Survival Analysis and Event Risk Prediction from Biomarkers

CS-E5890 - Statistical Genetics and Personalised Medicine

Aalto University

March 2, 2018

Tomi Peltola

tomi.peltola@aalto.fi

OUTLINE

OUTLINE

Analysis of time-to-event data

Examples from research at Aalto:

- time until depression diagnosis after giving birth,
- time until adverse cardiovascular event,
- time until cancer recurrence after treatment,
- graduation time after beginning of studies.

Survival analysis – example research questions

- Study the timing of events and change in the risk of events as a function of time.
- Study how subgroups of subjects differ in survival.
- Find associations of biomarkers measured at the beginning of the study to the outcomes.
- Predict the risk of the event based on biomarkers.

SURVIVAL ANALYSIS – SOME RESEARCH AT AALTO

Depression, depressive symptoms and treatments in women who have recently given birth: UK population study. Petersen, Peltola, Kaski, Walters, Hardoon. Submitted.

Hierarchical Bayesian Survival Analysis and Projective Covariate Selection in Cardiovascular Event Risk Prediction

Tomi Peltola Aki S. Havulinna Veikko Salomaa Aki Vehtari

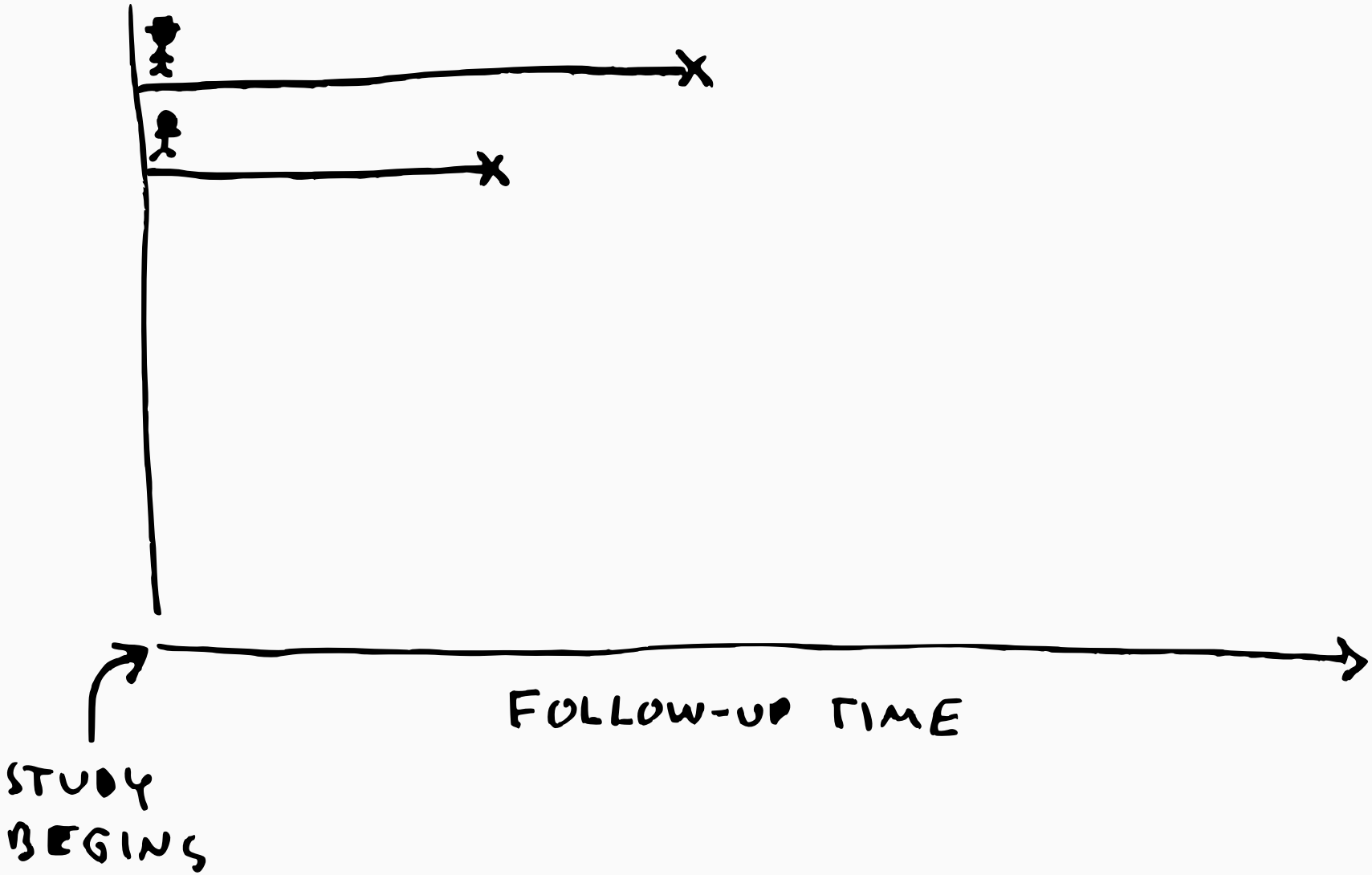
UAI Bayesian Bayesian Modeling Applications Workshop (BMAW 2014)

Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts

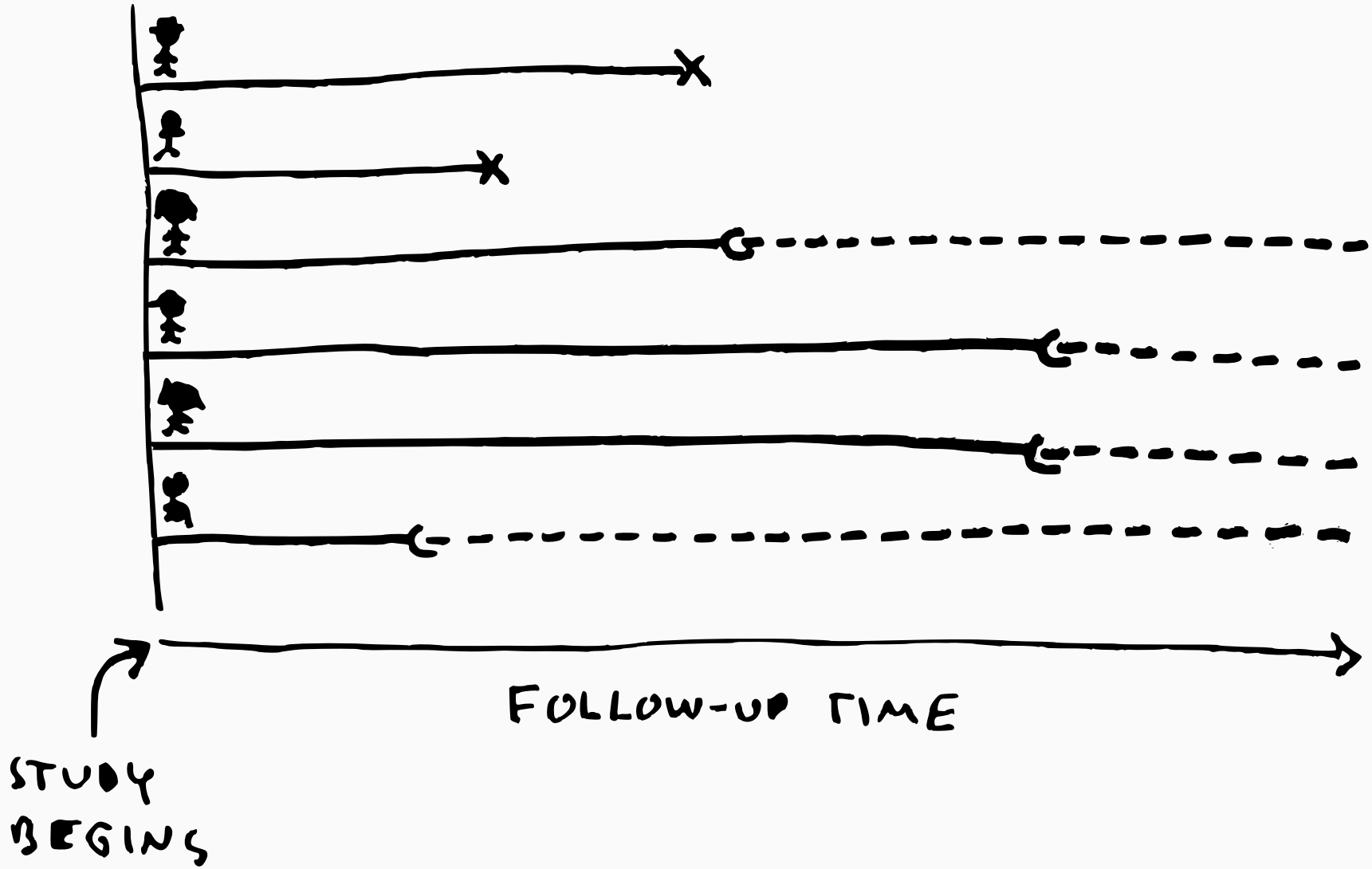
Heikki Joensuu, Aki Vehtari, Jaakko Riihimäki, Toshirou Nishida, Sonja E Steigen, Peter Brabec, Lukas Plank, Bengt Nilsson, Claudia Cirilli, Chiara Braconi, Andrea Bordoni, Magnus K Magnusson, Zdenek Linke, Jozef Sufliarsky, Massimo Federico, Jon G Jonasson, Angelo Paolo Dei Tos, Piotr Rutkowski

Lancet Oncol 2012; 13: 265–74

Time-to-event outcomes occur in **prospective study design**.

