Cohort heterogeneity and competing risks in survival analysis

ACC Coolen
Radboud University

Cohort heterogeneity and competing risks

- Regression for time-to-event data
- Consequences and fingerprints
- Informative censoring: the intiuition

Bayesian latent class models

- Rationale and definition
- Tests on synthetic data

Applications in cancer research

- Epidemiological cancer data
- Data from failed cancer trials

New directions

- Prospective latent class prediction
- Overfitting in multivariate survival analysis

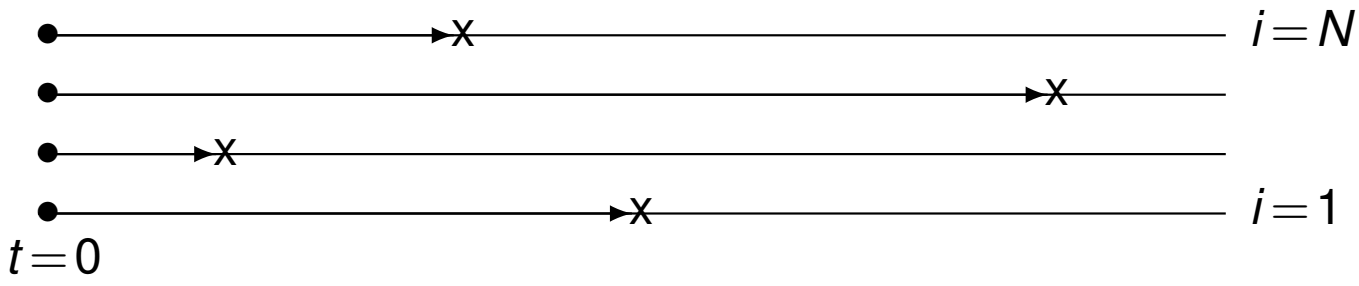
Regression for time-to-event data

- *Data* $D = \{(\boldsymbol{z}_1, t_1, r_1), \ldots, (\boldsymbol{z}_N, t_N, r_N)\}$

$\boldsymbol{z}_i = (z_{i1}, \ldots, z_{id})$: *d covariates* (*measured at* $t = 0$)

$t_i > 0$: *first failure time* (*death, onset of disease, ...*)

$r_i \in \{0, 1, \ldots, R\}$: *failure type* (*or 'risk'*)



- *Heterogeneity*

visible: variability in the available covariates

latent: variability in host or disease, not visible in the covariates
(individuals with same covariates $\boldsymbol{z}$ are not clones ...)

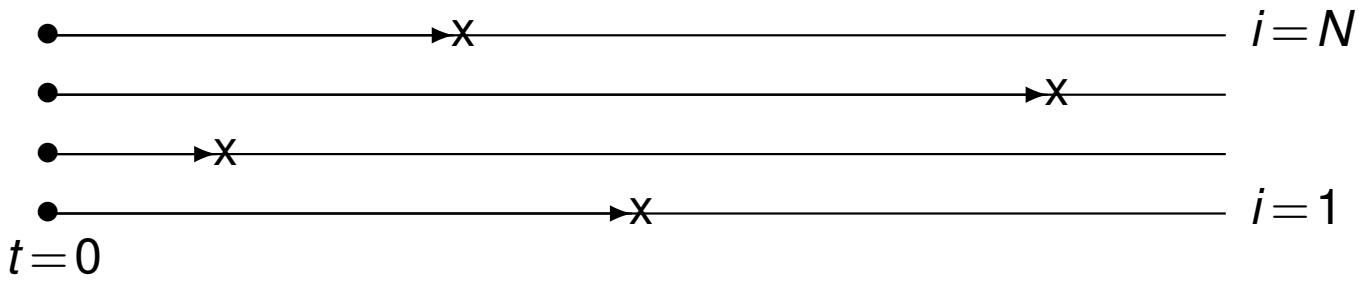
Regression for time-to-event data

- *Data* $D = \{(\boldsymbol{z}_1, t_1, r_1), \ldots, (\boldsymbol{z}_N, t_N, r_N)\}$

$\boldsymbol{z}_i = (z_{i1}, \ldots, z_{id})$: *d covariates* (*measured at* $t = 0$)

$t_i > 0$: *first failure time* (*death, onset of disease, ...*)

$r_i \in \{0, 1, \ldots, R\}$: *failure type* (*or 'risk'*)



- *Heterogeneity*

visible: variability in the available covariates

latent: variability in host or disease, not visible in the covariates
(individuals with same covariates $\boldsymbol{z}$ are not clones ...)

Competing risks, identifiability and interpretation

- *Competing risks*

  Informative censoring, i.e. event times of risks are *statistically dependent*: $p(t_1, \ldots, t_R|\boldsymbol{z}) \neq \prod_{r=1}^{R} p(t_r|\boldsymbol{z})$

  reported time: $t = \min\{t_1, \ldots, t_R\}$

- *Interpretation of crude hazard rates*

  Eliminating one risk can change hazard rate of others ...

  if hazard rate for risk 1 is low:

  (i) event 1 is intrinsically unlikely?
  (ii) or it is often preceded by event 2?

  to disentangle risks: need $p(t_1, \ldots, t_R|\boldsymbol{z})$

- *Tsiatis' identifiability problem (1975)*

  Joint event time distribution $p(t_1, \ldots, t_R|\boldsymbol{z})$ cannot be inferred from survival data alone ...

Competing risks, identifiability and interpretation

- *Competing risks*

  Informative censoring, i.e. event times of risks are *statistically dependent*: $p(t_1, \ldots, t_R|\boldsymbol{z}) \neq \prod_{r=1}^{R} p(t_r|\boldsymbol{z})$

  reported time: $t = \min\{t_1, \ldots, t_R\}$

- *Interpretation of crude hazard rates*

  Eliminating one risk can change hazard rate of others ...

  if hazard rate for risk 1 is low:

  (i) event 1 is intrinsically unlikely?
  (ii) or it is often preceded by event 2?

  to disentangle risks: need $p(t_1, \ldots, t_R|\boldsymbol{z})$

- *Tsiatis' identifiability problem (1975)*

  Joint event time distribution $p(t_1, \ldots, t_R|\boldsymbol{z})$ cannot be inferred from survival data alone ...

Competing risks, identifiability and interpretation

- *Competing risks*

  Informative censoring, i.e. event times of risks are *statistically dependent*: $p(t_1, \ldots, t_R|\boldsymbol{z}) \neq \prod_{r=1}^{R} p(t_r|\boldsymbol{z})$

  reported time: $t = \min\{t_1, \ldots, t_R\}$

- *Interpretation of crude hazard rates*

  Eliminating one risk can change hazard rate of others ...

  if hazard rate for risk 1 is low:

  (i) event 1 is intrinsically unlikely?
  (ii) or it is often preceded by event 2?

  to disentangle risks: need $p(t_1, \ldots, t_R|\boldsymbol{z})$

- *Tsiatis' identifiability problem (1975)*

  Joint event time distribution $p(t_1, \ldots, t_R|\boldsymbol{z})$ cannot be inferred from survival data alone ...

Conventional methods

for analysing time-to-event data

Kaplan-Meier estimators
Cox regression
.....

- not designed to handle disease/host heterogeneity, beyond variability in covariates

- to allow interpretation:

  have to assume different risks are uncorrelated, dangerous when many censoring events ... (older populations!)

random effects models, frailty models, latent class models

- usually constructed for primary risk only, so still cannot handle correlated risks

- do not exploit the link between latent heterogeneity and competing risks ...

Conventional methods

for analysing time-to-event data

Kaplan-Meier estimators
Cox regression
.....

- not designed to handle disease/host heterogeneity, beyond variability in covariates

- to allow interpretation:

  have to assume different risks are uncorrelated, dangerous when many censoring events ... (older populations!)

random effects models, frailty models,
latent class models

- usually constructed for primary risk only, so still cannot handle correlated risks

- do not exploit the link between latent heterogeneity and competing risks ...

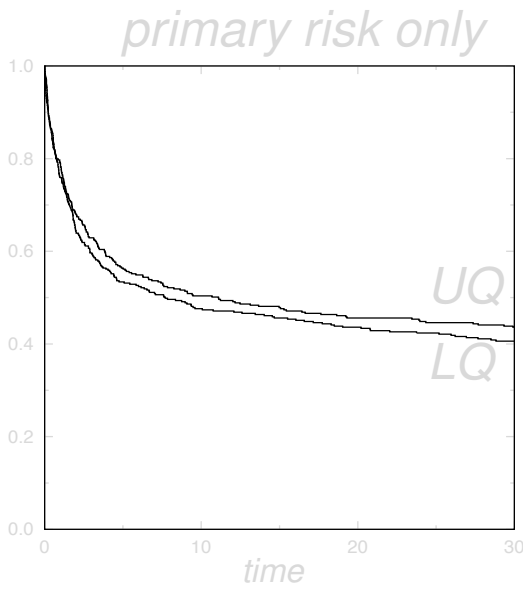
Consequences and fingerprints of latent heterogeneity

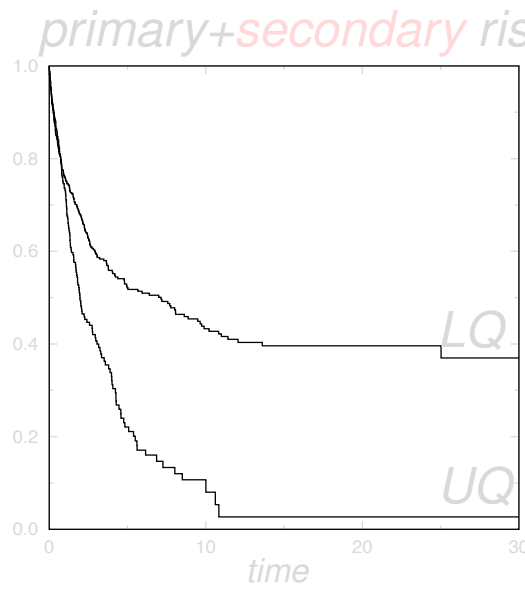
- *Violation of proportional hazards assumption*
- *Interpretation of time dependencies tricky*

  even if all *individual* hazard rates h_i are time-independent, cohort hazard rate will be time-dependent:

$$h(t) = \frac{\sum_{i=1}^{n} h_i e^{-h_i t}}{\sum_{i=1}^{n} e^{-h_i t}}$$

- *Interpreting cause-specific survival curves (KM, Cox) no longer possible ...*





Consequences and fingerprints of latent heterogeneity

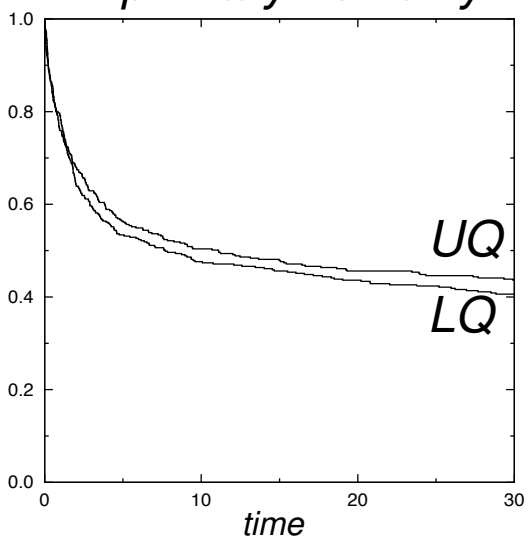
- *Violation of proportional hazards assumption*
- *Interpretation of time dependencies tricky*

  even if all *individual* hazard rates h_i are time-independent, cohort hazard rate will be time-dependent:

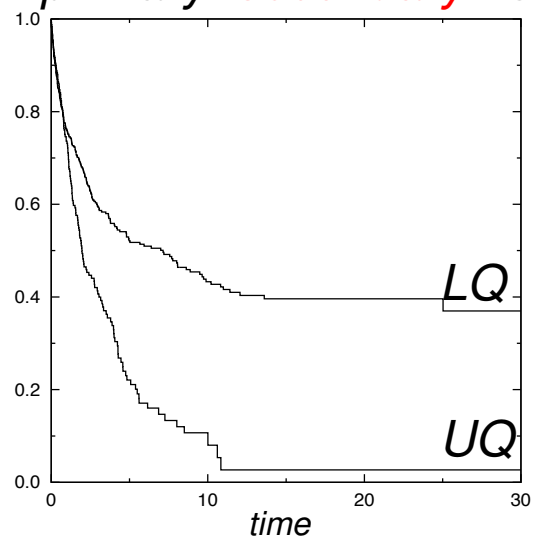
$$h(t) = \frac{\sum_{i=1}^{n} h_i e^{-h_i t}}{\sum_{i=1}^{n} e^{-h_i t}}$$

- *Interpreting cause-specific survival curves (KM, Cox) no longer possible ...*

primary risk only



primary+secondary risk



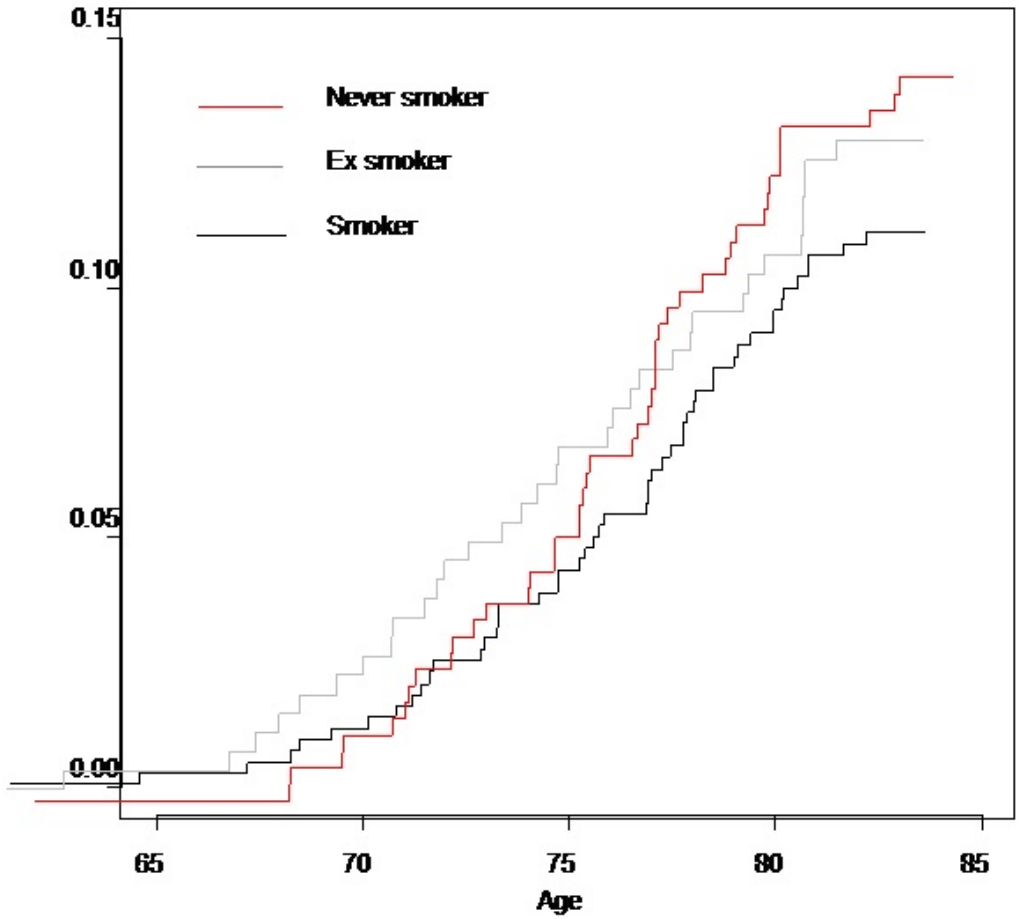
If in interpreting our data we assume
censoring risks <u>uncorrelated</u> with primary risk

censoring by competing risks
can give nonsensical results ...

– harmful drugs look beneficial
– beneficial drugs look harmful

– false protectivity of covariates
– false aetiology of covariates

(ULSAM prostate cancer data)



would we have spotted this
if the covariate represented
the expression of a specific gene?

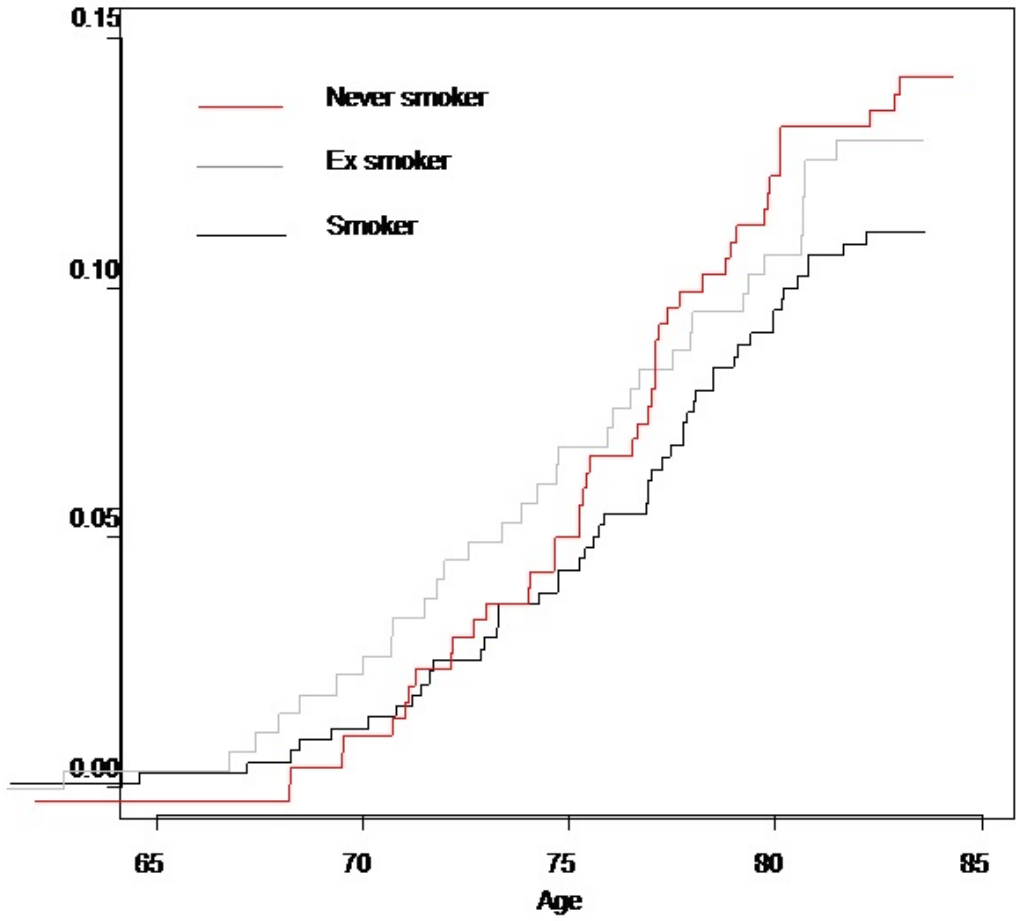
If in interpreting our data we assume
censoring risks uncorrelated with primary risk

censoring by competing risks
can give nonsensical results ...

