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

Mohamed Amine BAYAR¹,², Gwénaël LE TEUFF¹,², Marie-Cécile LE DELEY¹,²,³, Stefan MICHIELS¹,²

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

¹Service de Biostatistique et d’Epidémiologie, Gustave Roussy, Villejuif, France
²CESP - UPS - UVSQ, INSERM, Université Paris-Saclay, Villejuif, France
³Unité de Méthodologie et Biostatistique, Centre Oscar Lambret, Lille, France
• 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
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
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)
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 $\lambda^c_1$

<table>
<thead>
<tr>
<th>Survival</th>
<th>$\lambda^c_1$</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>median survival of 6 months</td>
<td>$2 \log(2)$</td>
<td>6 months</td>
</tr>
<tr>
<td>median survival of 1 year</td>
<td>$\log(2)$</td>
<td>1 year</td>
</tr>
<tr>
<td>median survival of 2 years</td>
<td>$\frac{\log(2)}{2}$</td>
<td>2 years</td>
</tr>
<tr>
<td>2-year survival rate of 75%</td>
<td>$\frac{\log(4)-\log(3)}{2}$</td>
<td>2 years</td>
</tr>
</tbody>
</table>

Trials within the same series are designed to achieve the same power (80% or 90%) for the same expected 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** $\lambda_1^C$

<table>
<thead>
<tr>
<th>Survival</th>
<th>$\lambda_1^C$</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>median survival of 6 months</td>
<td>$2 \log(2)$</td>
<td>6 months</td>
</tr>
<tr>
<td>median survival of 1 year</td>
<td>$\log(2)$</td>
<td>1 year</td>
</tr>
<tr>
<td>median survival of 2 years</td>
<td>$\frac{\log(2)}{2}$</td>
<td>2 years</td>
</tr>
<tr>
<td>2-year survival rate of 75%</td>
<td>$\frac{\log(4) - \log(3)}{2}$</td>
<td>2 years</td>
</tr>
</tbody>
</table>

Trials within the same series are designed to achieve the same power (80% or 90%) for the same expected 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 $\lambda_1^C$

<table>
<thead>
<tr>
<th>Survival</th>
<th>$\lambda_1^C$</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>median survival of 6 months</td>
<td>$2 \log(2)$</td>
<td>6 months</td>
</tr>
<tr>
<td>median survival of 1 year</td>
<td>$\log(2)$</td>
<td>1 year</td>
</tr>
<tr>
<td>median survival of 2 years</td>
<td>$\frac{\log(2)}{2}$</td>
<td>2 years</td>
</tr>
<tr>
<td>2-year survival rate of 75%</td>
<td>$\frac{\log(4) - \log(3)}{2}$</td>
<td>2 years</td>
</tr>
</tbody>
</table>

Trials within the same series are designed to achieve the same power (80% or 90%) for the same expected 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 $\lambda_1^C$

<table>
<thead>
<tr>
<th>Survival</th>
<th>$\lambda_1^C$</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>median survival of 6 months</td>
<td>$2 \log(2)$</td>
<td>6 months</td>
</tr>
<tr>
<td>median survival of 1 year</td>
<td>$\log(2)$</td>
<td>1 year</td>
</tr>
<tr>
<td>median survival of 2 years</td>
<td>$\frac{\log(2)}{2}$</td>
<td>2 years</td>
</tr>
<tr>
<td>2-year survival rate of 75%</td>
<td>$\frac{\log(4) - \log(3)}{2}$</td>
<td>2 years</td>
</tr>
</tbody>
</table>

Trials within the same series are designed to achieve the same power (80% or 90%) for the same expected HR of 0.5, 0.6 or 0.75
### Hypotheses of how treatments improve over time: future treatment effects

#### Relative characterization

<table>
<thead>
<tr>
<th>Hazard Ratio</th>
<th>P[HR ≤ 0.5]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.925</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.01</td>
</tr>
<tr>
<td>1.000</td>
<td>0.01</td>
</tr>
</tbody>
</table>

- **Historical distribution** derived from the meta-analysis of 698 RCTs on > 200 000 patients (Djulbegovic et al., 2012)
- **Other distributions** ± optimistic or pessimistic
## Simulation parameters

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

#### Relative characterization

<table>
<thead>
<tr>
<th>Hazard Ratio</th>
<th>$P[HR \leq 0.5]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[HR]$</td>
<td></td>
</tr>
<tr>
<td>0.925</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.01</td>
</tr>
<tr>
<td>1.000</td>
<td>0.01</td>
</tr>
</tbody>
</table>

