



# **Interim Decision Rules for Early Futility: A Comparative Assessment of Bayesian and Frequentist Approaches**

J. Kyle Wathen, PhD  
VP Scientific Strategy and Product Innovation  
Cytel

**Bayes 2025, Leiden**

**Oct 2025**

---

# Motivations for the Proposed Talk

## Interim decision-making is an integral part of efficient clinical trial design

Its benefits include:

- The incorporation of current and previously observed trial data
- The prediction of future outcomes based on Bayesian predictive probabilities compared to a frequentist approach which computes conditional power at the observed delta
- Decision-making for trial conduct using informed priors or more advanced modeling approach, eg Bayesian hierarchical model for the control data
- Bayesian predictive probabilities can incorporate past, current, and potential future data

---

# Potential Options for Futility Decision

## Frequentist Conditional Power

Conditional power thresholds based on Observed Difference – Native in East Horizon

---

## Bayesian Predictive Probabilities

Incorporating prior data directly into the prior, weighted – Requires R integration

---

## Bayesian Predictive Probabilities

Bayesian Hierarchical model for the response rate on control to include data for previous trials – Requires R Integration

---

## Difference in Observed Rates

If the observed difference is small, then stop for futility – Native to East Horizon

# An Oncology Case Study

## Standard Frequentist Design Parameters

Design Parameter	Input
Endpoint Type	Binary
Sample Size	210
1-sided Type 1 Error	0.025
Target Power	90%
Number of Interim Analyses	1 at 50% Information Fraction (IF)
Efficacy Boundary	Lan-DeMets (OBF)
Scenario Parameter	Input
Response Proportion for Control Arm	0.20
Response Proportion for Treatment Arm	0.40

### Start Simple

A group sequential design as a point of comparison for adaptive features and analysis options

# An Oncology Case Study

## Standard Frequentist Design Parameters

Design Parameter	Input
Endpoint Type	Binary
Sample Size	210, <b>250, 290, 310</b>
1-sided Type 1 Error	0.025
Number of Interim Analyses	1 at <b>40%</b> , 50%, <b>60</b> IF
Efficacy Boundary	Lan-DeMets (OBF)
<b>Futility based on conditional power</b>	<b>0.1, 0.01</b>
Scenario Parameter	Input
Response Proportion for Control Arm	0.20
Response Proportion for Treatment Arm	0.40

### Adding Additional Designs

Sample Sizes  
Information Fraction  
Futility: Conditional Power

Explore Trade-offs

= **24 designs added to the simulation**

# An Oncology Case Study

## Standard Frequentist Design Parameters

Design Parameter	Input
Endpoint Type	Binary
Sample Size	210, <b>250, 290, 310</b>
1-sided Type 1 Error	0.025
Number of Interim Analyses	1 at <b>40%</b> , 50%, <b>60</b> IF
Efficacy Boundary	Lan-DeMets (OBF)
<b>Futility based on observed delta</b>	0.01, 0.05, 0.1
Scenario Parameter	Input
Response Proportion for Control Arm	0.20
Response Proportion for Treatment Arm	0.40

### Adding Additional Designs

Sample Sizes

Information Fraction

**Futility: Observed Delta**

Explore Trade-offs

= **36 designs added to the simulation**

# Prior Study Data on Control

Study	Sample Size	Number of Responses
1	61	10
2	154	33
3	164	37
4	65	11
5	143	29
6	215	38
7	88	24
8	121	19
9	129	34
10	91	16
Total	1231	251

**Best Approach to Include?**  
Beta Binom – Prior mean equal 20%, Sample size of 1

Beta Binom – Prior mean equal to 20%, Sample size to 616 (50%), 1231 (100%)

Hierarchical Bayesian Model

**Overall Response Rate: 20.3%**

---

# Include External/Historical Data in Prediction

- External data available that is available at the interim may be utilized to make the futurity decision, aka “Totality of the data”
  - Variety of methods for inclusion of external data or historical
  - Why not include this in the simulation when conducting the prediction?
  
