MIRROS: Planning a Phase 3 Trial with Time-to-event Endpoint, a Cure Proportion, and a Futility Interim Analysis using Response

Kaspar Rufibach
Methods, Collaboration & Outreach Group, PD Data Sciences, Roche Basel
JSM 2022, Washington
8th August 2022 (talk given virtually)

Who

Rufibach *et al.* (2020):



Acute Myeloid Leukemia

Acute Myeloid Leukemia

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

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.

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).

Clinical development plan

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**)** $\Rightarrow$ minimize expected sample size.

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)

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.

MIRROS: key questions

Key questions of MIRROS

1. Base **interim** on OS or something else? If the latter, what?
2. How to compute **operating characteristics** of interim analysis?
3. Primary endpoint OS. Sample size with **cure proportion** in both arms?

Futility interim analysis

Futility interim analysis

Mitigate risk if drug does not work (sufficiently).

Planned after **120** patients are recruited.

Why not use OS for interim decision?

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

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

Intermediate endpoint

Complete response:

- Believed to be 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

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

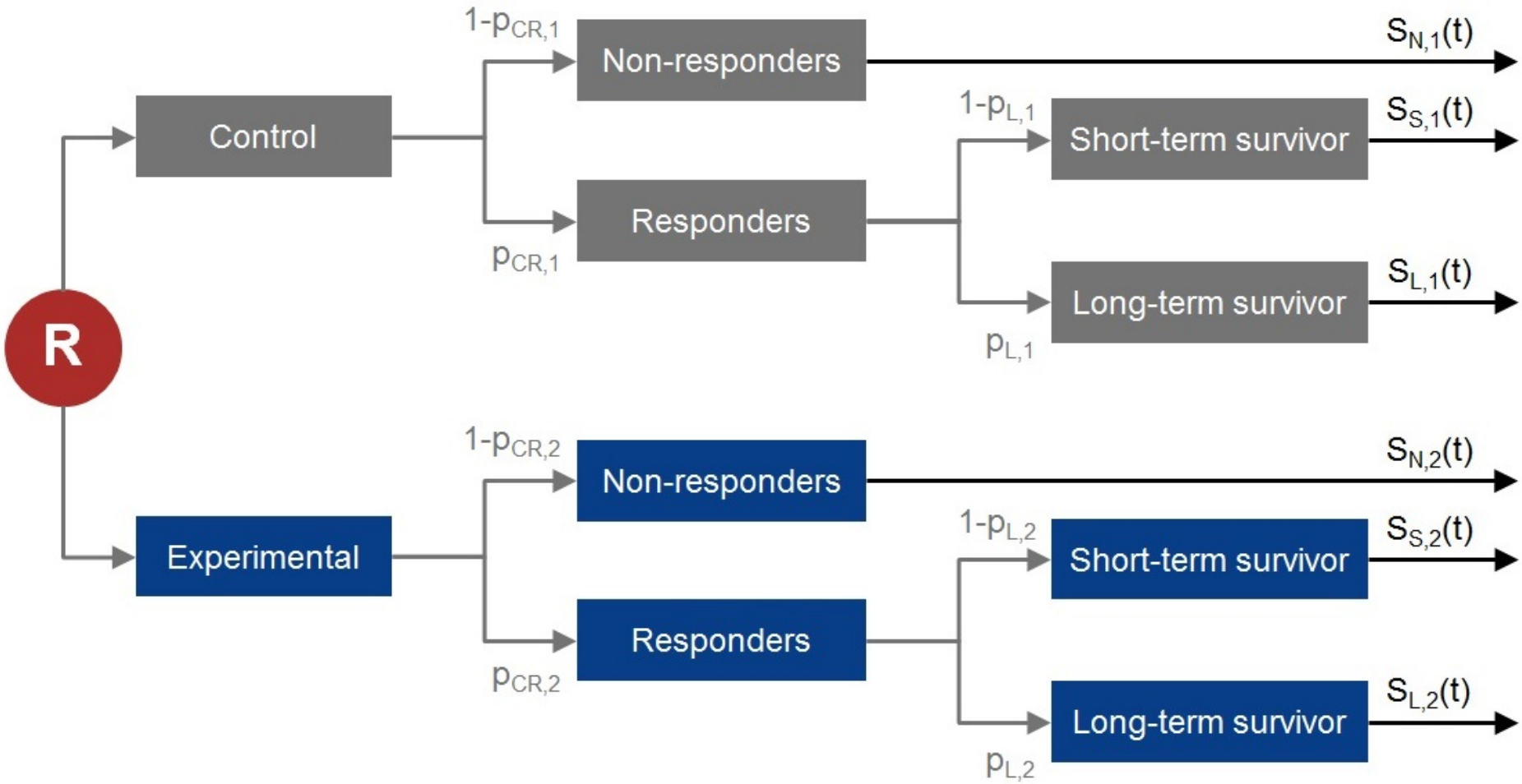
vs.