– harmful drugs look beneficial
– beneficial drugs look harmful

– false protectivity of covariates
– false aetiology of covariates

(ULSAM prostate cancer data)



would we have spotted this
if the covariate represented
the expression of a specific gene?

Link between cohort heterogeneity and informative censoring

Say 1000 people,
two risks, hazard rates h_A and h_B

- homogeneous cohort:
  all *individuals* have (h_A, h_B)

h_A, h_B

1000

- heterogeneous cohort,
  but *non-informative censoring*

$h_A\uparrow\ h_B\uparrow$ *250*

$h_A\downarrow\ h_B\uparrow$ *250*

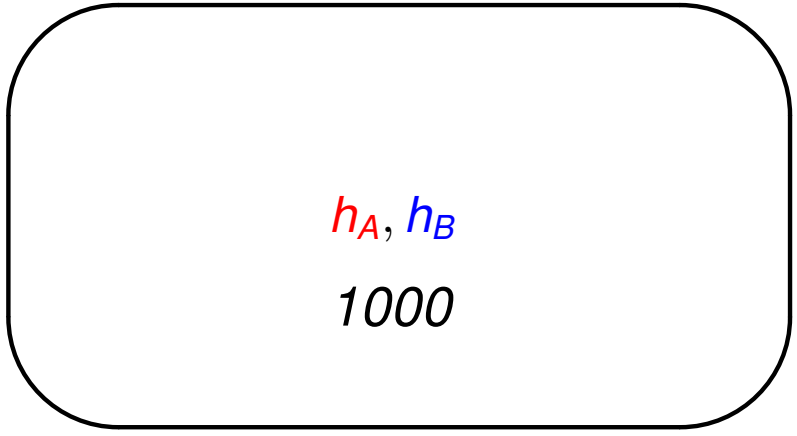
$h_A\uparrow\ h_B\downarrow$ *250*

$h_A\downarrow\ h_B\downarrow$ *250*

Link between cohort heterogeneity and informative censoring

Say 1000 people,
two risks, hazard rates h_A and h_B

- homogeneous cohort:
  all *individuals* have (h_A, h_B)

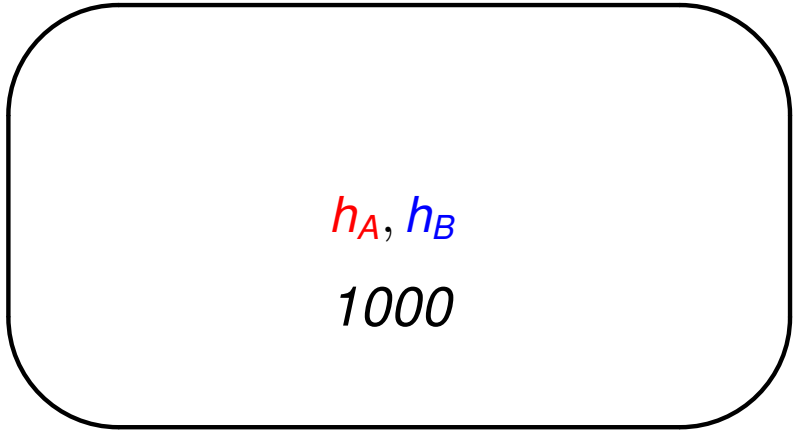


- heterogeneous cohort,
  but *non-informative censoring*

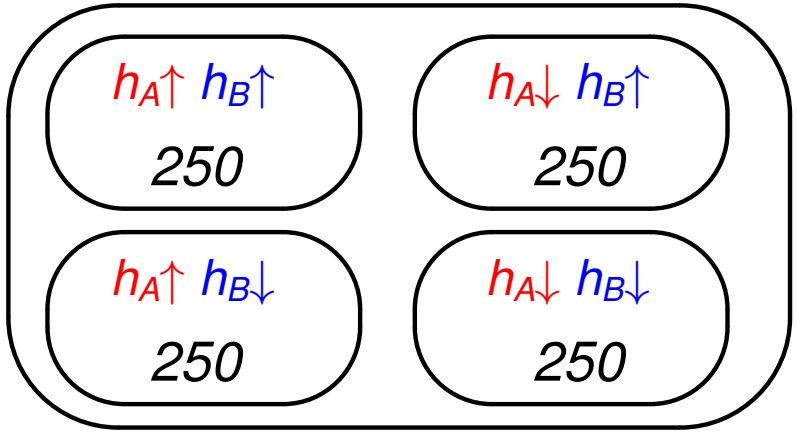
Link between cohort heterogeneity and informative censoring

Say 1000 people,
two risks, hazard rates h_A and h_B

- homogeneous cohort:
  all *individuals* have (h_A, h_B)



- heterogeneous cohort,
  but *non-informative censoring*



Heterogeneity and informative censoring

Say 1000 people,
two risks, hazard rates h_A and h_B

- homogeneous cohort:
  all *individuals* have (h_A, h_B)

h_A, h_B

1000

- heterogeneous cohort,
  informative cohort filtering

  result:
  underestimation of h_A

$h_A\uparrow\ h_B\uparrow$ *480*

$h_A\downarrow\ h_B\uparrow$ *20*

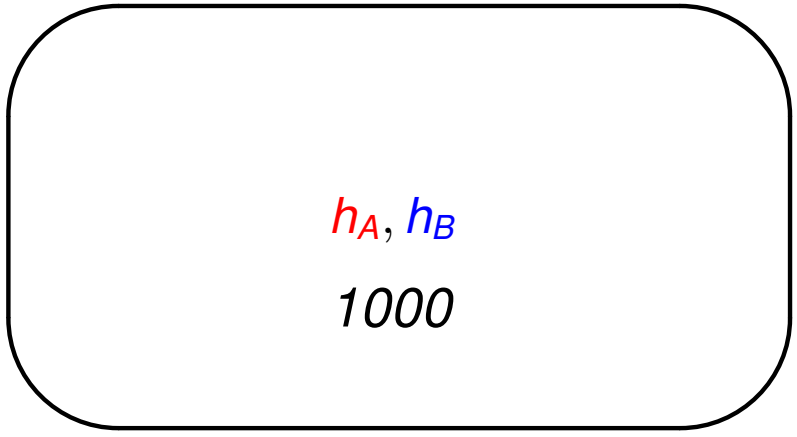
$h_A\uparrow\ h_B\downarrow$ *20*

$h_A\downarrow\ h_B\downarrow$ *480*

Heterogeneity and informative censoring

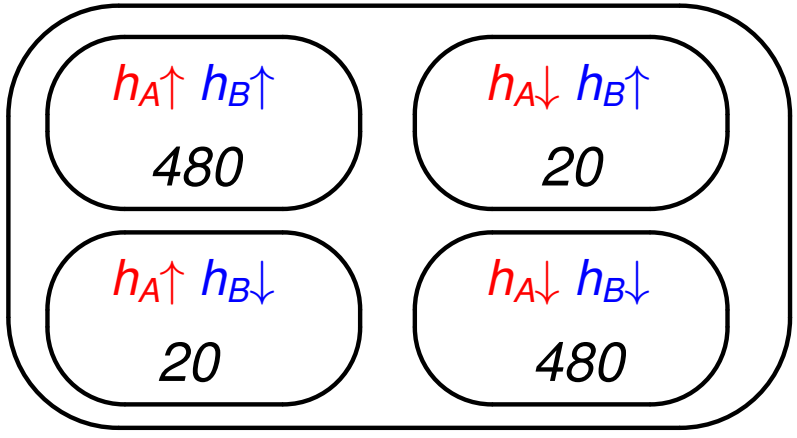
Say 1000 people,
two risks, hazard rates h_A and h_B

- homogeneous cohort:
  all *individuals* have (h_A, h_B)



- heterogeneous cohort,
  informative cohort filtering

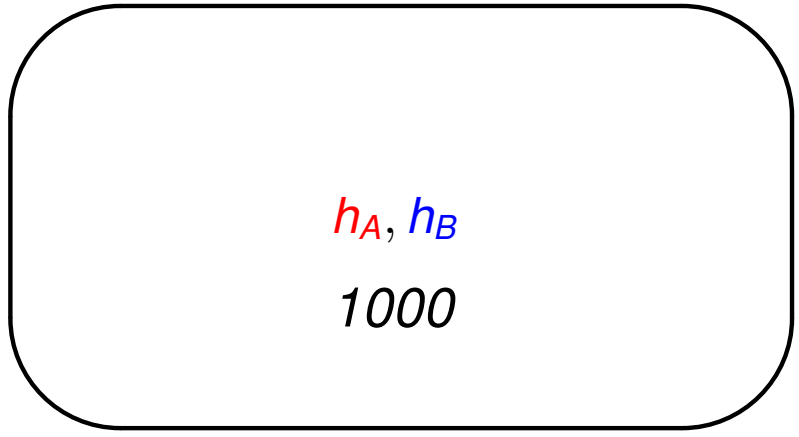
  result:
  underestimation of h_A



Heterogeneity and informative censoring

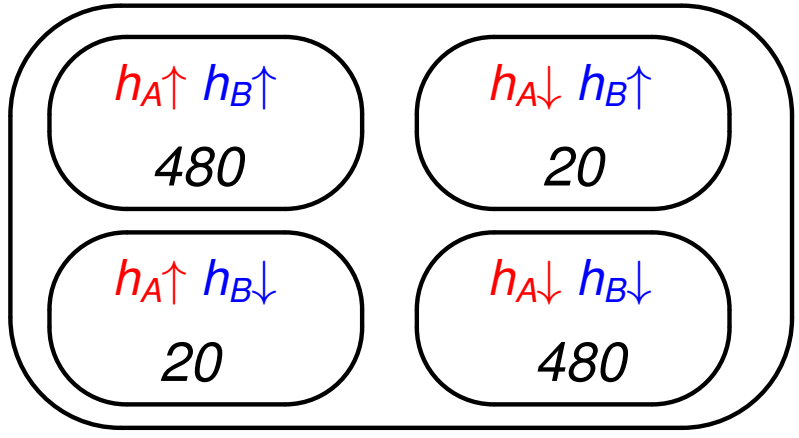
Say 1000 people,
two risks, hazard rates h_A and h_B

- homogeneous cohort:
  all *individuals* have (h_A, h_B)



- heterogeneous cohort,
  informative cohort filtering

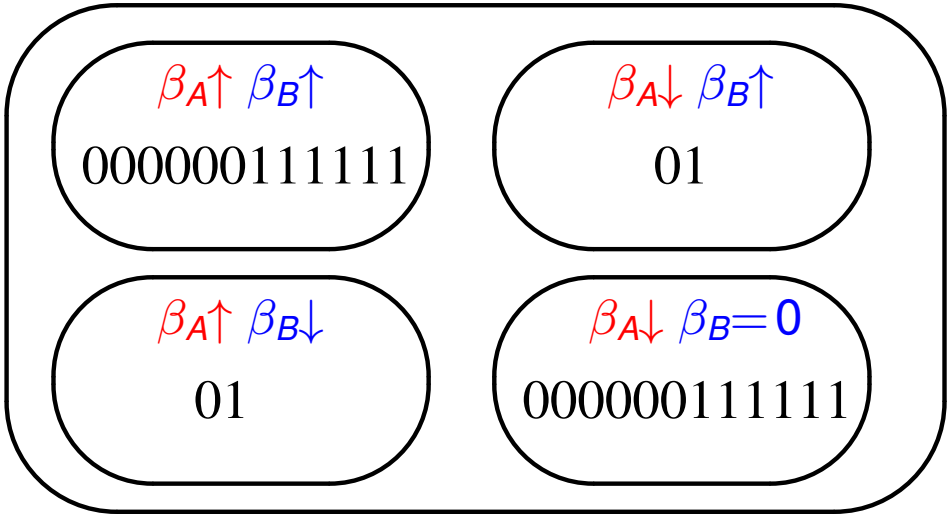
  result:
  underestimation of h_A



Heterogeneity and informative censoring

Say 28 people,
binary covariate: $z = 0,1$

association risk A: β_A
association risk B: β_B
(B: competing risk, strong)



without risk B:

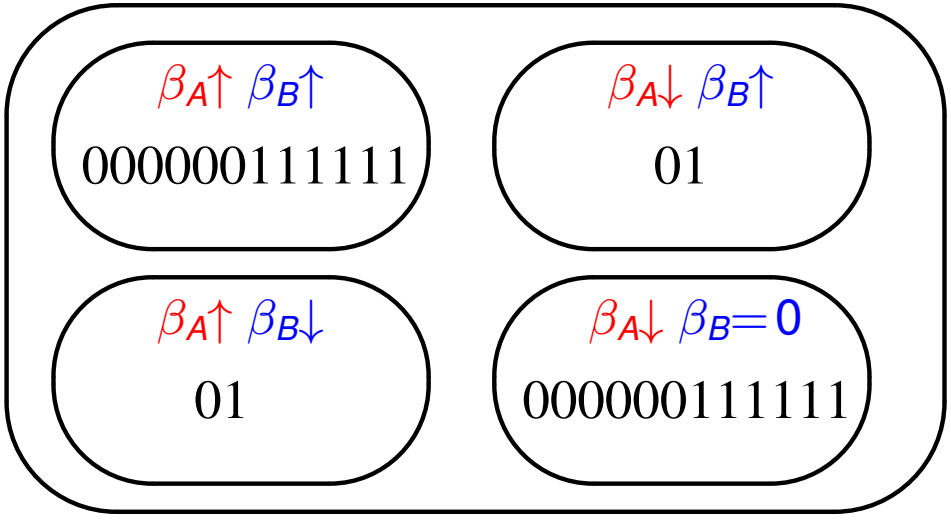
as many A deaths with $z = 0$ as for $z = 1$,

overall association $\beta_A = 0$

Heterogeneity and informative censoring

Say 28 people,
binary covariate: $z = 0,1$

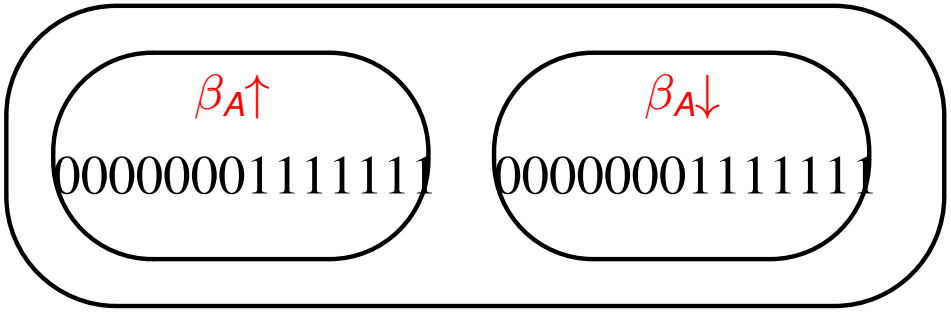
association risk A: β_A
association risk B: β_B
(B: competing risk, strong)



without risk B:

as many A deaths with $z = 0$ as for $z = 1$,

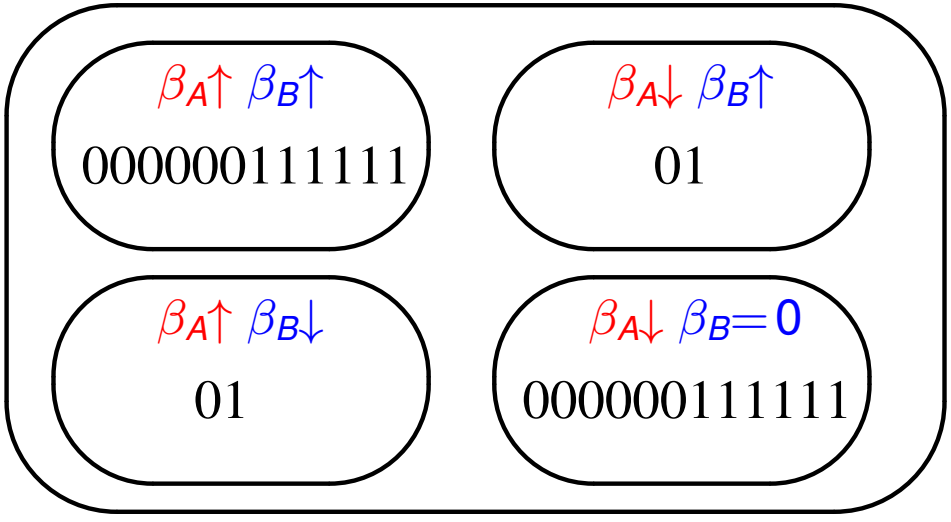
overall association $\beta_A = 0$



Heterogeneity and informative censoring

Say 28 people,
binary covariate: $z = 0,1$

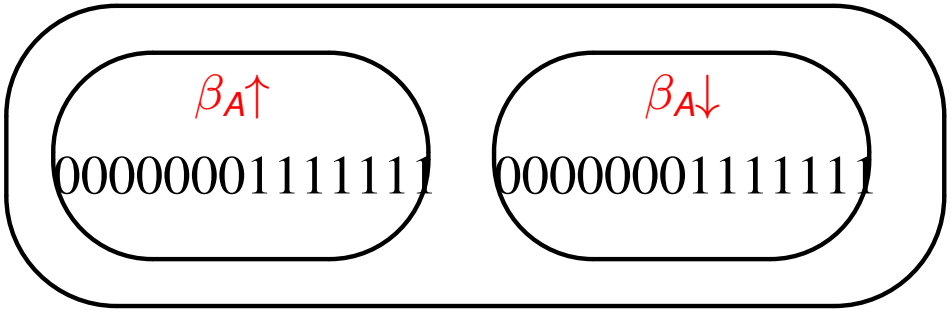
association risk A: β_A
association risk B: β_B
(B: competing risk, strong)



without risk B:

as many A deaths with
$z = 0$ as for $z = 1$,

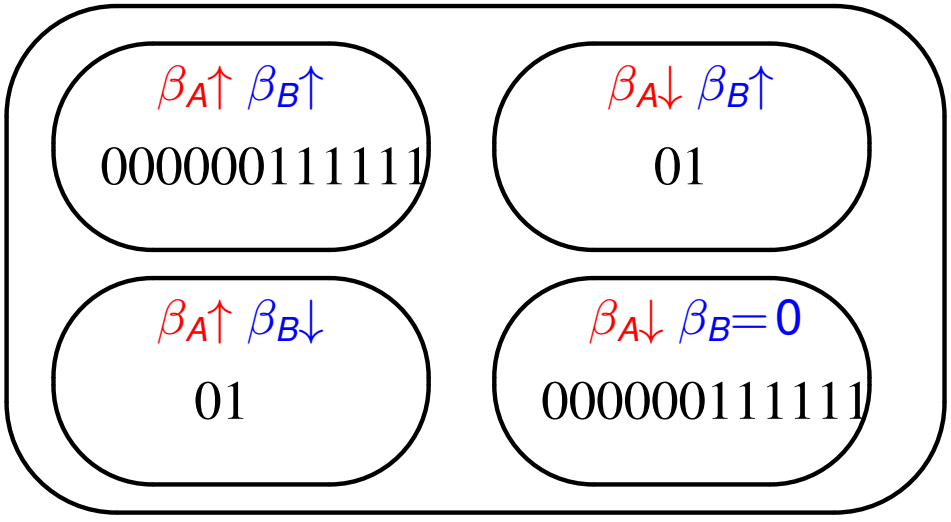
overall association $\beta_A = 0$



Heterogeneity and informative censoring

Say 28 people,
binary covariate: $z = 0,1$

associations risk A: β_A
associations risk B: β_B
(B: competing risk, strong)



