
Evaluating Bayesian and Frequentist Analysis Options for an Oncology Study Design with Non-Proportional Hazards Assumptions

Pantelis Vlachos and Valeria Mazzanti
Pantelis.Vlachos@cytel.com

Finding an Optimal Design

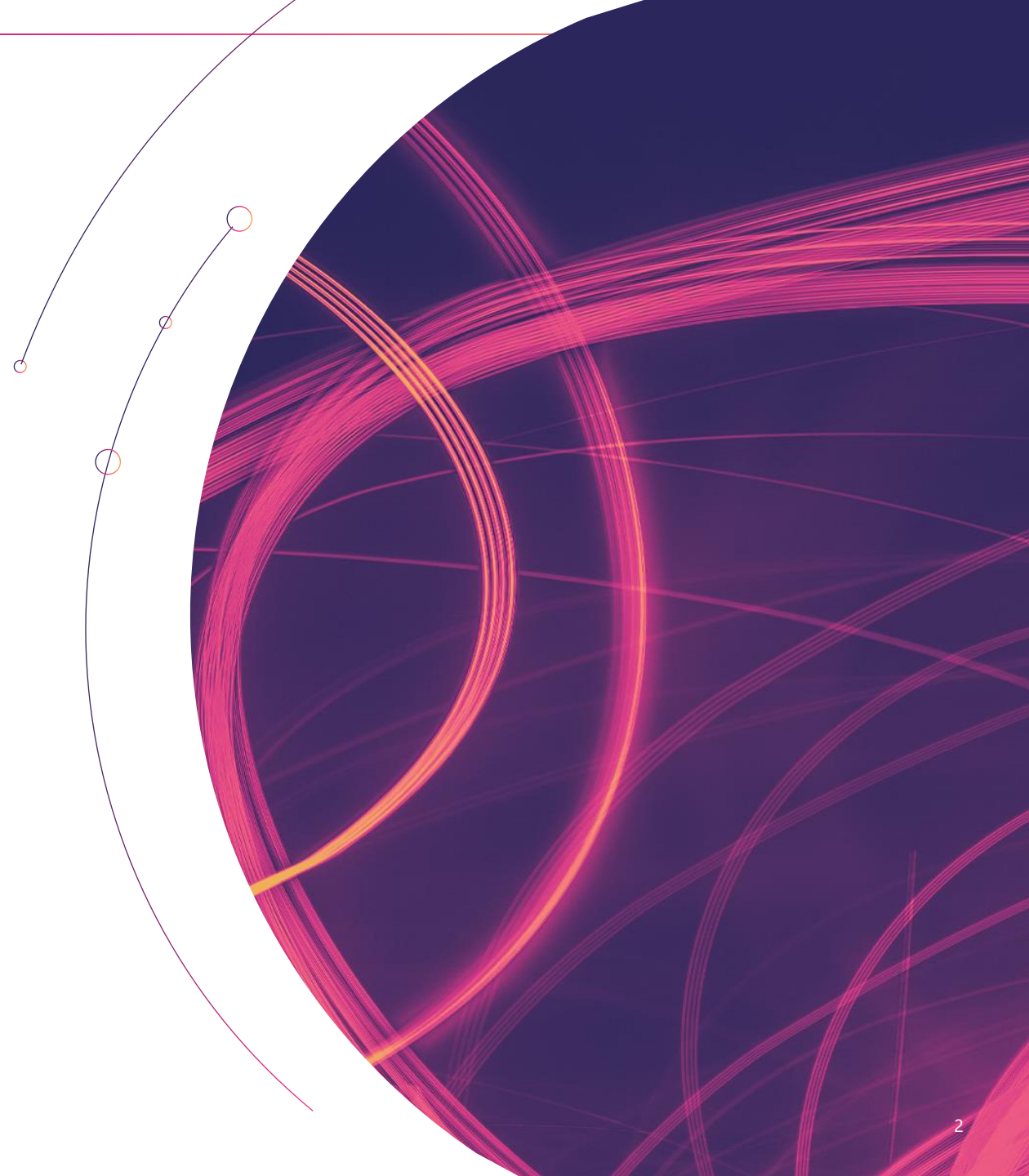
Case Study in Oncology

Collection of Designs Considered

Results/Conclusions

Q&A

Cytel



Cytel

Finding an Optimal Design



Trade-offs need to be considered

