Simulation-guided designs: a new approach to trial design and probabilistic risk assessment

Yannis Jemiai, Ph.D., CSO

Presented on 24 October 2023
At BASS XXX

Outline

Introduction

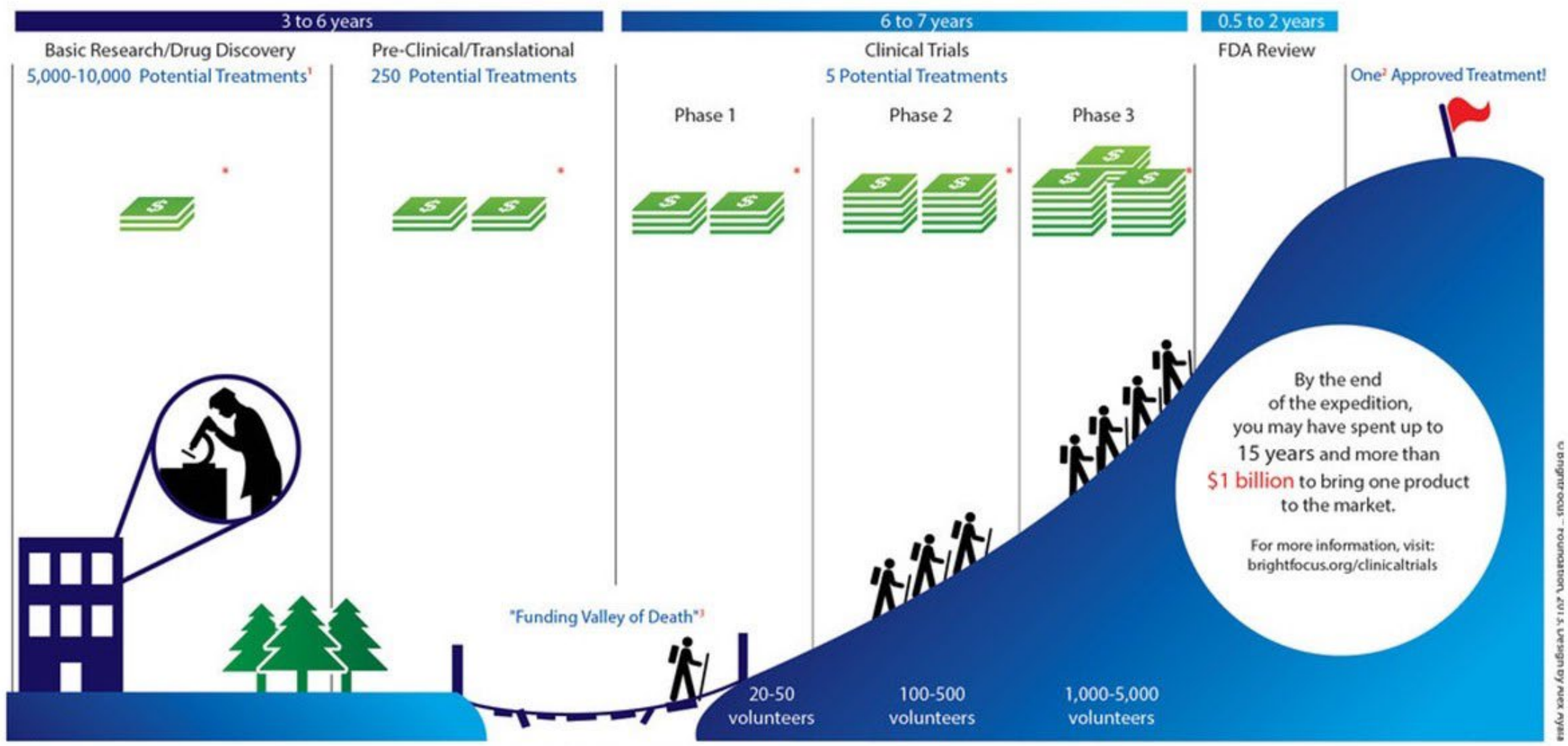
Challenges and limitations of traditional design process

Designing clinical trials with cloud-based simulations

Risk assessment and decision-making

Conclusion

Bringing a drug to market is lengthy, risky and expensive



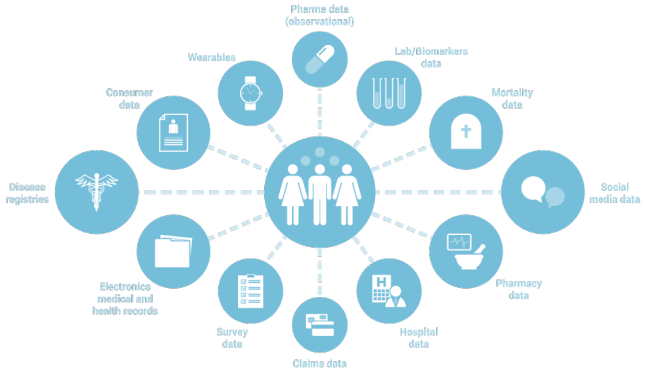
Designing modern clinical trials is increasingly complex

Multiple Objectives

Prioritize and balance

Complex decisions

Control Error Rates



Diverse Data Sources

Borrow Information

Novel data sources (RWD)

Control bias and confounding

Evidence-Based Decisions

Data-based / Quantitative

Assess Trade-offs

Cross-functional

Overcoming challenges

Scientific Innovation

Process Innovation

Technology Innovation

Design failures can have terrible consequences

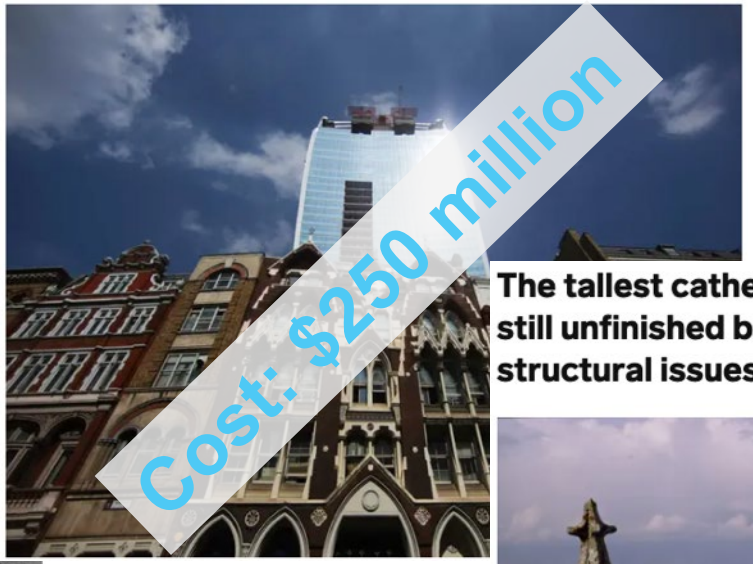
The Tacoma Narrows Bridge twisted back and forth in strong winds before collapsing in 1940. The failure is now a case study for physics students.

Cost: $11 million

All 10,344 window panes of the John Hancock Tower needed to be replaced because of poor glass integrity.

Cost: $175 million

A poorly designed tower in London reflected and magnified sunlight until it was melting cars and frying eggs.



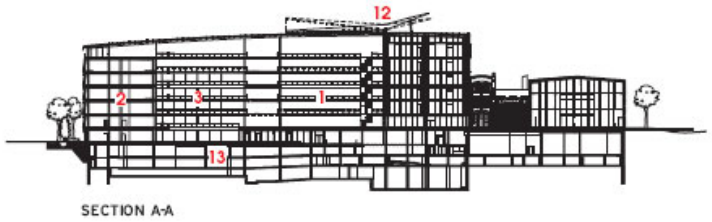
Cost: $250 million

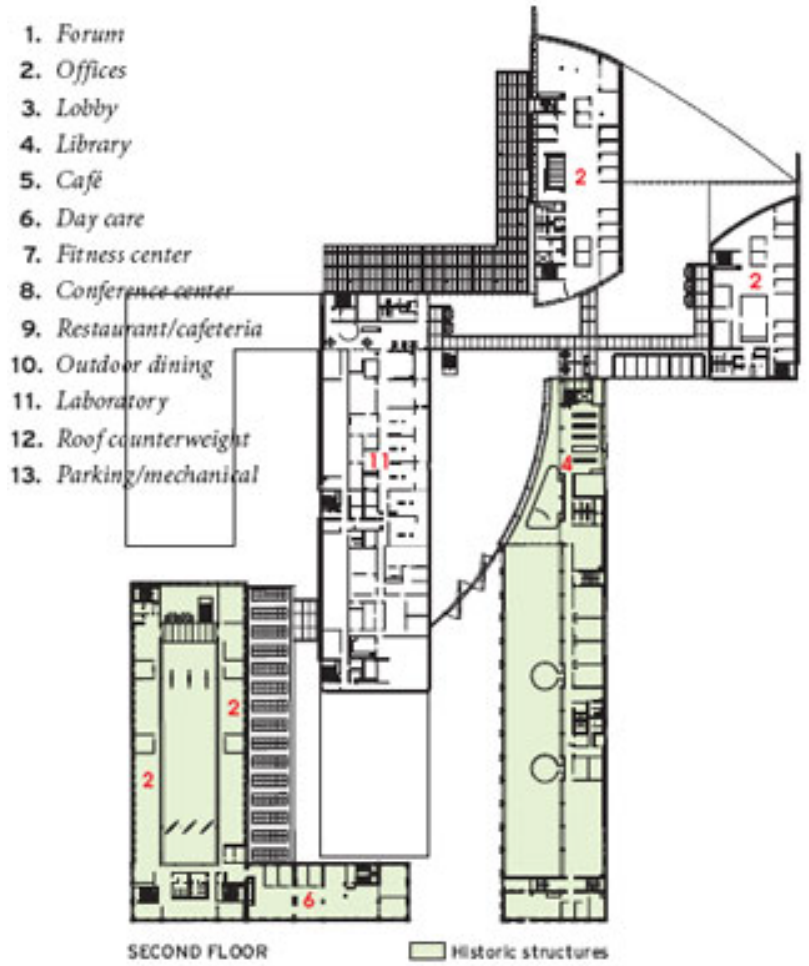
The tallest cathedral in the world is technically still unfinished because its height created structural issues.