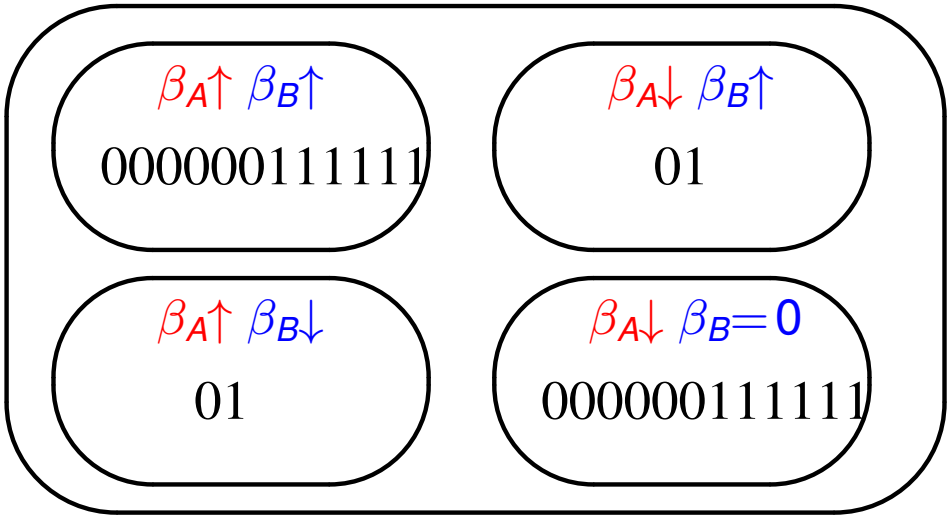
Effect of risk B:

what will we now observe for risk A?

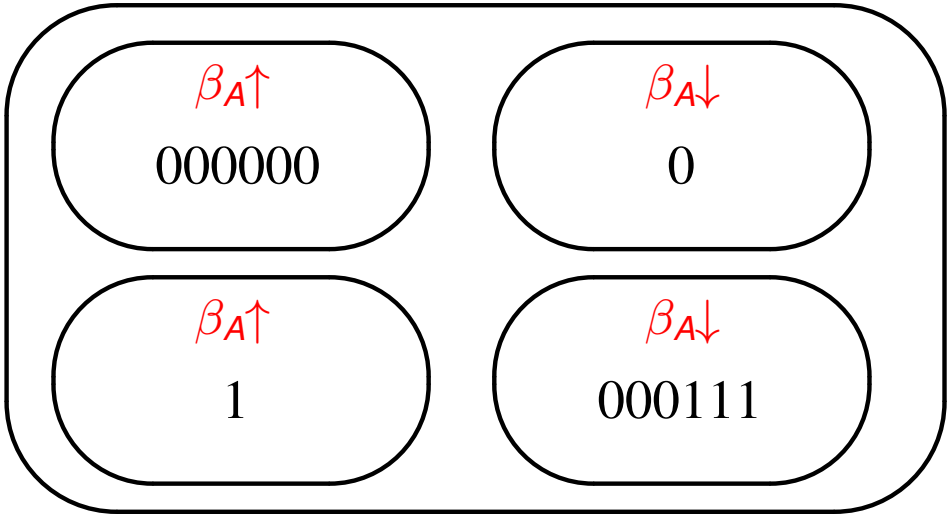
Heterogeneity and informative censoring

Say 28 people,
binary covariate: $z = 0,1$

associations risk A: β_A
associations risk B: β_B
(B: competing risk, strong)



Effect of risk B:

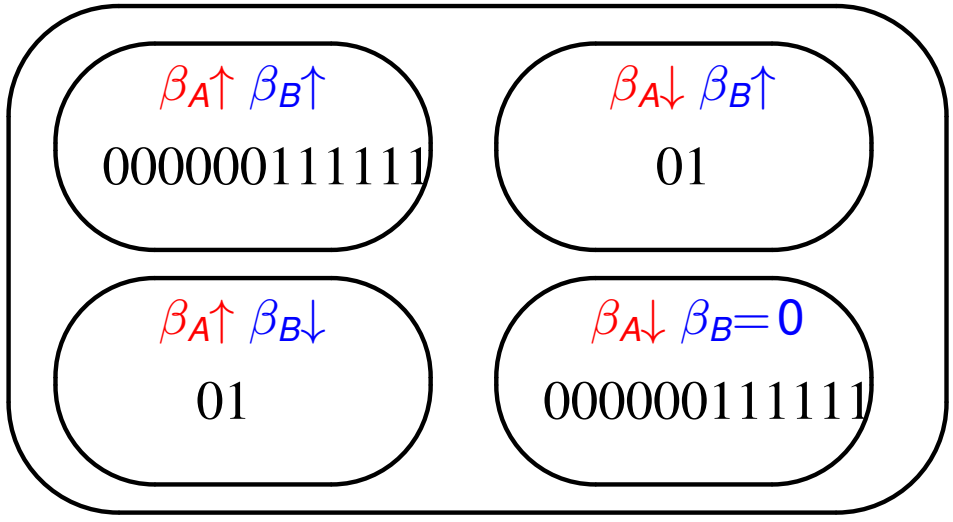


what will we now observe for risk A?

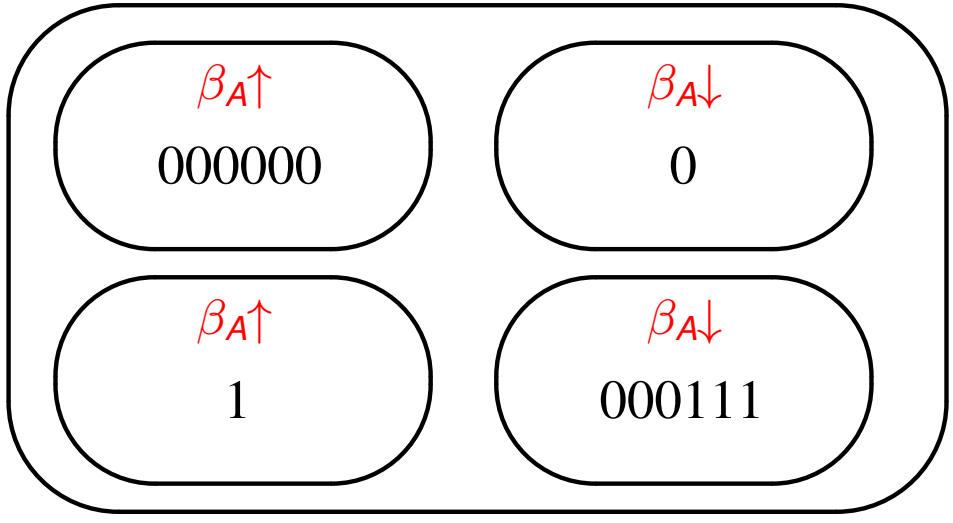
Heterogeneity and informative censoring

Say 28 people,
binary covariate: $z = 0,1$

associations risk A: β_A
associations risk B: β_B
(B: competing risk, strong)



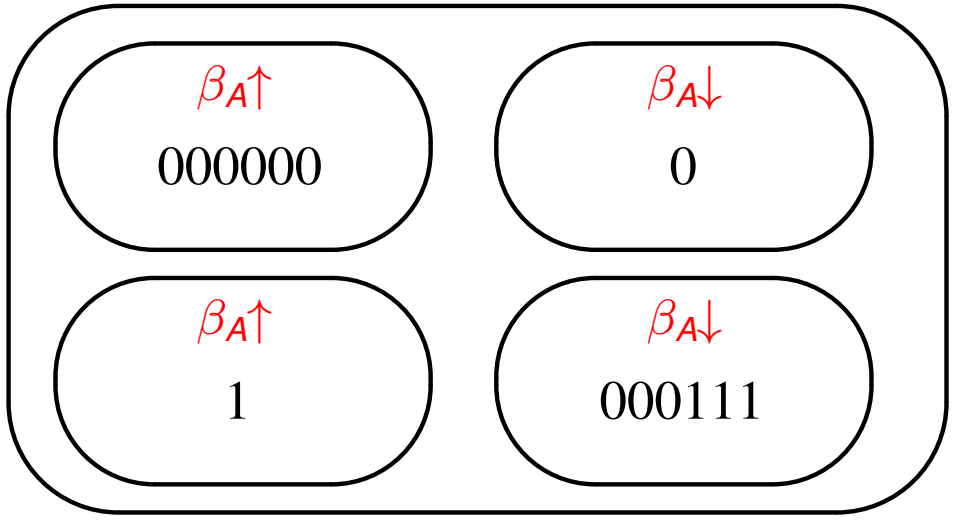
Effect of risk B:



what will we now observe for risk A?

Heterogeneity and informative censoring

what will we now observe for risk A?



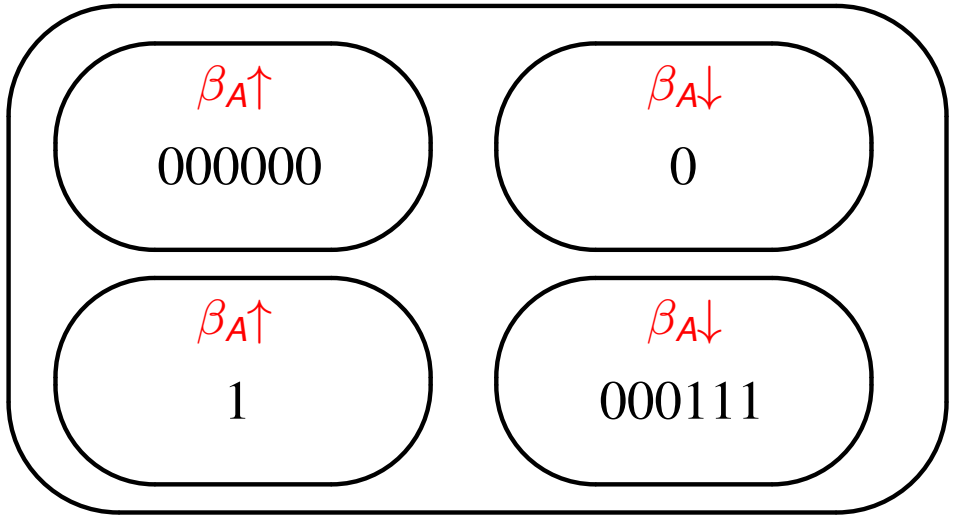
A survivors with $z=0$: 6
A survivors with $z=1$: 3

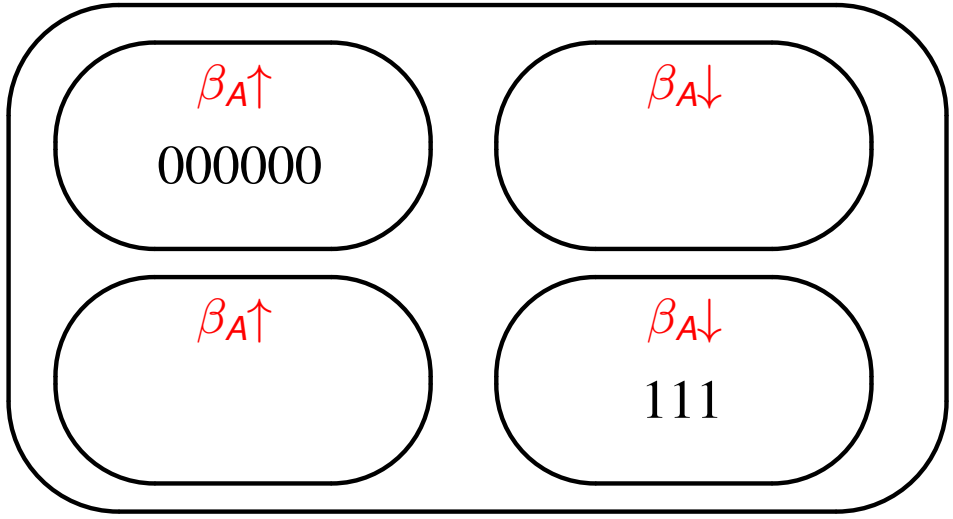
overall association $\beta_A > 0$,

false aetiology

Heterogeneity and informative censoring

what will we now observe for risk A?





A survivors with $z=0$: 6
A survivors with $z=1$: 3

overall association $\beta_A > 0$,

false aetiology

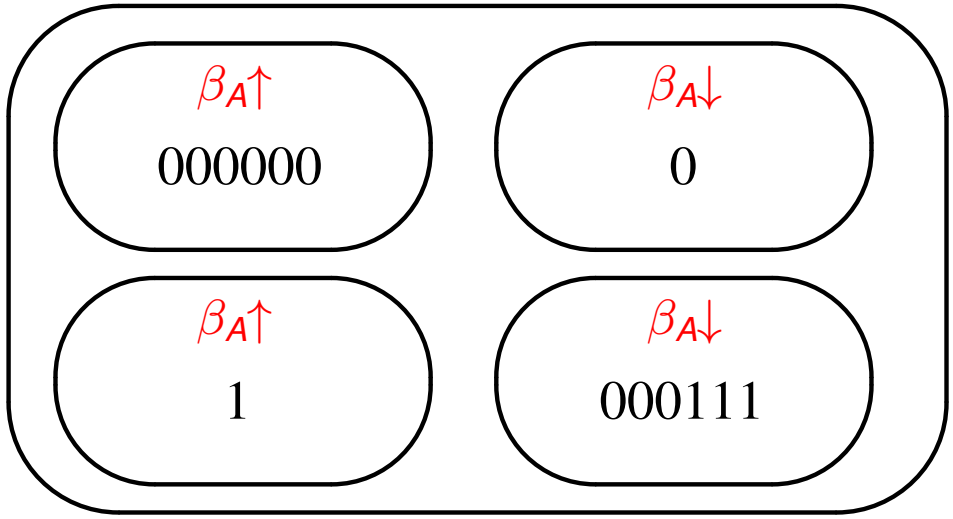
Heterogeneity and informative censoring

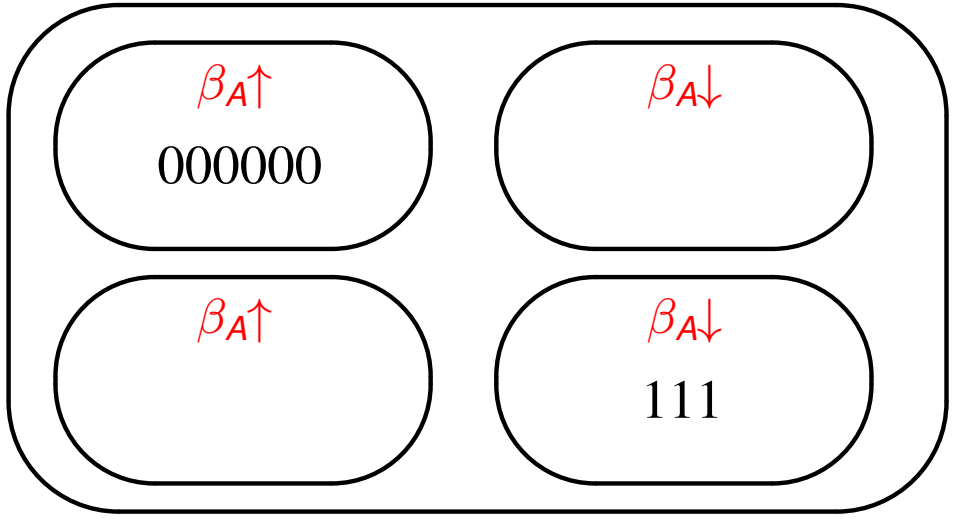
what will we now observe for risk A?