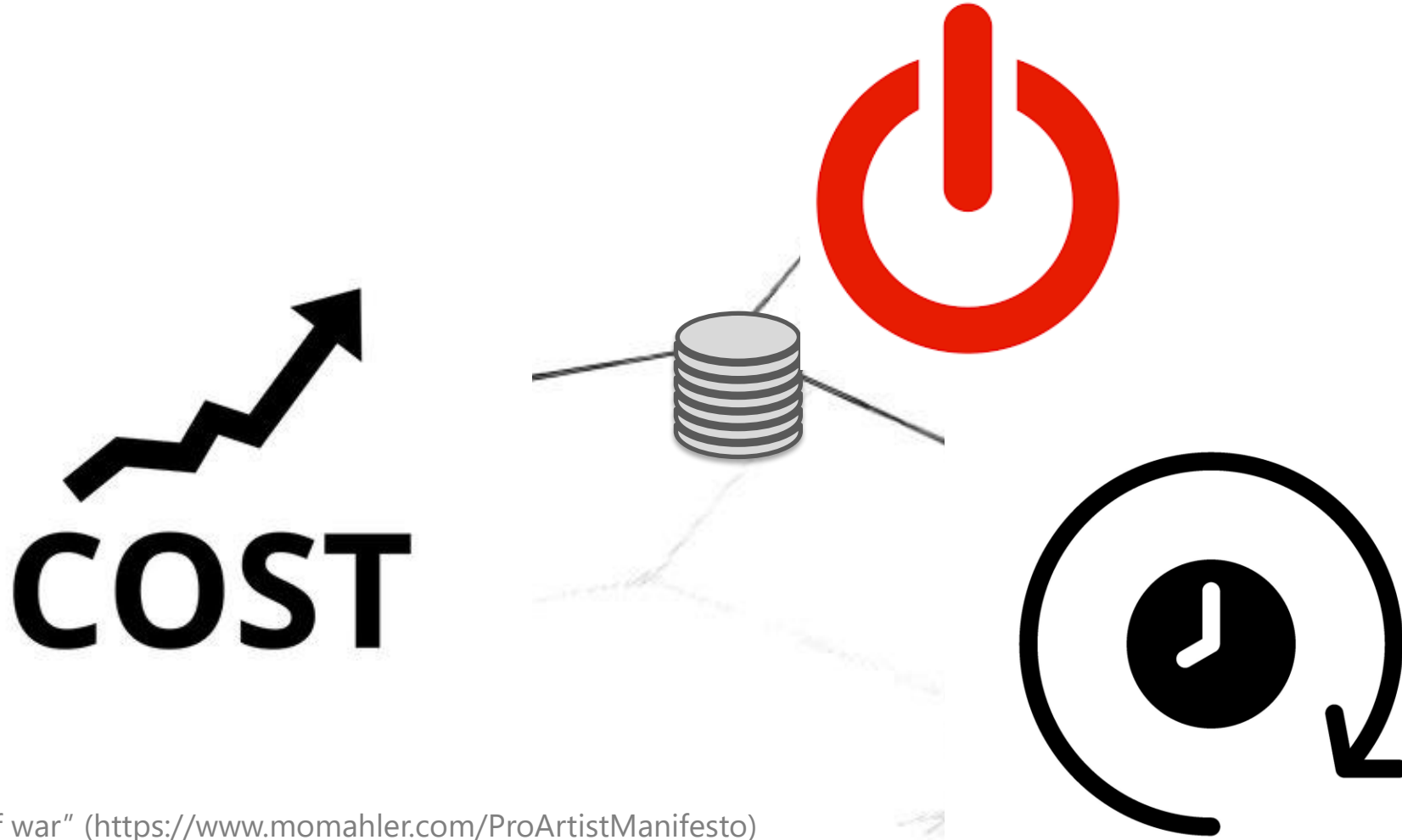


Image: "Three-way tug of war" (<https://www.momahler.com/ProArtistManifesto>)

Performance Score

$$\begin{aligned} & \textit{Score}(\textit{Design}|\theta) \\ &= w_P f(\textit{Power}) + w_T f(\textit{Time}) + w_C f(\textit{cost}) \end{aligned}$$

Conditional score for a Design given an assumed scenario θ is a weighted linear combination of Power, Time, and Cost/Sample Size

