

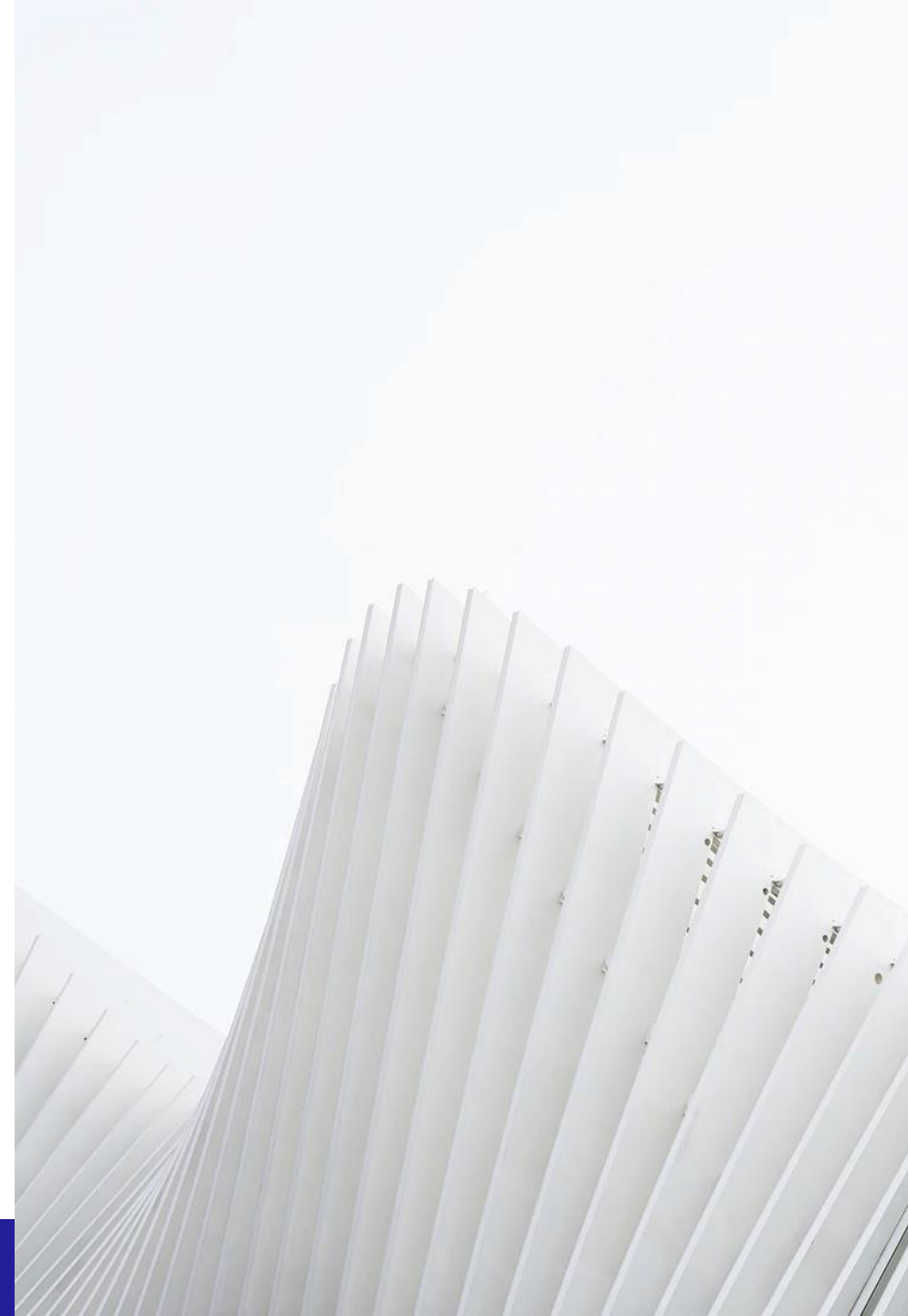


Simulation-based optimization of adaptive designs

Pantelis Vlachos, Cytel Inc, Geneva
Pantelis.Vlachos@Cytel.com

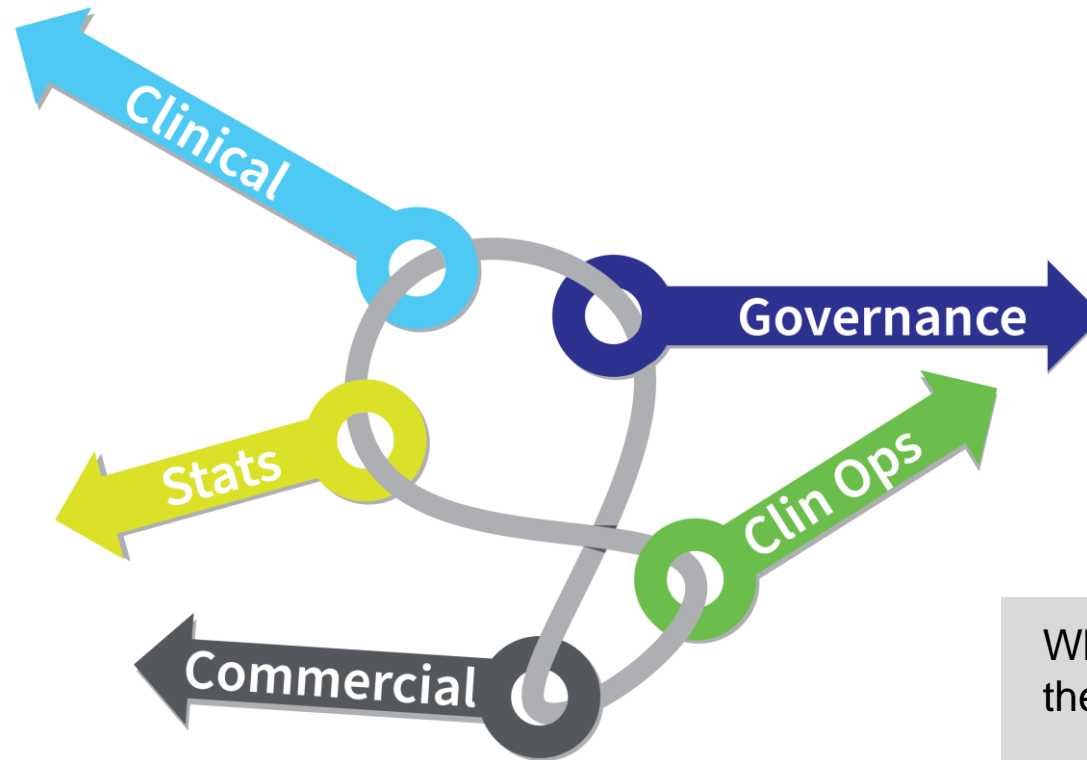
Agenda

- Motivation
- What is success in a clinical trial?
- What if priorities change?
- A case study
- Questions



How Do I Plan the Right Clinical Trial?

- What is the right patient population to treat?
- Is my medicine safe and effective?
- How many patients per arm do we need?
- Can I make a clear decision from the data?
- Will my endpoint be reimbursed?
- What is my benefit-risk profile?
- Is my asset better than the competition?
- How can we go faster?



What clinical trial design will get the right answers?

How can we accomplish these goals quickly and economically?

