Clinical trial optimization approaches to phase III trials with multiple objectives

Gautier Paux[1], Alex Dmitrienko[2]

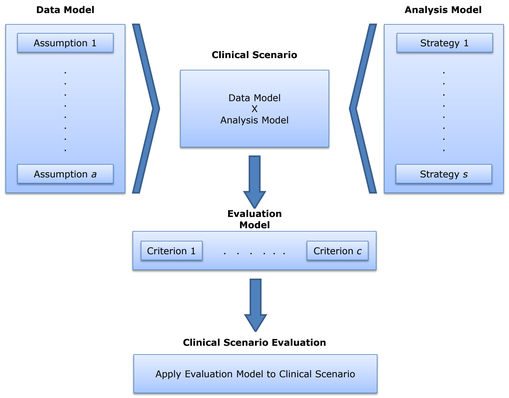
[1] Institut de Recherches Internationales (I.R.I.S.), Suresnes ,France, [2] Mediana Inc., Kansas City, USA

Problematic

- In modern drug development, clinical trial sponsors are increasingly interested in pursuing **multiple objectives** in Phase II or III clinical trials to enrich the product label
- **Sample size calculation and power evaluation** in these complex settings is more challenging and lead to high-dimensional statistical problems:
  - The familywise error rate needs to be controlled using a pre-specified efficient multiple testing procedure
  - There is not one unique definition of trial's success
  - In general, no closed-form expression is available and sample size calculations are assessed using simulation-based methods
- FDA draft guidance on Multiple Endpoints in Clinical Trials (2017) emphasized the use of clinical trial simulations to determine appropriate sample size to ensure that the study is adequately powered

Clinical Scenario Evaluation (CSE)

- The **CSE** approach (Benda et al, 2010; Friede et al, 2010) provides a structured framework for clinical trial simulations to facilitate the selection of a robust approach to clinical trial design and analysis which demonstrates optimal performance.
- The CSE framework decomposes the problem of clinical trial simulations into three components:
  - **Data model** is a structure for describing the data generation mechanism in a clinical trial (e.g., sample sizes, outcome distributions)
  - **Analysis model** is a structure for defining the analysis strategies applied to the data (e.g., statistical tests, multiplicity adjustments)
  - **Evaluation model** is a structure for specifying the measures for evaluating the performance of the analysis strategies (e.g. power)



- **CSE-based clinical trial optimization**
  - Enables clinical trial sponsors to transition from basic sample size calculations to clinical trial optimization
  - **Objective:** to quantify the impact of applicable design scenarios and analysis strategies on the selected evaluation criteria and identifying the parameter configurations that lead to **optimal performance**.
  - Key steps:
    - **Optimization criteria**: should be aligned with trial's/program's objectives
    - **Optimal parameters**: Important to assess the performance of analysis models under optimal and nearly optimal configuration of parameters
    - **Sensitivity assessment**: imperative to ensure that the conclusions are robust to reasonable deviations from the underlying statistical assumptions

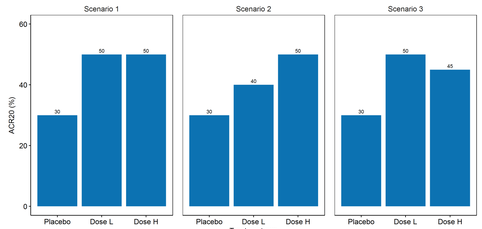
Case study outline

- Phase III trial in patients with rheumatoid arthritis
- **Design**: two doses of new treatment (Dose L and Dose H) vs. Placebo
- **Trial objective**: demonstrate that at least one dose is effective based on a single clinical endpoint
- **Primary endpoint**: ACR20 at Week 24

Case study

Data model

- Sample size is balanced between groups with 100 patients per arm
- Primary endpoint follows a binomial distribution with parameter π_i, i=0,1,2 for Placebo, Dose L and Dose H respectively



Data model scenarios

Analysis model

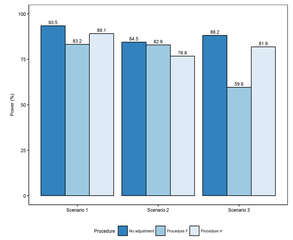
- **Tests**
  - H_1, null hypothesis of no treatment effect between Dose H vs. Placebo
  - H_2, null hypothesis of no treatment effect between Dose L vs. Placebo
  - Tested with a two-sample test for proportions
- **Multiple testing procedure (MTP)**
  - Procedure F: Fixed-sequence procedure starting with H_1
  - Procedure H: Hochberg procedure (data-driven)