Robustness

Details on optimizing algorithm in 2024 Bayes-Pharma presentation

Robustness (Design)

$$= \int_{\theta} \mathbf{Score}(\mathbf{Design}|\theta)g(\theta)d\theta$$

Unconditional score for a Design given an assumed distribution (prior) for the scenario θ

Cytel

Case Study in Oncology



An Oncology Study

Design Parameters

Design Parameter	Input
Endpoint	Progression Free Survival
Sample Size	500
1-sided Type 1 Error	0.025
Target Power	80%
Number of Interim Analyses	0 or 1 at 50% IF
Efficacy Boundary	Lan-deMets (OBF)
Enrollment Rate	8 patients per month
Monthly Hazard Rate for Control Arm	0.0833
Expected Hazard Ratio	0.75
Potential Delayed Treatment Effect by	<ul style="list-style-type: none">• 2 months• 4 months• 6 months• 9 months• 12 months

Delayed Treatment Effect

- Most likely 2-4 months
- May be as high as 12 months

Historical Data

- Can Bayesian approaches be used?

Note:

No futility stop as the delayed treatment effect may incorrectly stop for futility

Modestly Weighted Log-Rank Test

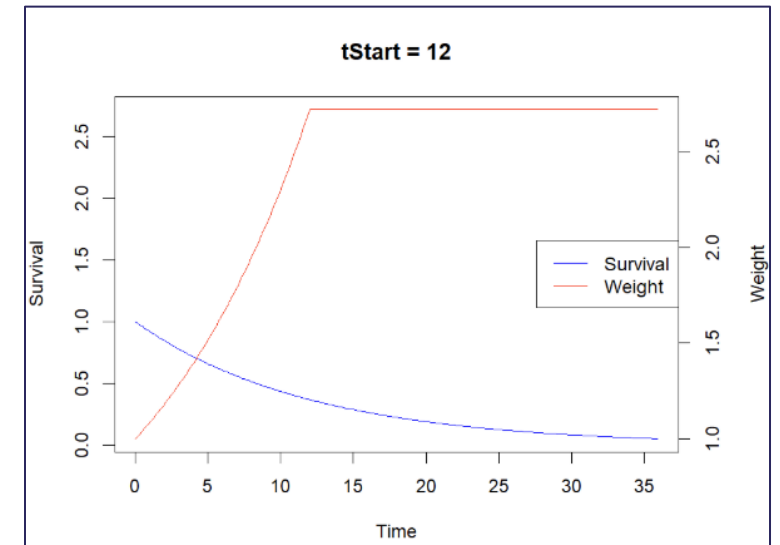
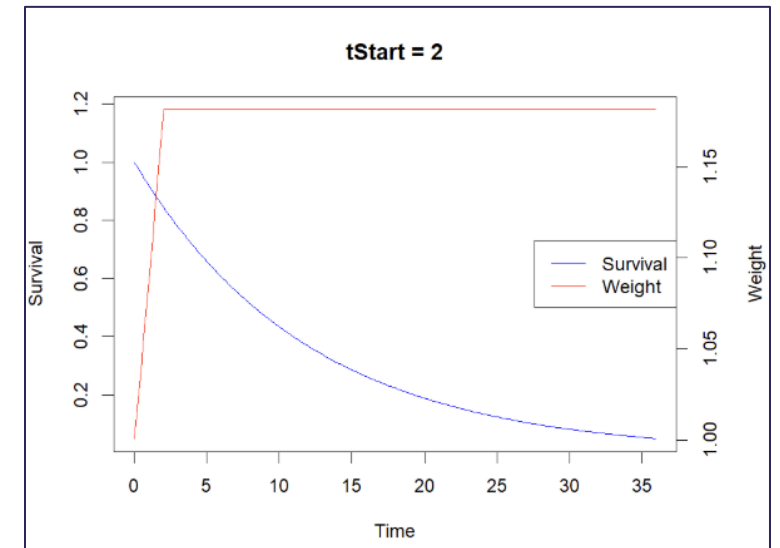
Background

In cases where survival curves cross or the treatment effect varies over time, the standard log-rank test might not perform optimally.

Weighting Mechanism

Unlike the traditional log-rank test, which applies equal weights across all time points, the modestly weighted log-rank test assigns **moderate, non-extreme weights** to event times.

This allows the test to capture differences in survival across a wider range of time points, rather than focusing only on early or late events.