Cost: N/A

source: https://www.businessinsider.com/biggest-structural-failures-disasters-history-2019-11

The design of a clinical trial is the **blueprint** for its eventual success



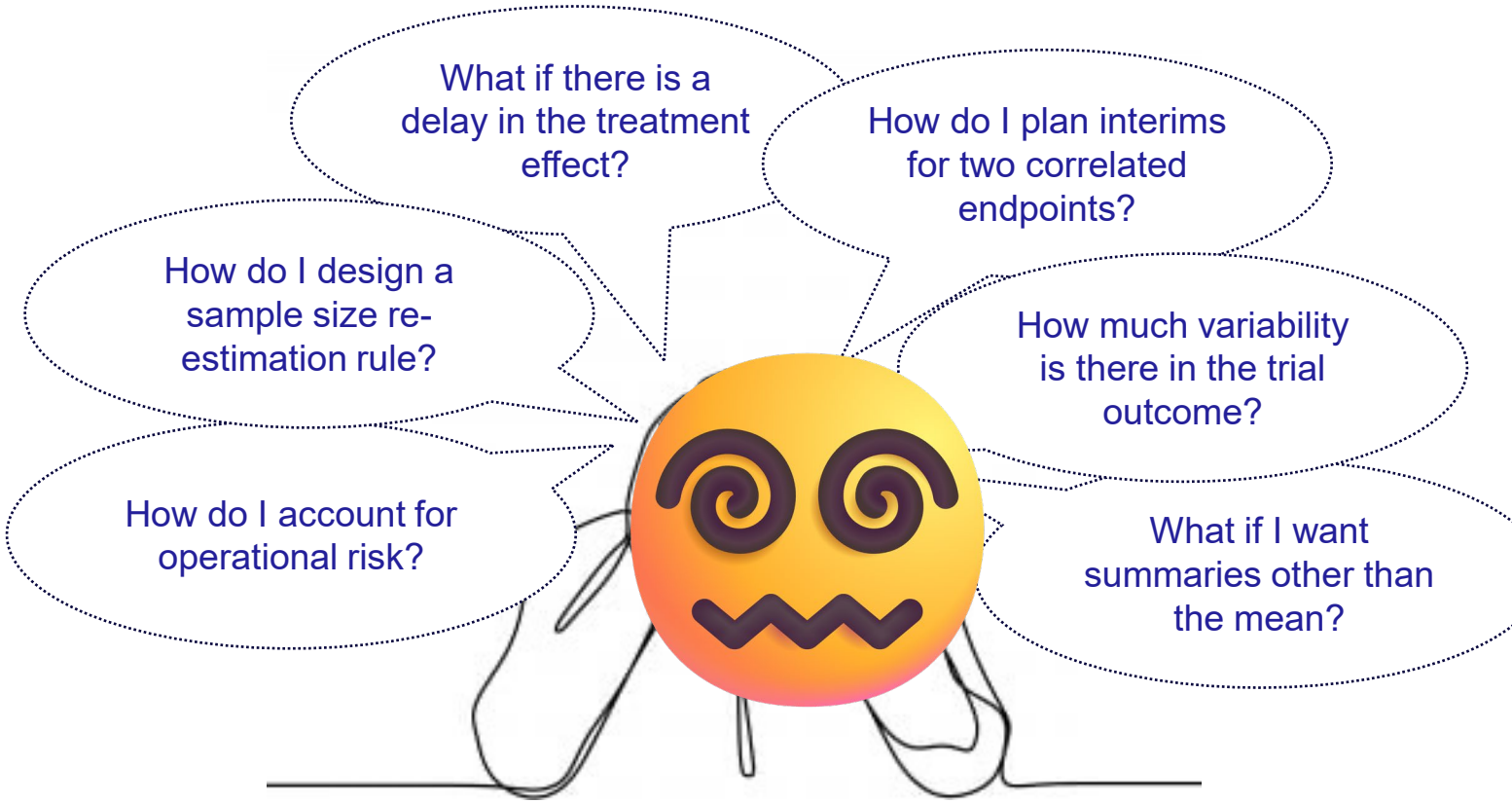


The traditional analytical approach to design has serious drawbacks and limitations

	Des16
Test Type	1-Sided
Specified α	0.025
Attained α	0.023
Power	0.874
Model Parameters	
Allocation Ratio (nt/nc)	2
Hazard Ratio (Alt.)	0.7
Var (Log HR)	Null
Boundary Parameters	
Spacing of Lo	ual
Efficacy Boundary	LD (OF)
Futility Boundary	Gm (1.5) (NB)
Accrual & Dropout Parameters	
Accrual Rate	22
Subjects are Followed	Until End of Study
No. of Accrual Periods	1
No. of Dropout Pieces	1
Sample Size	
Maximum	572
Expected Under H0	428.961
Expected Under H1	535.75
Events	
Maximum	374
Expected Under H0	256.474
Expected Under H1	339.484
Study Duration	
Maximum	29.573
Expected Under H0	20.007
Expected Under H1	27.425
Accrual Duration	
Maximum	26
Expected Under H0	19.498
Expected Under H1	24.352

Boundary Parameters

Team questions may be challenging to answer rapidly and effectively

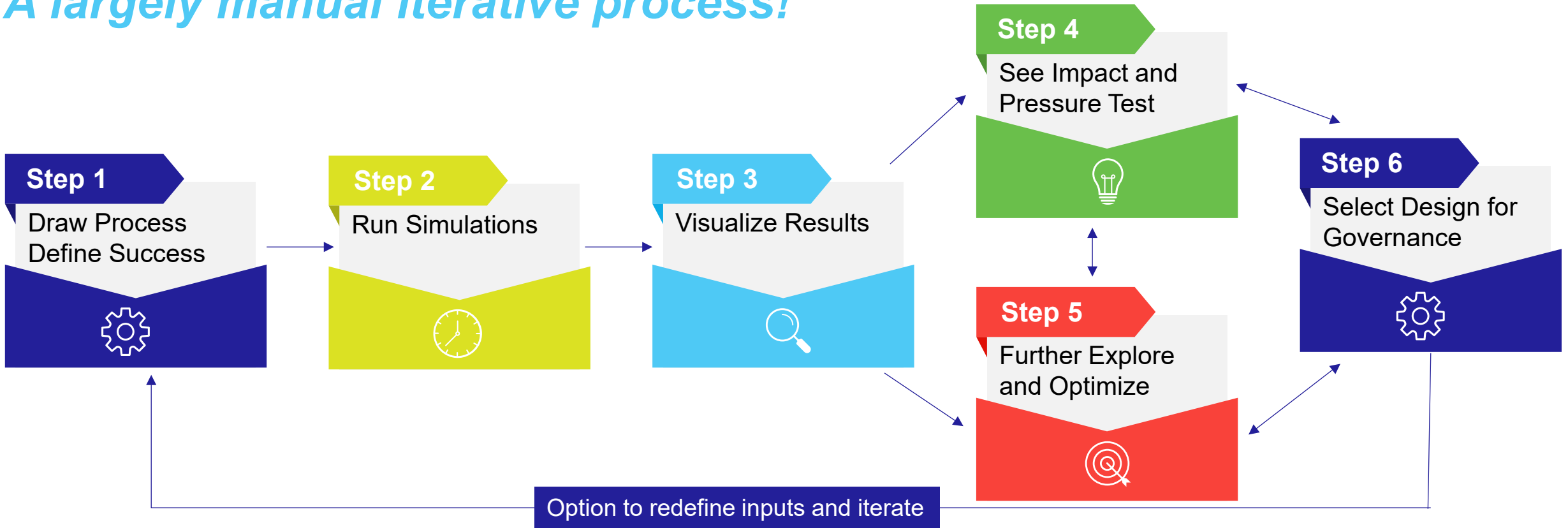


Instead simulations offer several advantages

Cytel

How does simulation guided study design work?

A largely manual iterative process!