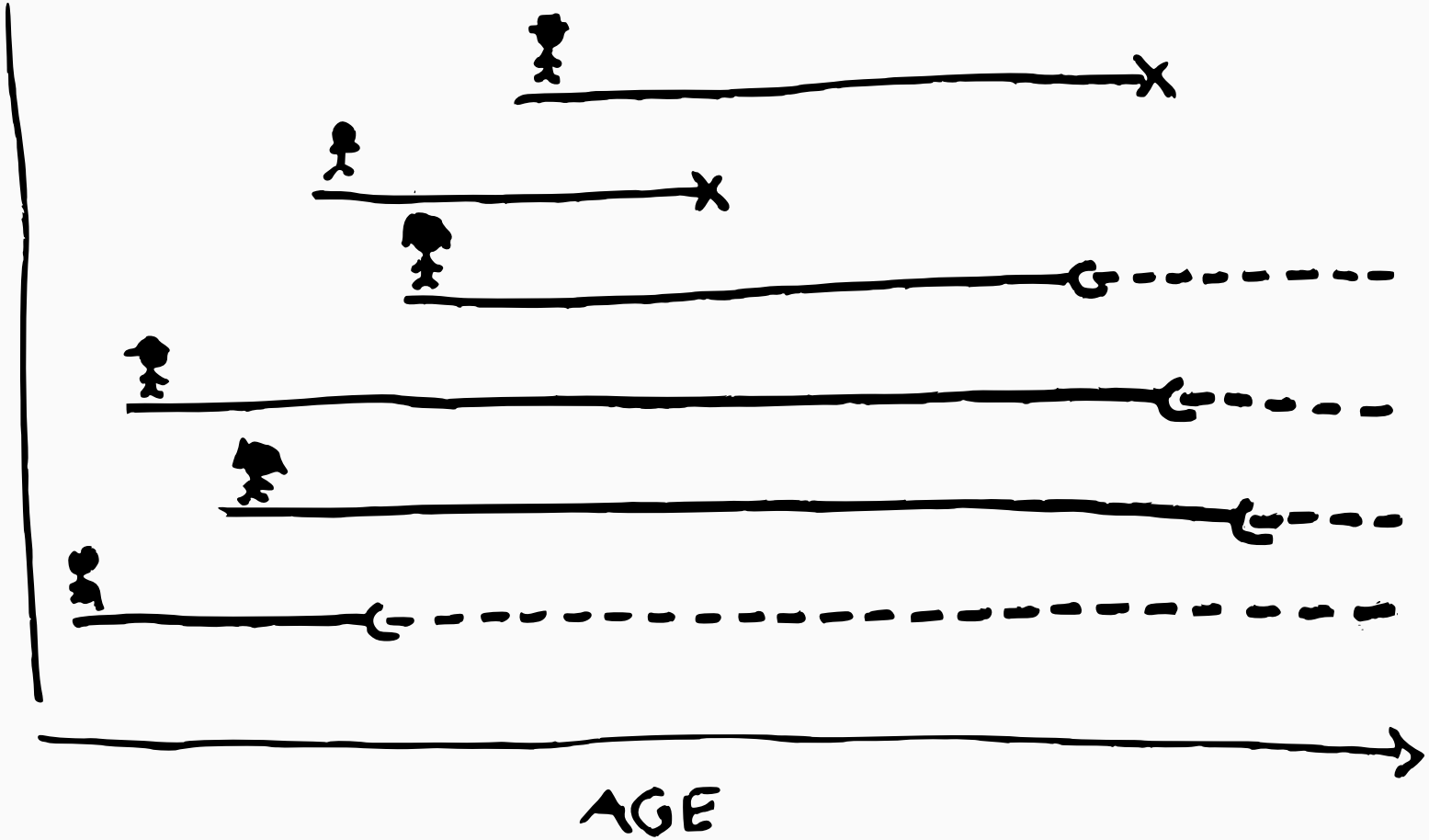


Complications: not everyone experiences the event.

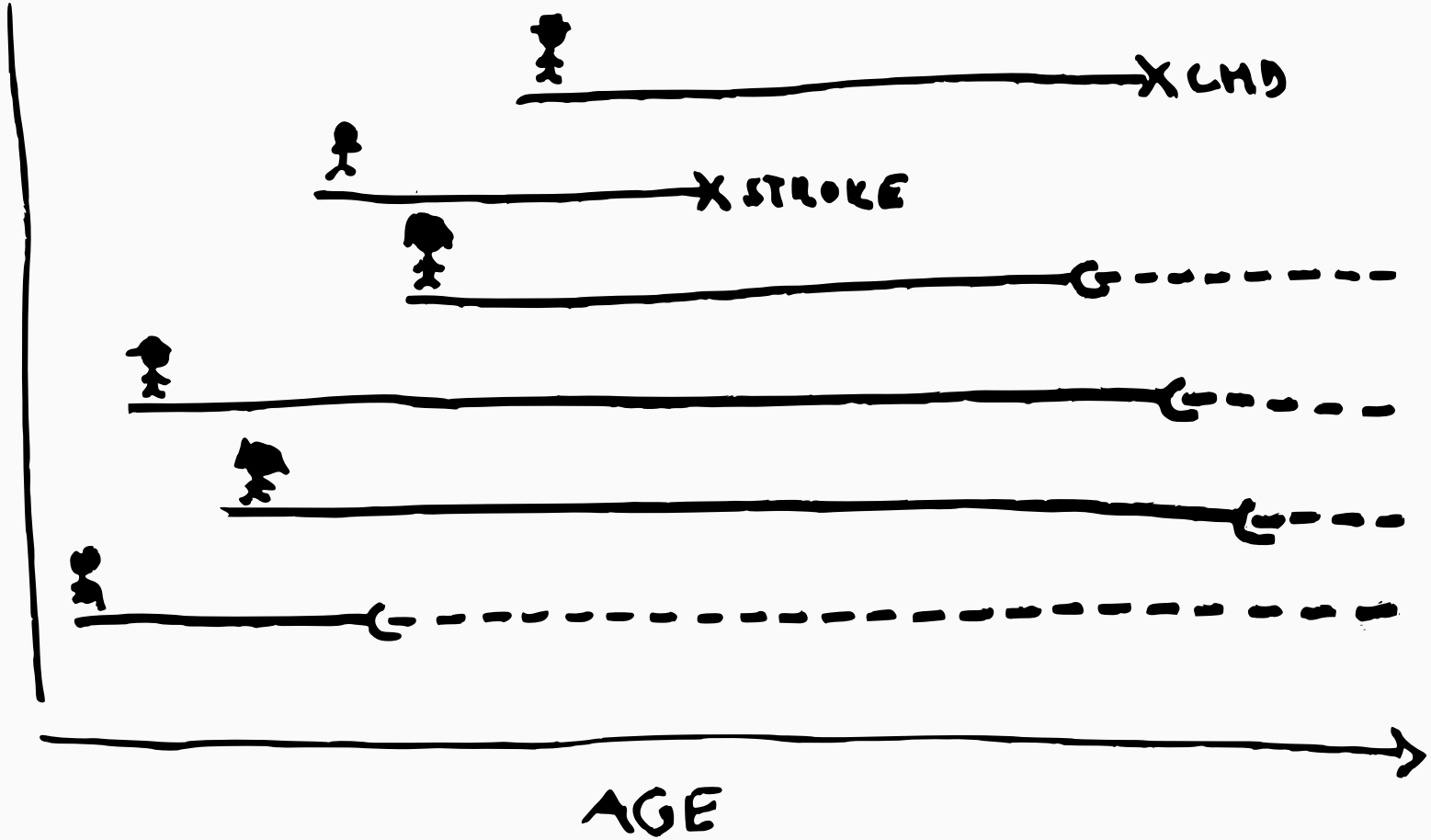


SURVIVAL ANALYSIS – CHOICE OF TIME AXIS

Complications: follow-up begins at different times for subjects.



