A Phase 3 trial with cure proportion, and some thoughts on NPH

Kaspar Rufibach
Methods, Collaboration & Outreach Group, PD Data Sciences, Roche Basel
PSI 1-day meeting: Non-proportional hazards and applications in immuno-oncology
29th April 2021 (virtual)

Who

Rufibach et al. (2020):



Meller et al. (2019):



Acute Myeloid Leukemia

Acute Myeloid Leukemia

Rare malignant blood disease.

Most common leukemia, lowest survival rate in adults: **median survival $\leq$ 1y**.

Recurrent **life-threatening infections**.

Chemotherapy: modest benefit without cure.

Stem cell transplant:

- "Bridge-to-transplant": Goal of any therapy. Needs **complete response** (CR) to initial therapy.
- Only way to survive AML.

Mirros

MDM2 **I**dasanutlin in **R**elapsed **R**efractory AML for **OS**.

- **Population**: R/R AML.
- **Comparison**: **Idasanutlin** + cytarabine vs. placebo + cytarabine.
- **Phase III, 2:1 randomized, double-blind, placebo-controlled** clinical trial.
- Primary endpoint: **overall survival**.
- Planned recruitment: 374 patients.

`https://clinicaltrials.gov/ct2/show/NCT02545283`

How to plan RCT when some patients may be cured?

Cure proportion model

See e.g. Sun et al. (2018).

Let

- S_i^*, f_i^*: survival and density functions of **uncured** patients.
- p_i: proportions of patients cured.

Survival and hazard function in each treatment arm ($t \geq 0$):

$$\begin{aligned} S_i(t) &= p_i + (1-p_i)S_i^*(t), \\ h_i(t) &= \frac{(1-p_i)f_i^*(t)}{p_i + (1-p_i)S_i^*(t)}. \end{aligned}$$

Ratio of hazard functions:

$$\theta(t) = h_2(t)/h_1(t) = \left(\frac{1-p_2}{1-p_1}\right)\frac{f_2^*(t)}{f_1^*(t)}\left(\frac{p_1 + (1-p_1)S_1^*(t)}{p_2 + (1-p_2)S_2^*(t)}\right).$$

Even if both S_i^* exponential $\Rightarrow$ $\theta(t)$ **depends on time** (if ≥ 1 p_i is > 0).

What if we simply ignored cure proportions?

Cure proportion model – assumptions

Assume effect size for S_i^*.

Cure proportion model – assumptions

Assume effect size for S_i^*.

Compute necessary events d using **Schoenfeld's formula**:

Cure proportion model – assumptions

Assume effect size for S_i^*.

Compute necessary events d using **Schoenfeld's formula**:

- Study will (typically) be **underpowered**.
- Time to clinical cutoff will be **underestimated**.

Cure proportion model – assumptions

Assume effect size for S_i^*.

Compute necessary events d using **Schoenfeld's formula**:

- Study will (typically) be **underpowered**.
- Time to clinical cutoff will be **underestimated**.

Control arm, based on **historical data**, H_0:

- Median OS 6m.
- Cure: 0.080.

Cure proportion model – assumptions

Assume effect size for S_i^*.

Compute necessary events d using **Schoenfeld's formula**:

- Study will (typically) be **underpowered**.
- Time to clinical cutoff will be **underestimated**.

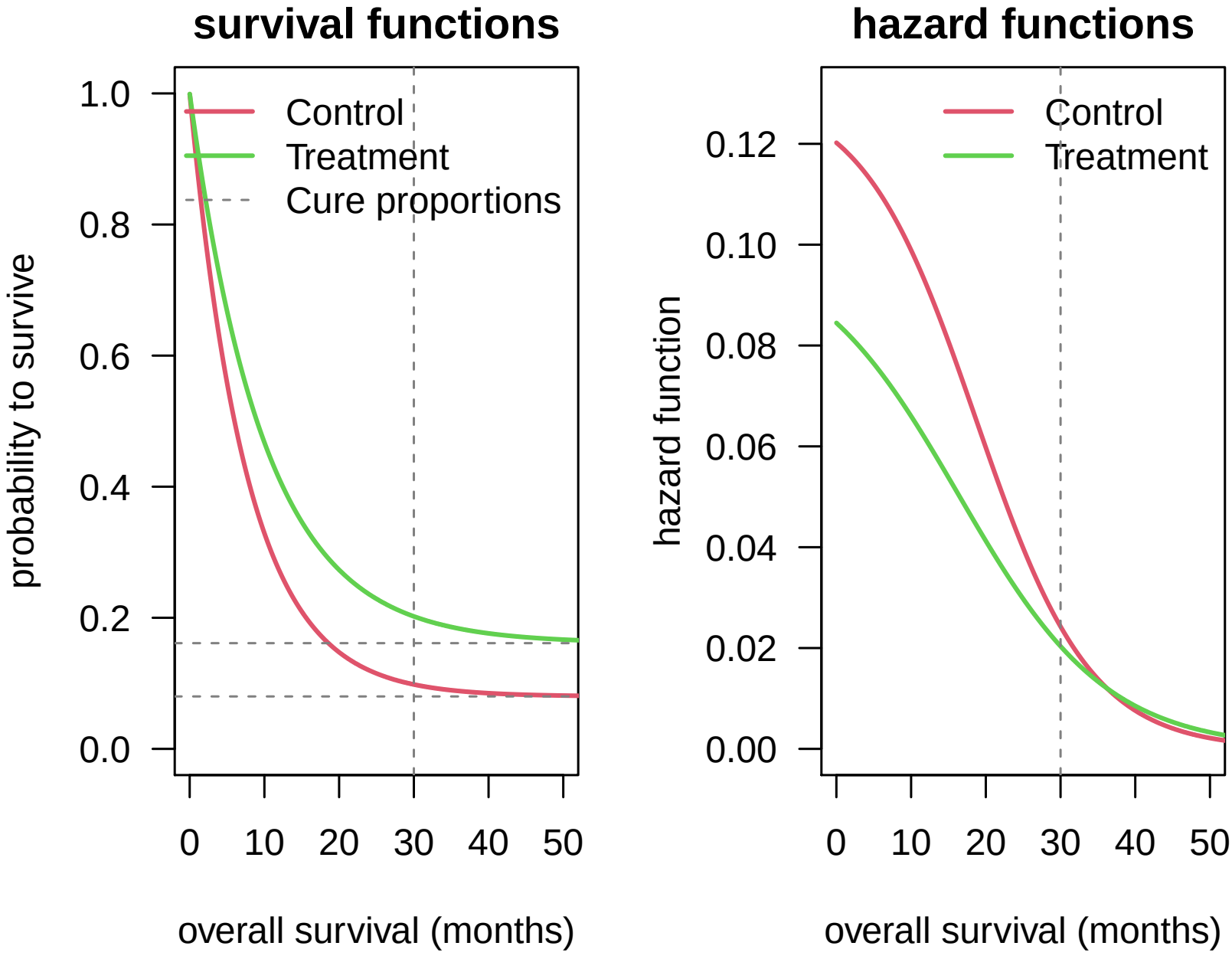
Control arm, based on **historical data**, H_0:

- Median OS 6m.
- Cure: 0.080.

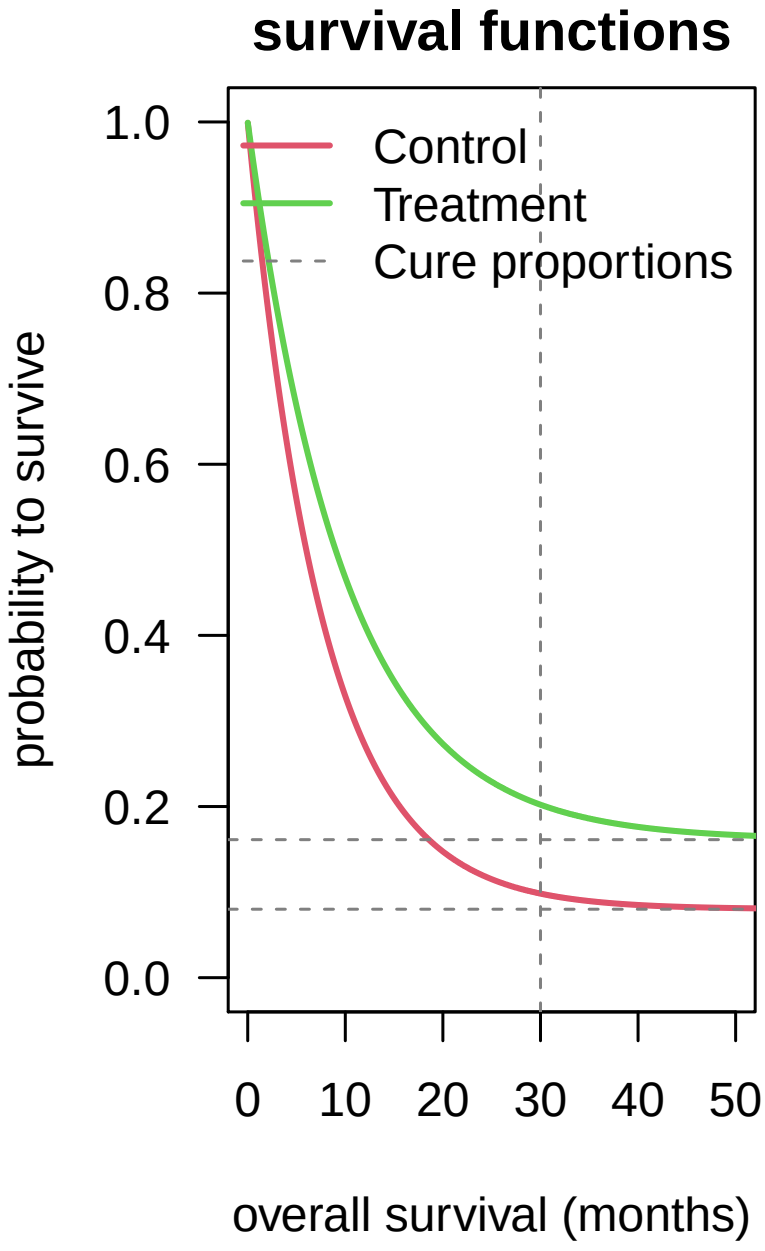
Targeted effect size treatment arm (for 85% power, H_1):

