Cytel

A confirmatory adaptive Phase II/III design with Bayesian decision rules for dose selection and sample size re-estimation – a case study

Jan Priel

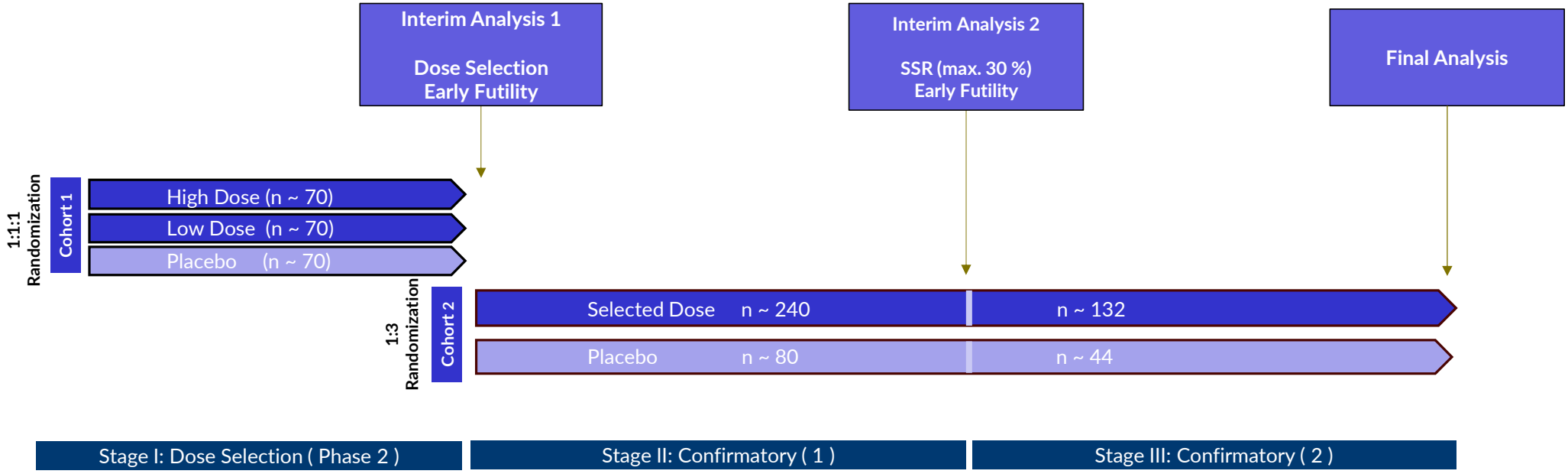
Statistical Strategic Consulting
Therapeutics Development Team

25 Oct 2023

Background Information

Indication	Radicular Leg Pain: Pain that radiates through the spine from the back and hips into the legs.
Experimental Treatment	Epidural steroid injection (ESI). Non-opioid option for treatment.
Prior Information	
Phase 2	Small randomized Phase 2 trial comparing two doses with Placebo. Both doses seem to have similar safety profiles. Efficacy was promising in both doses. Better effect in the higher dose. Efficacy data was re-analysed using Bayesian models to inform scenarios and simulation assumptions.
Project/Client requirements for the Phase II/III design	
Dose Selection	Start with two doses and drop one dose at the first interim analysis.
Early Futility	Non-binding early futility at interim analysis in case no dose shows signs of efficacy.
Interim Sample Size Increase	Uncertainty about treatment effects and standard deviation. Opportunity to increase sample size if necessary.
Regulatory Acceptability	Among other, there should be an established method to control the overall type I error.