Complications: multiple, related types of events.



OUTLINE

BASIC CONCEPTS IN SURVIVAL ANALYSIS

- Survival function
- Hazard function
- Censoring
- Truncation
- Kaplan–Meier estimator

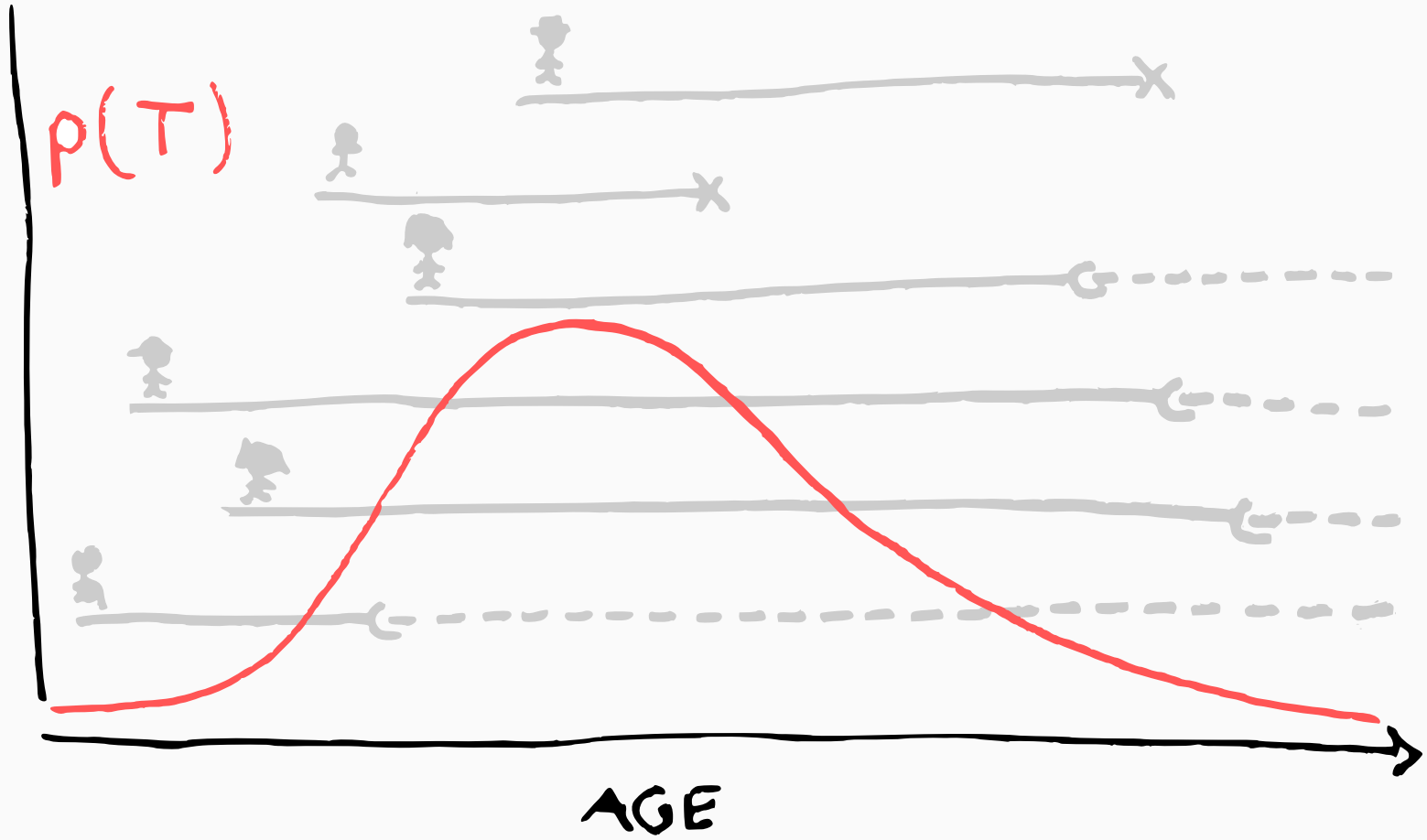
PROBABILITY MODELS – RECAP

Let $T \geq 0$ be a random variable denoting the event time.

- Probability density function $p_\theta(t)$.
- Cumulative distribution function: $F_\theta(t) = P_\theta(T \leq t) = \int_0^t p_\theta(s)ds$.

PROBABILITY MODELS – RECAP

Density function.



SURVIVAL FUNCTION $s(t)$

Let $T \geq 0$ be a random variable denoting the event time.

- Probability density function $p_\theta(t)$.
- Cumulative distribution function: $F_\theta(t) = P_\theta(T \leq t) = \int_0^t p_\theta(s)ds$.

The probability of surviving (at least) until t:

$$S_\theta(t) = P_\theta(T > t) = 1 - F_\theta(t)$$

HAZARD FUNCTION $h(t)$

Let $T \geq 0$ be a random variable denoting the event time.

- Probability density function $p_\theta(t)$.
- Cumulative distribution function: $F_\theta(t) = P_\theta(T \leq t) = \int_0^t p_\theta(s)ds$.

The probability of surviving (at least) until t:

$$S_\theta(t) = P_\theta(T > t) = 1 - F_\theta(t)$$

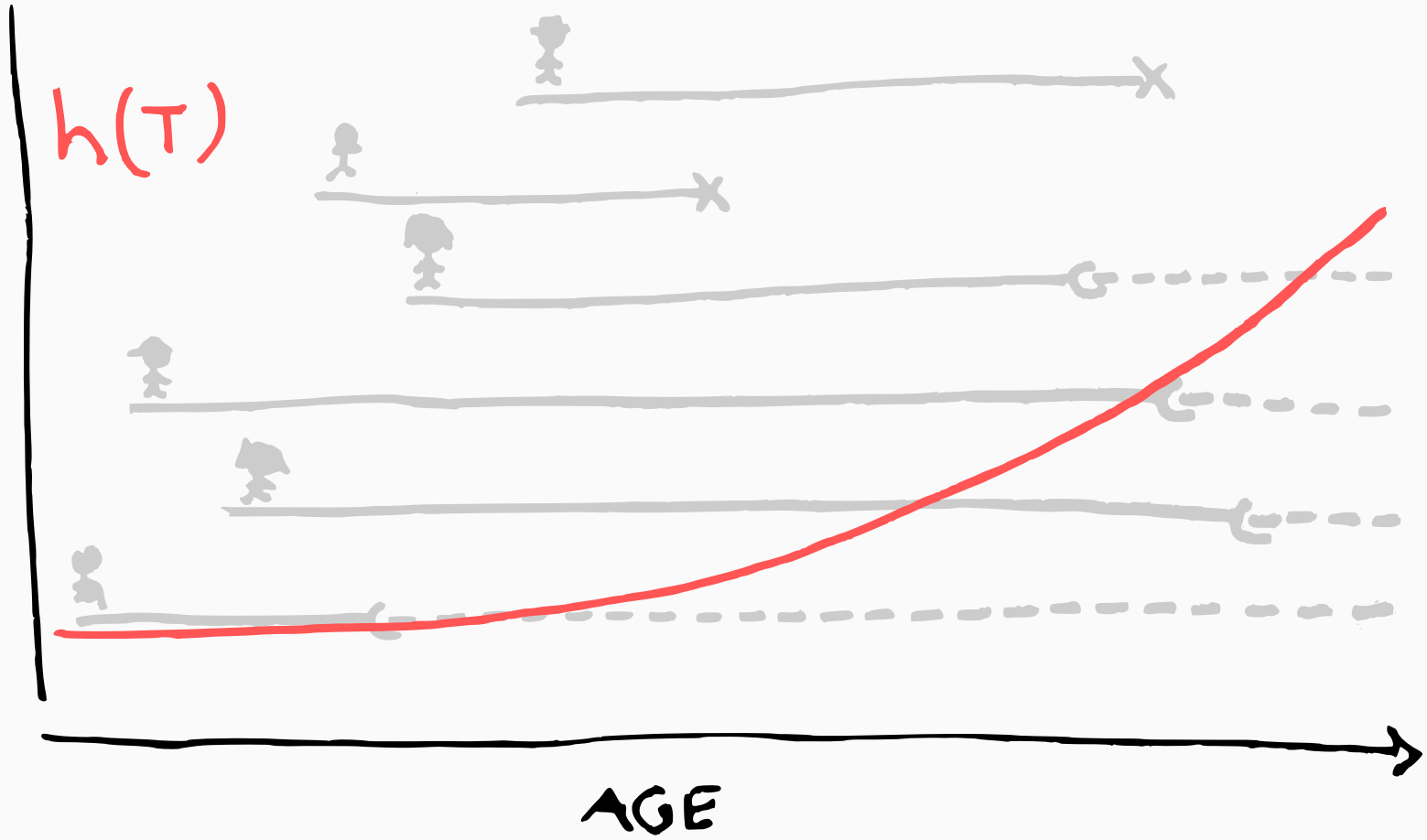
"Instantaneous event rate" at t having survived until t:

$$h(t) = \lim_{dt \to 0} \frac{P_\theta(t < T \leq t + dt \mid T > t)}{dt}$$

Cumulative hazard: $H(t) = \int_0^t h(s)ds$.

