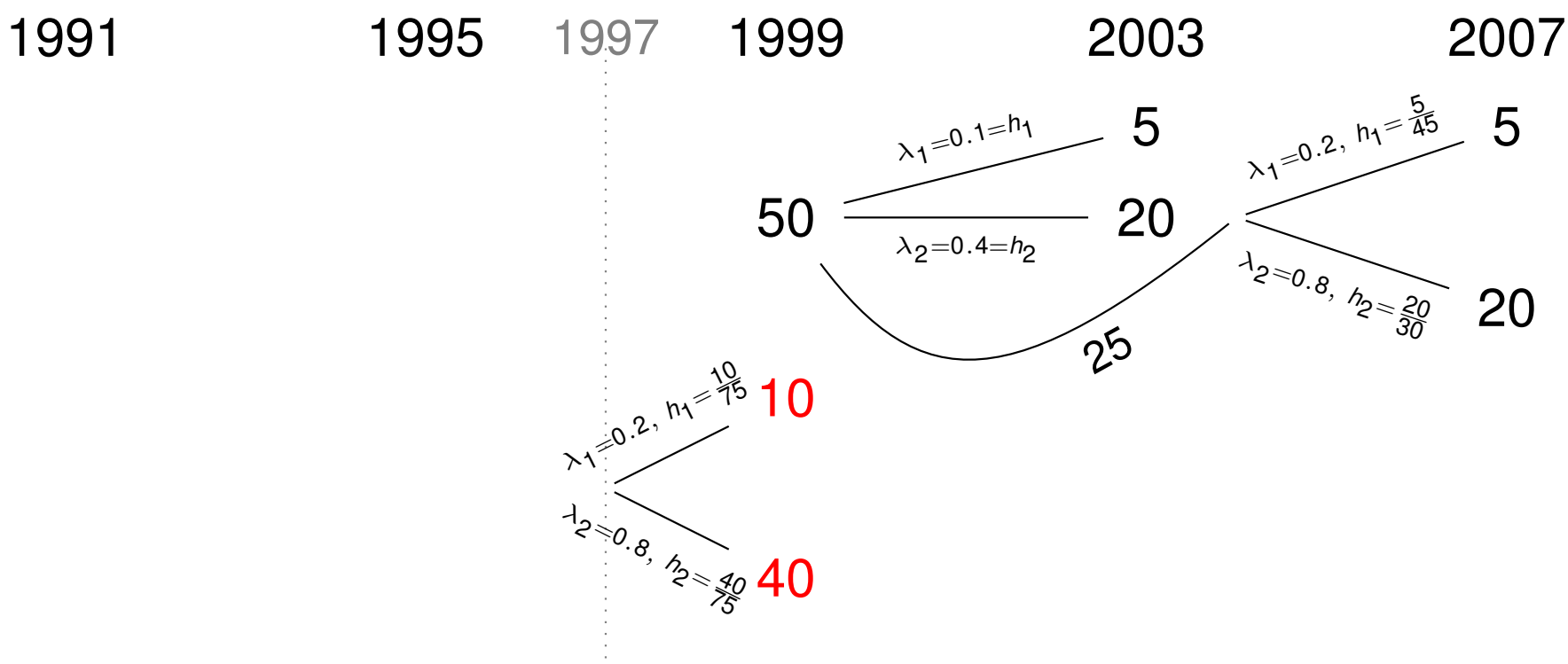

1991
1995
1997
1999
2003
2007
50
$\lambda_1=0.1=h_1$
5
$\lambda_2=0.4=h_2$
20
25
$\lambda_1=0.2,\ h_1=\frac{5}{45}$
5
$\lambda_2=0.8,\ h_2=\frac{20}{30}$
20


1991
1995
1997
1999
2003
2007
50
$\lambda_1=0.1=h_1$
5
$\lambda_2=0.4=h_2$
20
25
$\lambda_1=0.2$, $h_1=\frac{5}{45}$
5
$\lambda_2=0.8$, $h_2=\frac{20}{30}$
20