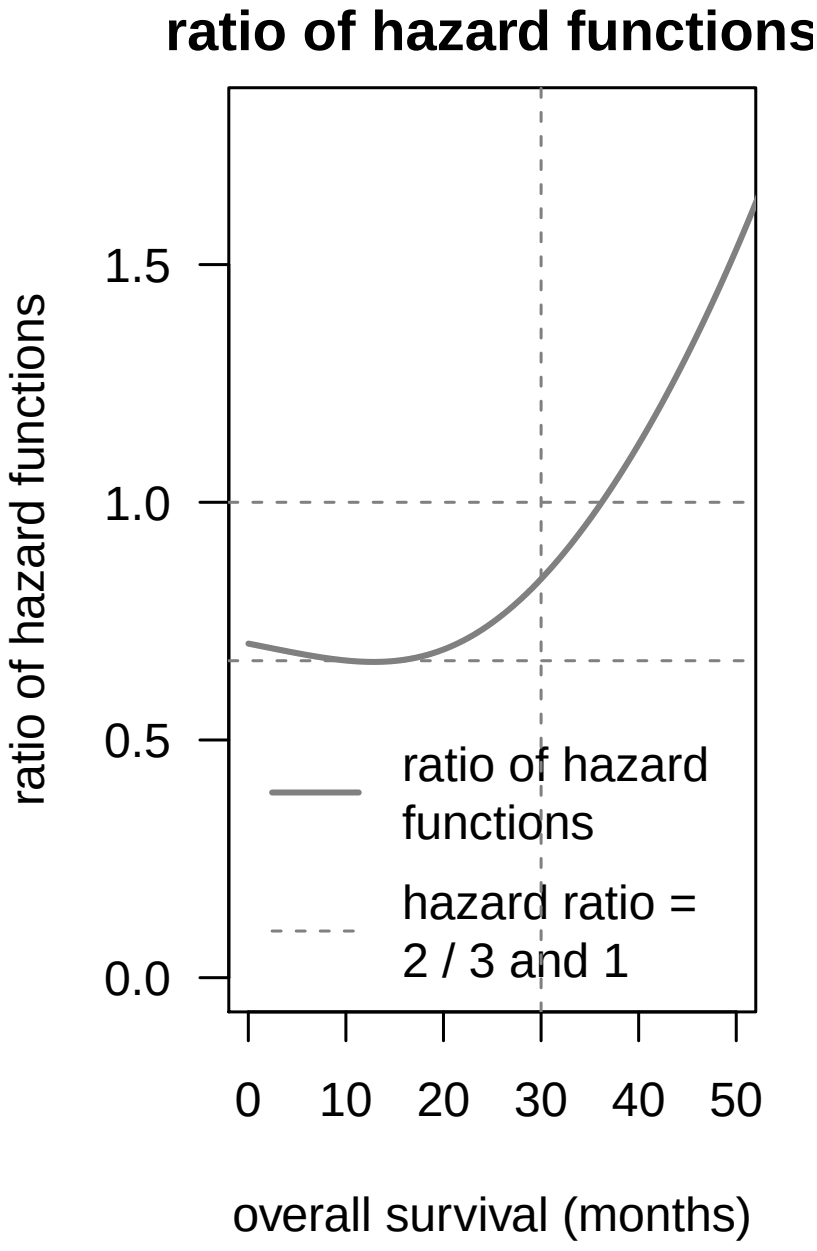
- Median OS 9m.
- Cure: 0.161.

Cure proportion model – assumptions



Cure proportion model – assumptions





Cure proportion model – sample size

To find sample size:

- Compute necessary events d_0 using Schoenfeld's formula.
- **Simulate** from assumed S_i's, compute power for grid of $d = d_0, \ldots, d_1$.
- Choose d such that (unweighted) logrank test gives targeted power.

MIRROS: 2-sided $\alpha = 0.05$, $\beta = 0.15$, some accrual and drop-out assumption.

Assumption	$S_1^{-1}(0.5)$	$S_2^{-1}(0.5)$	p_1	p_2	d	power	time

Cure proportion model – sample size

To find sample size:

- Compute necessary events d_0 using Schoenfeld's formula.
- **Simulate** from assumed S_i's, compute power for grid of $d = d_0, \ldots, d_1$.
- Choose d such that (unweighted) logrank test gives targeted power.

MIRROS: 2-sided $\alpha = 0.05$, $\beta = 0.15$, some accrual and drop-out assumption.

Assumption	$S_1^{-1}(0.5)$	$S_2^{-1}(0.5)$	p_1	p_2	d	power	time
MIRROS	6.0	9.0	0.080	0.161	275	0.852	**38.8**

Cure proportion model – sample size

To find sample size:

- Compute necessary events d_0 using Schoenfeld's formula.
- **Simulate** from assumed S_i's, compute power for grid of $d = d_0, \ldots, d_1$.
- Choose d such that (unweighted) logrank test gives targeted power.

MIRROS: 2-sided $\alpha = 0.05$, $\beta = 0.15$, some accrual and drop-out assumption.

Assumption	$S_1^{-1}(0.5)$	$S_2^{-1}(0.5)$	p_1	p_2	d	power	time
MIRROS	6.0	9.0	0.080	0.161	275	0.852	**38.8**
PH, no cure	6.0	9.0	0	0	246	0.858	**29.2**

Cure proportion model – sample size

To find sample size:

- Compute necessary events d_0 using Schoenfeld's formula.
- **Simulate** from assumed S_i's, compute power for grid of $d = d_0, \ldots, d_1$.
- Choose d such that (unweighted) logrank test gives targeted power.

MIRROS: 2-sided $\alpha = 0.05$, $\beta = 0.15$, some accrual and drop-out assumption.

Assumption	$S_1^{-1}(0.5)$	$S_2^{-1}(0.5)$	p_1	p_2	d	power	time
MIRROS	6.0	9.0	0.080	0.161	275	0.852	**38.8**
PH, no cure	6.0	9.0	0	0	246	0.858	**29.2**
MIRROS with #events for (PH, no cure)	6.0	9.0	0.080	0.161	246	**0.810**	33.7

Regulatory view on effect quantification

(European) health authorities: Emphasized many times that effect quantification in label must **not necessarily**

Regulatory view on effect quantification

(European) health authorities: Emphasized many times that effect quantification in label must **not necessarily**

- be tied to hypothesis test,

Regulatory view on effect quantification

(European) health authorities: Emphasized many times that effect quantification in label must **not necessarily**

- be tied to hypothesis test,
- provide inference with "significant" p-value.

Regulatory view on effect quantification

(European) health authorities: Emphasized many times that effect quantification in label must **not necessarily**

- be tied to hypothesis test,
- provide inference with "significant" p-value.

Reject H_0 using valid test.

Regulatory view on effect quantification

(European) health authorities: Emphasized many times that effect quantification in label must **not necessarily**

- be tied to hypothesis test,
- provide inference with "significant" p-value.

Reject H_0 using valid test.

Quantify effect using suitable **summary statistics**.

Cure proportion model – effect quantification

Cure proportion model – **no proportional hazards**. Unweighted logrank...

- ...**not most powerful** test, but loss modest (see above).
- ...**still valid** test, i.e. protects type I error.

How to quantify effect?

- **Kaplan-Meier** estimates provide entire information in data.
- Desire to summarize effect in one number.
- Hazard ratio from Cox regression and logrank test: if NPH, estimand and **power** depend on censoring distribution: accrual, dropout, follow-up pattern!

Rufibach (2019): extended discussion in **estimand** context.

Cure proportion model – estimation

Numerous parametric and nonparametric estimates of relevant quantities:
Cantor and Shuster (1992), Maller and Zhou (1992), Maller and Zhou (1996),
Tsodikov et al. (2003).

Obvious nonparametric estimate of cure proportion p, with $\widehat{S}$ Kaplan-Meier:

- $\widehat{S}(t_0)$ for some $t_0 > 0$.
- Maller and Zhou (1992): Kaplan-Meier evaluated at largest observed time, censored or event, consistently estimates p_0 under "sufficient follow-up" condition Tsodikov et al. (2003).
- Finite sample: likely not use latest observed time to evaluate the Kaplan-Meier estimate at. Rather **trade-off bias to reduce variability** of estimate.
- Choose milestone t_0 where clinically, cure seems very plausible.

What was done in MIRROS?

What was done in MIRROS?

Violation of PH only **very late**.

What was done in MIRROS?

Violation of PH only **very late**.

Power loss modest.

What was done in MIRROS?

Violation of PH only **very late**.

Power loss modest.

MIRROS statistical analysis plan:

What was done in MIRROS?

Violation of PH only **very late**.

Power loss modest.

MIRROS statistical analysis plan:

- Logrank test.

What was done in MIRROS?

Violation of PH only **very late**.

Power loss modest.

MIRROS statistical analysis plan:

- Logrank test.
- Hazard ratio.

What was done in MIRROS?

Violation of PH only **very late**.

Power loss modest.

MIRROS statistical analysis plan:

- Logrank test.
- Hazard ratio.
- Survival probabilities at milestones 6m, 12m, ...

What was done in MIRROS?

Violation of PH only **very late**.

Power loss modest.

MIRROS statistical analysis plan:

- Logrank test.
- Hazard ratio.
- Survival probabilities at milestones 6m, 12m, ...
- (Notorious) median OS.

What was NOT planned in MIRROS?

What was NOT planned in MIRROS?

Rerun of simulations with observed recruitment $\Rightarrow$ potential power impact.

Outcome of MIRROS

Trial was **negative**.

Outcome of MIRROS

Trial was **negative**.

