# Seamless phase 2/3 design with benefit-risk driven treatment selection using multi-criteria decision analysis (MCDA)

Marco Ratta[1,2,3], Donia Skanji[2], Zhaoyang Teng[2], Gaëlle Saint-Hilary[1,3], Mauro Gasparini[1], Pavel Mozgunov[3,4]

[1]Politecnico di Torino, Italy
[2]Servier International Research Institut (IRIS)
[3]Saryga, France
[4]University of Cambridge, United Kingdom

**PSI Conference**
**Belfast, June 16 2026**

# FDA project Optimus

SARYGA
STATISTICS & METHODOLOGY

U.S. FOOD & DRUG ADMINISTRATION

- Initiative by the **FDA Oncology Center of Excellence (2021)** to modernize dose optimization in oncology.
- Moves beyond **Maximum Tolerated Dose (MTD)** toward identifying the **Optimal Dose** based on both **efficacy and safety**.
- Emphasizes the need for a **comprehensive benefit-risk analysis** in dose selection.
- Promotes **early-phase studies with multiple dose levels**, integrating pharmacokinetics, pharmacodynamics, and clinical outcomes.
- Aims to improve **patient safety, treatment adherence**, and **overall therapeutic value**.

# Trial design

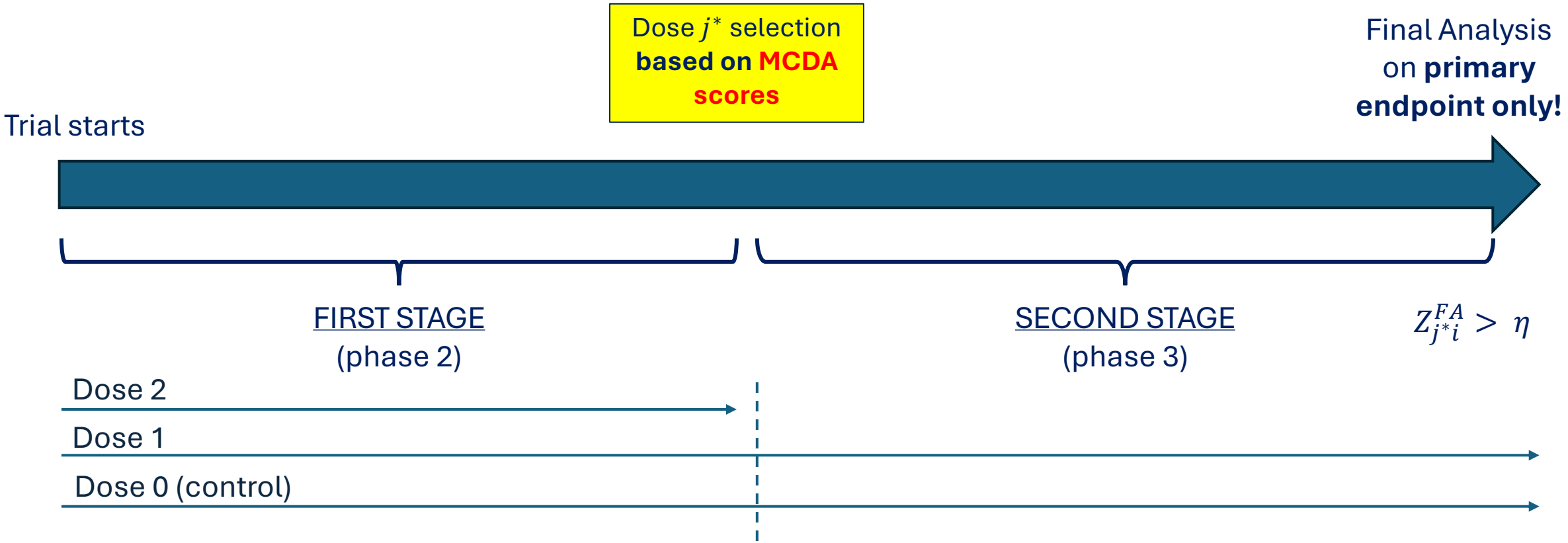
$j = 0,1,2,\dots,J$ : index of the dose ( $j = 0$ is the control arm)