Need to first define success

- Team input:
 - Endpoints
 - Power
 - Budget
 - Design
 - Decision Criteria
 - **Priorities**

Calculating Probability of Success (for example in EAST)

Clinical Study Description and Fixed Design Requirements

Phase III multicenter, randomized, placebo-controlled, parallel-arm clinical trial to evaluate the efficacy of Treatment versus Control in an acute Myeloid Leukemia study

Endpoint: Median OS

- Control median OS: 8 months
- Treatment effect: HR = 0.7
- Enrollment rate: 20 patients/month
- Sample Size: 451, Events: 331
- Power: 90%
- One-sided alpha: 2.5%

Design Operating Characteristics

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Test Parameters	
Design ID	fixed0.7-20subjs
Design Type	Superiority
Number of Looks	1
Test Type	1-Sided
Specified α	0.025
Power	0.90053
Model Parameters	
HR = λ_t/λ_c	
Under H0	1
Under H1	0.7
Med. Surv. Time Control (m_c)	8
Med. Surv. Time Treatment (m_t)	11.429
Var (Log HR)	Null
Allocation Ratio (n_t/n_c)	1
Accrual / Dropouts Parameters	
Accrual Rate	20
Dropout	No

Variable Follow-Up Design: All subjects are followed until failure, drop out or end of study.

Sample sizes and events have been rounded.

Sample Size Information

Sample Size (n)	451
Treatment (n_t)	226
Control (n_c)	225
Events (s)	331
Treatment (s_t)	153
Control (s_c)	178
Information (I)	82.75

Accrual and Study Duration

Accrual Duration	22.55
Max. Study Duration	31.145

Critical Points

Critical Point	-1.96
----------------	-------

Adding uncertainty in Treatment effect

HR	Pr(HR)
0.60	10%
0.65	15%
0.70	30%
0.75	20%
0.80	15%
1.0	10%

Question: What is the study probability of success if we are only partially certain about the true HR being 0.7

Assurance Calculation

- $PoS = \sum_x P(\text{reject } H_0 | HR = x)P(HR = x)$
- In EAST:

Design Type: Superiority Number of Looks: 1

Test Parameters Accrual / Dropouts

Type I Error (α): 0.025

Power: 0.9

No. of Events: 331

Allocation Ratio: 1
(n_t/n_c)

Hazard Ratio (Optional) Alternative

Hazard Ratio (λ_t/λ_c) 0.7

Ratio of Medians (m_t/m_c) 1.429

Med.Surv.Time	
Control	Treatment: Alt.
8	11.429

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success) 0.7466

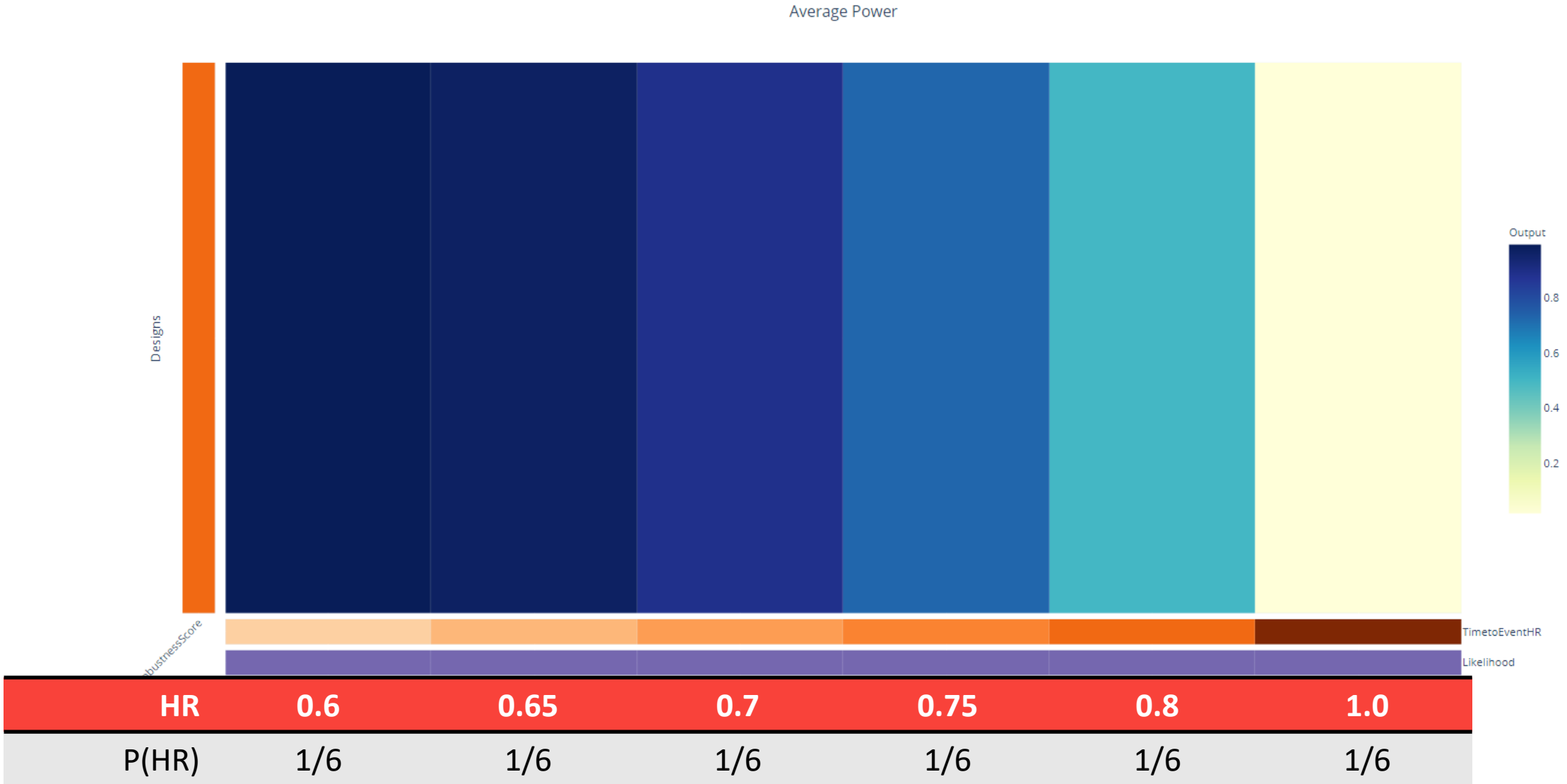
Prior Distribution for: Log Hazard Ratio (δ) Distribution: User Specified-R