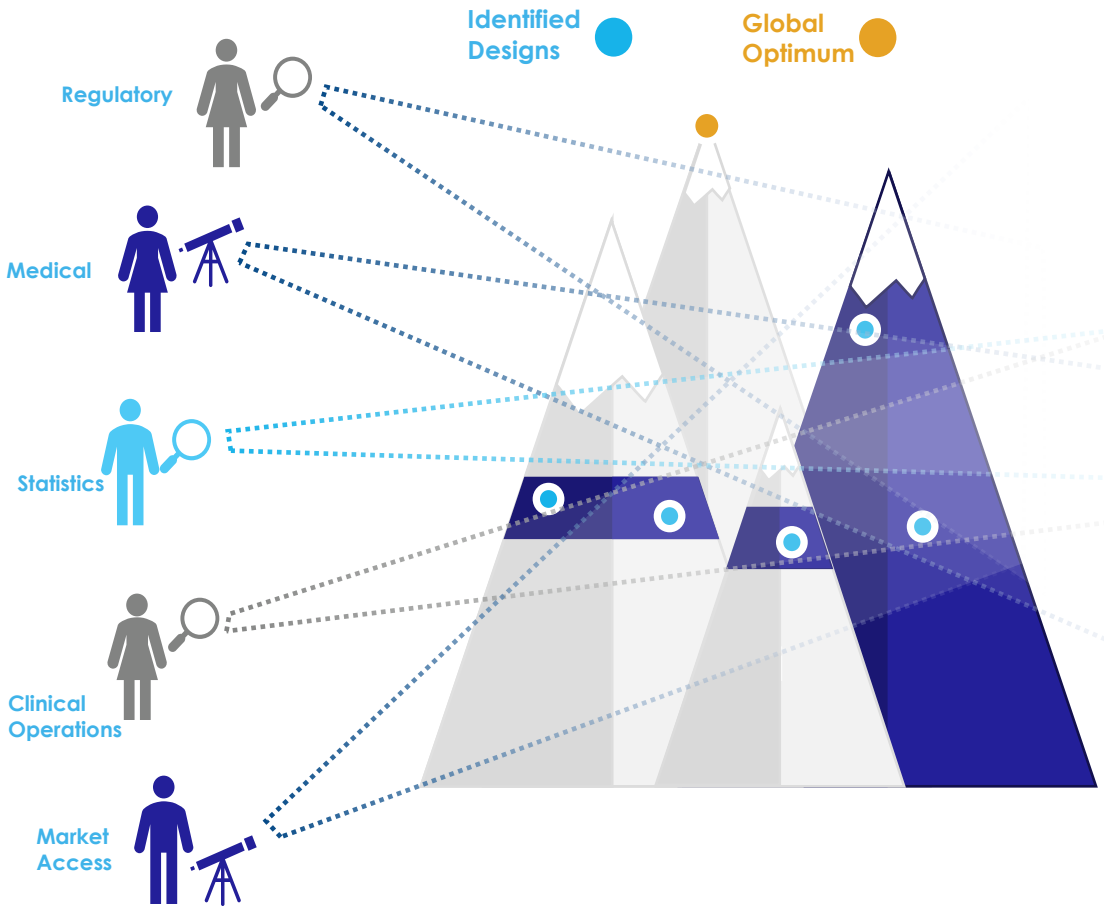


Scaling this process fits seamlessly with how teams work and provides great benefits

- Formalizes best practices in clinical study design – predefined goals, diligence, visualization
- Transparently delineates options, not just favorites
- Maximizes PTS, minimizes technical overhead
- Facilitates incorporation of clinical and commercial strategy into design

The core design challenge remains

Traditional approaches are suboptimal due to uncertain inputs and lack of sufficient tools



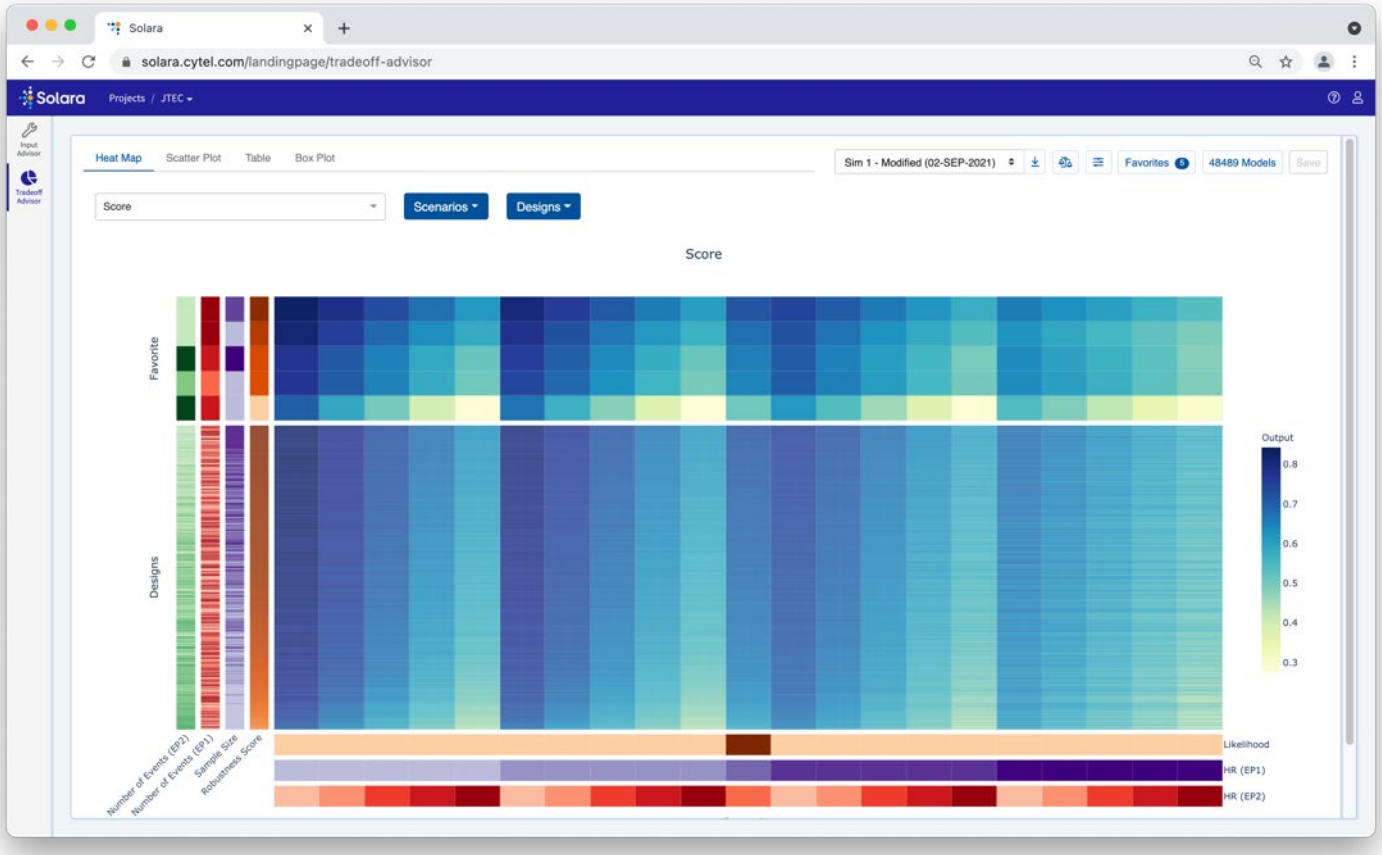
Challenges

1. Traditional methods of developing statistical operating characteristics do not generate the most efficient designs
2. Study design input estimates (e.g., treatment effect, enrollment rate, cost, market size) are prone to error and bias
3. Fragmented, iterative study design process takes months and does not properly balance enrollment, treatment effect, regulatory and market access dynamics

Impact: Inefficient study designs that take too long, are too costly and fail too often

A new improved study design process combines team inputs, simulation algorithms and massive cloud compute

High throughput screening for clinical trial designs



Many thousand simulation models can be explored in less than an hour

- ✓ Statisticians' time focused on finding the optimal design among tens of thousands of options
- ✓ Assess and mitigate uncertainty in all critical assumptions
- ✓ Evaluate stats, business, operational risks
- ✓ Interactive real-time collaboration among PDT members; answers all "what-if" questions

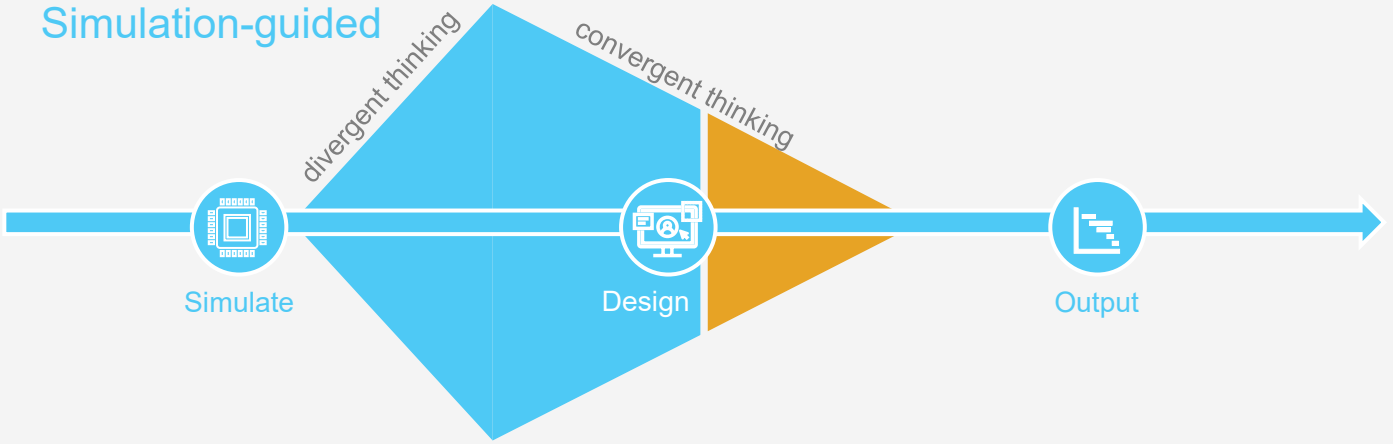
An evolution in trial design thinking

Traditional

Challenges

- Design possibilities often limited from the beginning
- Time and resource constraints restrict number of designs and scenarios that can be considered
- Binary study-by-study decision of what tool to use