- Inclusion of previous data in analysis is beyond the scope of the presentation

# Bayesian Model

For simplicity assume a conjugate model

$X \sim \text{Binary}(\pi_i)$  for  $i = C$  (Control) or  $E$  (Experimental)

## Analysis Priors

$\pi_C \sim \text{Beta}(0.20, 0.80)$

$\pi_E \sim \text{Beta}(0.20, 0.80)$  Assume prior response rate is same C/E

## Decision Criteria

Efficacy:

$p = \Pr(\pi_E > \pi_C \mid \text{data}) > P_U \rightarrow E \text{ is better than } C$

Futility:

$\text{Pred Prob}\{ [\Pr(\pi_E > \pi_C \mid \text{data at final}) > P_{U\text{Final}}] \mid \text{data at Interim} \} < PL$

**Prior 1:** Non-informative  $\pi_C \sim \text{Beta}(0.20, 0.80)$

**Prior 2:** 50%,  $\pi_C \sim \text{Beta}(126, 490)$

**Prior 3:** 100%,  $\pi_C \sim \text{Beta}(251, 980)$

# Hierarchical Bayesian Model

## Prediction

Include prior study data and the data at the interim to obtain samples of the control response rate for the current trial using JAGS with the following model:

$$\text{Data: } y_i \sim \text{Binomial}(n_i, p_i)$$

$$\text{Hierarchy: } \text{logit}(p_i) = \mu + u_i, \quad u_i \sim \mathcal{N}(0, \tau^2)$$

Priors: weakly informative on  $\mu$ ; half-uniform on  $\tau$

## Decision Criteria

Analysis prior

Efficacy:

$$p = \Pr(\pi_E > \pi_C \mid \text{data}) > P_U \rightarrow \text{E is better than C}$$

Futility:

$$\text{Pred Prob}\{ [\Pr(\pi_E > \pi_C \mid \text{data at final}) > P_{U\text{Final}}] \mid \text{data at Interim} \} < PL$$

# An Oncology Case Study

## Simple Bayesian Approach (Conjugate Prior)

Design Parameter	Input
Endpoint Type	Binary
Sample Size	210, 250, 290, 310
1-sided Type 1 Error	0.025
Number of Interim Analyses	1 at 40%, 50%, 60 IF
PU (At Interim Analysis)	PU = 0.995
PU Final	0.98
PL	0.1, 0.01
Scenario Parameter	Input
Response Proportion for Control Arm	0.20
Response Proportion for Treatment Arm	0.40

**Adding Additional Designs**  
Sample Sizes  
Information Fraction  
Futility Cutoff  
Prior

**24 designs added to the simulation for each Bayesian Approach**

Total Bayesian designs added to simulation: 96

Total Frequentist designs added to simulation: 60

**Total Designs: 156**

# Sensitivity Analysis

## Impact of different control response rates

Control Response Rate	Experimental Response Rate
0.20	<b>0.20</b> , 0.25, 0.30, <b>0.40</b>
0.15	<b>0.15</b> , 0.20, 0.25, <b>0.35</b>
0.25	<b>0.25</b> , 0.30, 0.35, <b>0.45</b>

### Futility Goals

In the **alternative**, teams want less than 1%-2% probability of declaring futility  
Minimum probability of futility in **null** cases of 50%

---

## Possible Drawback

- These Bayesian options are not currently native to East Horizon
- Not proficient in R
  - > Don't want to develop an entire code base for the simulations to include Bayesian options
- Already have R code available for both types of Bayesian analysis
- Don't want to use East Horizon for frequentist designs and develop all the code for the Bayesian designs

# Implementation of Bayesian Analysis

R Code Assistant - Binary - Analysis

Can I help you with your R code?

You can enter your own prompt below to generate R code that can be used in your input set.  
There is a limit of 25 prompts per session.