A survivors with $z=0$: 6
A survivors with $z=1$: 3

overall association $\beta_A > 0$,

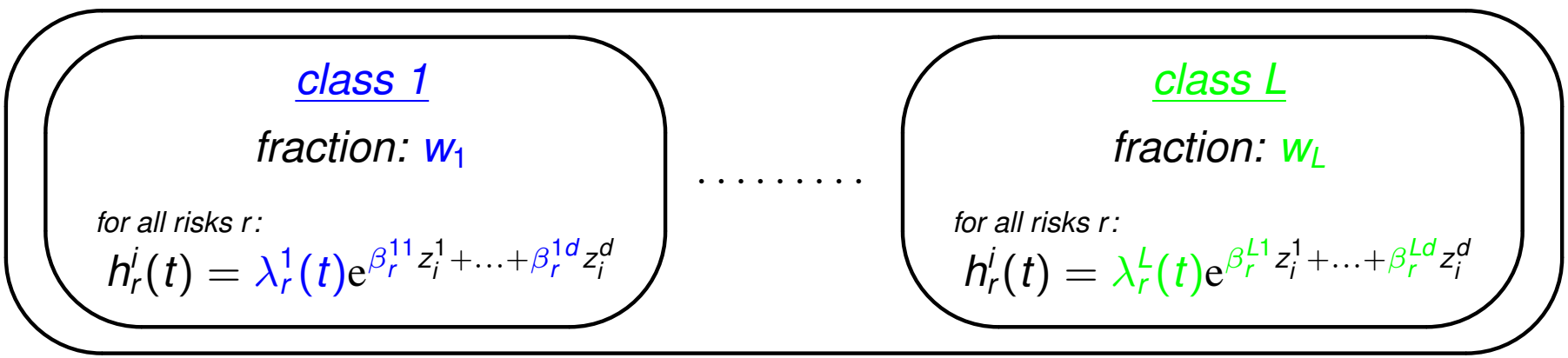
false aetiology





Bayesian latent class methods: rationale and definition

- model all risks simultaneously
- individuals with *same* covariates can have *distinct* associations and *distinct* base hazard rates
- risks are assumed independent only *at the level of individuals* (this removes Tsiatis' identifiability problem)
- competing risks, informative censoring: reflect correlated association parameters of different risks



proportional hazards within classes $\not\Longrightarrow$ *proportional hazards at cohort level*

independent risks within classes $\not\Longrightarrow$ *independent risks at cohort level*

Personalised cause-specific hazard rate model variants		
$M = 1$	Heterogeneous frailties Homogeneous associations Homogeneous base hazard rates	$h_r^i(t) = \lambda_r(t) \mathrm{e}^{\beta_r^{\ell 0} + \sum_\mu \beta_r^\mu z_i^\mu}$
$M = 2$	Heterogeneous frailties Heterogeneous associations Homogeneous base hazard rates	$h_r^i(t) = \lambda_r(t) \mathrm{e}^{\beta_r^{\ell 0} + \sum_\mu \beta_r^{\ell\mu} z_i^\mu}$
$M = 3$	Heterogeneous frailties Heterogeneous associations Heterogeneous base hazard rates	$h_r^i(t) = \lambda_r^\ell(t) \mathrm{e}^{\beta_r^{\ell 0} + \sum_\mu \beta_r^{\ell\mu} z_i^\mu}$

- Bayesian analysis and model selection:
  reliable error bars, and multiple classes *only if data demand it*
- reduces to standard Cox regression if no heterogeneity
  (Occam's Razor action of Bayesian model selection)
- non-primary events all contribute to latent class inference
- fully transparent interpretation,
  unlike some other competing risk approaches ...
- formulae for survival curves *decontaminated* for informative censoring,
  and *retrospective class allocation* of individuals

Personalised cause-specific hazard rate model variants		
$M = 1$	Heterogeneous frailties Homogeneous associations Homogeneous base hazard rates	$h_r^i(t) = \lambda_r(t) \mathrm{e}^{\beta_r^{\ell 0} + \sum_\mu \beta_r^\mu z_i^\mu}$
$M = 2$	Heterogeneous frailties Heterogeneous associations Homogeneous base hazard rates	$h_r^i(t) = \lambda_r(t) \mathrm{e}^{\beta_r^{\ell 0} + \sum_\mu \beta_r^{\ell\mu} z_i^\mu}$
$M = 3$	Heterogeneous frailties Heterogeneous associations Heterogeneous base hazard rates	$h_r^i(t) = \lambda_r^\ell(t) \mathrm{e}^{\beta_r^{\ell 0} + \sum_\mu \beta_r^{\ell\mu} z_i^\mu}$

- Bayesian analysis and model selection:
  reliable error bars, and multiple classes *only if data demand it*
- reduces to standard Cox regression if no heterogeneity
  (Occam's Razor action of Bayesian model selection)
- non-primary events all contribute to latent class inference
- fully transparent interpretation,
  unlike some other competing risk approaches ...
- formulae for survival curves *decontaminated* for informative censoring,
  and *retrospective class allocation* of individuals

Technicalities ...

- *censoring*
  modelled as 'risk' $r = 0$ with no associations

- *data likelihood*

$$p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell \; p(t,r|\boldsymbol{z},\ell), \qquad p(t,r|\boldsymbol{z},\ell) = \lambda_r^\ell(t) e^{\beta_r^\ell \cdot \boldsymbol{z} - \Lambda_0(t) - \sum_{r'=1}^{R} \exp(\beta_r^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}$$

- *base rates*
  spline construction for $\{\lambda_r^\ell(t)\}$, with K spline points

- *Bayesian model selection*
  K: baserate complexity
  L: number of latent classes
  M: heterogeneity complexity

- *numerical implementation*
  curvature estimation near parameter boundaries ...
  avoiding local minima in high-dim searches ...
  CPU efficiency ...

Technicalities ...

- *censoring*
  modelled as 'risk' $r = 0$ with no associations
- *data likelihood*

$$p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell \; p(t,r|\boldsymbol{z},\ell), \qquad p(t,r|\boldsymbol{z},\ell) = \lambda_r^\ell(t) e^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \Lambda_0(t) - \sum_{r'=1}^{R} \exp(\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}$$

- *base rates*
  spline construction for $\{\lambda_r^\ell(t)\}$, with K spline points
- *Bayesian model selection*
  K: baserate complexity
  L: number of latent classes
  M: heterogeneity complexity
- *numerical implementation*
  curvature estimation near parameter boundaries ...
  avoiding local minima in high-dim searches ...
  CPU efficiency ...

Technicalities ...

- *censoring*
  modelled as 'risk' $r = 0$ with no associations

- *data likelihood*

$$p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell \; p(t,r|\boldsymbol{z},\ell), \qquad p(t,r|\boldsymbol{z},\ell) = \lambda_r^\ell(t) e^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \Lambda_0(t) - \sum_{r'=1}^{R} \exp(\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}$$

- *base rates*
  spline construction for $\{\lambda_r^\ell(t)\}$, with K spline points

- *Bayesian model selection*
  K: baserate complexity
  L: number of latent classes
  M: heterogeneity complexity

- *numerical implementation*
  curvature estimation near parameter boundaries ...
  avoiding local minima in high-dim searches ...
  CPU efficiency ...

Technicalities ...

- *censoring*
  modelled as 'risk' $r = 0$ with no associations
- *data likelihood*

$$p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell \; p(t,r|\boldsymbol{z},\ell), \qquad p(t,r|\boldsymbol{z},\ell) = \lambda_r^\ell(t) e^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \Lambda_0(t) - \sum_{r'=1}^{R} \exp(\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}$$

- *base rates*
  spline construction for $\{\lambda_r^\ell(t)\}$, with K spline points
- *Bayesian model selection*
  K: baserate complexity
  L: number of latent classes
  M: heterogeneity complexity
- *numerical implementation*
  curvature estimation near parameter boundaries ...
  avoiding local minima in high-dim searches ...
  CPU efficiency ...

Technicalities ...

- *censoring*
  modelled as 'risk' $r = 0$ with no associations

- *data likelihood*

$$p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell \; p(t,r|\boldsymbol{z},\ell), \qquad p(t,r|\boldsymbol{z},\ell) = \lambda_r^\ell(t) e^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \Lambda_0(t) - \sum_{r'=1}^{R} \exp(\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}$$

- *base rates*
  spline construction for $\{\lambda_r^\ell(t)\}$, with K spline points

- *Bayesian model selection*
  K: baserate complexity
  L: number of latent classes
  M: heterogeneity complexity

- *numerical implementation*
  curvature estimation near parameter boundaries ...
  avoiding local minima in high-dim searches ...
  CPU efficiency ...

Upon determining parameters and hyper-parameters

explicit formulae for e.g.

- covariate-conditioned survival curves and hazard rates:

$$\textit{crude}: \qquad h_r(t|\boldsymbol{z}) = \frac{\sum_\ell w_\ell \; \lambda_r^\ell(t) \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}}{\sum_\ell w_\ell \; \mathrm{e}^{- \sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}},$$

$$\textit{decontaminated}: \qquad \tilde{h}_r(t|\boldsymbol{z}) = \frac{\sum_\ell w_\ell \; \lambda_r^\ell(t) \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \exp(\hat{\boldsymbol{\beta}}_r^\ell \cdot \boldsymbol{z}) \Lambda_r^\ell(t)}}{\sum_\ell w_\ell \; \mathrm{e}^{- \exp(\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z}) \Lambda_r^\ell(t)}}.$$

- cause-specific cumulative incidence function:

$$F_r(t|\boldsymbol{z}) = \int_0^t \mathrm{d}t' \; \mathrm{e}^{-\Lambda_0(t')} \sum_\ell w_\ell \; \lambda_r^\ell(t') \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t')}.$$

- class membership probabilities:

$$p(\ell|t, r, \boldsymbol{z}) = \frac{w_\ell p(t, r|\boldsymbol{z}, \ell)}{\sum_{\ell'=1}^L w_{\ell'} p(t, r|\boldsymbol{z}, \ell')}$$

Upon determining parameters and hyper-parameters

explicit formulae for e.g.

- covariate-conditioned survival curves and hazard rates:

$$\textit{crude}: \qquad h_r(t|\boldsymbol{z}) = \frac{\sum_\ell w_\ell \; \lambda_r^\ell(t) \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}}{\sum_\ell w_\ell \; \mathrm{e}^{-\sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}},$$

$$\textit{decontaminated}: \qquad \tilde{h}_r(t|\boldsymbol{z}) = \frac{\sum_\ell w_\ell \; \lambda_r^\ell(t) \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \exp(\hat{\boldsymbol{\beta}}_r^\ell \cdot \boldsymbol{z}) \Lambda_r^\ell(t)}}{\sum_\ell w_\ell \; \mathrm{e}^{-\exp(\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z}) \Lambda_r^\ell(t)}}.$$

- cause-specific cumulative incidence function:

$$F_r(t|\boldsymbol{z}) = \int_0^t \mathrm{d}t' \; \mathrm{e}^{-\Lambda_0(t')} \sum_\ell w_\ell \; \lambda_r^\ell(t') \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t')}.$$

- class membership probabilities:

$$p(\ell|t, r, \boldsymbol{z}) = \frac{w_\ell p(t, r|\boldsymbol{z}, \ell)}{\sum_{\ell'=1}^L w_{\ell'} p(t, r|\boldsymbol{z}, \ell')}$$

Upon determining parameters and hyper-parameters

explicit formulae for e.g.

- covariate-conditioned survival curves and hazard rates:

$$\textit{crude}: \qquad h_r(t|\boldsymbol{z}) = \frac{\sum_\ell w_\ell \; \lambda_r^\ell(t) \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}}{\sum_\ell w_\ell \; \mathrm{e}^{-\sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t)}},$$

$$\textit{decontaminated}: \qquad \tilde{h}_r(t|\boldsymbol{z}) = \frac{\sum_\ell w_\ell \; \lambda_r^\ell(t) \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \exp(\hat{\boldsymbol{\beta}}_r^\ell \cdot \boldsymbol{z}) \Lambda_r^\ell(t)}}{\sum_\ell w_\ell \; \mathrm{e}^{-\exp(\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z}) \Lambda_r^\ell(t)}}.$$

- cause-specific cumulative incidence function:

$$F_r(t|\boldsymbol{z}) = \int_0^t \mathrm{d}t' \; \mathrm{e}^{-\Lambda_0(t')} \sum_\ell w_\ell \; \lambda_r^\ell(t') \mathrm{e}^{\boldsymbol{\beta}_r^\ell \cdot \boldsymbol{z} - \sum_{r'=1}^R \exp(\boldsymbol{\beta}_{r'}^\ell \cdot \boldsymbol{z}) \Lambda_{r'}^\ell(t')}.$$

- class membership probabilities:

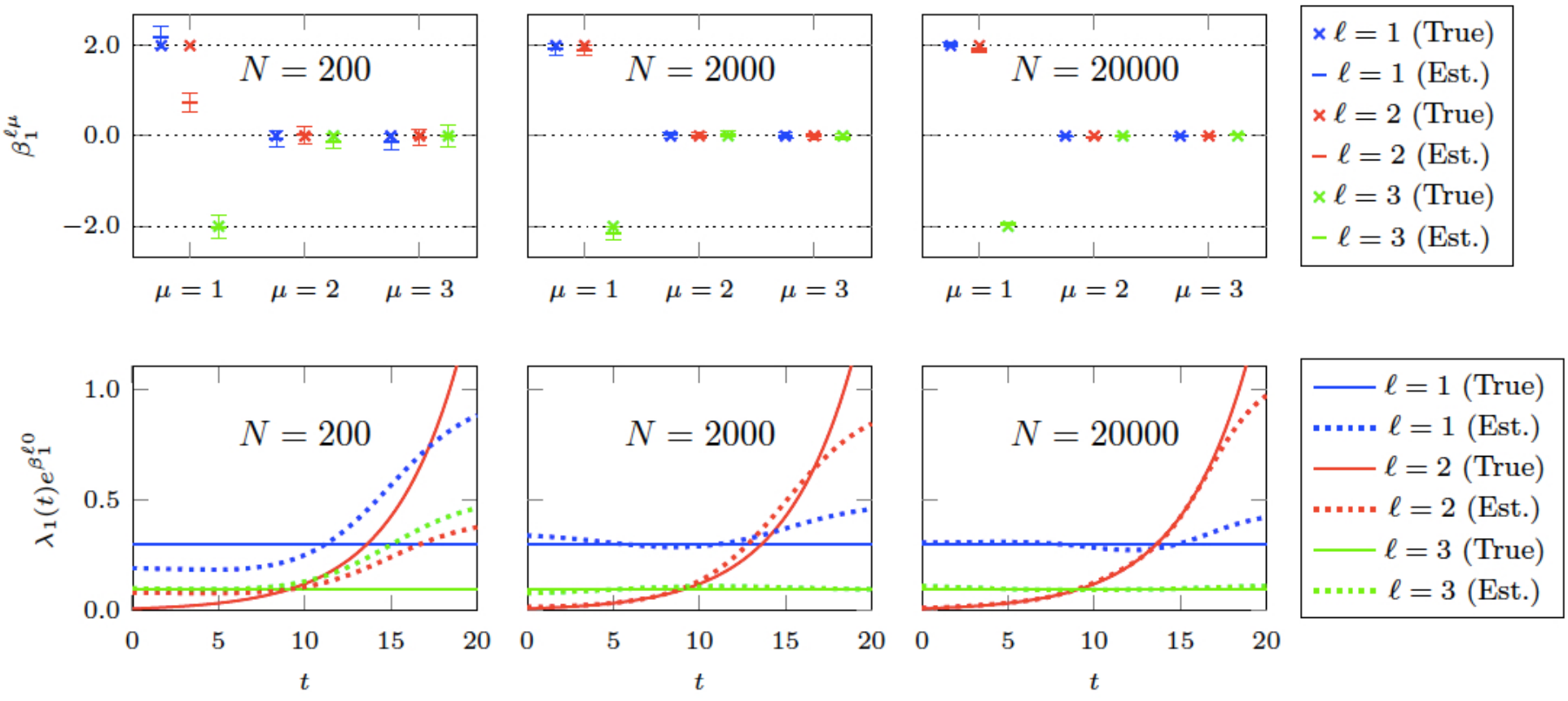
$$p(\ell|t, r, \boldsymbol{z}) = \frac{w_\ell p(t, r|\boldsymbol{z}, \ell)}{\sum_{\ell'=1}^L w_{\ell'} p(t, r|\boldsymbol{z}, \ell')}$$