Evaluation model

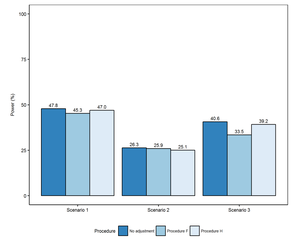
- **Disjunctive power:**
  - Probability to reject at least one null hypothesis
    $$\psi_D = P(Reject\ H_1\ or\ H_2)$$
  - Serve as a good starting point but does not take into account the number of rejected hypotheses and the relative importance of each rejected hypotheses
  - In Procedure F, $\psi_D = P(Reject\ H_1)$
- **Partition-based weighted power:**
  - Weighted power of each possible outcome
    $$\psi_{PW} = v_1 P(Reject\ H_1\ only) + v_2 P(Reject\ H_2\ only) + v_3 P(Reject\ H_1\ and\ H_2)$$
  - $v_i, i = 1,2,3$ are pre-specified non-negative values defining the relative importance of each outcome
- **Multiplicity penalty and head-to-head comparisons:**
  - Represents the amount of power loss due to the MTP, e.g. for Procedure F
    $$\psi_F = P_F(1,0) + P_F(2,0) + P_F(2,1)$$
    where $P_F(i,j)$ is the probability to reject i hypotheses with the unadjusted procedure and j hypotheses with the procedure F
  - Direct comparison based on the number of rejected hypotheses
    $$\psi_{FH} = P_{FH}(1,0) + P_{FH}(2,0) + P_{FH}(2,1)$$
    $$\psi_{HF} = P_{FH}(0,1) + P_{FH}(0,2) + P_{FH}(1,2)$$
    where $P_{FH}(i,j)$ is the probability to reject i hypotheses with the Procedure F and j hypotheses with the procedure H

Indirect comparisons

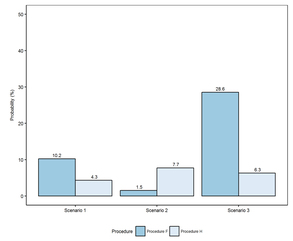
- Based on the disjunctive ψ_D, partition-based ψ_{PW} and multiplicity penalty ψ_F and ψ_H criteria



Disjunctive power ψ_D



Partition-based weighted power ψ_{PW} $v_1 = 0.15, v_2 = 0.25, v_3 = 0.60$,



Multiplicity penalty ψ_F and ψ_H

Consistent findings across criteria:

- Procedure H performs reasonably well under all pre-specified scenarios
- Procedure F is too sensitive to the shape of the true dose-response

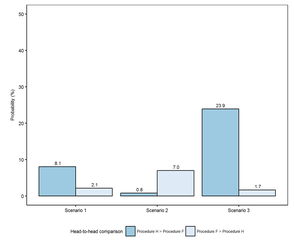
Direct comparisons

- Based on the head-to-head comparison criteria ψ_{FH} and ψ_{HF}

Number of rejected hypotheses		Procedure H		
		0	*1*	*2*
Procedure F	*0*	8.7	8.1	0.0
	1	2.1	8.2	0.0
	2	0.0	0.0	72.9

F > H	F = H	F < H

Head-to-head comparison matrix in Scenario 1

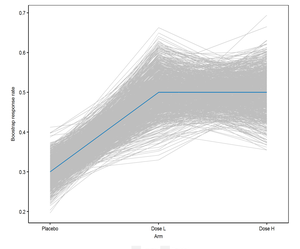


Head-to-head comparisons ψ_{FH} and ψ_{HF}

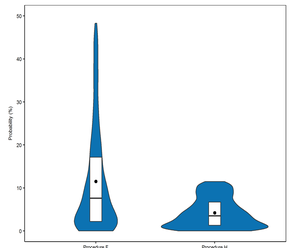
- **Scenario 1**: if a plateau response is observed, Procedure H tends to perform better than Procedure F ($\psi_{FH} = 2.1\%$ vs. $\psi_{HF} = 8.1\%$)
- **Scenario 2**: in positive linear dose-response assumption, Procedure F is superior to Procedure H ($\psi_{FH} = 7.0\%$ vs. $\psi_{HF} = 0.8\%$)
- **Scenario 3**: in non-monotonic dose-response relationship, substantially higher probability that Procedure H is superior to Procedure F ($\psi_{FH} = 23.9\%$ vs. $\psi_{HF} = 1.7\%$)