HAZARD FUNCTION $h(t)$

Hazard function – often more intuitive to think about than density.



RELATIONSHIPS BETWEEN THE FUNCTIONS

Fix one and others follow (assuming some conditions).

Basic relationships:

$$
\begin{aligned}
p_\theta(t) &= \frac{d}{dt} F_\theta(t), \\
S_\theta(t) &= 1 - F_\theta(t), \\
h_\theta(t) &= \frac{p_\theta(t)}{S_\theta(t)}.
\end{aligned}
$$

Another useful relationship:

$$
p_\theta(t) = h_\theta(t) S_\theta(t).
$$

ESTIMATION OF SURVIVAL FUNCTION – PARAMETRIC

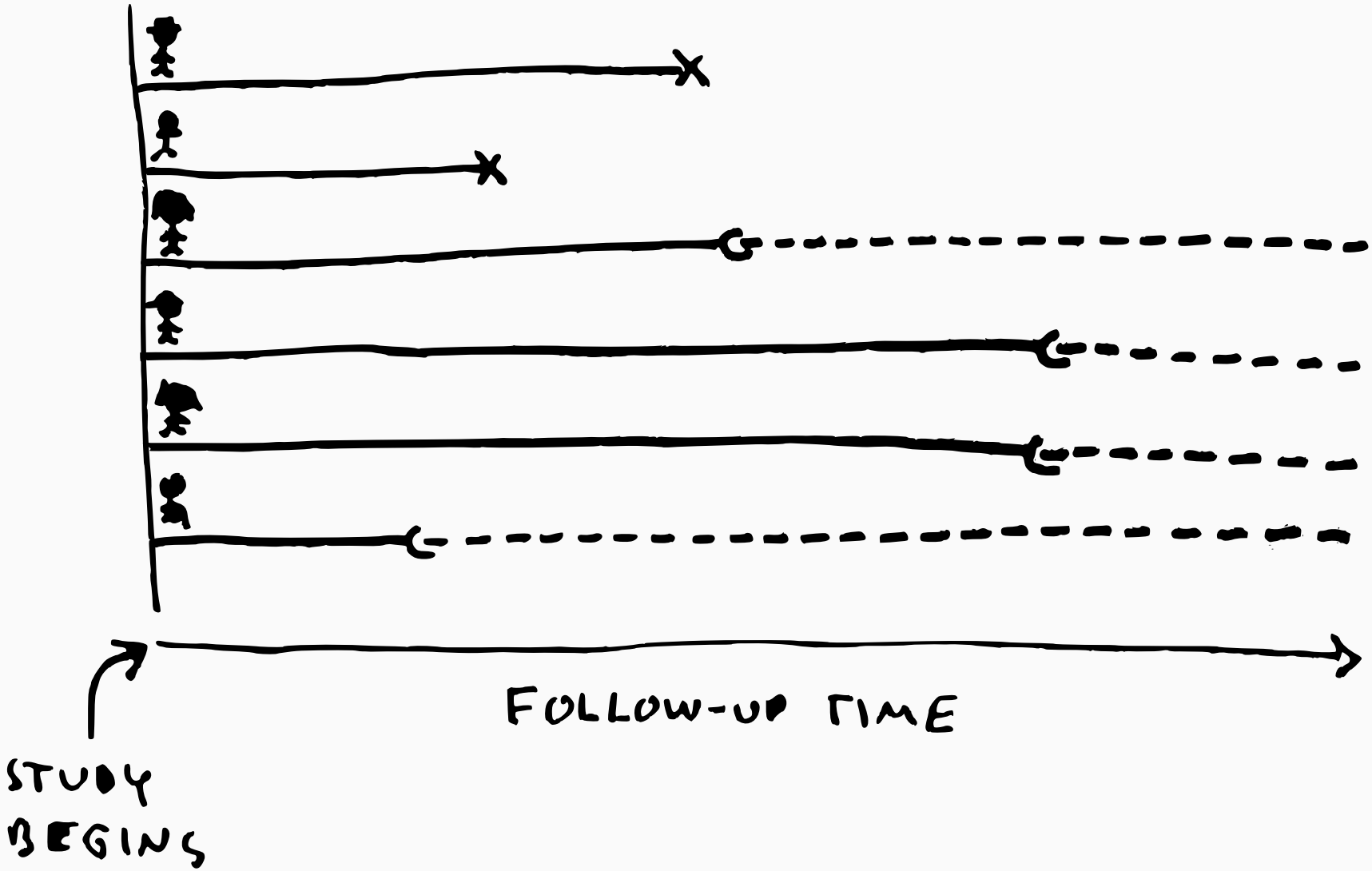
- Observed data of event times: t_i, $i = 1 \ldots, N$.
- Parametric assumption: $p_\theta(t)$.

The parameters θ can be estimated using, e.g., maximum likelihood (assuming iid):

$$\hat{\theta} = \arg\max_\theta L(\theta) = \arg\max_\theta \prod_i p_\theta(t_i) = \arg\max_\theta \sum_i \log p_\theta(t_i).$$

ESTIMATION OF SURVIVAL FUNCTION – CENSORING

Not everyone experiences the event: we don't know t_i for all i.



ESTIMATION OF SURVIVAL FUNCTION – PARAMETRIC

We don't know the event time t_i for all i.

But for such i, we do know that $t_i > c_i$, where c_i is a (right-)censoring time.

ESTIMATION OF SURVIVAL FUNCTION – PARAMETRIC

We don't know the event time t_i for all i.

But for such i, we do know that $t_i > c_i$, where c_i is a (right-)censoring time.

$$\hat{\theta} = \arg\max_{\theta} \prod_{i \in \mathcal{O}} p_\theta(t_i) \prod_{i \in \mathcal{C}} S_\theta(c_i) = \prod_i h_\theta(u_i)^{\delta_i} S_\theta(u_i),$$

where $u_i = \min(c_i, t_i)$ and $\delta_i = 1$ if $u_i = t_i$ and 0 otherwise.

ESTIMATION OF SURVIVAL FUNCTION – PARAMETRIC

We don't know the event time t_i for all i.

But for such i, we do know that $t_i > c_i$, where c_i is a (right-)censoring time.

$$\hat{\theta} = \arg\max_{\theta} \prod_{i \in \mathcal{O}} p_\theta(t_i) \prod_{i \in \mathcal{C}} S_\theta(c_i) = \prod_i h_\theta(u_i)^{\delta_i} S_\theta(u_i),$$

where $u_i = \min(c_i, t_i)$ and $\delta_i = 1$ if $u_i = t_i$ and 0 otherwise.

Practically, the outcome data is usually represented by pairs (u_i, δ_i).

(Note: above assumes non-informative censoring!)

ESTIMATION OF SURVIVAL FUNCTION – KAPLAN–MEIER ESTIMATOR

Simple and very popular counting based non-parametric estimator.

If there is no censoring (and no truncation):

$$\hat{S}(t) = \hat{P}(T > t) = \frac{\text{number of subjects alive after } t}{\text{total number of subjects in study}}$$

ESTIMATION OF SURVIVAL FUNCTION – KAPLAN–MEIER ESTIMATOR

Simple and very popular counting based non-parametric estimator.

If there is no censoring (and no truncation):

$$\hat{S}(t) = \hat{P}(T > t) = \frac{\text{number of subjects alive after } t}{\text{total number of subjects in study}}$$

However, with censoring we might not know the numerator.

ESTIMATION OF SURVIVAL FUNCTION – KAPLAN–MEIER ESTIMATOR

Simple and very popular counting based non-parametric estimator.

- Let $t_{(1)}, \ldots, t_{(N)}$ be sorted event times (smallest to largest) and assume $S(t)$ changes only at these.
- In particular, $P(T \geq t_{(j)}) = P(T > t_{(j-1)}) = S(t_{(j-1)})$.