Assumption on shape of S in treatment arm quite accurate.

Outcome of MIRROS

Trial was **negative**.

Assumption on shape of S in treatment arm quite accurate.

Relative effect vs. control not big enough.

Immunotherapy:

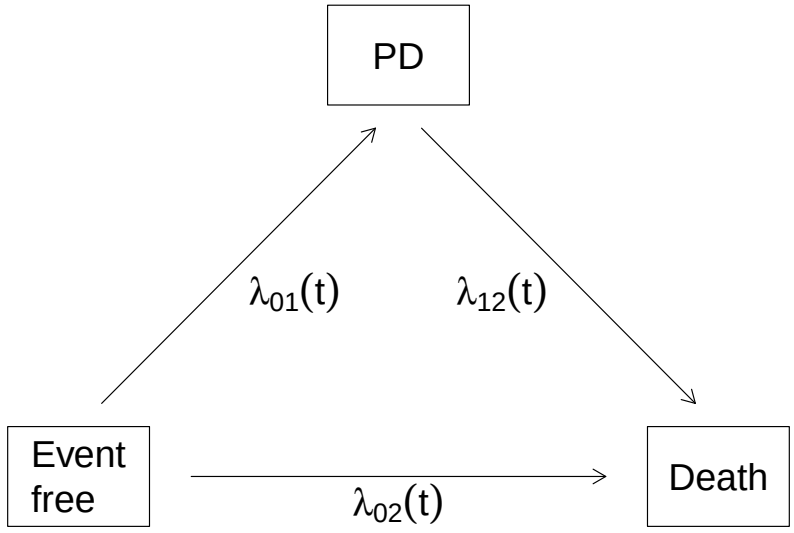
Immunotherapy:
1) no difference in PFS,

Immunotherapy:
1) no difference in PFS,
2) delayed effect for OS.

Immunotherapy:
1) no difference in PFS,
2) delayed effect for OS.

How to quantify effect?

Multistate model for PFS and OS



Standard **illness-death model without recovery**:

- Process $X(t) \in \{0, 1, 2\}, t \geq 0$ models the state occupied at time t.
- All patients in state 0 at time 0: $P(X(0) = 0) = 1$.
- PFS: waiting time in initial state 0, $\mathrm{PFS} = \inf\{t : X(t) \neq 0\}$.
- OS: time until reaching state 2, $\mathrm{OS} = \inf\{t : X(t) = 2\}$.

Multistate model formulation

Transition probabilities:

- **Full description** of multistate model by only assuming existence of intensities α_{01}, α_{02} and α_{12}.
- Formulas, even for **non-Markov** case: Aalen et al. (2008).

Multistate model formulation

Transition probabilities:

- **Full description** of multistate model by only assuming existence of intensities α_{01}, α_{02} and α_{12}.
- Formulas, even for **non-Markov** case: Aalen et al. (2008).

Meller et al. (2019):

- Formulas for P_{lm}'s assuming **Weibull** transition hazards for time-inhomogeneous Markov and semi-Markov.
- **Marginal** distributions:

$$\begin{aligned} S_{PFS}(t) &= P(\mathrm{PFS} > t) = P_{00}(0,t), \\ S_{OS}(t) &= P(\mathrm{OS} > t) = P_{00}(0,t) + P_{01}(0,t). \end{aligned}$$

- **Joint** distribution:

$$\begin{aligned} P(\mathrm{PFS} \leq u, \mathrm{OS} \leq v) &= P(X(u) \in \{1,2\}, X(v) = 2) \\ &= P(X(v) = 2 | X(u) = 1) \cdot P_{01}(0,u) + P_{02}(0,u). \end{aligned}$$

Multistate models to quantify effect when NPH?

Multistate models to quantify effect when NPH?

- NPH: focus on **hypothesis testing** or **effect quantification**?

Multistate models to quantify effect when NPH?

- NPH: focus on **hypothesis testing** or **effect quantification**?
- If only **one number** should comprehensively quantify effect:

Multistate models to quantify effect when NPH?

- NPH: focus on **hypothesis testing** or **effect quantification**?
- If only **one number** should comprehensively quantify effect:
  - Assume effect is time-constant $\Rightarrow$ PH.

Multistate models to quantify effect when NPH?

- NPH: focus on **hypothesis testing** or **effect quantification**?
- If only **one number** should comprehensively quantify effect:
  - Assume effect is time-constant $\Rightarrow$ PH.
  - Average over time.

Multistate models to quantify effect when NPH?

- NPH: focus on **hypothesis testing** or **effect quantification**?
- If only **one number** should comprehensively quantify effect:
  - Assume effect is time-constant $\Rightarrow$ PH.
  - Average over time.
  - Pick a timepoint for comparison.

Multistate models to quantify effect when NPH?

- NPH: focus on **hypothesis testing** or **effect quantification**?
- If only **one number** should comprehensively quantify effect:
  - Assume effect is time-constant $\Rightarrow$ PH.
  - Average over time.
  - Pick a timepoint for comparison.
- If >1 number "allowed" - which one to pick?

Multistate models to quantify effect when NPH?

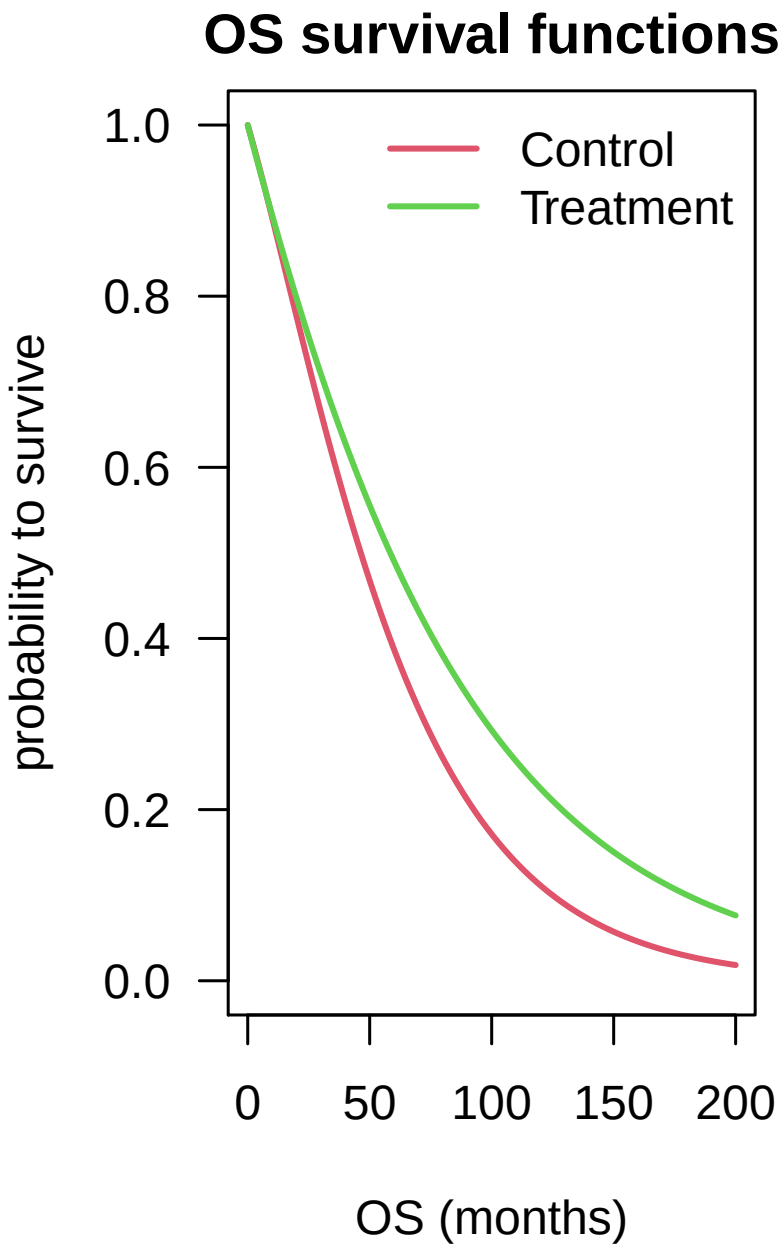
- NPH: focus on **hypothesis testing** or **effect quantification**?
- If only **one number** should comprehensively quantify effect:
  - Assume effect is time-constant $\Rightarrow$ PH.
  - Average over time.
  - Pick a timepoint for comparison.
- If >1 number "allowed" - which one to pick?
- Effect **transparent** to, e.g., patient?

