Multistate models to improve decision-making and effect quantification in clinical trials

Kaspar Rufibach
Methods, Collaboration & Outreach Group, Department of Biostatistics, Roche Basel
27th June 2019, Shanghai Biostatistics Forum

Generated 2019-06-27 at 06:01:09.

Acknowledgments

Survival prediction for early decision-making:

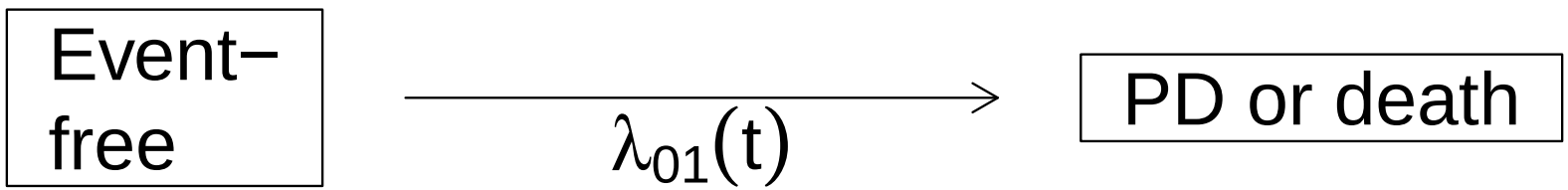
- Joint work with **Ulrich Beyer**, David Dejardin, Matthias Meller, Hans Ulrich Burger.
- Beyer et al. (2019), incl. R code and code documentation.

Agenda

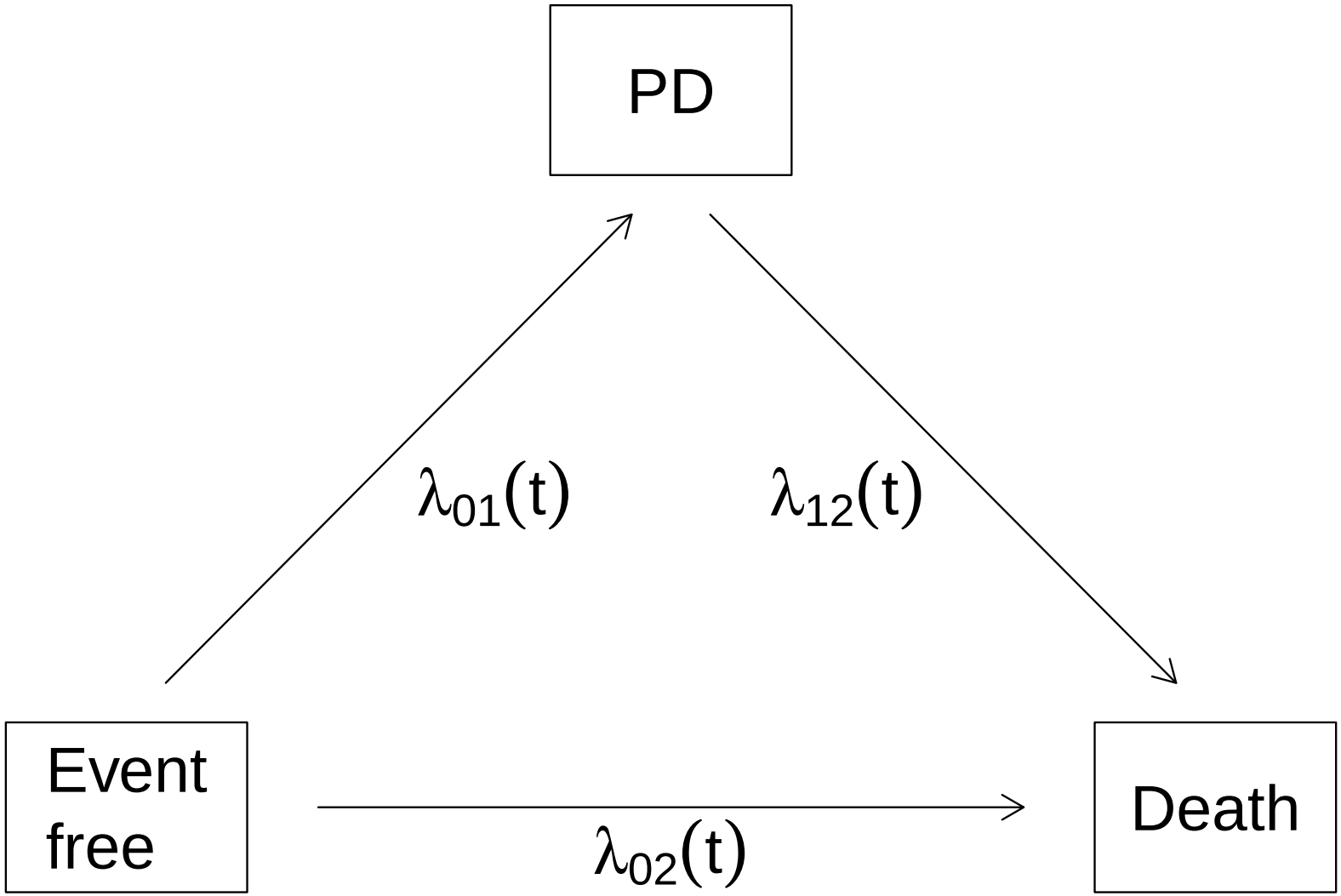
1. Multistate models
2. Survival prediction for early decision-making
3. Effect quantification for non-proportional hazards
4. Conclusions

Multistate models

Canonical extension of survival analysis



Canonical extension of survival analysis



Multistate models

Multistate model:

- 1-1 correspondence **hazard - probability** breaks down.
- Transition probabilities: (Markov) process $X(t)_{t\geq 0}$ with state space $\{0,1,2\} = \{\text{event-free, progression, death}\}$. Then,

$$P_{lj}(s,t) \quad := \quad P(X_t = j | X_s = l, \text{Past}).$$

- Estimate P_{lj}'s **nonparametrically** by **Aalen-Johansen** estimator.
- PFS: Kaplan-Meier of time-to-progression simply censoring death is **biased**!
- OS: Aalen-Johansen offers **higher precision** compared to simple Kaplan-Meier estimate, Andersen et al. (1993) (p. 315 and Fig. IV.4.16).
- Markov assumption **stronger** than what is needed for Kaplan-Meier though.

Prediction in multistate models

Rates (hazards, intensities):

- Modelling of effects of covariates on **transition hazards**.
- Hazard ratios (HR) from Cox regression.

Transition probabilities look at **cumulative effects**:

- Effects on transition probabilities may be different from what HRs suggest.
- **Intermediate** events in multistate model also contribute to cumulative effects.
- How to estimate such cumulative effects?

Prediction from multistate model!

Prediction in multistate models

General problem: estimate **conditional probability of some future clinical event**, given

- event history,
- set of values for prognostic factors of a patient.

Derive formulas for these conditional probabilities, or simulate.

Final result: survival function for OS, as function of

- covariates and
- relevant **cumulative hazards**.

Survival prediction for early decision-making

Decision-making in early oncology development

Contemporary decision-making in early oncology development:

- Single-arm Phase 1b trial for **experimental** drug with, e.g., 40 patients.
- Compute proportions of complete and overall response, duration of response.
- Compare proportions to "corresponding" quantities from literature for **control** treatment.

Meaningful PFS / OS typically not available!

Challenges and proposal

Endpoint in Phase 3 will be time-to-event, e.g. OS.

1. Response-type endpoint **meaningful for interpretation**?
2. Surrogacy? Meta-analyses **IF** data available. **Surrogacy poor** in many indications.
3. Immunotherapy (CIT): response proportions similar between experimental and control, but relevant OS effect.
4. **Non-randomized** comparison $\Rightarrow$ selection bias.

Proposal: Base decision-making on **OS prediction from multistate model**.

1. Predicted OS survival function for experimental arm, $S_{\mathbf{exp}}$, is what we are interested in.
2. Combine S_{exp} with S_{control} to get **predicted OS HR** based on multistate model.
3. Experimental drug might act on certain transitions only $\Rightarrow$ not captured through simple modelling of OS. Potential **efficiency gain**!
4. **Propensity scoring**.

Goal of this talk

Feasibility and usefulness of multistate model:

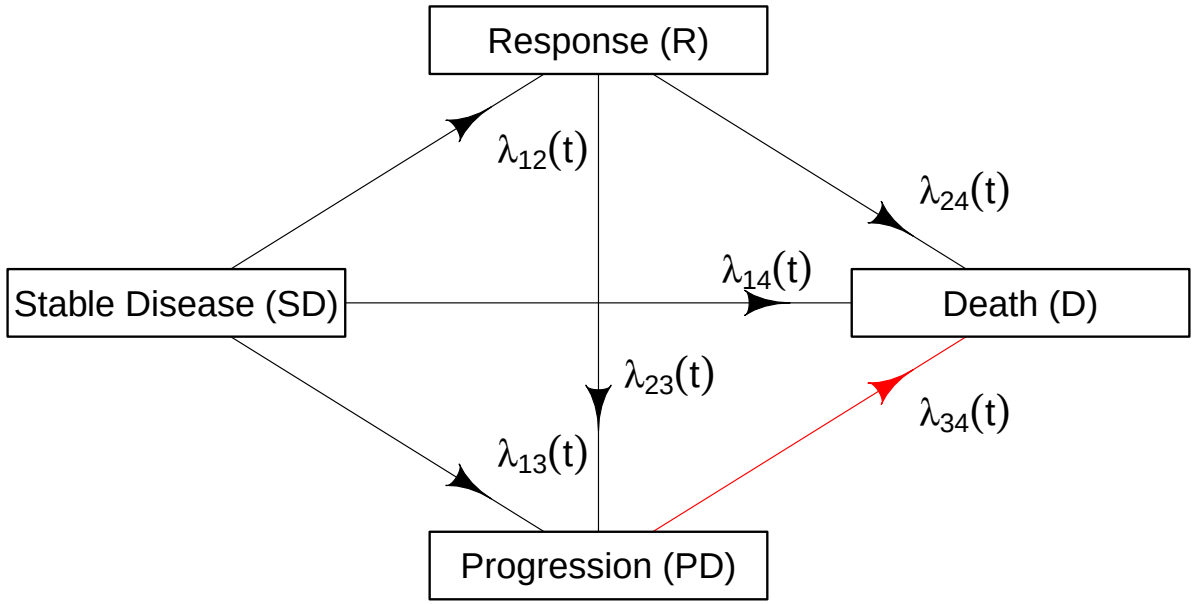
- **Idealized** scenario using retrospective data from Phase 3 RCTs.
- We have **long-term** follow up in both arms.
- **Control arm mimics historical control**.
- Randomization $\Rightarrow$ no selection bias.

If multistate model approach should be useful $\Rightarrow$ has to work in this **idealized** scenario.

Selection bias taken care of later.

Multistate model for early decision-making

- States: stable disease (SD), response (R), progression (PD), and death (D).
- States and transitions hazards λ_{ij}: **imposed model** for disease mechanism.



Early phase studies:

- Follow up of patient until PD or death without progression.
- **Post-progression data** very limited.
- Post-progression hazard λ_{34}: **Assumption** or **borrowing** from historical data.
- Transitions $1 \to 4, 2 \to 4$ rare, hazards $\approx$ same in both arms.

Computation of S_{exp}

Compute transition probabilities for each transition. Function of cumulative hazards.

S_{exp} = sum of transition probabilities that end in death:

$$S_{\text{exp}}(t) = 1 - \Big(P_{SD\to D}(0,t) + P_{SD\to \mathbf{PD}\to \mathbf{D}}(0,t) + P_{SD\to R\to D}(0,t) + P_{SD\to R\to \mathbf{PD}\to \mathbf{D}}(0,t)\Big).$$

λ_{34} corresponding to **PD** $\to$ **D** transition borrowed from historical data.

Historical borrowing for λ_{34}

Scenario 1: λ_{34} **completely borrowed** from historical control.

- Experimental treatment not expected to provide benefit beyond PD, e.g. antibody-dependent cellular cytotoxicity or chemotherapy.
- **Plug-in** hazard function estimate from historical control $\Rightarrow$ no post-PD information required for experimental arm.

Scenario 2: λ_{34} **proportional** to post-PD hazard from historical control.

- Experimental treatment expected to provide benefit beyond PD, e.g. comparing CIT with chemo or ADCC.
- How much post-PD deaths needed in experimental arm to reliably **estimate post-PD HR**?

