# Interim analysis and treatment selection in time-to-event randomized trials in rare diseases on a long-term horizon

Mohamed Amine BAYAR[1,2], Gwénaël LE TEUFF[1,2], Marie-Cécile LE DELEY[1,2,3], Stefan MICHIELS[1,2]

November 9, 2017 - Statistical methods and designs in clinical oncology - Symposium of the ONCOSTAT team

[1]Service de Biostatistique et d'Epidémiologie, Gustave Roussy, Villejuif, France
[2]CESP - UPS - UVSQ, INSERM, Université Paris-Saclay, Villejuif, France
[3]Unité de Méthodologie et Biostatistique, Centre Oscar Lambret, Lille, France

# Context

- Most cancers in children are rare
- $\approx$ 20% of cancers in adults are rare
- Precision medicine $\Rightarrow$ Many common cancers in adults become a set of rare cancers

$\Rightarrow$ Scarce resources for clinical research

Large randomized clinical trials (RCT) with standard one-sided 2.5% $\alpha$-level and 80% power for a reasonable effect size often no longer feasible (Parmar et al., 2016)

# Previous work (Bayar et al., 2016)

- Consider a trial as part of **a series of two-arm RCTs** rather than in isolation
- Assess **benefits** and **risks** on **a long period**
- Search for **the best compromise** between **evidence criteria** and **sample size** to achieve **the greatest therapeutic gain**

**Conclusion:**

Performing a series of small trials with relaxed $\alpha$-levels leads, on average, to larger survival benefits over a long research horizon compared with larger trials with a typical 2.5% one-sided $\alpha$-level

# Previous work (Bayar et al., 2016)

- Consider a trial as part of **a series of two-arm RCTs** rather than in isolation
- Assess **benefits** and **risks** on **a long period**
- Search for **the best compromise** between **evidence criteria** and **sample size** to achieve **the greatest therapeutic gain**

**Conclusion:**

Performing a series of small trials with relaxed $\alpha$-levels leads, on average, to larger survival benefits over a long research horizon compared with larger trials with a typical 2.5% one-sided $\alpha$-level

# Current objective

1. Design each trial within the series as a two-arm RCT with an interim analysis (IA)
2. Design each trial within the series as a two-stage three-arm RCT with treatment selection at interim
3. Compare the performance of the two previous series designs between them, and with other more traditional designs

# Simulation framework

## Basic simulation Model

- **Succession** of $K$ two-arm RCTs over 15 years
- Experimental arm $E$ vs. control arm $C$
- Time-to-event endpoint - One-sided log-rank Test
- **Treatment selected after each trial becomes the control of the next trial**
- Number of patients for each trial within the series computed with the current baseline selected from the previous trial (Kim and Tsiatis, 1990)

# Simulation framework

**Assumptions**

- Uniform accrual
- Exponential distribution of survival times $(\lambda_k^C, \lambda_k^E)$, for each trial $k$, $k \in [1, K]$ and $K$ depends on the course of the series
- No patient lost to follow-up (FU)
- Fixed FU time

# Simulation parameters

## Characteristics of the underlying disease

- Accrual rate: 50, 100, or 200 patients/year
- Hazard rate of the control arm of the first trial of the series $\lambda_1^c$

| Survival | $\lambda_1^c$ | Follow-up |
|---|---|---|
| median survival of 6 months | $2\ log(2)$ | 6 months |
| median survival of 1 year | $log(2)$ | 1 year |
| median survival of 2 years | $\frac{log(2)}{2}$ | 2 years |
| 2-year survival rate of 75% | $\frac{log(4)-log(3)}{2}$ | 2 years |

**Trials within the same series are designed to achieve** the same power (80% or 90%) for the same expected $\mathcal{HR}$ of 0.5, 0.6 or 0.75

# Simulation parameters

## Characteristics of the underlying disease

- Accrual rate: 50, 100, or 200 patients/year
- Hazard rate of the control arm of the first trial of the series $\lambda_1^C$

| Survival | $\lambda_1^C$ | Follow-up |
|---|---|---|
| median survival of 6 months | $2\ log(2)$ | 6 months |
| median survival of 1 year | $log(2)$ | 1 year |
| median survival of 2 years | $\frac{log(2)}{2}$ | 2 years |
| 2-year survival rate of 75% | $\frac{log(4)-log(3)}{2}$ | 2 years |

**Trials within the same series are designed to achieve** the same power (80% or 90%) for the same expected $\mathcal{HR}$ of 0.5, 0.6 or 0.75

# Simulation parameters

**Characteristics of the underlying disease**

- Accrual rate: 50, 100, or 200 patients/year
- Hazard rate of the control arm of the first trial of the series $\lambda_1^C$

| Survival | $\lambda_1^C$ | Follow-up |
|---|---|---|
| median survival of 6 months | $2\ log(2)$ | 6 months |
| median survival of 1 year | $log(2)$ | 1 year |
| median survival of 2 years | $\frac{log(2)}{2}$ | 2 years |
| 2-year survival rate of 75% | $\frac{log(4)-log(3)}{2}$ | 2 years |

**Trials within the same series are designed to achieve** the same power (80% or 90%) for the same expected $\mathcal{HR}$ of 0.5, 0.6 or 0.75

# Simulation parameters

## Characteristics of the underlying disease

- Accrual rate: 50, 100, or 200 patients/year
- Hazard rate of the control arm of the first trial of the series $\lambda_1^C$

| Survival | $\lambda_1^C$ | Follow-up |
|---|---|---|
| median survival of 6 months | $2\ log(2)$ | 6 months |
| median survival of 1 year | $log(2)$ | 1 year |
| median survival of 2 years | $\frac{log(2)}{2}$ | 2 years |
| 2-year survival rate of 75% | $\frac{log(4)-log(3)}{2}$ | 2 years |

## Trials within the same series are designed to achieve

the same power (80% or 90%) for the same expected $\mathcal{HR}$ of 0.5, 0.6 or 0.75

# Simulation parameters

**Hypotheses of how treatments improve over time: future treatment effects**

Relative characterization Hazard Ratio

| $\mathbb{E}[\mathcal{HR}]$ | $\mathbb{P}[\mathcal{HR} \leqslant 0.5]$ |
|---|---|
| 0.925 | 0.02 |
| 0.950 | 0.02 |
| 0.950 | 0.01 |
| 1.000 | 0.01 |

- Historical distribution derived from the meta-analysis of 698 RCTs on $>$ 200 000 patients (Djulbegovic et al., 2012)
- Other distributions $\pm$ optimistic or pessimistic

# Simulation parameters

**Hypotheses of how treatments improve over time: future treatment effects**

Relative characterization Hazard Ratio

| $\mathbb{E}[\mathcal{HR}]$ | $\mathbb{P}[\mathcal{HR} \leqslant 0.5]$ |
|---|---|
| 0.925 | 0.02 |
| 0.950 | 0.02 |
| 0.950 | 0.01 |
| 1.000 | 0.01 |

- Historical distribution derived from the meta-analysis of 698 RCTs on $>$ 200 000 patients (Djulbegovic et al., 2012)
- Other distributions $\pm$ optimistic or pessimistic

# Simulation parameters

## Hypotheses of how treatments improve over time: future treatment effects

Relative characterization

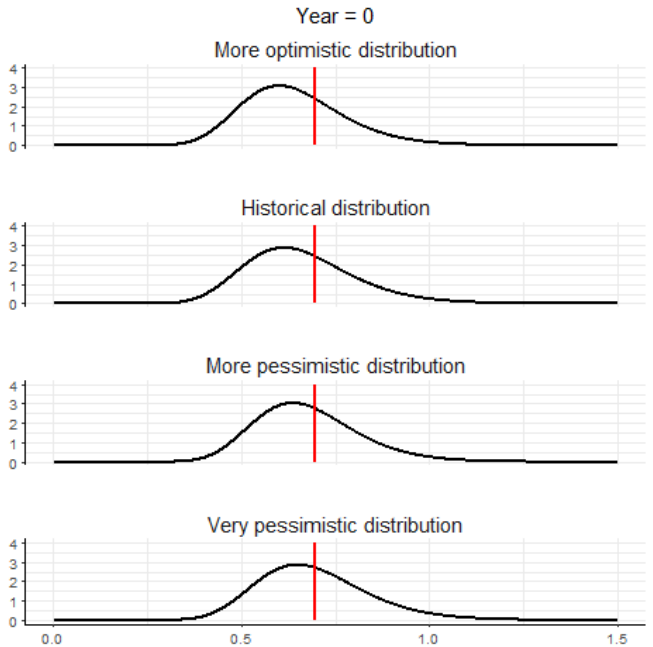
Hazard Ratio

| $\mathbb{E}[\mathcal{HR}]$ | $\mathbb{P}[\mathcal{HR} \leqslant 0.5]$ |
|---|---|
| 0.925 | 0.02 |
| 0.950 | 0.02 |
| 0.950 | 0.01 |
| 1.000 | 0.01 |

Absolute characterization

Hazard Rate $\lambda^E(t)$

Hazard rate of the control arm of the first trial of the series $\lambda_1^C = \log(2)$



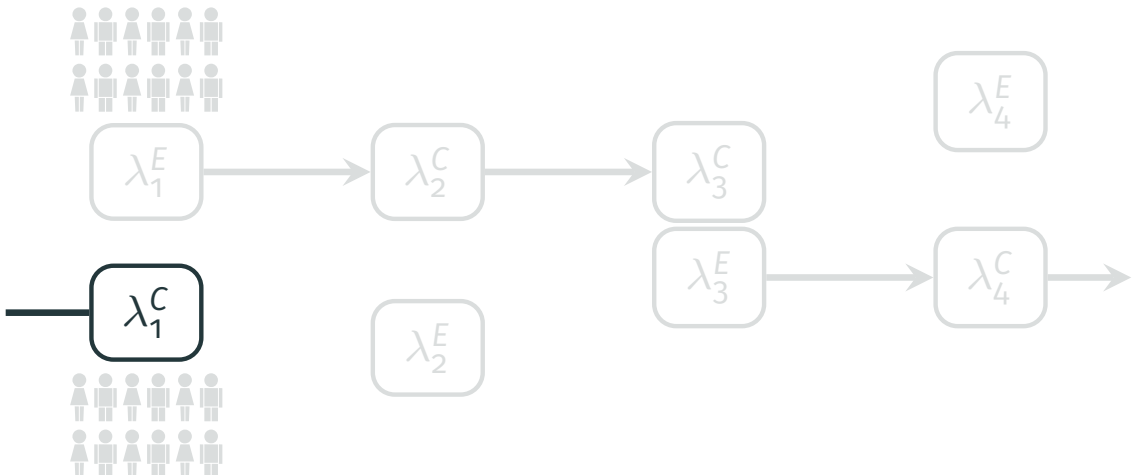
# Simulation parameters

**288 possible combinations of simulation parameters**

- 3 accrual rates
- 4 baseline survivals
- 4 hypotheses of how treatments improve over time
- 2 powers to be achieved for 3 expected $\mathcal{HR}$s

# Illustration of one possible series

Hazard rate of the control arm of the first trial of the series

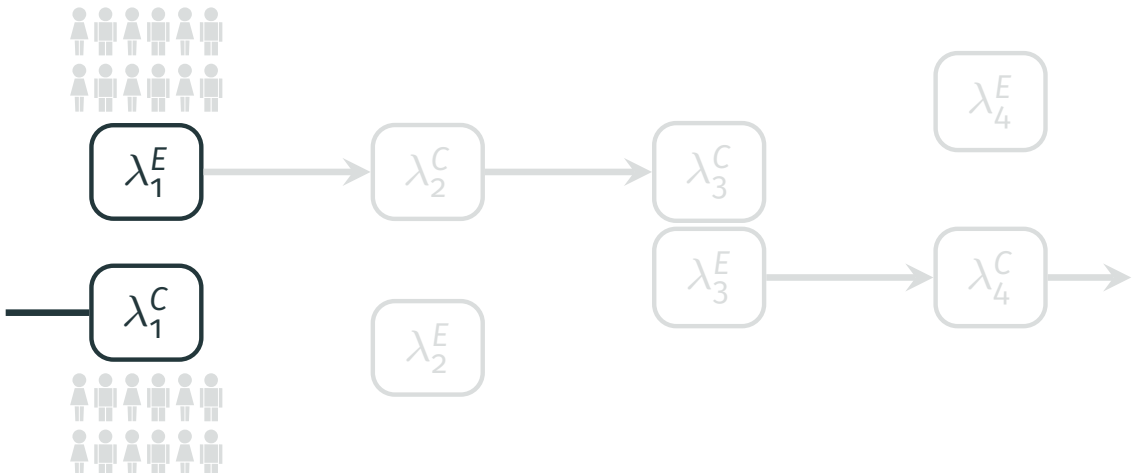



$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Illustration of one possible series

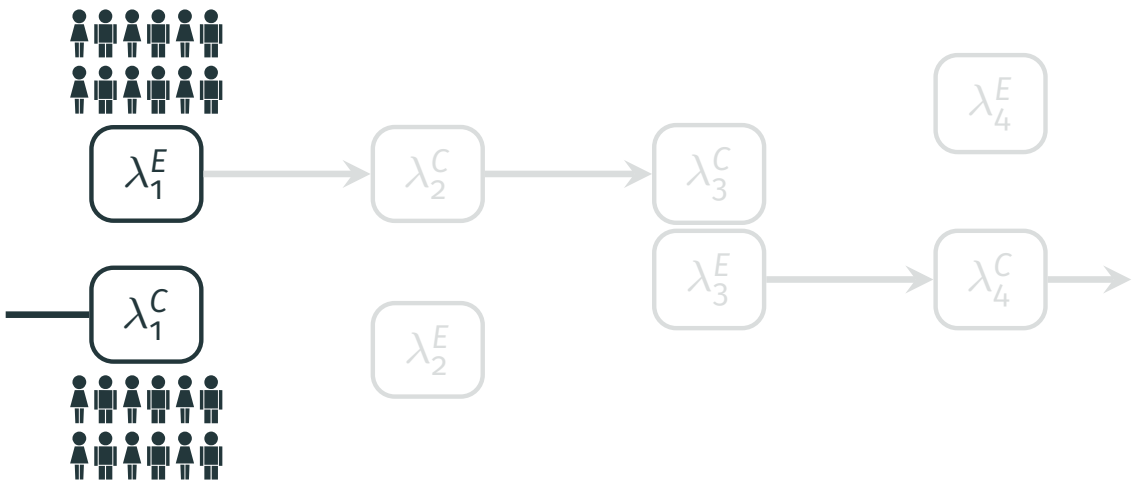
Distribution of future treatment effects