$i = 1,2,\dots,I$ : index of safety/efficacy endpoints ( $i = 1$ is the primary efficacy endpoint)

$\theta_{ji}$ : j-th parameter of interest for the i-th dose

$\delta_{ji}$ : treatment difference in the i-th endpoint related to the j-th dose

$Z_{ji}^{\star}$: Test statistic for the i-th endpoint related to the j-th dose at the time point $\star = IA, FA$



# Structure of the parameters MLEs

For each dose $j$ the maximum likelihood estimators of the parameters of interest are supposed to follow a multivariate normal distribution

$$\begin{pmatrix} \hat{\theta}_{j1} \\ \vdots \\ \hat{\theta}_{jI} \end{pmatrix} = Normal\left( \begin{pmatrix} \theta_{j1} \\ \vdots \\ \theta_{jI} \end{pmatrix}, \begin{pmatrix} \sigma_{j1}^2 & \cdots & \rho_{1I}\sigma_{j1}\sigma_{jI} \\ \vdots & \ddots & \vdots \\ \rho_{1I}\sigma_{j1}\sigma_{jI} & \cdots & \sigma_{jI}^2 \end{pmatrix} \right)$$

- The standard errors $\sigma_{ji}$ depend on the **Fisher information** available
- For each arm $j$, **correlations $\boldsymbol{\rho_{pq}}$** between parameters $\theta_{jp}$ and $\theta_{jq}$ are assumed to be known

Note that for many endpoints an asymptotic normal distribution for a transformation of the parameters of interest holds, e.g. the **log odds** for binary endpoints or the **log hazard** for survival endpoints

# Multi-criteria decision analysis (MCDA)

Multi-criteria decision analysis (MCDA) is a structured approach for **decision making** that involve balancing multiple factors.

For each dose *j*, the true MCDA score is constructed as follows:

- A **factor-specific score** is constructed plugging $\theta_{ji}$ in a *partial value function* $u(x)$,

$$u(x) = \begin{cases} 0 \\ \frac{x - \theta_i^L}{\theta_i^U - \theta_i^L} \\ 1 \end{cases}$$

$\theta_i^L, \theta_i^U$ represent the maximum and minimum plausible values for the parameters

- An aggregated **MCDA score** is constructed

$$MCDA_j = \sum_{i=1}^{I} \omega_i u(\theta_{ji}) \qquad \sum_{i=1}^{I} \omega_i = 1$$

A trivial way to estimate the MCDA scores consists in using the MLEs $\hat{\theta}_{ji}^{IA}$ instead of the true parameters $\theta_{ji}$, however the variability of the estimates is not taken into account.

# Probabilistic MCDA

In order to account for the **variability in the parameters estimates**, we can work in a Bayesian framework, following this procedure:

- Non informative **prior distributions** are assigned to the parameters, namely $\theta_{ji}^{prior}$
- Once data are observed **posterior distributions** $\theta_{ji}^{post}$ are obtained via Bayesian rule
- The MCDA score is found transforming $\theta_{ji}^{post}$ as follows