**Bayesian Model**

For simplicity assume a conjugate model  
 $X \sim \text{Binary}(\pi_i)$  for  $i = C$  (Control) or  $E$  (Experimental)

**Priors**

$\pi_C \sim \text{Beta}(a, b)$ ; Vary amount of prior data utilized

- Prior 1:** Non-informative  $\pi_C \sim \text{Beta}(0.20, 0.80)$
- Prior 2:** 50%,  $\pi_C \sim \text{Beta}(126, 490)$
- Prior 3:** 100%,  $\pi_C \sim \text{Beta}(251, 980)$

Reflect prior response rate of 20%

$\pi_E \sim \text{Beta}(0.20, 0.80)$  Assume prior response rate is same C/E

**Decision Criteria**

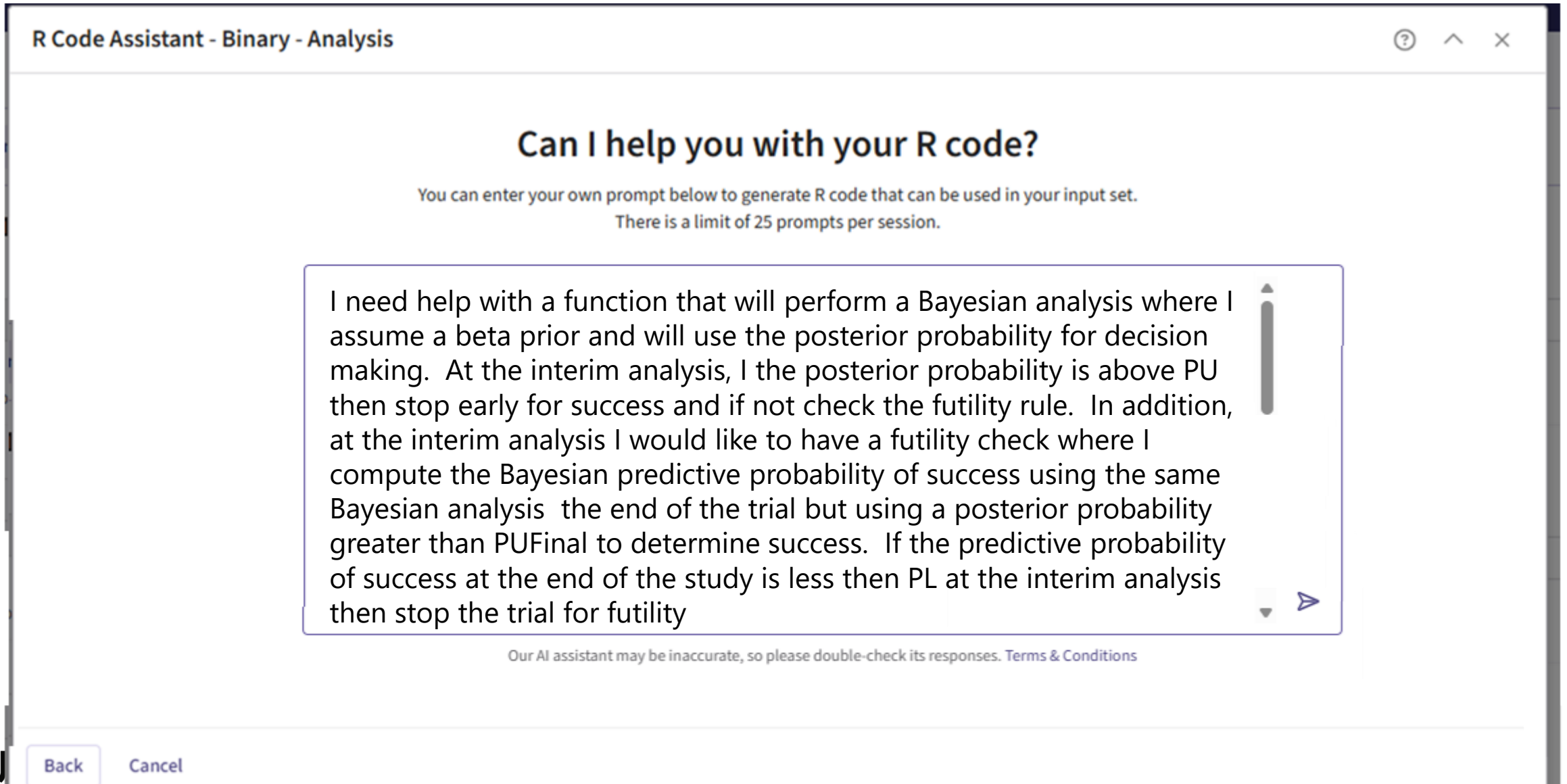
Efficacy:  
 $p = \text{Pr}(\pi_E > \pi_C | \text{data}) > P_U \rightarrow E \text{ is better than } C$

Futility:  
 $\text{Pred Prob}\{ \text{Pr}(\pi_E > \pi_C | \text{data at final}) > P_{\text{Final}} | \text{data at Interim} \} < PL$

Our AI assistant may be inaccurate, so please double-check its responses. [Terms & Conditions](#)

Back Cancel

# Implementation of Bayesian Analysis



R Code Assistant - Binary - Analysis

Can I help you with your R code?