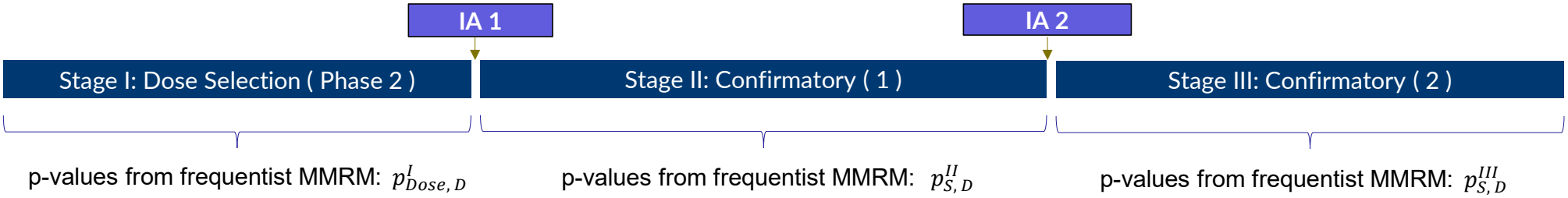
Outline of Design



Two Primary Endpoints: Change from Baseline of (7-days mean) Worst Daily Leg Pain (WDLP) after 60 Days and after 90 Days.
Final Analysis: Stagewise (Frequentist) Mixed Model for Repeated Measures adjusting for Baseline WDLP and Baseline Pain Duration.

Final Analysis - Stagewise Testing

Claim success for selected dose $S \in \{High\ Dose,\ Low\ Dose\}$ and Day $\mathrm{D} \in \{Day\ 60,\ Day\ 90\}$ if both $H_{S,D}$ and $H_{\cap,D} = H_{High\ Dose,D} \cap H_{Low\ Dose,\ D}$ can be rejected at $\frac{\alpha}{2}$.



Combination p-value to reject $H_{S,D}$: $p^{comb}_{S,D} = \phi(\omega_I \phi^{-1}(p^{I}_{S,D}) + \omega_{II} \phi^{-1}(p^{II}_{S,D}) + \omega_{III} \phi^{-1}(p^{III}_{S,D}))$

Combination p-value to reject $H_{\cap,D}$: $p^{comb}_{\cap,D} = \phi(\omega_I \phi^{-1}(p^{I}_{\cap,D}) + \omega_{II} \phi^{-1}(p^{II}_{S,D}) + \omega_{III} \phi^{-1}(p^{III}_{S,D}))$

For stage I intersection Hypothesis we used the Simes adjustment:
$p^{I}_{\cap,D} = \min(2 * \min(p^{I}_{Low\ Dose,D}, p^{I}_{High\ Dose,D}), \max(p^{I}_{Low\ Dose,D}, p^{I}_{High\ Dose,D}))$

Pre-specified weights:
$\omega_I^2 + \omega_{II}^2 + \omega_{III}^2 = 1$

1. Posch et al (2005), Testing and estimation in flexible group sequential designs with adaptive treatment selection. Statistics in Medicine; 24: 3697-3714.

Interim Analysis 1 – Dose Selection

Posterior Probabilities from Bayesian MMRM:

$P_{Low\ Dose,\ 60\ Days}$:= P(Difference between Low Dose and Placebo after 60 Days < 0 | Data at IA1)
$P_{Low\ Dose,\ 90\ Days}$:= P(Difference between Low Dose and Placebo after 90 Days < 0 | Data at IA1)
$P_{High\ Dose,\ 60\ Days}$:= P(Difference between High Dose and Placebo after 60 Days < 0 | Data at IA1)
$P_{High\ Dose,\ 90\ Days}$:= P(Difference between High Dose and Placebo after 90 Days < 0 | Data at IA1)

Dose Selection Rule:

Start with Day 60: If $|\boldsymbol{P_{High\ Dose,\ 60\ Days} - P_{Low\ Dose,\ 60\ Days}}| > X$
→ Select Dose with higher posterior probability for Day 60

else look at Day 90: If $|\boldsymbol{P_{High\ Dose,\ 90\ Days} - P_{Low\ Dose,\ 90\ Days}}| > X$
→ Select Dose with higher posterior probability for Day 90

else: →Select the High Dose.

Higher X values increase selection preference for the High Dose. X ~ 10% calibrated by simulations.