$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

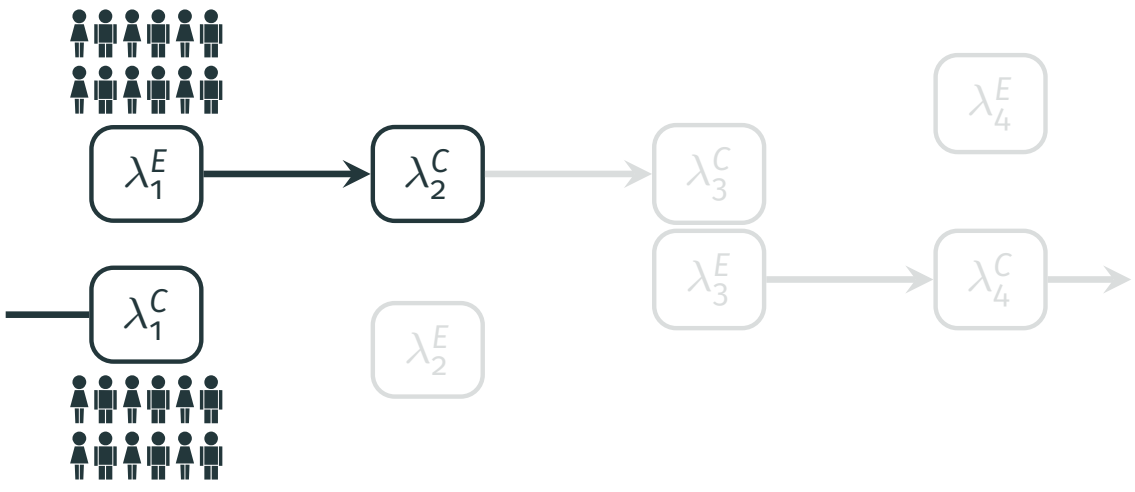
# Illustration of one possible series




$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Illustration of one possible series

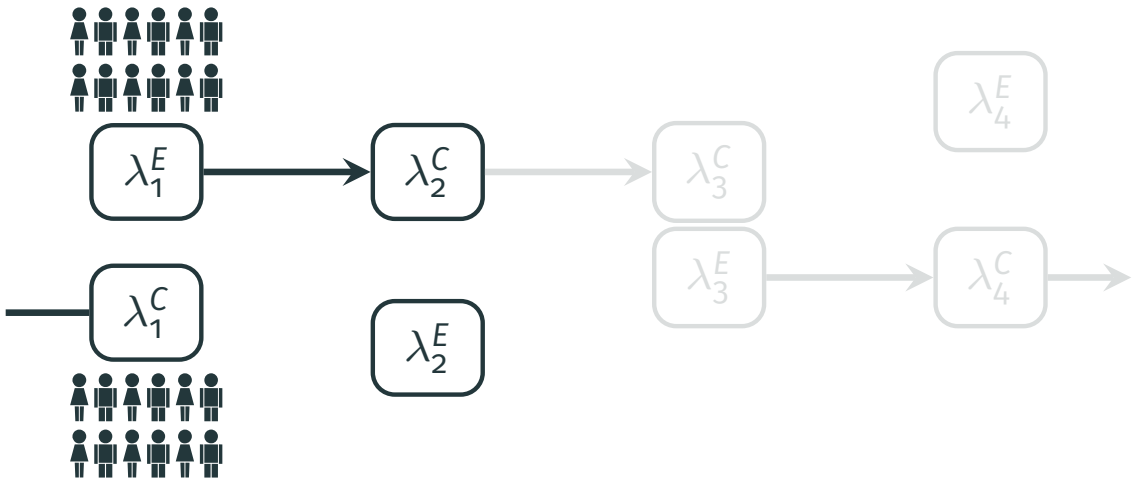



$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Illustration of one possible series

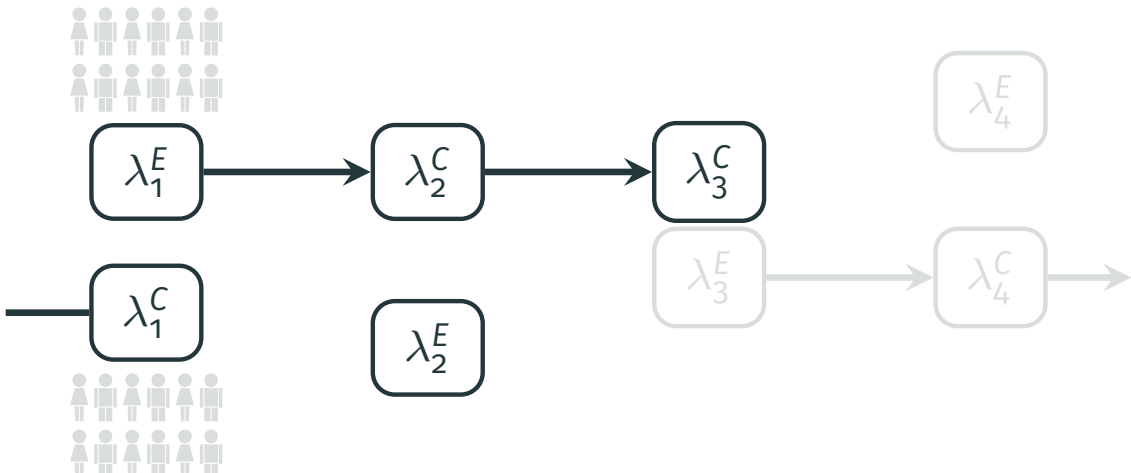
Distribution of future treatment effects




$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Illustration of one possible series

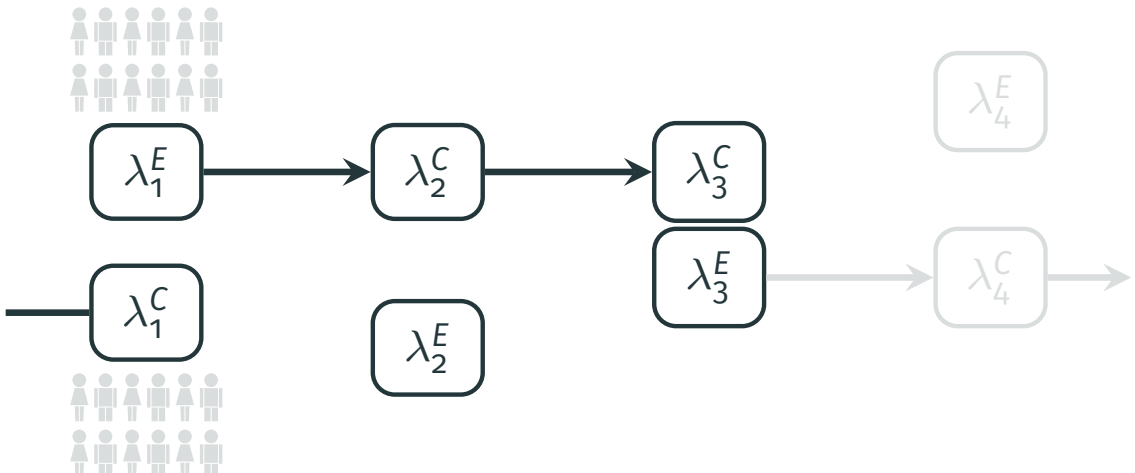



$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Illustration of one possible series

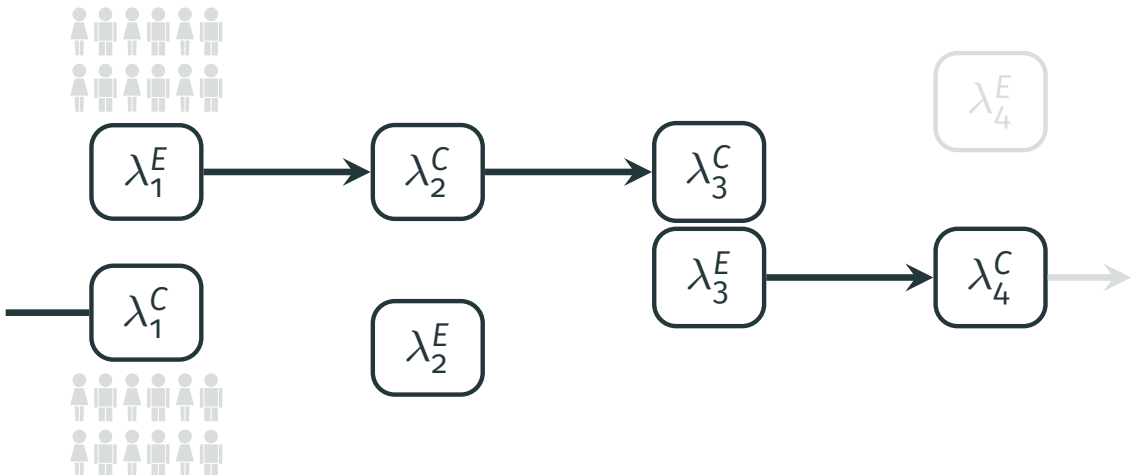
Distribution of future treatment effects




$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Illustration of one possible series

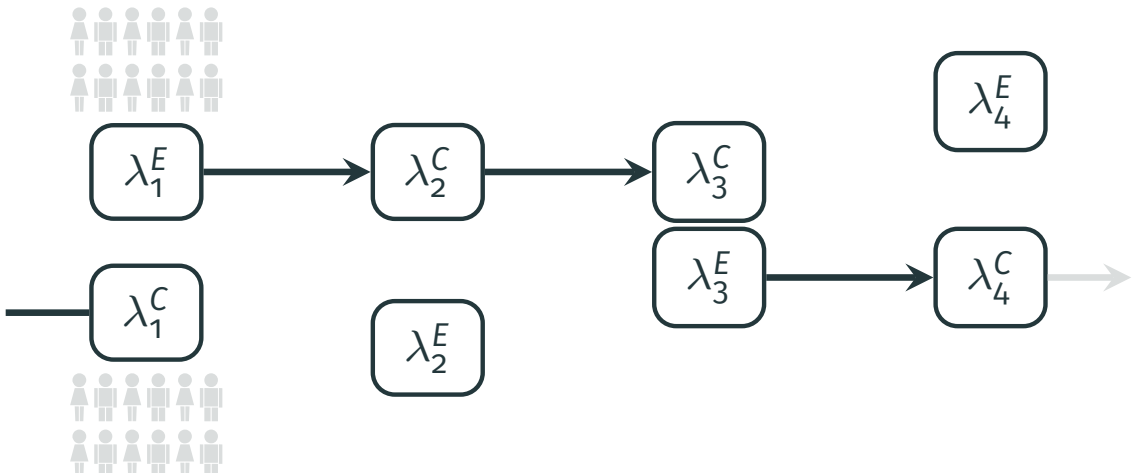



$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Illustration of one possible series

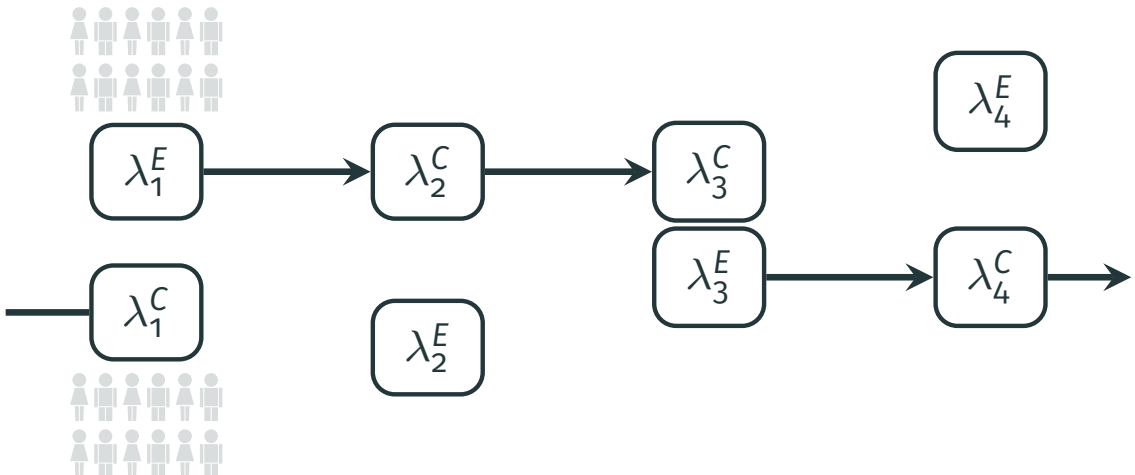
Distribution of future treatment effects




$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Illustration of one possible series




$\mathcal{HR}_{series}$

$$\frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}$$

# Performance metrics

At the end of the 15-year research period

**Total survival benefit**

$$\frac{1}{\mathcal{HR}_{series}} - 1 = \frac{\lambda^{Control}_{First\ trial}}{\lambda^{Selected}_{Last\ trial}} - 1$$

**Example**

For a series of RCTs,

At baseline, median survival = 12 months $\Rightarrow \lambda^{Control}_{First\ trial} = 0.69$

After 15 year, $\lambda^{Selected}_{Last\ trial} = 0.46 \Rightarrow$ median survival = 18 months

$\mathcal{HR}_{series} = 0.67 \Leftrightarrow$ Total survival benefit $= 50\%$

# Performance metrics

10 000 repetitions of the 15-year research period

**Expected total survival benefit (Gain)**

$$\mathbb{E}\left[\frac{\lambda^{Control}_{First\ trial}}{\lambda^{Selected}_{Last\ trial}} - 1\right]$$