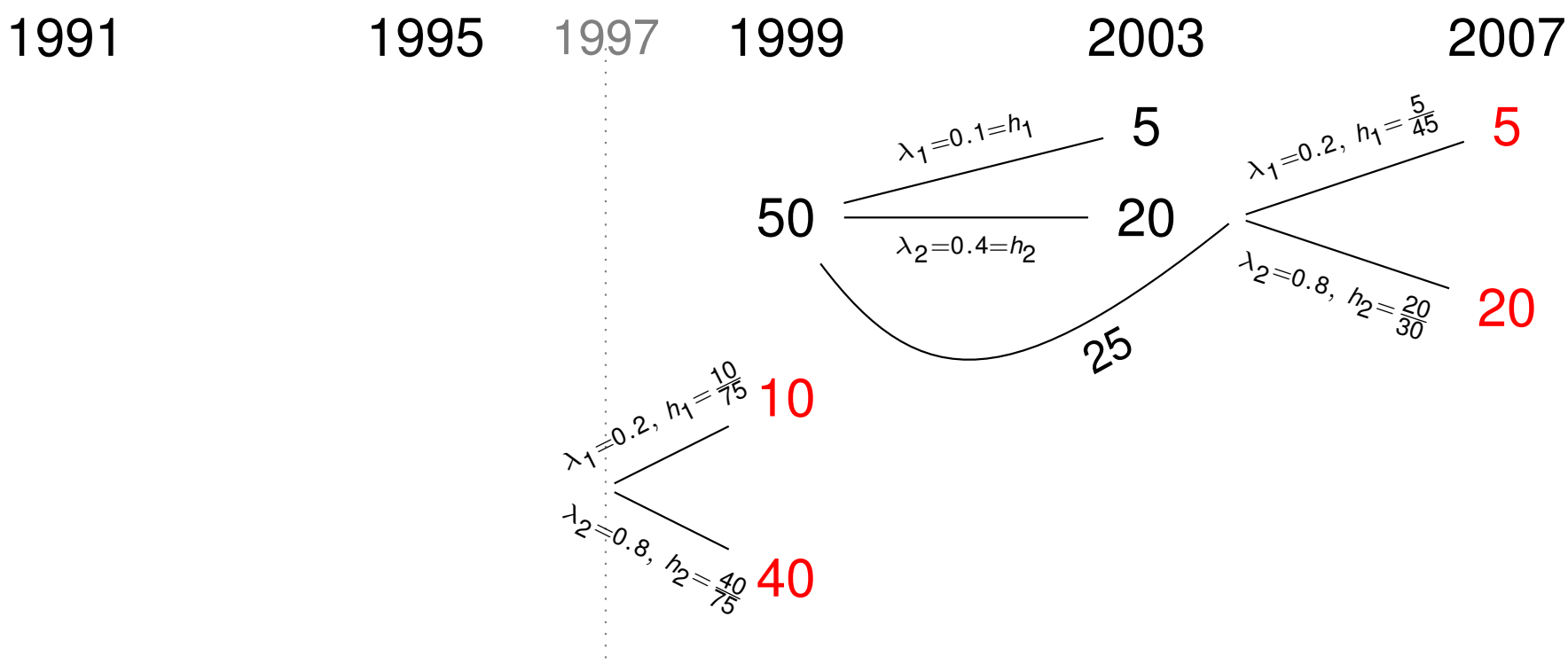

1991
1995
1997
1999
2003
2007
50
$\lambda_1=0.1=h_1$
5
$\lambda_2=0.4=h_2$
20
25
$\lambda_1=0.2,\ h_1=\frac{5}{45}$
5
$\lambda_2=0.8,\ h_2=\frac{20}{30}$
20


1991
1995
1997
1999
2003
2007
50
$\lambda_1=0.1=h_1$
5
20
$\lambda_2=0.4=h_2$
25
$\lambda_1=0.2,\ h_1=\frac{5}{45}$
5
$\lambda_2=0.8,\ h_2=\frac{20}{30}$
20

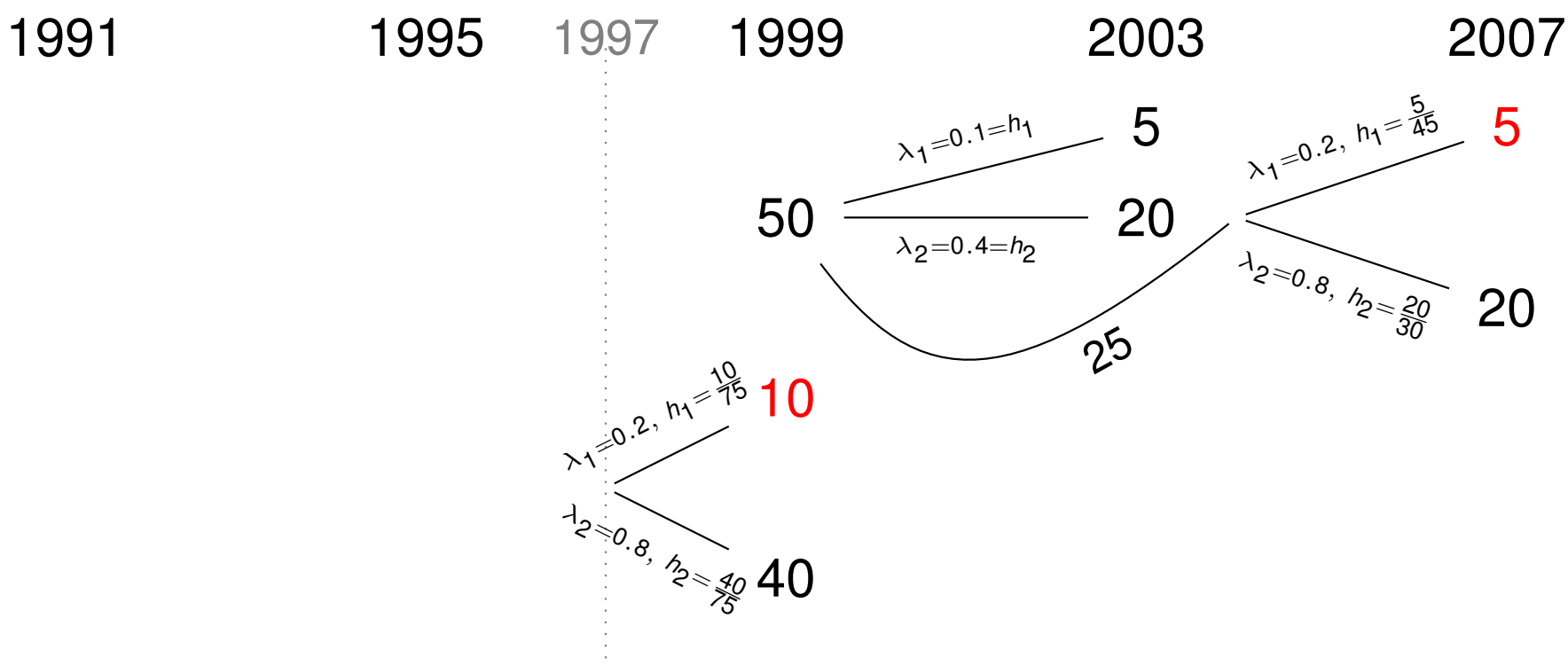

1991
1995
1997
1999
2003
2007
100
$\lambda_1=0.25=h_1$
$\lambda_2=0.25=h_2$
25
25
50
$\lambda_1=0.2$, $h_1=\frac{10}{75}$
$\lambda_2=0.8$, $h_2=\frac{40}{75}$
10
40
50
$\lambda_1=0.1=h_1$
$\lambda_2=0.4=h_2$
5
20
25
$\lambda_1=0.2$, $h_1=\frac{5}{45}$
$\lambda_2=0.8$, $h_2=\frac{20}{30}$
5
20



- Pseudo-individual: $\widehat{\mathrm{P}}(V<6)=1-\frac{50}{75}=\frac{1}{3}$, $\widehat{\mathrm{P}}(V\leq 6)=1$