Simulation-guided



Benefit

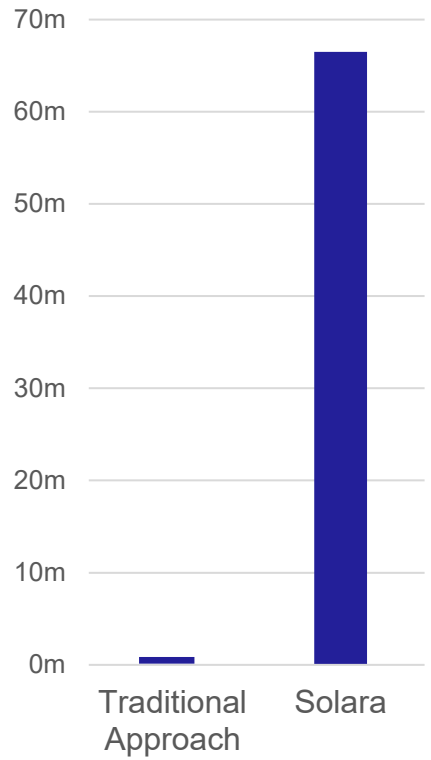
- Optimal designs modeled against business strategy
- Cross-functional collaboration on design selection
- Accelerate speed to market

Simulation guided design of the PLUTO* trial

Case Study

TRADITIONAL APPROACH	NEW APPROACH
Assumption	Evidence Based
Variable, Reactive	Predictable, Proactive
Finite number of designs	Infinite scenarios
Shared domain expertise	Intelligent inputs and assumptions

66.5M trials simulated for PLUTO

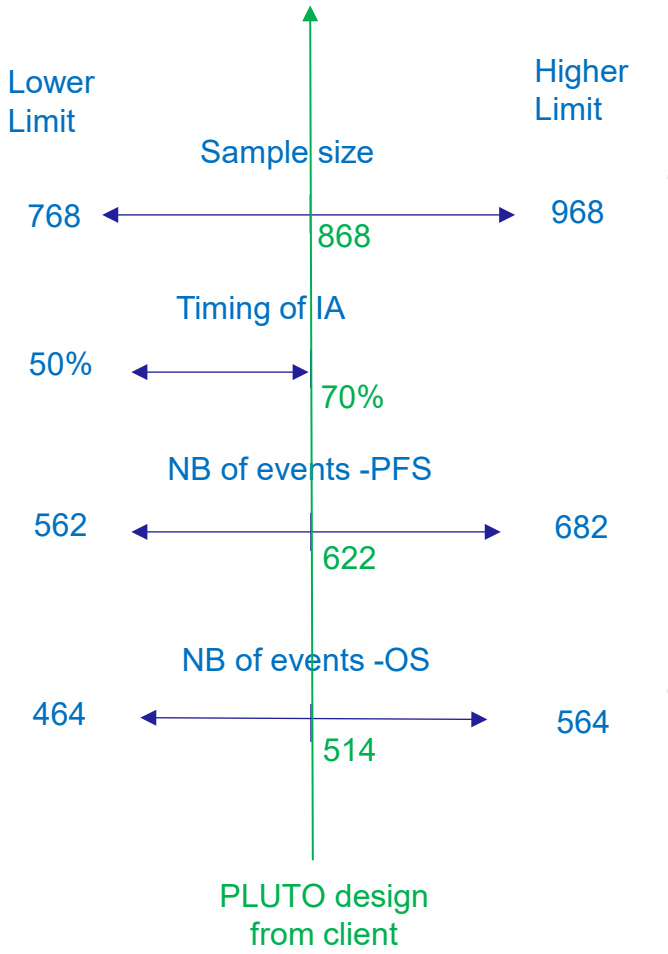


- ☑ **Maintain power**
  No material reduction in power
- ☑ **Reduce timelines**
  4.5-month reduction
- ☑ **Reduce cost**
  $8.1M estimated savings

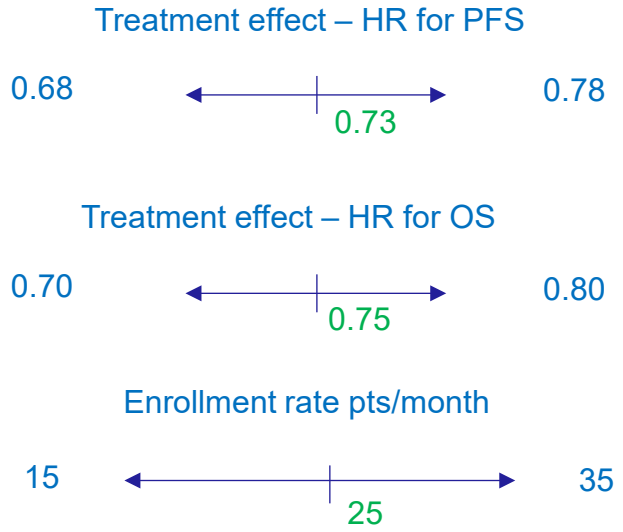
Cytel

* Fictitious study name to preserve confidentiality

Simulation exploration– PLUTO



Design exploration =7,393 designs

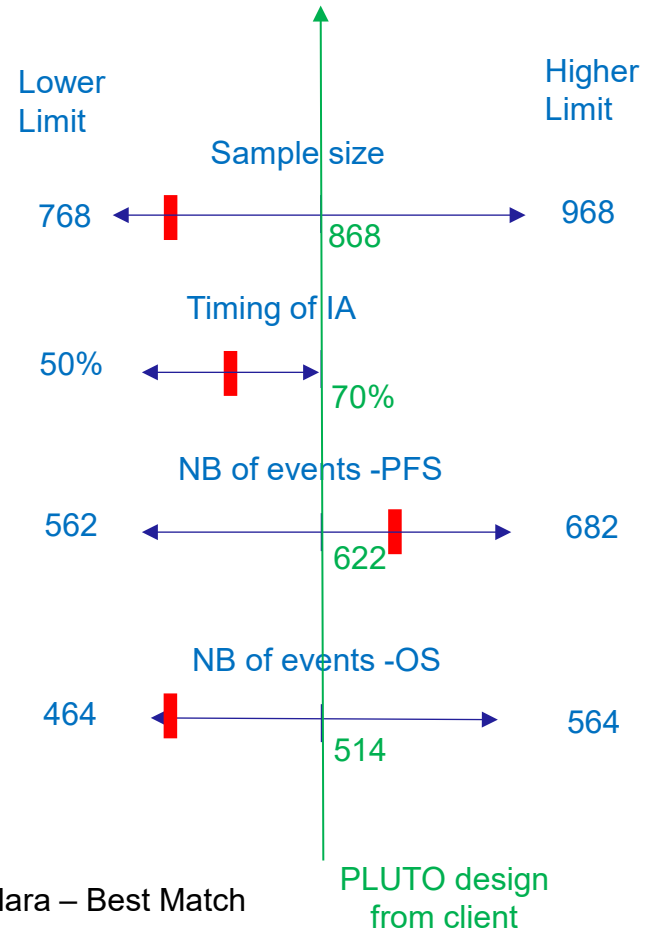


Scenario variations =9 scenarios

7,393 designs x 9 scenarios x1000 simulations
= **66.5 million trials simulated**
v.s. industry avg of 5 designs x 3 scenarios x 1000 simulations = 15,000 trials simulated

What changes can be made to impact study PLUTO

Case Study



Reduce Sample Size → reduce cost and duration → But lose power

Earlier timing of IA → Better probability to stop at IA2 for efficacy → Gain Power

Changed ratio of PFS (+20) /OS (-50) events → Gain Power and reduce slightly duration

Pro: Cost savings of $8.1m associated with 80 patients and 5 months reduction
Con : Very minor reduction of 0.3% power to 79.7%