**Probability of a detrimental effect (Risk)**

Probability that the event rate associated with the treatment selected at the end of the 15 years is worse than the baseline event rate

$$\mathbb{P}\left[\lambda^{Selected}_{Last\ trial} > \lambda^{Control}_{First\ trial}\right]$$

# Series of two-arm RCTs with an interim analysis (IA)

IA performed when $^1/_2$ of the required events are expected to be attained

- Wieand stopping rule for futility (Wieand et al., 1994)
  $\widehat{\mathcal{HR}} \geqslant 1 \Rightarrow$ stop the trial for futility
- OBF $\beta-$spending stopping rule for futility (O'Brien and Fleming, 1979)
- OBF $\alpha-$spending stopping rule for efficacy
- Combining the latter two

# Series of two-stage three-arm RCTs with treatment selection at interim (Posch et al., 2005)

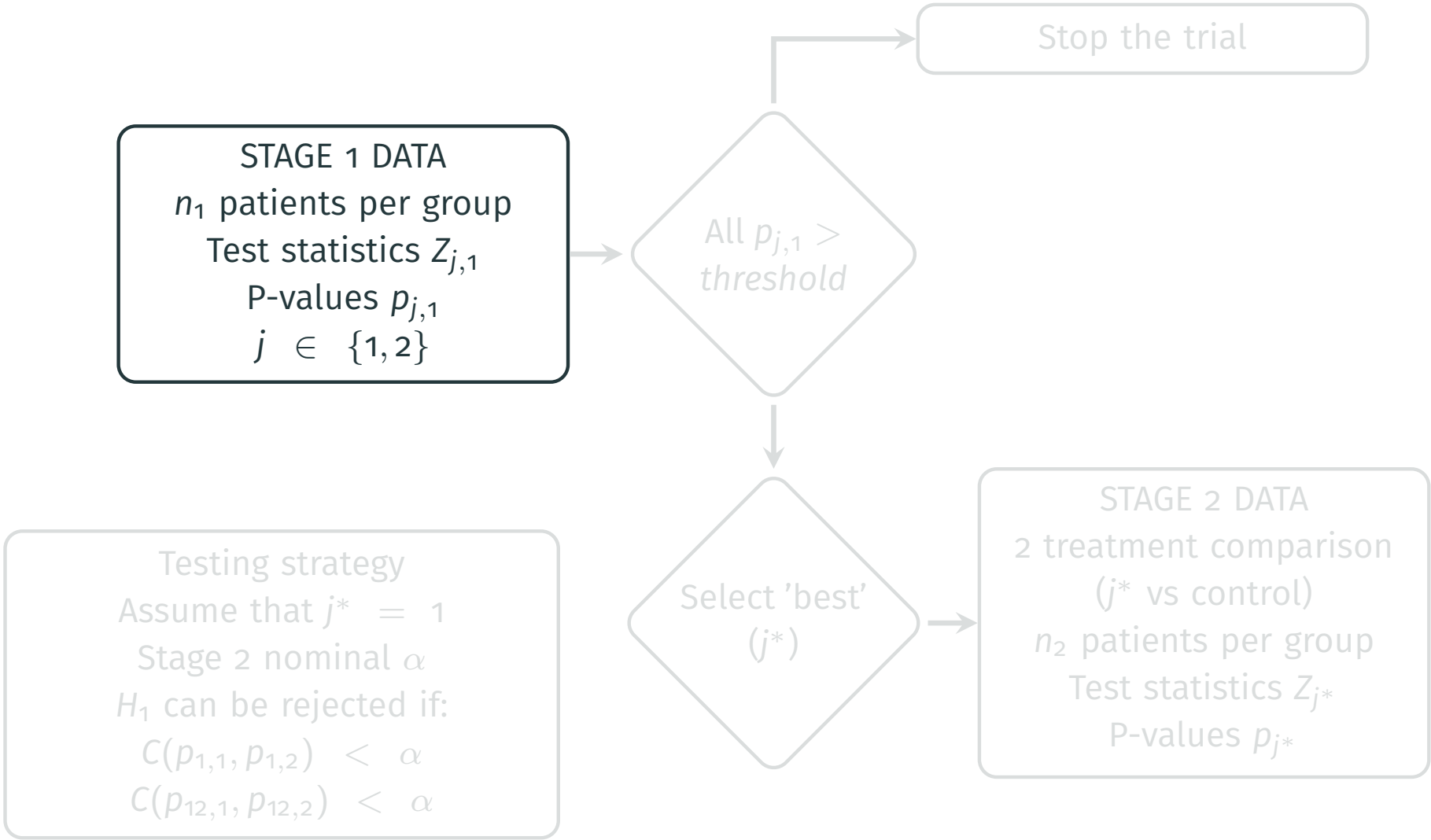
**At the first stage**

2 experimental treatments $J_1 = \{1, 2\}$ are compared to the control and the best is selected for the second stage

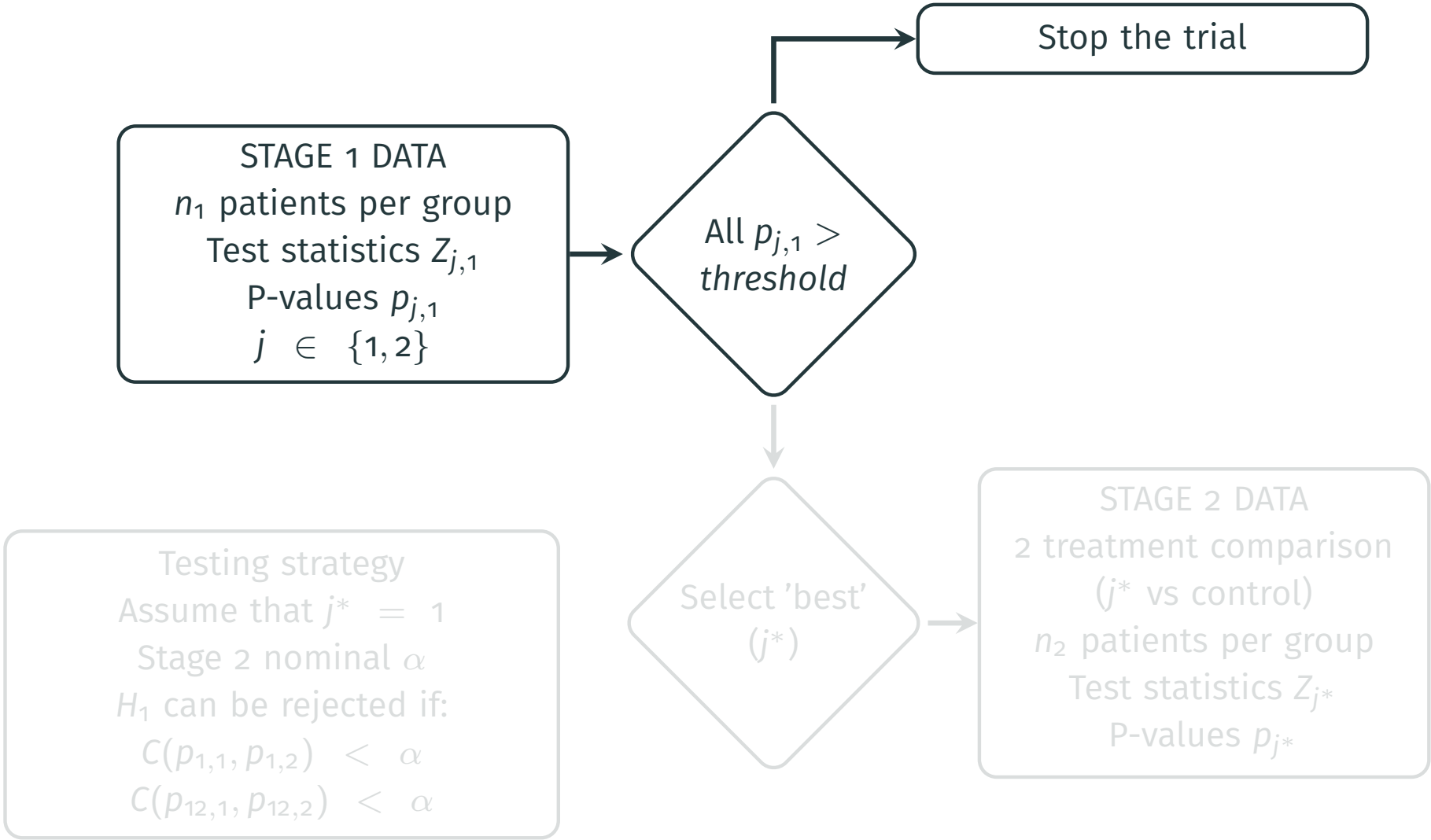
**At the second stage**

Selected treatment compared to the control, combining data from both stages at the multiple level $\alpha$

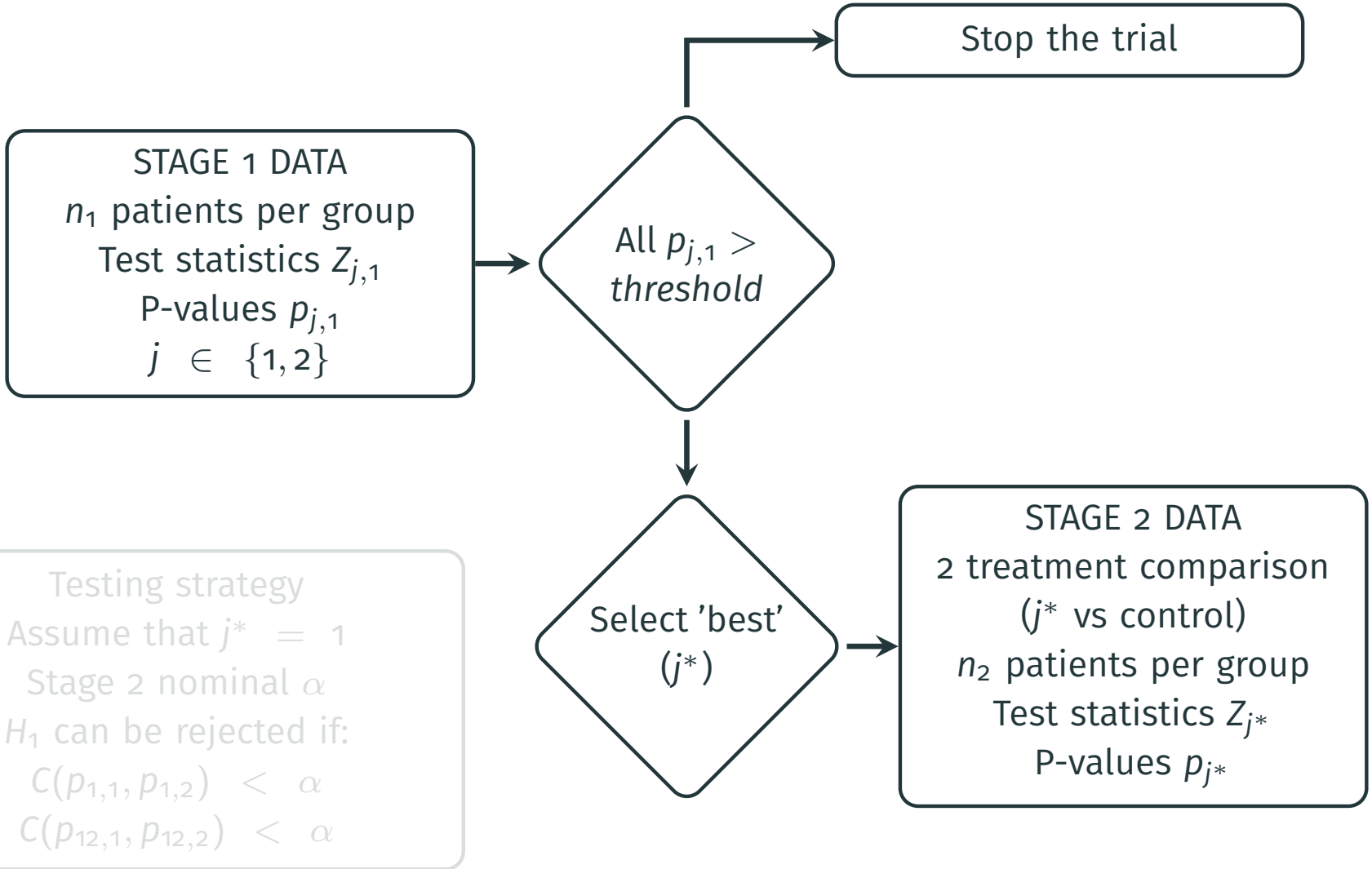
- Closed testing procedure for multiple testing
- Simes test for intersection hypotheses
- Weighted inverse normal combination function for stagewise p-values combination