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

Hypotheses of how treatments improve over time: future treatment effects

Relative characterization
Hazard Ratio

<table>
<thead>
<tr>
<th>$\mathbb{E}[HR]$</th>
<th>$\mathbb{P}[HR \leq 0.5]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.925</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.01</td>
</tr>
<tr>
<td>1.000</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Absolute characterization
Hazard Rate $\lambda^E(t)$
Hazard rate of the control arm of the first trial of the series $\lambda^C_1 = \log(2)$

- More optimistic distribution
- Historical distribution
- More pessimistic distribution
- Very pessimistic distribution
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 HRs
Illustration of one possible series

$\mathcal{HR}_{\text{series}}$

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

$\frac{\lambda^E_1}{\lambda^C_1} \times \frac{\lambda^C_2}{\lambda^C_2} \times \frac{\lambda^E_3}{\lambda^C_3} \times \frac{\lambda^C_4}{\lambda^C_4} = \frac{\lambda^C_4}{\lambda^C_1}$
Illustration of one possible series

Distribution of future treatment effects

\[ \mathcal{HR}_{\text{series}} \]

\[ \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C} \]
Illustration of one possible series

\[ \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C} \]
Illustration of one possible series

\[ \mathcal{H}_\text{series} \]

\[ \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C} \]
Illustration of one possible series

\[ \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C} \]
Illustration of one possible series

\[
\mathcal{HR}_{\text{series}} \quad \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C}
\]
Illustration of one possible series

\[ \mathcal{HR}_{\text{series}} = \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^E} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^E} = \frac{\lambda_4^C}{\lambda_1^C} \]
Illustration of one possible series

\[ \mathcal{HR}_{\text{series}} = \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C} \]
Illustration of one possible series

\[ \mathcal{H}_\text{series} = \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^E} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^E} = \frac{\lambda_4^C}{\lambda_1^C} \]

Distribution of future treatment effects
Illustration of one possible series

\[ \mathcal{H}_\text{series} = \frac{\lambda_1^E}{\lambda_1^C} \times \frac{\lambda_2^C}{\lambda_2^C} \times \frac{\lambda_3^E}{\lambda_3^C} \times \frac{\lambda_4^C}{\lambda_4^C} = \frac{\lambda_4^C}{\lambda_1^C} \]
At the end of the 15-year research period

**Total survival benefit**

\[
\frac{1}{HR_{\text{series}}} - 1 = \frac{\lambda_{\text{Control}}}{\lambda_{\text{First trial}}} - 1 = \frac{\lambda_{\text{Selected}}}{\lambda_{\text{Last trial}}} - 1
\]

**Example**

For a series of RCTs,

At baseline, median survival = 12 months \(\Rightarrow\) \(\lambda_{\text{Control}}^{\text{First trial}} = 0.69\)

After 15 year, \(\lambda_{\text{Selected}}^{\text{Last trial}} = 0.46\) \(\Rightarrow\) median survival = 18 months

\(HR_{\text{series}} = 0.67\) \(\Leftrightarrow\) Total survival benefit = 50%
Performance metrics

10 000 repetitions of the 15-year research period

**Expected total survival benefit (Gain)**

\[
\mathbb{E} \left[ \frac{\lambda_{\text{Control \ First trial}}}{\lambda_{\text{Selected \ Last trial}}} - 1 \right]
\]

**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

\[
P \left[ \lambda_{\text{Selected \ Last trial}} > \lambda_{\text{Control \ First trial}} \right]
\]
Series of two-arm RCTs with an interim analysis (IA)

IA performed when $\frac{1}{2}$ of the required events are expected to be attained

- **Wieand stopping rule for futility** (Wieand et al., 1994)
  \[ \widehat{HR} \geq 1 \Rightarrow \text{stop the trial for futility} \]

- **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)
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 $HR$ 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 $HR$ 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 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%

Expected total survival benefit, %

Expected total survival benefit, %

Expected total survival benefit, %

Expected total survival benefit, %

Probability of a detrimental effect, %

Probability of a detrimental effect, %

Probability of a detrimental effect, %

Probability of a detrimental effect, %

α-level

α-level

α-level

α-level

α-level

α-level

Interim analysis
- none
- weand
- futility
- efficacy
- futility and efficacy

Interim analysis
- none
- weand
- futility
- efficacy
- futility and efficacy

Interim analysis
- none
- weand
- futility
- efficacy
- futility and efficacy

Interim analysis
- none
- weand
- futility
- efficacy
- futility and efficacy
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 $HR$ of 0.6

More optimistic

Historical

More pessimistic

Very pessimistic
Accrual rate = 50 patients/year
Baseline median survival = 1 year
Historical distribution of treatment effects
90% power for an expected $HR$ of 0.6
### Results - Optimal designs

\[
\text{Argmax } E \left[ \frac{\lambda_{\text{Control \ First trial}}}{\lambda_{\text{Selected \ Last trial}}} - 1 \right]
\]

subject to \( P \left[ \frac{\lambda_{\text{Selected \ Last trial}}}{\lambda_{\text{Control \ First trial}}} > 1 \right] < 1.0\%