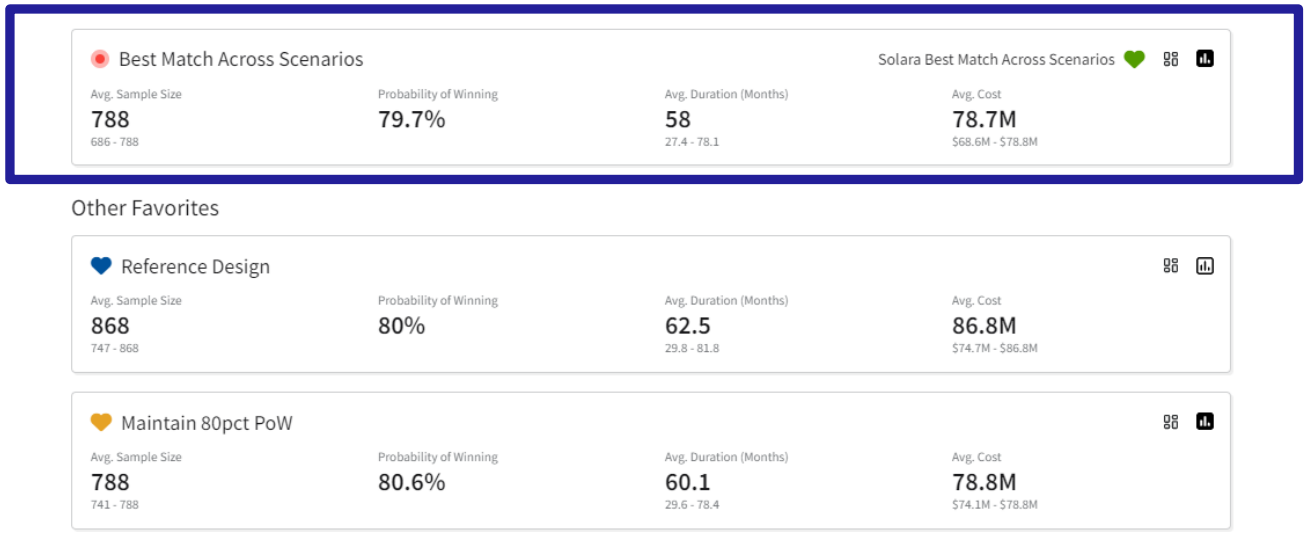
Summary of trial design considerations for PLUTO

Case Study

Maintained appropriate power under the base scenario while reducing average sample size significantly (**80 fewer patients on average**), and average study duration by close to **5 months**.

Key Assumptions for Simulation:

- Timing of IA can be moved
- IA stopping rule can be modified (efficacy, futility)
- Number of events for 2 endpoints can be changed



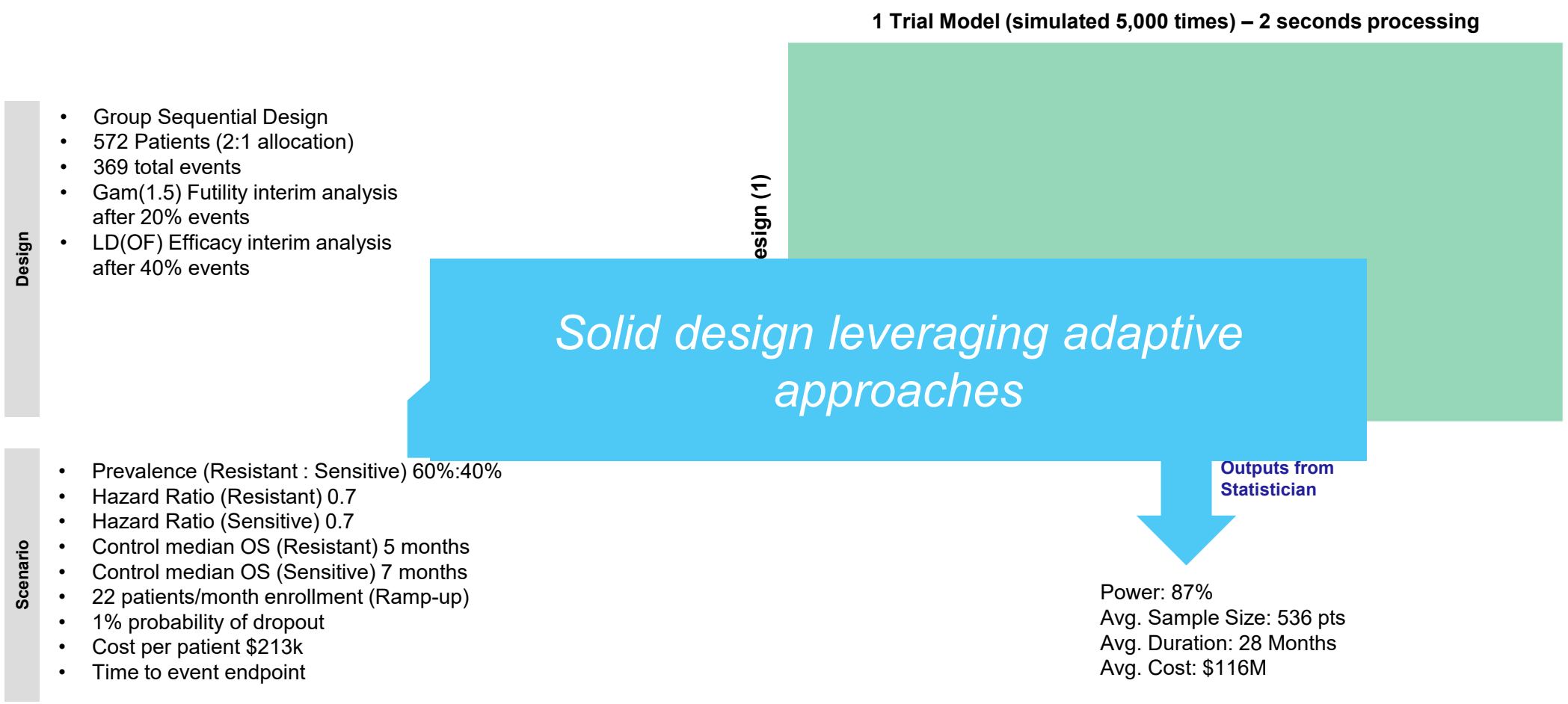
Recommended Design Changes:

- reduce the planned sample size
- vary the number of events for each of the endpoints
- add a futility rule and change time of IA earlier

Cytel

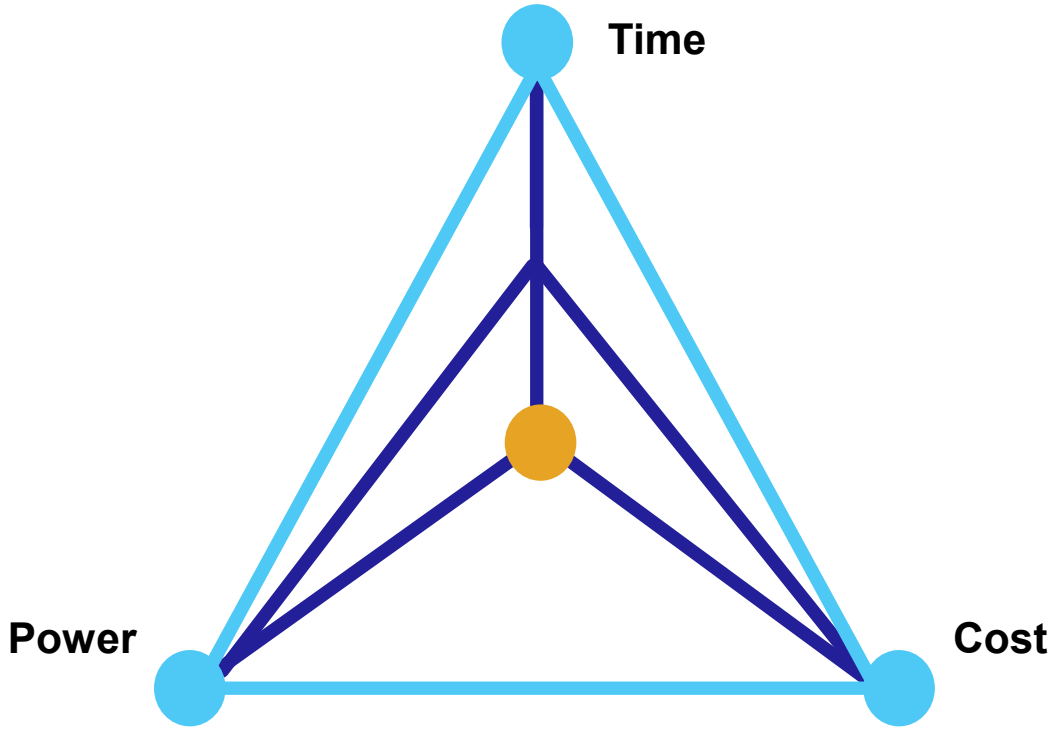
Expanding the Design Space

Initial reference design

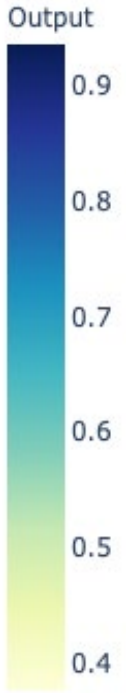


Coloring to highlight strategic priorities

Product Development Team Chooses Relative Weighting of Cost, Time and Power



Designs with darker coloring are more closely aligned with the team's preferences