Example 1: Cleopatra

Cleopatra

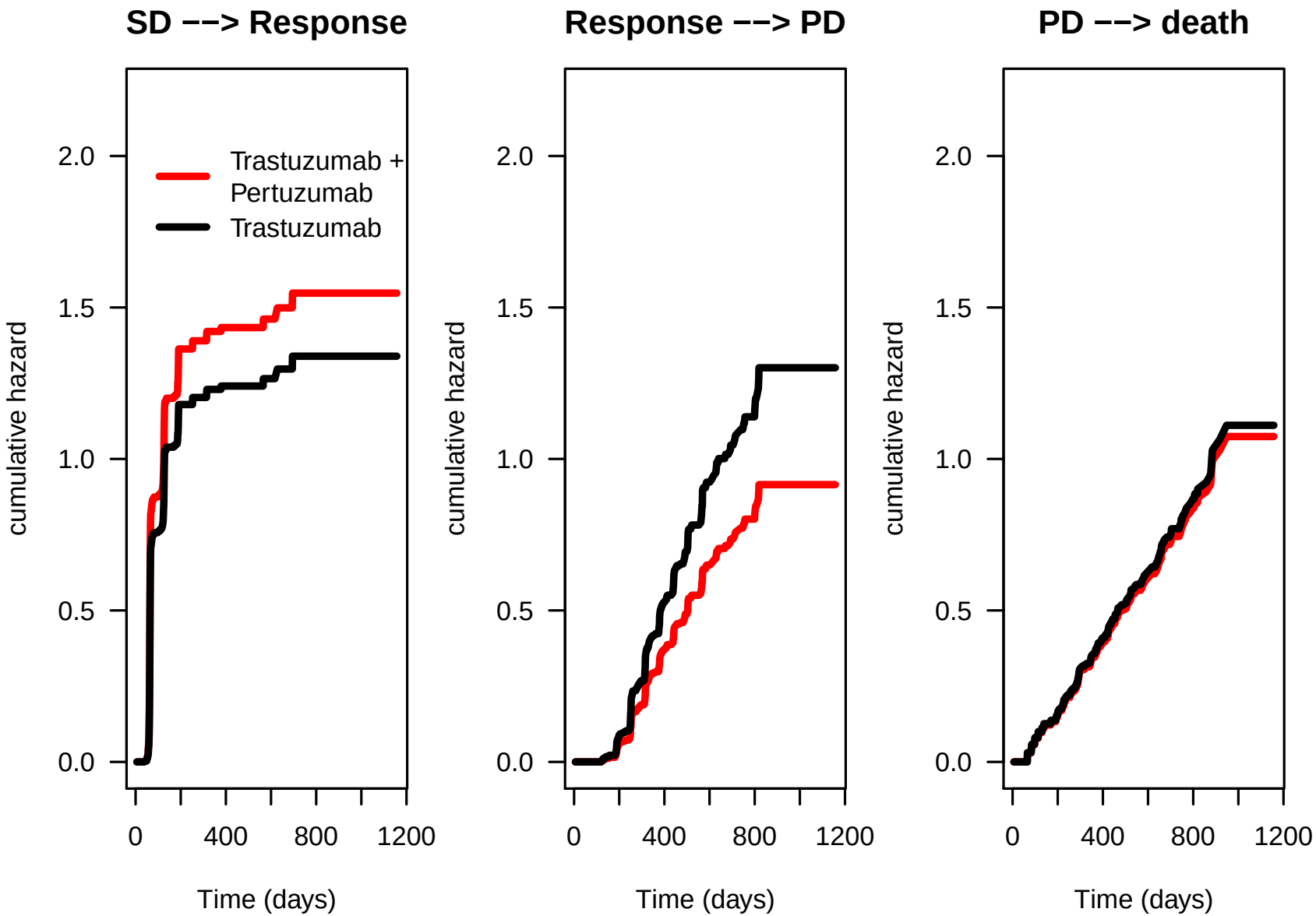
Baselga and Cortes (2012), Swain and Baselga (2015).

Previously untreated HER2-positive metastatic breast cancer patients.

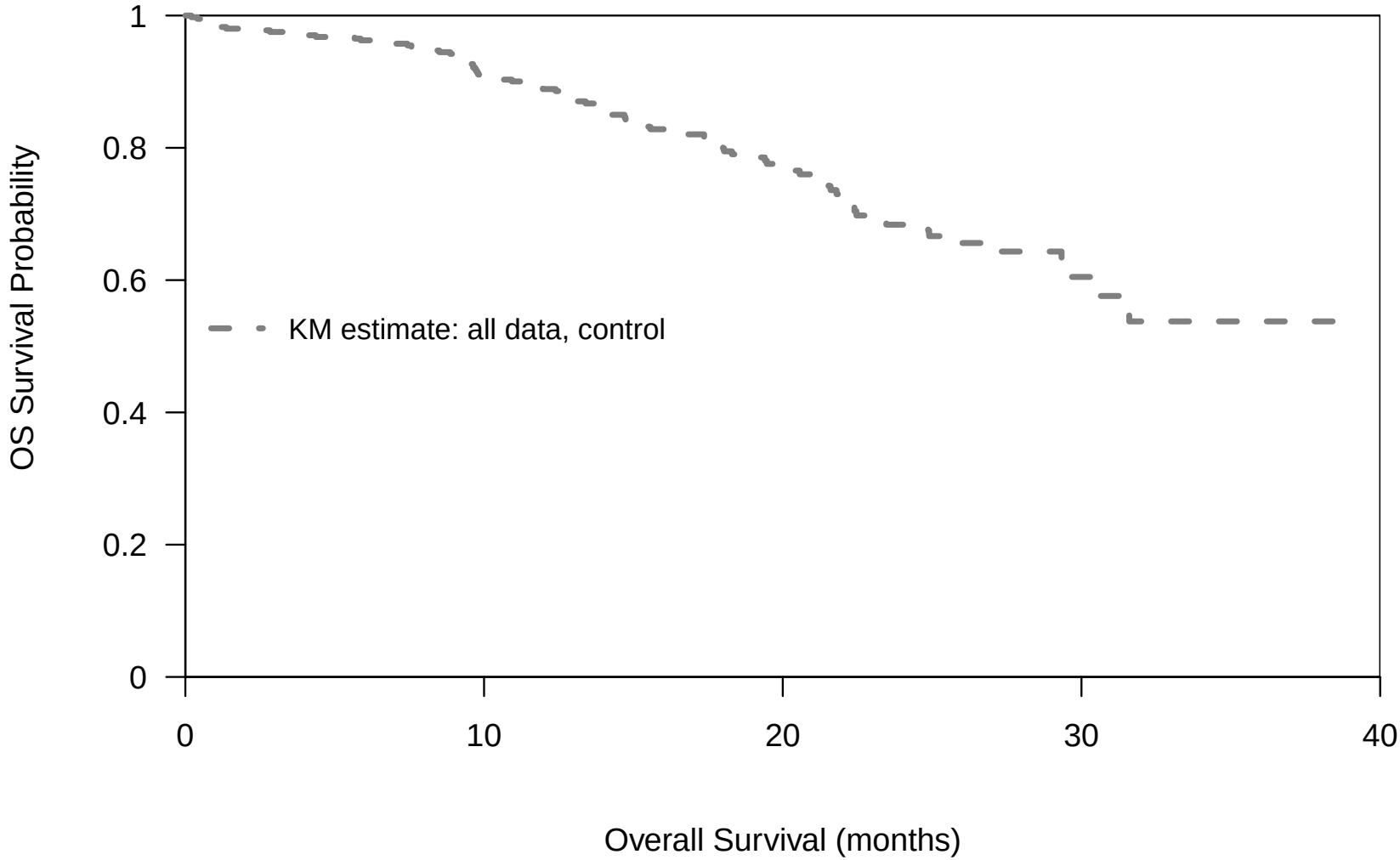
	Pertuzumab+Trastuzumab	Trastuzumab	HR (95% CI)
Survival	N=402	N=406	
Overall Survival			**0.64**
			(0.47,0.88)
Progression-free Survival			0.62 (0.51,0.75)
Response	N=343	N=336	
Objective Response	275 (80.2%)	233 (69.3%)	
Stable Disease	50 (14.6%)	70 (20.8%)	
Progressive Disease	13 (3.8%)	28 (8.3%)	
Duration of Response	N=275	N=233	
Median (months, 95% CI)	**20.2 (16.0,24.0)**	**12.5 (10.0-15.0)**	

- Moderate difference in response.
- Prolonged **duration of response** in experimental arm.
- Clear OS benefit.
- Experimental treatment induces antibody-dependent cellular cytotoxicity $\Rightarrow$ no benefit beyond PD expected $\Rightarrow$ λ_{34} **same in both arms**.

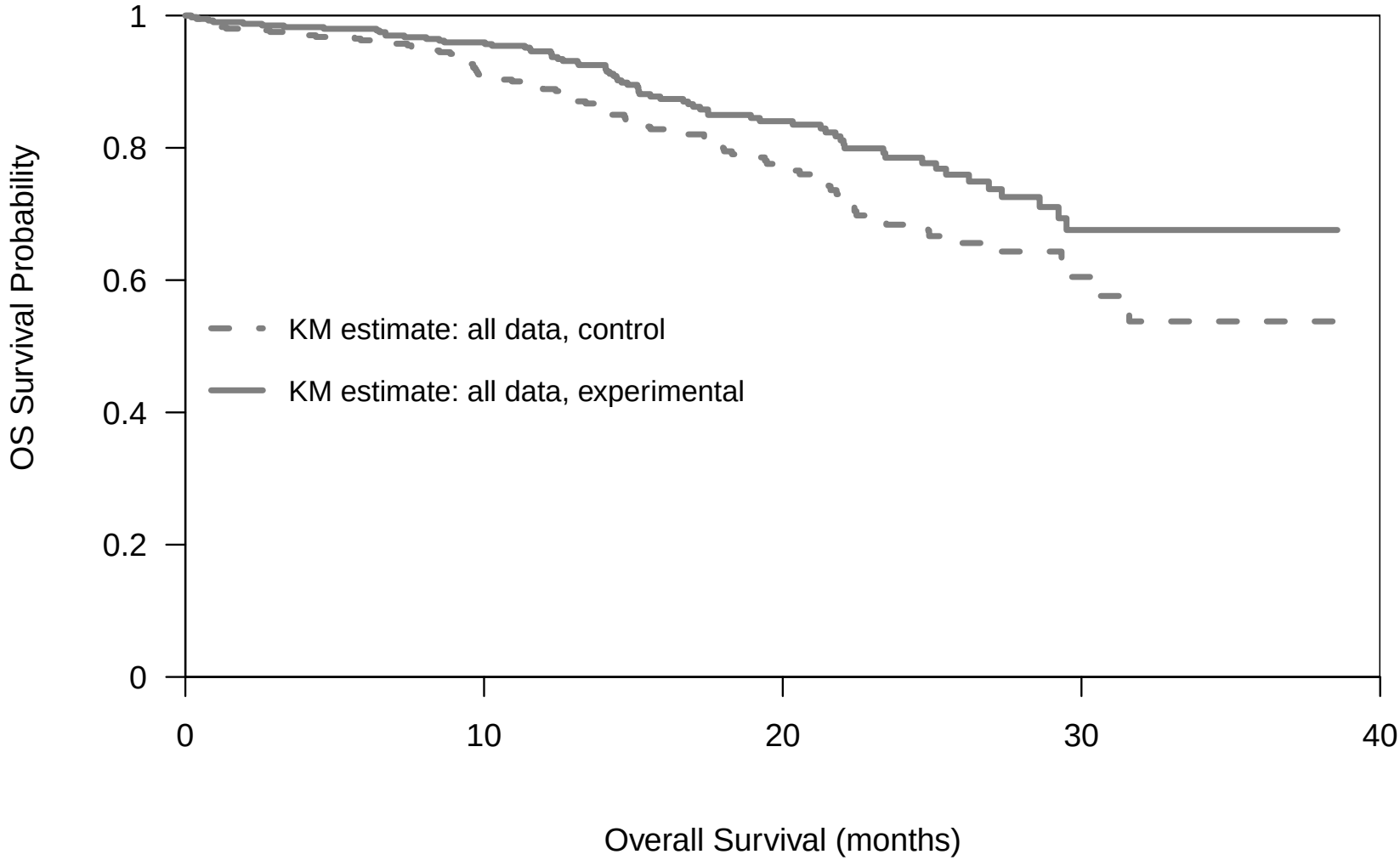
Cleopatra: cumulative hazards of interest