Tests on synthetic data

inference of classes and parameters

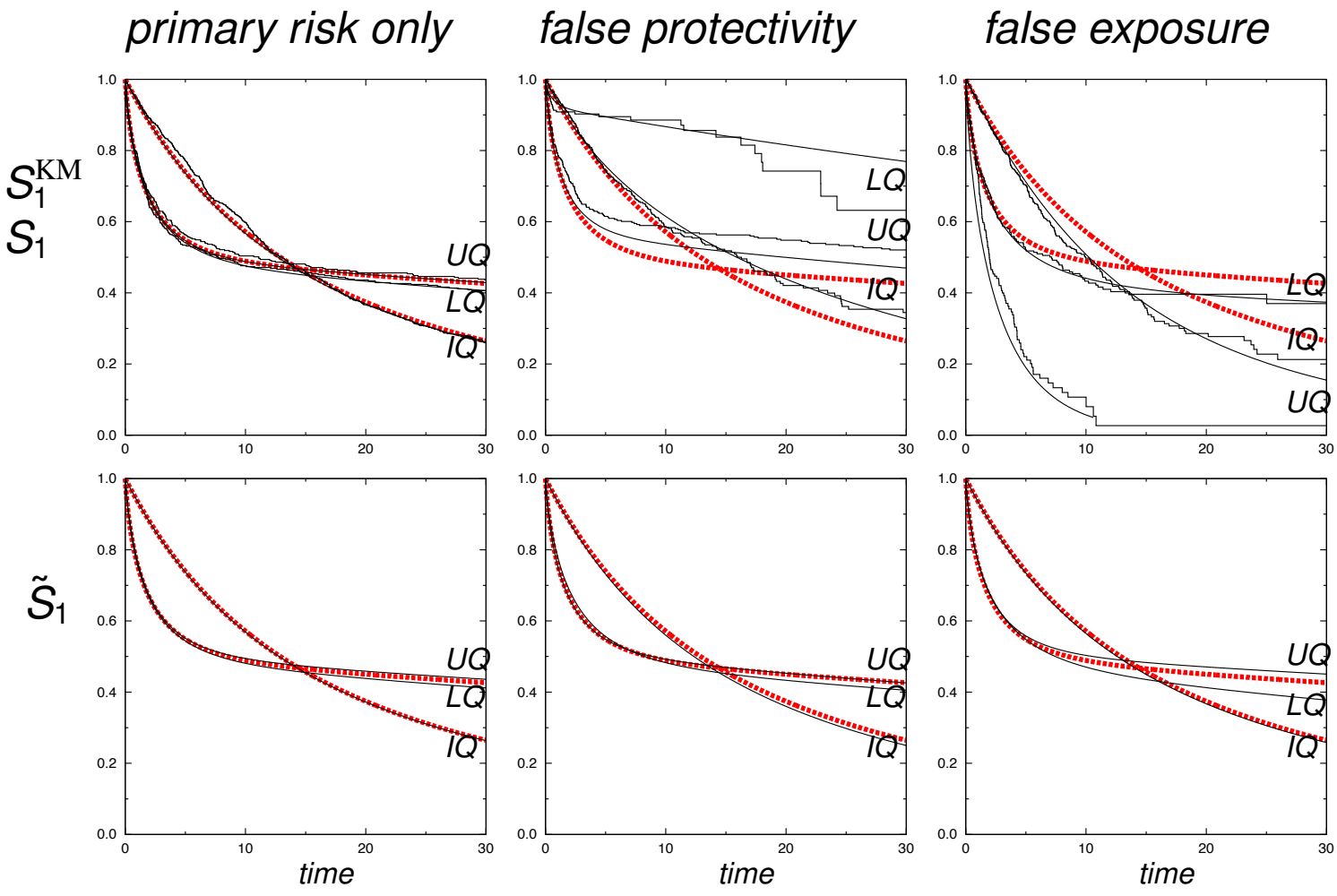
3 classes:

red, blue, green



Tests on synthetic data

decontaminating survival curves for informative censoring



S_1^{KM} : *Kaplan-Meier*

S_1 : *crude survival curve*

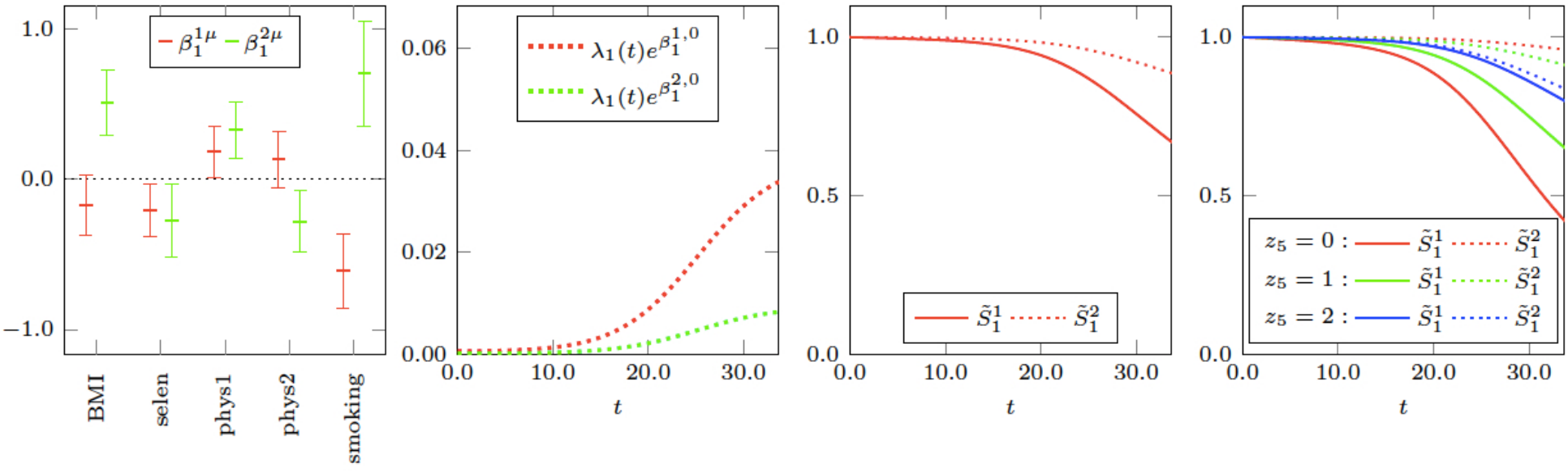
red dashed: true survival curves

$\tilde{S}_1$: *decontaminated curves*

Prostate cancer data

(ULSAM data base, $n = 2047$)

Cox regression:
smoking is protective against PC



negative association with smoking *only* in *extremely frail* subgroup of patients

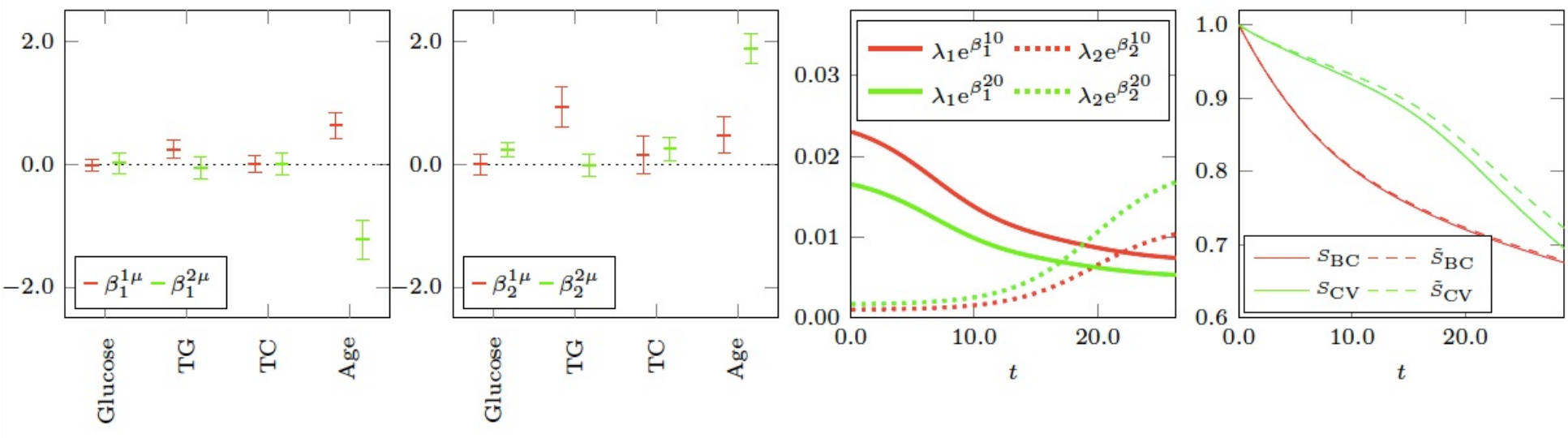
red class: high overall frailty
green class: low overall frailty

Breast cancer data

(AMORIS data base, $N = 1798$)

Cox regression finds no significant associations
(proportional hazards violated)

BC death *CV death* *solid: BC dashed: CV*



red class: predominantly younger women
green class: predominantly older women

Applications to failed cancer trials

- *failed clinical trials*

  often some drug benefit, but not enough in view of costs ...
  (in the absence of a biomarker to select patients)

- *two possibilities*

  1. there exist measurable differences between individuals that explain response variation, we just don't know what they are ...

     subgroups with distinct quantitative characteristics,
     cohort is in principle stratifiable

  2. there are no measurable differences between individuals to explain response variation: cohort not stratifiable

- *Bayesian Latent class analysis*

  – rational method for determining whether cohort is stratifiable
  – retrospective class assignment: tool for identifying latent classes

Applications to failed cancer trials

- *failed clinical trials*

  often some drug benefit, but not enough in view of costs ...
  (in the absence of a biomarker to select patients)

- *two possibilities*

  1. there exist measurable differences between individuals that explain response variation, we just don't know what they are ...

     subgroups with distinct quantitative characteristics,
     cohort is in principle stratifiable

  2. there are no measurable differences between individuals to explain response variation: cohort not stratifiable

- *Bayesian Latent class analysis*

  – rational method for determining whether cohort is stratifiable
  – retrospective class assignment: tool for identifying latent classes

Applications to failed cancer trials

- *failed clinical trials*

  often some drug benefit, but not enough in view of costs ...
  (in the absence of a biomarker to select patients)

- *two possibilities*

  1. there exist measurable differences between individuals that explain response variation, we just don't know what they are ...

     subgroups with distinct quantitative characteristics,
     cohort is in principle stratifiable

  2. there are no measurable differences between individuals to explain response variation: cohort not stratifiable

- *Bayesian Latent class analysis*

  – rational method for determining whether cohort is stratifiable
  – retrospective class assignment: tool for identifying latent classes

Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial

Timothy S Maughan, Richard A Adams, Christopher G Smith, Angela M Meade, Matthew T Seymour, Richard H Wilson, Shelley Idziaszczyk, Rebecca Harris, David Fisher, Sarah L Kenny, Edward Kay, Jenna K Mitchell, Ayman Madi, Bharat Jasani, Michelle D James, John Bridgewater, M John Kennedy, Bart Claes, Diether Lambrechts, Richard Kaplan, Jeremy P Cheadle, on behalf of the MRC COIN Trial Investigators

Lancet 2011; 377: 2103–14
Published Online
June 4, 2011

Summary

Background In the Medical Research Council (MRC) COIN trial, the epidermal growth factor receptor (EGFR)-targeted antibody cetuximab was added to standard chemotherapy in first-line treatment of advanced colorectal cancer with the aim of assessing effect on overall survival.

outcome:

Interpretation This trial has not confirmed a benefit of addition of cetuximab to oxaliplatin-based chemotherapy in first-line treatment of patients with advanced colorectal cancer. Cetuximab increases response rate, with no evidence of benefit in progression-free or overall survival in *KRAS* wild-type patients or even in patients selected by additional mutational analysis of their tumours. The use of cetuximab in combination with oxaliplatin and capecitabine in first-line chemotherapy in patients with widespread metastases cannot be recommended.

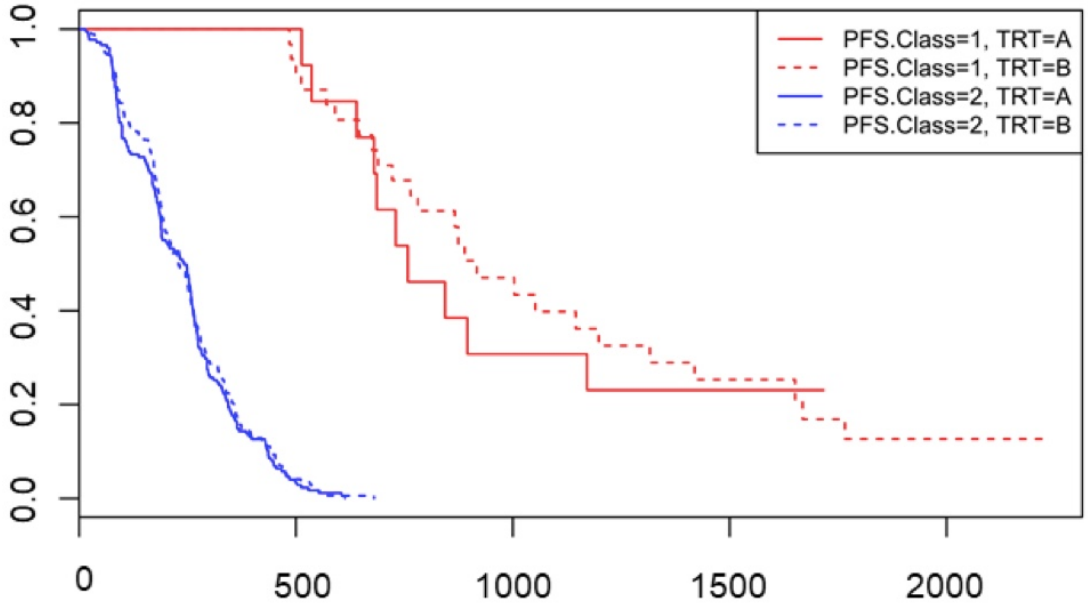
The COIN trial (colorectal cancer)

first analysis: $n = 398$
validation: $n = 1630$

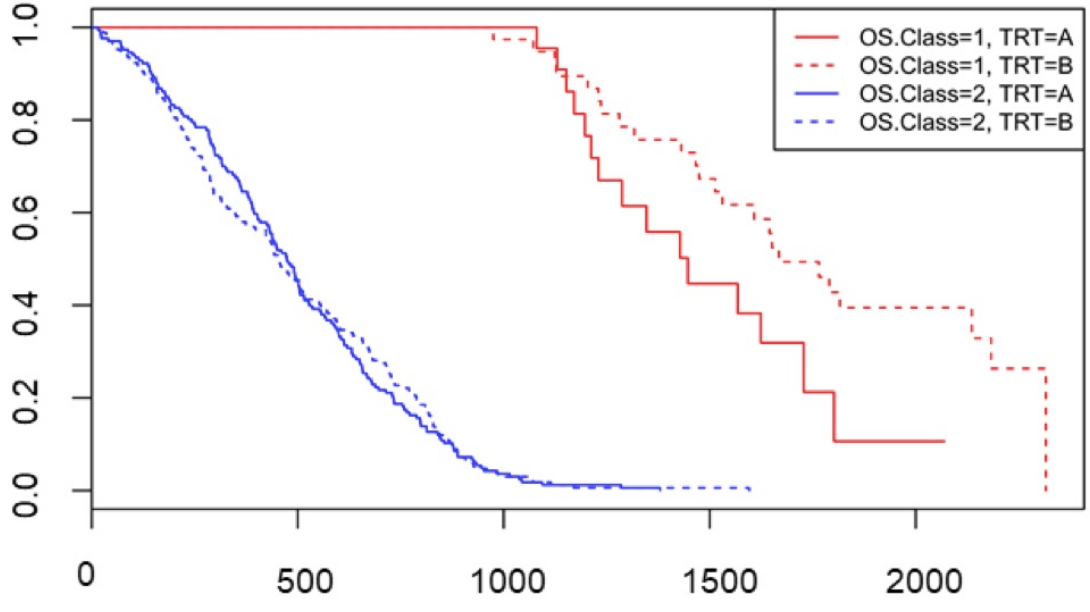
HR [95% CI]	$\beta(0)$	FRET eff	Her2-Her3	Cetux	KRASmut
Cox (M1L1K5), *lnZ=-2419.82*					
	-1.89	0.9 [0.7-1.0]	1.1 [0.9-1.5]	0.8 [0.7-0.9]	1.3 [1.1-1.7]
		p=0.3	p=0.4	p=0.03	p=0.006
Model M2L2K4(R), *lnZ=-2418.064*					
class I, W=31%	*-2.57*	1.8 [0.8-4.6]	0.8 [0.4-1.7]	0.5 [0.3-1.0]	1.5 [0.9-2.6]
alloc$[p_1 > 0.5]$*: N=59*		p=0.2	p=0.6	p=0.05	p=0.2
class II, W=69%	*-1.56*	0.5 [0.4-0.8]	1.4 [0.9-2.1]	1.0 [0.7-1.4]	1.3 [0.9-1.9]
alloc$[p_2 > 0.5]$*: N=339*		p=0.006	p=0.1	p=0.8	p=0.1

- two sub-cohorts, with similar base hazard rates, but distinct overall frailties and associations.
- method provides retrospective class assignment
- new tools to identify *a priori* the responders to Cetuximab?

PFS


PFS.Class=1, TRT=A
PFS.Class=1, TRT=B
PFS.Class=2, TRT=A
PFS.Class=2, TRT=B
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000

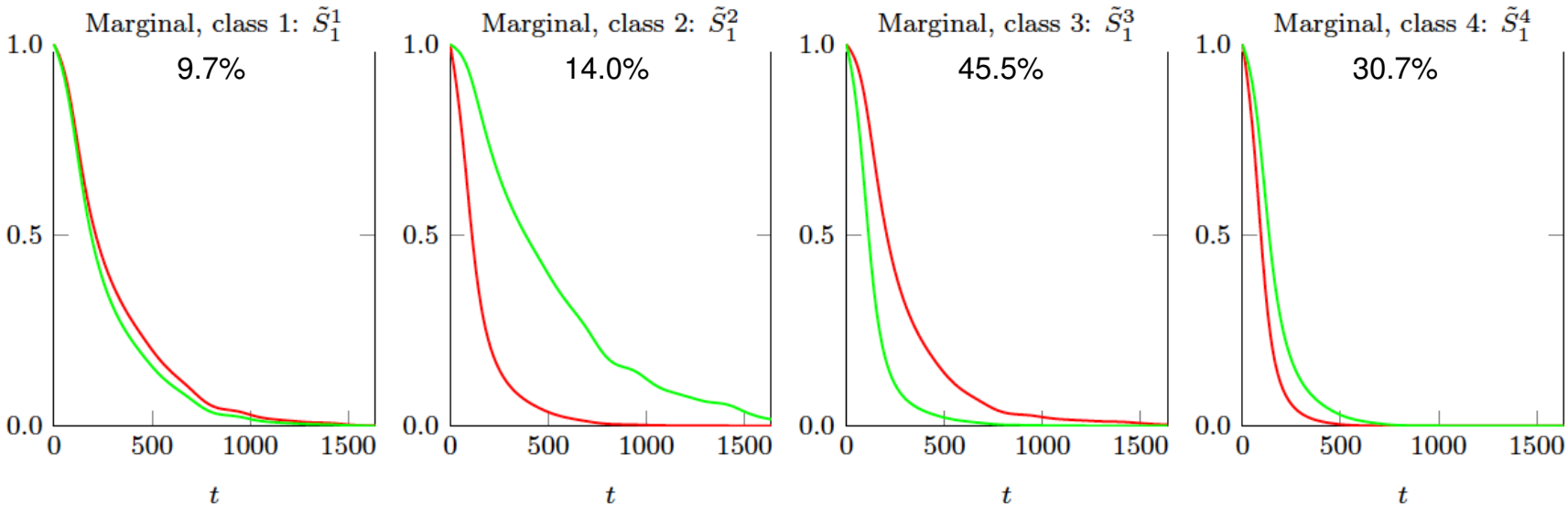
OS


OS.Class=1, TRT=A
OS.Class=1, TRT=B
OS.Class=2, TRT=A
OS.Class=2, TRT=B
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000

The TOPICAL trial (lung cancer)