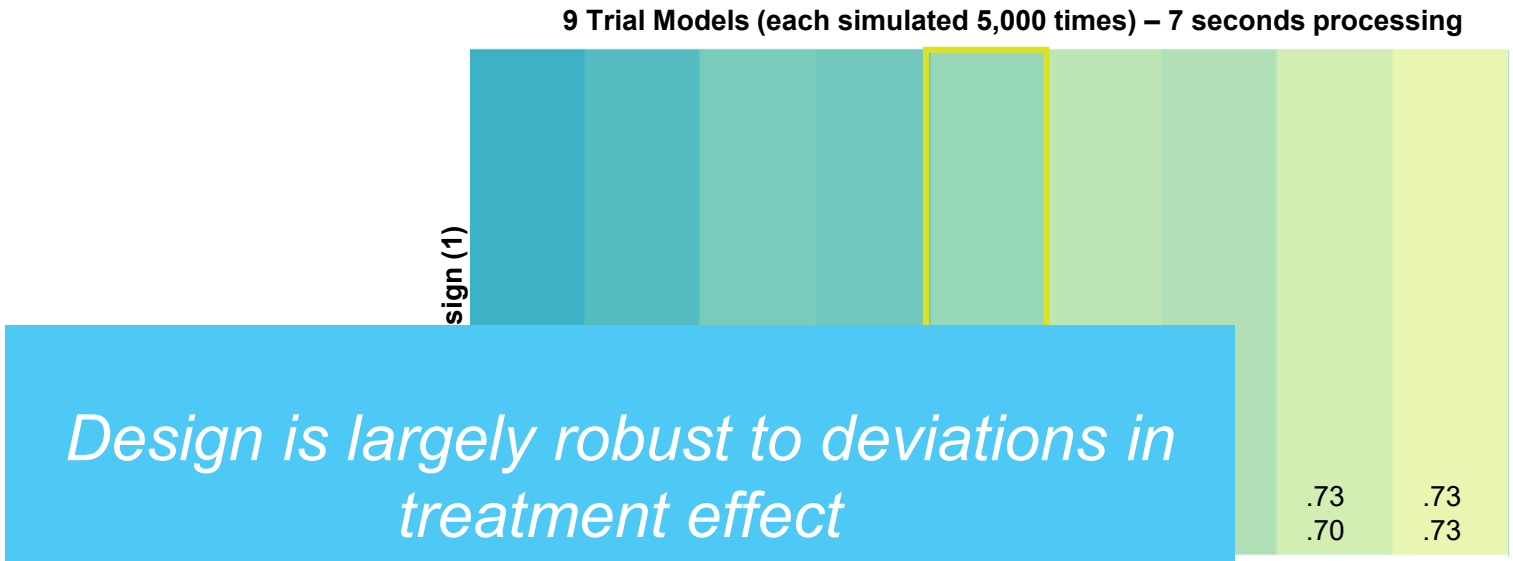


Cytel

What happens if the treatment effect is different than planned?

Design

- Group Sequential Design
- 572 Patients (2:1 allocation)
- 369 total events
- Gam(1.5) Futility interim analysis after 20% events
- LD(OF) Efficacy interim analysis after 40% events



Design is largely robust to deviations in treatment effect

Scenario

- Prevalence (Resistant : Sensitive) 60%:40%
- Hazard Ratio (Resistant) 0.7
- Hazard Ratio (Sensitive) 0.7
- Hazard Ratio (Resistant) 0.67, 0.7, 0.73
- Hazard Ratio (Sensitive) 0.67, 0.7, 0.73
- Control median OS (Resistant) 5 months
- Control median OS (Sensitive) 7 months
- 22 patients/month enrollment (Ramp-up)
- 1% probability of dropout
- Cost per patient $213k
- Time to event endpoint

Power: 87% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

Can we optimize the design by modifying the number of patients or total events?

144 Trial Models (each simulated 5,000 times) – 1 minute processing

Design

- Group Sequential Design
- 570, 550, 472, 610 Patients (2:1 allocation)
- 369, 858, 309, 890 total events
- Gam(1.5) Futility interim analysis after 20% events
- LD(OF) Efficacy interim analysis after 40% events

Opportunity for cost savings at the expense of time and power

Scenario

- Prevalence (Resistant : Sensitive) 60
- Hazard Ratio (Resistant) 0.67, 0.7, 0.73
- Hazard Ratio (Sensitive) 0.67, 0.7, 0.73
- Control median OS (Resistant) 5 months
- Control median OS (Sensitive) 7 months
- 22 patients/month enrollment (Ramp-up)
- 1% probability of dropout
- Cost per patient $213k
- Time to event endpoint

Reference Design

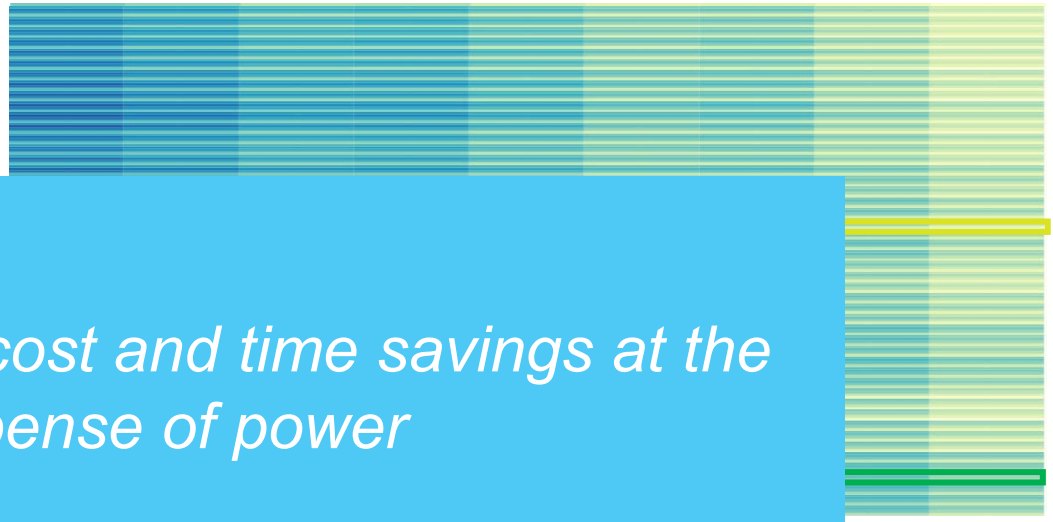
Power: 94% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

New Design

Power: 93% - 77%
Avg. Sample Size: **472-475 pts**
Avg. Duration: **29-30 Months**
Avg. Cost: **$101-102M**

Can we further optimize by modifying the timing of interim analyses?

1,728 Trial Models (each simulated 5,000 times) – 4 minutes processing



Design

- Group Sequential Design
- 500,550, 572, 600 Patients (2:1 allocation)
- 360, 369, 380, 390 total events
- **Gam(1.5) Futility interim analys**
- **LD(OF) Efficacy interim analysi**

Opportunity for cost and time savings at the expense of power

Scenario

- Prevalence (Resistant : Sensitiv
- Hazard Ratio (Resistant) 0.67, 0.7, 0.73
- Hazard Ratio (Sensitive) 0.67, 0.7, 0.73
- Control median OS (Resistant) 5 months
- Control median OS (Sensitive) 7 months
- 22 patients/month enrollment (Ramp-up)
- 1% probability of dropout
- Cost per patient $213k
- Time to event endpoint

Reference Design

Power: 94% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

New Design

Power: 93% - 77%
Avg. Sample Size: **473-481 pts**
Avg. Duration: **26-27 Months**
Avg. Cost: **$102-104M**

Can we further optimize by modifying our approach to stopping rules?

Design

- Group Sequential Design
- 500,550, 572, 600 Patients (2:1 allocation)
- 360, 369, 380, 390 total events
- **Gam(-3, -1, 1.5, 2)** Futility interim analyse
- **LD(OF) & Gam(-4, -3, -2)** Efficacy, futility interim analysis after 40%, 50%, 60%, 7

Scenario

- Prevalence (Resistant : Sensitive
- Hazard Ratio (Resistant) 0.67, 0
- Hazard Ratio (Sensitive) 0.67, 0.7, 0.73
- Control median OS (Resistant) 5 months
- Control median OS (Sensitive) 7 months
- 22 patients/month enrollment (Ramp-up)
- 1% probability of dropout
- Cost per patient $213k
- Time to event endpoint

17,280 Trial Models (each simulated 5,000 times) – 17 minutes processing

Opportunity for cost and time savings with better power

Reference Design

Power: 94% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

New Design

Power: **94% - 82%**
Avg. Sample Size: **492-496 pts**
Avg. Duration: **26-27 Months**
Avg. Cost: **$107-108M**

Process expands the field of options to enable identification of better designs

Reference Design

Power: 94% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

- 572 Patients (2:1 allocation)
- 369 total events
- Group Sequential Design
- Gam(1.5) Futility interim analysis after 20% events
- LD(OF) Efficacy interim analysis after 40% events

Given the team inputs, what is the fastest design?