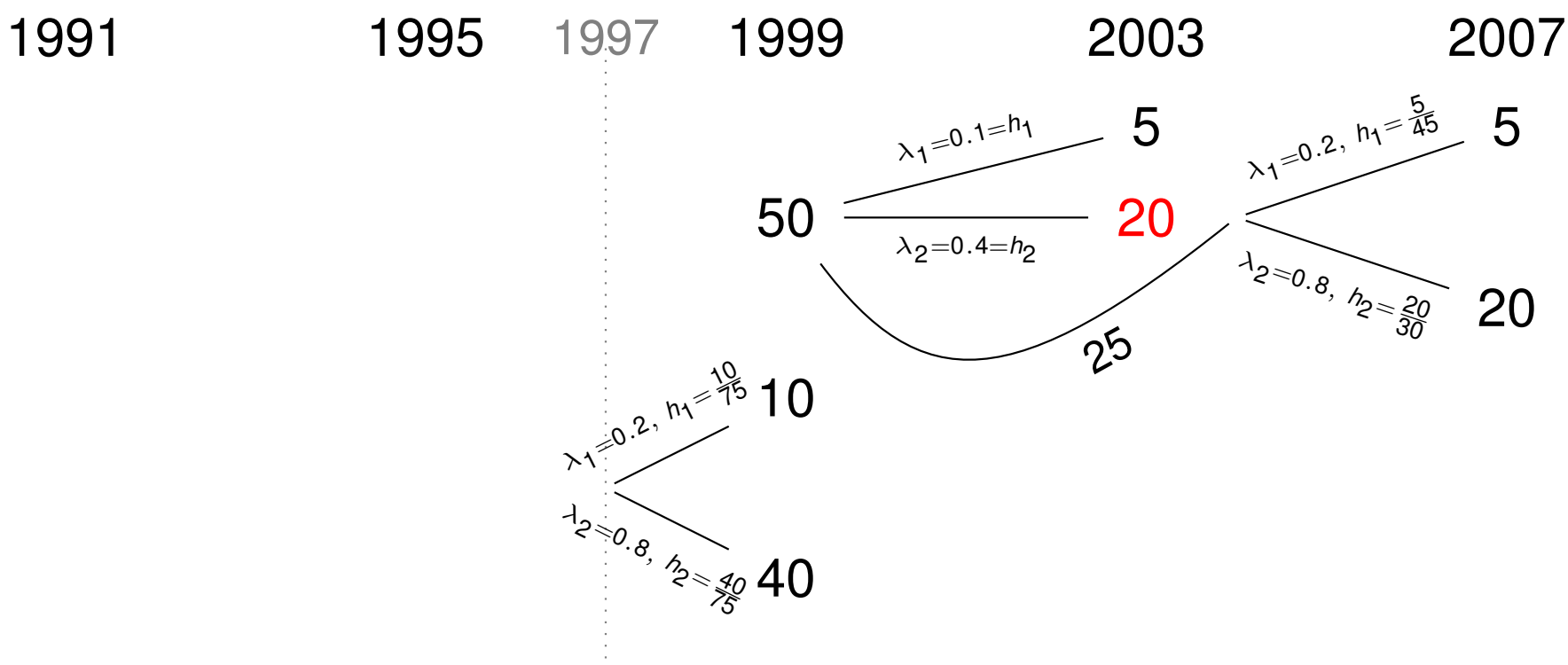
- Pseudo-individual: $\widehat{\mathrm{P}}(V<6)=1-\frac{50}{75}=\frac{1}{3}$, $\widehat{\mathrm{P}}(V\leq 6)=1$



- Pseudo-individual: $\widehat{\mathrm{P}}(V<6)=1-\frac{50}{75}=\frac{1}{3}$, $\widehat{\mathrm{P}}(V\leq 6)=1$

$$\widehat{h_1}(8|II)=\frac{5+10}{75+3\times 20}=\frac{15}{135}=\frac{1}{9}$$

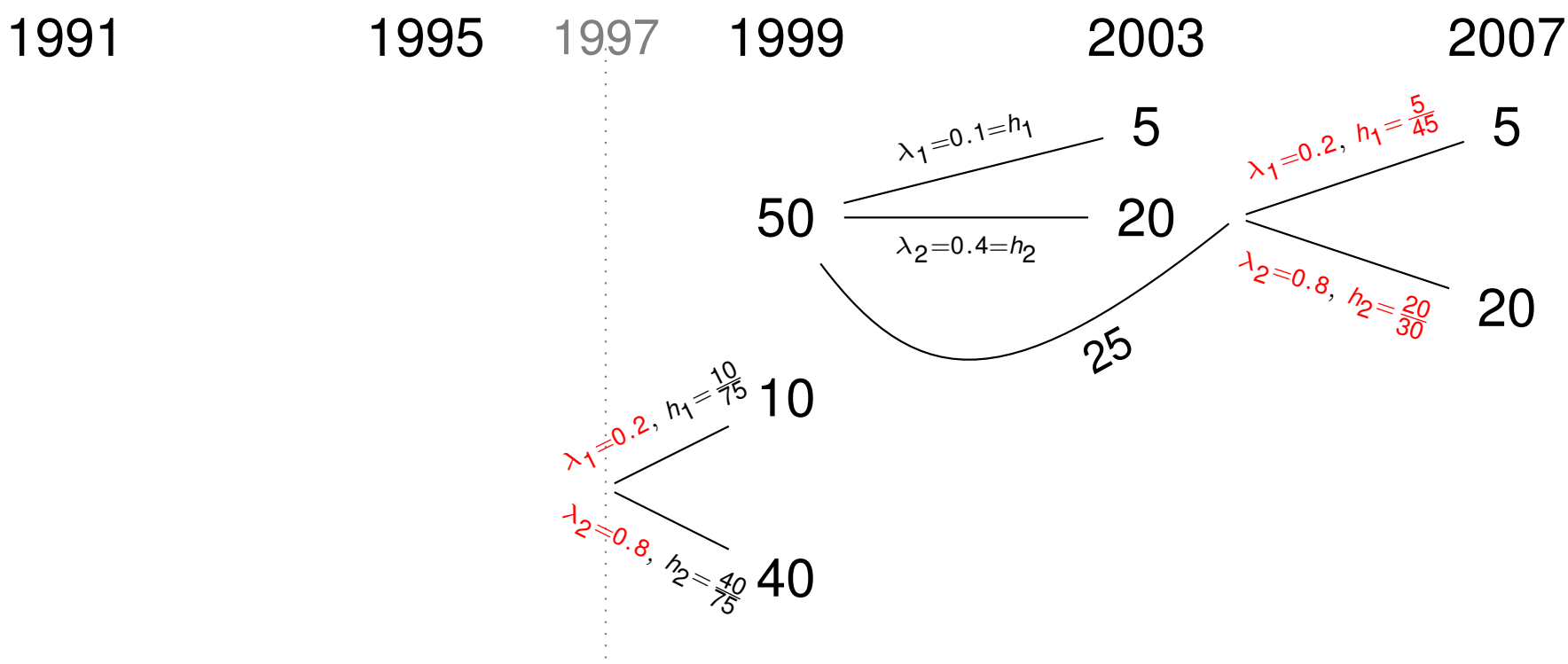
$$\widehat{h_2}(8|II)=\frac{20+40}{75+3\times 5}=\frac{60}{90}=\frac{2}{3}$$