Then,

$$\begin{aligned} S(t_{(j)}) &= P(T > t_{(j)}) = P(T > t_{(j)} \text{ and } T \geq t_{(j)}) \\ &= P(T \geq t_{(j)}) P(T > t_{(j)} \mid T \geq t_{(j)}) \\ &= P(T > t_{(j-1)}) P(T > t_{(j)} \mid T \geq t_{(j)}) \\ &= S(t_{(j-1)}) P(T > t_{(j)} \mid T \geq t_{(j)}). \end{aligned}$$

ESTIMATION OF SURVIVAL FUNCTION – KAPLAN–MEIER ESTIMATOR

Estimator:

$$\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)})\hat{P}(T > t_{(j)} \mid T \geq t_{(j)}) = \prod_{i=1}^{j} \hat{P}(T > t_{(i)} \mid T \geq t_{(i)}),$$

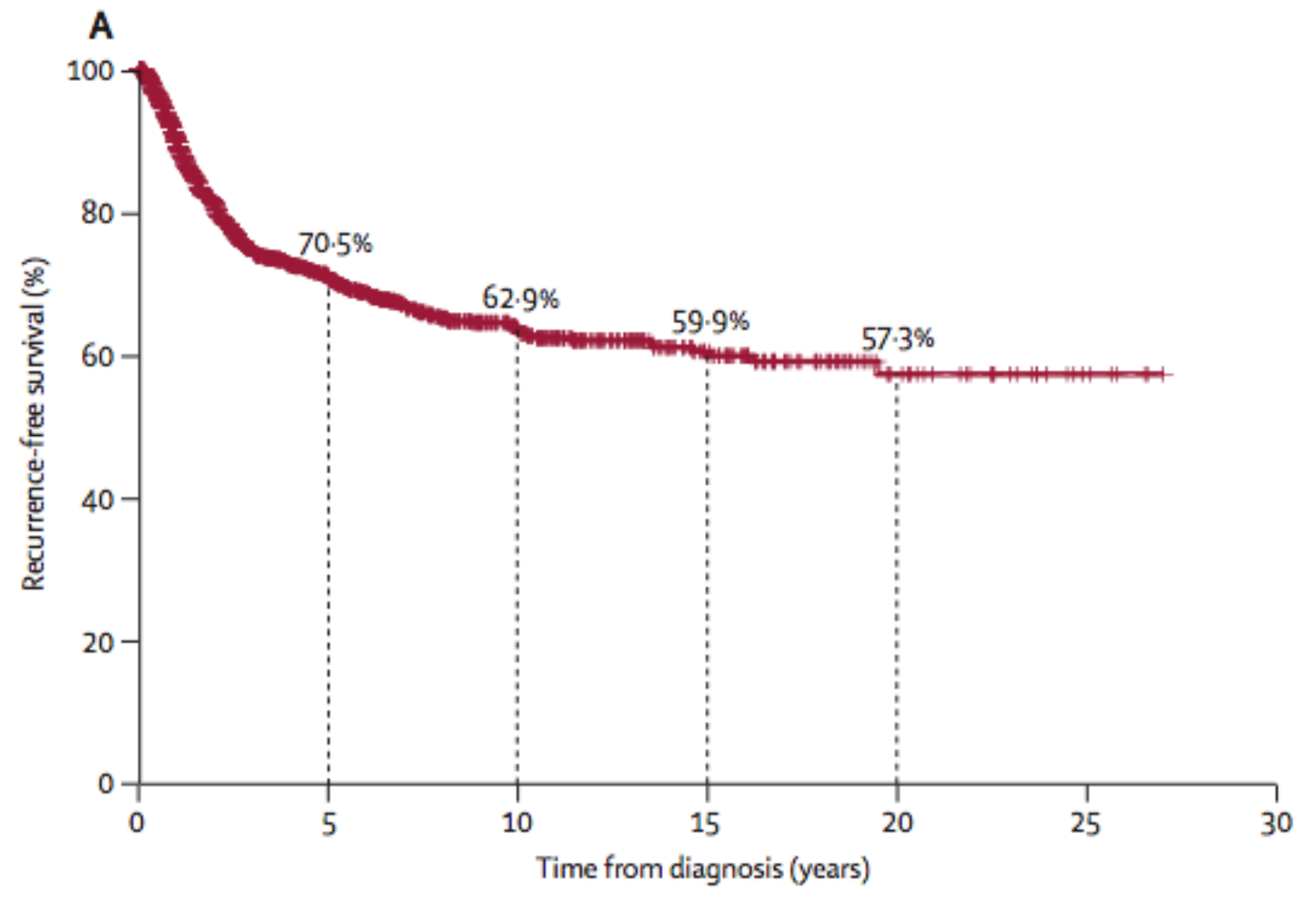
where each factor is

$$\hat{P}(T > t_{(i)} \mid T \geq t_{(i)}) = \frac{Y_i - d_i}{Y_i},$$

where Y_i is the number at risk and d_i the number of events at $t_{(i)}$.

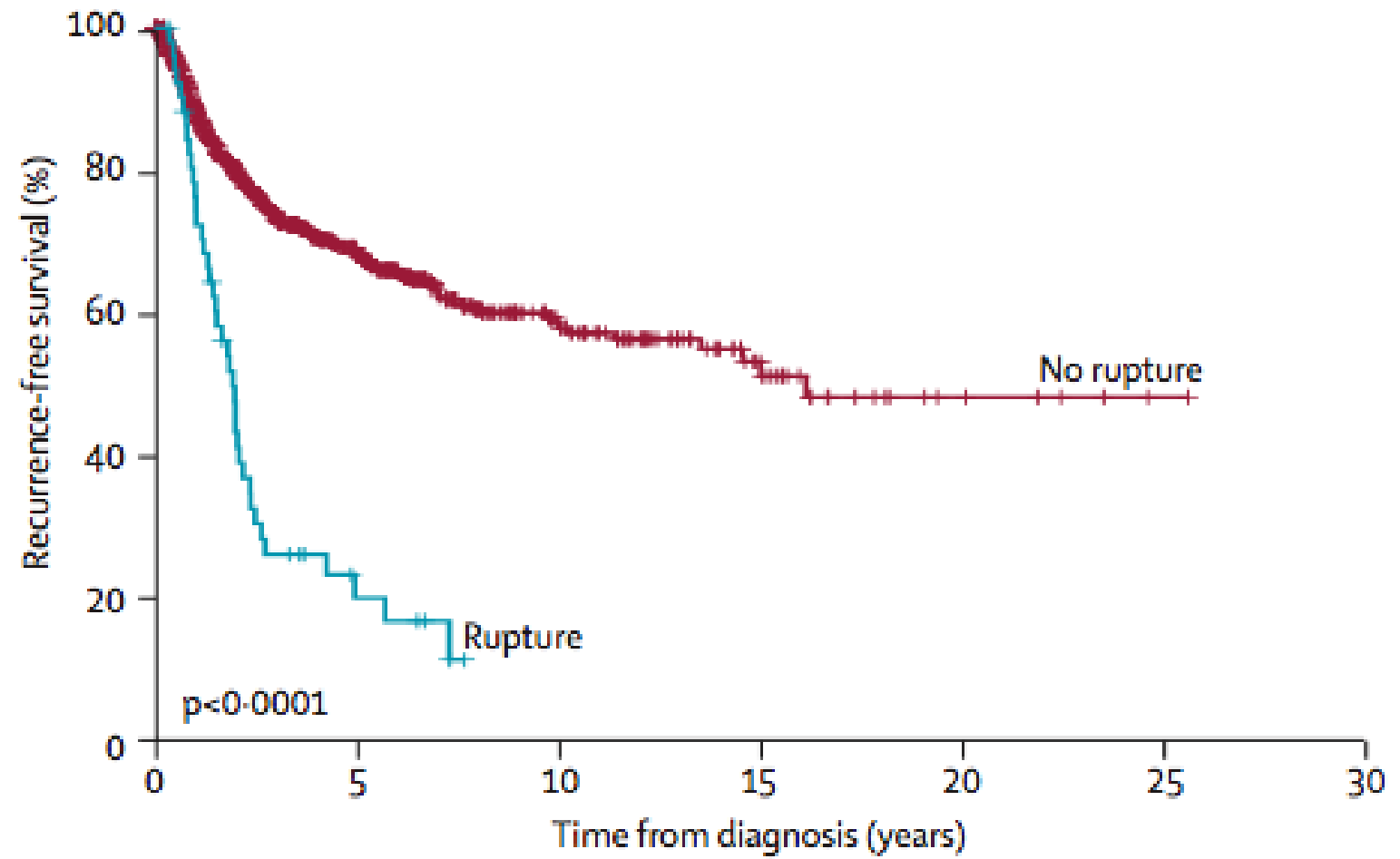
Gastrointestinal stromal tumour (GIST) after surgery recurrence, $N = 2560$

(Joensuu et al., 2012)



Gastrointestinal stromal tumour (GIST) after surgery recurrence, $N = 2560$

(Joensuu et al., 2012)