Balanced Optimized

Power: **94% - 82%**
Avg. Sample Size: **492-496 pts**
Avg. Duration: **26-27 Months**
Avg. Cost: **$107-108M**

- 550 Patients (2:1 allocation)
- 390 total events
- Group Sequential Design
- Gam(2) Futility interim analysis after 20% events
- Gam(-4) Efficacy interim analysis after 50% events

Proposed Designs

Reference Design

Power: 94% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

- 572 Patients (2:1 allocation)
- 369 total events
- Group Sequential Design
- Gam(1.5) Futility interim analysis after 20% events
- LD(OF) Efficacy interim analysis after 40% events

Speed Optimized

Power: **90% - 74%**
Avg. Sample Size: **469-473 pts**
Avg. Duration: **24-25 Months**
Avg. Cost: **$104M**

- 550 Patients (2:1 allocation)
- 360 total events
- Group Sequential Design
- Gam(2) Futility interim analysis after 30% events
- Gam(-2) Efficacy interim analysis after 50% events

Cost Optimized

Power: **91% - 74%**
Avg. Sample Size: **437-445 pts**
Avg. Duration: **26-27 Months**
Avg. Cost: **$96-97M**

- 500 Patients (2:1 allocation)
- 360 total events
- Group Sequential Design
- Gam(2) Futility interim analysis after 30% events
- Gam(-2) Efficacy interim analysis after 40% events

Power Optimized

Power: **96% - 83%**
Avg. Sample Size: **509-542 pts**
Avg. Duration: **28 Months**
Avg. Cost: **$114-116M**

- 600 Patients (2:1 allocation)
- 390 total events
- Group Sequential Design
- Gam(-3.3) Futility interim analysis after 10% events
- Gam(-2) Efficacy interim analysis after 50% events

Balanced Optimized

Power: **94% - 82%**
Avg. Sample Size: **492-496 pts**
Avg. Duration: **26-27 Months**
Avg. Cost: **$107-108M**

- 550 Patients (2:1 allocation)
- 390 total events
- Group Sequential Design
- Gam(2) Futility interim analysis after 20% events
- Gam(-4) Efficacy interim analysis after 50% events

How does our design perform if the drop-out rate is higher than we expect?

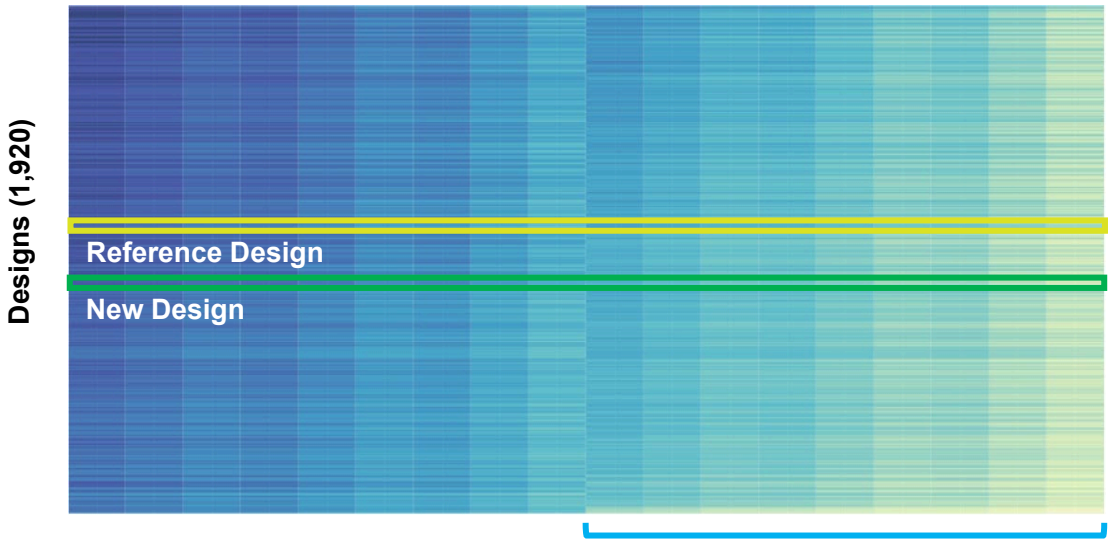
Design

- Group Sequential Design
- 500,550, 572, 600 Patients (2:1 allocation)
- 360, 369, 380, 390 total events
- Gam(-3.3, -1, 1.5, 2) Futility interim analysis after 10%, 20%, 30% events
- LD(OF) & Gam(-4, -3, -2) Efficacy interim analysis after 40%, 50%, 60%, 70% events

Scenario

- Prevalence (Resistant : Sensitive) 60%:40%
- Hazard Ratio (Resistant) 0.67, 0.7, 0.73
- Hazard Ratio (Sensitive) 0.67, 0.7, 0.73
- Control median OS (Resistant) 5 months
- Control median OS (Sensitive) 7 months
- 22 patients/month enrollment
- 1%, 5% probability of dropout
- Cost per patient $213k
- Time to event endpoint

34,560 Trial Models (each simulated 5,000 times) – 15 minutes processing



Reference Design

Power: 94% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

Higher Dropouts

Power: 93% - 79%
Avg. Sample Size: **534-539 pts**
Avg. Duration: **46-51 Months**
Avg. Cost: **$116M-$130M**

Balanced

Power: 94% - 82%
Avg. Sample Size: 492-496 pts
Avg. Duration: 26-27 Months
Avg. Cost: $107-108M

Higher Dropouts

Power: **92% - 77%**
Avg. Sample Size: **513-517 pts**
Avg. Duration: **45-51 Months**
Avg. Cost: **$111-127M**

Power: **94% - 78%**
Avg. Sample Size: **483-505 pts**
Avg. Duration: **24-26 Months**
Avg. Cost: **$107-111M**

Alternative Option
Higher Dropouts

Power: **93% - 79%**
Avg. Sample Size: **507-519 pts**
Avg. Duration: **34-35 Months**
Avg. Cost: **$112-123M**

How does our design perform if the enrollment is faster than expected?

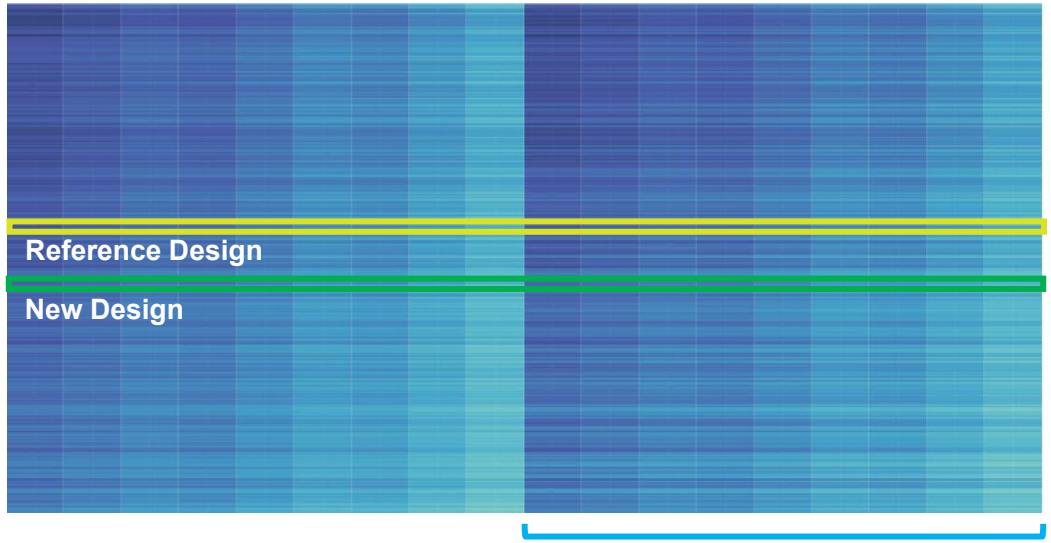
Design

- Group Sequential Design
- 500,550, 572, 600 Patients (2:1 allocation)
- 360, 369, 380, 390 total events
- Gam(-3.3, -1, 1.5, 2) Futility interim analysis after 10%, 20%, 30% events
- LD(OF) & Gam(-4, -3, -2) Efficacy interim analysis after 40%, 50%, 60%, 70% events

Scenario

- Prevalence (Resistant : Sensitive) 60%:40%
- Hazard Ratio (Resistant) 0.67, 0.7, 0.73, 0.76
- Hazard Ratio (Sensitive) 0.67, 0.7, 0.73,0.76
- Control median OS (Resistant) 5 months
- Control median OS (Sensitive) 7 months
- 22, 23, 28 patients/month enrollment
- 22 patients/month enrollment
- 1% probability of dropout
- Cost per patient $213k
- Time to event endpoint

51,840 Trial Models (each simulated 5,000 times) – 20 minutes processing



Reference Design

Power: 94% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

Faster Enrollment

Power: 94% - 79%
Avg. Sample Size: **536-541 pts**
Avg. Duration: **24 Months**
Avg. Cost: **$116M-$117M**

Balanced

Power: 94% - 82%
Avg. Sample Size: 492-496 pts
Avg. Duration: 26-27 Months
Avg. Cost: $107-108M

Faster Enrollment

Power: 94% - 80%
Avg. Sample Size: **514-519 pts**
Avg. Duration: **23 Months**
Avg. Cost: **$111-112M**

Power: **96% - 81%**
Avg. Sample Size: **450-470 pts**
Avg. Duration: **28-30 Months**
Avg. Cost: **$98-101M**