Quantitative sensitivity assessment

- **Key features:**
  - How robust are the MTPs when the dose-response curve is slightly different from the flat dose-response function under Scenario 1 ?
  - Quantitative deviations from the main data model by randomly perturbing the dose-response curve
- **Parametric bootstrap approach:**
  - Let $\pi = (\pi_1, \pi_2, \pi_3)$ be the assumed dose-response function
  - **Step 1**: Generate k "true" dose-response functions, i.e., $\pi_i^* = (\pi_{i1}^*, \pi_{i2}^*, \pi_{i3}^*)$, $i = 1, \dots, k$, with $\pi_{ij}^* \sim Beta(\alpha_j, \beta_j)$, $j = 1,2,3$ where α_j and β_j are determined as follows, with c being the CV:
    $$\begin{cases} \pi_j = \dfrac{\alpha_j}{\alpha_j + \beta_j} \\ c^2 = \dfrac{\beta_j}{\alpha_j(\alpha_j + \beta_j + 1)} \end{cases}$$
  - **Step 2**: Compute the criteria for each true dose-response function (10,000 simulation runs)
  - **Step 3**: Summarize the criteria over k dose-response function
- **Results ($c = 0.1$):**



Simulated dose-response in Scenario 1



Empiric distribution of ψ_F and ψ_H

- Procedure F is quite sensitive to quantitative deviations from the assumed dose-response function

Optimization

- **Key features:**
  - An equally weighted Hochberg procedure has been considered
  - Different hypothesis weights could have been specified, e.g., w_1 and w_2 for H_1 and H_2 respectively, where $w_1 + w_2 = 1$
  - Optimal weights can be determined using a grid search for w_1 from 0 to 1, e.g., $w_1^*(i) = \underset{w_1 \in (0,1)}{\operatorname{argmin}}\, \psi_H(w_1)$ is the optimal weight for Scenario i
  - $I_\eta(i) = \left\{w_1 \in (0,1) : \psi_H(w_1) \le \frac{1}{\eta} * \psi_H\left(w_1^*(i)\right)\right\}$ is the 100η% optimal interval for $w_1^*(i)$ $(0 < \eta < 1)$
  - An optimal value across all scenarios can be determined using a joint optimal interval $I_\eta = \cap_i I_\eta(i)$

Mediana R package

- The **Mediana R package** is based on the CSE framework and
  - Provides a standardized framework for clinical trial simulations
  - Supports a broad class of data, analysis and evaluation models
  - Provides a flexible framework easily extensible to define custom options in data, analysis and evaluation models
  - High performance computing with simulation runs distributed among multiple processor cores
  - Word-based reporting with CSE summary and simulation results
- Web-based manual with multiple case studies implementation accessible at http://gpaux.github.io/Mediana
- Extensively used in a recently released book: **Clinical Trial Optimization Using R**, edited by A. Dmitrienko and E. Pulkstenis.

Conclusion

- The CSE approach is an efficient framework that enables clinical trials optimization with complex multiple objectives
- It is critical to choose adequate optimization criteria that lead to relevant choice of optimal multiple testing procedure
- The robustness of optimal procedure's performance should be assessed by randomly perturbing the statistical assumptions
- The Mediana R package provides an turnkey solution to facilitate a systematic quantitative assessment of performance in clinical trials with multiplicity issues
- More complex case studies are presented in Dmitrienko, A. and Pulkstenis, E. (2017), including clinical trials with several sources of multiplicity

References

- Benda, N. *et al.* (2010). Aspects of modernizing drug development using clinical scenario planning and evaluation. Drug Information Journal. 44, 299–315.
- Dmitrienko, A. *et al.*. (2016). Power calculations in clinical trials with complex clinical objectives. Journal of the Japanese Society of Computational Statistics. 28, 15–50.
- Dmitrienko, A. and Pulkstenis, E. (2017). Clinical Trial Optimization Using R. New-York : CRC Press.
- Friede, T. *et al.*. (2010). Refinement of the clinical scenario evaluation framework for assessment of competing development strategies with an application to multiple sclerosis. Drug Information Journal 44 :713–718.