Modestly Weighted Log-Rank Test Continued

Advantages

- More balanced power across different survival curve scenarios
- Useful in clinical trials where treatment effects may vary over time
- Can be a robust alternative to Fleming-Harrington weighted log-rank tests when choosing optimal weights is difficult.

Disadvantages

- Does not fully optimize power for extreme cases
- Choosing correct weights can be difficult
- Inference interpretation may become more complex
- Not optimal for small sample sizes
- Less common for regulatory approvals as FDA/EMA may be less familiar than they are with standard log-rank test

Add to East Horizon

- R Function for the computation utilizing **simtrial::wlr**
- Avoid hard coding of numerical values where possible
- User Inputs For
 - The initial delay period where weights increase; after this, weights are constant at the final weight in the delay period:

tStar = 12

References:

Freidlin, Boris, and Edward L Korn. 2019. "Methods for Accommodating Nonproportional Hazards in Clinical Trials: Ready for the Primary Analysis?" *Journal of Clinical Oncology* 37 (35): 3455–59

Magirr, Dominic. 2021. "Non-Proportional Hazards in Immuno-Oncology: Is an Old Perspective Needed?" *Pharmaceutical Statistics* 20 (3): 512–27

simtrial: Anderson, Keaven; <https://cran.r-project.org/web/packages/simtrial/>

Bayesian Model

Model Assumptions

- Denote the median time-to-event by θ_j for $j = C$ or E , for control treatment or experimental treatment, respectively
- **Prior**
 - $\theta_j \sim \text{Inverse Gamma}(a_j, b_j)$
 - Set a_j, b_j Such that the prior mean is 12 months and variance 1000 (non-informative)
- **Posterior**
 - $\text{TTT}_j = \text{Total time on test for arm } j$
 - $E_j = \text{Number of events for arm } j$
 - $\theta_j \mid \text{TTT}_j, E_j \sim \text{Inverse Gamma}(a_j + E_j, b_j + \ln(2) * \text{TTT}_j)$
- $p = \Pr\left(\frac{\theta_C}{\theta_E} < 1.0 \mid \text{data}\right)$
- At Interim Analysis if $p > 0.997 \rightarrow \text{Stop for early efficacy}$
- At Final Analysis if $p > 0.975 \rightarrow \text{Efficacy}$

Add to East Horizon

- R function for the computation
- Avoid hard coding of numerical values where possible
- User Inputs For
 - Prior parameter values for Control and Experimental
 - **dPriorACtrl = 2.144**
 - **dPriorBCtrl = 13.728**
 - Interim Analysis Cutoff: **dPUIA = 0.997**
 - Final Analysis Cutoff: **dPUFA = 0.975**
 - Other user inputs that are not discussed
 - **bCompareMedians = FALSE**
 - **dDelta = 1** (Cutoff for HR)