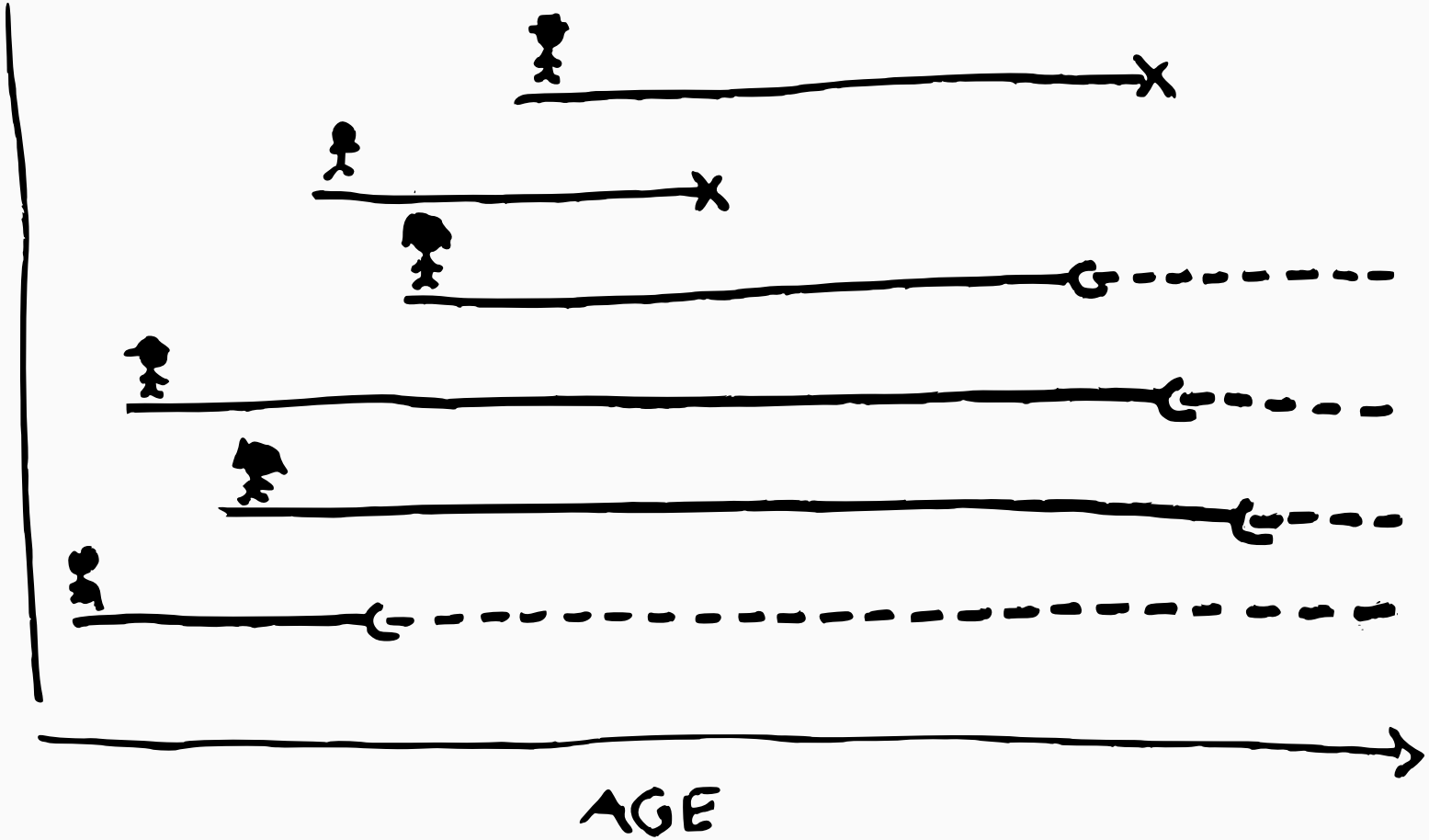
Estimation of survival function – Kaplan–Meier estimator

Remarks:

- Censoring is assumed non-informative.
- Kaplan–Meier estimator can be shown to be a non-parametric maximum likelihood estimator.
- `survival` R-package contains basic estimators.
- Log-rank test can be used to test differences between KM curves.
- Remember bias-variance tradeoff! Non-parametric estimator has larger variance but smaller bias than parametric estimator.

TRUNCATION

Left-truncation: follow-up begins at different times for subjects.



TRUNCATION

Left-truncation: follow-up begins at different times for subjects.

Let e be the entry time (start of the follow-up for a subject).

- In parametric models: $p_\theta(t \mid T \geq e) = \frac{p_\theta(t)}{S_\theta(e)}$.
- In non-parametric models: subjects enter risk sets at different times. (Some care required in interpretation.)

These assume that entry times e are not related to the risk of the event.

EXERCISE 1

Weibull distribution is

$$p(t) = \frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1} \exp(-(t/\lambda)^{\alpha}).$$

Calculate

- survival function $S(t) = \int_t^{\infty} p(s)ds = ?$,
- hazard function $h(t) = \frac{p(t)}{S(t)} = ?$.

What kind of shapes the hazard function takes for $\alpha = 1$, $\alpha > 1$, and $\alpha < 1$ ($\lambda = 1$)?

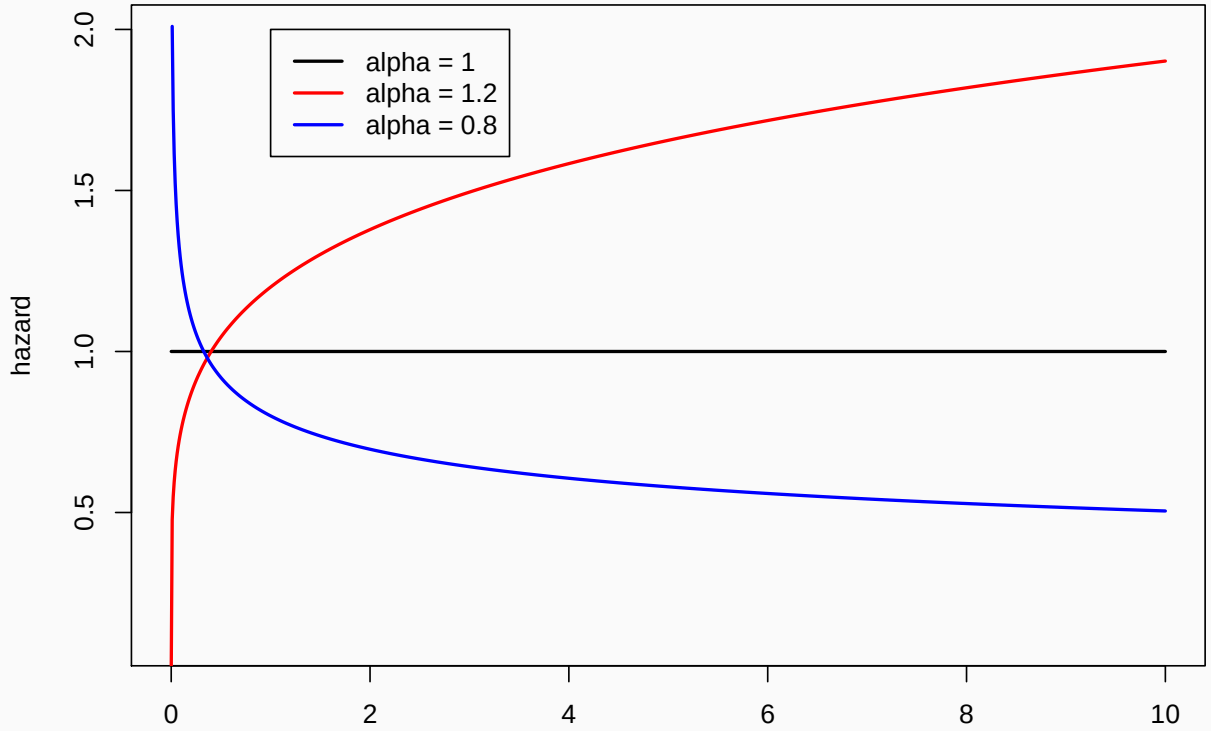
Exercise 1: solution

Weibull distribution is

$$p(t) = \frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1} \exp(-(t/\lambda)^{\alpha}).$$

Calculate

- survival function $S(t) = \int_t^{\infty} p(s)ds = \exp(-(t/\lambda)^{\alpha})$,
- hazard function $h(t) = \frac{p(t)}{S(t)} = \frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1}$.



OUTLINE

Example setting:

- N study subjects with M biomarkers measured from blood samples at the beginning of the study.
- Y years of follow-up for cardiovascular (CVD) events from hospital registries.
- Data for $i = 1, \ldots, N$:
  - u_i is event or censoring time,
  - δ_i is event indicator,
  - $\boldsymbol{x}_i$ is a vector containing the biomarker measurements.
- Goals:
  - Find which biomarkers are associated to the risk of CVD.
  - Predict the risk of CVD.