You can enter your own prompt below to generate R code that can be used in your input set.  
There is a limit of 25 prompts per session.

I need help with a function that will perform a Bayesian analysis where I assume a beta prior and will use the posterior probability for decision making. At the interim analysis, I the posterior probability is above  $P_U$  then stop early for success and if not check the futility rule. In addition, at the interim analysis I would like to have a futility check where I compute the Bayesian predictive probability of success using the same Bayesian analysis the end of the trial but using a posterior probability greater than  $P_{UFinal}$  to determine success. If the predictive probability of success at the end of the study is less then  $P_L$  at the interim analysis then stop the trial for futility

Our AI assistant may be inaccurate, so please double-check its responses. [Terms & Conditions](#)

Back Cancel

I need help with a function that will perform a Bayesian analysis where I assume beta prior and will use the posterior probability for decision making. At the interim analysis, if the posterior probability is above PU then stop early for success and if not check the futility rule. In addition, at the interim analysis I would like to have a futility check where I compute the Bayesian predictive probability of success using the same Bayesian analysis at the end of the trial but using a posterior probability greater than PUFinal to determine success. Can you help me right this code



Generated

```
1 # Check and verify your code before running.
2 # Run the code to see the results.
3 # Extract the function before saving or uploading to East Horizon.
4 # Visit this help page for more information: https://cytel-inc.github.io/CyneRe
5
6 PerformBayesianAnalysis <- function(SimData, DesignParam, LookInfo = NULL, User
7   # Initialize variables
8   nError <- 0
9   nDecision <- 0
10  dTestStatistic <- 0
11
12  # Step 1: Determine the number of patients in the analysis
13  if (!is.null(LookInfo)) {
14    nLookIndex <- LookInfo$CurrLookIndex
15    nQtyOfPatsInAnalysis <- LookInfo$CumCompleters[nLookIndex]
16  } else {
17    nLookIndex <- 1
18    nQtyOfPatsInAnalysis <- nrow(SimData)
19  }
20
```

Enter R code prompt here...

Our AI assistant may be inaccurate, so please double-check its responses. [Terms & Conditions](#)