File Information for δ

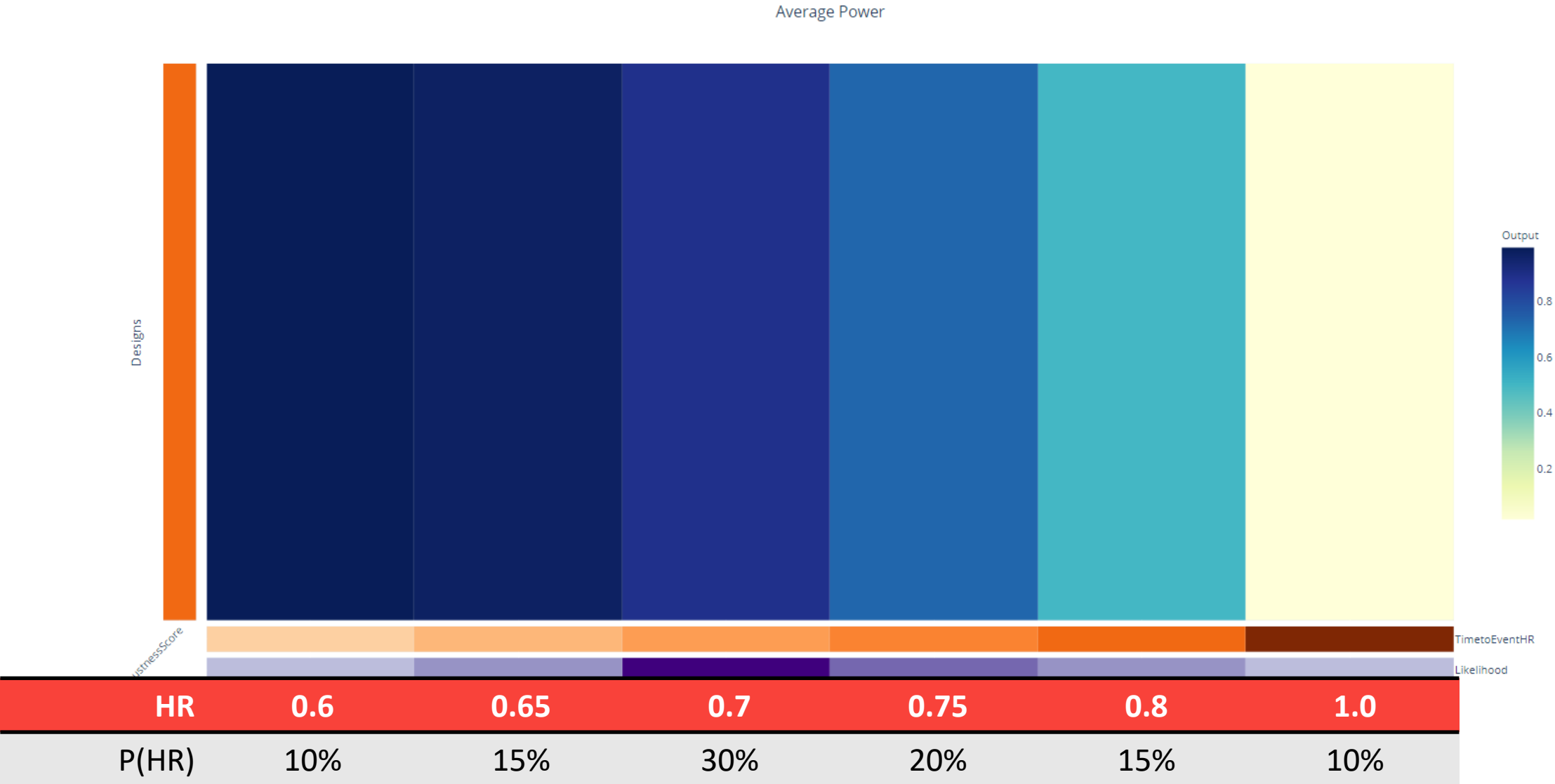
R File: C:\Users\Pantelis.vlachos\Desktop\ Browse...

R Function: HR View...

An Alternative display – flat prior on HR – assurance = 0.68



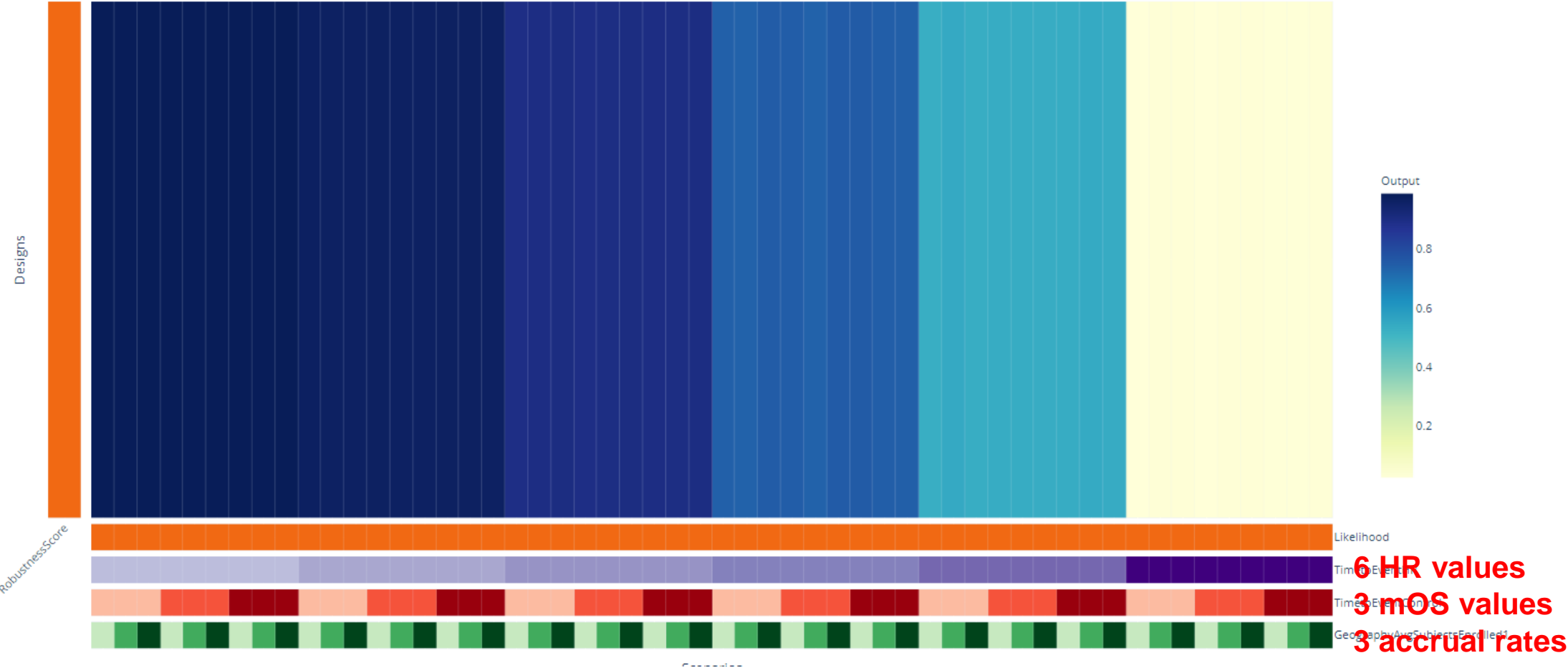
Prior used in EAST example – assurance = 0.74



What if we are also uncertain about control mOS and Accrual

Equal prior weights (1/54) – Flat prior – PoS = 0.69

Average Power



54 scenarios

Assuming informative prior HR and flat prior for Ctrl mOS, Accruals – PoS = 0.75

Fixed Sample ♡ ✕

Superiority 2-Arms / Score: 0.91 / Robustness Score: 0.751 / Likelihood: 0.033 / Followup Time: Until End of Study

Summary ⬆

Designs Scenarios Financials

Number of Interim Analyses	NA
Events	330
Sample Size	451
Allocation Ratio	1
Test Statistic	Logrank
Type 1 Error	0.025
Critical HR	0.806