Regression modelling – proportional hazard (PH) model

Parametrize hazard with linear model:

$$h_\theta(t \mid \boldsymbol{x}_i) = h_{0,\alpha}(t) \exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{w}),$$

where

- $h_{0,\alpha}(t)$ is a baseline hazard function (same for all i),
- the biomarkers modify the hazard multiplicatively by $\exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{w})$.

REGRESSION MODELLING – PROPORTIONAL HAZARD (PH) MODEL

Parametrize hazard with linear model:

$$h_\theta(t \mid \boldsymbol{x}_i) = h_{0,\alpha}(t) \exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{w}),$$

where

- $h_{0,\alpha}(t)$ is a baseline hazard function (same for all i),
- the biomarkers modify the hazard multiplicatively by $\exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{w})$.

Proportional hazards assumption:

$$\frac{h_\theta(t \mid \boldsymbol{x}_i)}{h_\theta(t \mid \boldsymbol{x}_j)} = \exp((\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{w}).$$

Note: doesn't depend on t! The hazard ratio $exp(w_j)$ is the change in the hazard given a unit difference in jth covariate.

Maximum likelihood estimation:

1. Assume $h_{0,\alpha}(t)$ is a known function with parameter α.
   $\Rightarrow h_\theta(t \mid \boldsymbol{x}_i) = h_{0,\alpha}(t) \exp(\boldsymbol{x}_i^\mathrm{T} \boldsymbol{w})$ with $\theta = (\alpha, \boldsymbol{w})$.
2. Derive the corresponding survival function $S_\theta(t \mid \boldsymbol{x}_i)$.
3. The likelihood $L(\theta) = \prod_i h_\theta(t_i \mid \boldsymbol{x}_i)^{\delta_i} S_\theta(t_i \mid \boldsymbol{x}_i)$.
4. Find $\hat{\theta} = \arg\max_\theta \log L(\theta)$ using some optimization routine.

Tractability depends on the assumed parametric form.

Cox PH model avoids estimating the baseline hazard h_0.

Maximizes the partial likelihood (assuming no tied event times):

$$L(\boldsymbol{w}) = \prod_{i \in \mathcal{D}} \frac{h_\theta(t_i \mid \boldsymbol{x}_i)}{\sum_{j \in \mathcal{R}(t_i)} h_\theta(t_j \mid \boldsymbol{x}_j)} = \prod_{i \in \mathcal{D}} \frac{\exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{w})}{\sum_{j \in \mathcal{R}(t_i)} \exp(\boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{w})},$$

where

- $\mathcal{D}$ is the set of subjects that had event,
- $\mathcal{R}(t_i)$ is the set of subjects at risk at time t_i (alive just before t_i).

Cox PH model avoids estimating the baseline hazard h_0.

Maximizes the partial likelihood (assuming no tied event times):

$$L(\boldsymbol{w}) = \prod_{i \in \mathcal{D}} \frac{h_\theta(t_i \mid \boldsymbol{x}_i)}{\sum_{j \in \mathcal{R}(t_i)} h_\theta(t_j \mid \boldsymbol{x}_j)} = \prod_{i \in \mathcal{D}} \frac{\exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{w})}{\sum_{j \in \mathcal{R}(t_i)} \exp(\boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{w})},$$

where

- $\mathcal{D}$ is the set of subjects that had event,
- $\mathcal{R}(t_i)$ is the set of subjects at risk at time t_i (alive just before t_i).

Interpretation of the factors:

$$P(\text{individual dies at instant } t_i \mid \text{one death at instant } t_i)$$
$$= \frac{P(\text{individual dies at instant } t_i \mid \text{survival to } t_i)}{P(\text{one death at instant } t_i \mid \text{survival to } t_i)}.$$

REGRESSION MODELLING – ACCELERATED FAILURE TIME

Parametric, non-proportional hazard models.

1. Assume a parametric model $p_\theta(t)$ for event times.
2. Introduce covariates via

$$p_\theta(\exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w})t)|J|$$

(J is the Jacobian correction for the change of measure).

Time runs faster for those with high risk covariates.

Parametric, non-proportional hazard models.

1. Assume a parametric model $p_\theta(t)$ for event times.
2. Introduce covariates via

$$p_\theta(\exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w})t)|J|$$

(J is the Jacobian correction for the change of measure).

Time runs faster for those with high risk covariates.

These can also be presented as

$$\log T = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{w} + \epsilon,$$

with ϵ some noise model.

Remark: Weibull model is both an accelerated failure time model and proportional hazards model.

Could predict, e.g., mean or median survival time, but often not very interesting.

Rather predict survival or risk probabilities at chosen time t:

- given point estimate of the parameters $\hat{\theta}$: $S_{\hat{\theta}}(t \mid \boldsymbol{x}^*)$.
- given posterior distribution $p(\theta \mid \mathcal{D})$ of the parameters (or samples of it):

$$\mathbb{E}[S(t \mid \boldsymbol{x}^*)] = \int S_\theta(t \mid \boldsymbol{x}^*) p(\theta \mid \mathcal{D}) d\theta \approx \frac{1}{M} \sum_{m=1}^{M} S_{\theta^{(m)}}(t \mid \boldsymbol{x}^*),$$

  where $\theta^{(m)}, m = 1, \ldots, M$ are samples from the posterior.

Note: predictions require estimate of the baseline hazard. There are ways to do this for the Cox model.

Remarks:

- Assumed non-informative censoring.
- Exploring covariate associations (e.g., hazard ratios) to the event doesn't require the baseline hazard. There are significance tests for the hazard ratios.
- There are tests for the PH assumption, and quality and calibration measures for the survival probability predictions, e.g., C-index.
- Remember bias-variance tradeoff! Non-parametric estimator has larger variance but smaller bias than parametric estimator.

EXERCISE 2

Weibull hazard is (with scale $\lambda = 1$)

$$h_{0,\alpha}(t) = \alpha t^{\alpha-1}.$$

Introduce covariate effects by multiplying $h_{0,\alpha}(t)$ with $\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{w})$.

Derive the survival function

$$S_\theta(t \mid \boldsymbol{x}) = \exp(-\int_0^t h(s)ds) = ?,$$

and write down the likelihood for

- a (right-)censored observation at t:

$$L_\theta(t) = S_\theta(t \mid \boldsymbol{x}_i) = ?,$$

- an observed event at t:

$$L_\theta(t) = h_\theta(t \mid \boldsymbol{x})S_\theta(t \mid \boldsymbol{x}) = ?.$$

EXERCISE 2: SOLUTION

Weibull hazard is

$$h_{0,\alpha}(t) = \alpha t^{\alpha-1}.$$

Introduce covariate effects by multiplying $h_{0,\alpha}(t)$ with $\exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w})$.

Derive the survival function

$$S_\theta(t \mid \boldsymbol{x}) = \exp(-\int_0^t h(s)ds) = \exp\left(-\exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w})t^\alpha\right),$$

and write down the likelihood for

- a (right-)censored observation at t:

$$L_\theta(t) = S_\theta(t \mid \boldsymbol{x}) = \exp\left(-\exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w})t^\alpha\right),$$

- an observed event at t:

$$L_\theta(t) = h_\theta(t \mid \boldsymbol{x})S_\theta(t \mid \boldsymbol{x}) = \alpha t^{\alpha-1}\exp\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w} - \exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w})t^\alpha\right).$$

OUTLINE

Hierarchical Bayesian Survival Analysis and Projective Covariate Selection in Cardiovascular Event Risk Prediction

Tomi Peltola Aki S. Havulinna Veikko Salomaa Aki Vehtari

UAI Bayesian Bayesian Modeling Applications Workshop (BMAW 2014)

Setting:

- FINRISK 1997 study by National Institute for Health and Welfare (THL): random samples of approx. 10,000 Finnish adults.
- Biomarker measurements from blood samples collected at the beginning of the study.
- 15 years of follow-up from hospital registries for cardiovascular events.
- Here, focus on 401 diabetic individuals with 55 candidate biomarkers.