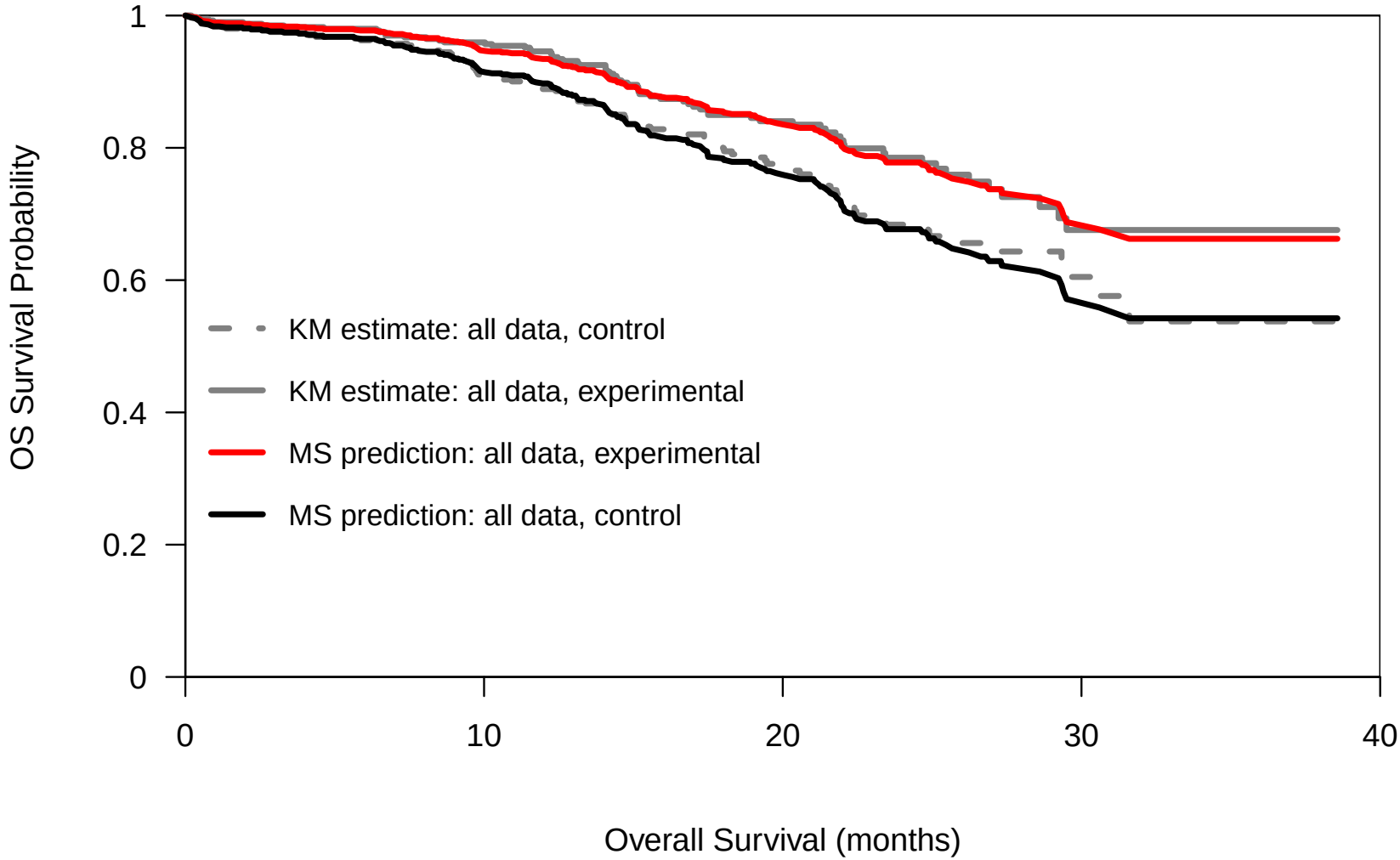
Cleopatra: estimates / predictions of S_{exp}



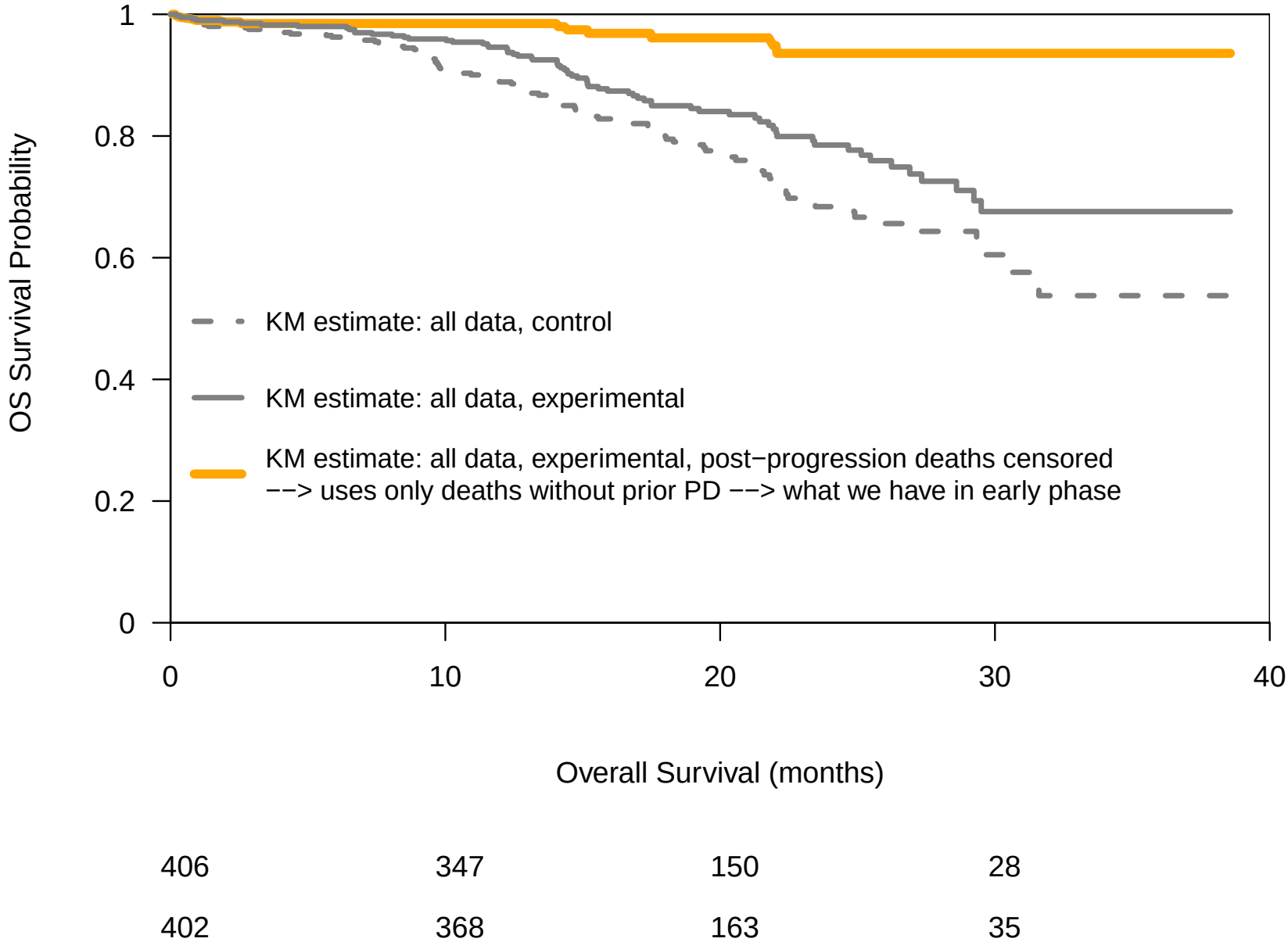
Cleopatra: estimates / predictions of S_{exp}



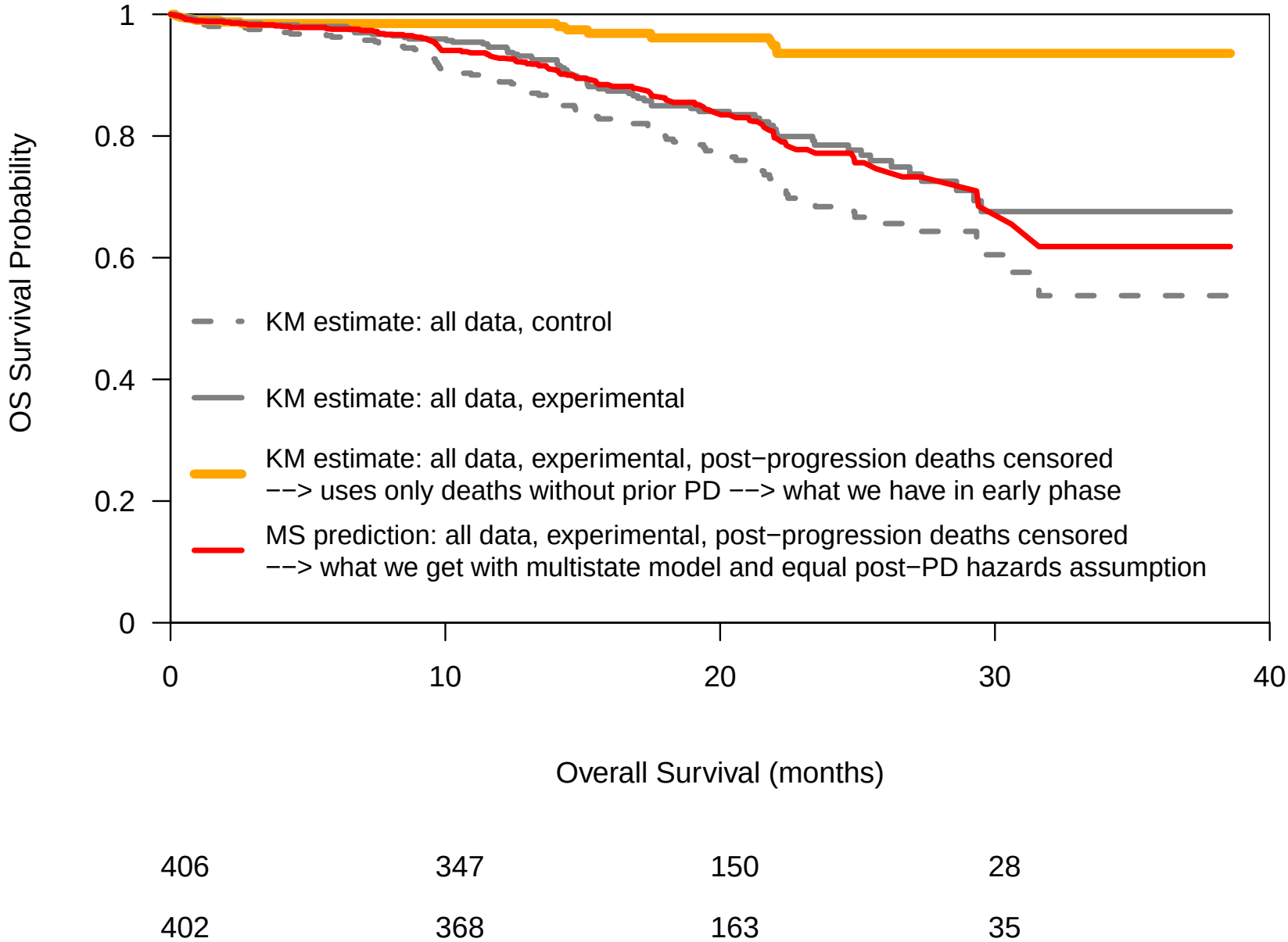
Cleopatra: estimates / predictions of S_{exp}



Cleopatra: estimates / predictions of S_{exp}



Cleopatra: estimates / predictions of S_{exp}



Conclusions for Cleopatra

For estimated / predicted survival function in experimental arm, based on **all data**:

- Majority of patients dies after observed PD.
- KM estimate of simply censoring post-PD deaths does not work $\Rightarrow$ very **few deaths observed**.
- Multistate model prediction assuming post-PD hazards as in control provides good prediction.

Operating characteristics of early phase decision based on multistate prediction?

OS prediction from mimicked early phase data

Sample early phase trial from **Cleopatra experimental arm**:

- 40 patients,
- 6 months uniform recruitment,
- analysis 15 months after first patient entered,
- censor post-PD follow up **one day after PD**,
- estimate $\lambda_{12}, \lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}$ from this data,
- **borrow $\widehat{\lambda}_{34}$ from historical data** = Cleopatra control arm in idealized scenario,
- compute prediction of S_{exp} as described above.

Resampling of operating characteristics

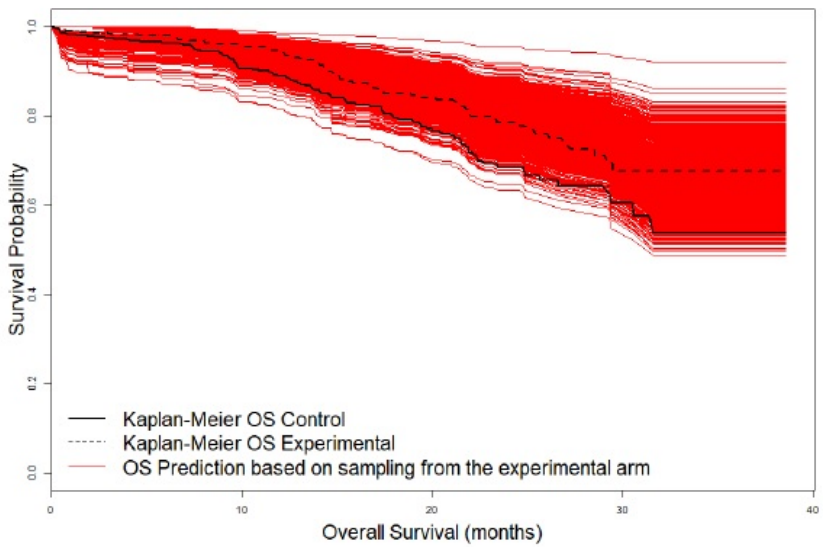
Setup:

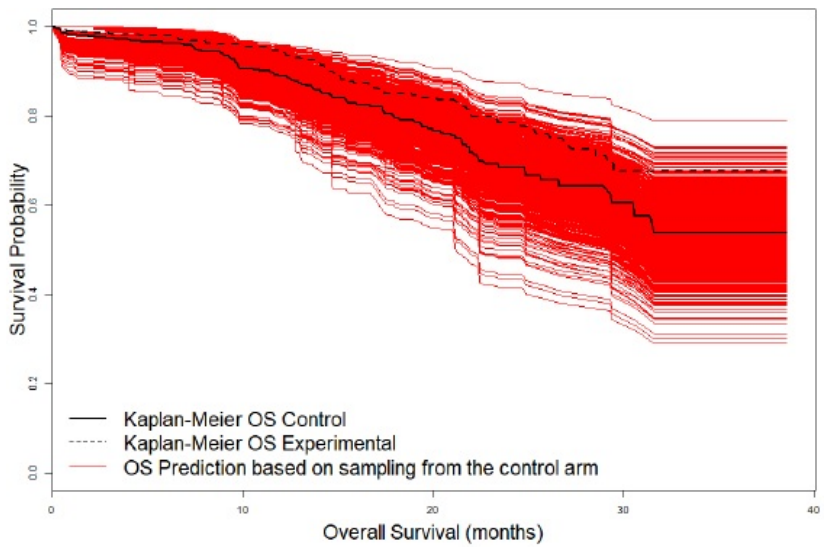
- Use all data in control arm $\Rightarrow$ corresponds to historical control.
- **False-positive** decision: Sample early phase trial from Cleopatra control arm.
- **False-negative** decision: Sample early phase trial from Cleopatra experimental arm.
- Approximate HR by fitting exponential distribution to both arms $\Rightarrow \widehat{HR}$.
- Decision to move to Phase 3: $\widehat{HR} \leq$ boundary $\in \{0.80, 0.85, 0.90, 1.00\}$.
- Repeat 1000 times.