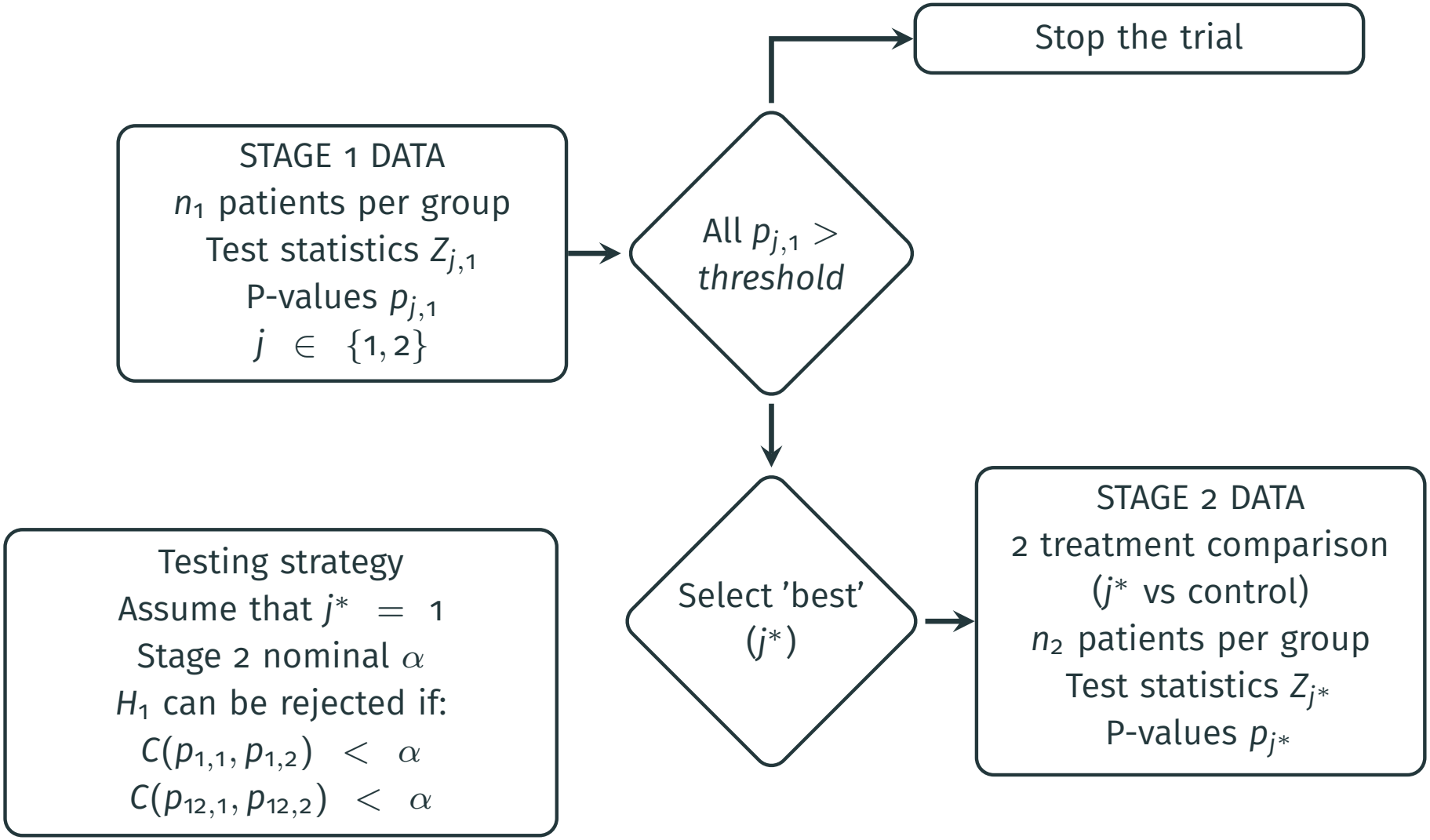

Decision making and testing strategy for the two-stage adaptive treatment selection design, adapted from Dmitrienko et al. (2009)




Decision making and testing strategy for the two-stage adaptive treatment selection design, adapted from Dmitrienko et al. (2009)




Decision making and testing strategy for the two-stage adaptive treatment selection design, adapted from Dmitrienko et al. (2009)




Decision making and testing strategy for the two-stage adaptive treatment selection design, adapted from Dmitrienko et al. (2009)

# Design parameters

## Series of 2-arm trials with interim analysis

Interim analysis

- No interim analysis
- Wieand stopping rule for futility
- OBF $\beta-$spending stopping rule for futility
- OBF $\alpha-$spending stopping rule for efficacy
- Combining the latter two

One-sided $\alpha$-level:

0.025, 0.05, 0.1, 0.2

## Series of 3-arm trials with selection at interim

First stage threshold:

0.05, 0.1, 0.15, 0.2

Second stage $\alpha$-level:

0.025, 0.05, 0.1, 0.2

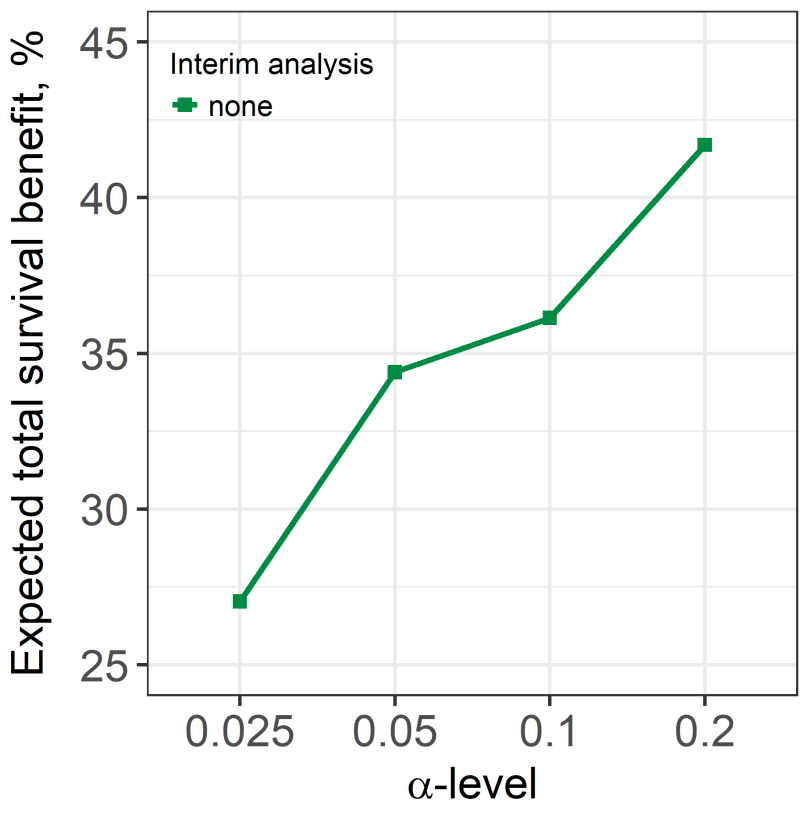
# Results

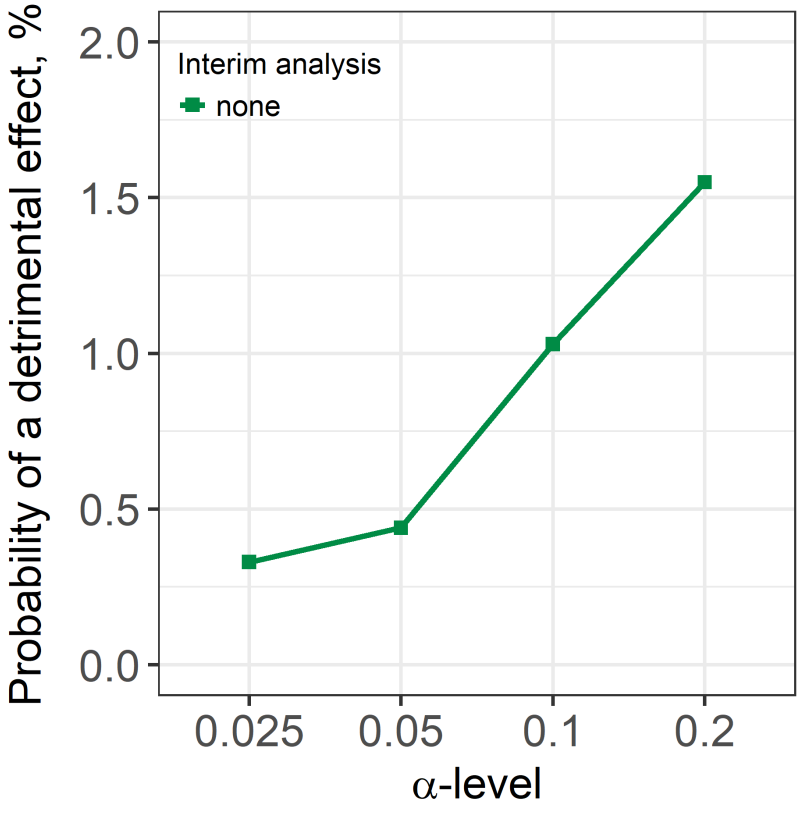
**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

**90% power for an expected $\mathcal{HR}$ of 0.6**



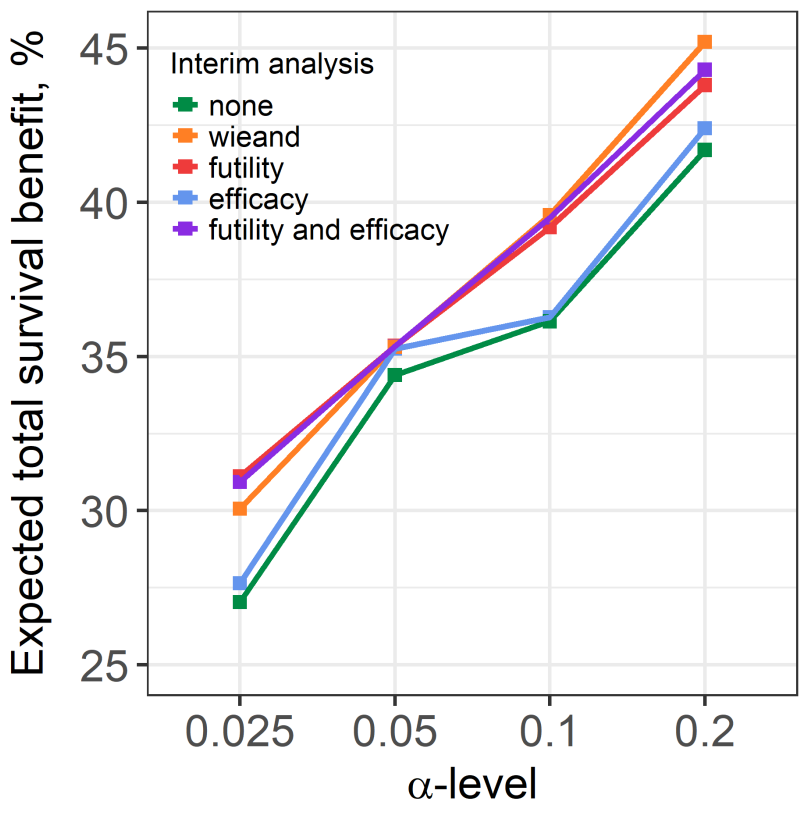
# Results - Inclusion of an interim analysis

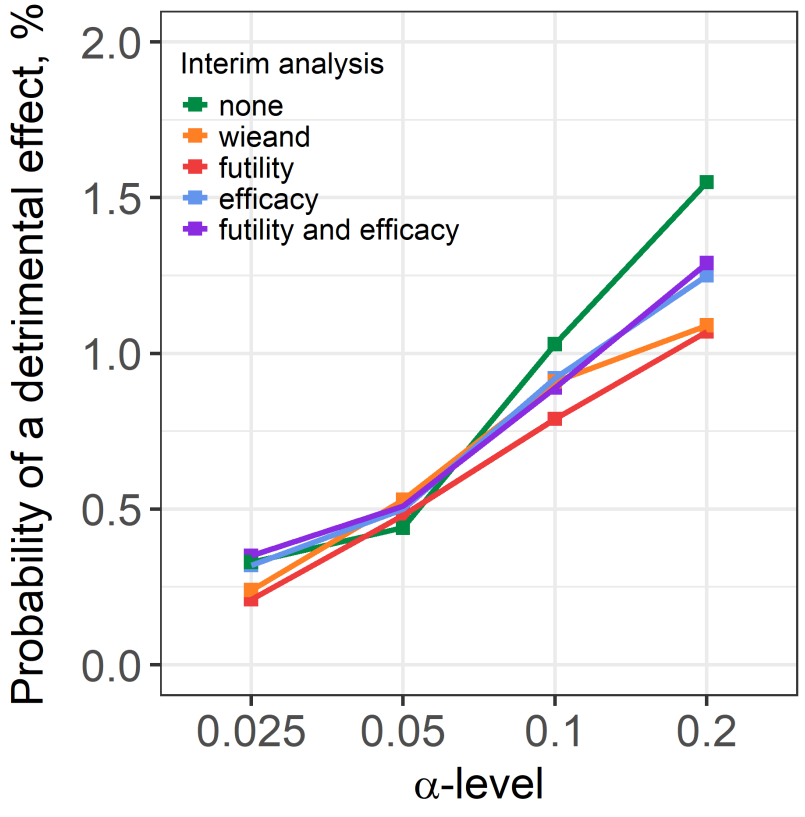
**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

**90% power for an expected $\mathcal{HR}$ of 0.6**



# Results - Inclusion of an interim analysis

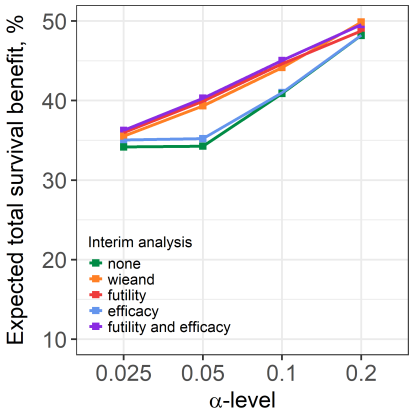
**Accrual rate = 50 patients/year**

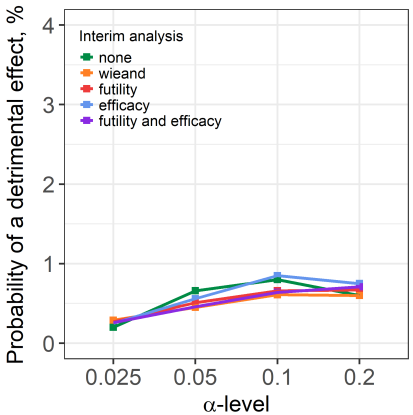
**Different baseline hazards rates**

**Historical distribution of treatment effects**

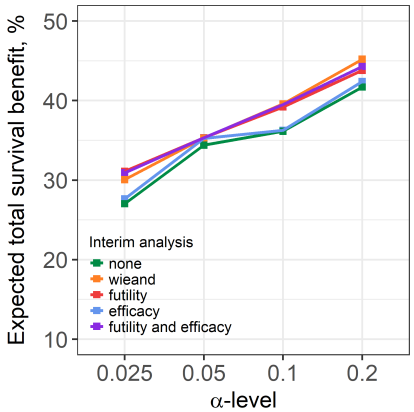
**90% power for an expected $\mathcal{HR}$ of 0.6**