- Pseudo-individual: $\widehat{\mathrm{P}}(V<6)=1-\frac{50}{75}=\frac{1}{3}$, $\widehat{\mathrm{P}}(V\leq 6)=1$

$$\widehat{h_1}(8|II)=\frac{5+10}{75+3\times 20}=\frac{15}{135}=\frac{1}{9}$$

$$\widehat{h_2}(8|II)=\frac{20+40}{75+3\times 5}=\frac{60}{90}=\frac{2}{3}$$



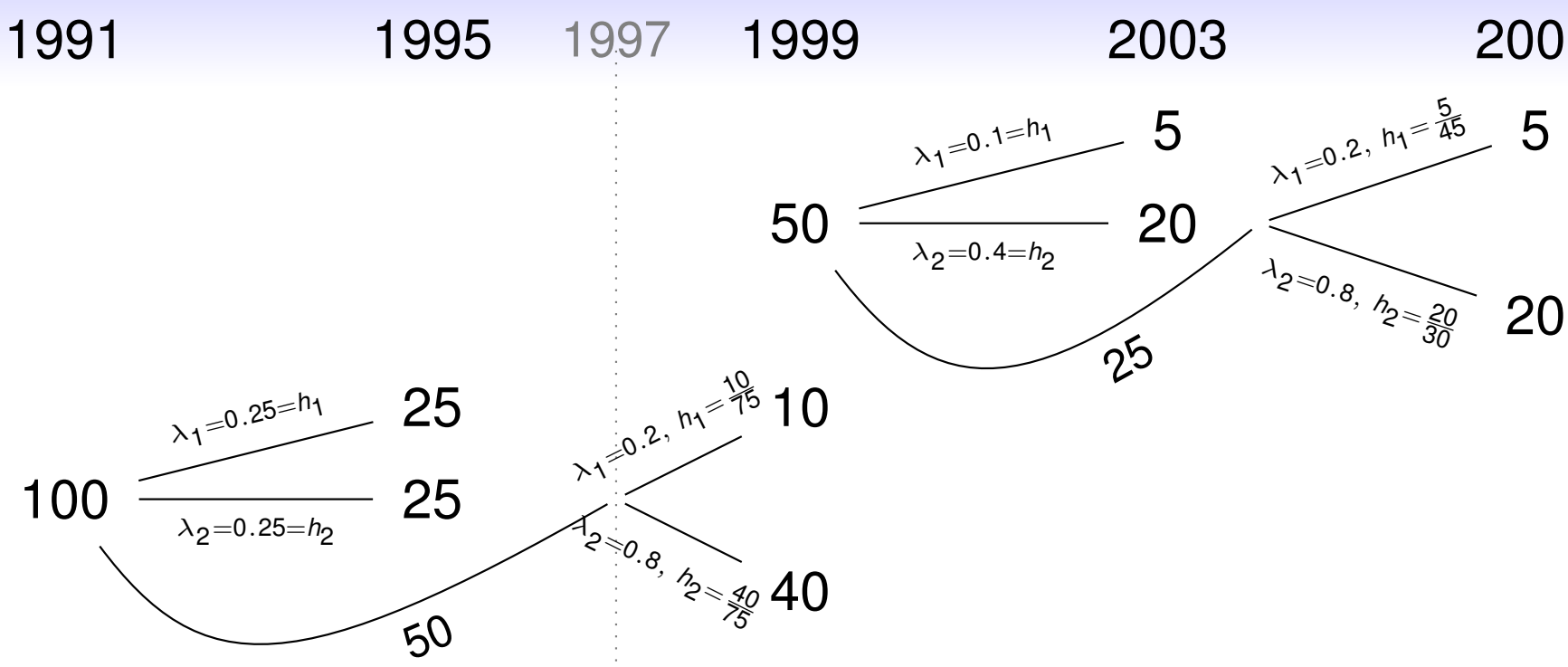
- Pseudo-individual: $\widehat{\mathrm{P}}(V<6)=1-\frac{50}{75}=\frac{1}{3}$, $\widehat{\mathrm{P}}(V\leq 6)=1$

$$\widehat{h_1}(8|II)=\frac{5+10}{75+3\times 20}=\frac{15}{135}=\frac{1}{9}$$

$$\widehat{h_2}(8|II)=\frac{20+40}{75+3\times 5}=\frac{60}{90}=\frac{2}{3}$$

Interpretation

- Pseudo-individual approach
  - Assumption: all individuals $\geq$ 1997 same cause-specific hazard
  - Estimate: subdistribution hazard for those that remained in single period



- Pseudo-individual: $\widehat{\mathrm{P}}(V<6)=1-\frac{50}{75}=\frac{1}{3}$, $\widehat{\mathrm{P}}(V\leq 6)=1$

$$\widehat{h_1}(8|II)=\frac{5+10}{75+3\times 20}=\frac{15}{135}=\frac{1}{9}$$
$$\widehat{h_2}(8|II)=\frac{20+40}{75+3\times 5}=\frac{60}{90}=\frac{2}{3}$$