Resampling easily allows for **quantification of uncertainty**.

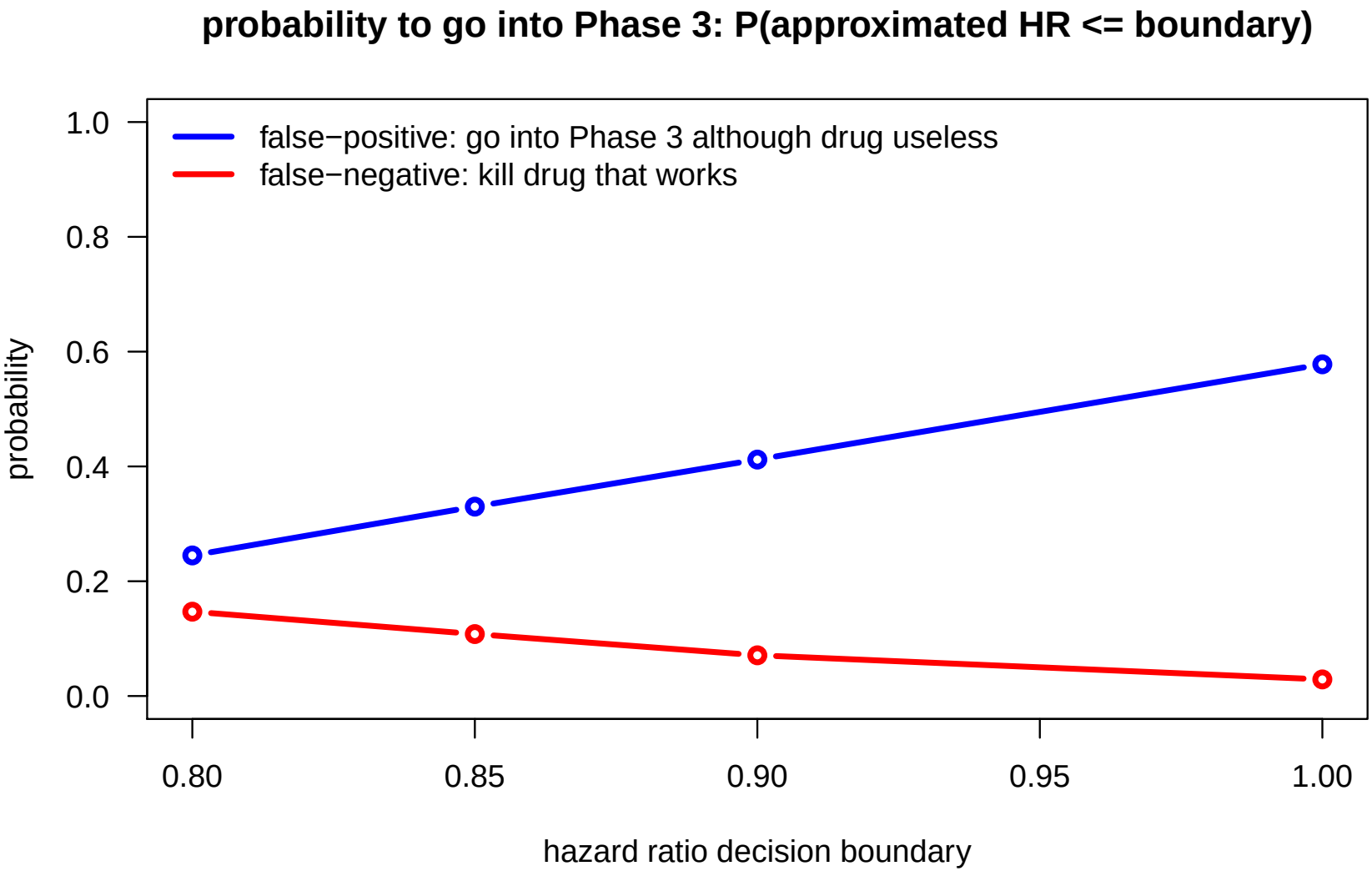
Cleopatra: operating characteristics

Sampled from **experimental** and **control** arm.





Cleopatra: operating characteristics



Decision based on response: $\approx 10\%$ difference, some prolongation of DOR $\Rightarrow$ moved to Phase 3.

Example 2: OAK

OAK

Rittmeyer et al. (2017).

Previously treated non-small-cell lung cancer.

	Atezolizumab	Chemotherapy	HR (95% CI)
Survival	N=425	N=425	
Overall Survival			**0.73 (0.62,0.87)**
Progression-free Survival			0.95 (0.82,1.10)
Response	N=425	N=425	
Objective Response	**58 (13.6%)**	**57 (13.4%)**	
Stable Disease	150 (35%)	177 (42%)	
Progressive Disease	187 (44%)	117 (28%)	
Duration of Response	N=58	N=57	
Median (months, 95% CI)	**26.3 (10,NE)**	**6.2 (4.9-7.6)**	

- **No observed difference in response**.
- **Prolonged duration of response** in experimental arm.
- Control: no benefit post-PD expected.
- Experimental: CIT and post-PD treatment allowed $\Rightarrow$ continued **benefit expected after treatment/PD** $\Rightarrow$ post-PD hazards expected to be **different**.

OS prediction when post-PD hazards assumed proportional

Define random variable $Z = 1$ if patient is in experimental, $Z = 0$ if in control.

Assumption:

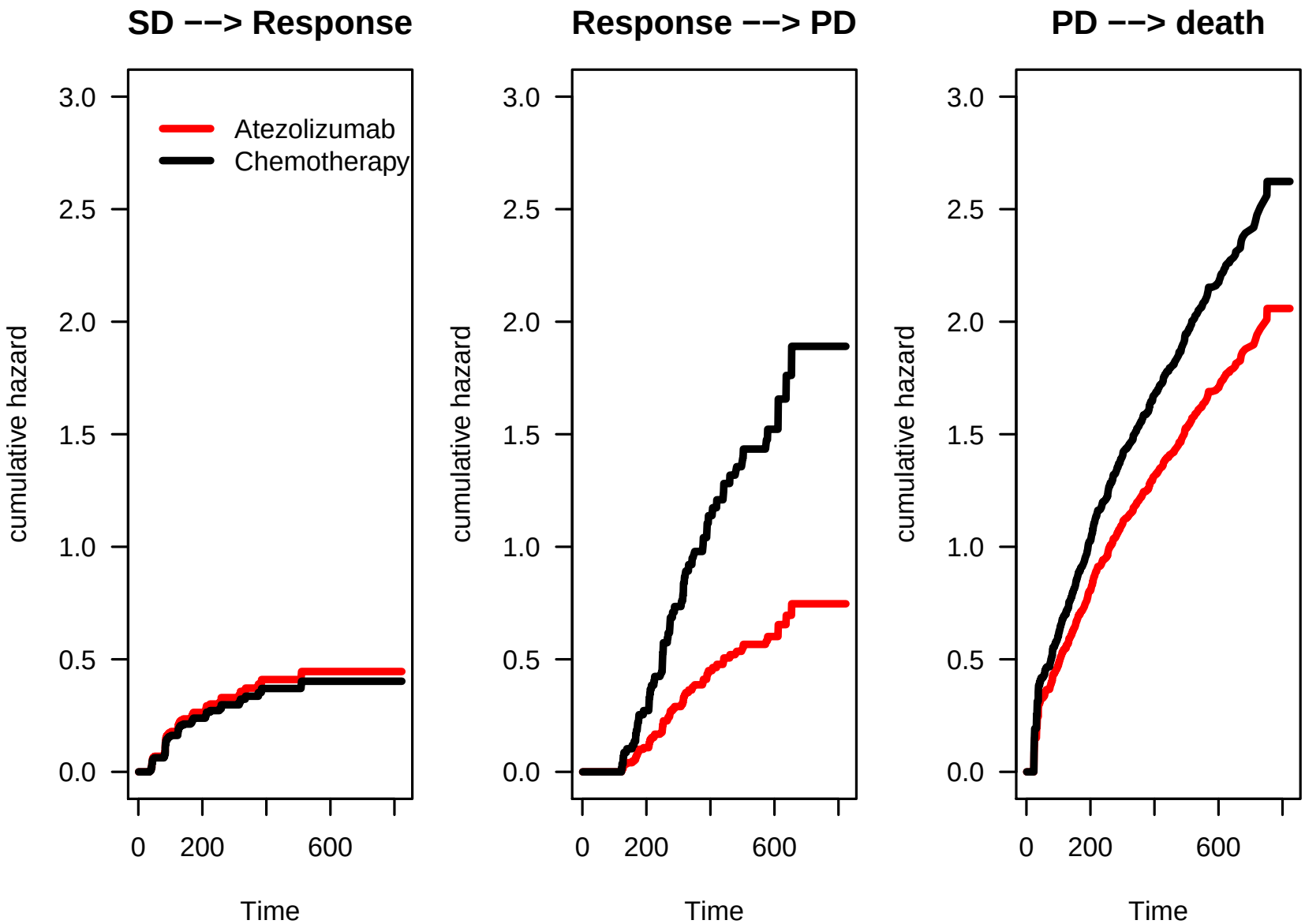
$$\lambda_{34}(t \mid Z) \quad = \quad \lambda_{34,0}(t) \exp(\beta_{34} Z).$$

Baseline hazard $\lambda_{34,0}$ **estimated from both arms combined**.

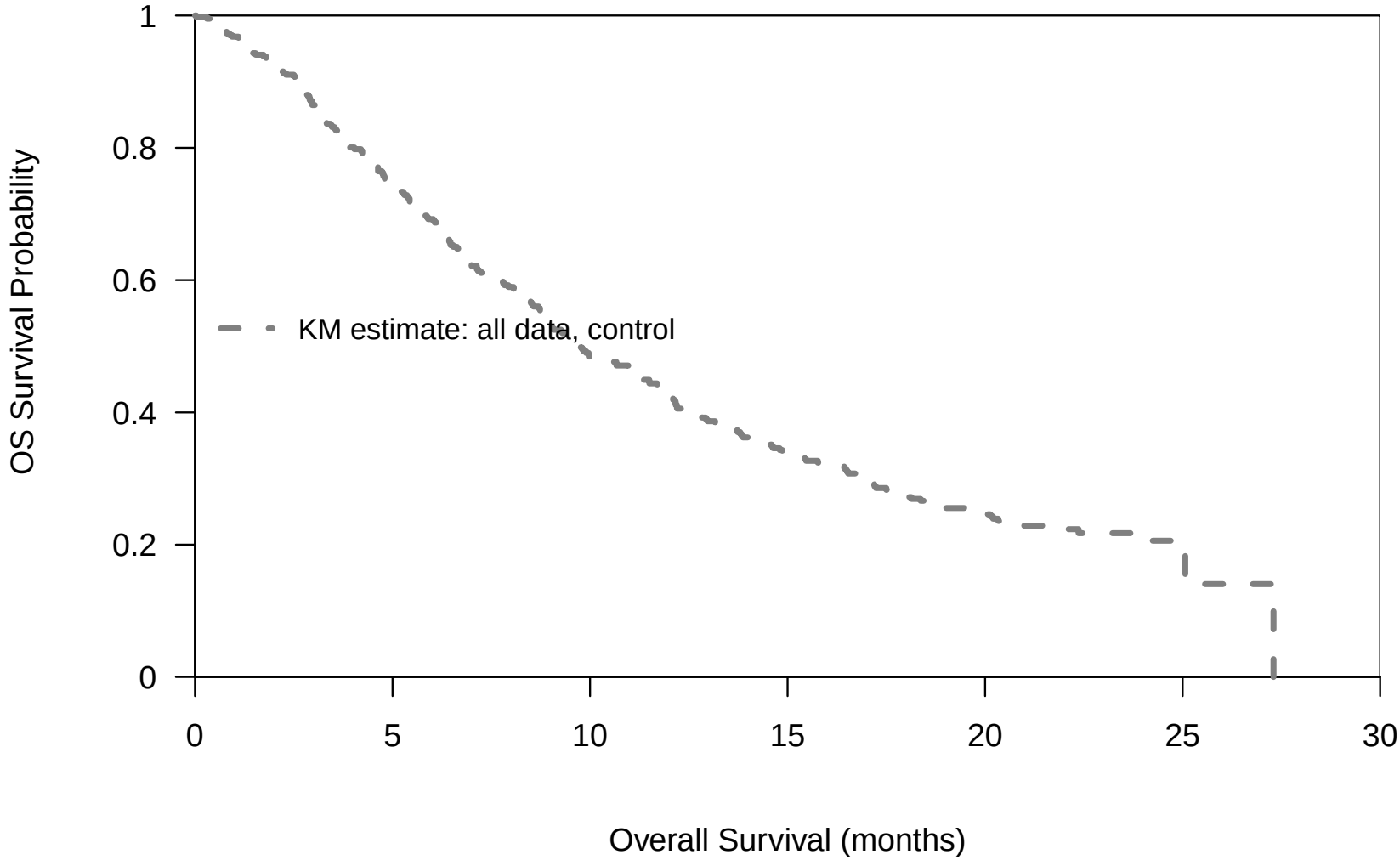
How much post-PD data needed in experimental arm to estimate β_{34}?