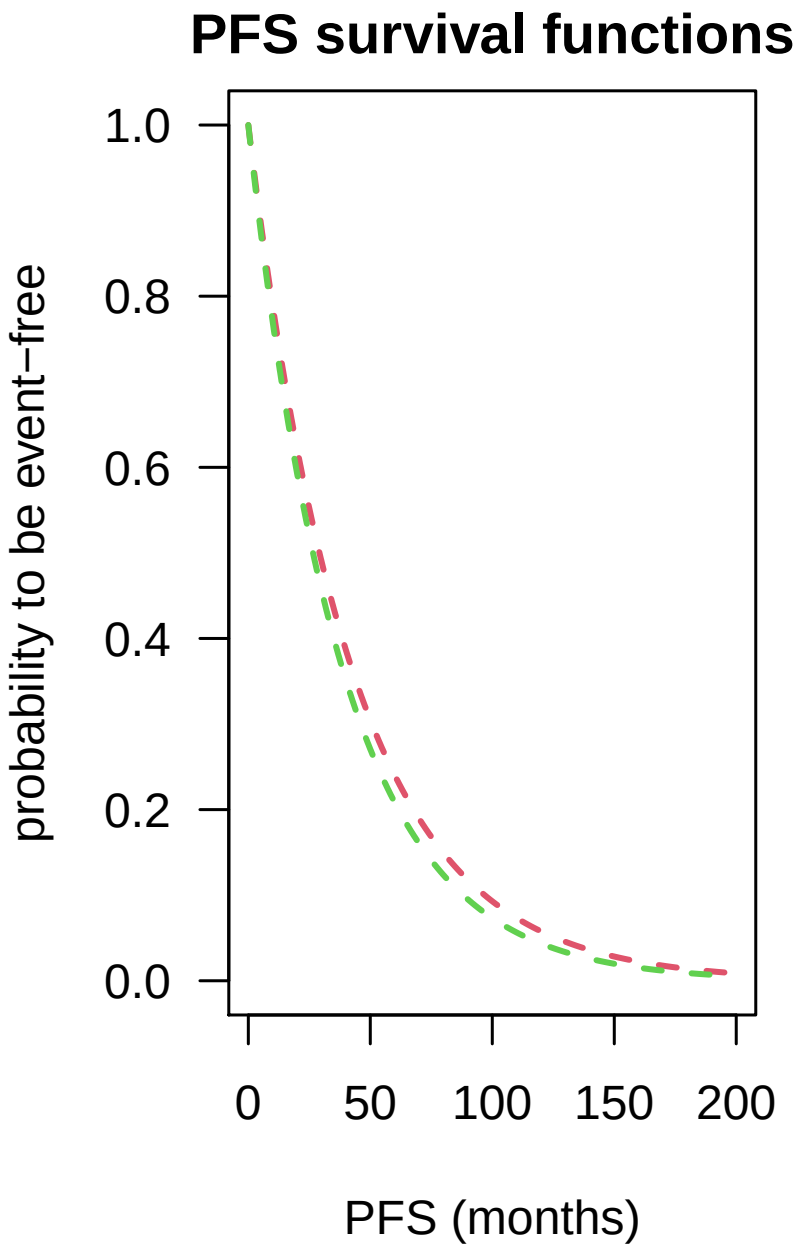
Multistate models to quantify effect when NPH?

- NPH: focus on **hypothesis testing** or **effect quantification**?
- If only **one number** should comprehensively quantify effect:
  - Assume effect is time-constant $\Rightarrow$ PH.
  - Average over time.
  - Pick a timepoint for comparison.
- If >1 number "allowed" - which one to pick?
- Effect **transparent** to, e.g., patient?

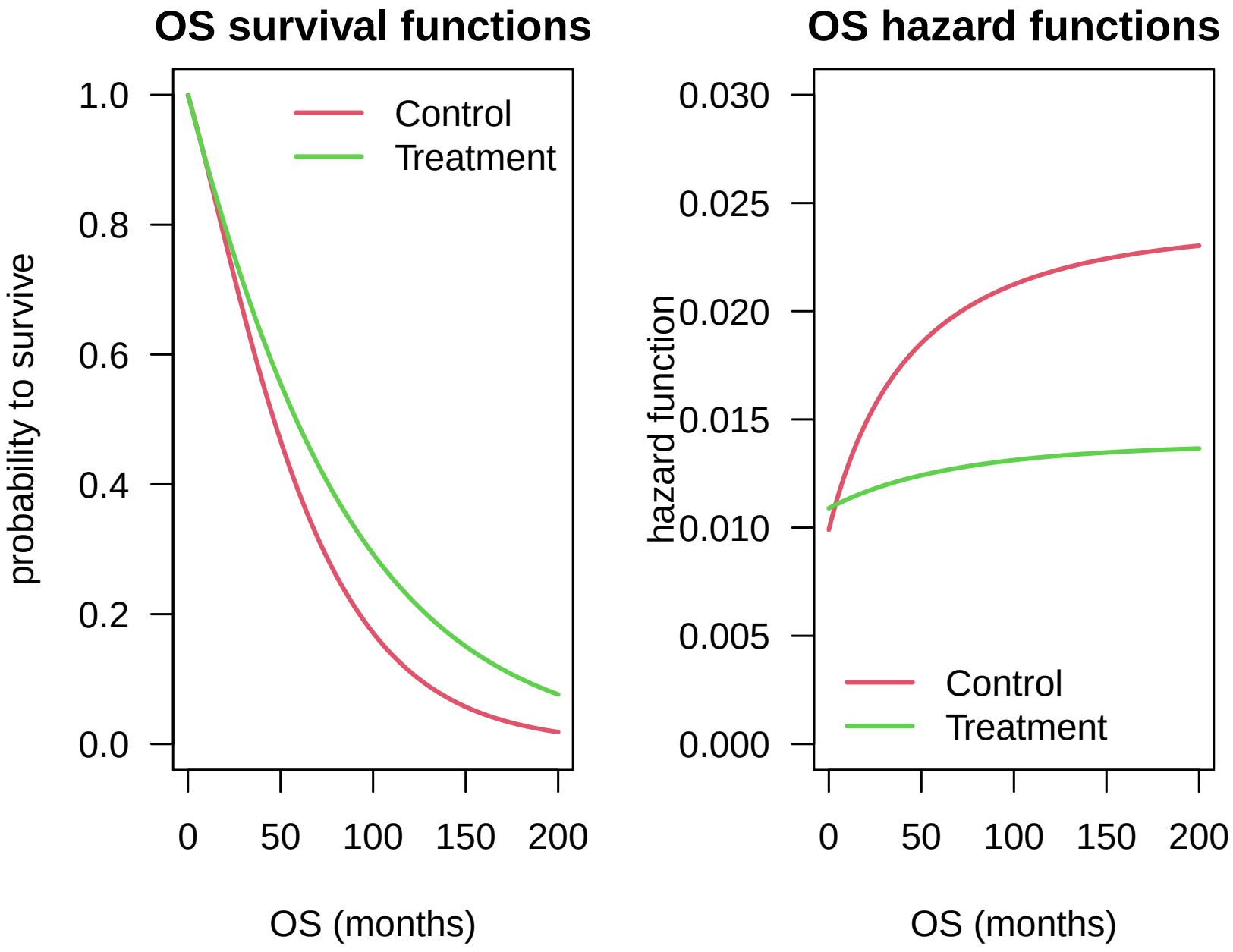
Proposal	focus	parameters to pre-specify	interpretation
Piecewise exponential hazard	estimation	interval limits, hazard ratio on each interval	✓
Subgroupwise hazard	estimation	prevalence of each sub-group, hazard ratio in each group	✓
Max-combo tests	testing	number of weight functions, one hazard ratio	?
RMST	both	upper limit, effect size	recalibration needed