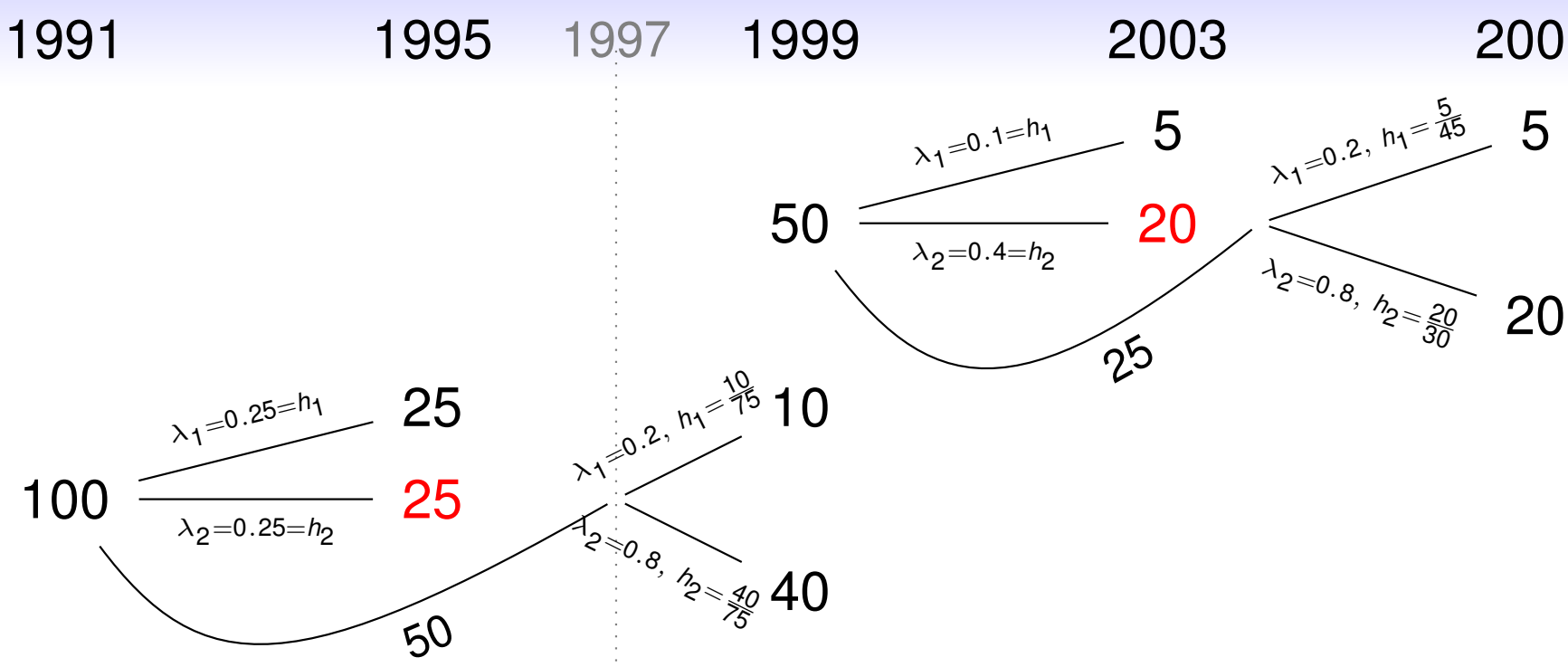
- Internal:

$$\widehat{h_1}(8|II)=\frac{5+10}{75+45}=\frac{15}{120}=\frac{1}{8}$$
$$\widehat{h_2}(8|II)=\frac{20+40}{75+30}=\frac{60}{105}=\frac{4}{7}$$

oucru

wellcome

UNIVERSITY OF OXFORD



- Pseudo-individual: $\widehat{\mathrm{P}}(V<6)=1-\frac{50}{75}=\frac{1}{3}$, $\widehat{\mathrm{P}}(V\leq 6)=1$

$$\widehat{h_1}(8|II)=\frac{5+10}{75+3\times 20}=\frac{15}{135}=\frac{1}{9}$$
$$\widehat{h_2}(8|II)=\frac{20+40}{75+3\times 5}=\frac{60}{90}=\frac{2}{3}$$

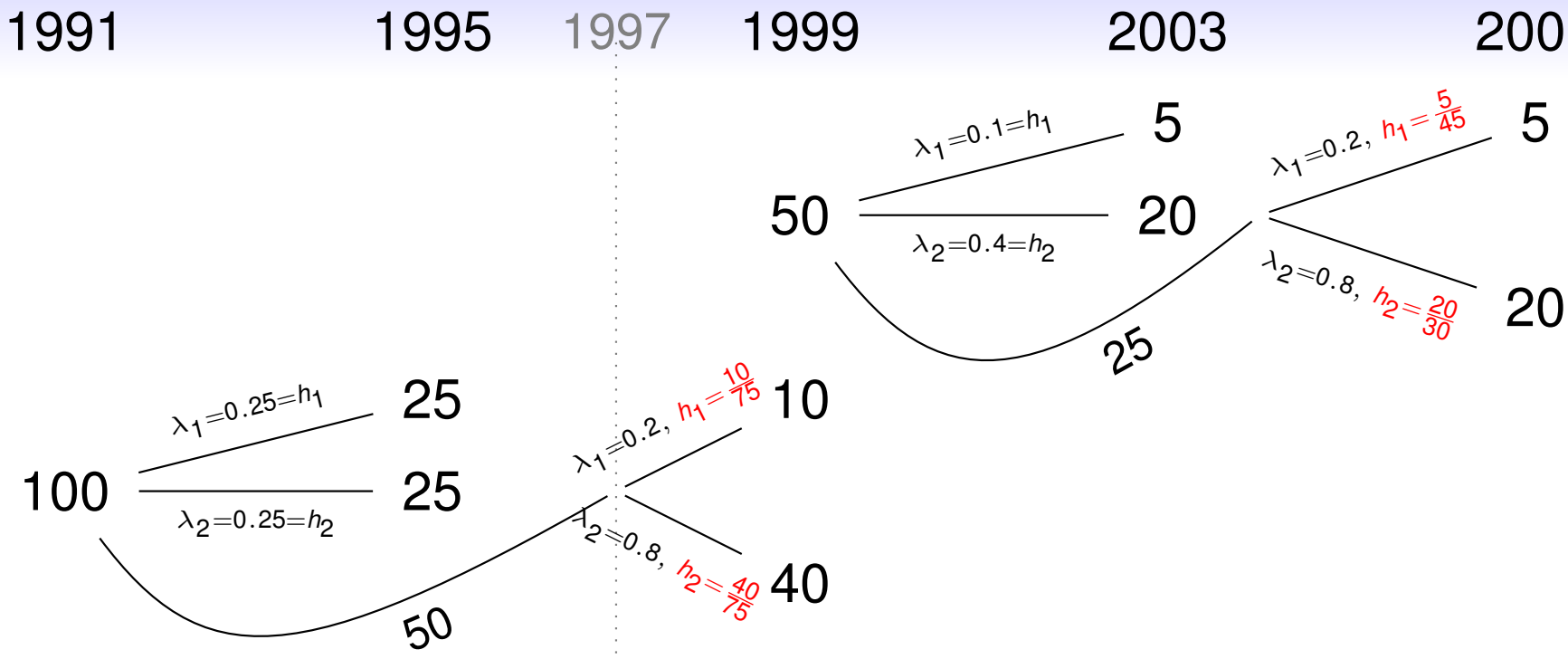
- Internal:

$$\widehat{h_1}(8|II)=\frac{5+10}{75+45}=\frac{15}{120}=\frac{1}{8}$$
$$\widehat{h_2}(8|II)=\frac{20+40}{75+30}=\frac{60}{105}=\frac{4}{7}$$

oucru

wellcome

UNIVERSITY OF OXFORD



- Pseudo-individual: $\widehat{\mathrm{P}}(V<6)=1-\frac{50}{75}=\frac{1}{3}$, $\widehat{\mathrm{P}}(V\leq 6)=1$

$$\widehat{h_1}(8|II)=\frac{5+10}{75+3\times 20}=\frac{15}{135}=\frac{1}{9}$$
$$\widehat{h_2}(8|II)=\frac{20+40}{75+3\times 5}=\frac{60}{90}=\frac{2}{3}$$