Back

Cancel

```

123 # Function to compute Bayesian predictive probability
124 ComputeBayesianPredictiveProbabilitywithBayesianAnalysis <- function(
125   datas, dataE, priorAlphas, priorBetas, priorAlphaE, priorBetaE,
126   nQtyOfPatsS, nQtyOfPatsE, nSimulations, finalBoundary, dLowerCutoffForFutility,
127   lAnalysisParams
128 ) {
129   posteriorAlphas <- priorAlphas + sum(datas)
130   posteriorBetas <- priorBetas + length(datas) - sum(datas)
131
132   posteriorAlphaE <- priorAlphaE + sum(dataE)
133   posteriorBetaE <- priorBetaE + length(dataE) - sum(dataE)
134
135   successfulTrials <- 0
136
137   nRepToCheck <- ceiling(nSimulations/2) # Check half way through and if greater than dLowerCutoffForFutility then no need to continue
138   for (i in 1:nSimulations) {
139     posteriorRates <- rbeta(1, posteriorAlphas, posteriorBetas)
140     posteriorRateE <- rbeta(1, posteriorAlphaE, posteriorBetaE)
141
142     remainingDataS <- rbinom(nQtyOfPatsS - length(datas), 1, posteriorRates)
143     combinedDataS <- c(datas, remainingDataS)
144
145     remainingDataE <- rbinom(nQtyOfPatsE - length(dataE), 1, posteriorRateE)
146     combinedDataE <- c(dataE, remainingDataE)
147
148     result <- ProbExpGreaterCtrlBeta(
149       combinedDataS, combinedDataE,
150       lAnalysisParams$dAlphaCtrl, lAnalysisParams$dBetaCtrl,
151       lAnalysisParams$dAlphaExp, lAnalysisParams$dBetaExp
152     )
153
154
155     if (result$dPostProb > finalBoundary) {
156       successfulTrials <- successfulTrials + 1
157     }
158
159     if( i == nRepToCheck )
160     {
161       if( successfulTrials/nRepToCheck > 2*dLowerCutoffForFutility ) # We will not be able to hit the boundary so stop the prediction
162       {
163
164         return(list(predictiveProbabilityS = successfulTrials/nRepToCheck))
165       }
166     }
167   }
168
169   predictiveProbabilityS <- successfulTrials / nSimulations
170   return(list(predictiveProbabilityS = predictiveProbabilityS))
171 }
172

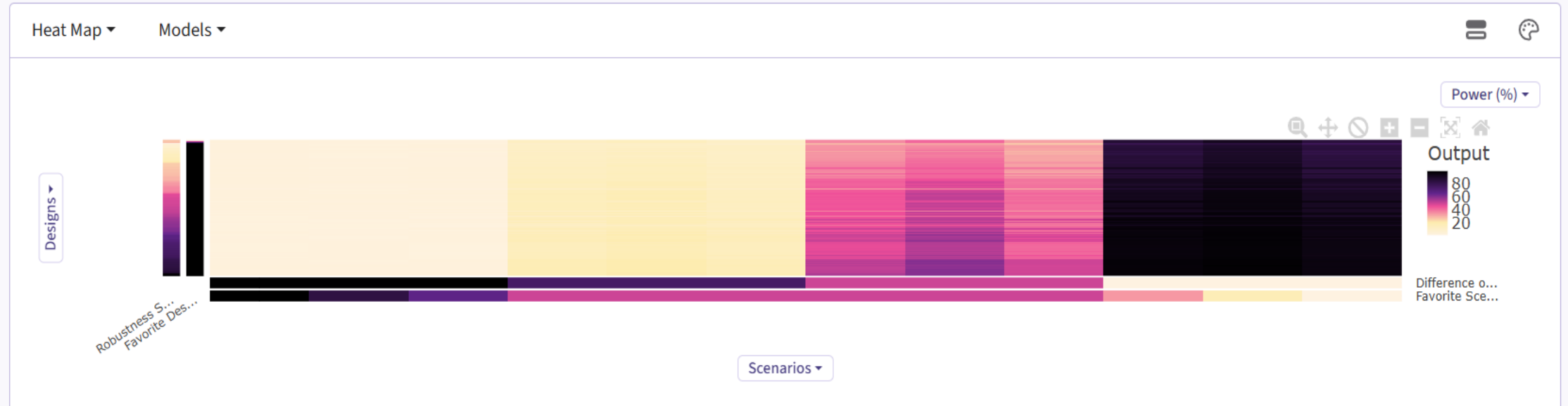
```

- Bayes 2025 C...
- Overview
- Inputs
- Results**
- Compare
- Files

Result - Input 9 - Sensitivity x +

Result - Input 9 - Sensitivity ... View 1 ... Dashboard Details

1,872 Models ⓘ ⌵ ⚙️ 🔍 ⬇️ Download Save



☐	Model ID	ⓘ	Favorites				
			Favorite Designs	Favorite Scenarios	Avg Accrual Duration ...	Avg Completers	Avg Follow Up Time
☐	1	ⓘ	♡	🚩 Null 20	16.68	129.49	0.97
☐	2	ⓘ	♡	🚩 Null 20	19.93	155.5	0.98
☐	3	ⓘ	♡	🚩 Null 20	23.25	182.47	0.98

All Models

# Sensitivity Analysis

## Impact of different control response rates

Control Response Rate	Experimental Response Rate
0.20	<b>0.20</b> , 0.25, 0.30, <b>0.40</b>
0.15	<b>0.15</b> , 0.20, 0.25, <b>0.35</b>
0.25	<b>0.25</b> , 0.30, 0.35, <b>0.45</b>

### Futility Goals

In the **alternative**, teams want less than 1%-2% probability of declaring futility  
Minimum probability of futility in **null** cases of 50%

# Find Designs



Filters Test Scenarios

New Filter Set ▾ Save As

Add Filter... ▾

**TEAM PRIORITIES (%)**

Power	Sample Size	Duration
<input type="text" value="40"/>	<input type="text" value="30"/>	<input type="text" value="30"/>

Note: Performance metrics in Find Designs are based on the original scenario weights from the simulation. You can modify Team priorities independently within this feature.