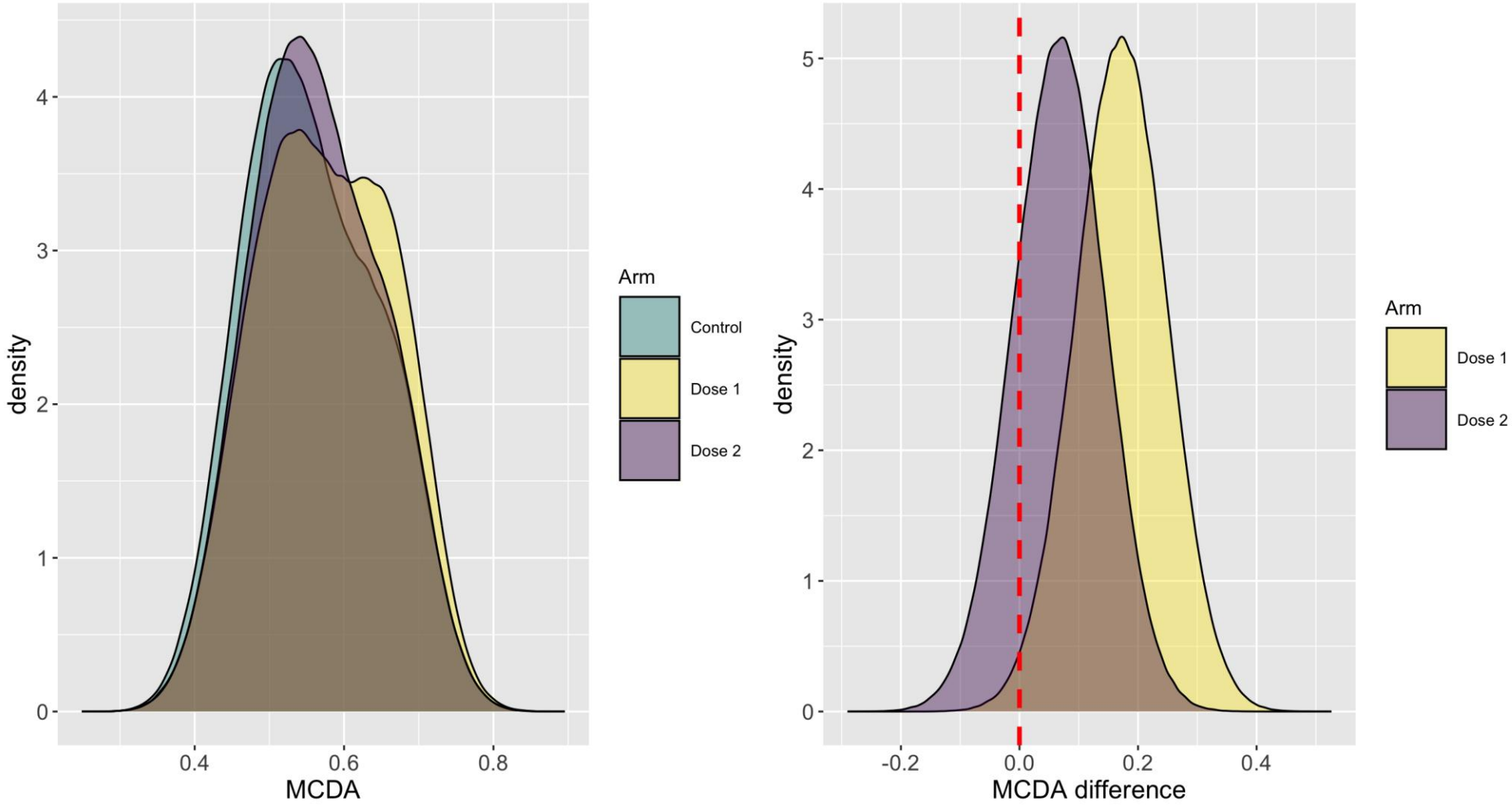
$$MCDA_j = \sum_{j=1}^{J} w_j u\left(\theta_{ji}^{post}\right)$$

DOSE SELECTION RULE

- For each active dose, the posterior probability that its MCDA score is better than control is found and the dose with higher probability is selected

$$P_j = \mathbb{P}\left(MCDA_j > MCDA_0\right) \qquad S = \underset{j=1,\dots,J}{\operatorname{argmax}} P_j$$