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

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

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

Mechanistic simulation model



Mechanistic simulation model

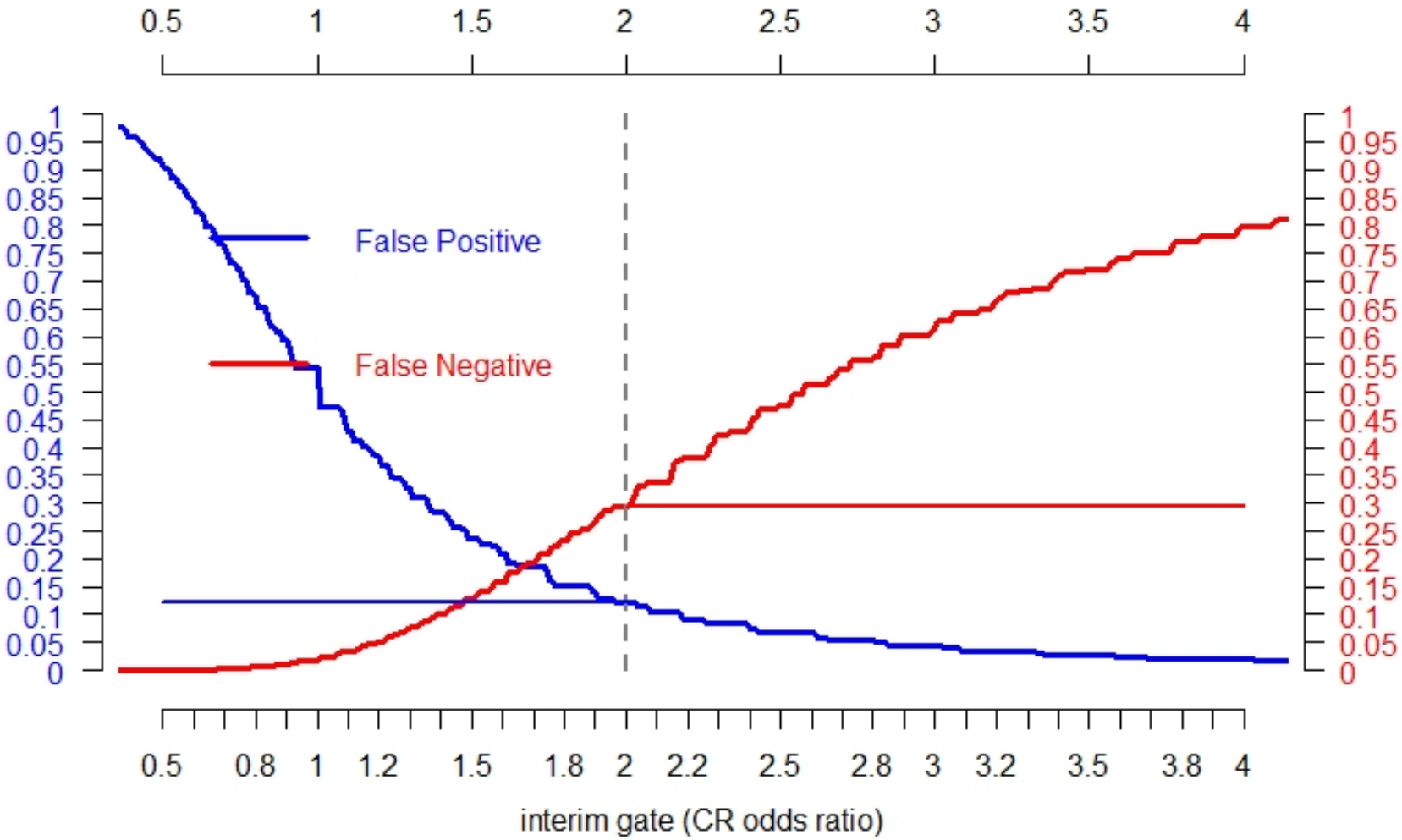
Connects CR to OS.