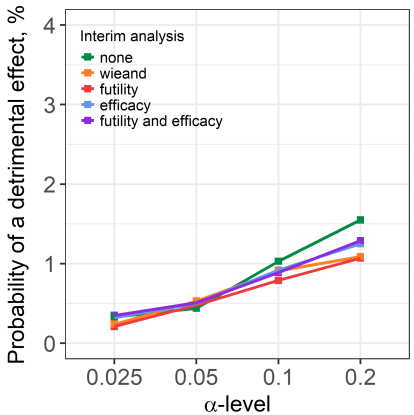
**Median survival of 6 months**



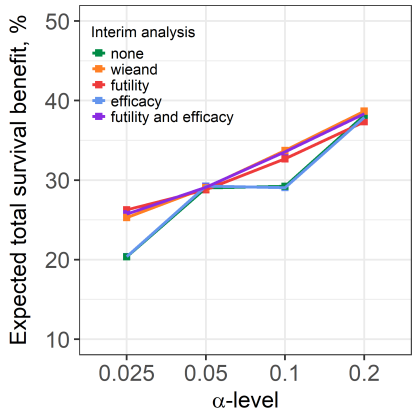



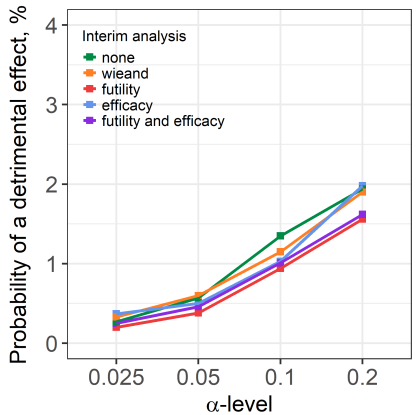

**Median survival of 1 year**



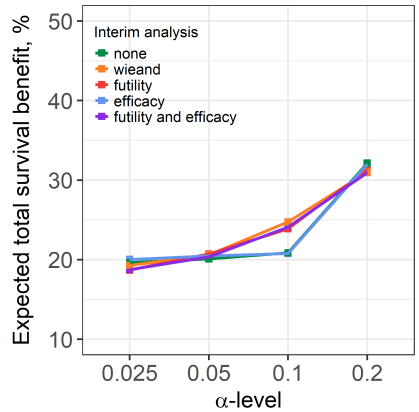



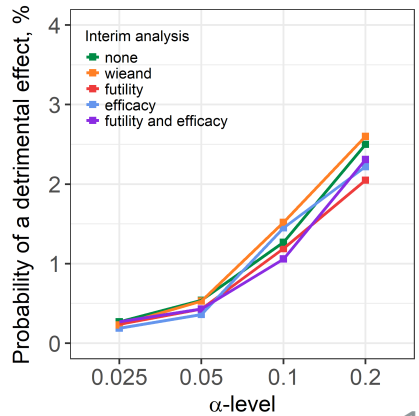

**Median survival of 2 years**




**2-year survival rate of 75%**



# Results - Inclusion of an interim analysis

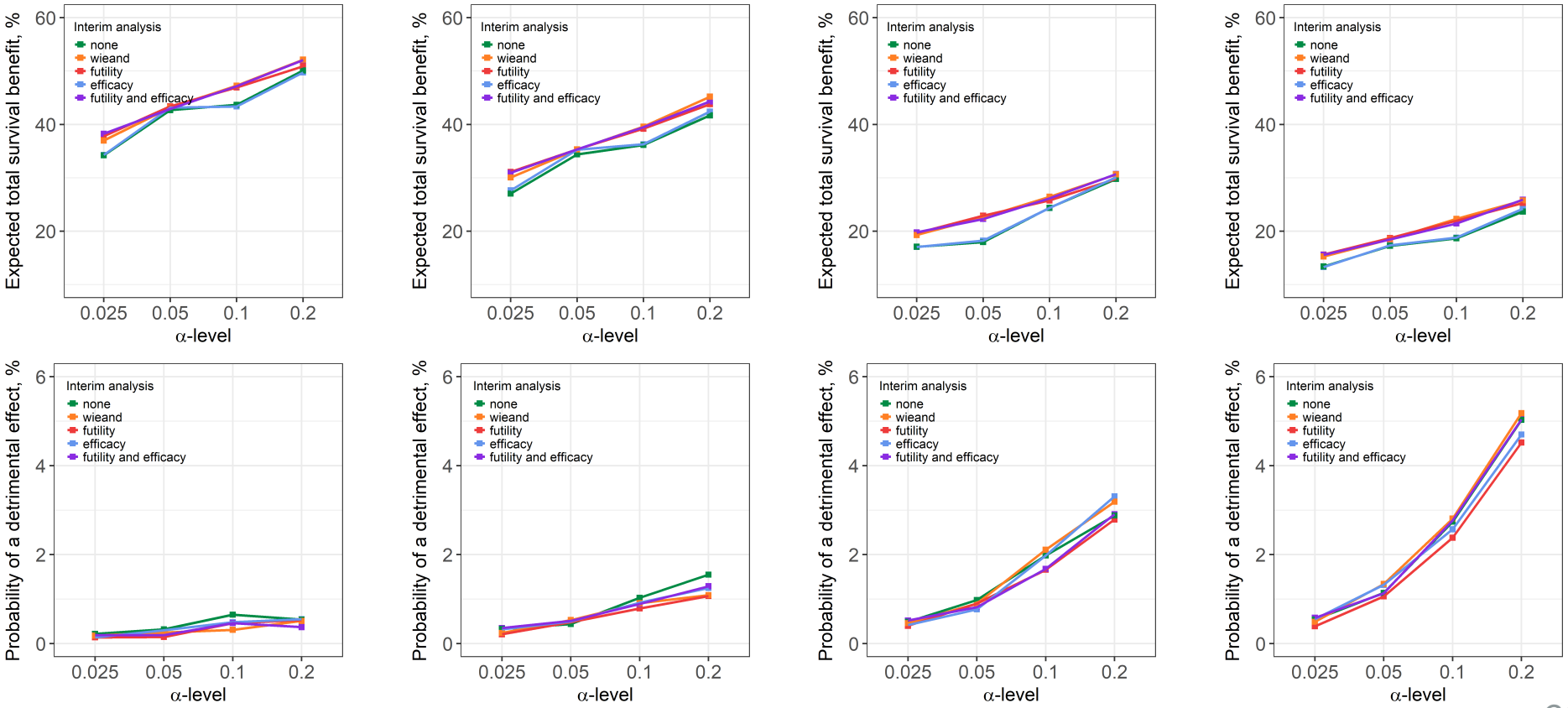
**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Different distributions of treatment effects**

**90% power for an expected $\mathcal{HR}$ of 0.6**

**More optimistic** | **Historical** | **More pessimistic** | **Very pessimistic**



# Results - Designs comparison

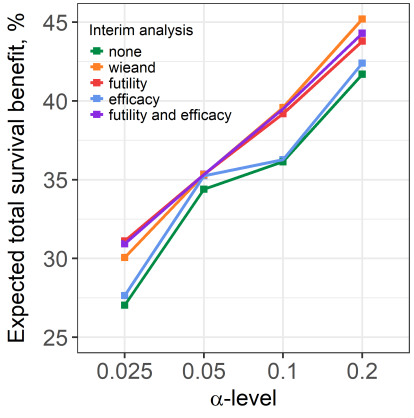
**Accrual rate = 50 patients/year**

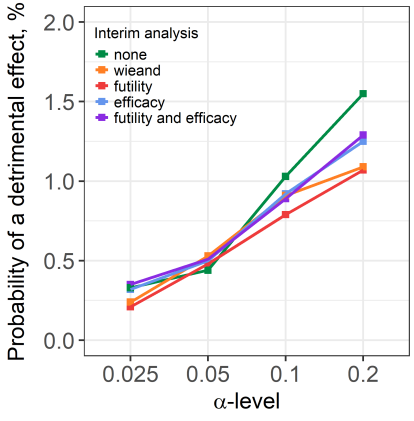
**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

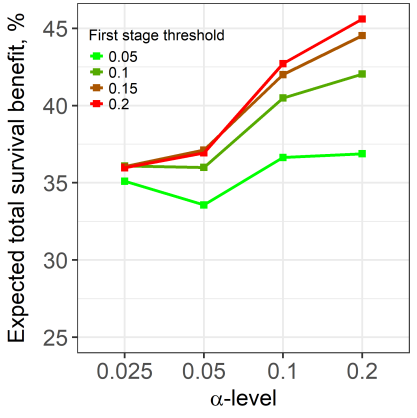
**90% power for an expected $\mathcal{HR}$ of 0.6**

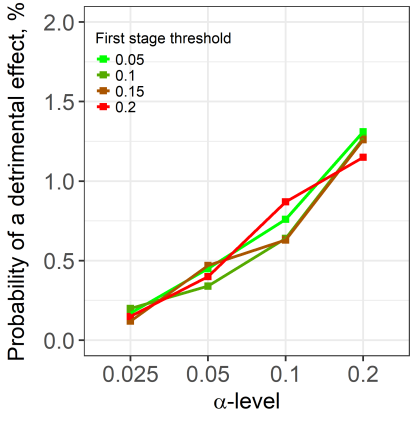
**Series of 2-arm RCTs with interim analysis**




**Series of 3-arm RCTs with selection at interim**



# Results - Optimal designs

$$\text{Argmax } \mathbb{E}\left[\frac{\lambda^{Control}_{First\ trial}}{\lambda^{Selected}_{Last\ trial}} - 1\right]$$

$$\text{subject to } \mathbb{P}\left[\lambda^{Selected}_{Last\ trial} > \lambda^{Control}_{First\ trial}\right] < 1.0\%$$

**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

**90% power for an expected $\mathcal{HR}$ of 0.6**

| Optimal design | Traditional design | Series of 2-arm RCTs with no interim analysis | Series of 2-arm RCTs with interim analysis | Series of 3-arm RCTs with selection at interim |
|---|---|---|---|---|
| $\alpha$-level | 0.025 | 0.1 | 0.1 | 0.1 |
| Interim analysis | None | None | Wieand | threshold = 0.2 |
| Number of trial | 3.0 | 4.0 | 4.7 | 4.3 |
| Gain | 27.0% | 36.1% | 39.6% | 42.7% |
| Risk | 0.33% | 1.03% | 0.91% | 0.87% |

# Results - Optimal designs

$$\text{Argmax } \mathbb{E}\left[\frac{\lambda^{Control}_{First\ trial}}{\lambda^{Selected}_{Last\ trial}} - 1\right]$$

$$\text{subject to } \mathbb{P}\left[\lambda^{Selected}_{Last\ trial} > \lambda^{Control}_{First\ trial}\right] < 1.0\%$$

**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

**90% power for an expected $\mathcal{HR}$ of 0.6**

| Optimal design | Traditional design | Series of 2-arm RCTs with no interim analysis | Series of 2-arm RCTs with interim analysis | Series of 3-arm RCTs with selection at interim |
|---|---|---|---|---|
| $\alpha$-level | 0.025 | 0.1 | 0.1 | 0.1 |
| Interim analysis | None | None | Wieand | threshold = 0.2 |
| Number of trial | 3.0 | 4.0 | 4.7 | 4.3 |
| Gain | 27.0% | 36.1% | 39.6% | 42.7% |
| Risk | 0.33% | 1.03% | 0.91% | 0.87% |

# Results - Optimal designs

$$\text{Argmax } \mathbb{E}\left[\frac{\lambda^{Control}_{First\ trial}}{\lambda^{Selected}_{Last\ trial}} - 1\right]$$

$$\text{subject to } \mathbb{P}\left[\lambda^{Selected}_{Last\ trial} > \lambda^{Control}_{First\ trial}\right] < 1.0\%$$

**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

**90% power for an expected $\mathcal{HR}$ of 0.6**

| Optimal design | Traditional design | Series of 2-arm RCTs with no interim analysis | Series of 2-arm RCTs with interim analysis | Series of 3-arm RCTs with selection at interim |
|---|---|---|---|---|
| $\alpha$-level | 0.025 | 0.1 | 0.1 | 0.1 |
| Interim analysis | None | None | Wieand | threshold = 0.2 |
| Number of trial | 3.0 | 4.0 | 4.7 | 4.3 |
| Gain | 27.0% | 36.1% | 39.6% | 42.7% |
| Risk | 0.33% | 1.03% | 0.91% | 0.87% |

# Results - Optimal designs

$$\text{Argmax } \mathbb{E}\left[\frac{\lambda^{Control}_{First\ trial}}{\lambda^{Selected}_{Last\ trial}} - 1\right]$$

$$\text{subject to } \mathbb{P}\left[\lambda^{Selected}_{Last\ trial} > \lambda^{Control}_{First\ trial}\right] < 1.0\%$$

**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

**90% power for an expected $\mathcal{HR}$ of 0.6**

| Optimal design | Traditional design | Series of 2-arm RCTs with no interim analysis | Series of 2-arm RCTs with interim analysis | Series of 3-arm RCTs with selection at interim |
|---|---|---|---|---|
| $\alpha$-level | 0.025 | 0.1 | 0.1 | 0.1 |
| Interim analysis | None | None | Wieand | threshold = 0.2 |
| Number of trial | 3.0 | 4.0 | 4.7 | 4.3 |
| Gain | 27.0% | 36.1% | 39.6% | 42.7% |
| Risk | 0.33% | 1.03% | 0.91% | 0.87% |

# Results - Optimal designs

$$\text{Argmax } \mathbb{E}\left[\frac{\lambda^{Control}_{First\ trial}}{\lambda^{Selected}_{Last\ trial}} - 1\right]$$

$$\text{subject to } \mathbb{P}\left[\lambda^{Selected}_{Last\ trial} > \lambda^{Control}_{First\ trial}\right] < 1.0\%$$

**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

**90% power for an expected $\mathcal{HR}$ of 0.6**

| Optimal design | Traditional design | Series of 2-arm RCTs with no interim analysis | Series of 2-arm RCTs with interim analysis | Series of 3-arm RCTs with selection at interim |
|---|---|---|---|---|
| $\alpha$-level | 0.025 | 0.1 | 0.1 | 0.1 |
| Interim analysis | None | None | Wieand | threshold = 0.2 |
| Number of trial | 3.0 | 4.0 | 4.7 | 4.3 |
| Gain | 27.0% | 36.1% | 39.6% | 42.7% |
| Risk | 0.33% | 1.03% | 0.91% | 0.87% |