Interim Analysis 1 – Futility Rule

80% Bayesian Credible Intervals from Bayesian MMRM for:

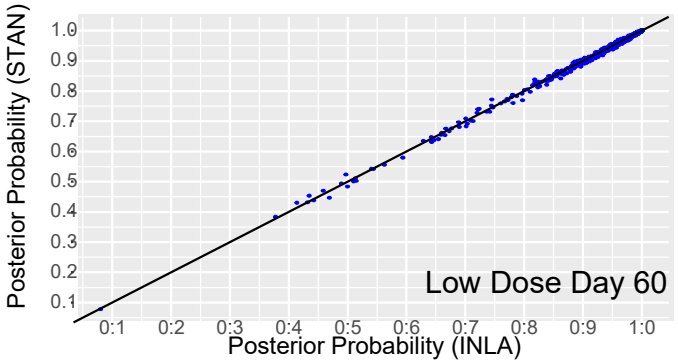
- Difference between Low Dose and Placebo after 60 Days
- Difference between Low Dose and Placebo after 90 Days
- Difference between High Dose and Placebo after 60 Days
- Difference between High Dose and Placebo after 90 Days

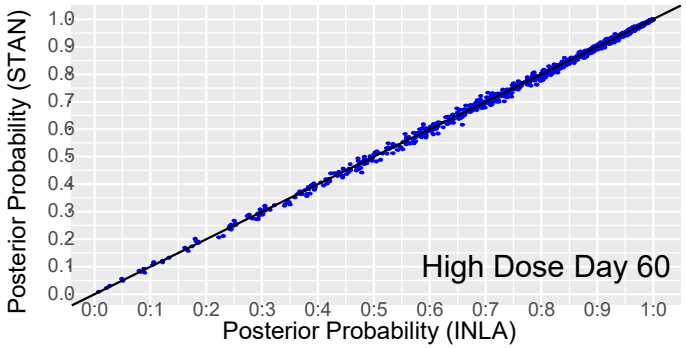
Futility Rule:

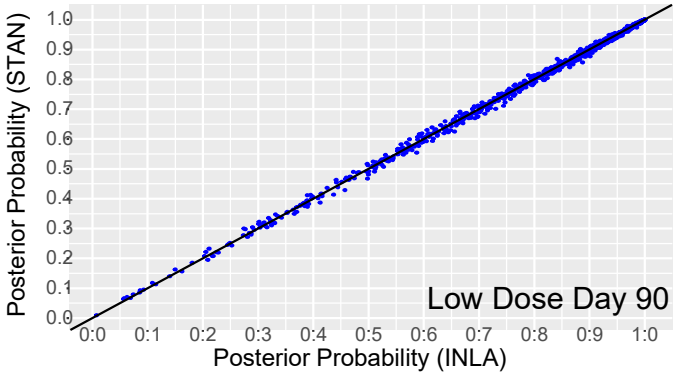
- If all four intervals exclude the difference of -1.0 (lower bounds higher than -1.0)

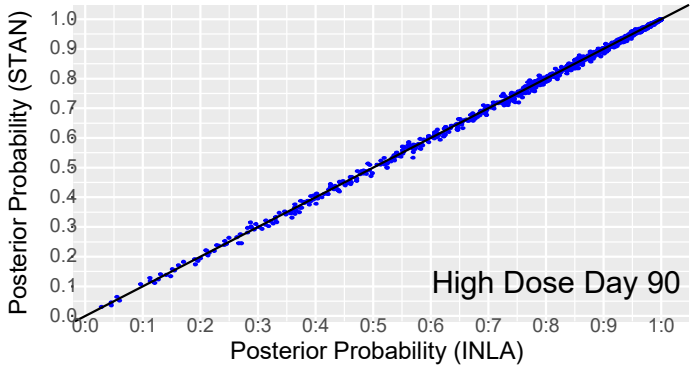
  → Recommend stopping for Futility (non-binding)

Posterior Probabilities – MCMC vs. INLA









- Posterior probability (dose selection) from R-INLA matched those from RStan
- In our case, INLA lead to ~ 10 times faster simulations for the dose selection part.
- Full posterior sampling implemented in R-INLA which can then be used for the PPoS calculation.