**Accrual rate = 50 patients/year**

**Baseline median survival = 1 year**

**Historical distribution of treatment effects**

**90% power for an expected HR of 0.6**

<table>
<thead>
<tr>
<th>Optimal design</th>
<th>Traditional design</th>
<th>Series of 2-arm RCTs with no interim analysis</th>
<th>Series of 2-arm RCTs with interim analysis</th>
<th>Series of 3-arm RCTs with selection at interim</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )-level</td>
<td>0.025</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Interim analysis</td>
<td>None</td>
<td>None</td>
<td>Wieand</td>
<td>threshold = 0.2</td>
</tr>
<tr>
<td>Number of trial</td>
<td>3.0</td>
<td>4.0</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Gain</td>
<td>27.0%</td>
<td>36.1%</td>
<td>39.6%</td>
<td>42.7%</td>
</tr>
<tr>
<td>Risk</td>
<td>0.33%</td>
<td>1.03%</td>
<td>0.91%</td>
<td>0.87%</td>
</tr>
</tbody>
</table>
Results - Optimal designs

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

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

Accrual rate = 50 patients/year
Baseline median survival = 1 year
Historical distribution of treatment effects
90% power for an expected HR of 0.6

<table>
<thead>
<tr>
<th>Optimal design</th>
<th>Traditional design</th>
<th>Series of 2-arm RCTs with no interim analysis</th>
<th>Series of 2-arm RCTs with interim analysis</th>
<th>Series of 3-arm RCTs with selection at interim</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-level</td>
<td>0.025</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Interim analysis</td>
<td>None</td>
<td>None</td>
<td>Wieand</td>
<td>threshold = 0.2</td>
</tr>
<tr>
<td>Number of trial</td>
<td>3.0</td>
<td>4.0</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Gain</td>
<td>27.0%</td>
<td>36.1%</td>
<td>39.6%</td>
<td>42.7%</td>
</tr>
<tr>
<td>Risk</td>
<td>0.33%</td>
<td>1.03%</td>
<td>0.91%</td>
<td>0.87%</td>
</tr>
</tbody>
</table>
Results - Optimal designs

\[
\text{Argmax } \mathbb{E} \left[ \frac{\lambda_{\text{Control, First trial}}}{\lambda_{\text{Selected, Last trial}}} - 1 \right]
\]

subject to \( \mathbb{P} \left[ \frac{\lambda_{\text{Selected, Last trial}}}{\lambda_{\text{Control, First trial}}} > 1.0\% \right] < 1.0\% \)

Accrual rate = 50 patients/year
Baseline median survival = 1 year
Historical distribution of treatment effects
90% power for an expected HR of 0.6

<table>
<thead>
<tr>
<th>Optimal design</th>
<th>Traditional design</th>
<th>Series of 2-arm RCTs with no interim analysis</th>
<th>Series of 2-arm RCTs with interim analysis</th>
<th>Series of 3-arm RCTs with selection at interim</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )-level</td>
<td>0.025</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Interim analysis</td>
<td>None</td>
<td>None</td>
<td>Wieand</td>
<td>threshold = 0.2</td>
</tr>
<tr>
<td>Number of trial</td>
<td>3.0</td>
<td>4.0</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Gain</td>
<td>27.0%</td>
<td>36.1%</td>
<td>39.6%</td>
<td>42.7%</td>
</tr>
<tr>
<td>Risk</td>
<td>0.33%</td>
<td>1.03%</td>
<td>0.91%</td>
<td>0.87%</td>
</tr>
</tbody>
</table>
Results - Optimal designs

\[
\text{Argmax } \mathbb{E} \left[ \frac{\lambda_{\text{Control First trial}}}{\lambda_{\text{Selected Last trial}}} - 1 \right]
\]

subject to \( \mathbb{P} \left[ \lambda_{\text{Selected Last trial}} > \lambda_{\text{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 HR of 0.6