- Internal:

$$\widehat{h_1}(8|II)=\frac{5+10}{75+45}=\frac{15}{120}=\frac{1}{8}$$
$$\widehat{h_2}(8|II)=\frac{20+40}{75+30}=\frac{60}{105}=\frac{4}{7}$$

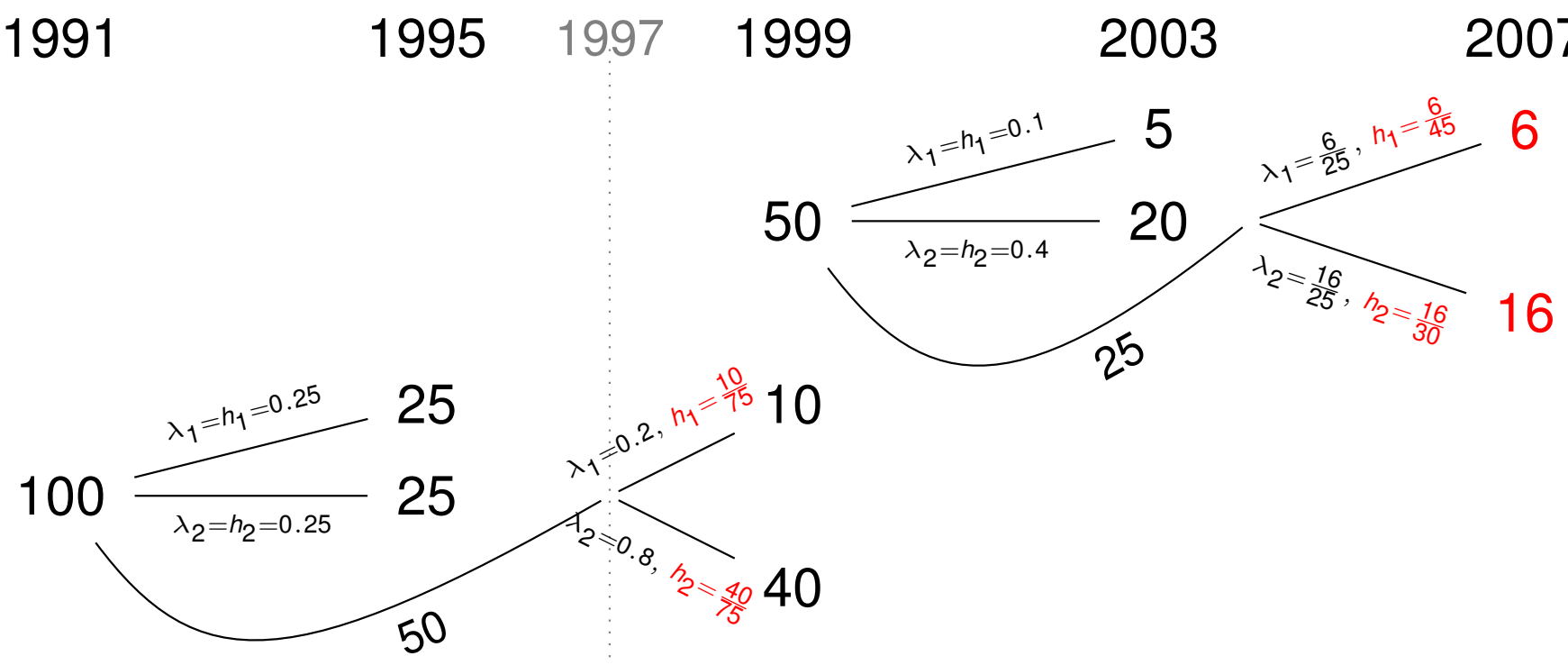
oucru

wellcome

UNIVERSITY OF OXFORD

Interpretation

- Internal approach
  - Assumption: all individuals $\geq$ 1997 same subdistribution hazard
  - Cause-specific hazards differ