(True) PFS and OS for hypothetical clinical trial

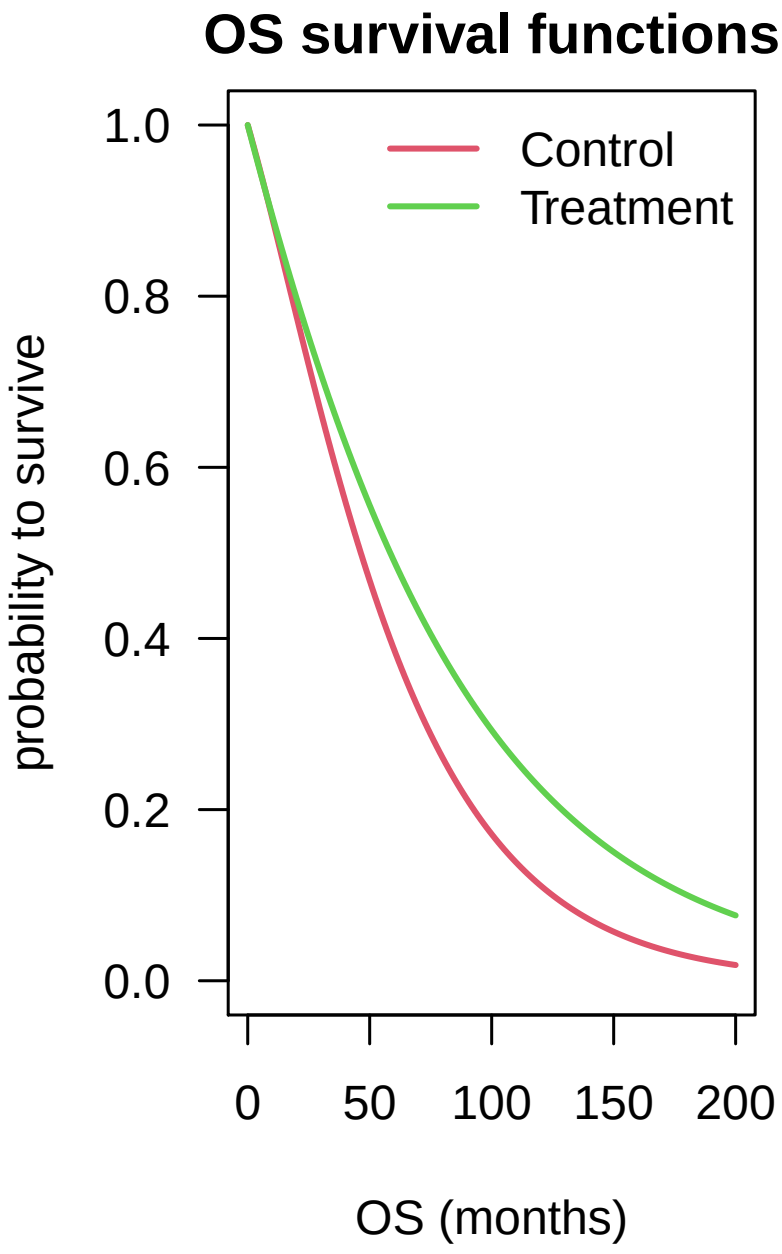


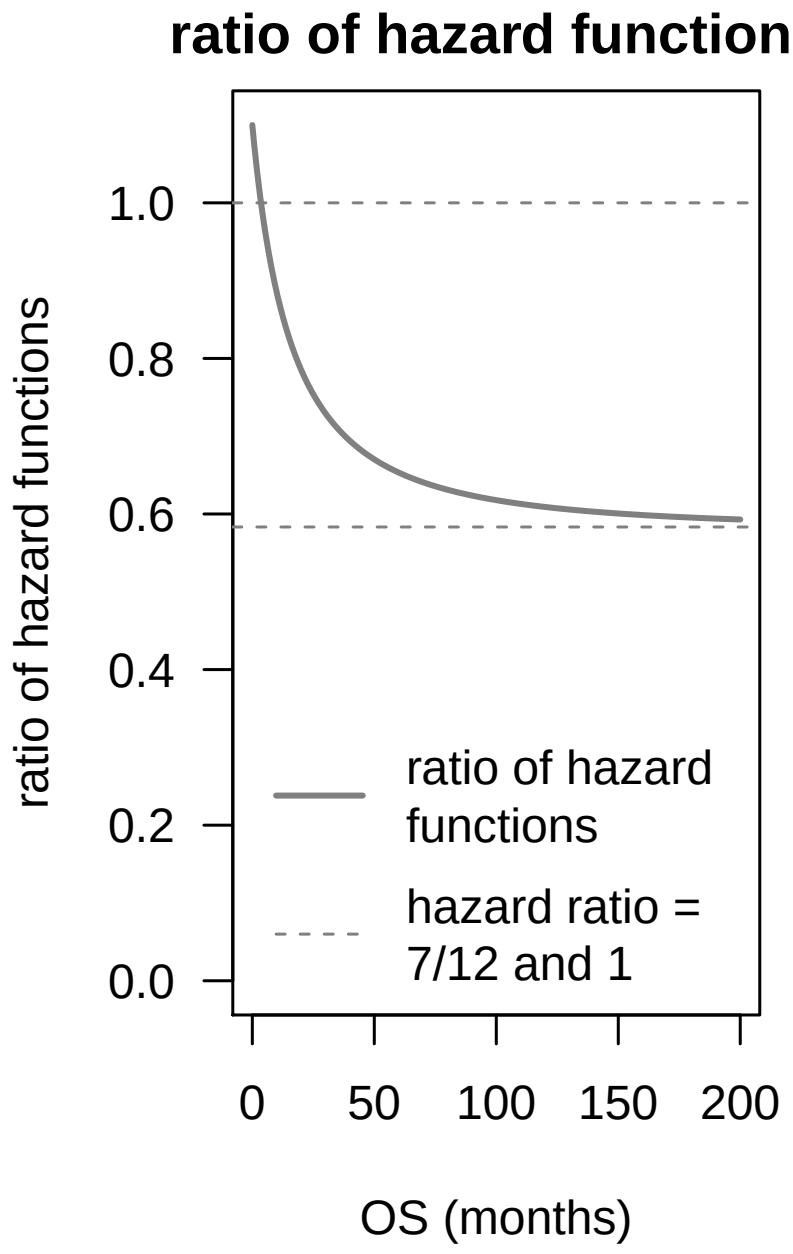


(True) OS for hypothetical clinical trial



(True) OS for hypothetical clinical trial





Data was generated according to...

Transition	Control arm	Treatment arm

Data was generated according to...

Transition	Control arm	Treatment arm
$0 \to 1$	$\lambda_{01}^{c} = \log(2)/50$	$\lambda_{01}^{t} = \lambda_{01}^{c} \cdot \mathbf{1.1}$

Data was generated according to...

Transition	Control arm	Treatment arm
$0 \to 1$	$\lambda_{01}^c = \log(2)/50$	$\lambda_{01}^t = \lambda_{01}^c \cdot \mathbf{1.1}$
$0 \to 2$	$\lambda_{02}^c = \log(2)/70$	$\lambda_{02}^t = \lambda_{02}^c \cdot \mathbf{1.1}$

Data was generated according to...

Transition	Control arm	Treatment arm
$0 \to 1$	$\lambda_{01}^c = \log(2)/50$	$\lambda_{01}^t = \lambda_{01}^c \cdot \mathbf{1.1}$
$0 \to 2$	$\lambda_{02}^c = \log(2)/70$	$\lambda_{02}^t = \lambda_{02}^c \cdot \mathbf{1.1}$
$1 \to 2$	$\lambda_{12}^c = \log(2)/20$	$\lambda_{12}^t = \lambda_{12}^c \cdot \mathbf{0.4}$

Multistate models for NPH?

Multistate models for NPH:

Multistate models for NPH?

Multistate models for NPH:

- Potential option with **straightforward** (?) effect quantification.

Multistate models for NPH?

Multistate models for NPH:

- Potential option with **straightforward** (?) effect quantification.
- Easy to plan trial via simulation.

Multistate models for NPH?

Multistate models for NPH:

- Potential option with **straightforward** (?) effect quantification.
- Easy to plan trial via simulation.
- PD $\rightarrow$ death transition non-randomized comparison!

Multistate models for NPH?

Multistate models for NPH:

- Potential option with **straightforward** (?) effect quantification.
- Easy to plan trial via simulation.
- PD $\rightarrow$ death transition non-randomized comparison!

Open questions - **ongoing research**:

Multistate models for NPH?

Multistate models for NPH:

- Potential option with **straightforward** (?) effect quantification.
- Easy to plan trial via simulation.
- PD $\rightarrow$ death transition non-randomized comparison!

Open questions - **ongoing research**:

- How to define **trial success**? Combine transition-specific test statistics.

Multistate models for NPH?

Multistate models for NPH:

- Potential option with **straightforward** (?) effect quantification.
- Easy to plan trial via simulation.
- PD $\rightarrow$ death transition non-randomized comparison!

Open questions - **ongoing research**:

- How to define **trial success**? Combine transition-specific test statistics.
- Interim analyses?

Multistate models for NPH?

Multistate models for NPH:

- Potential option with **straightforward** (?) effect quantification.
- Easy to plan trial via simulation.
- PD $\rightarrow$ death transition non-randomized comparison!

Open questions - **ongoing research**:

- How to define **trial success**? Combine transition-specific test statistics.
- Interim analyses?
- How to inform transition-specific quantities?

Conclusions

Conclusions

Conclusions

- Account for **power loss** and **cutoff delay** if you have cure proportions (or NPH).

Conclusions

- Account for **power loss** and **cutoff delay** if you have cure proportions (or NPH).
- NPH $\Rightarrow$ large zoo of alternatives $\Rightarrow$ assumption needed on shape of survival functions $\Rightarrow$ use **simulations extensively**!

Conclusions

- Account for **power loss** and **cutoff delay** if you have cure proportions (or NPH).
- NPH $\Rightarrow$ large zoo of alternatives $\Rightarrow$ assumption needed on shape of survival functions $\Rightarrow$ use **simulations extensively**!
- Think about **how to quantify effect**.

Conclusions

- Account for **power loss** and **cutoff delay** if you have cure proportions (or NPH).
- NPH $\Rightarrow$ large zoo of alternatives $\Rightarrow$ assumption needed on shape of survival functions $\Rightarrow$ use **simulations extensively**!
- Think about **how to quantify effect**.
- Number of events: metric related to PH! **Delayed separation** $\Rightarrow$ number of events not necessarily sufficiently informative.

Conclusions

- Account for **power loss** and **cutoff delay** if you have cure proportions (or NPH).
- NPH $\Rightarrow$ large zoo of alternatives $\Rightarrow$ assumption needed on shape of survival functions $\Rightarrow$ use **simulations extensively**!
- Think about **how to quantify effect**.
- Number of events: metric related to PH! **Delayed separation** $\Rightarrow$ number of events not necessarily sufficiently informative.
- Power optimization $\Leftrightarrow$ assumptions might also be off!

Conclusions

- Account for **power loss** and **cutoff delay** if you have cure proportions (or NPH).
- NPH $\Rightarrow$ large zoo of alternatives $\Rightarrow$ assumption needed on shape of survival functions $\Rightarrow$ use **simulations extensively**!
- Think about **how to quantify effect**.
- Number of events: metric related to PH! **Delayed separation** $\Rightarrow$ number of events not necessarily sufficiently informative.
- Power optimization $\Leftrightarrow$ assumptions might also be off!
- Mitigation against (large) family of alternatives appears meaningful.

Conclusions

- Account for **power loss** and **cutoff delay** if you have cure proportions (or NPH).
- NPH $\Rightarrow$ large zoo of alternatives $\Rightarrow$ assumption needed on shape of survival functions $\Rightarrow$ use **simulations extensively**!
- Think about **how to quantify effect**.
- Number of events: metric related to PH! **Delayed separation** $\Rightarrow$ number of events not necessarily sufficiently informative.
- Power optimization $\Leftrightarrow$ assumptions might also be off!
- Mitigation against (large) family of alternatives appears meaningful.
- **Multistate models** may offer way to plan trial and quantify effect if NPH.

Conclusions

- Account for **power loss** and **cutoff delay** if you have cure proportions (or NPH).
- NPH $\Rightarrow$ large zoo of alternatives $\Rightarrow$ assumption needed on shape of survival functions $\Rightarrow$ use **simulations extensively**!
- Think about **how to quantify effect**.
- Number of events: metric related to PH! **Delayed separation** $\Rightarrow$ number of events not necessarily sufficiently informative.
- Power optimization $\Leftrightarrow$ assumptions might also be off!
- Mitigation against (large) family of alternatives appears meaningful.
- **Multistate models** may offer way to plan trial and quantify effect if NPH.

Power optimization $\Leftrightarrow$ pragmatism.

Resources

MIRROS trial design:

- Paper with Dominik Heinzmann and Annabelle Monnet: Rufibach et al. (2020).
- Reproduce simulations and **plan your own trial**: `https://github.com/numbersman77/integratePhase2.git`.

Multistate model for PFS and OS:

- Paper with Matthias Meller and Jan Beyersmann: Meller et al. (2019).