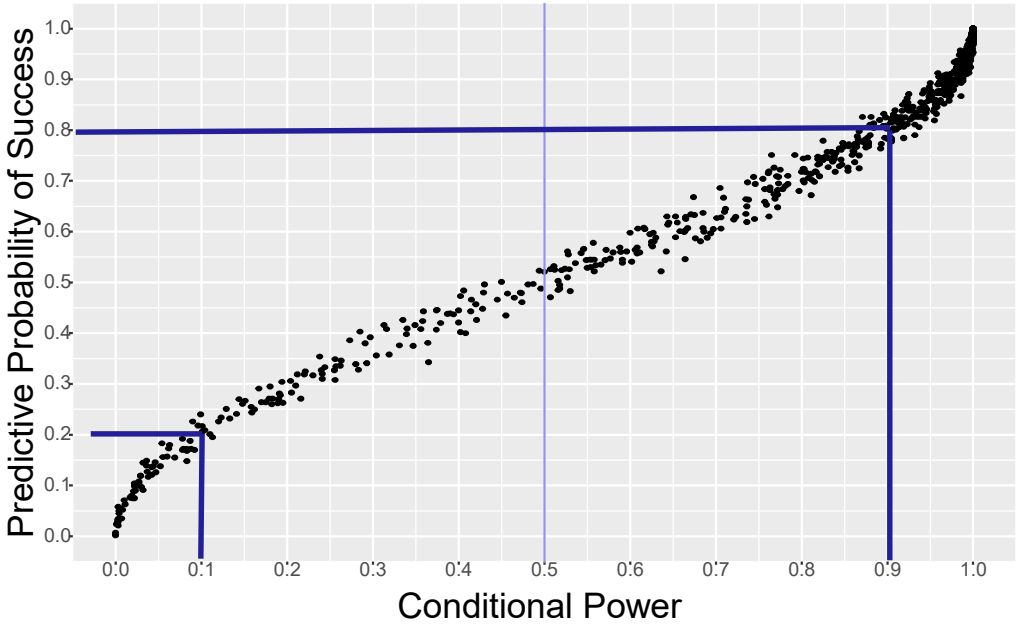
Interim Analysis 2 - Zones

Interim Analysis 2 - Zones				
Day 60 / Day 90	**Futility** PPoS < 10%	**Unfavorable** 10% ≤ PPoS < 50%	**Promising** 50% ≤ PPoS < 90%	**Favorable** PPoS ≥ 90%
Futility PPoS < 10%	**Stop**	**Continue**	**Re-estimate for Day 60**	**Continue**
Unfavorable 10% ≤ PPoS < 50%	**Stop**	**Continue**	**Re-estimate for Day 60**	**Continue**
Promising 50% ≤ PPoS < 90%	**Stop**	**Re-estimate for Day 90**	**Re-estimate for both Days**	**Continue**
Favorable PPoS ≥ 90%	**Stop**	**Continue**	**Re-estimate for Day 60**	**Continue**

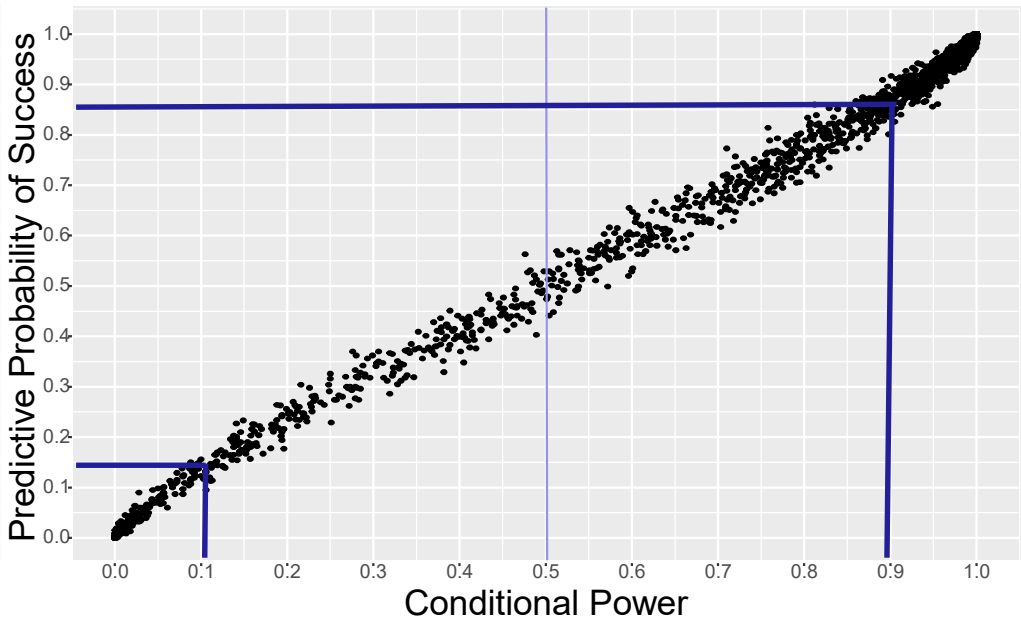
PPoS = Predictive Probability of Success (of selected Dose).