BIOMARKER BASED RISK PREDICTION – IDENTIFYING BIOMARKERS

Model:

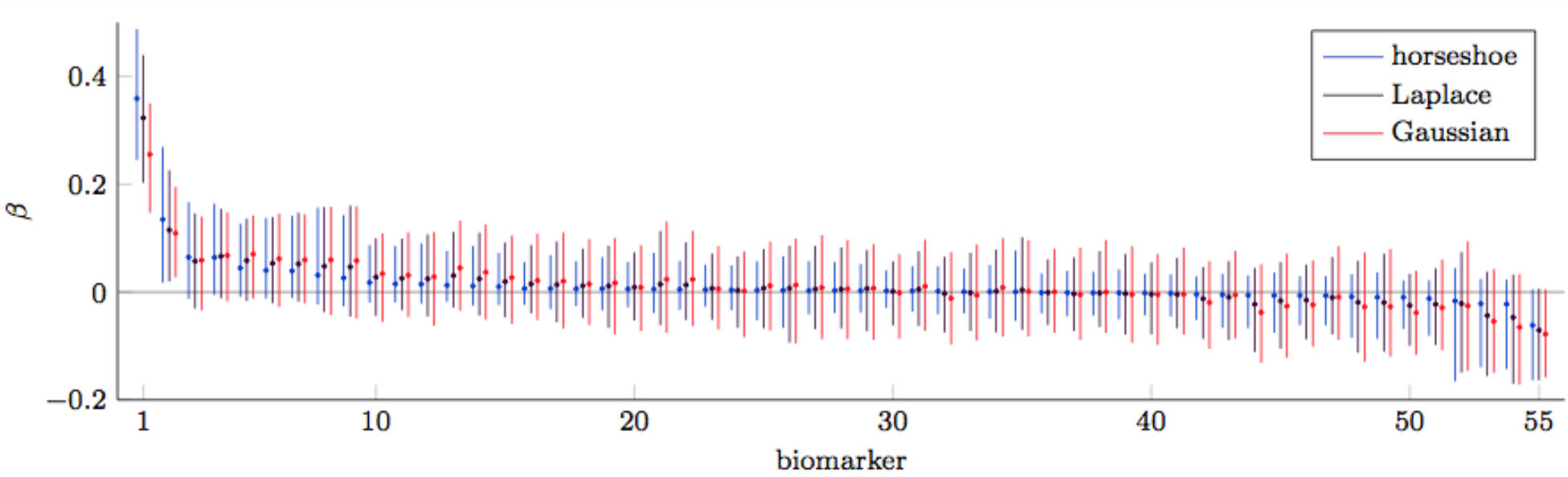
- Bayesian linear survival model,
- Weibull observation model,
- shrinkage priors (and transfer learning),
- implemented in Stan probabilistic programming language.

```
model {
  // observed event times t_obs with covariates X_obs:
  t_obs ~ weibull(alpha, exp(mu + X_obs * w));
  // censoring times t_cen with covariates X_cen:
  target += weibull_lccdf(t_cen | alpha, exp(mu + X_cen * w));

  // horseshoe prior on regression weights:
  w ~ horseshoe_prior(0, 1); // see full impl. for actual code
  // ...
}
```

See https://github.com/to-mi/stan-survival-shrinkage for full implementation.

BIOMARKER BASED RISK PREDICTION – IDENTIFYING BIOMARKERS



- Interesting biomarkers could be identified by weights that are far from zero,
- but we applied covariate selection via projection methods in the paper (Juho Piironen and Aki Vehtari have interesting recent research on these and the horseshoe shrinkage prior).

Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts

Heikki Joensuu, Aki Vehtari, Jaakko Riihimäki, Toshirou Nishida, Sonja E Steigen, Peter Brabec, Lukas Plank, Bengt Nilsson, Claudia Cirilli, Chiara Braconi, Andrea Bordoni, Magnus K Magnusson, Zdenek Linke, Jozef Sufliarsky, Massimo Federico, Jon G Jonasson, Angelo Paolo Dei Tos, Piotr Rutkowski

Lancet Oncol 2012; 13: 265–74

- Recurrence of gastrointestinal stromal tumour (GIST) after surgery.
- Pooled set of 2,560 patients from various studies.
- Biomarkers: tumour size, mitosis count, location, presence of rupture.

BIOMARKER BASED RISK PREDICTION – CANCER RECURRENCE RISK

Non-linear Bayesian proportional hazards model using Gaussian processes:

$$h_\theta(t \mid \boldsymbol{x}_i) = h_{0,\alpha}(t) \exp(f_i(\boldsymbol{x}_i)),$$

where

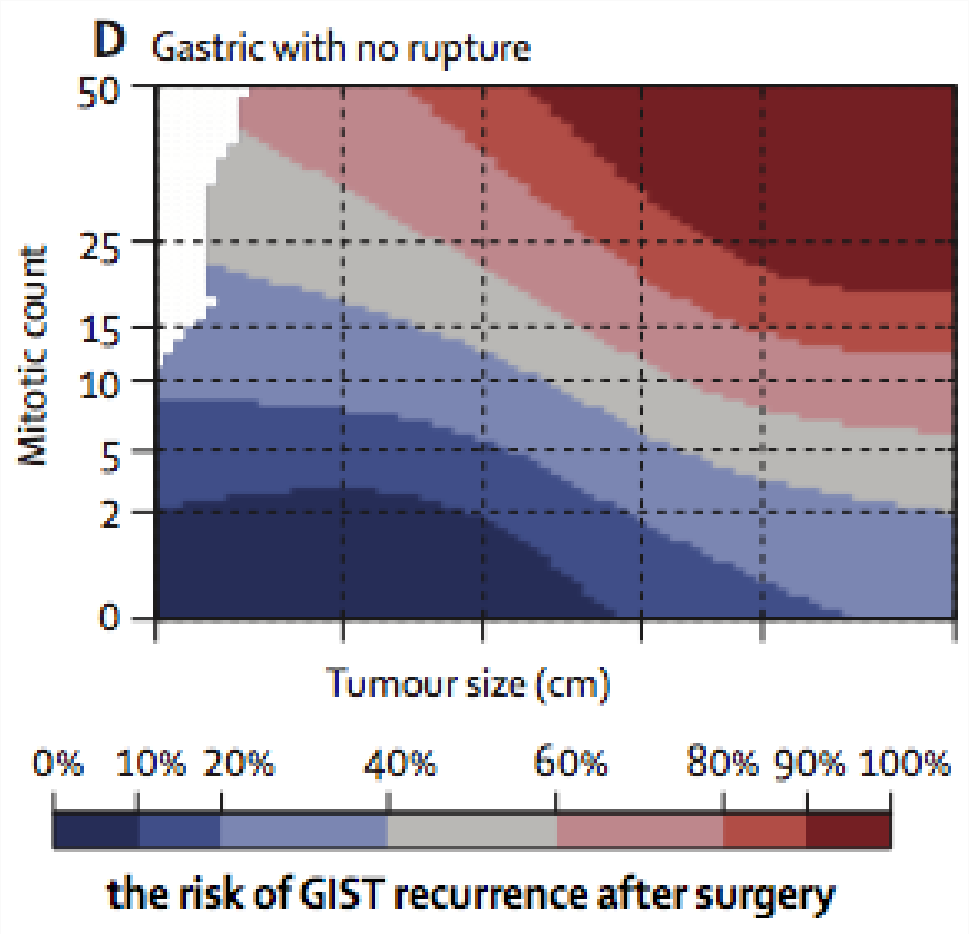
- baseline hazard $h_{0,\alpha}(t)$ is piece-wise constant in M time intervals,
- $\log \boldsymbol{\alpha} \sim N(\mathbf{0}, \boldsymbol{K}_t)$, where $\alpha_j = h_{0,\alpha}(t)$ in jth time interval,
- $\boldsymbol{f} = [f_1, \ldots, f_N]^{\mathrm{T}} \sim N(\mathbf{0}, \boldsymbol{K}_{\boldsymbol{x}})$.

Implemented in GPstuff (Matlab toolbox):
http://research.cs.aalto.fi/pml/software/gpstuff/.

(GP 1D regression demo: http://www.tmpl.fi/gp/.)

BIOMARKER BASED RISK PREDICTION – CANCER RECURRENCE RISK

Contour map shows risk of recurrence within 10 years after surgery based on mitosis count and tumour size:



GIST risk calculator: http://gistrisk.com.

OUTLINE

SUMMARY

Survival analysis = Analysis of time-to-event data

Important concepts:

- Survival and hazard functions
- Censoring and truncation
- Kaplan–Meier estimator
- Cox proportional hazard regression

Important applications and goals in (personalized) medicine:

- Prospective cohort studies and clinical trials,
- Electronic health record data,
- Biomarker identification,
- Personalized risk prediction.

SUMMARY

Lots of things not mentioned: multivariate and competing outcomes, joint analysis with longitudinal data, censoring models, frailty and cure-rate models, time-varying coefficients and covariates, model criticism and checking...

References (books): Kleinbaum: Survival Analysis: A Self-Learning Text; Klein and Moeschberger: Survival Analysis: Techniques for Censored and Truncated Data, 2nd Edition.

Questions?

no stick figures were harmed
in the making of this presentation