# Results - Optimal designs

For the 288 possible combinations of simulation parameters

## Series of 2-arm RCTs with interim analysis

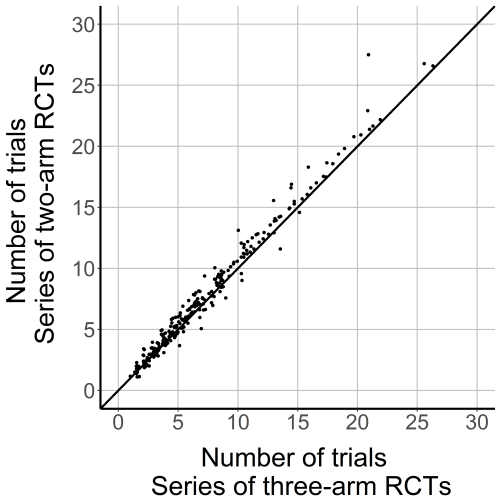
| | One-sided $\alpha$-level | | | | |
|---|---|---|---|---|---|
| Interim Analysis | 0.025 | 0.05 | 0.1 | 0.2 | |
| No interim analysis | **0.0%** | 0.3% | 1.4% | 1.0% | 2.8% |
| Wieand stopping rule for futility | 5.6% | 4.5% | 4.5% | **25.8%** | **40.4%** |
| OBF $\beta$—spending stopping rule for futility | 5.2% | 5.9% | 2.1% | 2.4% | 15.7% |
| OBF $\alpha$—spending stopping rule for efficacy | 0.3% | 2.1% | 0.7% | 1.0% | 4.2% |
| Combining the latter two | 5.2% | 11.5% | 4.5% | 15.7% | **36.9%** |
| | 16.4% | **24.4%** | 13.2% | **46.0%** | Total |

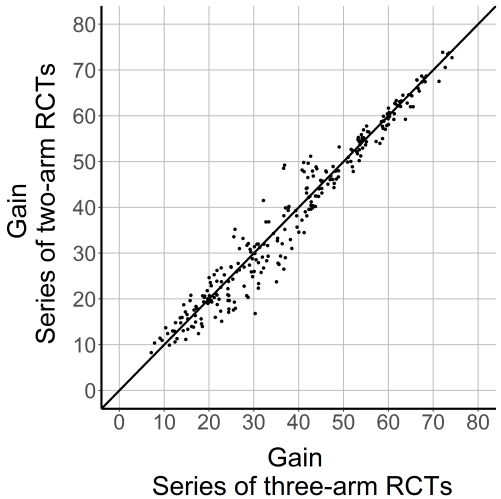
## Series of 3-arm RCTs with selection at interim

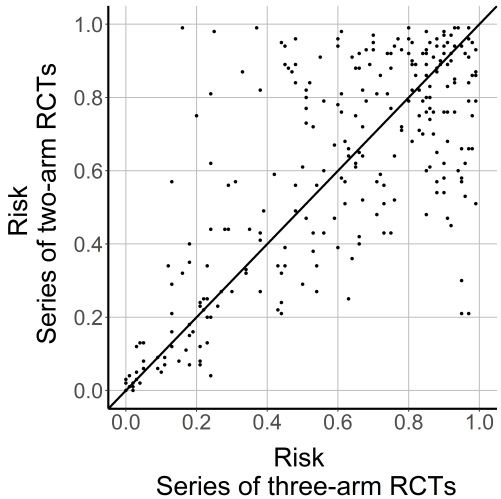
| | Second stage $\alpha$-level | | | | |
|---|---|---|---|---|---|
| First stage threshold | 0.025 | 0.05 | 0.1 | 0.2 | |
| 0.05 | 0.4% | 0.0% | 0.7% | 0.7% | 1.8% |
| 0.1 | 3.7% | 0.7% | 1.1% | 0.4% | 5.9% |
| 0.15 | **10.7%** | 8.1% | 5.1% | 2.2% | **26.1%** |
| 0.2 | 4.4% | 6.6% | **12.9%** | **42.3%** | **66.2%** |
| | 19.1% | 15.4% | 19.9% | **45.6%** | Total |

# Results - Optimal designs

Comparison of the performance of optimal designs for the 288 possible combinations of simulation parameters




For the same number of trials, a series of 3-arm RCTs test twice more experimental treatments than a series of 2-arm RCTs

# Conclusion

Even when including interim analysis or two-stage design with treatment selection at interim, we still recommend to relax $\alpha$-level

Our recommendation is only valid when considering a series of trials run over a relatively long research horizon and when the supply of new treatments is large

# References


Bayar, M. A., Le Teuff, G., Michiels, S., Sargent, D. J., and Le Deley, M.-C. (2016). New insights into the evaluation of randomized controlled trials for rare diseases over a long-term research horizon: a simulation study. *Statistics in medicine*, 35(19):3245–3258.

Djulbegovic, B., Kumar, A., Glasziou, P. P., Perera, R., Reljic, T., Dent, L., Raftery, J., Johansen, M., Di Tanna, G. L., Miladinovic, B., et al. (2012). New treatments compared to established treatments in randomized trials. *The Cochrane Library*.

Dmitrienko, A., Tamhane, A. C., and Bretz, F. (2009). *Multiple testing problems in pharmaceutical statistics*. CRC Press.

Kim, K. and Tsiatis, A. A. (1990). Study duration for clinical trials with survival response and early stopping rule. *Biometrics*, pages 81–92.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, pages 549–556.

Parmar, M. K., Sydes, M. R., and Morris, T. P. (2016). How do you design randomised trials for smaller populations? a framework. *BMC medicine*, 14(1):183.

Posch, M., Koenig, F., Branson, M., Brannath, W., Dunger-Baldauf, C., and Bauer, P. (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in medicine*, 24(24):3697–3714.

Wieand, S., Schroeder, G., and O'Fallon, J. R. (1994). Stopping when the experimental regimen does not appear to help. *Statistics in medicine*, 13(13-14):1453–1458.

# Q & A

Please, send any additional questions or comments to:
Mohamedamine.BAYAR@gustaveroussy.fr

GUSTAVE ROUSSY
CANCER CAMPUS
GRAND PARIS

cesp

UNIVERSITÉ PARIS SUD

# Simulation parameter

**Hypotheses of how treatments improve over time : future treatment effects**

Relative characterization
Hazard Ratio

| $\mathbb{E}[\mathcal{HR}]$ | $\mathbb{P}[\mathcal{HR} \leqslant 0.5]$ |
|---|---|
| 0.925 | 0.02 |
| 0.950 | 0.02 |
| 0.950 | 0.01 |
| 1.000 | 0.01 |

Absolute characterization
Hazard Rate

$$\lambda^E(t) \sim log\mathcal{N}\left(\mu(t), \sigma^2\right)$$
$$\mu(t) = a \times t + b$$

$$\mathbb{E}[\lambda^E(t)] = e^{\mu(t)+\frac{1}{2}\sigma^2}$$
$$\mathbb{SD}[\lambda^E(t)] = e^{\mu(t)+\frac{1}{2}\sigma^2}\sqrt{e^{\sigma^2}-1}$$

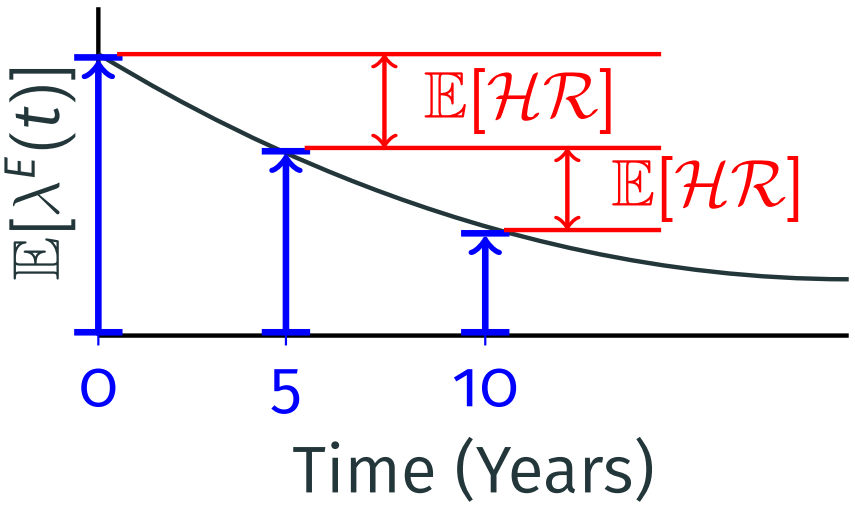
# Simulation parameter

**Hypotheses of how treatments improve over time : future treatment effects**

Relative characterization
Hazard Ratio

| $\mathbb{E}[\mathcal{HR}]$ | $\mathbb{P}[\mathcal{HR} \leqslant 0.5]$ |
|---|---|
| 0.925 | 0.02 |
| 0.950 | 0.02 |
| 0.950 | 0.01 |
| 1.000 | 0.01 |

Absolute characterization
Hazard Rate



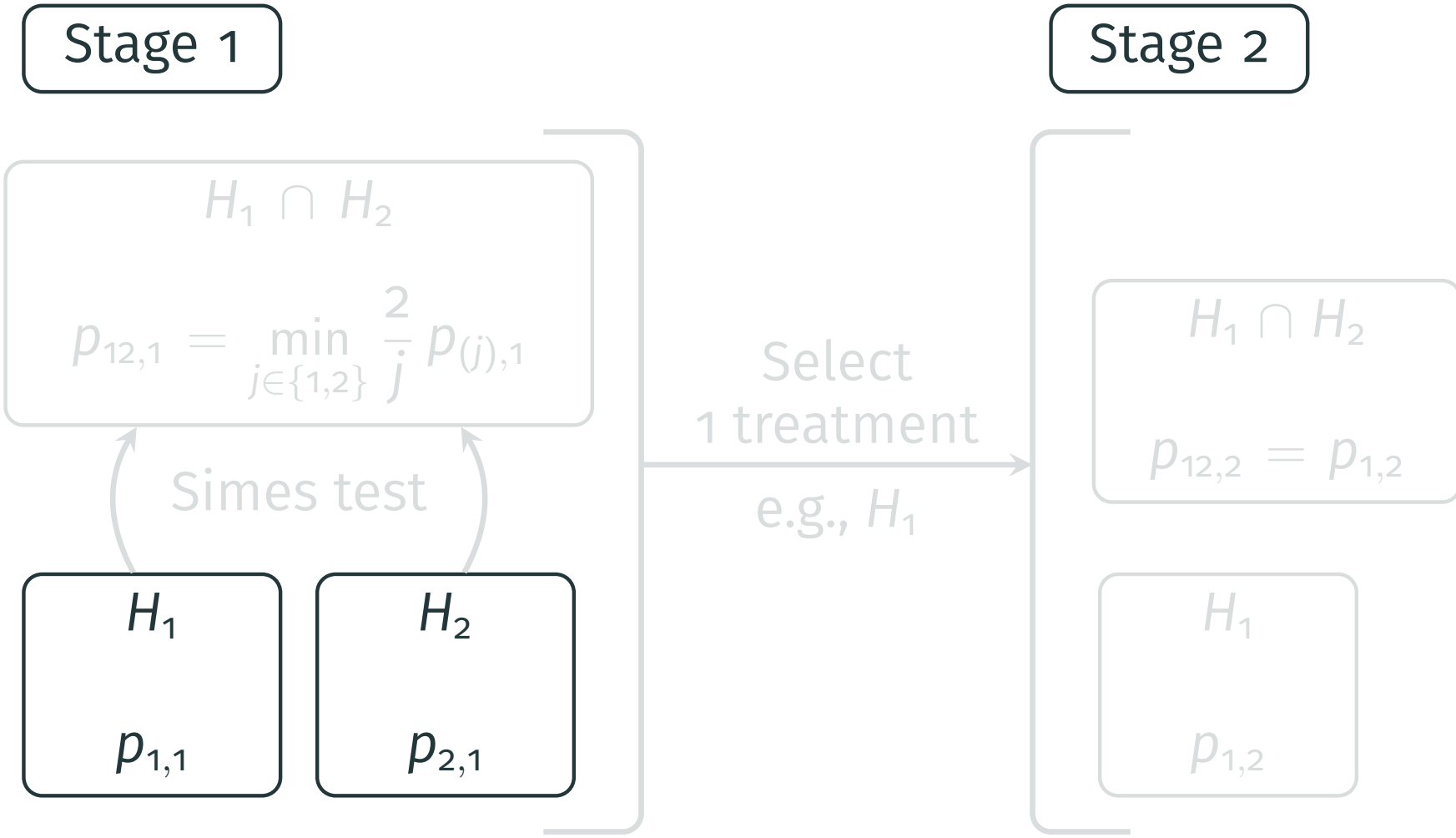



**Figure 1:** P-value definitions of the closed testing procedure using the Simes test for intersection hypotheses, adapted from Dmitrienko et al. (2009)