<table>
<thead>
<tr>
<th>Optimal design</th>
<th>Traditional design</th>
<th>Series of 2-arm RCTs with no interim analysis</th>
<th>Series of 2-arm RCTs with interim analysis</th>
<th>Series of 3-arm RCTs with selection at interim</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )-level</td>
<td>0.025</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Interim analysis</td>
<td>None</td>
<td>None</td>
<td>Wieand</td>
<td>threshold = 0.2</td>
</tr>
<tr>
<td>Number of trial</td>
<td>3.0</td>
<td>4.0</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Gain</td>
<td>27.0%</td>
<td>36.1%</td>
<td>39.6%</td>
<td>42.7%</td>
</tr>
<tr>
<td>Risk</td>
<td>0.33%</td>
<td>1.03%</td>
<td>0.91%</td>
<td>0.87%</td>
</tr>
</tbody>
</table>
**Results - Optimal designs**

\[
\text{Argmax } \mathbb{E} \left[ \frac{\lambda_{\text{Control, First trial}}}{\lambda_{\text{Selected, Last trial}}} - 1 \right]
\]

subject to \( \mathbb{P} \left[ \lambda_{\text{Selected, Last trial}} > \lambda_{\text{Control, First trial}} \right] < 1.0\% \)

Accrual rate = 50 patients/year  
Baseline median survival = 1 year  
Historical distribution of treatment effects  
90% power for an expected HR of 0.6

<table>
<thead>
<tr>
<th>Optimal design</th>
<th>Traditional design</th>
<th>Series of 2-arm RCTs with no interim analysis</th>
<th>Series of 2-arm RCTs with interim analysis</th>
<th>Series of 3-arm RCTs with selection at interim</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )-level</td>
<td>0.025</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Interim analysis</td>
<td>None</td>
<td>None</td>
<td>Wieand</td>
<td>threshold = 0.2</td>
</tr>
<tr>
<td>Number of trial</td>
<td>3.0</td>
<td>4.0</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Gain</td>
<td>27.0%</td>
<td>36.1%</td>
<td>39.6%</td>
<td>42.7%</td>
</tr>
<tr>
<td>Risk</td>
<td>0.33%</td>
<td>1.03%</td>
<td>0.91%</td>
<td>0.87%</td>
</tr>
</tbody>
</table>
For the 288 possible combinations of simulation parameters

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

<table>
<thead>
<tr>
<th>Interim Analysis</th>
<th>One-sided $\alpha$-level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.025</td>
</tr>
<tr>
<td>No interim analysis</td>
<td>0.0%</td>
</tr>
<tr>
<td>Wieand stopping rule for futility</td>
<td>5.6%</td>
</tr>
<tr>
<td>OBF $\beta$—spending stopping rule for futility</td>
<td>5.2%</td>
</tr>
<tr>
<td>OBF $\alpha$—spending stopping rule for efficacy</td>
<td>0.3%</td>
</tr>
<tr>
<td>Combining the latter two</td>
<td>5.2%</td>
</tr>
<tr>
<td></td>
<td>16.4%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>First stage threshold</th>
<th>Second stage $\alpha$-level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.025</td>
</tr>
<tr>
<td>0.05</td>
<td>0.4%</td>
</tr>
<tr>
<td>0.1</td>
<td>3.7%</td>
</tr>
<tr>
<td>0.15</td>
<td>10.7%</td>
</tr>
<tr>
<td>0.2</td>
<td>4.4%</td>
</tr>
<tr>
<td></td>
<td>19.1%</td>
</tr>
</tbody>
</table>
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.


Q & A

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

Hypotheses of how treatments improve over time: future treatment effects

Relative characterization
Hazard Ratio

<table>
<thead>
<tr>
<th>$\mathbb{E}[HR]$</th>
<th>$\mathbb{P}[HR \leq 0.5]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.925</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.01</td>
</tr>
<tr>
<td>1.000</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Absolute characterization
Hazard Rate

$\lambda^E(t) \sim \log N(\mu(t), \sigma^2)$

$\mu(t) = a \times t + b$

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

$\text{SD}[\lambda^E(t)] = e^{\mu(t) + \frac{1}{2}\sigma^2} \sqrt{e^{\sigma^2} - 1}$
Hypotheses of how treatments improve over time: future treatment effects

Relative characterization
Hazard Ratio

<table>
<thead>
<tr>
<th>$E[HR]$</th>
<th>$P[HR \leq 0.5]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.925</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.01</td>
</tr>
<tr>
<td>1.000</td>
<td>0.01</td>
</tr>
</tbody>
</table>

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)
\[ H_1 \cap H_2 \]

\[ p_{12,1} = \min_{j \in \{1, 2\}} \frac{2}{j} p(j), 1 \]

Stage 1

Simes test

Stage 2

Select 1 treatment e.g., \( H_1 \)

\[ H_1 \cap H_2 \]

\[ p_{12,2} = p_{1,2} \]

\[ H_1 \]

\[ H_2 \]

\[ p_{1,1} \]

\[ p_{2,1} \]

\[ p_{1,2} \]

**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 $HR$ of 0.6

**Accrual rate = 50 patients/year**

**Accrual rate = 100 patients/year**

**Accrual rate = 200 patients/year**
Comparison to a disease with 2-year median survival - Scenario 3

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

Hazard rate vs Year of diagnosis