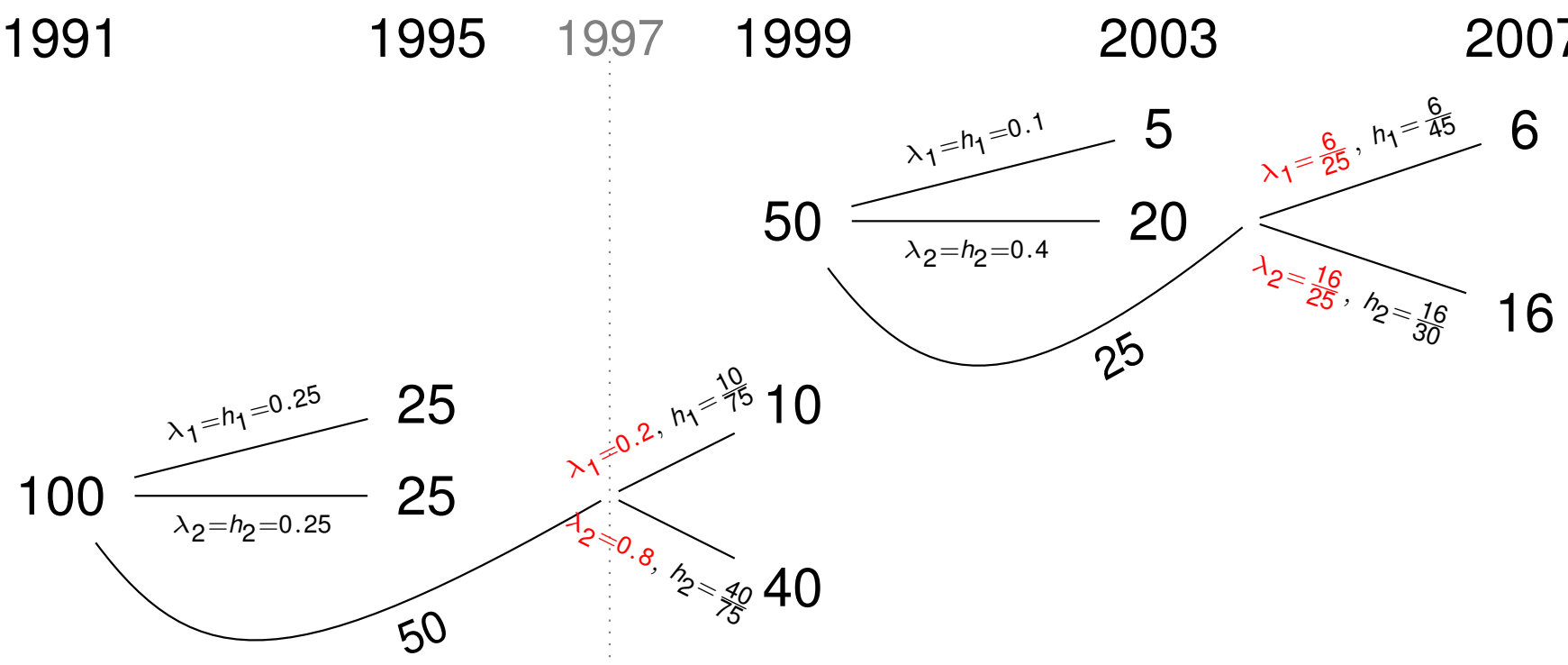


- Pseudo-individual: $\widehat{\mathrm{P}}(V<8)=\widehat{\mathrm{P}}(V\leq 6)=1-\frac{50}{75}=\frac{1}{3}$

$$\widehat{h_1}(8|II)=\frac{6+10}{75+3\times 20}=\frac{16}{135}$$
$$\widehat{h_2}(8|II)=\frac{16+40}{75+3\times 5}=\frac{56}{90}=\frac{28}{45}$$

- Internal:

$$\widehat{h_1}(8|II)=\frac{6+10}{45+75}=\frac{16}{120}=\frac{2}{15}$$
$$\widehat{h_2}(8|II)=\frac{16+40}{30+75}=\frac{56}{105}=\frac{8}{15}$$

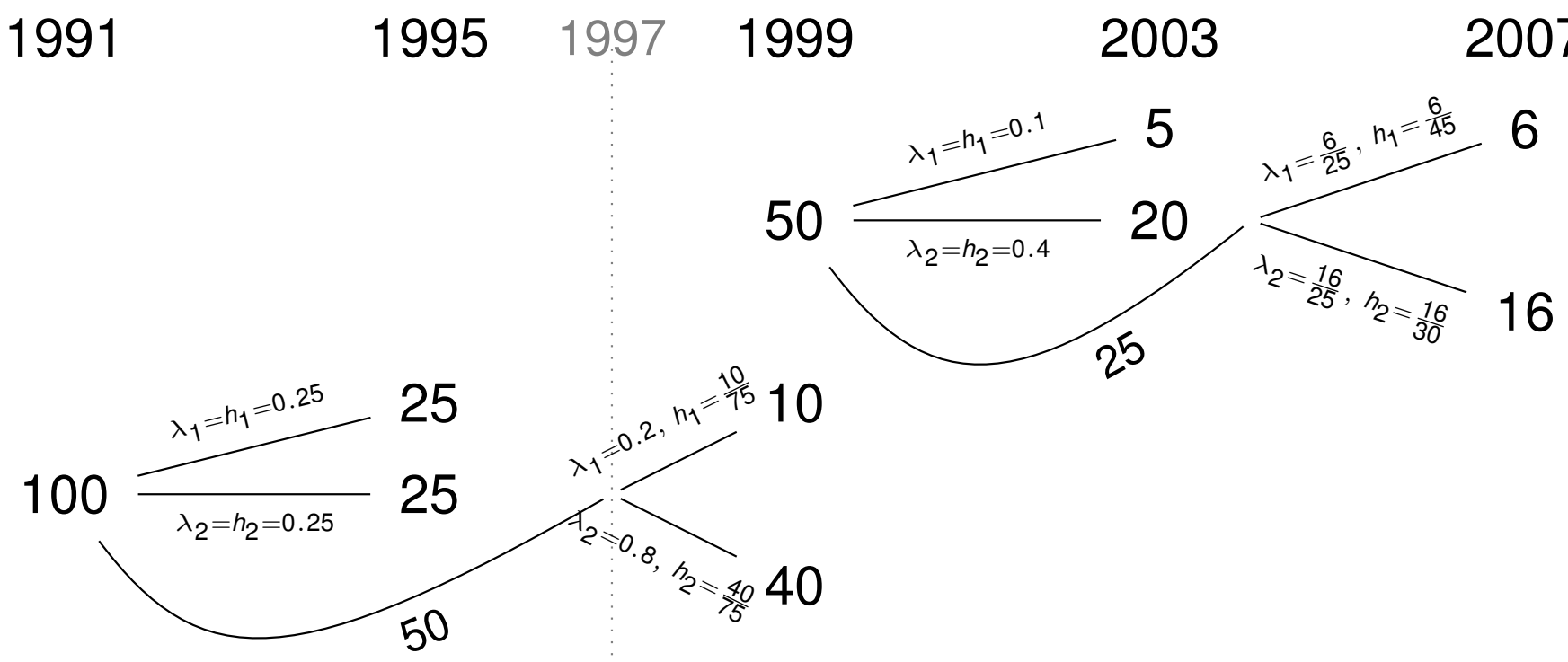


- Pseudo-individual: $\widehat{\mathrm{P}}(V<8)=\widehat{\mathrm{P}}(V\leq 6)=1-\frac{50}{75}=\frac{1}{3}$

$$\widehat{h_1}(8|II)=\frac{6+10}{75+3\times 20}=\frac{16}{135}$$
$$\widehat{h_2}(8|II)=\frac{16+40}{75+3\times 5}=\frac{56}{90}=\frac{28}{45}$$

- Internal:

$$\widehat{h_1}(8|II)=\frac{6+10}{45+75}=\frac{16}{120}=\frac{2}{15}$$
$$\widehat{h_2}(8|II)=\frac{16+40}{30+75}=\frac{56}{105}=\frac{8}{15}$$