$n = 580$

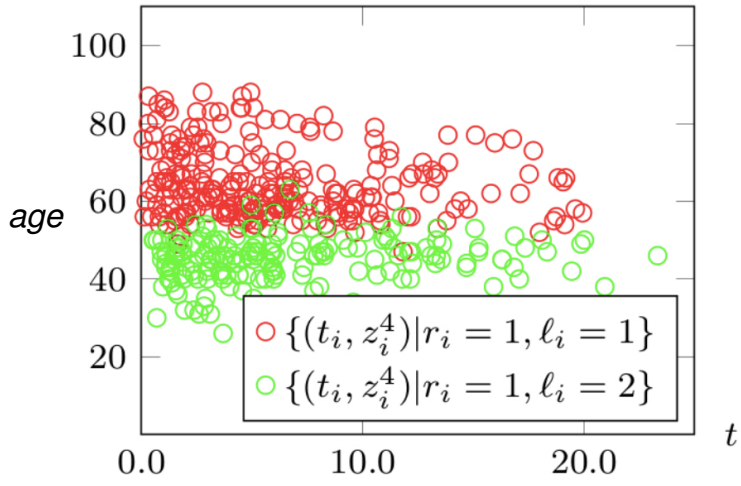
Risk 1				
	Class 1	Class 2	Class 3	Class 4
Covariate	HR, 95% CI, p-value	HR, 95% CI, p-value	HR, 95% CI, p-value	HR, 95% CI, p-value
AGE	0.77, [0.34,1.72], 0.521	2.93, [1.49,5.73], 0.002	0.59, [0.17,1.98], 0.390	0.92, [0.45,1.87], 0.819
Male	0.79, [0.36,1.74], 0.560	1.78, [0.92,3.42], 0.086	0.88, [0.33,2.35], 0.806	3.12, [1.44,6.74], 0.004
ECOG 2-3	0.40, [0.13,1.19], 0.099	1.49, [0.86,2.57], 0.156	1.54, [0.81,2.94], 0.186	1.75, [0.72,4.28], 0.216
Stage IV	1.34, [0.75,2.39], 0.326	1.46, [0.80,2.67], 0.219	1.96, [0.85,4.55], 0.116	1.20, [0.67,2.15], 0.539
Adenocarcinoma	7.20, [2.61,19.85], < 0.001	0.44, [0.24,0.82], 0.009	1.46, [0.65,3.31], 0.361	0.68, [0.33,1.39], 0.291
Ex-smoker	2.04, [0.56,7.48], 0.281	0.19, [0.06,0.63], 0.006	8.39, [2.12,33.18], 0.002	0.69, [0.27,1.77], 0.438
Smoker	5.04, [1.50,16.98], 0.009	0.30, [0.09,1.05], 0.060	4.99, [1.31,18.96], 0.018	1.19, [0.47,3.00], 0.717
CCI 4+	1.41, [0.65,3.06], 0.386	1.47, [0.80,2.67], 0.211	0.87, [0.25,2.96], 0.818	1.21, [0.55,2.63], 0.636
Good	0.32, [0.15,0.65], 0.002	0.23, [0.12,0.46], < 0.001	0.43, [0.21,0.87], 0.019	1.35, [0.70,2.61], 0.366
Tarceva	1.48, [0.79,2.76], 0.223	0.11, [0.05,0.22], < 0.001	3.95, [0.94,16.66], 0.061	0.45, [0.20,1.00], 0.050



survival curves: green=erlotinib, red=placebo

Prospective latent class prediction

If any of the covariates correlate with retrospective class membership:
(e.g. Amoris)



replace

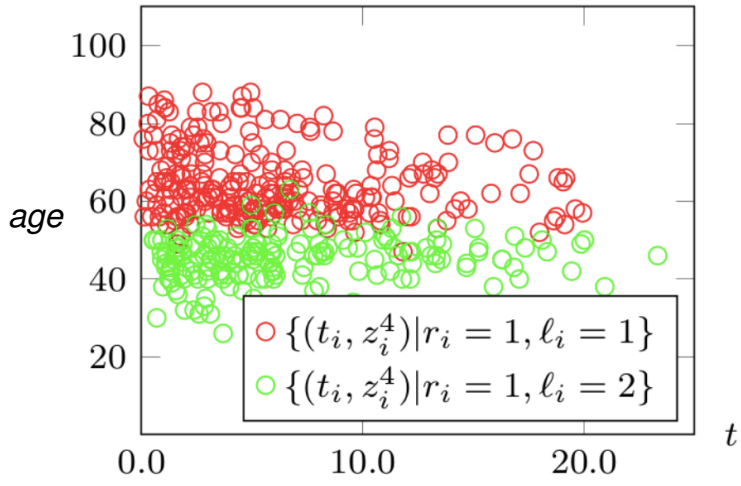
$$p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell \; p(t,r|\boldsymbol{z},\ell) \quad \rightarrow \quad p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell(\boldsymbol{z}) \; p(t,r|\boldsymbol{z},\ell)$$

- suitable parametrisation $w_\ell(\boldsymbol{z})$
- *prospective* class prediction,
  i.e. objective data-driven stratification to rescue failed trials

increasingly complex models,
many parameters,
danger of *overfitting* ...

Prospective latent class prediction

If any of the covariates correlate with retrospective class membership:
(e.g. Amoris)



replace

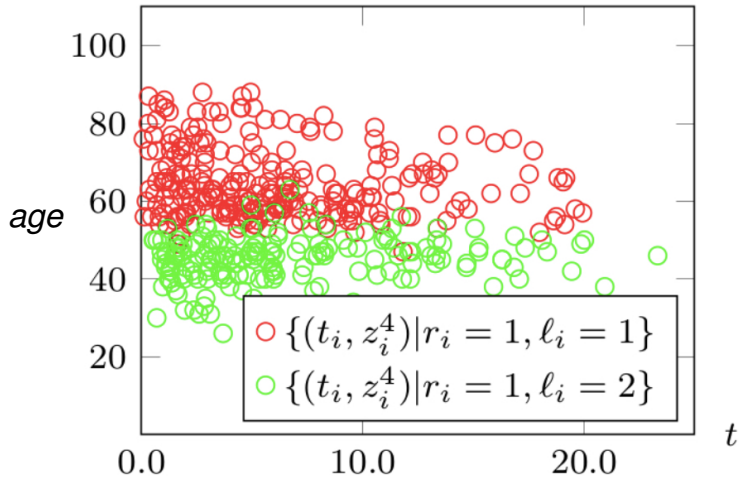
$$p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell \; p(t,r|\boldsymbol{z},\ell) \quad \rightarrow \quad p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell(\boldsymbol{z}) \; p(t,r|\boldsymbol{z},\ell)$$

- suitable parametrisation $w_\ell(\boldsymbol{z})$
- *prospective* class prediction,
  i.e. objective data-driven stratification to rescue failed trials

increasingly complex models,
many parameters,
danger of *overfitting* ...

Prospective latent class prediction

If any of the covariates correlate with retrospective class membership:
(e.g. Amoris)



replace

$$p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell \; p(t,r|\boldsymbol{z},\ell) \quad \rightarrow \quad p(t,r|\boldsymbol{z}) = \sum_{\ell=1}^{L} w_\ell(\boldsymbol{z}) \; p(t,r|\boldsymbol{z},\ell)$$

- suitable parametrisation $w_\ell(\boldsymbol{z})$
- *prospective* class prediction,
  i.e. objective data-driven stratification to rescue failed trials

increasingly complex models,
many parameters,
danger of *overfitting* ...

overfitting in Cox regression

ML method ...

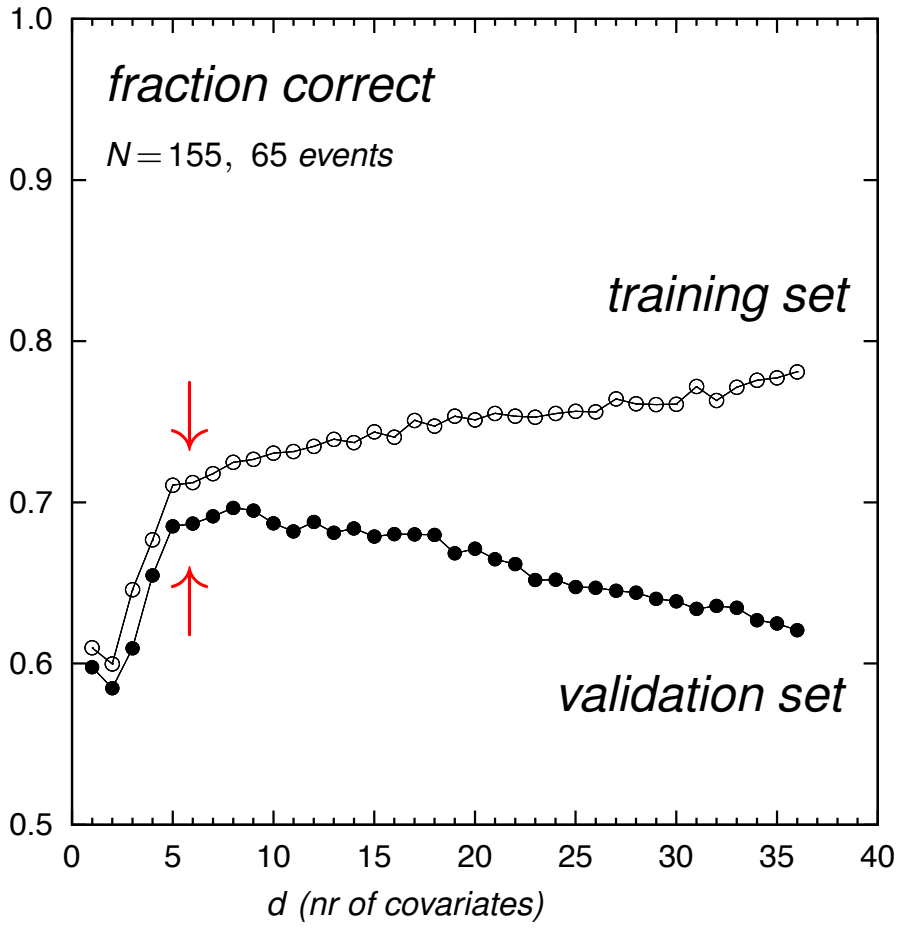
p-values, z-scores,
confidence intervals,
don't measure overfitting!

rule of thumb: $d_{\max} = events/10$

- too optimistic ...
- must depend on β ...
- covariate correlations ...

What happens in overfitting regime?

Can we predict the optimal point?



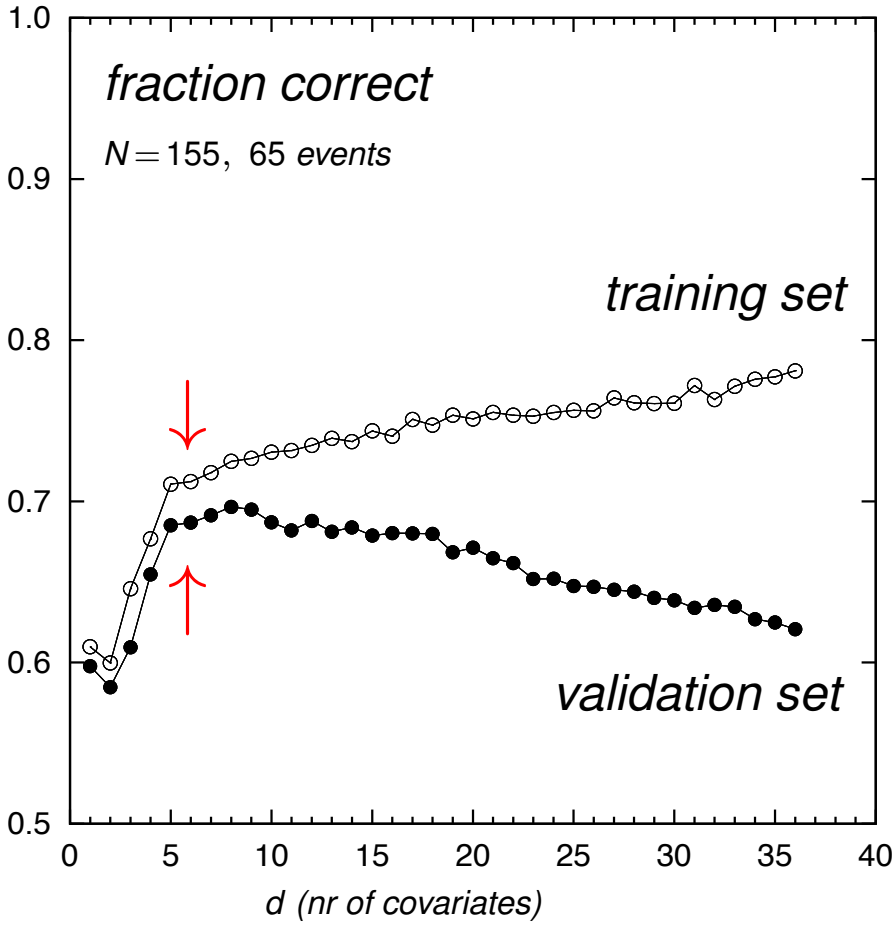
overfitting in Cox regression

ML method ...

p-values, z-scores,
confidence intervals,
don't measure overfitting!

rule of thumb: $d_{\max} = events/10$

- too optimistic ...
- must depend on β ...
- covariate correlations ...



What happens in overfitting regime?

Can we predict the optimal point?

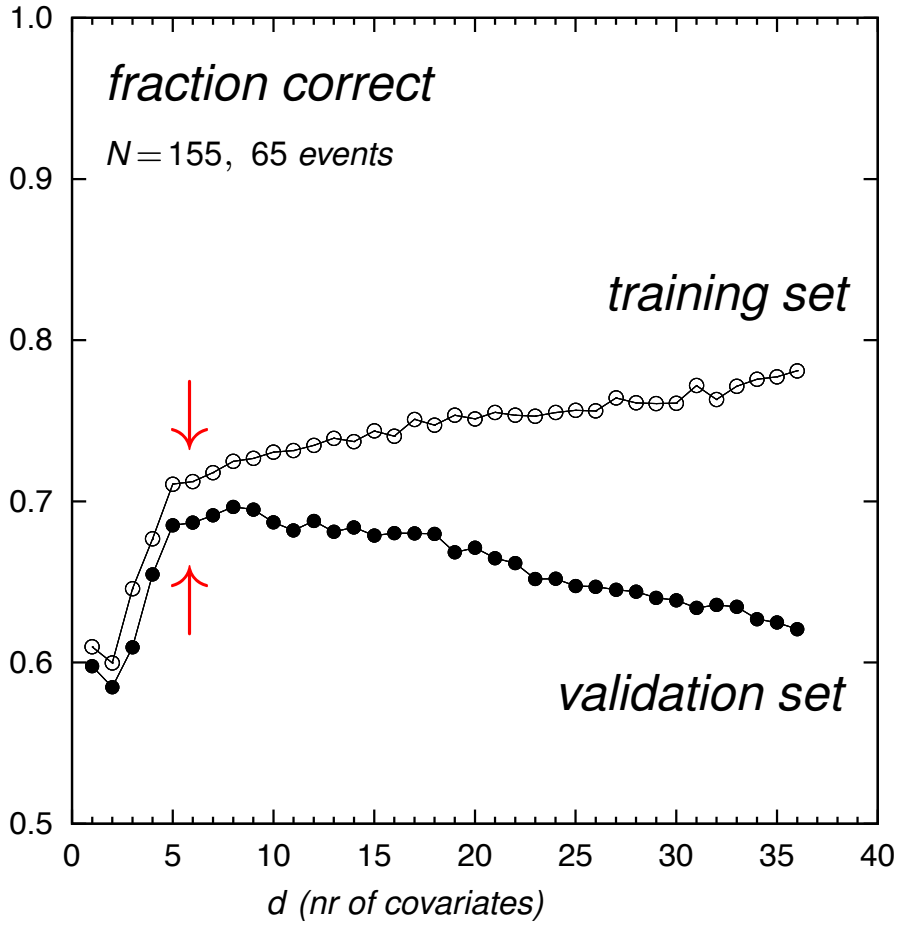
overfitting in Cox regression

ML method ...

p-values, z-scores,
confidence intervals,
don't measure overfitting!

rule of thumb: $d_{\max} = events/10$

- too optimistic ...
- must depend on β ...
- covariate correlations ...

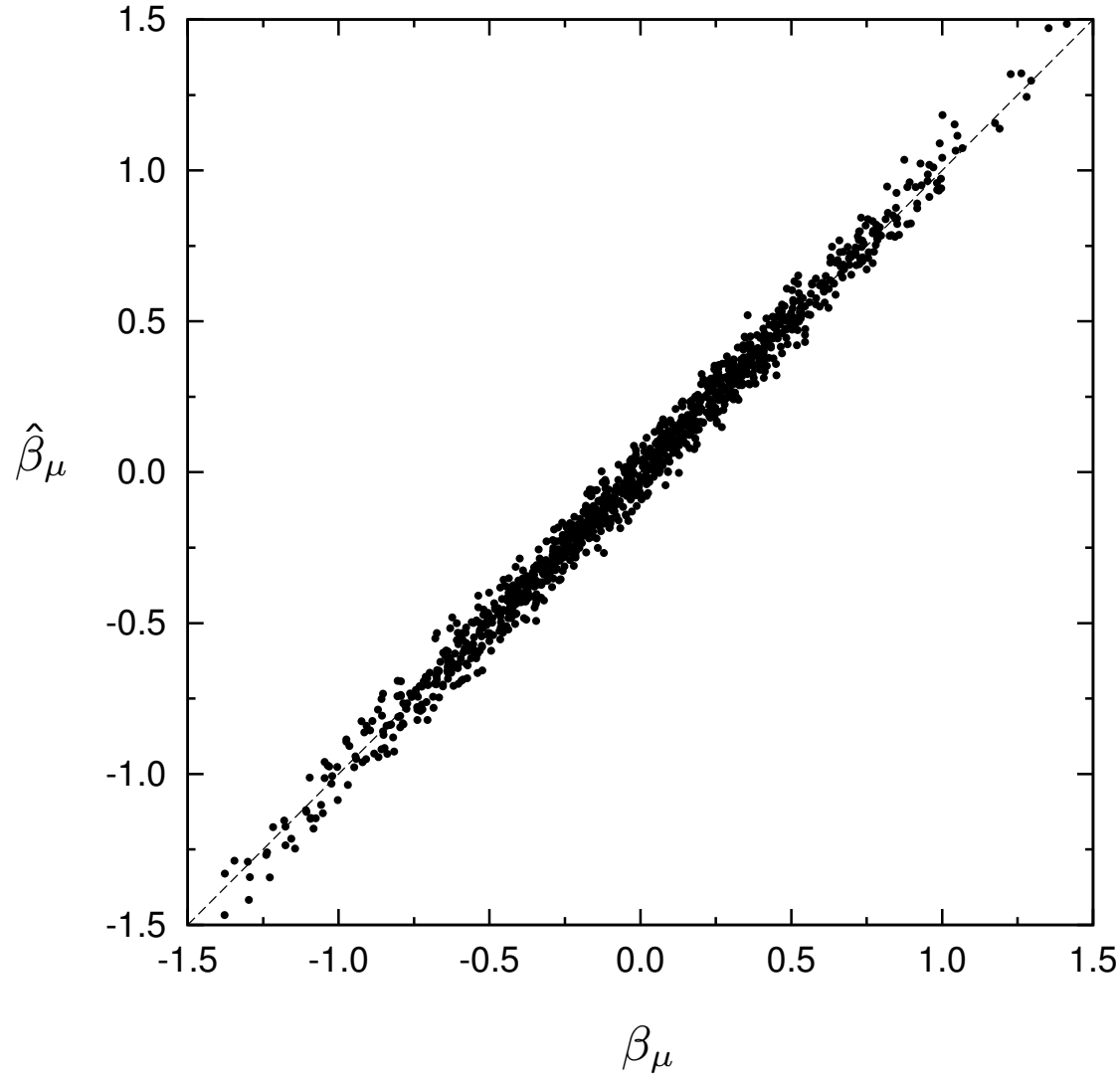


What happens in overfitting regime?