Output

Avg. Study Duration	30.984 Months
Power	90.2%
Avg. Sample Size	451
Avg. Number of Events	330
Avg. Accrual Duration	22.498 Months
Observed HR	0.702
Follow-up Time	10.117 Months

Average Sample Size, Dropouts and Analysis Times ⬇

Simulation Boundaries and Boundary Crossing Probabilities ⬇

Expanding from Fixed to Adaptive Designs

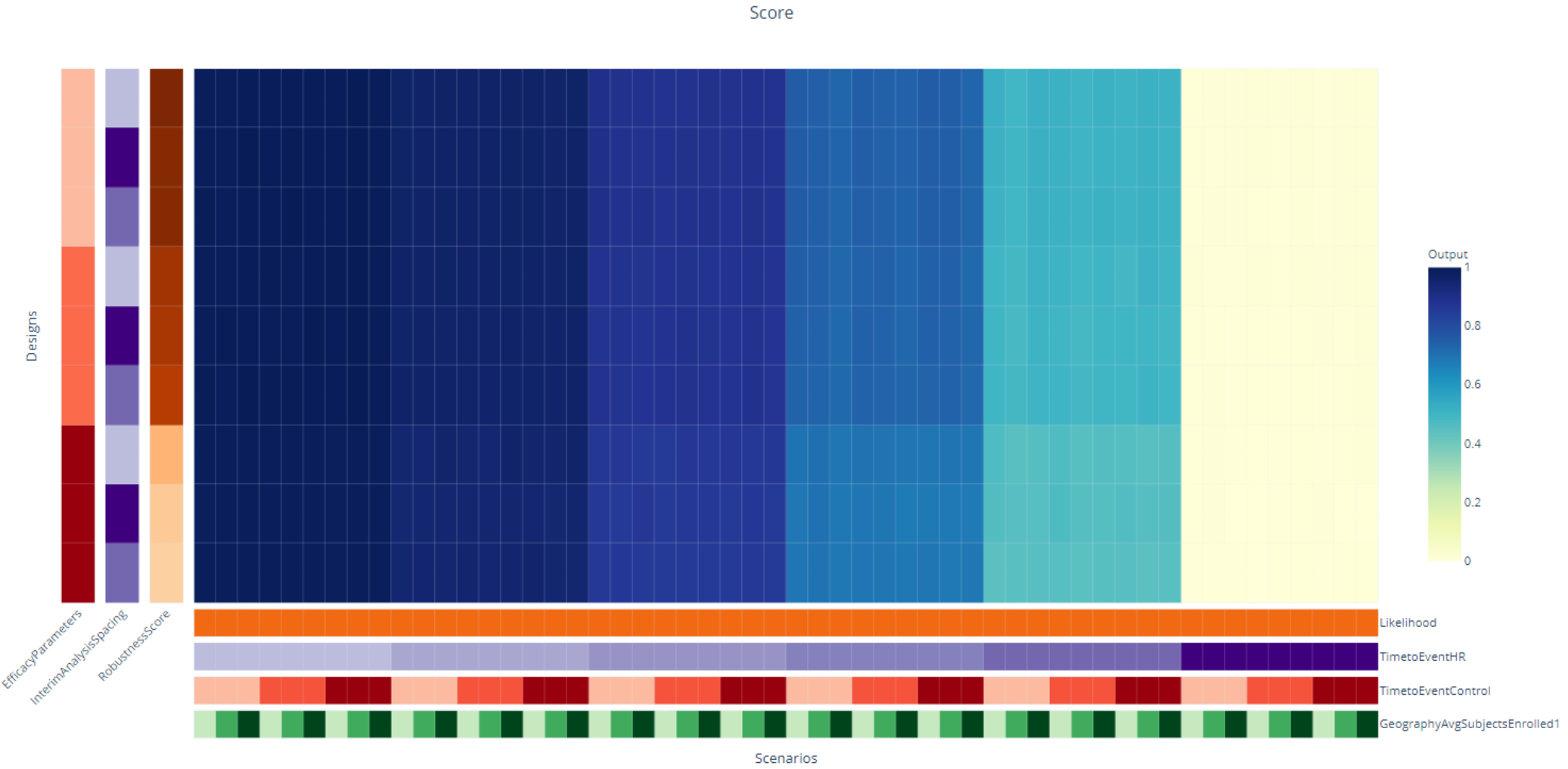
Clinical Study Description and Fixed Design Requirements

Phase III multicenter, randomized, placebo-controlled, parallel-arm clinical trial to evaluate the efficacy of Treatment versus Control in an acute Myeloid Leukemia study

Endpoint: Median OS

- Control median OS: 8 months
- Treatment effect: HR = 0.7
- Enrollment rate: 20 patients/month
- 1 Interim Analysis for Efficacy at either **40%, 50% or 60%** IF
- Alpha-spending according to Gamma rule **(-4,-2,1)**
- Sample Size: 451, Events: 331
- Power: 90%
- One-sided alpha: 2.5%

Using flat prior on unknowns (HR, Ctrl mOS, Accrual) we now have 1 PoS calculation for each possible design



Probability of Success of each design, flat priors

Fixed	IF	GSD								
		40			50			60		
	gamma	-4	-2	1	-4	-2	1	-4	-2	1
68.3%	Probability of Success	68.8%	68.3%	66.3%	68.7%	68.1%	65.9%	68.7%	68.2%	66.0%

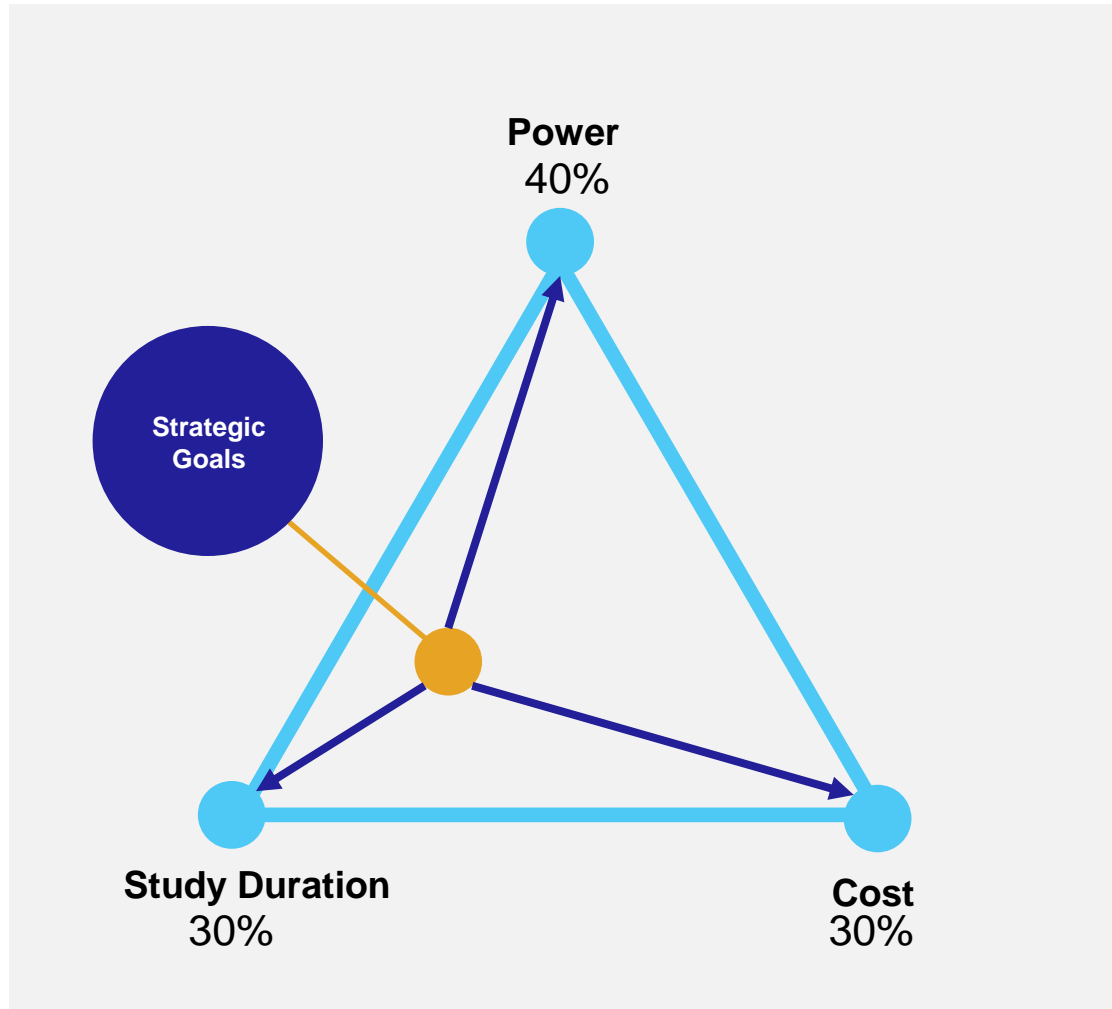
Probability of Success of each design, informative prior for HR, flat prior for Ctrl mOS and Accrual