Show Results Clear All

**Search for Designs**  
Please add filters on the left to search for Designs.

# Find Designs



Filters

Test Scenarios

New Filter Set ▾

Save As

Add Filter...



POWER

Power

Weighted Power

Probability of Stopping Early

Efficacy Probability of Stopping Early

Futility Probability of Stopping Early

TIME

Max Study Duration

Min Time Difference between Interim Analysis

Average Study Duration

Show Results

Clear All

## Search for Designs

Please add filters on the left to search for Designs.

# Find Designs



Filters Test Scenarios

New Filter Set Save As

Add Filter...

FUTILITY PROBABILITY OF STOPPING EARLY (%)

Alt 20 v 40

IA1

0 15.1

TEAM PRIORITIES (%)

Power 40 Sample Size 30 Duration 30

Note: Performance metrics in Find Designs are based on the original scenario weights from the simulation. You can modify Team priorities independently within this feature.

Show Results Clear All

156 Results of Null 20

Sort by: Best

<input type="checkbox"/>	Avg. Sample Size 107 (92 - 210)	Power 1.6%	Avg. Duration (Months) 13.4 (11.5 - 27.1)
<input type="checkbox"/>	Avg. Sample Size 121 (92 - 210)	Power 2.3%	Avg. Duration (Months) 15.3 (11.5 - 27.1)
<input type="checkbox"/>	Avg. Sample Size 122 (92 - 210)	Power 2%	Avg. Duration (Months) 15.3 (11.5 - 27.1)
<input type="checkbox"/>	Avg. Sample Size 123 (114 - 210)	Power 1.9%	Avg. Duration (Months) 15.4 (14.1 - 27.1)
<input type="checkbox"/>	Avg. Sample Size 123 (92 - 210)	Power 2.1%	Avg. Duration (Months) 15.5 (11.5 - 27.1)
<input type="checkbox"/>			Avg. Duration (Months) 15.5 (11.5 - 27.1)
<input type="checkbox"/>			Avg. Duration (Months) 15.8 (11.5 - 27.1)
<input type="checkbox"/>			Avg. Duration (Months) 15.6 (13.5 - 32.1)
<input type="checkbox"/>			Avg. Duration (Months) 16.4 (14.1 - 27.1)
<input type="checkbox"/>	Avg. Sample Size 131 (114 - 210)	Power 2.1%	Avg. Duration (Months) 16.5 (14.1 - 27.1)

**Some designs had a 15.1% chance of stopping for futility in the alternative case 20% vs 40%**



# Find Designs



Filters Test Scenarios

New Filter Set Save As

Add Filter...

FUTILITY PROBABILITY OF STOPPING EARLY (%)

Alt 20 v 40

IA1

0 1

TEAM PRIORITIES (%)

Power Sample Size Duration

40 30 30

Note: Performance metrics in Find Designs are based on the original scenario weights from the simulation. You can modify Team priorities independently within this feature.

Show Results Clear All

## 74 Results of Null 20

Sort by: Best

	Avg. Sample Size	Power	Avg. Duration (Months)	
<input type="checkbox"/>	159 (114 - 210)	2.4%	20.2 (14.1 - 27.1)	
<input type="checkbox"/>	160 (114 - 210)	2.5%	20.3 (14.1 - 27.1)	
<input type="checkbox"/>	161 (134 - 210)	2.4%	20.3 (16.6 - 27.1)	
<input type="checkbox"/>	161 (114 - 210)	2.5%	20.5 (14.1 - 27.1)	
<input type="checkbox"/>	162 (134 - 210)	2.5%	20.5 (16.6 - 27.1)	
<input type="checkbox"/>	163 (134 - 210)	2.4%	20.5 (16.6 - 27.1)	
<input type="checkbox"/>	163 (92 - 210)	2.5%	20.9 (11.5 - 27.1)	
<input type="checkbox"/>	164 (92 - 210)	2.5%	20.9 (11.5 - 27.1)	
<input type="checkbox"/>	165 (124 - 290)	1.8%	20.7 (15.5 - 37.1)	
<input type="checkbox"/>	164 (92 - 210)	2.5%	21 (11.5 - 27.1)	