Can we predict the optimal point?

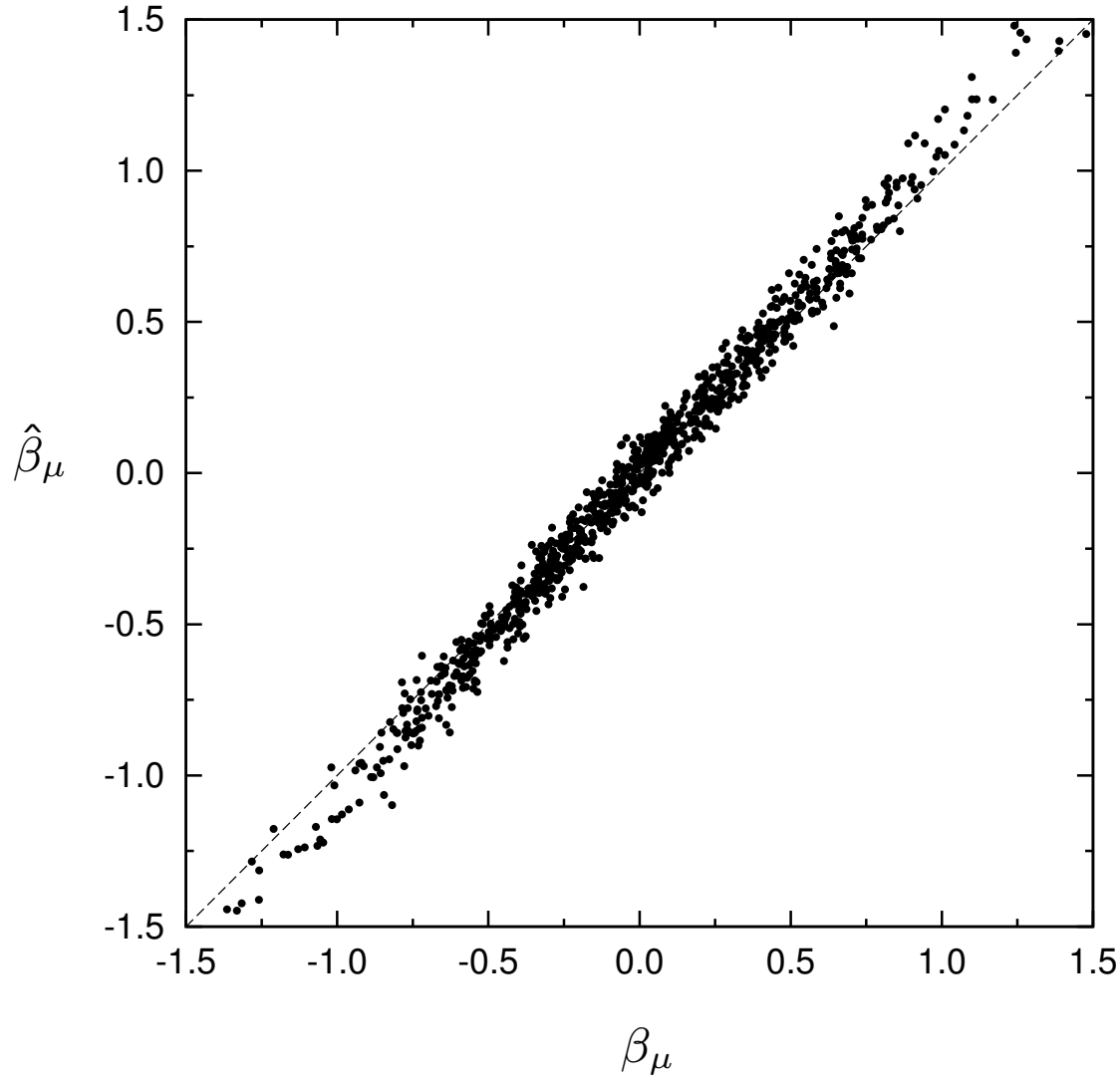
$N = 500$,
predicted versus true regression coefficients
(synthetic data, no censoring)

$d/N = 0.002$



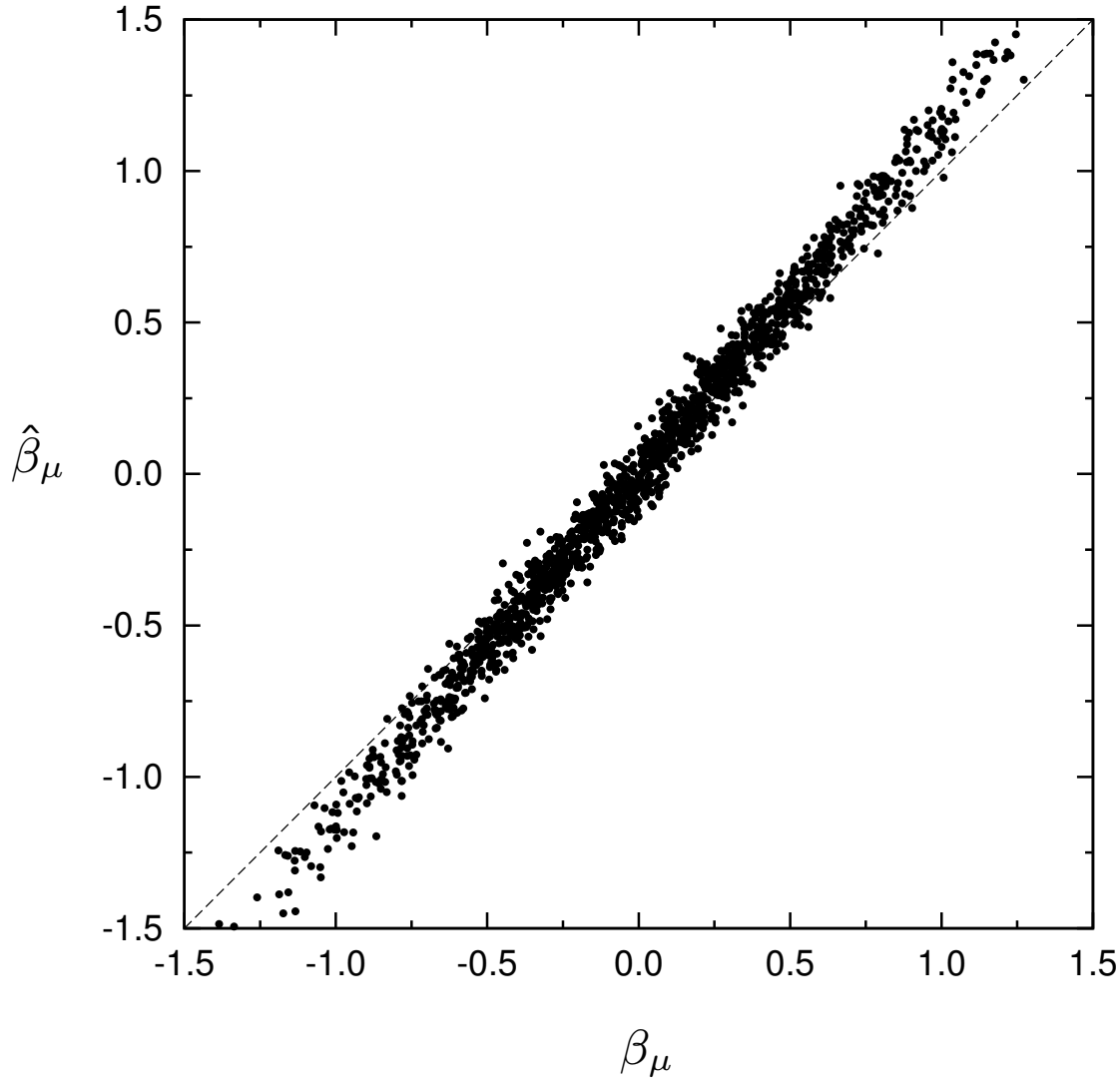
$N = 500$,
predicted versus true regression coefficients
(synthetic data, no censoring)

$d/N = 0.10$



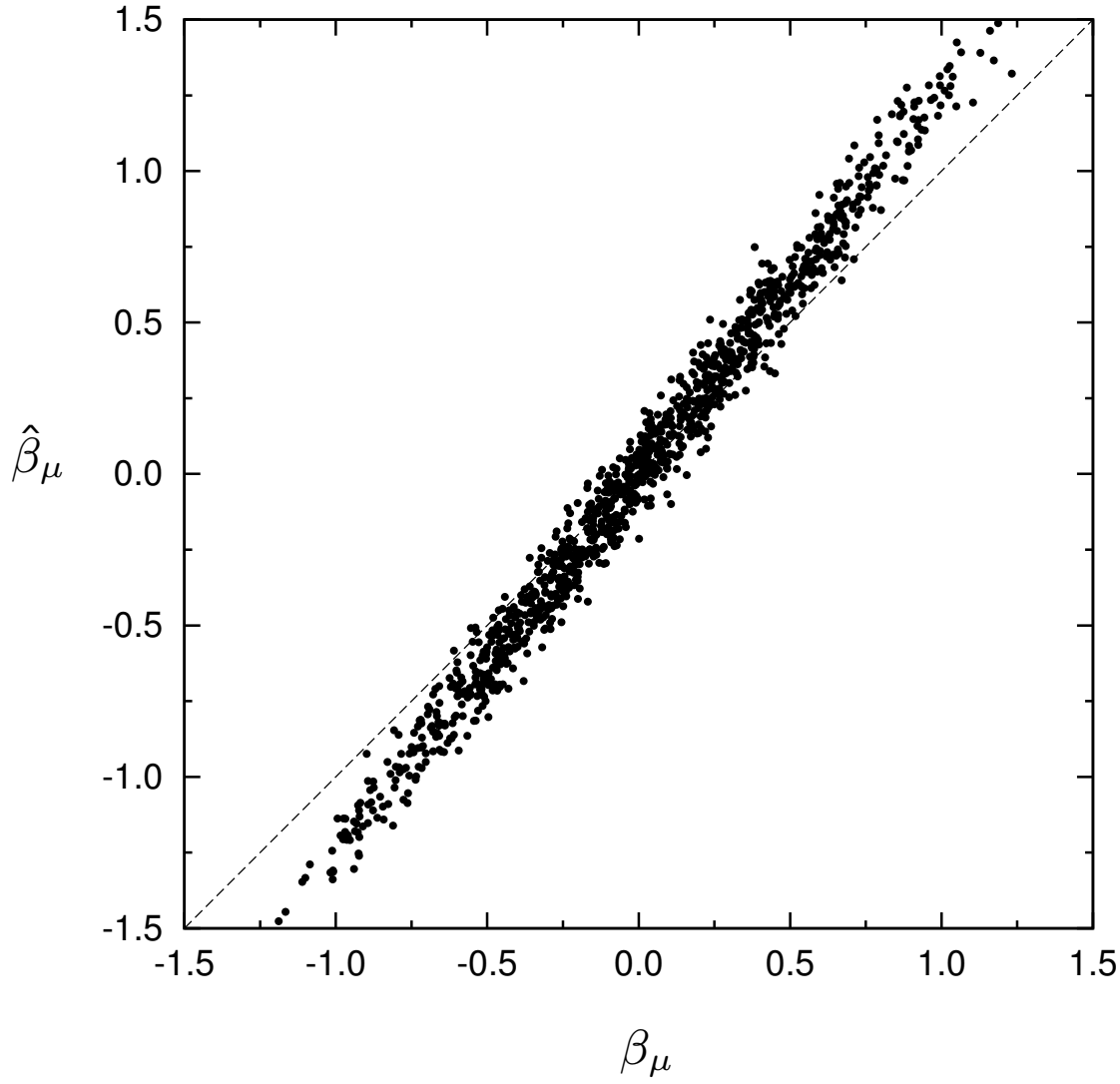
$N = 500$,
predicted versus true regression coefficients
(synthetic data, no censoring)

$d/N = 0.20$



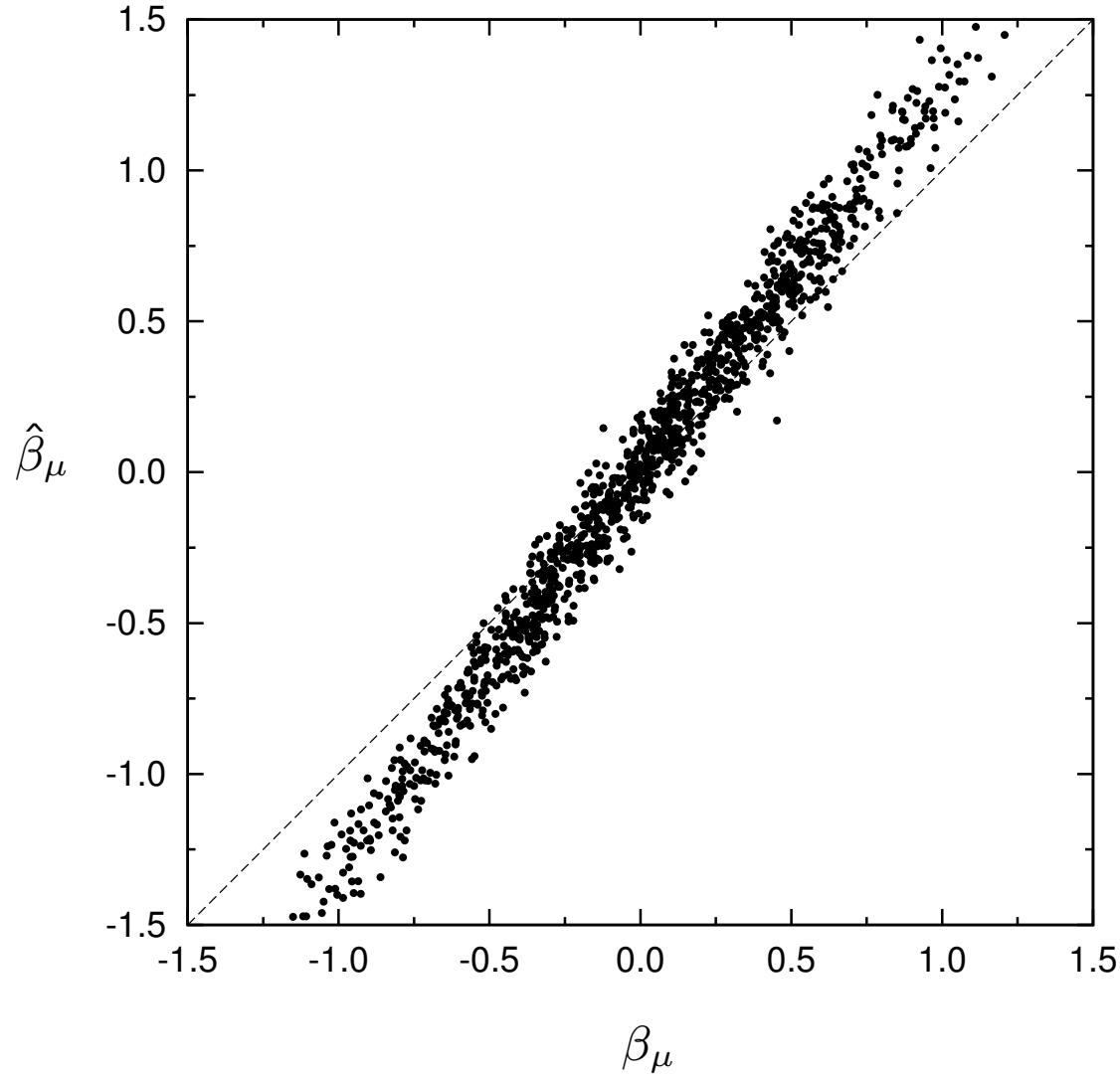
$N = 500$,
predicted versus true regression coefficients
(synthetic data, no censoring)