Thank you for your attention.

kaspar.rufibach@roche.com

http://www.kasparrufibach.ch

numbersman77

numbersman77

References I

- Aalen, O., Borgan, **Ø**. and Gjessing, H. (2008). *Survival and event history analysis: a process point of view*. Springer Science & Business Media.
- Aalen, O. O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal* **1987** 177–190.
- Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics* **5** 141–150.
- Andersen, P. K., Borgan, **Ø**., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer.
- Beyersmann, J., Allignol, A. and Schumacher, M. (2012). *Competing Risks and Multistate Models with R*. Springer.
- Bose, P., Vachhani, P. and Cortes, J. E. (2017). Treatment of relapsed/refractory acute myeloid leukemia. *Current Treatment Options in Oncology* **18** 17.
- Breems, D. A., Van Putten, W. L., Huijgens, P. C., Ossenkoppele, G. J., Verhoef, G. E., Verdonck, L. F., Vellenga, E., De Greef, G. E., Jacky, E., Van der Lelie, J., Boogaerts, M. A. and Lowenberg, B. (2005). Prognostic index for adult patients with acute myeloid leukemia in first relapse. *J. Clin. Oncol.* **23** 1969–1978.
- Cantor, A. B. and Shuster, J. J. (1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Stat Med* **11** 931–937.

References II

- Fleischer, F., Gaschler-Markefski, B. and Bluhmki, E. (2009). A statistical model for the dependence between progression-free survival and overall survival. *Stat. Med.* **28** 2669–2686.
- Hunsberger, S., Zhao, Y. and Simon, R. (2009). A comparison of phase II study strategies. *Clin. Cancer Res.* **15** 5950–5955.
- Li, Y. and Zhang, Q. (2015). A Weibull multi-state model for the dependence of progression-free survival and overall survival. *Stat Med* **34** 2497–2513.
- Maller, R. and Zhou, X. (1996). *Survival Analysis with Long-term Survivors*. John Wiley & Sons, United States.
- Maller, R. A. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika* **79** 731–739.
- Meller, M., Beyersmann, J. and Rufibach, K. (2019). Joint modeling of progression-free and overall survival and computation of correlation measures. *Statistics in medicine* **38** 4270–4289.
- Parmar, M. K., Barthel, F. M., Sydes, M., Langley, R., Kaplan, R., Eisenhauer, E., Brady, M., James, N., Bookman, M. A., Swart, A. M., Qian, W. and Royston, P. (2008). Speeding up the evaluation of new agents in cancer. *J. Natl. Cancer Inst.* **100** 1204–1214.
- Rufibach, K. (2019). Treatment effect quantification for time-to-event endpoints – estimands, analysis strategies, and beyond. *Pharmaceutical Statistics* **18** 144–164.

References III

- Rufibach, K., Heinzmann, D. and Monnet, A. (2020). Integrating phase 2 into phase 3 based on an intermediate endpoint while accounting for a cure proportion – with an application to the design of a clinical trial in acute myeloid leukemia. *Pharmaceutical Statistics* **19** 44–58. Code available on github: `https://github.com/numbersman77/integratePhase2.git`.
- Sun, S., Liu, G., Lyu, T., Xue, F., Yeh, T. M. and Rao, S. (2018). Design considerations in clinical trials with cure rate survival data: A case study in oncology. *Pharm Stat* **17** 94–104.
- Tsodikov, A. D., Ibrahim, J. G. and Yakovlev, A. Y. (2003). Estimating Cure Rates From Survival Data: An Alternative to Two-Component Mixture Models. *J Am Stat Assoc* **98** 1063–1078.
- Weber, E. M. and Titman, A. C. (2019). Quantifying the association between progression-free survival and overall survival in oncology trials using kendall's τ. *Statistics in medicine* **38** 703–719.

Backup slides.

Standard of care

No standard regimen for relapsed or refractory (R/R) AML. Breems et al. (2005)

No new drug approved for treatment of AML in **over 50 years**! Bose et al. (2017)

Standard of care

No standard regimen for relapsed or refractory (R/R) AML. Breems et al. (2005)

No new drug approved for treatment of AML in **over 50 years**! Bose et al. (2017)

THIS is unmet medical need!

Idasanutlin

p53: Tumor suppressor, many mechanisms of anticancer function.

Mouse double minute 2 homolog (MDM2): Negative regulator of p53 tumor suppressor.

Idasanutlin: binds to MDM2 $\Rightarrow$ prevents p53 - MDM2 interaction $\Rightarrow$ (re-)activation of p53 $\Rightarrow$ **reinstalls anti-tumor capacity of p53**.

Clinical development plan for Idasanutlin

Need for acceleration:

- Very high unmet medical need in R/R AML.
- Early phase results with Idasanutlin encouraging.
- Competitive landscape and economic constraints: Lean program only way to receive internal approval for pivotal trial.
- Willingness to trade-off risk reduction from randomized P2 against increased speed.

Skip or integrate Phase 2?

Assume we have **successful P1**.

Purpose of futility interim: optimize **P(stopping @ interim | H_0)**.

Hunsberger et al. (2009):

- **Integrate** P2 into P3: futility interim based on **intermediate** endpoint.
- **Skip** P2: futility interim based on **P3 primary** endpoint.

If trial

- stops at futility interim: basically performed randomized P2.
- passes futility interim: P3 pivotal trial well on its way.

Key advantage of setup: Decision to proceed to full P3 part based on randomized comparison. Parmar et al. (2008)

Futility interim analysis

Mitigate risk if drug does not work (sufficiently).

Planned after **120** patients are recruited.

Why not use OS for interim decision?

- 53 (under H_0) and 46 deaths (under H_1) expected at interim. Substantial uncertainty.
- Cures have not happened yet at the interim.
- Confounding by early (mainly safety-related) deaths.

Bottom line: interim is **too early for OS** to be meaningful endpoint.

Intermediate endpoint

Complete response:

- Sufficiently associated with OS.
- CR **necessary** for good OS / cure: Patient needs CR to have chance for cure, via bridge-to-transplant.
- Odds ratio as effect measure.

Intermediate endpoint

Complete response:

- Sufficiently associated with OS.
- CR **necessary** for good OS / cure: Patient needs CR to have chance for cure, via bridge-to-transplant.
- Odds ratio as effect measure.

Futility interim is **non-binding**.

Intermediate endpoint

Complete response:

- Sufficiently associated with OS.
- CR **necessary** for good OS / cure: Patient needs CR to have chance for cure, via bridge-to-transplant.
- Odds ratio as effect measure.

Futility interim is **non-binding**. Why do we need to model it at all?

Intermediate endpoint

Complete response:

- Sufficiently associated with OS.
- CR **necessary** for good OS / cure: Patient needs CR to have chance for cure, via bridge-to-transplant.
- Odds ratio as effect measure.

Futility interim is **non-binding**. Why do we need to model it at all?

- How to choose interim boundary on CR?

Intermediate endpoint

Complete response:

- Sufficiently associated with OS.
- CR **necessary** for good OS / cure: Patient needs CR to have chance for cure, via bridge-to-transplant.
- Odds ratio as effect measure.

Futility interim is **non-binding**. Why do we need to model it at all?

- How to choose interim boundary on CR?
- Decision-makers want to be able to trade-off

$$\textbf{False Positive} = \mathrm{P}(\text{continue @ interim} \mid H_0)$$

vs.

$$\textbf{False Negative} = \mathrm{P}(\text{stop @ interim} \mid H_1).$$

Intermediate endpoint

Complete response:

- Sufficiently associated with OS.
- CR **necessary** for good OS / cure: Patient needs CR to have chance for cure, via bridge-to-transplant.
- Odds ratio as effect measure.

Futility interim is **non-binding**. Why do we need to model it at all?

- How to choose interim boundary on CR?
- Decision-makers want to be able to trade-off

$$\textbf{False Positive} = \mathrm{P}(\text{continue @ interim} \mid H_0)$$

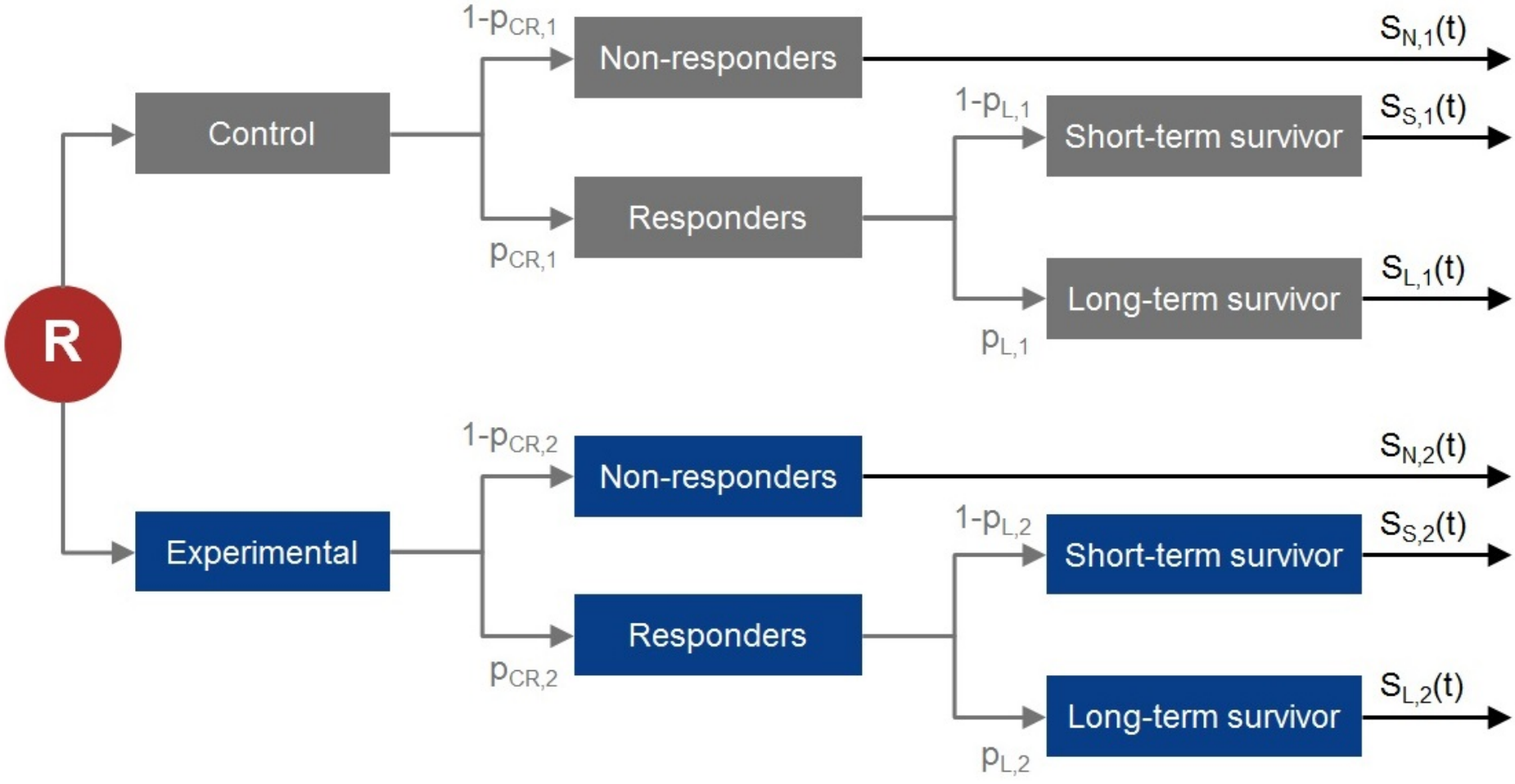
vs.