Fixed	IF	GSD								
		40			50			60		
	gamma	-4	-2	1	-4	-2	1	-4	-2	1
68.3 %	Probability of Success (equal weights)	68.8%	68.3%	66.3%	68.7%	68.1%	65.9%	68.7%	68.2%	66.0%
73.5 %	Probability of Success (unequal weights)	73.9%	73.4%	71.1%	73.8%	73.2%	70.7%	73.9%	73.2%	70.8%

Recap

- We started with $PoS = \sum_x P(\text{reject } H_0 | HR = x)P(HR = x)$
- We defined a scenario as $\{HR = x, mOS_C = y, r_{acc} = z\}$ and
- arrived at $PoS = \sum_x P(\text{reject } H_0 | \text{Scenario} = s)P(\text{Scenario} = s)$

What if our priorities extend beyond maximizing PoS?



Models can be scored on performance criteria that reflect strategic goals

The score is a weighted function of performance criteria

$$w_P (Power - P_{min}) / (P_{max} - P_{min}) \\ + w_T (T_{max} - Time) / (T_{max} - T_{min}) \\ + w_C (C_{max} - Cost) / (C_{max} - C_{min})$$

Selecting general design-agnostic criteria enable broad strategic comparisons

Scoring is meant to surface areas of interest in the design map that merit further exploration

Robustness score of each design, informative prior for HR, flat prior for Ctrl mOS and Accrual

Fixed	IF	GSD								
		40			50			60		
	gamma	-4	-2	1	-4	-2	1	-4	-2	1
	Robustness (equal weights)	44.9%	49.1%	54.2%	45.9%	48.9%	51.7%	44.8%	46.3%	47.5%
	Robustness (unequal weights)	46.1%	50.6%	56.1%	47.5%	50.8%	54.0%	46.8%	48.6%	49.9%

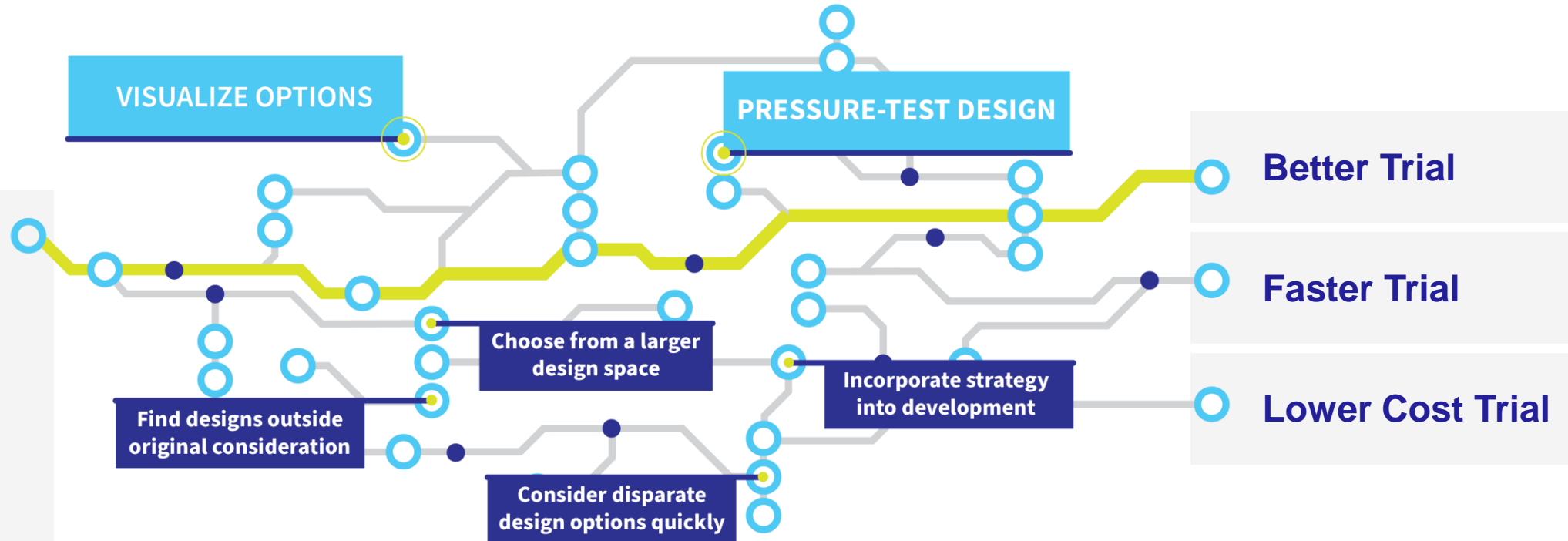
Score = 40%*Power + 30%*Duration + 30%*Sample Size