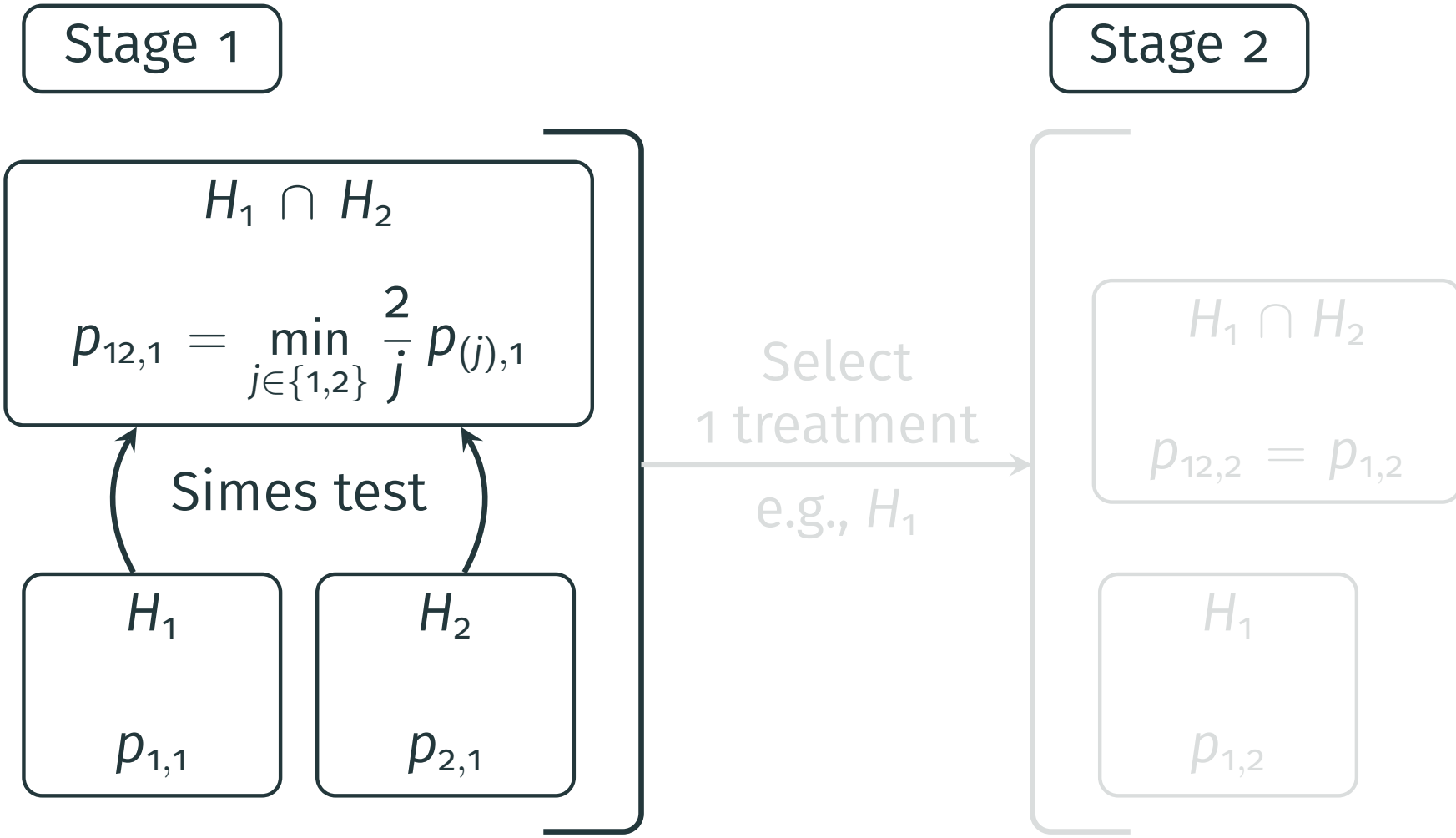



**Figure 1:** P-value definitions of the closed testing procedure using the Simes test for intersection hypotheses, adapted from Dmitrienko et al. (2009)

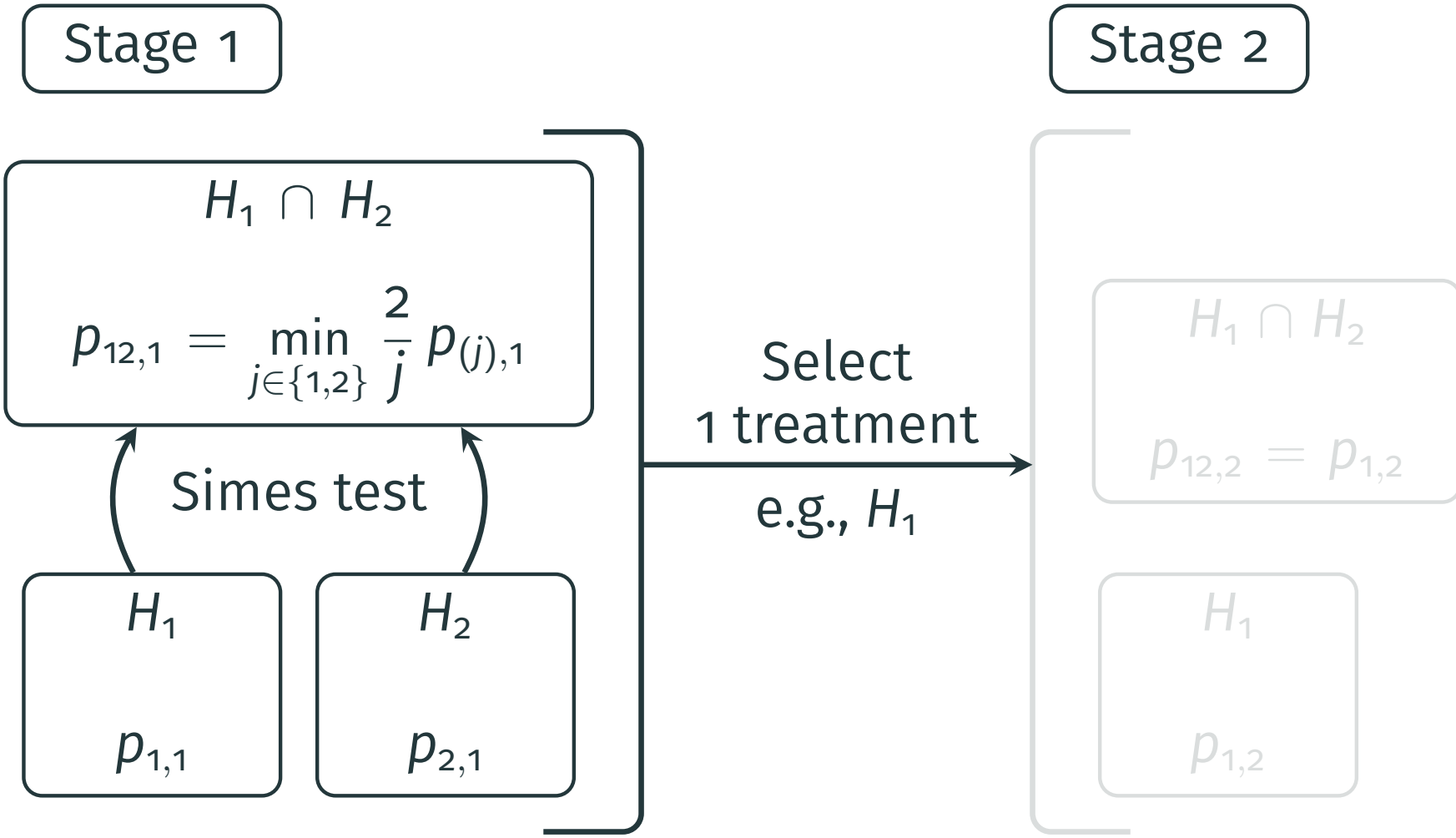



**Figure 1:** P-value definitions of the closed testing procedure using the Simes test for intersection hypotheses, adapted from Dmitrienko et al. (2009)

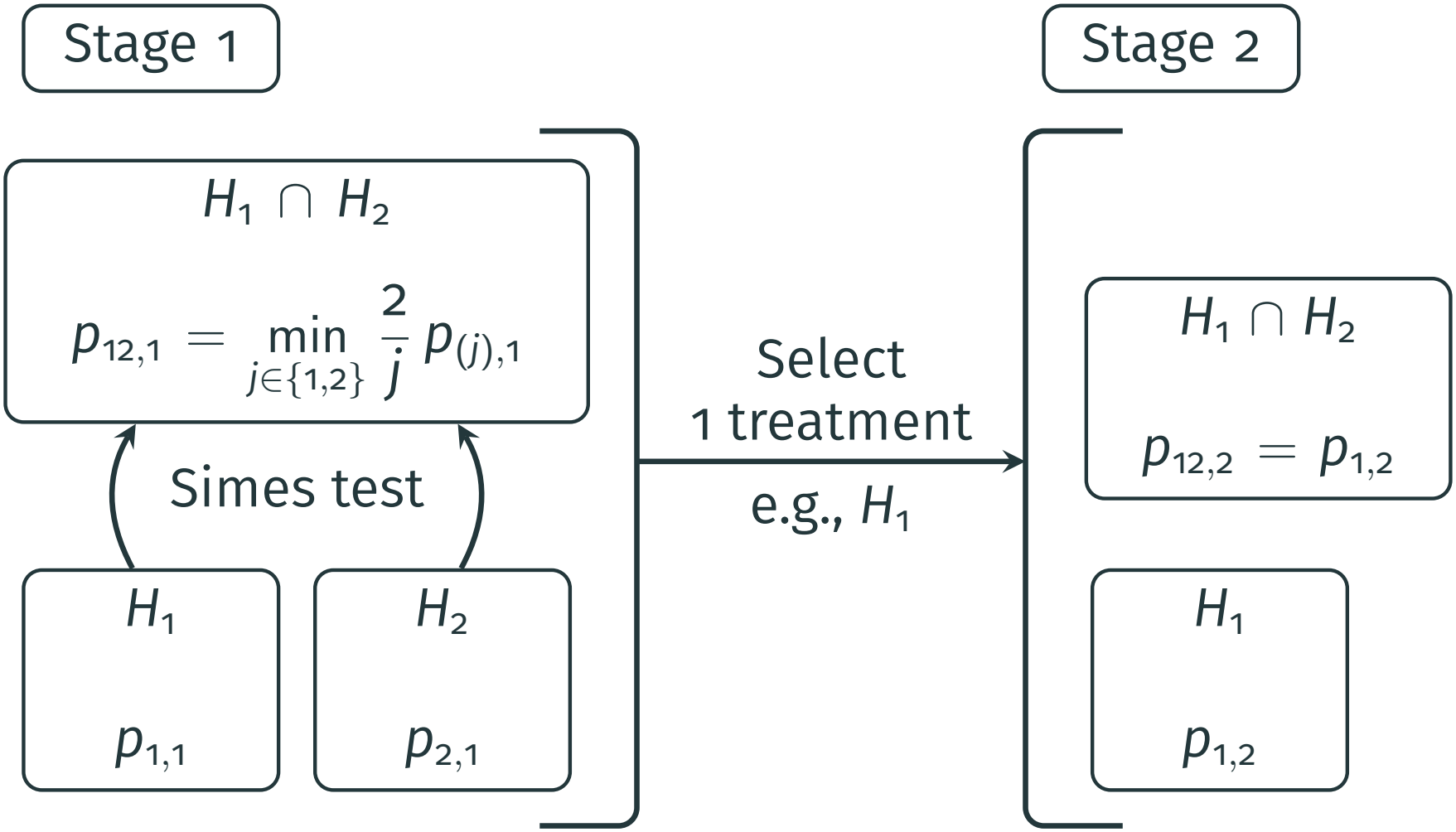



**Figure 1:** P-value definitions of the closed testing procedure using the Simes test for intersection hypotheses, adapted from Dmitrienko et al. (2009)

# Results - Sensitivity analysis

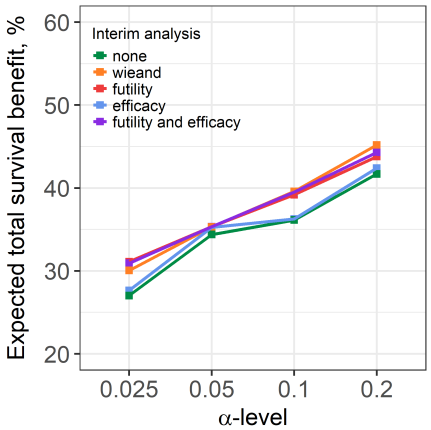
***Accrual rate = 50, 100, 200 patients/year***

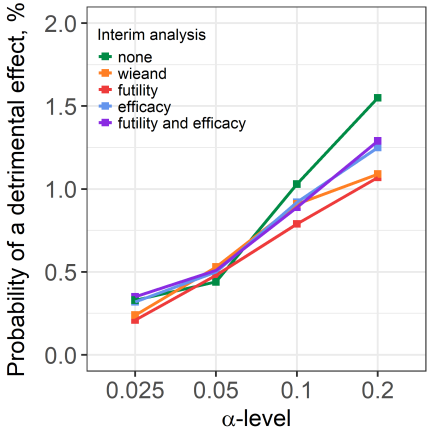
**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

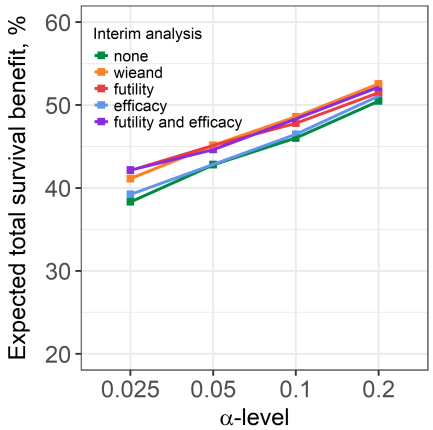
**90% power for an expected $\mathcal{HR}$ of 0.6**