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

Use "real-world data" (or basic epidemiology) to inform these.

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%.

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".
- **No option to stop for efficacy** $\Rightarrow$ no adaptive design theory for type I error correction needed.

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^*.

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

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

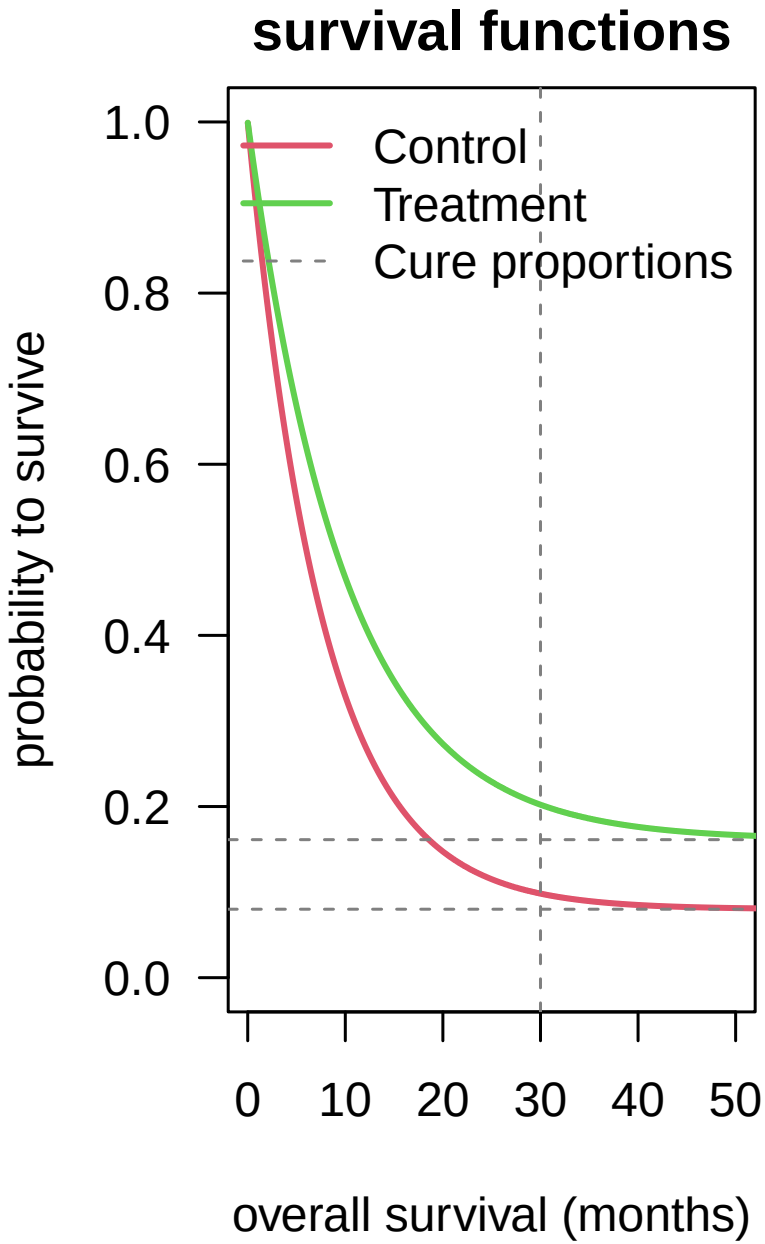
Control arm, based on **historical data**, H_0:

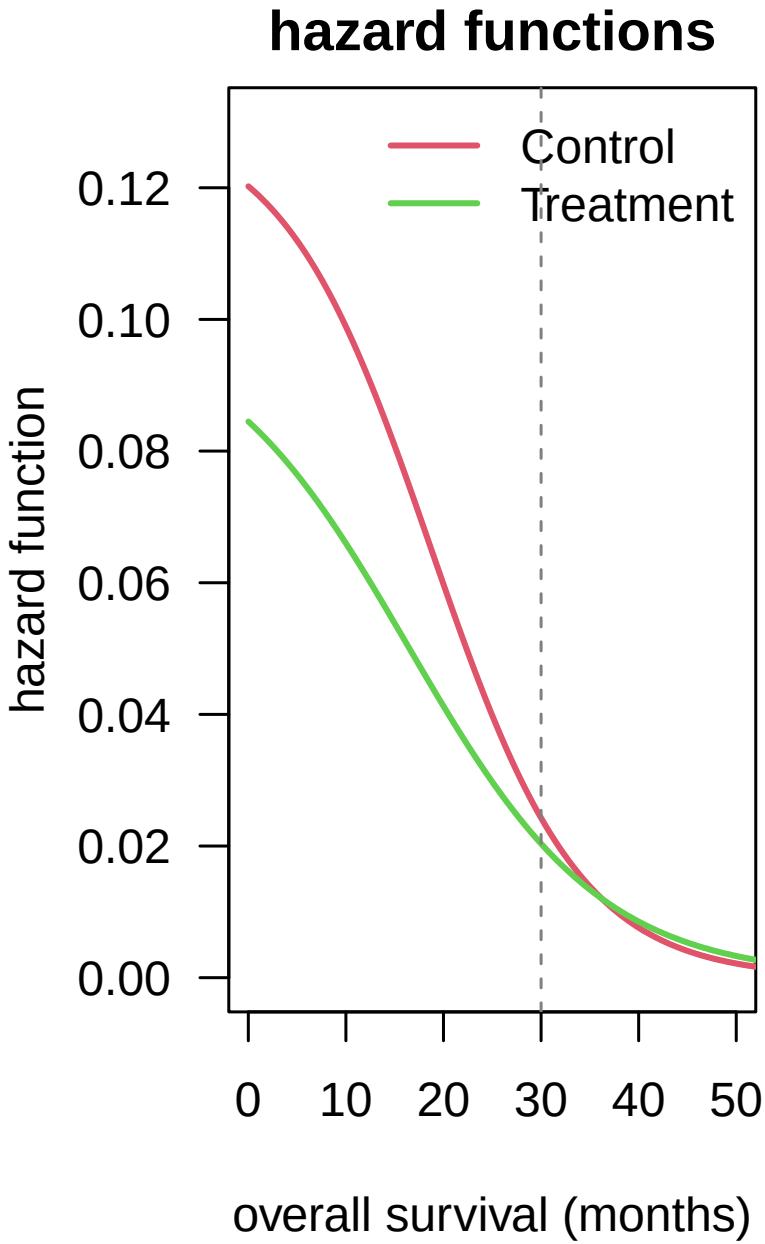
- Median OS 6m.
- Cure: 0.080.

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

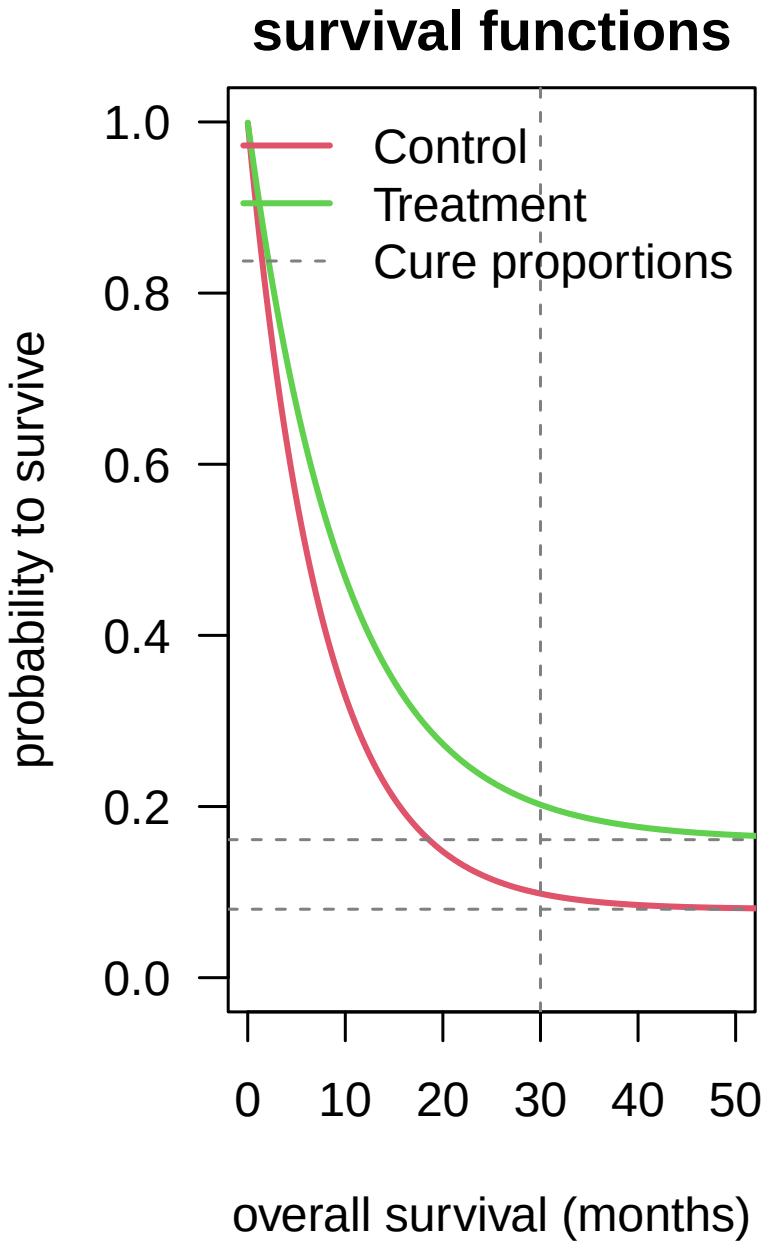
- Median OS 9m.
- Cure: 0.161.