Find the Right Path for Your Study

TRIAL DESIGN SIMPLIFIED AND SCALED

Define Success

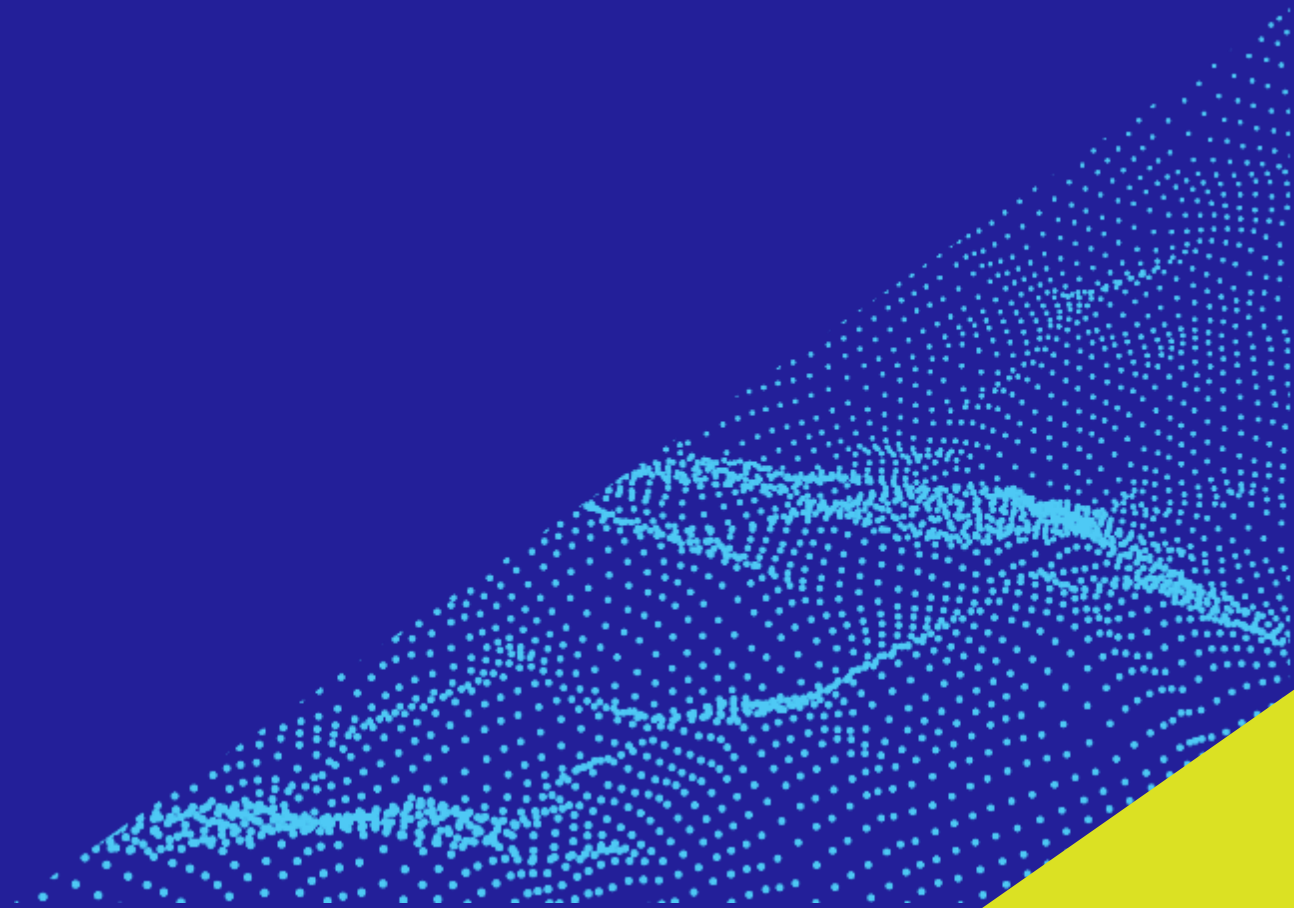
Team input:
Endpoints
Power
Budget
Design
Decision Criteria
Priorities



ACCELERATE TO VALUE



A case study in Multiple Myeloma



Multiple Myeloma Ph 3 Study

Reference Design	Inputs
Planned Sample Size	800
Planned Number of Events	227
Allocation Ratio	1:1
Targeted Treatment Effect (HR)	0.65
Control Median Survival Time	20 months
Type-1 error (1-sided)	0.025
Target Power	85%
Number of Interim Analyses	1
Timing of Interim Analysis	70%
Efficacy Stopping Rule	LD-OBF
Futility Stopping Rule	LD-OBF

Primary Outcome:

Progression Free Survival

Optimization Aim:

Maintain adequate power while minimizing time to market

Questions of interest:

- What is an optimal design that accounts for uncertainty on patient recruitment?
- How will treatment effect variations impact the trial?
- What study design would most optimize cost/sample size?

Cytel Simulation Plan Template

Design Options

Type 1 error: 1 sided 0.025

Allocation Ratios: 1:1

Number of subjects: 700:800:20

Number of events (if TTE): 130,162, 182, 210, 227, 263

Statistical Design: GSD, GSD with SSR

Number of interim analyses: 11A

Timing of interim analyses: 65%, 70%, 75%

Efficacy Stopping Rules/Alpha Spending Function: OBF

Futility Stopping Rules/Beta Spending Function: OBF, none

Promising Zone (if applicable): min = 0.3, max = 0.8, 0.9

Target Conditional Power (if applicable): 90%, 99%

Max Number of Subjects/Events (if applicable): 1.2, 1.3, 1.4

Population Scenarios

True underlying control response rates: 20m PFS (vary?)

True underlying treatment effects: 0.60, 0.65, 0.67

Dropout rate: 0

Enrollment Patterns

Enrollment Rates: (Number of periods, starting at time, average enrollment rate)