Analysis Variations

East Horizon

Fixed Log-rank test statistic

Group Sequential design using Log-rank test statistic

Group Sequential design with SSR option using Log-rank test statistic

Fixed and Group Sequential design with Max Combo test

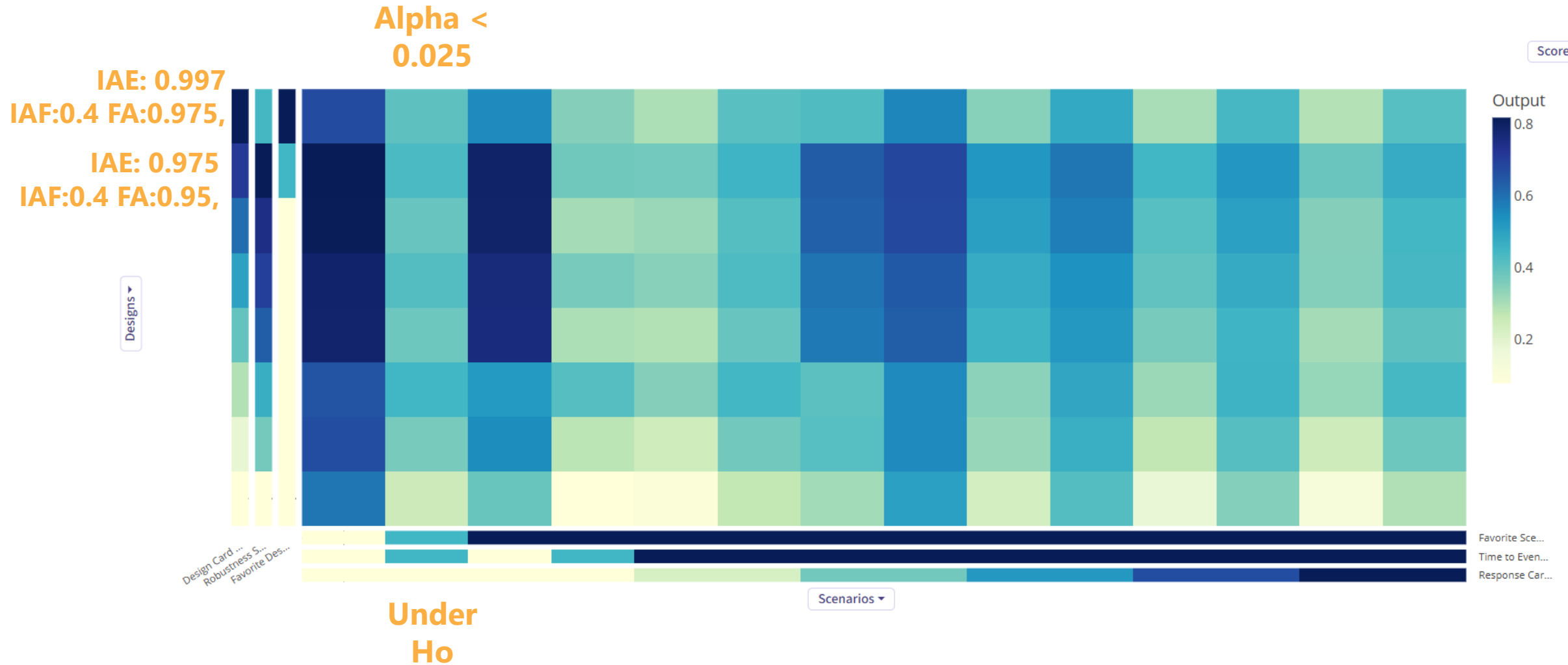
East Horizon + R Integration

Group Sequential design using Modestly Weighted Log-rank test statistic

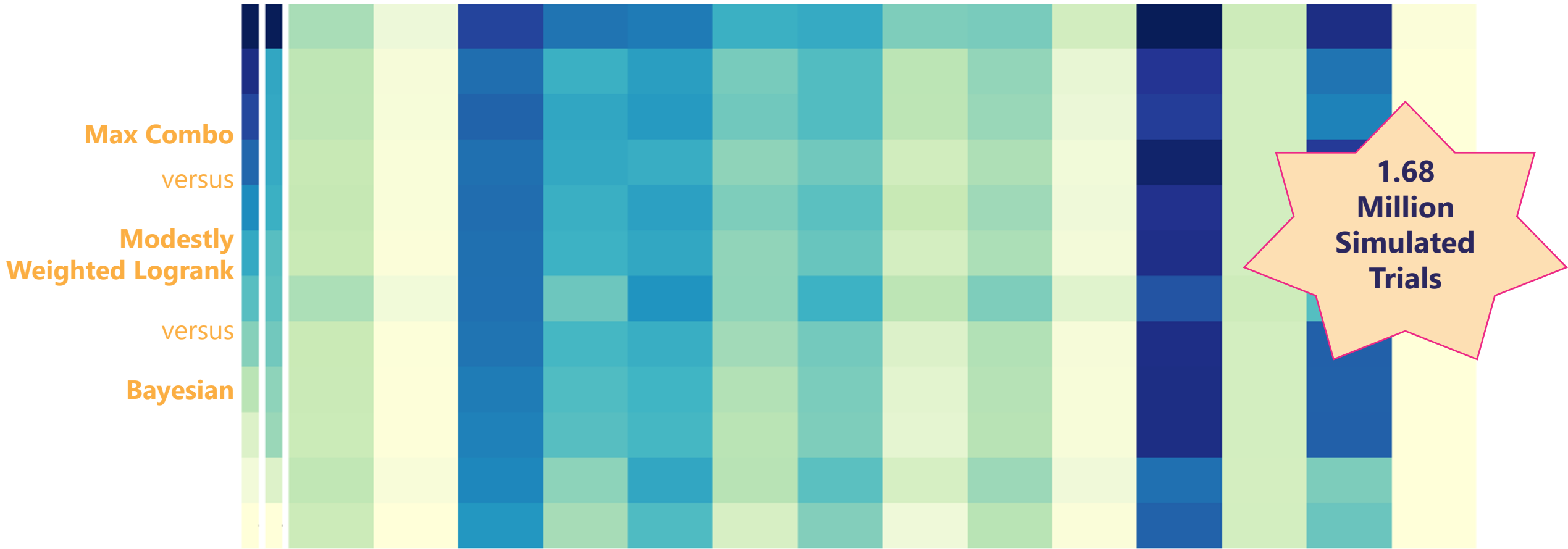