$$\textbf{False Negative} = \mathrm{P}(\text{stop @ interim} \mid H_1).$$

If futility based on OS $\Rightarrow$ conditional power.

If CR is intermediate endpoint: **mechanistic simulation model**.

Mechanistic simulation model



Mechanistic simulation model

Connects CR to OS.

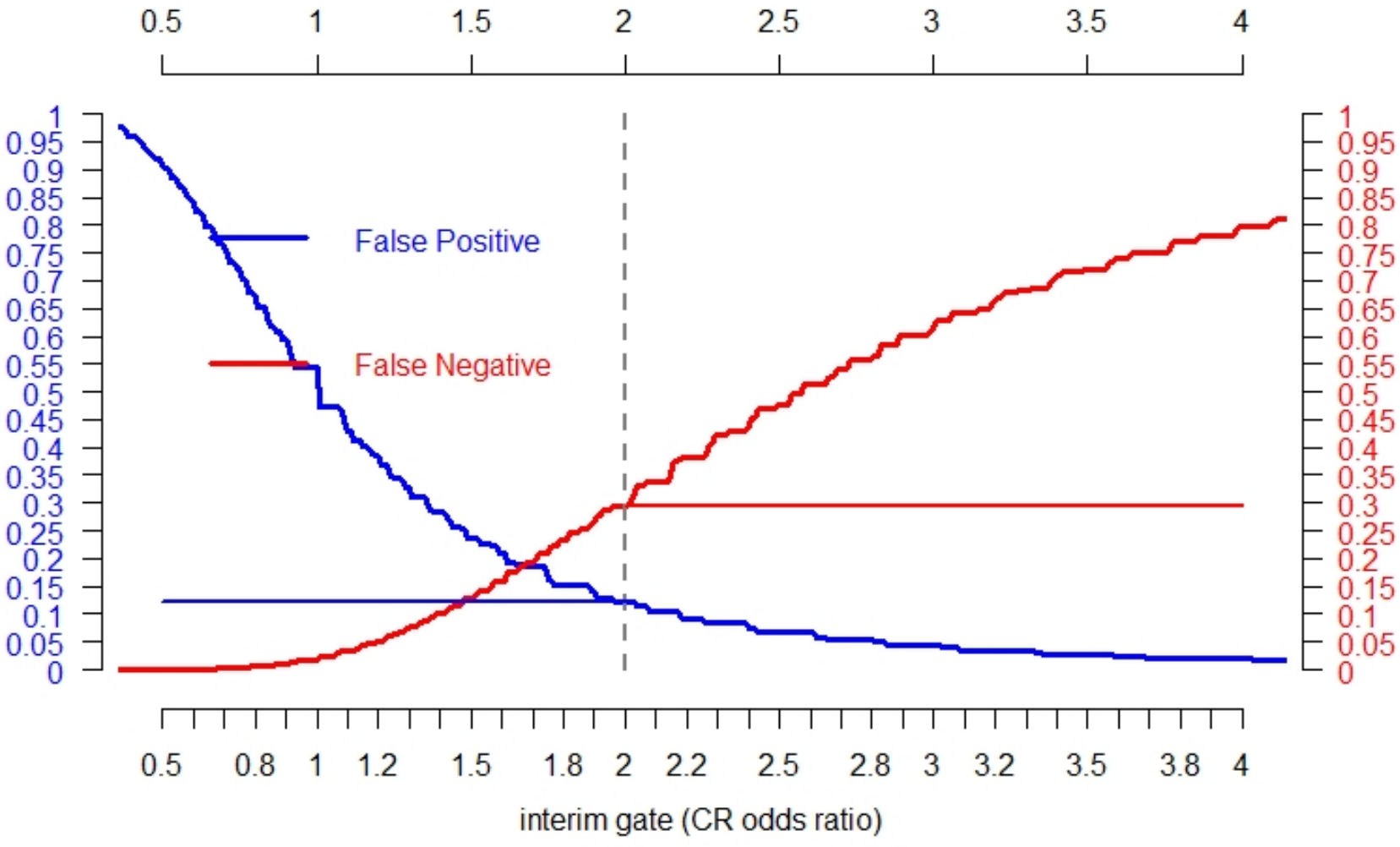
Need to inform all assumptions:

Quantity	Control arm	Treatment arm
Survival function of non-responders	$S_{N,1}$	$S_{N,2}$
Probability to have CR	$p_{CR,1}$	$p_{CR,2}$
Probability to be long-term responder | CR	$p_{L,1}$	$p_{L,2}$
Survival function of short-term responders	$S_{S,1}$	$S_{S,2}$
Survival function of long-term responders	$S_{L,1}$	$S_{L,2}$
#patients recruited per month	n_{1j}	n_{2j}
Months of recruitment	$j = 1, \ldots, N$	
Total #patients recruited	$n_1 = \sum_{j=1}^{N} n_{1j}$	$n_2 = \sum_{j=1}^{N} n_{2j}$
Drop-out rate per month	τ_1	τ_2

Align parameters such that **mechanistic simulation model can reproduce sample size**!

P(CR) control: 0.16. Assume OR = 2.5 to improve on this with treatment $\Rightarrow$
P(CR tmt) = 0.323. P(longterm survivor) = 0.5. This gives cure proportions.

Operating characteristics of various interim boundaries



False Positive = P(continue @ interim | no effect)
False Negative = P(stop @ interim | alternative used for powering)

Operating characteristics of various interim boundaries

Sweet spot: **odds ratio of 2**,

- False Positive = P(continue @ interim | no effect) $\approx$ 12%,
- False Negative = P(stop @ interim | alternative assumed for powering) $\approx$ 30%.

Interim decision:

- Based on independent data monitoring committee (iDMC) recommendation, i.e. sponsor **blinded**,
- **non-binding**,
- included safety criterion (molecule class toxicity) and criteria for early deaths $\Rightarrow$ OS component.

Power loss of adding futility interim

Can easily get that from simulations.

Power loss of adding futility interim

Can easily get that from simulations.

- Targeted power: 85%.

Power loss of adding futility interim

Can easily get that from simulations.

- Targeted power: 85%.
- Power taking into account futility interim:

Power loss of adding futility interim

Can easily get that from simulations.

- Targeted power: 85%.
- Power taking into account futility interim: **63%!**

Power loss of adding futility interim

Can easily get that from simulations.

- Targeted power: 85%.
- Power taking into account futility interim: **63%!**
- Illustrates risk-appetite. Futility interim somehow becomes "informal efficacy interim".

Power loss of adding futility interim

Can easily get that from simulations.

- Targeted power: 85%.
- Power taking into account futility interim: **63%!**
- Illustrates risk-appetite. Futility interim somehow becomes "informal efficacy interim".
- Do we always compute the power loss when adding futility interims? Do we **increase number of events** to account for it?

Power loss of adding futility interim

Can easily get that from simulations.

- Targeted power: 85%.
- Power taking into account futility interim: **63%!**
- Illustrates risk-appetite. Futility interim somehow becomes "informal efficacy interim".
- Do we always compute the power loss when adding futility interims? Do we **increase number of events** to account for it?

Who cares anyway $\Rightarrow$ interim **passed**!

Implementation features

A (industry) clinical trial is **not a pre-specified static** undertaking!

- Not clear whether p53 mutant patients (≈15%) also benefit from Idasanutlin.
  - Still included, as evidence unclear and high unmet medical need.
  - But testing too late for randomization, i.e. could not stratify for p53 status.
  - Adds uncertainty to recruitment assumptions.
- Decision-makers sceptical about interim gate based on CR only. Additionally engineered EFS criterion (not discussed here).
- Evolvement of gating criteria:

Date	Milestone	OR CR $\geq$ 2.5	OR CR $\geq$ 2 + EFS HR $\leq$ 1	OS HR $\leq$ 0.9	OS HR $\leq$ 0.8
22.04.2014	CHMP meeting	x	x		
27.01.2015	FDA type C mtg	x	x		
08.04.2015	LSPC team proposal			x	
09.04.2015	LSPC decision				x
24.04.2015	CHMP BP	x	x		x
27.08.2015	LSCP decision	x	x		

Implementation features

A (industry) clinical trial is **not a pre-specified static** undertaking!

- Biomarker development: typically in Phase 2! Recommendation on biomarker development by iDMC.
- Seamless designs in general: sponsor does not get to see data for a **long time**. Unease for decision-makers.
- No accrual suspension for interim $\Rightarrow$ data cleaning and decision needs to come fast.