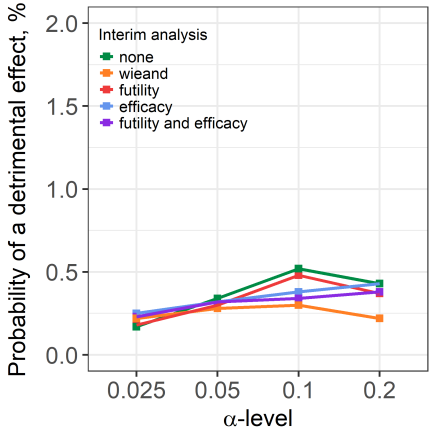
**Accrual rate = 50 patients/year**



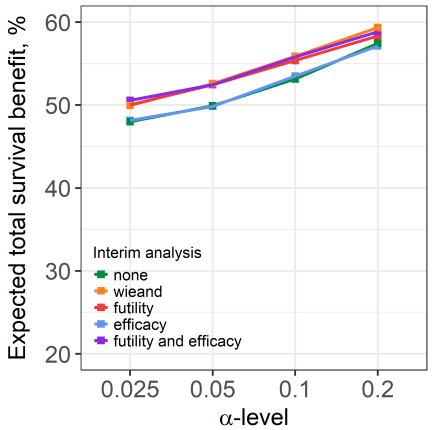



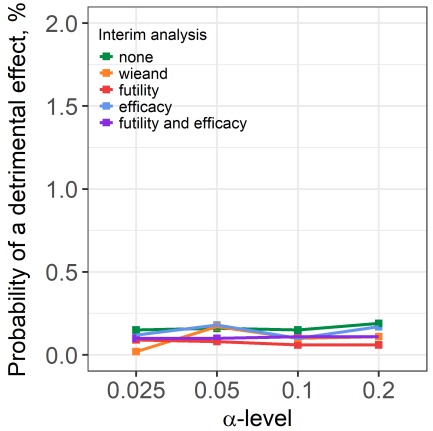

**Accrual rate = 100 patients/year**

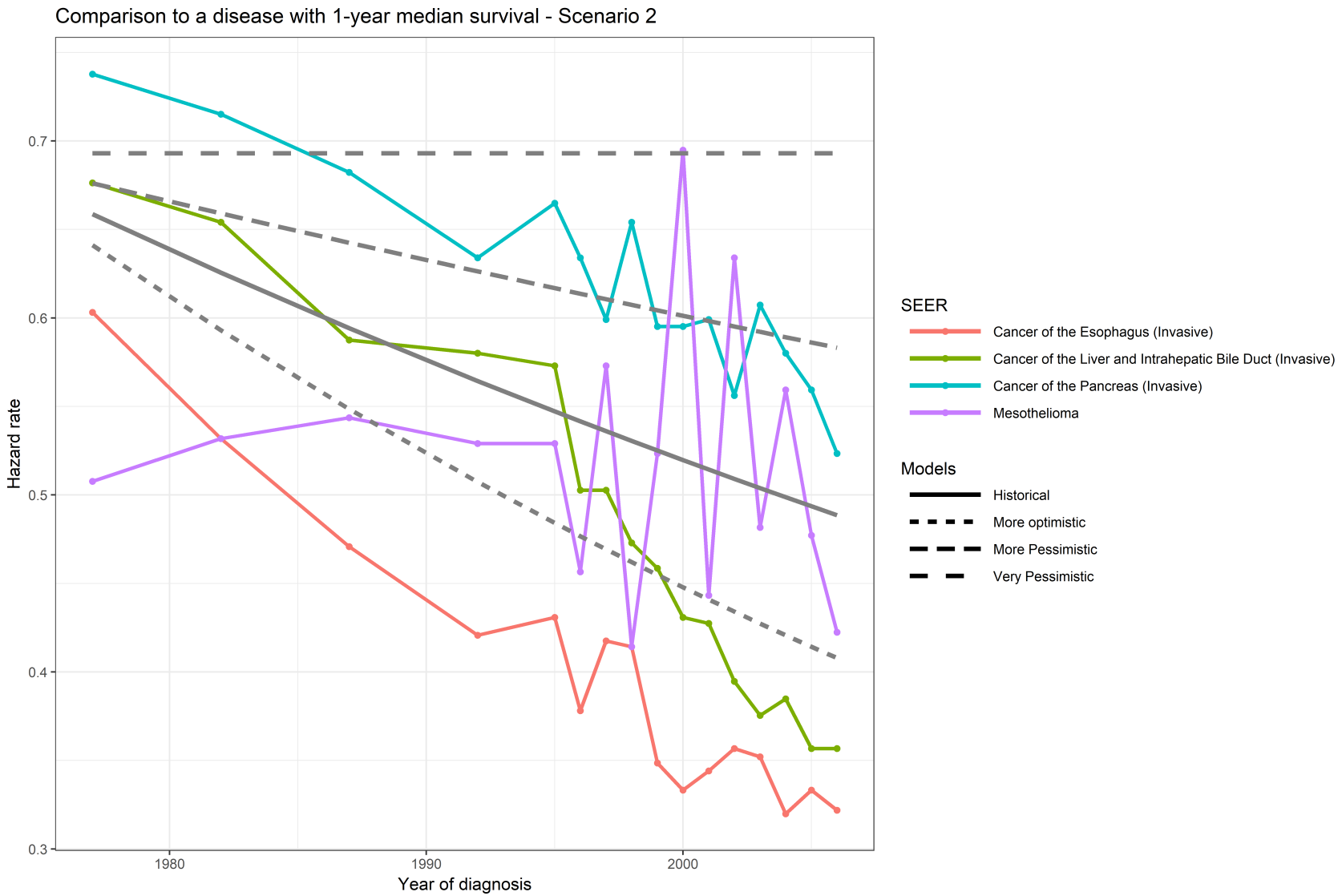


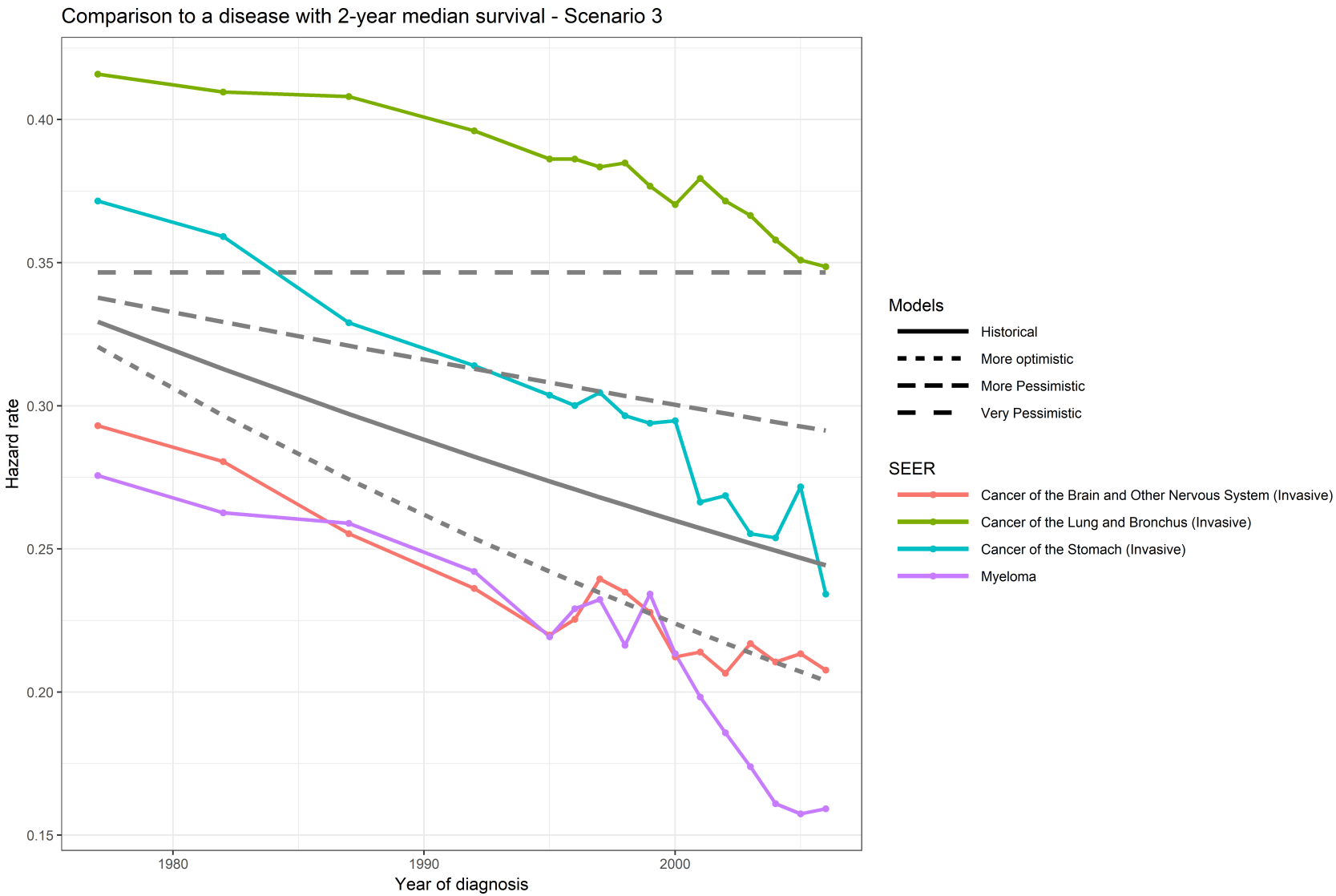



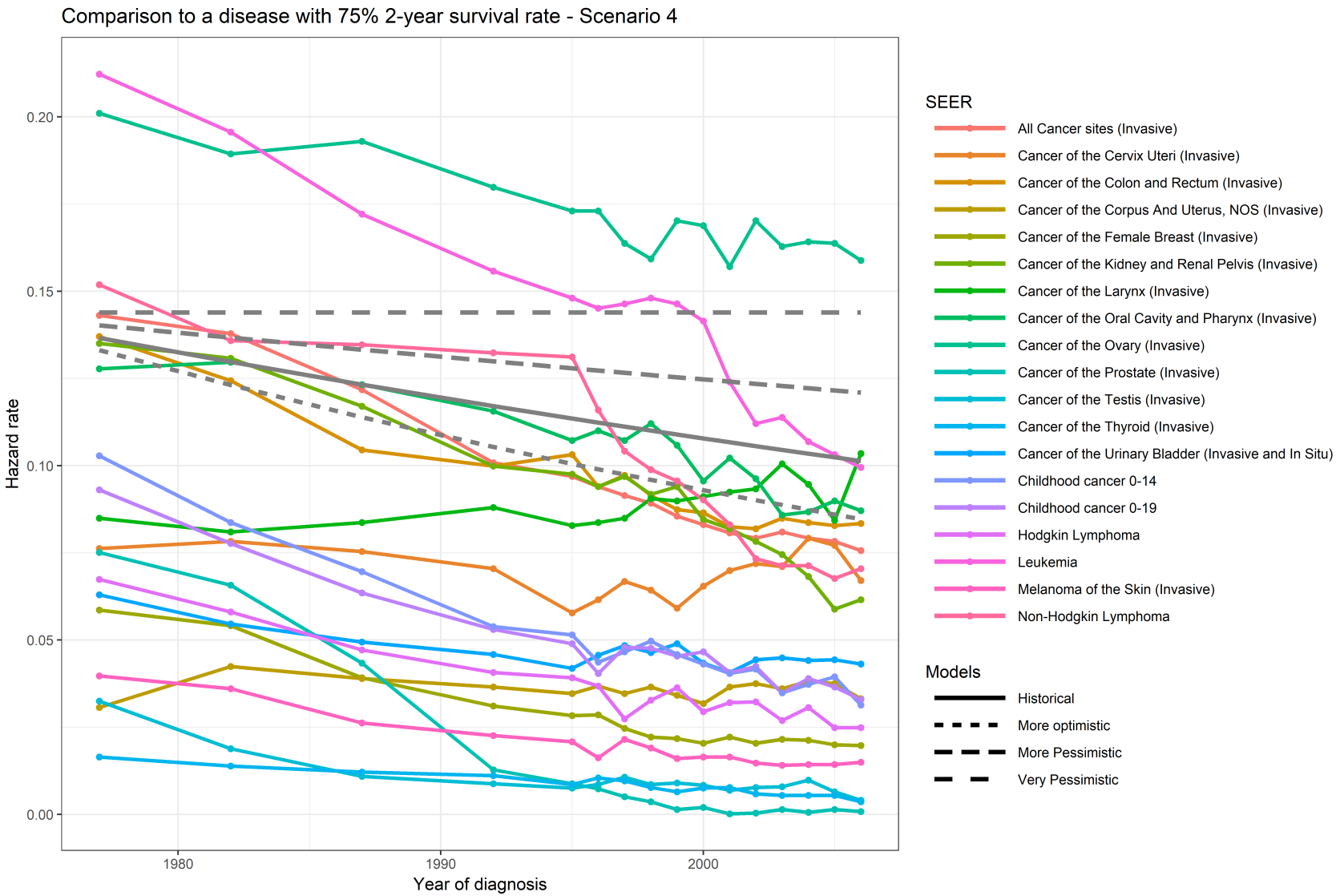

**Accrual rate = 200 patients/year**




Comparison to a disease with 1-year median survival - Scenario 2
Hazard rate
0.7
0.6
0.5
0.4
0.3
1980
1990
2000
Year of diagnosis
SEER
Cancer of the Esophagus (Invasive)
Cancer of the Liver and Intrahepatic Bile Duct (Invasive)
Cancer of the Pancreas (Invasive)
Mesothelioma
Models
Historical
More optimistic
More Pessimistic
Very Pessimistic


Comparison to a disease with 2-year median survival - Scenario 3
Hazard rate
0.40
0.35
0.30
0.25
0.20
0.15
1980
1990
2000
Year of diagnosis
Models
Historical
More optimistic
More Pessimistic
Very Pessimistic
SEER
Cancer of the Brain and Other Nervous System (Invasive)
Cancer of the Lung and Bronchus (Invasive)
Cancer of the Stomach (Invasive)
Myeloma


Comparison to a disease with 75% 2-year survival rate - Scenario 4
Hazard rate
0.20
0.15
0.10
0.05
0.00
1980
1990
2000
Year of diagnosis
SEER
All Cancer sites (Invasive)
Cancer of the Cervix Uteri (Invasive)
Cancer of the Colon and Rectum (Invasive)
Cancer of the Corpus And Uterus, NOS (Invasive)
Cancer of the Female Breast (Invasive)
Cancer of the Kidney and Renal Pelvis (Invasive)
Cancer of the Larynx (Invasive)
Cancer of the Oral Cavity and Pharynx (Invasive)
Cancer of the Ovary (Invasive)
Cancer of the Prostate (Invasive)
Cancer of the Testis (Invasive)
Cancer of the Thyroid (Invasive)
Cancer of the Urinary Bladder (Invasive and In Situ)
Childhood cancer 0-14
Childhood cancer 0-19
Hodgkin Lymphoma
Leukemia
Melanoma of the Skin (Invasive)
Non-Hodgkin Lymphoma
Models
Historical
More optimistic
More Pessimistic
Very Pessimistic