East Horizon + R Integration

Group Sequential design using R integration using Bayesian analysis

Finding the best Bayesian Design



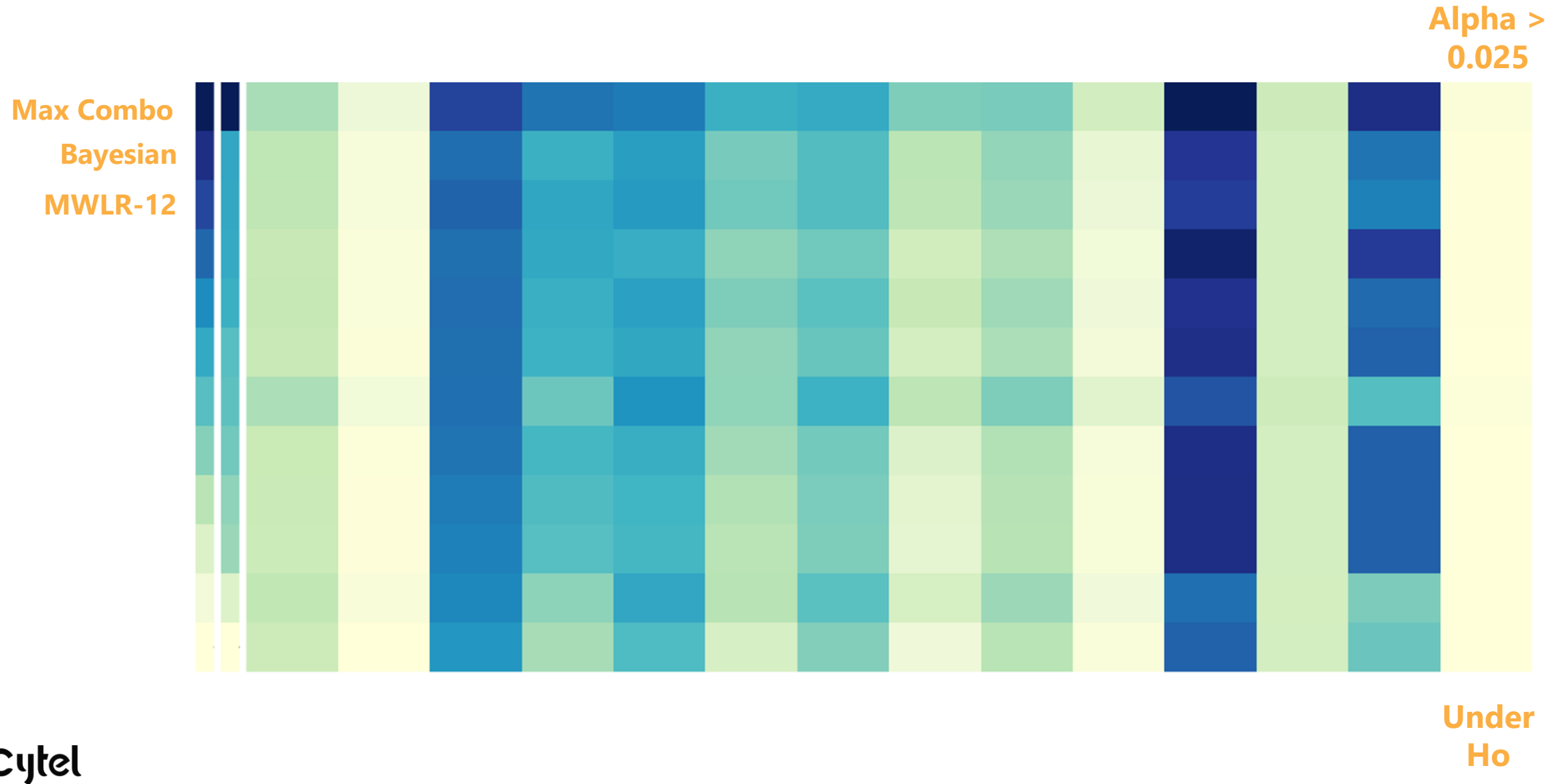
Finding the Best Design Overall



Cytel

 east horizon +  = 

Finding the Best Design Overall



Keep only designs that maintain type I error

Filters | Test Scenarios

New Filter Set Save As

Add Filter... ▼

POWER (%) 🗑️

Null ▼

2.46 ∞ 2.7

TEAM PRIORITIES (%)