Typical early phase follow up: Post-PD deaths censored 180 **days after recruitment** in experimental arm.

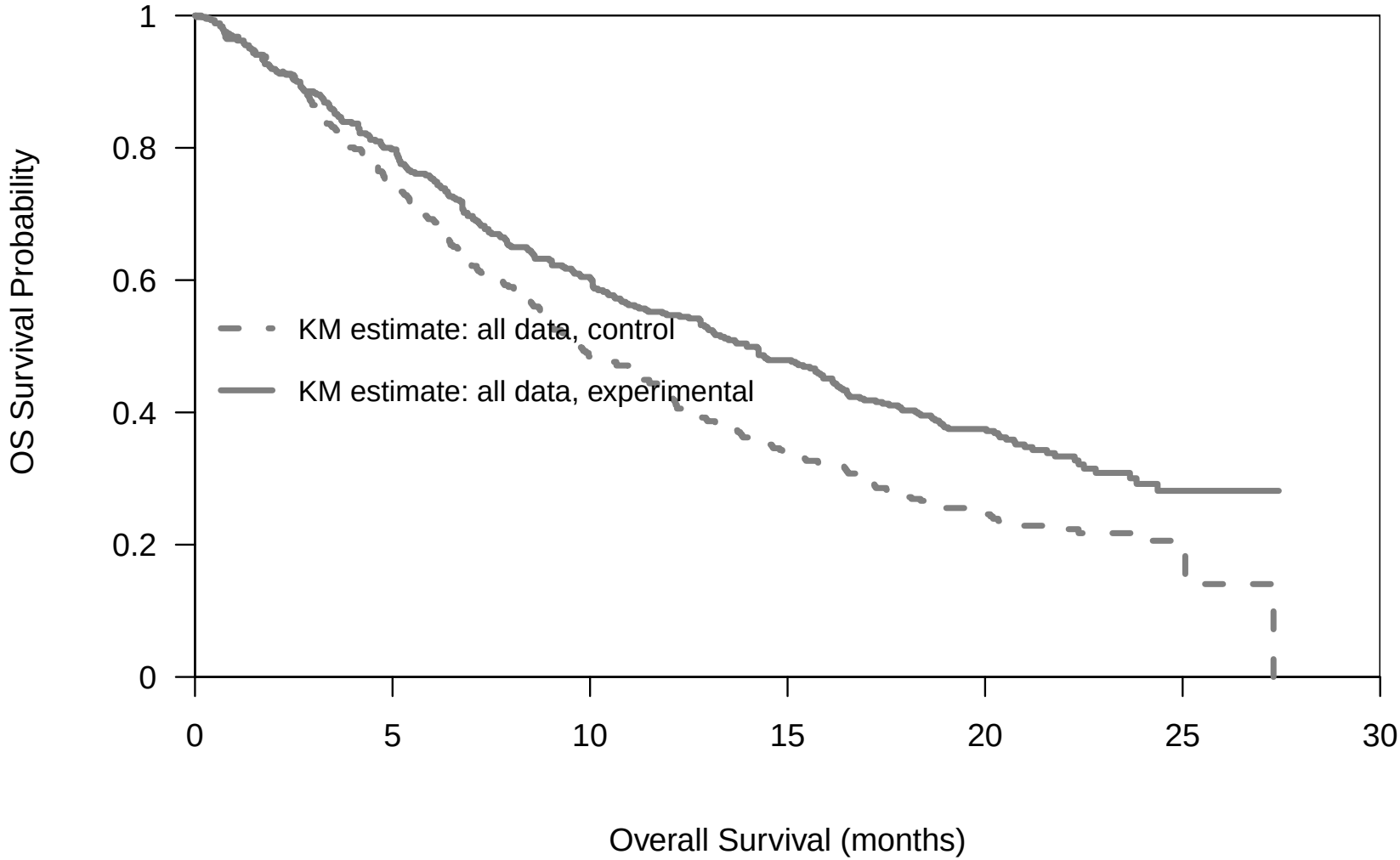
Oak: cumulative hazards of interest



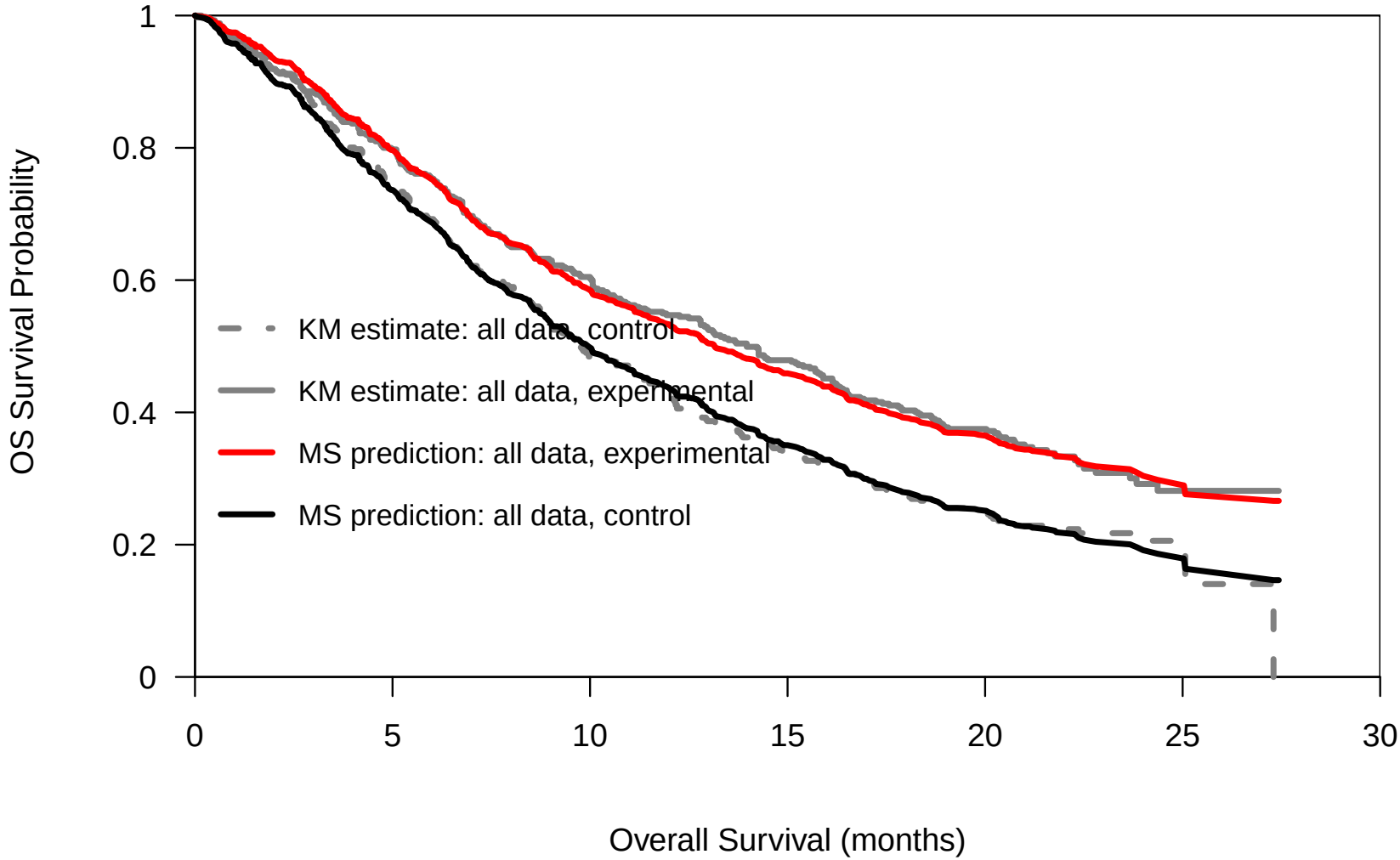
OAK: estimates / predictions of S_{exp}



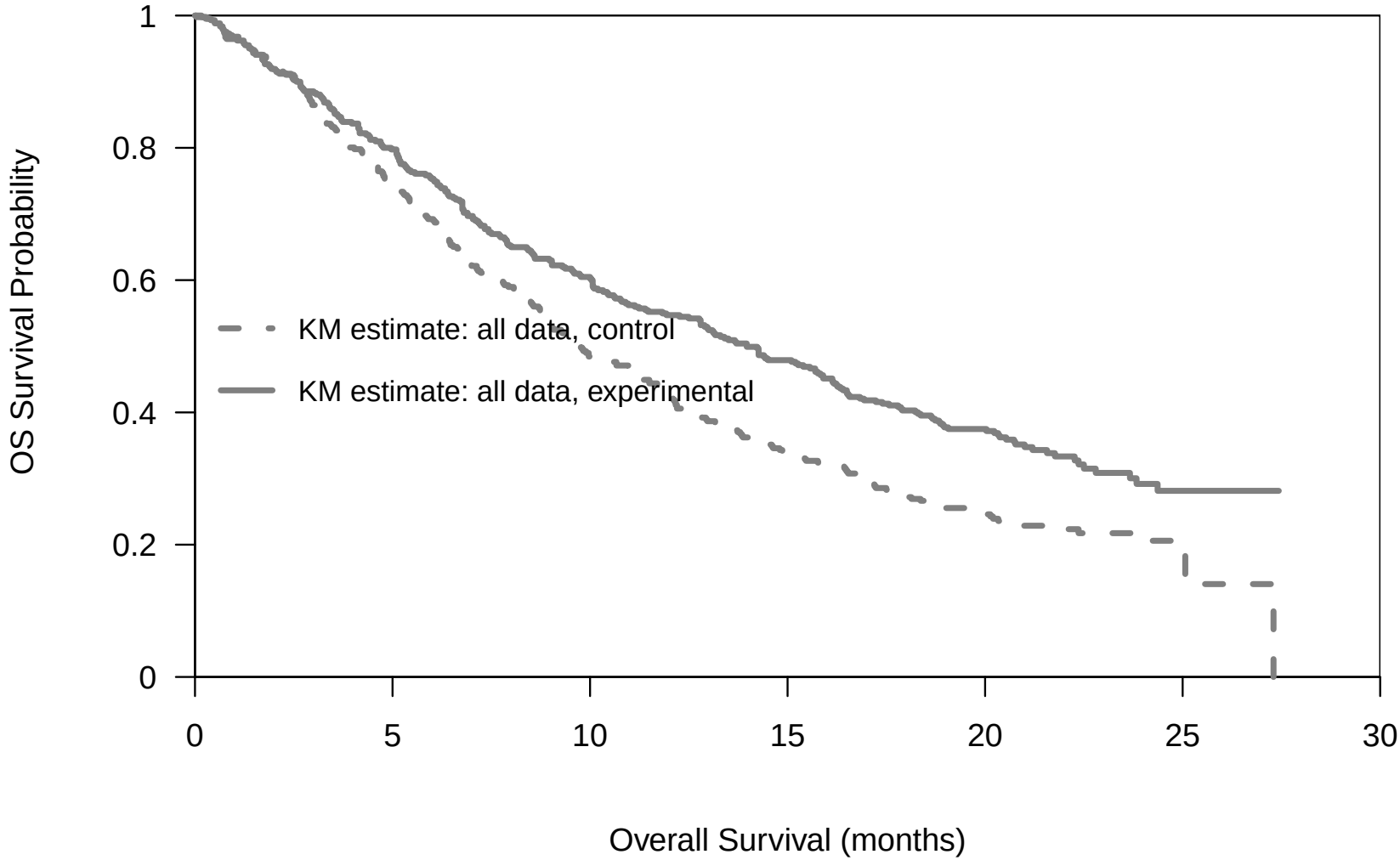
OAK: estimates / predictions of S_{exp}



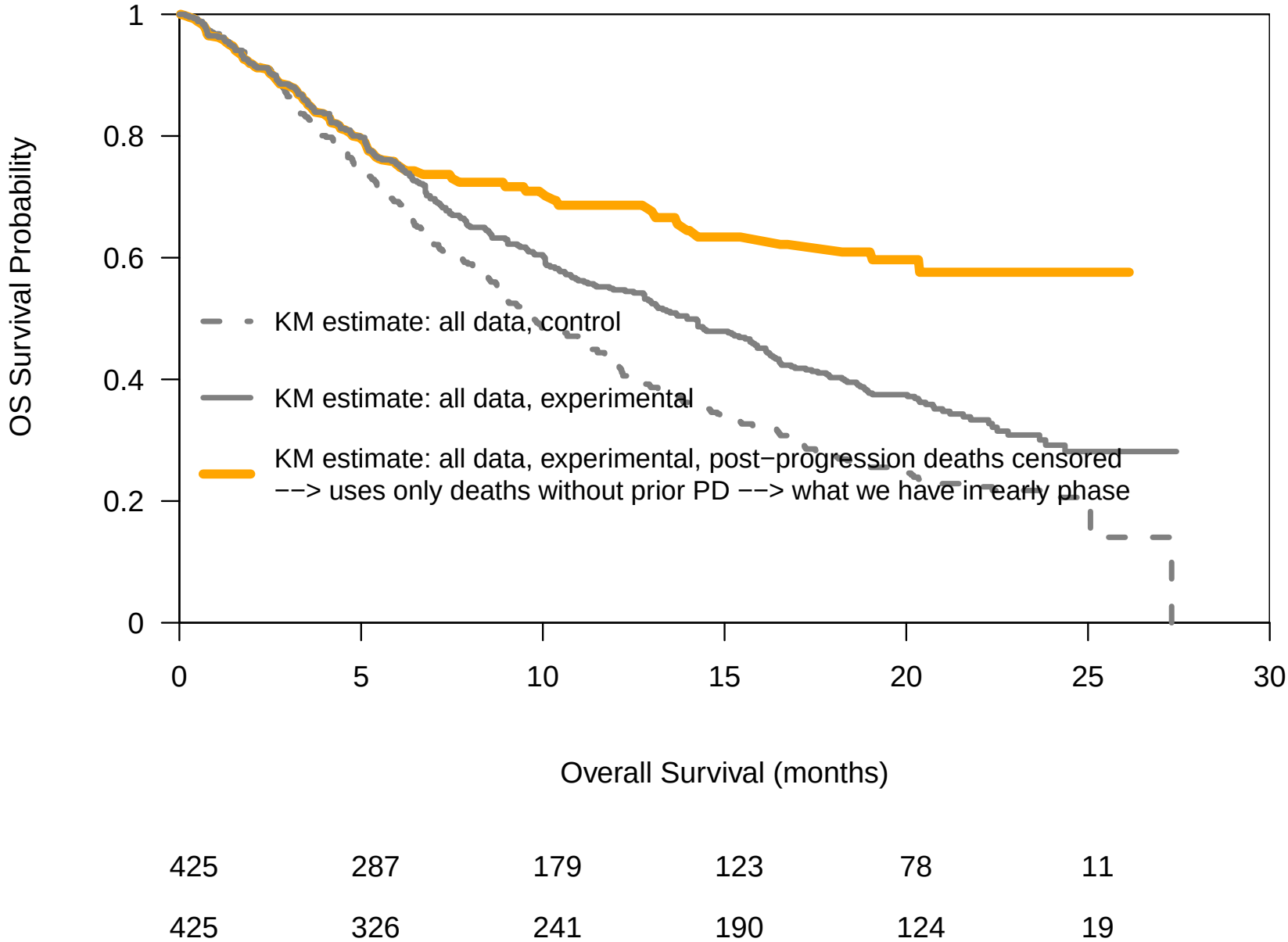
OAK: estimates / predictions of S_{exp}



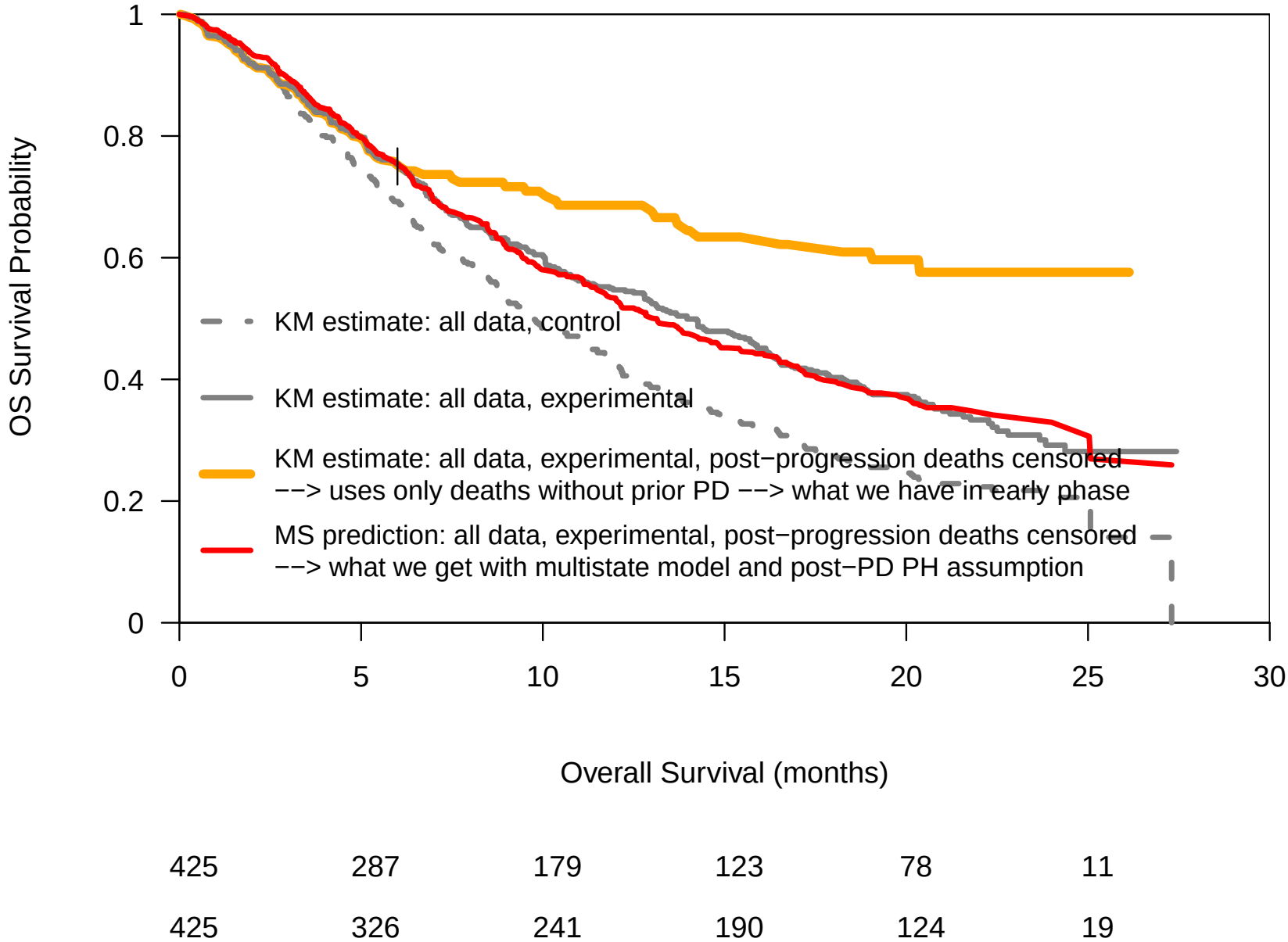
OAK: estimates / predictions of S_{exp}



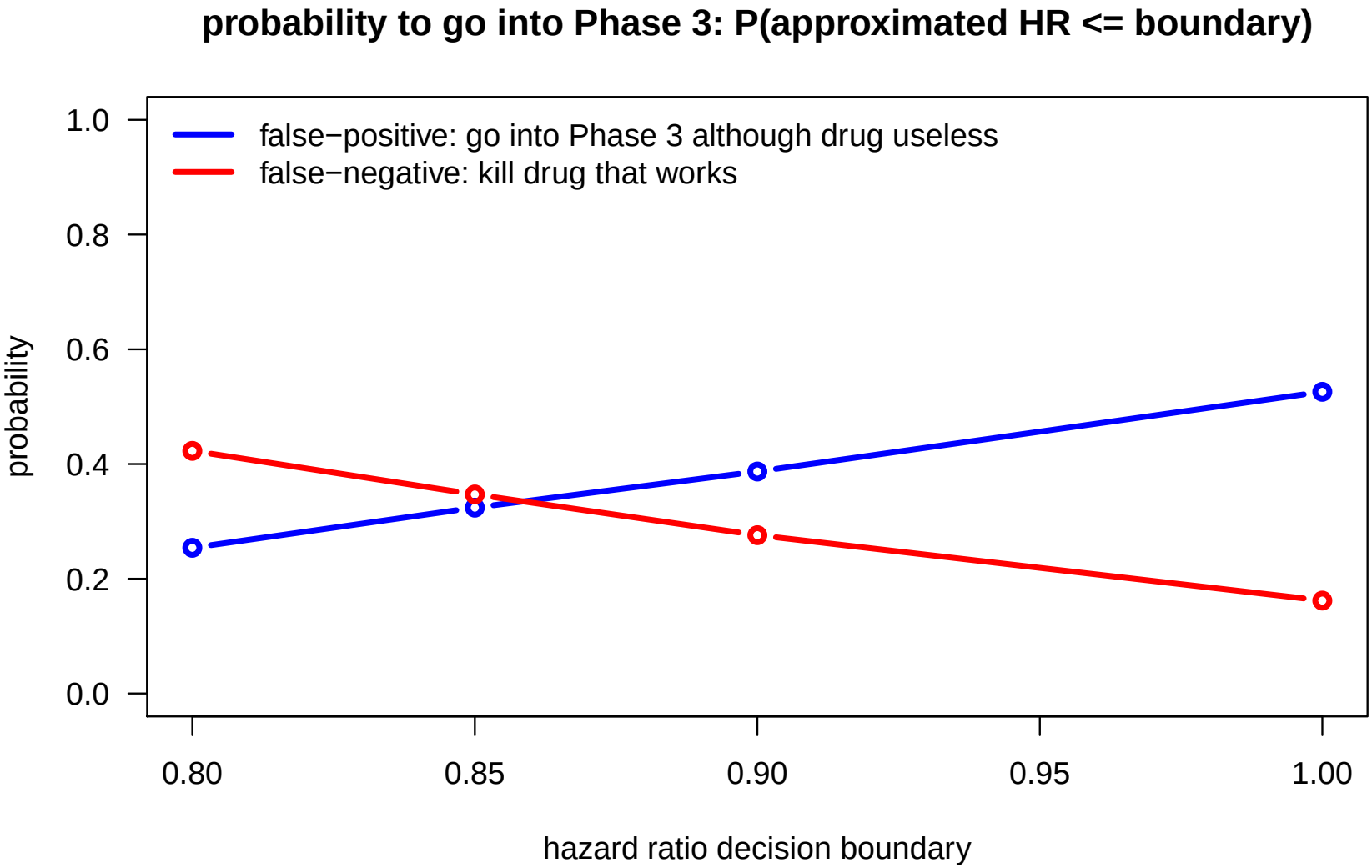
OAK: estimates / predictions of S_{exp}



OAK: estimates / predictions of S_{exp}



Oak: operating characteristics



Response: no difference, some prolongation of DOR. Likely would NOT have started Phase 3. Competitor info...