# Visual representation of the selection rule




$P_1 = 0.99$
$P_2 = 0.81$

Dose 1 is selected

# Distribution of the Z statistic (an idea..)

The interim test statistics $Z_{S1}^{FA}$ can be written as

$$f_{Z_{S1}^{FA}}(x) = \sum_{j=1}^{J} \int_{-\infty}^{+\infty} f_{Z_{j1}^{FA} \mid Z_{j1}^{IA}}\left(z \mid Z_{j1}^{IA} = x\right) f_{Z_{j1}^{IA} \mid S}(x \mid S = j)\, f_S(j)\, dz$$

**KEY ISSUE**

The distribution of the selection random variable $S = \underset{j=1,\ldots,J}{\operatorname{argmax}} P_j$, has no analytical form, because:

- It is based on a **probabilistic criterion** (based on Bayesian posterior probabilities)
- The posterior distribution of the MCDA scores is **discontinuous** due to the truncation introduced by the linear partial value function $u(\cdot)$

# Approximation of the selection criterion

SARYGA
STATISTICS & METHODOLOGY

**Hypothesis:**

The upper and lower bounds of the linear partial value function $u(\cdot)$, namely $\theta_i^U$ and $\theta_i^L$ are chosen so that the posterior probability $P(\theta_i^L < \theta_i < \theta_i^U) \approx 1$

**Equivalent selection rule:**

An equivalence selection rule is:

$$S = \underset{j=1,\ldots,J}{\operatorname{argmax}}\, \gamma_j$$

$$\gamma_j = \frac{\sum_{I=1}^{I} \omega_i \left[\prod_{k\neq i}\left(\theta_k^U - \theta_k^L\right)\right]\left(\hat{\theta}_{ji} - \hat{\theta}_{0i}\right)}{\sqrt{\sum_{p,q} \rho_{pq}\omega_p\omega_q\left[\prod_{k\neq p}(\theta_k^U - \theta_k^L)\right]\left[\prod_{k\neq q}(\theta_k^U - \theta_k^L)\right]\left(\sigma_{jq}\sigma_{jp} + \sigma_{0q}\sigma_{0p}\right)}}$$

- More **practical**, as does not need for Monte Carlo simulation
- An **analytical expression for the Z statistic** at final analysis can be derived

# Strong control of type I Error control

Since the hypothesis testing at final analysis is **only on the primary endpoint**, the null hypothesis is

$$H_0\colon\ \theta_{01} = \theta_{11} = \theta_{21} = \cdots = \theta_{I1},$$
$$\theta_{ji} \in \mathbb{R}, \forall j, \forall i > 1$$

INFINITE MANY NULL CONFIGURATIONS !

HOW DO WE CHOOSE THE CONFIGURATION FOR TYPE I ERROR CONTROL ?

It can be proven that The distribution of $Z_{S1}^{FA}$ is

- Stochastically increasing with $\delta_{ji}, \forall j, \forall i$
- Stochastically decreasing with $\sigma_{ji}^2, \forall j, \forall i$

**Worst Case Configuration**

$$\mathcal{C}\colon\quad \theta_{ji} = \theta_{0i} = \underset{\theta_{ji}}{\operatorname{argmax}}\, \sigma_{ji}^2 \quad \forall j > 0, \forall i$$

**Strong control** of type I error is achieved if

$$\eta = \left| \underset{q}{\operatorname{argmin}}\, F\left(Z_{S1}^{FA} > q\right) - \alpha \right| \quad \text{under} \quad \mathcal{C}$$

# Power

**Power** is defined as the joint probability of these two events:

- Select a treatment which is effective on the primary endpoint
- Reject the null hypothesis for the selected treatment on the primary endpoint

However, as per the «null configurations», we have also **many alternative configurations**:

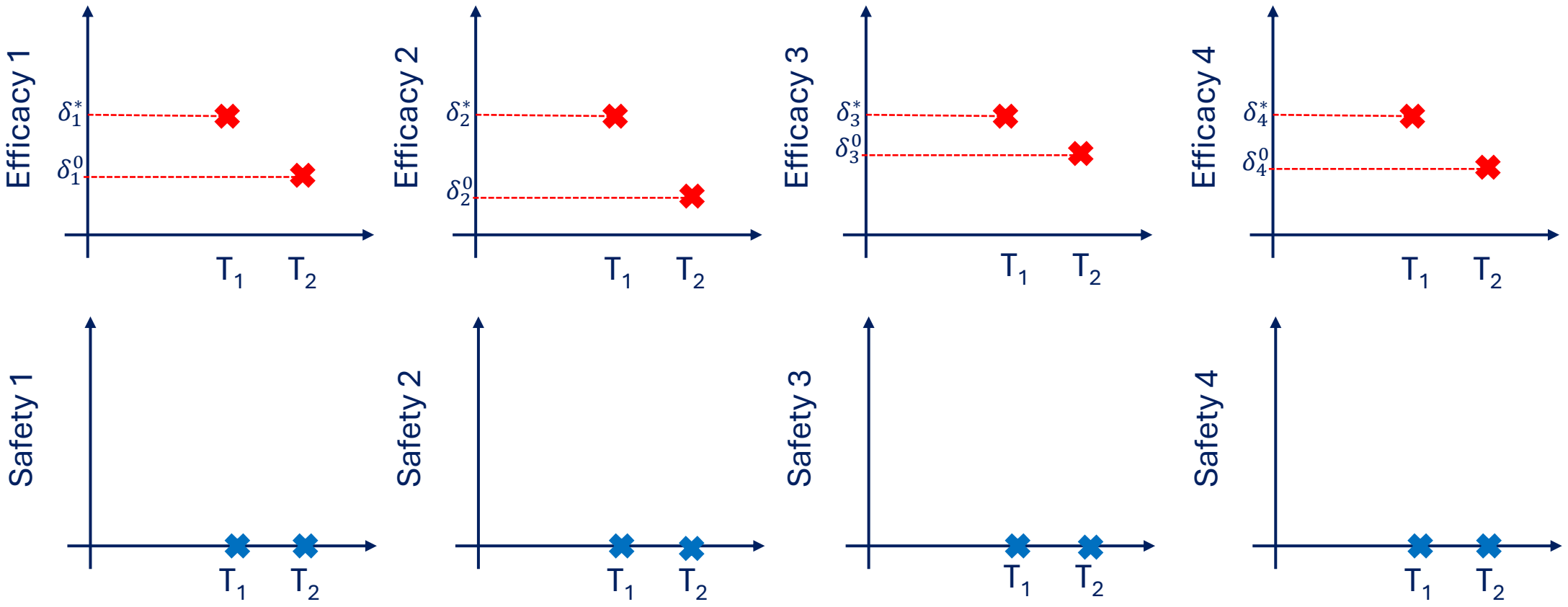
$$H_A\colon\ \theta_{j1} > \theta_{01} \text{ for at least one } j = 1, \dots, J$$
$$\theta_{ji} \in \mathbb{R}, \forall j, \forall i > 1$$

HOW DO WE CHOOSE THE CONFIGURATION FOR POWER ?

1. Assume that there exists **one single dose** which is superior in all efficacy endpoints at a level $\delta_i^*$
2. Assume that the efficacy of all the other active doses is at the **maximum non relevant level** $\delta_i^0$
3. All the doses are as safe as the control

**Least Favorable Configuration**

# Least Favorable Configuration (LFC)




The LFC represents the **best possible configuration** that the stakeholder would not be happy to select at the time of the interim analysis.