# Find Designs



Filters Test Scenarios

IA1 ▼

50  95.3

FUTILITY PROBABILITY OF STOPPING EARLY (%) 🗑️

Null 20 ▼

IA1 ▼

50  96.1

FUTILITY PROBABILITY OF STOPPING EARLY (%) 🗑️

Null 15 ▼

IA1 ▼

50  97.8

Show Results Clear All

## 3 Results of Null 20

Sort by: Best ↕

<input type="checkbox"/>	Avg. Sample Size 161 (134 - 210)	Power 2.4%	Avg. Duration (Months) 20.3 (16.6 - 27.1)	<span>♥</span> <span>☰</span>
<input type="checkbox"/>	Avg. Sample Size 162 (134 - 210)	Power 2.5%	Avg. Duration (Months) 20.5 (16.6 - 27.1)	<span>♥</span> <span>☰</span>
<input type="checkbox"/>	Avg. Sample Size 163 (134 - 210)	Power 2.4%	Avg. Duration (Months) 20.5 (16.6 - 27.1)	<span>♥</span> <span>☰</span>

**Futility constraints added  
N = 210**

# Find Designs



Filters
Test Scenarios

**Final Selection Crit** Save As

Add Filter... ▼

**MAX SAMPLE SIZE** 🗑️

Alt 15 v 36 ▼

310 
|
|
○
 310

**FUTILITY PROBABILITY OF STOPPING EARLY (%)** 🗑️

Null 25 ▼

IA1 ▼

50 
|
|
○
○
 95.3

Show Results
Clear All

## 5 Results of Null 20

Sort by: Best ↓↑

<input type="checkbox"/>	Avg. Sample Size 193 (164 - 310)	Power 1.7%	Avg. Duration (Months) 24.2 (20.3 - 39.6)
<input type="checkbox"/>	Avg. Sample Size 194 (164 - 310)	Power 1.9%	Avg. Duration (Months) 24.3 (20.3 - 39.6)
<input type="checkbox"/>	Avg. Sample Size 197 (164 - 310)	Power 2%	Avg. Duration (Months) 24.7 (20.3 - 39.6)
<input type="checkbox"/>	Avg. Sample Size 232 (194 - 310)	Power 2.1%	Avg. Duration (Months) 29.2 (24.1 - 39.6)
<input type="checkbox"/>	Avg. Sample Size 249 (194 - 310)	Power 2.2%	Avg. Duration (Months) 31.4 (24.1 - 39.6)

**Futility constraints added**  
**N = 310**  
**1 Conditional Power Option meets the criteria**

---

# Simulation Summary

- Hierarchical Bayesian model included in the prediction for futility had the most robust performance
  - Also applied if the same model is used for analysis (simulations not discussed)
- The Beta-Binomial model with historical data in the prediction for futility had robust performance
  - **MAJOR CAUTION** – Not true in general, especially in therapeutic areas that have high trial to trial variation in the control response rate
  - Not true if the same prior is used for analysis
- Conditional Power and Observed difference both meet the criteria if the futility probability in the alternative is increased to 2%-3%
  - In general, required larger average sample size and slightly longer average trial duration

---

# Summary

- Hybrid approach of East Horizon and the R integration removed limitations and allowed both frequentist and Bayesian options to be considered in the same software platform
- Use of the coding assistant reduced the code development time to under 1 hour for the Bayesian analysis, predictive probabilities, and hierarchical model/JAGS code
- Comparison of standard native approaches to newly published or exploratory options is greatly simplified with the R integration
  - Patient arrival time, data generation, analysis
- General flow:
  - R integration → CyneRgy → Anything in CyneRgy appears as a native option in East Horizon
  - R integration → Private Client Repo → Appears as a native option in East Horizon for that client

Cytel



**Thank you!**

J. Kyle Wathen

[kyle.wathen@cytel.com](mailto:kyle.wathen@cytel.com)