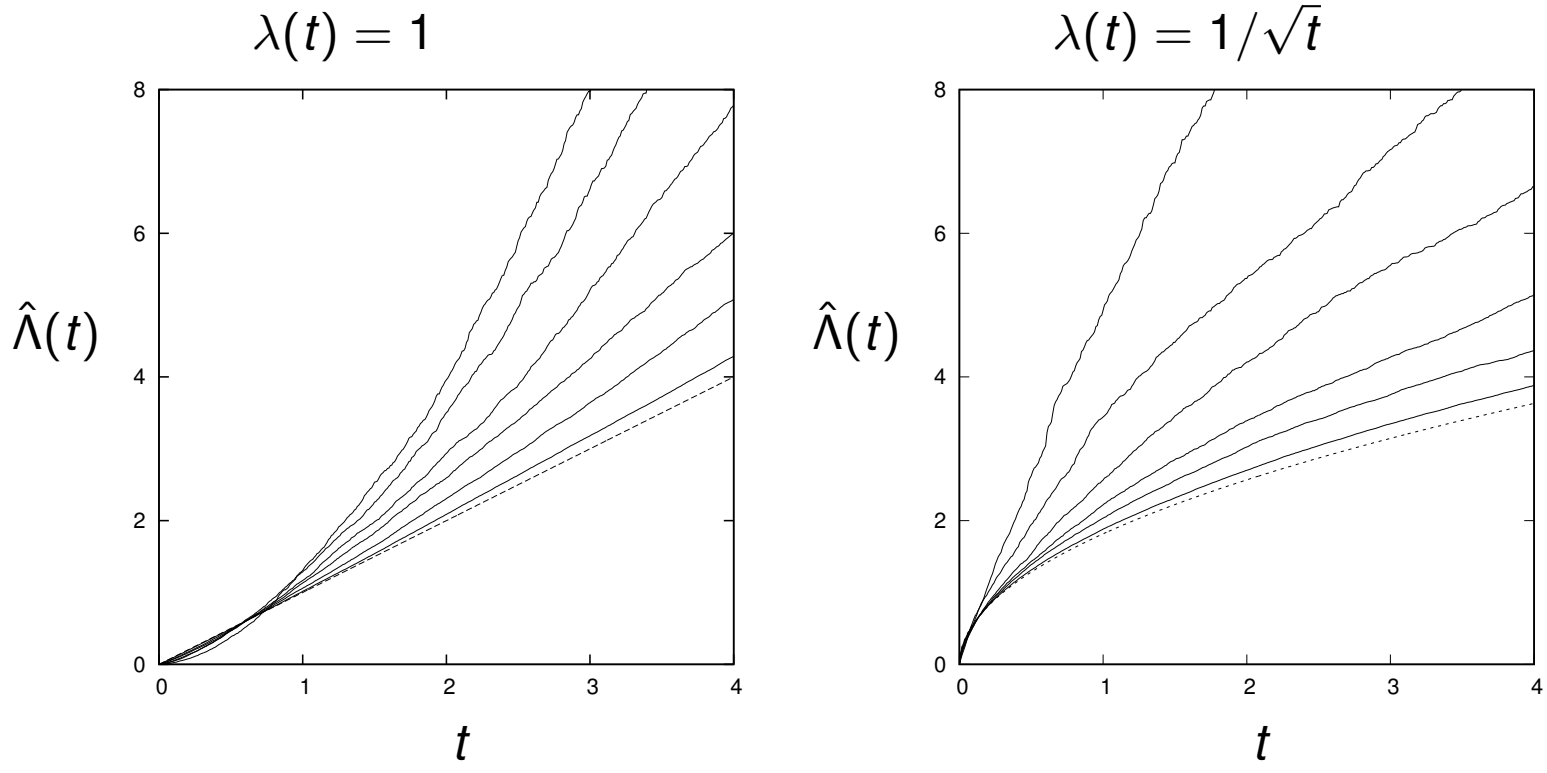
$d/N = 0.30$



$N = 500$,
predicted versus true regression coefficients
(synthetic data, no censoring)

$d/N = 0.40$





Base hazard rates underestimated for short times,
and over-estimated for large times ...

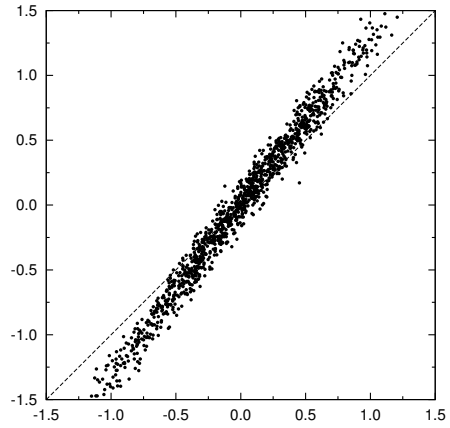
$d/N = 0.05, 0.15, 0.25, 0.35, 0.45, 0.55$
(lower to upper curves)

Gaussian association pars, $\langle \beta_\mu^2 \rangle = 0.25$, $N = 400$

Bad news

Overfitting *more dangerous*
than finite sample noise ...

we always inflate associations
(whether positive or negative)



Good news

Unlike pure noise,
deterministic bias may be predictable ...

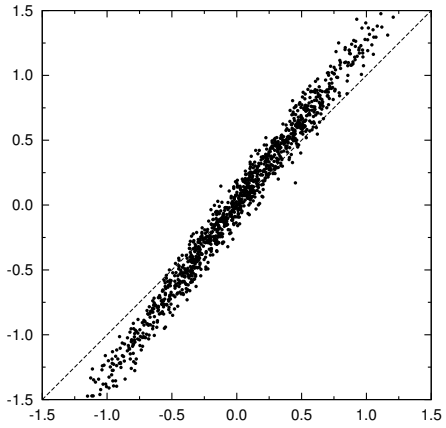
Roadmap for research ...

- Predict asymptotic impact of overfitting, in terms of
  – ratio d/N
  – correlations among covariates
  – true association strengths β
- Based on information theory and replica analysis
  (in spirit of Gardner theory of binary classifiers)
- Overfitting correction of Cox parameters

Bad news

Overfitting *more dangerous*
than finite sample noise ...

we always inflate associations
(whether positive or negative)



Good news

Unlike pure noise,
deterministic bias may be predictable ...

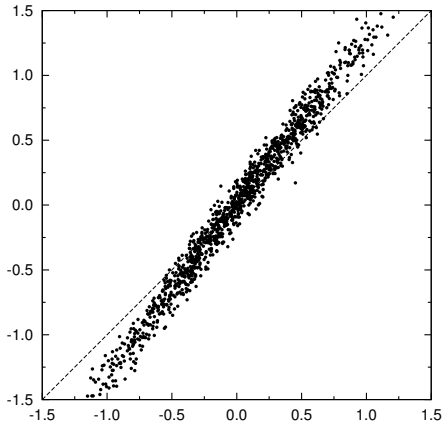
Roadmap for research ...

- Predict asymptotic impact of overfitting, in terms of
  - ratio d/N
  - correlations among covariates
  - true association strengths β
- Based on information theory and replica analysis
  (in spirit of Gardner theory of binary classifiers)
- Overfitting correction of Cox parameters

Bad news

Overfitting *more dangerous*
than finite sample noise ...

we always inflate associations
(whether positive or negative)



Good news

Unlike pure noise,
deterministic bias may be predictable ...

Roadmap for research ...

- Predict asymptotic impact of overfitting, in terms of
  – ratio d/N
  – correlations among covariates
  – true association strengths β
- Based on information theory and replica analysis
  (in spirit of Gardner theory of binary classifiers)
- Overfitting correction of Cox parameters

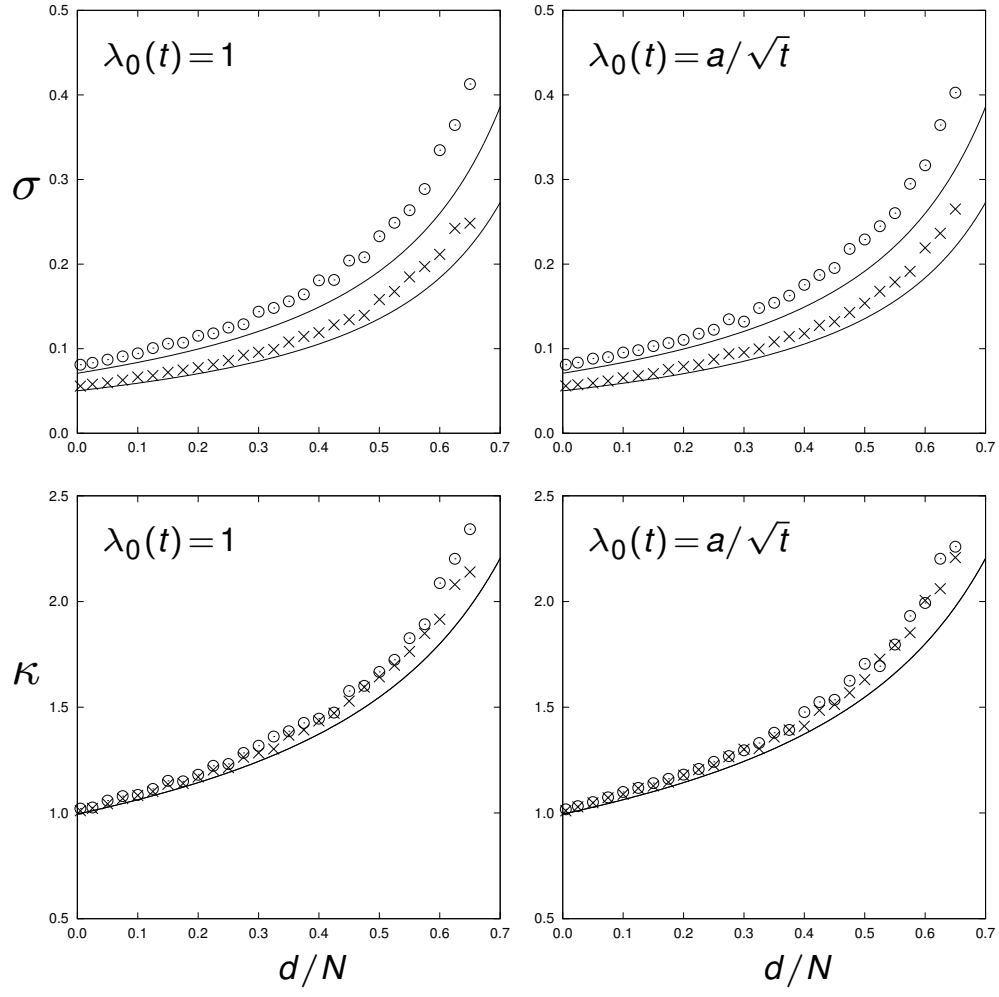
Overfitting correction for multivariate Cox regression

regression up to $d/N \sim 0.5$,
compared to $d/N \sim 0.1$
(variational replica theory)

width σ and slope κ
of data clouds

lines: theory
for $S = 0.5$ and $\langle t \rangle = 1$

o: $N = 200$
x: $N = 400$



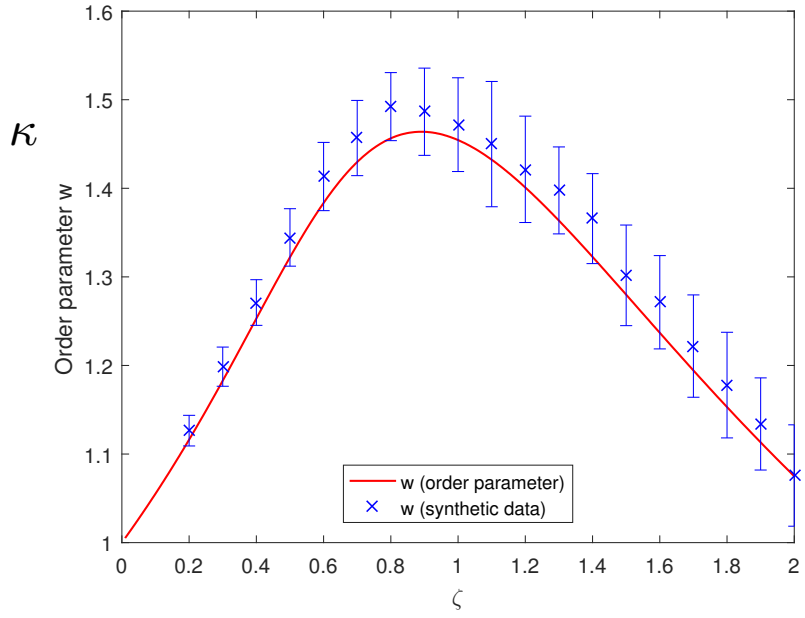
from ML to MAP inference,

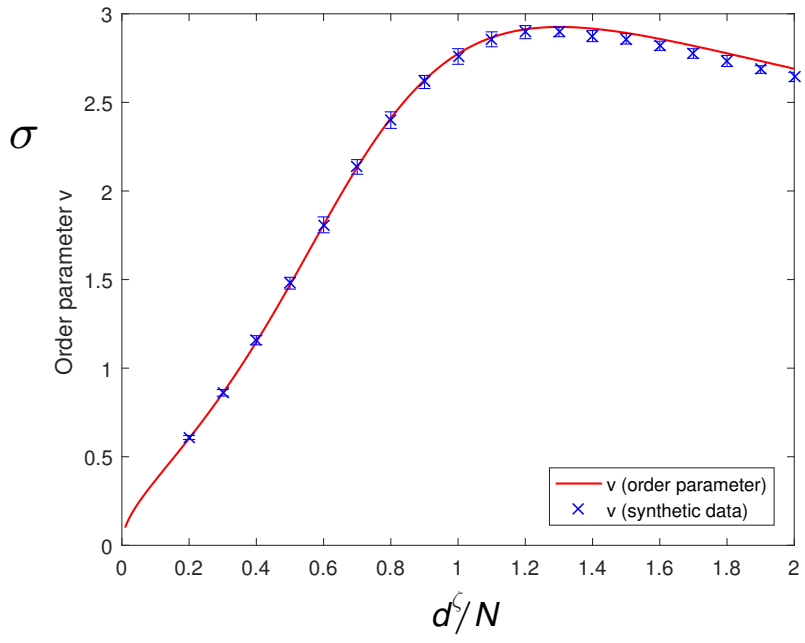
prior: $p(\boldsymbol{\beta}) \propto \exp(-\frac{1}{2}\eta p \boldsymbol{\beta}^2)$

regression up to $d/N \sim 2$ or more ...,
compared to $d/N \sim 0.1$
(variational replica theory)

width σ and slope κ
of data clouds

lines: theory,
markers: MAP Cox regression





IMMB @ King's College London

James Barrett, Mark Rowley,
Mansoor Sheikh, Alexander Mozeika

Cancer Division @ King's College London

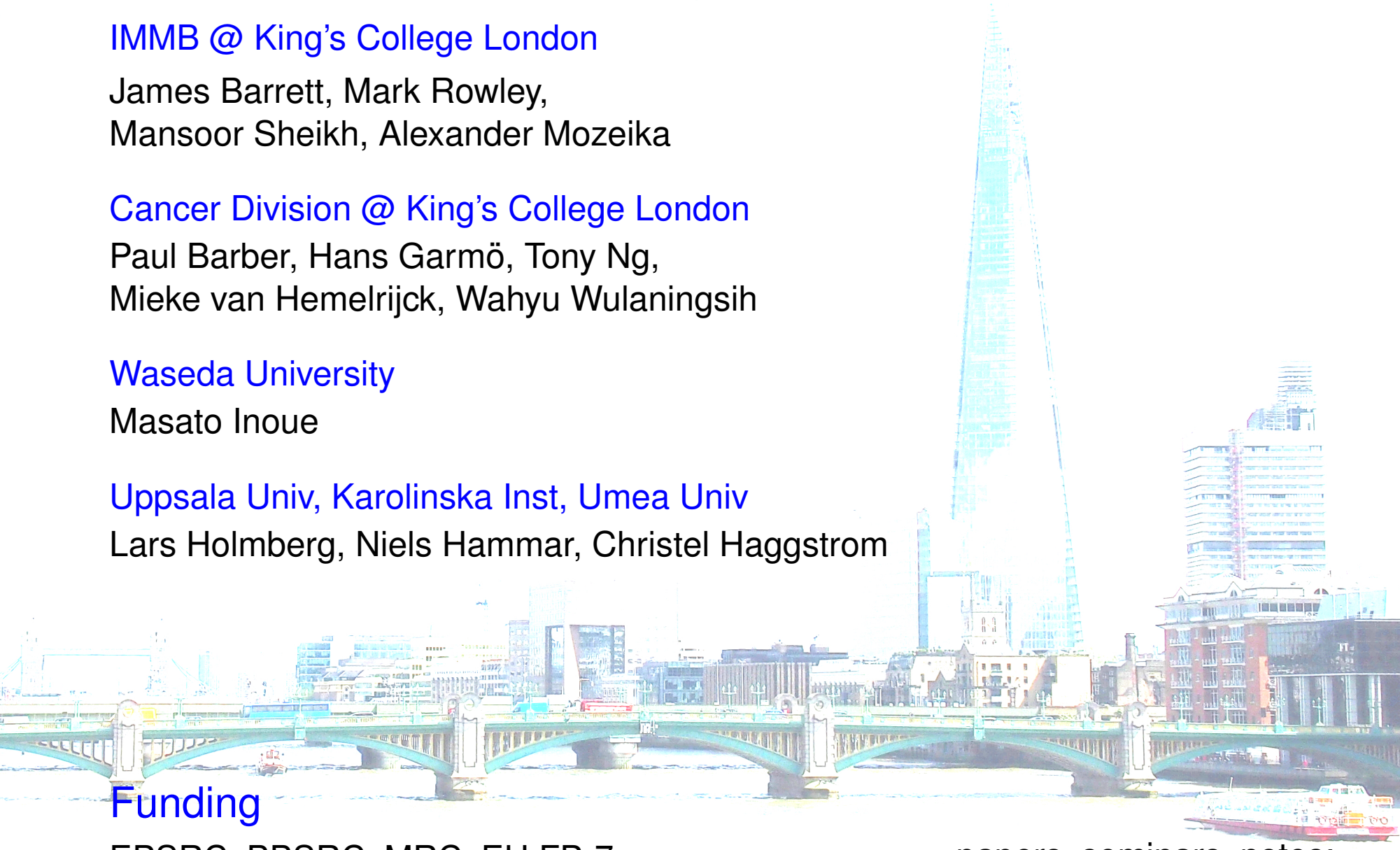
Paul Barber, Hans Garmö, Tony Ng,
Mieke van Hemelrijck, Wahyu Wulaningsih

Waseda University

Masato Inoue

Uppsala Univ, Karolinska Inst, Umea Univ

Lars Holmberg, Niels Hammar, Christel Haggstrom



Funding

EPSRC, BBSRC, MRC, EU FP-7,
Prostate UK, Ana Leaf Foundation

papers, seminars, notes:
https://toncoolen.wixsite.com/accc