20pts/mo, 25pts/mo, 30pts/mo

Average Cost per Patient

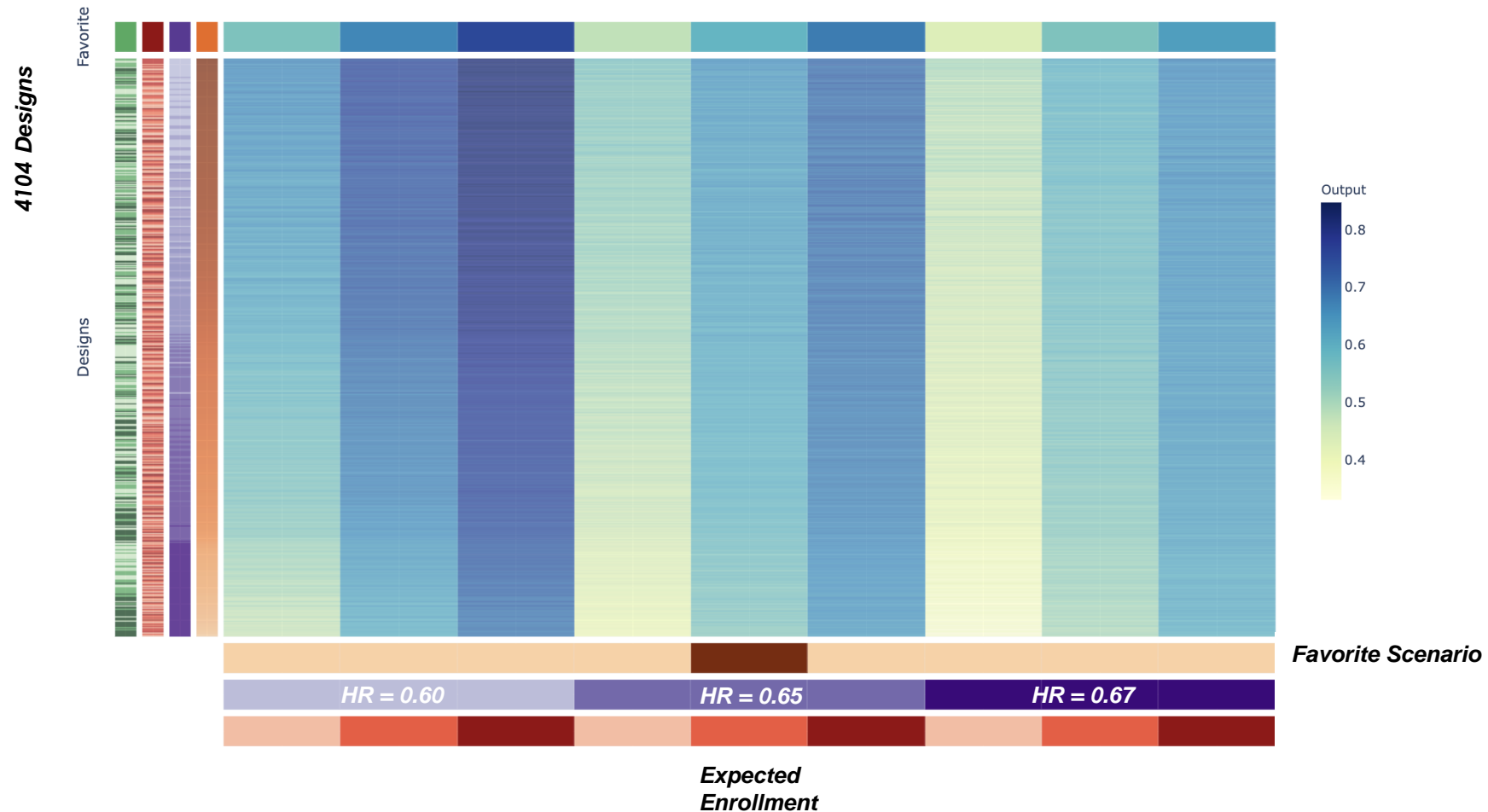
\$100,000

Total number of design options in combination with scenarios (i.e., Models) =
4,104 designs x 9 scenarios = 36,936 models

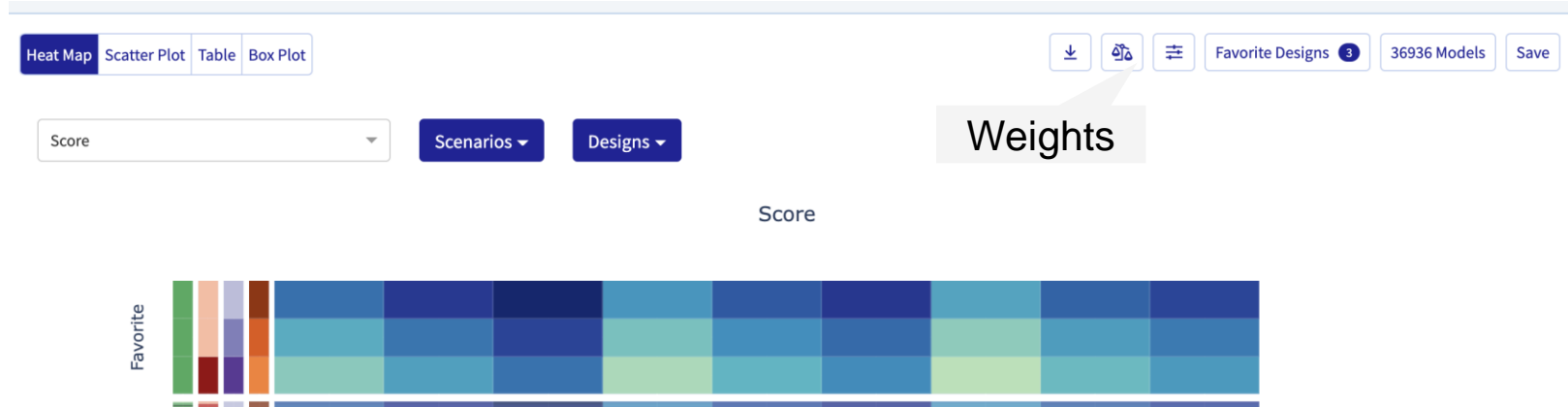
Multiple Myeloma Study

~37 Million Simulated Trials

9 Scenarios



Update weights on score and scenarios



Weights

Weights must add up to 100%.

	Weight (%)	Threshold
Maximize Probability of Winning	0	0.85
Minimize Sample Size	10	
Minimize Study Duration	90	

Apply Cancel Reset All

Weights

Weights must add up to 100%.

Set	Weight (%)
<input checked="" type="radio"/> Response Set 1	100

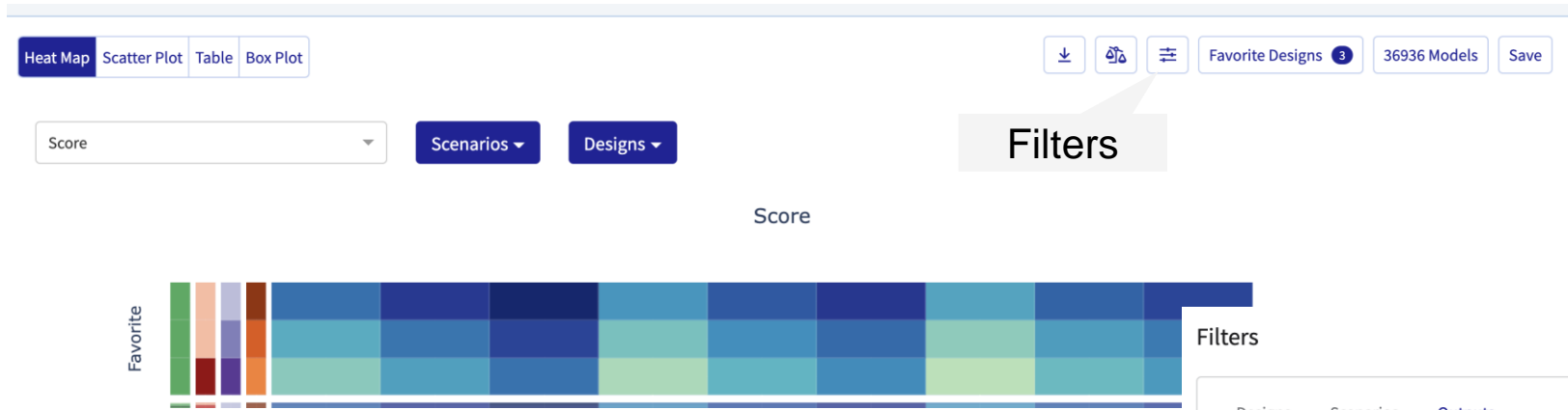
Probability Dropout Control	Weight (%)
0	100

Probability Dropout Treatment	Weight (%)
0	100

Time to Event Control	Weight (%)
20	100

Time to Event HR	Weight (%)
0.6	33.33
0.65	33.33
0.67	33.33

Filter on Scenarios/Designs/Outputs of Interest



Filters

Designs **Scenarios** Outputs

Avg Subjects Enrolled (Geography1)