How many post-PD deaths needed to estimate HR of $3 \to 4$ transition?

How many post-PD deaths needed?

Assumption:

$$\lambda_{34}(t \mid Z) \quad = \quad \lambda_{34,0}(t) \exp(\beta_{34} Z).$$

How many post-PD deaths needed in **experimental** arm to reliably estimate λ_{34}?

Planning stage: only data for control arm are available.

- Fit multistate model to control data.
- **Simulate** assuming potential differences in transition hazards for experimental arm.
- Considered hazard (ratios) should end up in a **clinical meaningful OS effect**.

Several scenarios for different post-PD follow up time can be simulated.

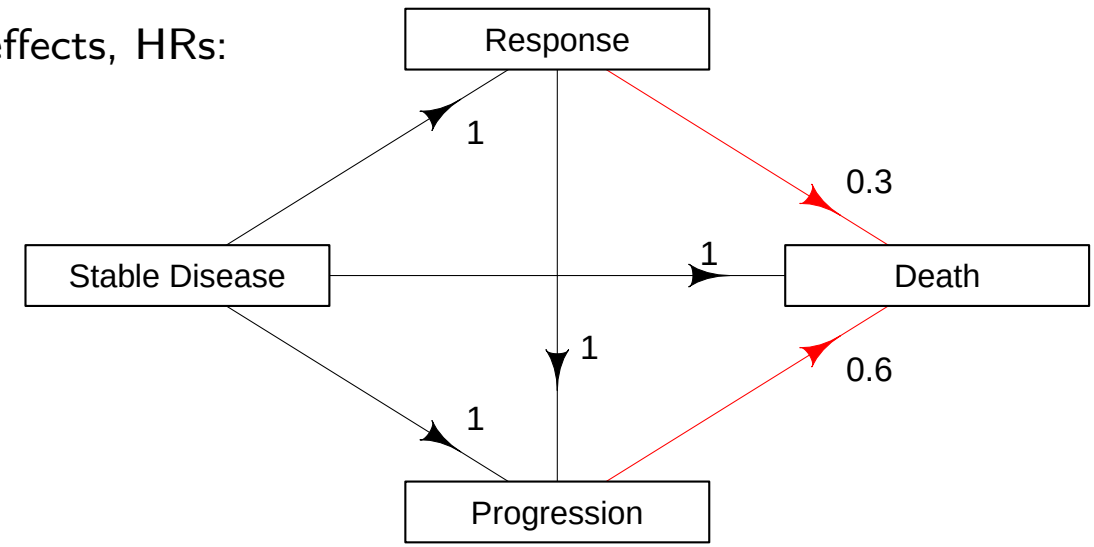
Purpose and simulation details

Goal: **NOT** power computation for hypothesis test – sample size too small anyway.