# Sample Size determination

SARYGA
STATISTICS & METHODOLOGY

- Make some assumptions on the **control parameters** $\boldsymbol{\theta_{ji}}$
- Choose the **target treatment effects** on all efficacy endpoints, namely $\delta_1^*, \delta_2^*, \delta_3^*, \ldots$
- Choose the **maximum non inteteresting treatment effects** on all efficacy endpoints, namely $\delta_1^0, \delta_2^0, \delta_3^0, \ldots$
- Target the **information fraction of the interim analysis** on the primary endpoint (or expected information fraction if the timing is nor driven by the primary endpoint)
- Target a **nominal level** $\boldsymbol{1 - \beta}$ **for the rejection rate** under the Least Favorable Configuration (LFC)
- Find numerically the minimum sample size so that

$$F\left(Z_{S1}^{FA} > \eta\right) > 1 - \beta \qquad \text{under LFC}$$

# Design used in the simulations

- Two experimental doses, one common control arm
- 500 patients enrolled, with accrual rate 20 patients per arm per month
- Primary endpoint is OS
- For the MCDA we use:
  - Two efficacy endpoints: Overall Survival (OS) – primary endpoint
    Overall Response Rate (ORR)
  - Two safety endpoints: Serious Adverse Events (SerAE)
    Severe Adverse Events (SevAE)
- Weights $\omega = (0.1, 0.4, 0.25, 0.25)$ used in the probabilistic MCDA approach
- Interim Analysis after 183 patients overall are evaluable on ORR
- Final analysis after 179 OS events are observed across the selected arm and the control arm → 80% Power under the following Least Favorable Configuration (LFC)

| LFC | Haz | ORR | SerAE | SevAE |
|---|---|---|---|---|
| **Control** | 0.10 | 0.3 | 0.3 | 0.3 |
| **Dose 1** | 0.08 | 0.4 | 0.3 | 0.3 |
| **Dose 2** | 0.06 | 0.6 | 0.3 | 0.3 |

# Stallard & Todd (2003) – competing approach

STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:689–703 (DOI: 10.1002/sim.1362)

Sequential designs for phase III clinical trials incorporating treatment selection

Nigel Stallard*,† and Susan Todd

Medical & Pharmaceutical Statistics Research Unit, The University of Reading, P.O. Box 240, Earley Gate, Reading, Berkshire RG6 6FN, U.K.

- The best performing treatment **on the primary endpoint** is selected
- Only primary endpoint is tested at the time of the final analysis
- Critical value $\eta$ is computed in order to control type I error at the nominal level under the least favorable configuration (**expressed only in terms of primary endpoint**)

# Operating characteristics used in the simulation study

SARYGA
STATISTICS & METHODOLOGY

- Probability that **dose 2 is selected** and continued until final analysis.
  (dose 2 is always the «best dose» on the primary endpoint in the simulation setting)

- **Type I Error**: probability to reject the null hypothesis that the selected treatment is effective on the primary endpoint, when none of the active treatment is effective on the primary endpoint.

- **Power**: probability to jointly make the «right selection» and reject the null hypothesis for the selected treatment when the best treatment has a treatment effect of HR(OS)=0.6

- **Uninteresting Success**: probability to jointly make the «wrong selection» and reject the null hypothesis for the selected treatment when the best treatment has a treatment effect of HR(OS)=0.6

- **Average Toxicity**: average toxicity (on the probability scale) for the selected dose for each safety endpoint considered

# Type I Error analysis

- Dose 1 is equivalent to control
- Varying characteristics of dose 2
- Selection based on probabilistic MCDA.
- FA after **174 OS events**, IF after **183 patients on ORR**.