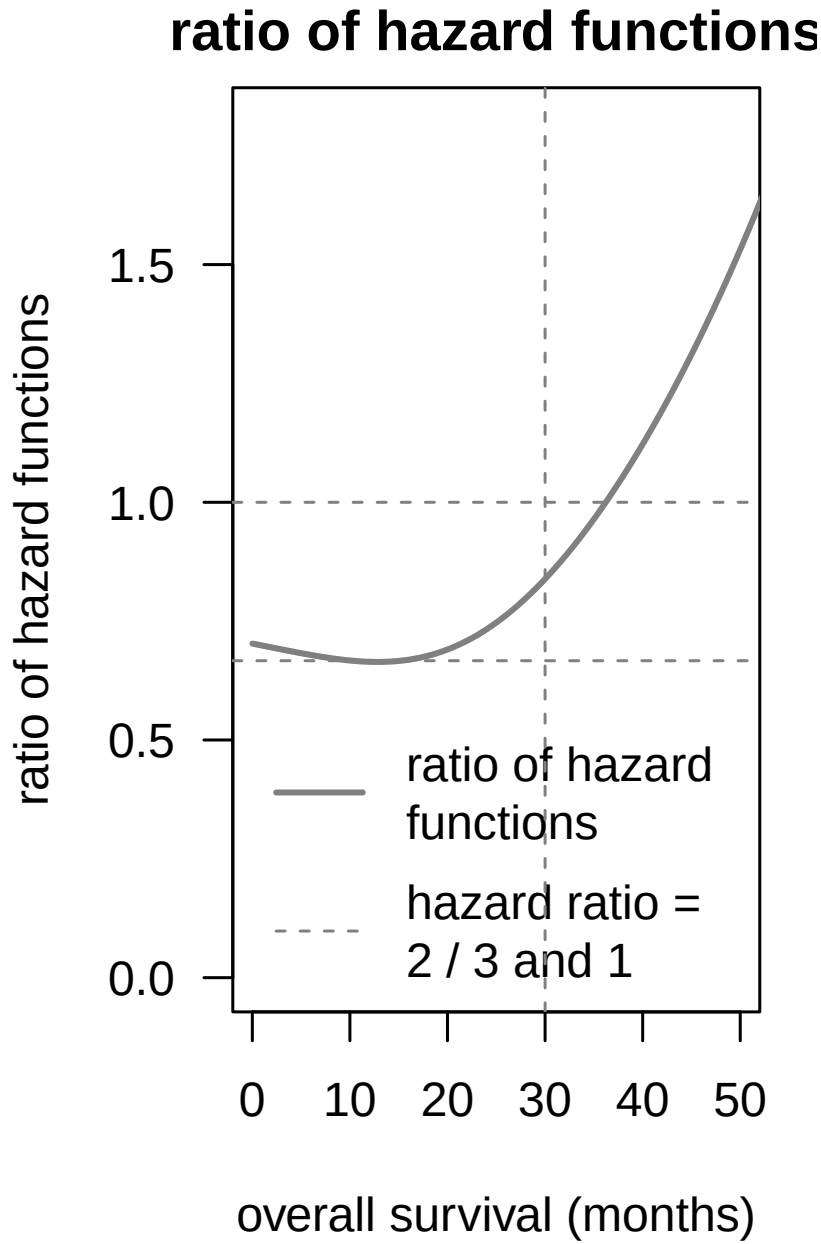
Cure proportion model – assumptions





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
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

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.