Time to Event HR

Designs Scenarios **Outputs**

Avg Accrual Duration	<input type="text" value="16.559"/>	<input type="range" value="16.559 33.466"/>	<input type="text" value="33.466"/>
Avg Number of Events	<input type="text" value="102.544"/>	<input type="range" value="102.544 246.571"/>	<input type="text" value="246.571"/>
Power	<input type="text" value=".85"/>	<input type="range" value="0.523 0.997"/>	<input type="text" value="0.997"/>
Avg Sample Size	<input type="text" value="415.114"/>	<input type="range" value="415.114 787.244"/>	<input type="text" value="787.244"/>
Avg Study Duration	<input type="text" value="16.594"/>	<input type="range" value="16.594 33.566"/>	<input type="text" value="33.566"/>
Avg Follow Up Time	<input type="text" value="7.109"/>	<input type="range" value="7.109 12.546"/>	<input type="text" value="12.546"/>
Observed HR	<input type="text" value="0.582"/>	<input type="range" value="0.582 0.695"/>	<input type="text" value="0.695"/>

Test Designs

Current Scenario

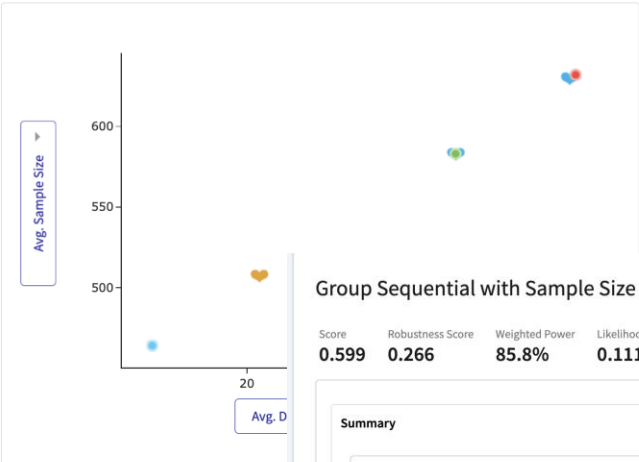
Change Scenario

Team Priorities



Solara Designs for Reference Scenario

Power



- Best Match**

Avg. Sample Size: **583** (519 - 910) | Power: **85.4%** | Avg. Duration (Months): **23.3** (20.7 - 28)
- Shortest Duration**

Avg. Sample Size: **464** (418 - 700) | Power: **63.7%** | Avg. Duration (Months): **18.5** (16.7 - 21)
- Lowest Sample Size**

Avg. Sample Size: **464** (418 - 700) | Power: **63.7%** | Avg. Duration (Months): **18.5** (16.7 - 21)
- Best Match Across Scenarios**

Avg. Sample Size: **632** (543 - 1008) | Power: **90.7%** | Avg. Duration (Months): **25.2** (21.7 - 30.3)
- Reference Design**

Avg. Sample Size: **630** (589 - 800) | Power: **86.4%** | Avg. Duration (Months): **25.1** (23.6 - 28.7)

Group Sequential with Sample Size Re-Estimation

Score	Robustness Score	Weighted Power	Likelihood	PFS: Superiority 2-Arms, Until End of Study
0.599	0.266	85.8%	0.111	

Summary

Designs	Scenarios	Financials
Number of Interim Analysis		1 (70%)
Number of Events		182
Sample Size		700
Allocation Ratio		1
1-Sided Type 1 Error		0.025
Test Statistics		Logrank
Critical HR		0.743
Efficacy Boundary Family: Spending Functions		LD (OF)
Futility Boundary Family: Spending Functions		LD (OF) (Non-Binding)

Output	
Avg Study Duration	23.311 Months
Power	85.4%
Avg Sample Size	582.974
Avg Number of Events	158.211
Avg Dropouts	0
Avg Accrual Duration	23.274 Months
Observed HR	0.64
Avg Follow Up Time	9.435 Months
Power Promising	0.84

Find Designs

Explore Test Designs **Find Designs**

Sim 2 - Modified (22-SEP-2022) ▾

Filters
Test Scenarios

POWER PROMISING (%) 🗑️

Base ▾

90
|
|
○
 100

POWER (%) 🗑️

Worse Trt Effect ▾

65
|
|
○
 94.1

MAX STUDY DURATION (MONTHS) 🗑️

Base ▾

20.92
|
|
○
 27

POWER (%) 🗑️

Base ▾

80
|
|
○
 96.8

7 Results of Base

Sort by: Best ⬆️⬆️

<input type="checkbox"/>	Avg. Sample Size 560 (508 - 1,120)	Power 81.1%	Avg. Duration (Months) 22.3 (20.2 - 26.2)
<input type="checkbox"/>	Avg. Sample Size 563 (486 - 840)	Power 81.6%	Avg. Duration (Months) 22.5 (19.4 - 26.8)
<input type="checkbox"/>	Avg. Sample Size 566 (488 - 1,008)	Power 81.8%	Avg. Duration (Months) 22.6 (19.5 - 26.8)
<input type="checkbox"/>	Avg. Sample Size 563 (508 - 1,120)	Power 81.1%	Avg. Duration (Months) 22.5 (20.2 - 26.8)
<input type="checkbox"/>	Avg. Sample Size 568 (487 - 936)	Power 82%	Avg. Duration (Months) 22.7 (19.4 - 26.7)
<input type="checkbox"/>	Avg. Sample Size 566 (468 - 864)	Power 81.3%	Avg. Duration (Months) 22.6 (18.7 - 26.9)
<input type="checkbox"/>	Avg. Sample Size 572 (469 - 960)	Power 81.1%	Avg. Duration (Months) 22.8 (18.7 - 27)

Favorited Designs

← Favorite Designs

Summary

Only Show Differences

♥ Best Match
♥ Shortest Duration SSR
♥ Reference Design

Outputs

Score	0.599	0	0.489
Avg Study Duration	23.311 Months	20.226 Months	25.149 Months
Power	85.4%	72.2%	86.4%
Avg Sample Size	582.974	506.533	629.05
Avg Number of Events	158.211	123.539	180.012
Avg Accrual Duration	23.274 Months	20.186 Months	25.11 Months
Observed HR	0.64	0.652	0.651
Avg Follow Up Time	9.435 Months	8.39 Months	10.036 Months
Power Promising	0.84	0.873	NA

Study Design

Statistical Design	Group Sequential with Sample Size Re-Estimation	Group Sequential with Sample Size Re-Estimation	Group Sequential
Robustness Score	0.266	0.088	0.319
Weighted Probability of Winning	85.8%	73.8%	88.1%

Selected Scenario

Time to Event Control 20

Time to Event HR 0.65

Probability Dropout Control 0

Probability Dropout Treatment 0

Response Scenario Name Response Set 1

Enrollment Scenario Name Enrollment Set 1

Likelihood 0.111

[Change Scenario](#)

View favorited designs under different scenarios



Simulation-based optimization of adaptive designs

Pantelis Vlachos, Cytel Inc, Geneva
Pantelis.Vlachos@Cytel.com





Your team's input

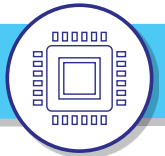
- Study endpoints, budget, ranges for sample size / enrollment / treatment effect, design options (e.g. fixed or adaptive)
- Team priorities (speed/cost/power)



Proprietary statistical design algorithms

- Trusted and validated for over 30 years
- Industry standard platform used by the FDA

=



Massive cloud compute power

- Parallel processing for near real-time design space generation



What it does:

- Based on team inputs, calculates study datasets for different designs and scenarios of interest
- Monte Carlo simulations scaled and applied across 1000s of permutations
- Helps teams visualize their options and select the best fit for their needs
- ***Routinely finds better designs than the manual process***

Thank you.

Cytel