Power: 40 | Sample Size: 30 | Duration: 30

7 Results of Expected

Sort by: Best ↕

<input type="checkbox"/>	Avg. Sample Size 447 (294 - 500)	Power 79.8%	Avg. Duration (Months) 55.8 (36.6 - 61.2)
<input type="checkbox"/>	Avg. Sample Size 457 (294 - 500)	Power 79.5%	Avg. Duration (Months) 57.1 (36.6 - 61.2)
<input type="checkbox"/>	Avg. Sample Size 458 (294 - 500)	Power 79.5%	Avg. Duration (Months) 57.1 (36.6 - 61.2)
<input type="checkbox"/>	Avg. Sample Size 458 (294 - 500)	Power 79.3%	Avg. Duration (Months) 57.2 (36.6 - 61.2)
<input type="checkbox"/>	Avg. Sample Size 458 (294 - 500)	Power 79%	Avg. Duration (Months) 57.2 (36.6 - 61.2)
<input type="checkbox"/>	Avg. Sample Size 459 (294 - 500)	Power 78.2%	Avg. Duration (Months) 57.3 (36.6 - 61.2)
<input type="checkbox"/>	Avg. Sample Size 490 (500 - 500)	Power 79.6%	Avg. Duration (Months) 61.2 (61.2 - 61.2)

And the winner is...

INPUTS

Designs	Scenarios	Financials
File Name (Design: Test)	AnalyzeUsingBayesInvGamma.R	
Function Name (Design: Test)	AnalyzeUsingBayesInvGamma	
Variables (Design: Test)	dPUIA (0.997), dPUFA (0.975), dPriorACtrl (2.144), dPriorBCtrl (13.728), dPriorAExp (2.144), dPriorBExp (13.728), bCompareMedians (0), dDelta (1)	
Number of Interim Analysis	1 (50%)	
Number of Events	380	
Sample Size	500	
Allocation Ratio	1	
1-Sided Type 1 Error	0.025	
Test Statistics	User Specified - R	
Critical HR	0.817	
Efficacy Boundary Family: Spending Functions	LD (OF)	
Target HR	0	

OUTPUT

Avg Study Duration	55.808 Months
Power	79.8%
Avg Sample Size	446.672
Avg Number of Events	338.295
Avg Dropouts	0
Avg Accrual Duration	55.623 Months
Observed HR	0.74
Avg Follow Up Time	10.4 Months





Thank you

Pantelis Vlachos

Pantelis.Vlachos@cytel.com

VP Customer Success

Cytel Inc. | Geneva, CH

<https://calendly.com/pantelis-vlachos/30min>



Unleash the Power of R-Infused Cytel Software

Cytel Software

- Cytel Software allows for confident and quick design through validated workflows and pre-specified and verified design types.
- Our software is a foundational, systematic approach to prioritizing, selecting, and comparing different complex design options.
- Our software automates certain aspects of simulation and selection, based on user criteria, and offers validation based on informed priors.
- Our solution contains cloud-native computing resources for simulation needs, as well as advanced file management and script editing functionality.

A Combined Solution

As Cytel continues to enable meaningful integration points of custom code in its state-of-the-art software, R&D organizations can enjoy the best of both worlds:

confident and fast results in a regulatory-trusted platform, with maximum flexibility in methods and statistical approaches.

Examples of useful R Integration in existing study design schema:

- Novel trial design & endpoint prioritization
- Bayesian methods, including assurance
- Variation in patient data simulation
- Variation in testing methods
- Variation in accrual and effect distributions

R Coding



- Coding in R allows for almost limitless flexibility in terms of methods, tests, and hypotheses.
- It is heavily dependent on the user's coding ability and requires time for writing and validation.
- Design with custom code demands additional resources for communicating results and design selection.
- Running code-based simulations requires the maintenance of a computing grid that requires highly-skilled engineers at an additional cost.