Rather: find cutoff timepoint from which on OS HR estimate remains **stable**.

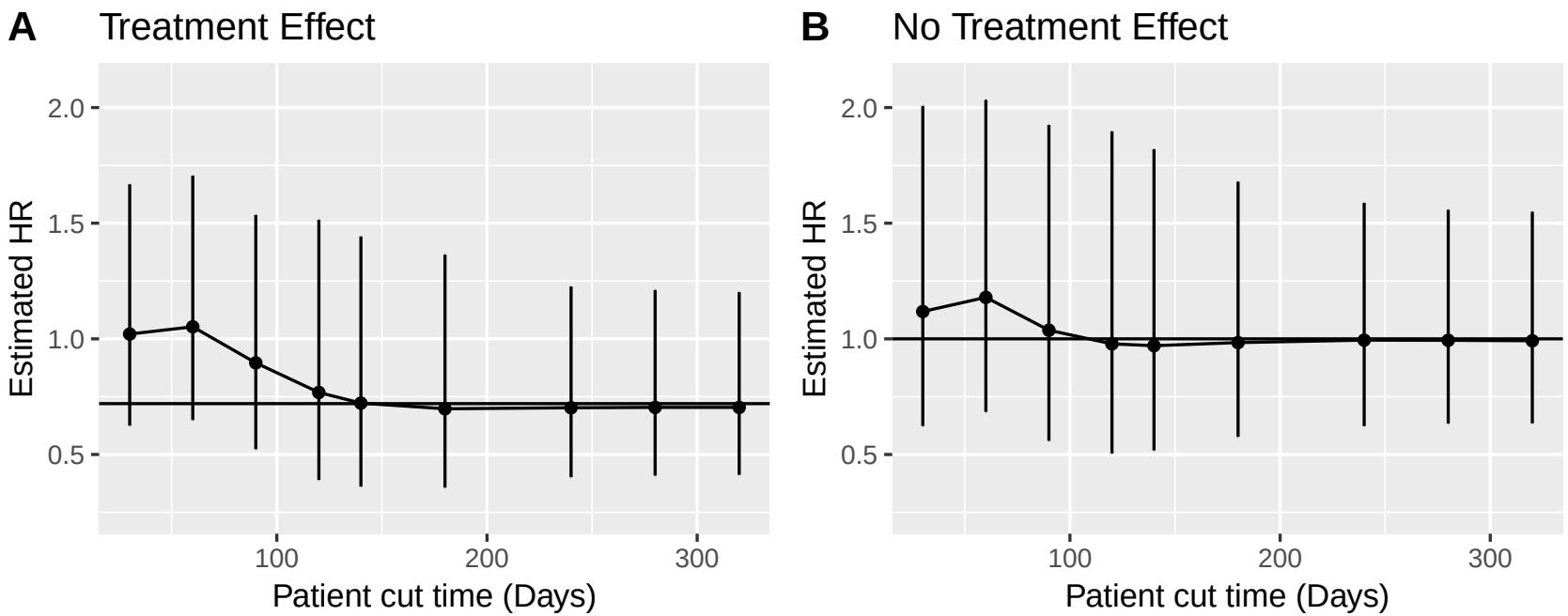
Mimick Oak:

- Simulate 40 patient from experimental arm as before.
- Treatment effects, HRs:



- Resulting **OS HR = 0.73**. Close to Oak OS HR.
- Follow up post-PD for experimental arm truncated at 30, 60, 90, 120, 150, 180 and 240 days after recruitment.
- Repeat 1000 times.

Stability of hazard ratio estimate



180-240 days appear sufficient to obtain stable point estimate over time.

Conclusions for early-decision making proposal

Conclusions

Early phase decision-making based on **multistate OS prediction**:

- Beyer et al. (2019) has **3rd CIT example** with post-PD hazard different between arms.
- Assumption on $\lambda_{34} \Rightarrow$ need to understand **disease and treatment**.
- **Avoids difficulty in interpretation of response-type endpoints**, especially DOR.
- Feasibility assessed in **idealized scenario** where experimental arm OS is available.
- Recommendation **how much post-PD follow-up** needed to estimate λ_{34}.

Open points:

- Use of **real-world data** as historical control $\Rightarrow$ selection bias. Combine proposal with propensity scoring.
- Needs **long-term individual-patient** data in control arm!
- Add **covariates**: baseline and pre-PD, or via joint models.
- Using states based on response $\Rightarrow$ dichotomization. Alternatives?

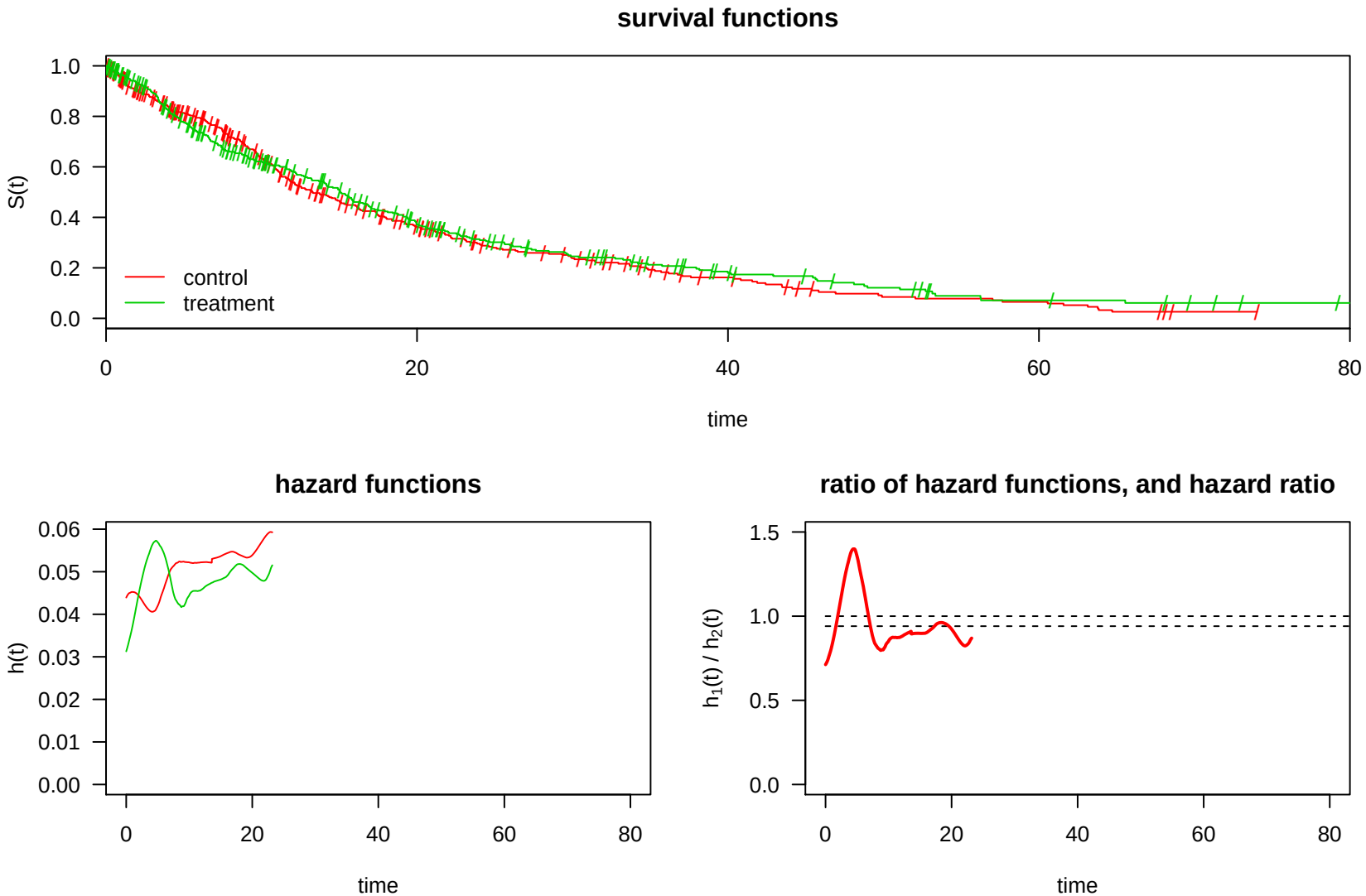
Effect quantification for non-proportional hazards

A fictional clinical trial

Simulated clinical trial:

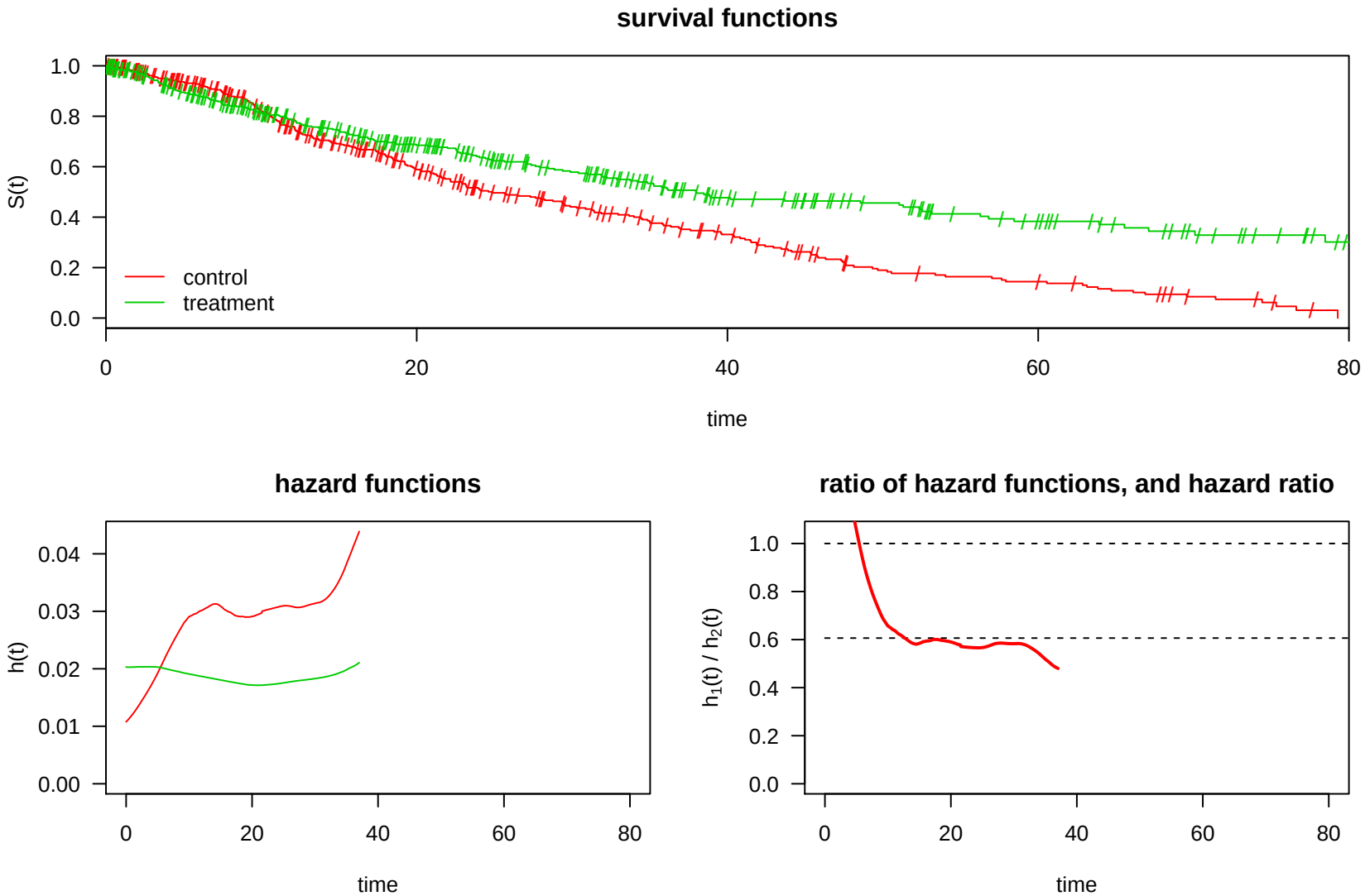
- 1:1 randomized, 400 and 400 patients per arm.
- No administrative censoring, but drop-out.

PFS for simulated clinical trial



- Estimated hazard ratio: 0.94, 95% confidence interval [0.80, 1.11].
- Test for PH: $p = 0.24$.

OS for simulated clinical trial



- Estimated hazard ratio: 0.61, 95% confidence interval [0.50, 0.74].
- Test for PH: $p < 0.0001$.

Summarize treatment effect

Non-proportional hazards for OS. How to summarize effect of treatment?