Characteristics of dose 2

| | Haz | ORR | Tox 1 | Tox 2 |
|---|---|---|---|---|
| S1 | 0.1 | 0.3 | 0.3 | 0.3 |
| S2 | 0.1 | 0.5 | 0.3 | 0.3 |
| S3 | 0.1 | 0.7 | 0.3 | 0.3 |
| S4 | 0.1 | 0.3 | 0.45 | 0.4 |
| S5 | 0.1 | 0.7 | 0.6 | 0.5 |

| | | | Our Approach | | Stallard & Todd | |
|---|---|---|---|---|---|---|
| | $\gamma_1$ | $\gamma_2$ | P(sel. T2) | t1e | P(sel. T2) | t1e |
| **S1** | 0 | 0 | 50.3 | 2.2 | 50.2 | 2.2 |
| **S2** | 0 | 1.56 | 94.1 | 2.2 | 49.6 | 2.5 |
| **S3** | 0 | 3.14 | 99.9 | 2.0 | 48.7 | 2.5 |
| **S4** | 0 | -1.25 | 10.6 | 2.2 | 50.7 | 2.4 |
| **S5** | 0 | 0.74 | 76.5 | 2.3 | 50.2 | 2.5 |

- Type I Error is strictly controlled below 2.5%, while it is controlled exactly at 2.5% for Stallard & Todd
- Probability of selecting dose 2 depends on $\gamma_2 - \gamma_1$

# Power analysis

- FA after **179 events** are observed (control + selected arm)
- IA after **183 responses** on secondary efficacy endpoint
- Almost **20% IF** on primary endpoint

Characteristics of control arm

| Haz | ORR | Tox1 | Tox2 |
|---|---|---|---|
| 0.1 | 0.3 | 0.3 | 0.3 |

Characteristics of dose 1

| Haz | ORR | Tox1 | Tox2 | MCDA |
|---|---|---|---|---|
| 0.08 | 0.4 | 0.3 | 0.3 | 1.23 |

Characteristics of dose 2

| | | | | | Our Approach | | | | | Stallard & Todd | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Haz | ORR | Tox1 | Tox2 | $\gamma_2$ | Avg Tox 1 | Avg Tox 2 | Sel T2 | Unint. Success | Pow | Avg Tox 1 | Avg Tox 2 | Sel T2 | Unint. Success | Pow |
| 0.06 | 0.6 | 0.3 | 0.3 | 2.83 | 0.30 | 0.30 | 95.9 | 1.7 | 79.2 | 0.30 | 0.30 | 67.9 | 11.7 | 57.7 |
| 0.06 | 0.7 | 0.3 | 0.3 | 3.64 | 0.30 | 0.30 | 99.4 | 0.2 | 82.2 | 0.30 | 0.30 | 67.7 | 12.2 | 57.4 |
| 0.06 | 0.6 | 0.5 | 0.5 | 0.92 | 0.38 | 0.38 | 39.2 | 20.9 | 33.5 | 0.44 | 0.44 | 66.9 | 12.5 | 57.3 |
| 0.06 | 0.7 | 0.55 | 0.55 | 1.29 | 0.44 | 0.44 | 54.5 | 15.9 | 46.5 | 0.48 | 0.48 | 68.3 | 12.2 | 57.9 |
| 0.06 | 0.3 | 0.4 | 0.4 | -0.5 | 0.31 | 0.31 | 16.6 | 30.7 | 3.5 | 0.37 | 0.37 | 68.1 | 11.8 | 57.9 |

✓ Power may be lower of higher wrt Stallard & Todd depending on the benefit-risk profile

✓ Arms with high level of toxicity are reccomended less on average with our approach

# Summary

SARYGA
STATISTICS & METHODOLOGY

- We constructed a seamless phase 2/3 design where:
  - Selection at IA is based on **probabilistic linear MCDA**
  - Final hypothesis testing is based on **primary endpoint**

- An **analytical expression** for the distribution of the test statistic at the final analysis is not available, but can be approximated under reasonable assumptions

- **Strong control of type I error** at the nominal level $\alpha$ is achieved computing the critical value of the final test under a «worst case configuration»

- Results show that including MCDA in the selection step of the design effectively helps in **limiting the probability to select potentially toxic doses**, still achieving good power if the most effective dose has also exhibits a good benefit-risk profile

# Thank you for the attention!