Predictive Probability of Success vs. Conditional Power

Example scenario:
Sample Size at IA 2 ~ **310** with Final Sample Size ~ 706



Example scenario:
Sample Size at IA 2 ~ **530** with Final Sample Size ~ 706



Main Operating Characteristics

Scenario	Treatment Effect				Selection Rate Low Dose (%)	Selection Rate High Dose (%)	Success Rate Low Dose (%)			Success Rate High Dose (%)			Success Rate (Any) (%)
	Low Dose Day 60	Low Dose Day 90	High Dose Day 60	High Dose Day 90			Day 60	Day 90	Any	Day 60	Day 90	Any	
Null	0	0	0	0	43.1	56.9	0.4	0.3	0.5	0.8	1.1	1.5	2.0
Base 1	-0.5	-0.5	-1.0	-0.5	16.1	83.9	10.4	8.2	11.5	79.8	45.4	79.9	91.4
Base 2	-0.5	-0.5	-1.0	-1.0	3.6	96.4	2.0	1.8	2.1	93.3	92.6	93.8	95.9
Base 3	-1.0	-0.5	-0.5	-0.5	49.5	50.5	45.9	19.4	45.9	35.7	29.8	38.4	84.3
Base 4	-1.0	-1.0	-0.5	-0.5	58.4	41.6	55.7	55.6	56.1	29.7	28.9	33.2	89.3
Low Dose Selection 1	-1.5	-0.5	-0.5	-0.5	55.4	44.6	54.8	23.1	54.8	34.4	29.5	36.8	91.6
Low Dose Selection 2	-1.5	-1.5	-0.5	-0.5	65.6	34.4	65.5	65.5	65.5	26.1	23.7	28.3	93.8

1) In the Null scenario, since the interim futility rules are non-binding, all success rates are calculated ignoring early futility.
2) In these scenarios, a standard deviation of 2.5 was assumed.

Additional Operating Characteristics

Scenario	Treatment Effect				Interim Analysis 1	Interim Analysis 2				Final Analysis
	Low Dose Day 60	Low Dose Day 90	High Dose Day 60	High Dose Day 90	Futility (%)	Futility (%)	Unfavorable (%)	Promising (%)	Favorable (%)	Success Rate (Any) (%)
Null	0	0	0	0	64.3	28.8	4.2	1.9	0.8	2.0
Base 1	-0.5	-0.5	-1.0	-0.5	5.8	2.5	3.9	10.1	77.8	91.4
Base 2	-0.5	-0.5	-1.0	-1.0	2.5	1.3	2.2	8.4	85.6	95.9
Base 3	-1.0	-0.5	-0.5	-0.5	5.1	5.8	7.1	21.3	60.7	84.3
Base 4	-1.0	-1.0	-0.5	-0.5	2.5	4.9	5.5	17.4	69.7	89.3
Low Dose Selection 1	-1.5	-0.5	-0.5	-0.5	0.6	3.2	7.2	14.1	74.9	91.6
Low Dose Selection 2	-1.5	-1.5	-0.5	-0.5	0.1	3.3	3.4	10.1	83.1	93.8

Gain in power with SSR: With the budget constraint to increase the sample size by max. 30%, the gain in power in the promising zone was approx. 5% to 10% in the scenarios of interest.

Summary

- Designs with stagewise hypothesis testing are attractive due to their flexibility while simultaneously controlling the overall type I error.
- The stagewise testing ensures overall type I error control even if rules other than the pre-specified rules are used.
- The price for this flexibility in stagewise testing is typically a loss of power.
- Interim decision rules can be based on Bayesian methods such as posterior and predictive probabilities coming from Bayesian models.
- Fast estimations of powerful Bayesian models can be performed with R-INLA to speed up simulations.

Thank you!