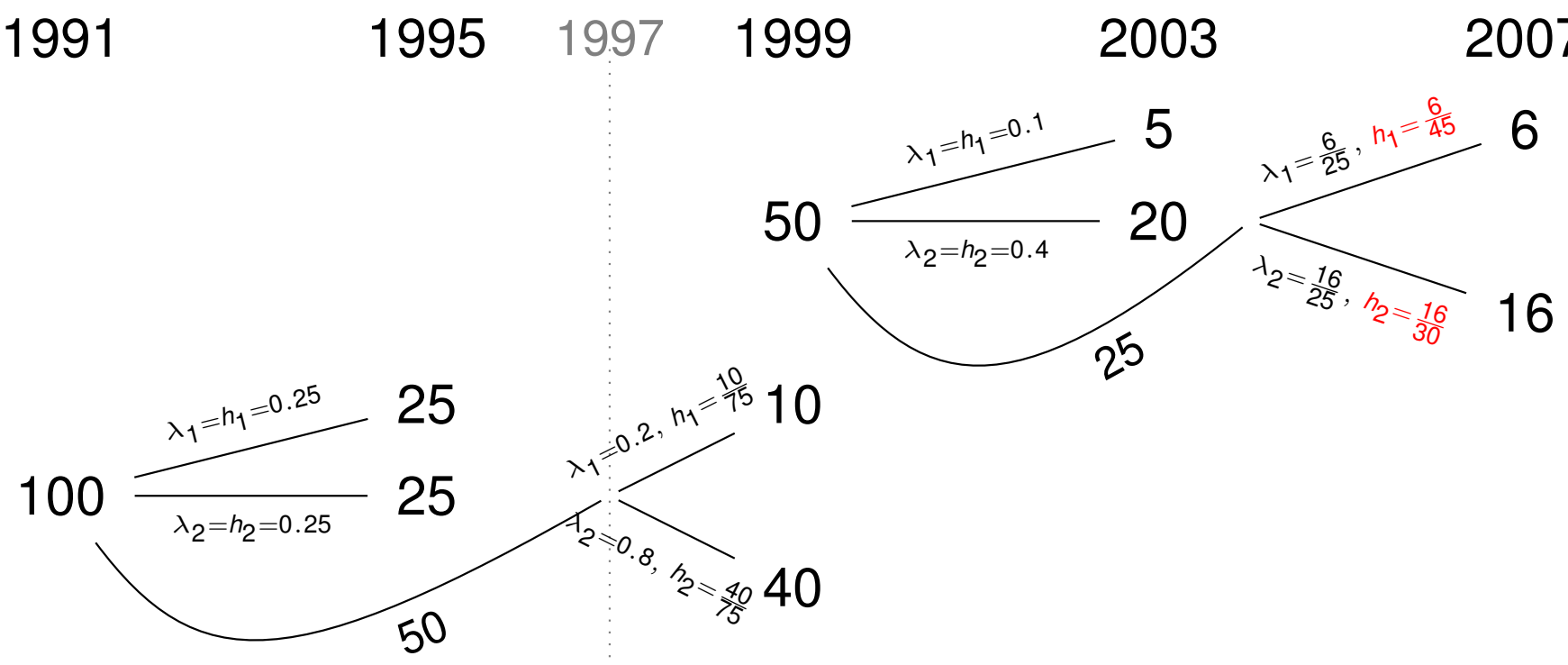


- Pseudo-individual: $\widehat{\mathrm{P}}(V<8)=\widehat{\mathrm{P}}(V\leq 6)=1-\frac{50}{75}=\frac{1}{3}$

$$\widehat{h}_1(8|II)=\frac{6+10}{75+3\times 20}=\frac{16}{135}$$
$$\widehat{h}_2(8|II)=\frac{16+40}{75+3\times 5}=\frac{56}{90}=\frac{28}{45}$$

- Internal:

$$\widehat{h}_1(8|II)=\frac{6+10}{45+75}=\frac{16}{120}=\frac{2}{15}$$
$$\widehat{h}_2(8|II)=\frac{16+40}{30+75}=\frac{56}{105}=\frac{8}{15}$$



- Pseudo-individual: $\widehat{\mathrm{P}}(V<8)=\widehat{\mathrm{P}}(V\leq 6)=1-\frac{50}{75}=\frac{1}{3}$

$$\widehat{h}_1(8|II)=\frac{6+10}{75+3\times 20}=\frac{16}{135}$$
$$\widehat{h}_2(8|II)=\frac{16+40}{75+3\times 5}=\frac{56}{90}=\frac{28}{45}$$

- Internal:

$$\widehat{h}_1(8|II)=\frac{6+10}{45+75}=\frac{16}{120}=\frac{2}{15}$$
$$\widehat{h}_2(8|II)=\frac{16+40}{30+75}=\frac{56}{105}=\frac{8}{15}$$

Time-varying covariables

- Effect of $\mathbf{Z}(t)$ on subdistribution hazard depends on
  - Effect of $\mathbf{Z}(t)$ on other event types
  - History of $\mathbf{Z}(t)$

Time-varying covariables

- Effect of $\mathbf{Z}(t)$ on subdistribution hazard depends on
  - Effect of $\mathbf{Z}(t)$ on other event types
  - History of $\mathbf{Z}(t)$
- Pseudo-individual approach
  - Covariables per pseudo-individual all time-fixed
  - Aligns cause-specific hazards
  - Quantifies subdistribution hazard for constant value of $\mathbf{Z}(t)$

Time-varying covariables

- Effect of $\mathbf{Z}(t)$ on subdistribution hazard depends on
  - Effect of $\mathbf{Z}(t)$ on other event types
  - History of $\mathbf{Z}(t)$
- Pseudo-individual approach
  - Covariables per pseudo-individual all time-fixed
  - Aligns cause-specific hazards
  - Quantifies subdistribution hazard for constant value of $\mathbf{Z}(t)$
- Internal approach
  - Aligns subdistribution hazards
  - Problematic for internal covariables
    - Cause-specific hazards differ $\rightarrow$ no relation with etiology
    - Value of covariable after competing event required


Per Kragh Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding.
Statistical Models Based on Counting Processes.
Springer Verlag, New York, 1993.


Ronald B. Geskus.
Data analysis with competing risks and intermediate states.
CRC Press, Boca Raton, 2015.


J. D. Kalbfleisch and R. L. Prentice.
The Statistical Analysis of Failure Time Data.
Second Edition. Wiley, New York, 2002.


Beyersmann, J. and Schumacher, M.

Time-dependent covariates in the proportional subdistribution hazards model for competing risks.
Biostatistics, 9(4):765–776, 2008.


Deslandes, E. and Chevret, S.

Joint modeling of multivariate longitudinal data and the dropout process in a competing risk setting: application to ICU data.
BMC Medical Research Methodology, 10:69, 2010.
http://www.biomedcentral.com/1471-2288/10/69


Ronald B. Geskus.

Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring.
Biometrics, 67(1):39–49, 2011.


J. van der Helm, R. Geskus, C. Sabin et al..

Effect of HCV infection on cause-specific mortality after HIV seroconversion, before and after 1997.
Gastroenterology, 144(4):751–760, 2013.


A. Latouche, R. Porcher and S. Chevret.

A note on including time-dependent covariate in regression model for competing risks data.
Biometrical Journal, 47(6):807–814, 2005.