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**.

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?

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

Health authority feedback

Health authority feedback to design

FDA:

- **Preferred randomized P2**.
- 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, supported hazard ratio as appropriate effect measure.

Running the trial

Outcome

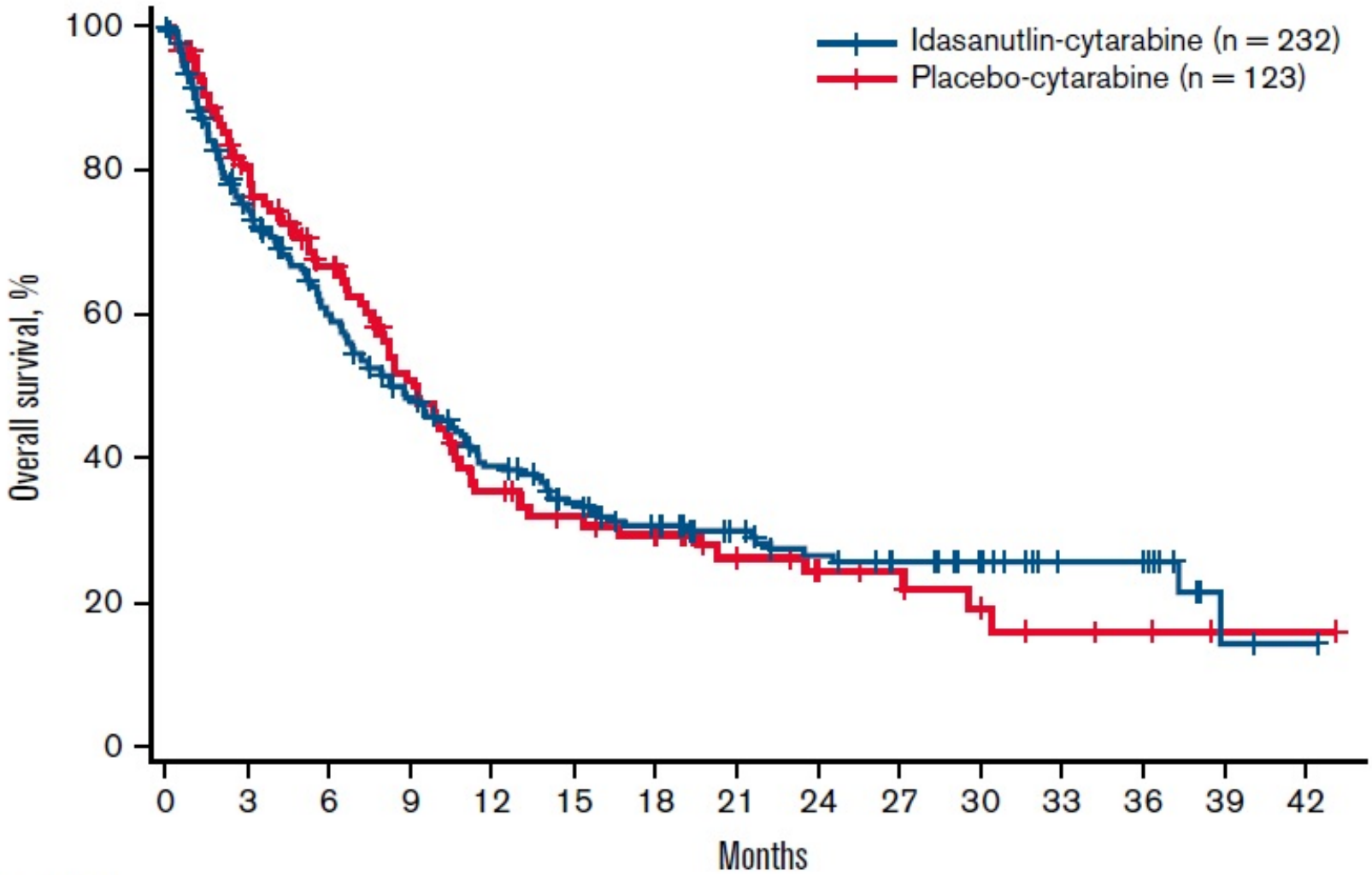
Futility interim was **passed**.