Data was generated according to:

Transition	Control arm	Treatment arm
$0 \to 1$	$\lambda_{01}^c = \log(2)/25$	$\lambda_{01}^t = \lambda_{01}^c \cdot \mathbf{1}$
$0 \to 2$	$\lambda_{02}^c = \log(2)/30$	$\lambda_{02}^t = \lambda_{02}^c \cdot \mathbf{0.8}$
$1 \to 2$	$\lambda_{12}^c = \log(2)/15$	$\lambda_{12}^t = \lambda_{12}^c \cdot \mathbf{0.4}$

	coef	HR = exp(coef)	95% CI	p-value
transition event-free –> PD	-0.04	0.96	[0.77, 1.19]	0.72
transition event-free –> death	-0.09	0.91	[0.70, 1.18]	0.49
transition PD –> death	-1.09	0.34	[0.24, 0.46]	< 0.0001

Conclusions

Multistate models

Multistate models useful:

- Canonical **extension of survival analysis**.
- Get more **insight** in how disease and drug work.
- **Competing risk** simplest multistate model.
- **Prediction** in well-specified, as opposed to black-box, model.
- **Jointly** model three key oncology endpoints: response, PFS, OS. Applications by no means restricted to oncology though!

Many potential applications:

- Improved **early stage decision-making** ⇒ Beyer et al. (2019).
- Improved **communication** of effect and optimized **sample size** computation.
- **Event-tracking** with transition-specific covariates and taking into account every patient's history.
- Bivariate modelling of PFS and OS to help inform **surrogacy** questions ⇒ Meller et al. (2019).

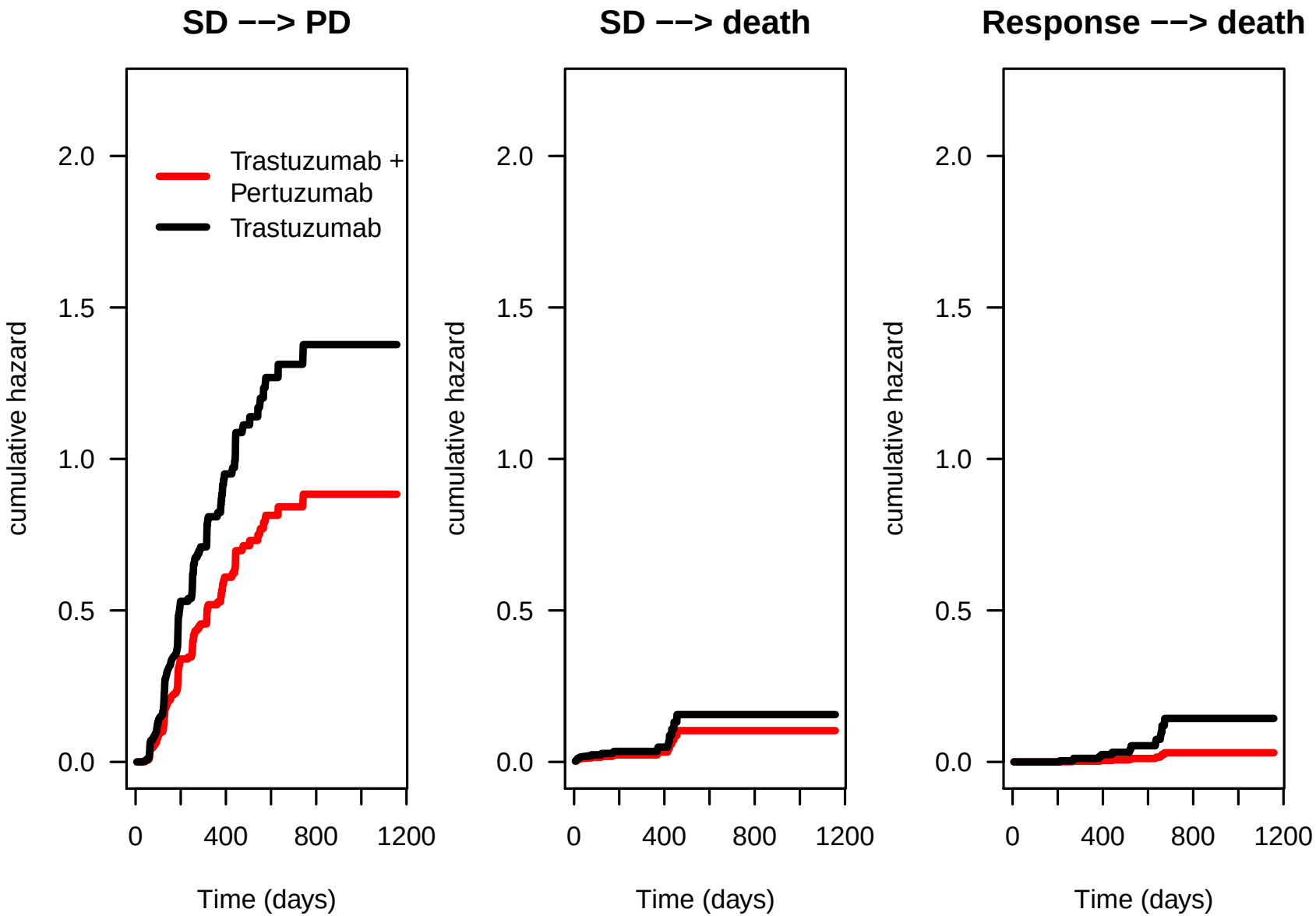
Thank you for your attention

References I

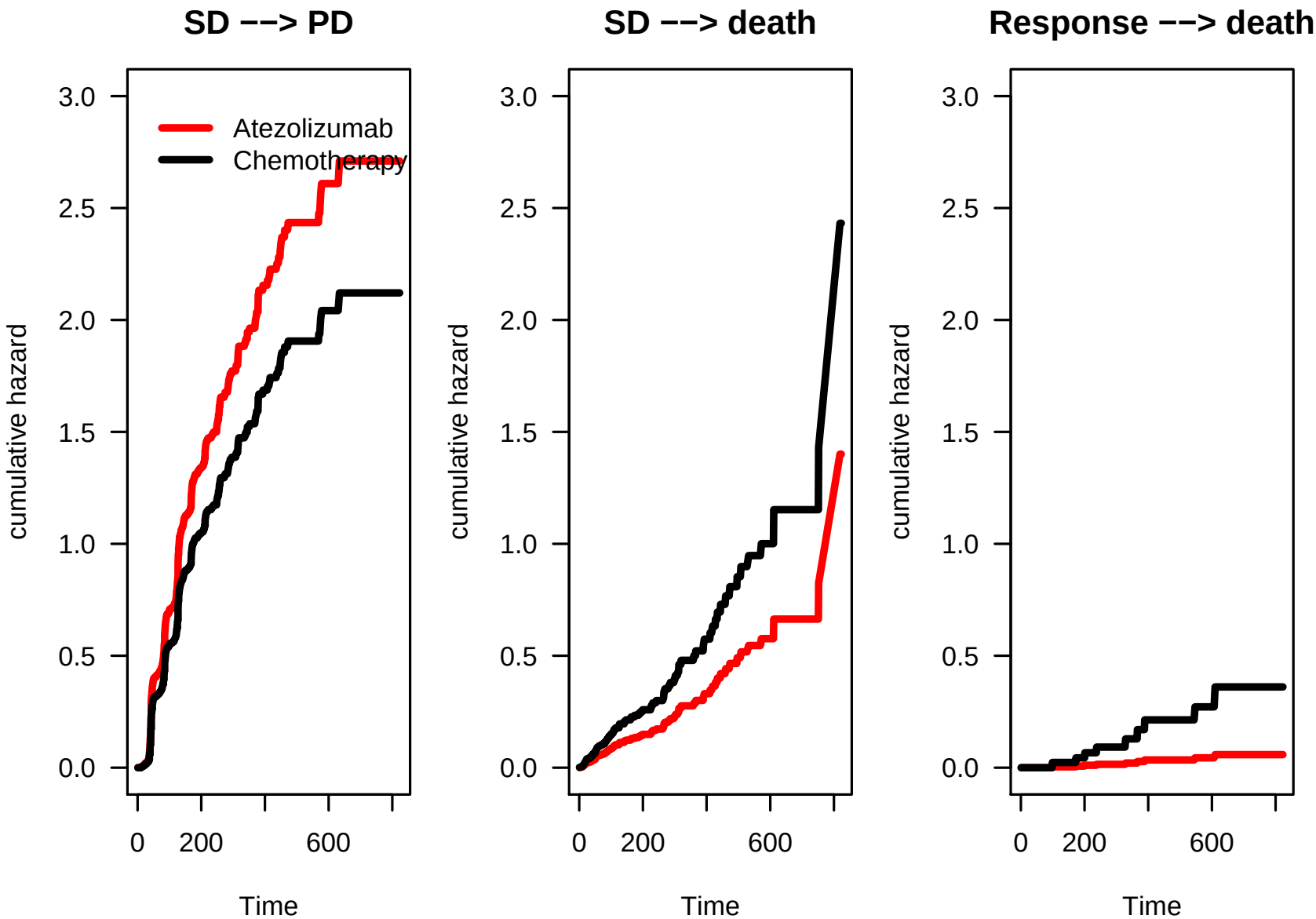
- Andersen, P. K., Borgan, r., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer.
- Baselga, J. and Cortes, J. e. a. (2012). Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N. Engl. J. Med.* **366** 109–119.
- Beyer, U., Dejardin, D., Meller, M., Rufibach, K. and Burger, H. U. (2019). A multistate model for early decision making in oncology. *Biom J, to appear* .
- Meller, M., Beyersmann, J. and Rufibach, K. (2019). Joint modelling of progression-free and overall survival and computation of correlation measures. *Stat. Med., accepted* .
- Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., Von Pawel, J., Gadgeel, S. M., Hida, T., Kowalski, D. M., Dols, M. C. et al. (2017). Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (oak): a phase 3, open-label, multicentre randomised controlled trial. *The Lancet* **389** 255–265.
- Swain, S. M. and Baselga, J. (2015). Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *N. Engl. J. Med.* **372** 724–734.

Backup

Cleopatra: cumulative hazards of secondary interest

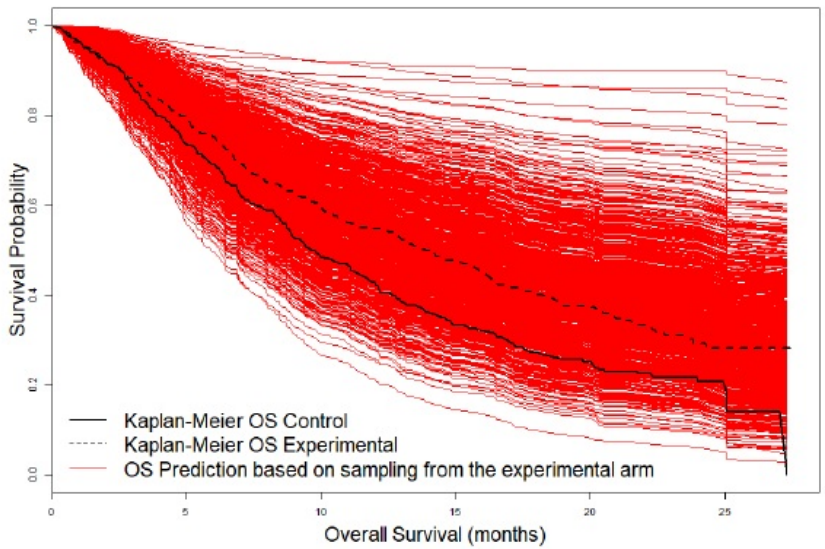


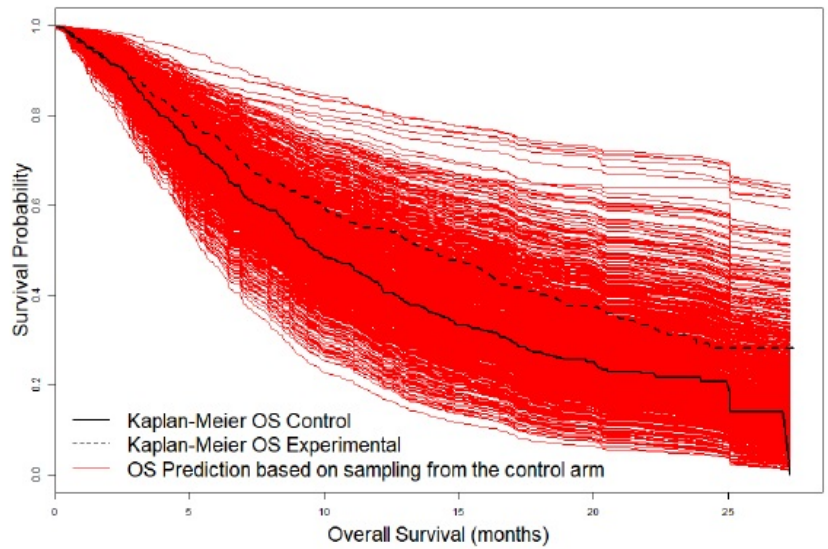
Oak: cumulative hazards of secondary interest



Oak: operating characteristics

Sampled from **experimental** and **control** arm.





Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 3.5.1 (2018-07-02)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages: nls2 / proto / bindrcpp / diagram / shape / ggplot2 / rocheBCE / muhaz / flexsurv / reporttools / xtable / mstate / etm / dplyr / mvna / prodlim / biostatKR / survival

This document was generated on 2019-06-27 at 06:01:09.