Health authority feedback

FDA:

- **Preferred randomized P2**.
- Challenged lack of stratification on p53 mutation status.
- Companion Diagnostic component with blinded P2 data $\Rightarrow$ not clear how to decide on development.
- **Challenged assumptions**, asked for additional sensitivity analyses.
- Concerns of early events driving interim analysis. OS not part of futility decision, but early tox deaths are.
- US sites only opened after passing the IA.

EMA:

- Agreed to accelerated development due to high unmet need.
- PH assumption discussed, support hazard ratio as appropriate effect measure.

Why two models?

We have two models:

- Cure proportion model to derive sample size,
- mechanistic simulation model to explore interim operating characteristics.

Why?

Reasons:

- Futility interim analysis has no implication on type I error $\Rightarrow$ independent of key design characteristic.
- Cure proportion model:
  - Simple,
  - depends on less assumptions than mechanistic model,
  - Robust model to plan sample size.
- Mechanistic simulation model:
  - Interim setup has potential to be changed before or while study is running. Prefer not to have these changes interfere with sample size.
  - Only used for (internal) decision-making via iDMC, no filing relevance $\Rightarrow$ can "afford" more modeling.

Advantages of multistate model

Multistate model:

Advantages of multistate model

Multistate model:

- Assumptions on $X(t)$ induce properties of transition intensities, (joint) probabilites, and thus PFS and OS.

Advantages of multistate model

Multistate model:

- Assumptions on $X(t)$ induce properties of transition intensities, (joint) probabilites, and thus PFS and OS.
- No progression after death.

Advantages of multistate model

Multistate model:

- Assumptions on $X(t)$ induce properties of transition intensities, (joint) probabilites, and thus PFS and OS.
- No progression after death.
- PFS = OS easily possible.

Advantages of multistate model

Multistate model:

- Assumptions on $X(t)$ induce properties of transition intensities, (joint) probabilites, and thus PFS and OS.
- No progression after death.
- PFS = OS easily possible.
- Estimation: No assumption about "in-/dependence" of PFS and OS.

Advantages of multistate model

Multistate model:

- Assumptions on $X(t)$ induce properties of transition intensities, (joint) probabilites, and thus PFS and OS.
- No progression after death.
- PFS = OS easily possible.
- Estimation: No assumption about "in-/dependence" of PFS and OS.

Multistate = (most?) parsimonious model

Correlation coefficient

$$\mathrm{Corr}(\mathrm{PFS}, \mathrm{OS}) \quad = \quad \frac{\mathrm{Cov}(\mathrm{PFS}, \mathrm{OS})}{\sqrt{\mathrm{Var}(\mathrm{PFS})\,\mathrm{Var}(\mathrm{OS})}} \quad = \quad \frac{\mathbb{E}(\mathrm{PFS} \cdot \mathrm{OS}) - \mathbb{E}(\mathrm{PFS})\,\mathbb{E}(\mathrm{OS})}{\sqrt{\mathrm{Var}(\mathrm{PFS})\,\mathrm{Var}(\mathrm{OS})}}.$$

Mean, variance of PFS and OS: via survival functions.

$\mathbb{E}(\mathrm{PFS} \cdot \mathrm{OS})$: Use

$$P(\mathrm{PFS} \cdot \mathrm{OS} > t) \quad = \quad P(\mathrm{PFS} > \sqrt{t}) + \int_{(0,\sqrt{t}]} P_{11}(u, t/u; u) P(\mathrm{PFS} > u-) \alpha_{01}(u)\ \mathrm{d}u.$$

Proof: manipulations using law of total probability.

Estimation and inference for Markov models

Parametric:

- Plug parametric assumption in formulas for $P_{lm}(s,t)$, S_{PFS}, S_{OS}, $\mathrm{Corr}(\mathrm{PFS},\mathrm{OS})$.
- Estimate parameters using **Counting Process Likelihood**, Andersen et al. (1993). Product of patient-specific likelihood-contributions to each state transition.
- Inference via delta method or bootstrap (results comparable).

Nonparametric:

- Transition probabilities: Aalen-Johansen estimator, Aalen and Johansen (1978).
- Plug in estimates into formulas for $\mathrm{PFS}, \mathrm{OS}, \mathrm{Corr}(\mathrm{PFS},\mathrm{OS})$.
- Challenge: need to **extrapolate tail beyond where we have data**.
- Inference via bootstrap.

Estimation and inference for Markov models

LFTM in Fleischer et al. (2009) and Li and Zhang (2015):

- Group patients depending on their path from 0 to 1 or 2, or censored.
- Likelihood uses **assumption of independence** of TTP, OS_{orig}. Cannot tell from (even uncensored!) data! Aalen (1987): "artificial problem", as LFTM not needed, see also Beyersmann et al. (2012).

Weber and Titman (2019):

- Kendall's τ, based on multistate, nonparametric, and copula models.
- Use again LFTM for estimation.

Multistate model formulation

Transition probabilities:

- **Full description** of multistate model by only assuming existence of intensities α_{01}, α_{02} and α_{12}.
- Formulas, even for **non-Markov** case: Aalen et al. (2008).

Multistate model formulation

Transition probabilities:

- **Full description** of multistate model by only assuming existence of intensities α_{01}, α_{02} and α_{12}.
- Formulas, even for **non-Markov** case: Aalen et al. (2008).

Meller et al. (2019):

- Embed PFS and OS in multistate model framework,
- formulas for P_{lm}'s assuming **Weibull** transition hazards for time-inhomogeneous Markov and semi-Markov (explicit),
- inference via **counting process likelihood**,
- $P(\mathrm{PFS} \leq u, \mathrm{OS} \leq v)$ for X non-Markov (generic).

Multistate model formulation

Transition probabilities:

- **Full description** of multistate model by only assuming existence of intensities α_{01}, α_{02} and α_{12}.
- Formulas, even for **non-Markov** case: Aalen et al. (2008).

Meller et al. (2019):

- Embed PFS and OS in multistate model framework,
- formulas for P_{lm}'s assuming **Weibull** transition hazards for time-inhomogeneous Markov and semi-Markov (explicit),
- inference via **counting process likelihood**,
- $P(\mathrm{PFS} \leq u, \mathrm{OS} \leq v)$ for X non-Markov (generic).

Allows derivation of any functional of PFS and OS.

Multistate model formulation

Transition probabilities:

- **Full description** of multistate model by only assuming existence of intensities α_{01}, α_{02} and α_{12}.
- Formulas, even for **non-Markov** case: Aalen et al. (2008).

Meller et al. (2019):

- Embed PFS and OS in multistate model framework,
- formulas for P_{lm}'s assuming **Weibull** transition hazards for time-inhomogeneous Markov and semi-Markov (explicit),
- inference via **counting process likelihood**,
- $P(\mathrm{PFS} \leq u, \mathrm{OS} \leq v)$ for X non-Markov (generic).

Allows derivation of any functional of PFS and OS.

Exemplary application: **Pearson correlation**.

Multistate model for PFS and OS

Marginal distributions:

$$\begin{aligned} S_{PFS}(t) &= P(\mathrm{PFS} > t) = P_{00}(0,t), \\ S_{OS}(t) &= P(\mathrm{OS} > t) = P_{00}(0,t) + P_{01}(0,t), \end{aligned}$$

Multistate model for PFS and OS

Marginal distributions:

$$\begin{aligned} S_{PFS}(t) &= P(\mathrm{PFS} > t) = P_{00}(0,t), \\ S_{OS}(t) &= P(\mathrm{OS} > t) = P_{00}(0,t) + P_{01}(0,t), \end{aligned}$$

Joint distribution:

$$\begin{aligned} P(\mathrm{PFS} \leq u, \mathrm{OS} \leq v) &= P(X(u) \in \{1,2\}, X(v) = 2) \\ &= P(X(v) = 2 | X(u) = 1) \cdot P_{01}(0,u) + P_{02}(0,u). \end{aligned}$$

Multistate model for PFS and OS

Marginal distributions:

$$\begin{aligned} S_{PFS}(t) &= P(\mathrm{PFS} > t) = P_{00}(0,t), \\ S_{OS}(t) &= P(\mathrm{OS} > t) = P_{00}(0,t) + P_{01}(0,t), \end{aligned}$$

Joint distribution:

$$\begin{aligned} P(\mathrm{PFS} \le u, \mathrm{OS} \le v) &= P(X(u) \in \{1,2\}, X(v) = 2) \\ &= P(X(v) = 2 | X(u) = 1) \cdot P_{01}(0,u) + P_{02}(0,u). \end{aligned}$$

X **inhomogeneous Markov**: $P(X(v) = 2 | X(u) = 1) = P_{12}(u,v)$ independent of progression time $t_1 \le u$.

Multistate model for PFS and OS

Marginal distributions:

$$\begin{aligned} S_{PFS}(t) &= P(\mathrm{PFS} > t) = P_{00}(0,t), \\ S_{OS}(t) &= P(\mathrm{OS} > t) = P_{00}(0,t) + P_{01}(0,t), \end{aligned}$$

Joint distribution:

$$\begin{aligned} P(\mathrm{PFS} \le u, \mathrm{OS} \le v) &= P(X(u) \in \{1,2\}, X(v) = 2) \\ &= P(X(v) = 2 | X(u) = 1) \cdot P_{01}(0,u) + P_{02}(0,u). \end{aligned}$$

X **inhomogeneous Markov**: $P(X(v) = 2|X(u) = 1) = P_{12}(u,v)$ independent of progression time $t_1 \le u$.

X **non-Markov**:

- Integrate $P_{12}(u,v;t_1)$ over conditional distribution of all possible progression times $t_1 \le u$.
- Formula tedious (see Meller et al. (2019)) $\Rightarrow$ **simulate** in applications.

Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 4.0.5 (2021-03-31)

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

Other packages: MASS / mstate / prodlim / reporttools / xtable / biostatKR / survival

This document was generated on 2021-04-29 at 10:51:25.