Trial was **negative**.

Assumption on shape of S in treatment arm quite accurate.

Control arm did better than assumed: Relative effect vs. control not big enough.

OS survival function estimates



No. at risk															
Idasanutlin-cytarabine	232	163	123	95	73	56	46	37	30	25	19	11	10	2	1
Placebo-cytarabine	123	88	67	48	33	27	24	16	12	10	6	4	3	1	1

Conclusions

Conclusions futility interim

Skipping / integrating P2 into P3 allows for **acceleration** and **risk-mitigation**:

- If you stop at interim not much is lost.
- If you continue you accelerate potentially dramatically.

Mechanistic simulation model:

- Associate binary intermediate to time-to-event primary endpoint.
- Explore interim analysis **operating characteristics**.

Conclusions effect quantification in cure model

- 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!

Power optimization $\Leftrightarrow$ pragmatism.

Resources

MIRROS trial design:

- Design and statistical methodology: Rufibach *et al.* (2020).
- Reproduce simulations and **plan your own trial**: `https://github.com/numbersman77/integratePhase2.git`.
- Clinical paper describing futility interim: Montesinos *et al.* (2020).
- Clinical paper with final analysis: Konopleva *et al.* (2022).

Received: 4 January 2019 | Revised: 13 June 2019 | Accepted: 11 July 2019

DOI: 10.1002/pst.1969

MAIN PAPER

WILEY

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

Kaspar Rufibach[1] | **Dominik Heinzmann**[2] | **Annabelle Monnet**[2]

Thank you for your attention.

kaspar.rufibach@roche.com
http://go.roche.com/dss-mco

http://www.kasparrufibach.ch
numbersman77
numbersman77

References I

Bose, P., Vachhani, P., and Cortes, J. E. (2017). Treatment of relapsed/refractory acute myeloid leukemia. *Current Treatment Options in Oncology*, **18**(3), 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**(9), 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**(7), 931–937.

Konopleva, M. Y., Röllig, C., Cavenagh, J., Deeren, D., Girshova, L., Krauter, J., Martinelli, G., Montesinos, P., SchÃ€fer, J. A., Ottmann, O., Petrini, M., Pigneux, A., Rambaldi, A., Recher, C., Rodriguez-Veiga, R., Taussig, D., Vey, N., Yoon, S.-S., Ott, M., Muehlbauer, S., Beckermann, B. M., Catalani, O., Genevray, M., Mundt, K., Jamois, C., Fenaux, P., and Wei, A. H. (2022). Idasanutlin plus cytarabine in relapsed or refractory acute myeloid leukemia: results of the MIRROS trial. *Blood Advances*, **6**(14), 4147–4156.

Maller, R. and Zhou, X. (1996). *Survival Analysis with Long-term Survivors*. John Wiley & Sons, United States.

References II

Maller, R. A. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika*, **79**(4), 731–739.

Montesinos, P., Beckermann, B. M., Catalani, O., Esteve, J., Gamel, K., Konopleva, M. Y., Martinelli, G., Monnet, A., Papayannidis, C., Park, A., Récher, C., Rodríguez-Veiga, R., Röllig, C., Vey, N., Wei, A. H., Yoon, S.-S., and Fenaux, P. (2020). Mirros: a randomized, placebo-controlled, phase iii trial of cytarabine ± idasanutlin in relapsed or refractory acute myeloid leukemia. *Future Oncology*, **16**(13), 807–815. PMID: 32167393.

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**(17), 1204–1214.

Rufibach, K. (2019). Treatment effect quantification for time-to-event endpoints – estimands, analysis strategies, and beyond. *Pharmaceutical Statistics*, **18**, 144–164.

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**(1), 44–58. Code available on github: `https://github.com/numbersman77/integratePhase2.git`.

References III

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**(2), 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**(464), 1063–1078.

Backup slides.

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**.

Cure proportion model – estimation

Numerous parametric and nonparametric estimates of relevant quantities: Cantor and Shuster (1992); Maller and Zhou (1992, 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.

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.

Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 4.1.1 (2021-08-10)

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

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

This document was generated on 2022-08-02 at 18:21:14.