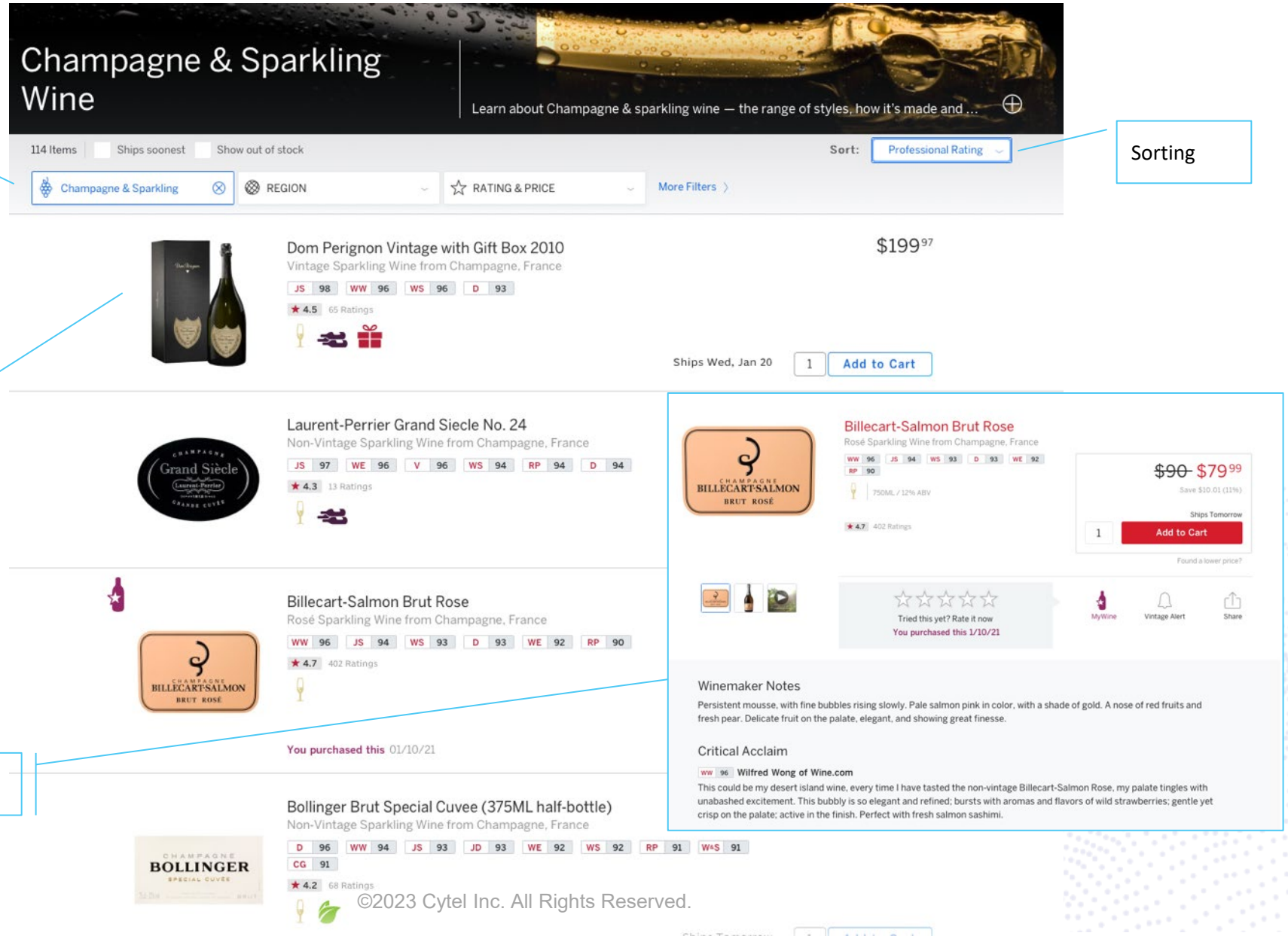
Alternative Option
Faster Enrollment

Power: 96% - 81%
Avg. Sample Size: **468-479 pts**
Avg. Duration: **24-25 Months**
Avg. Cost: **$101-103M**

Cytel

Evaluating the Design Space

Multi-criteria decisions are everywhere



Filters

Sorting

Highest Score

Details

Sensitivity analysis allows team to understand key factors that drive design performance and ensure robustness

Sensitivity analysis of a design across many scenarios of interest is key to avoid failure modes

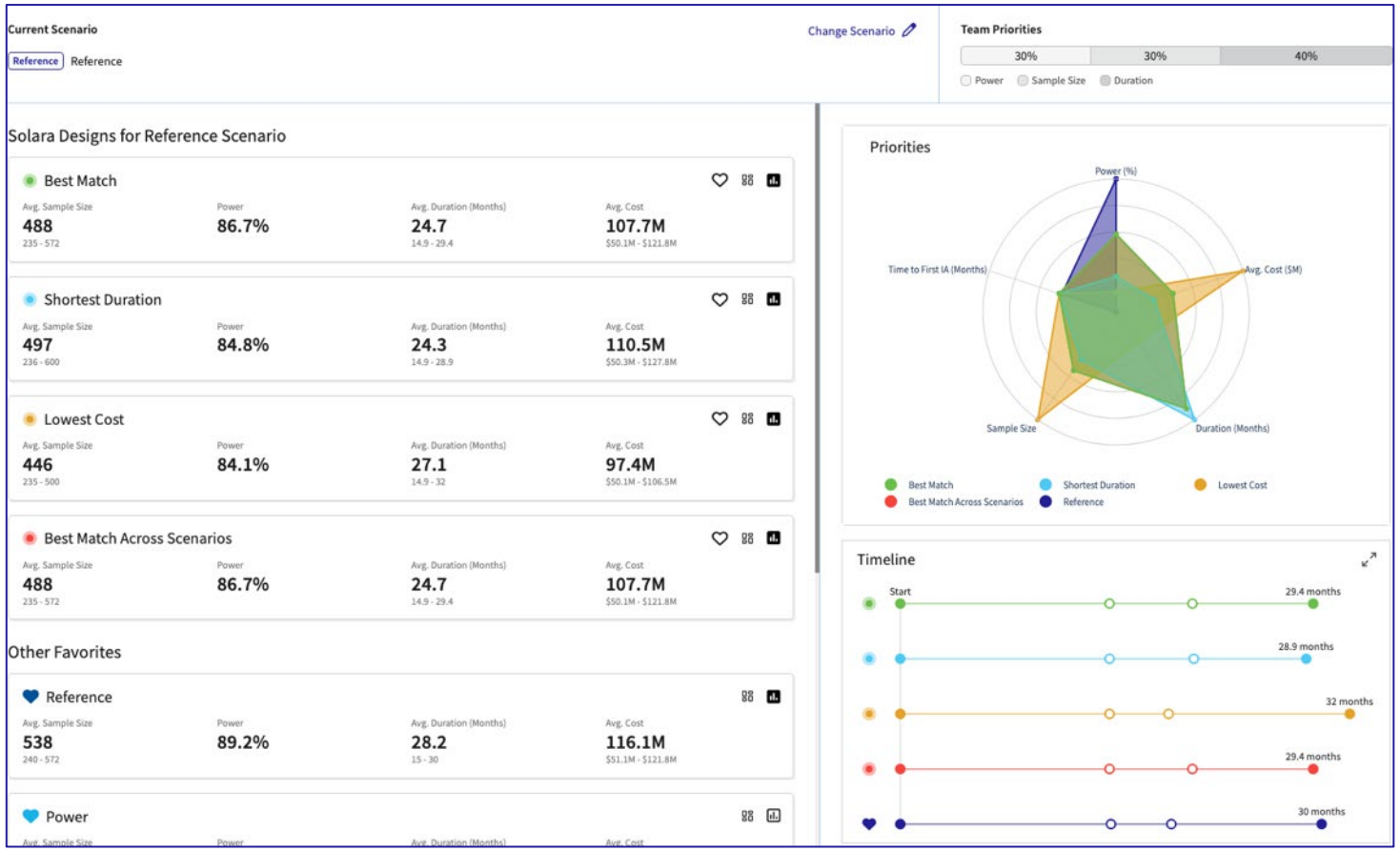
Design robustness can also be measured as a weighted average of model scores and used to identify optimal design candidates

Weights reflect beliefs and interests of the study team

Visualize options and tradeoffs in real time facilitate communication within broader development team

Visualization tools to enable the product development team members to engage with the design space

Pressure Test Designs & Explore the Design Space



Visualization tools provide the clinical development team with the ability to navigate the design space created by the statistician

- Review fastest, cheapest, highest power, balanced options
- Stress test designs (e.g. higher enrollment, higher drop-outs, different treatment effects)
- Compare designs against strategic priorities
- Compare timelines of events

Incorporating team assumptions about the parameters of the design space

Design Space Parameters

Assumptions

Proposed Designs

Reference Design

Power: 94% - 78%
Avg. Sample Size: 529-536 pts
Avg. Duration: 28 Months
Avg. Cost: $114-116M

- 572 Patients (2:1 allocation)
- 369 total events
- Group Sequential Design
- Gam(1.5) Futility interim analysis after 20% events
- LD(OF) Efficacy interim analysis after 40% events

Speed Optimized

Power: **90% - 74%**
Avg. Sample Size: **469-473 pts**
Avg. Duration: **24-25 Months**
Avg. Cost: **$104M**

- 550 Patients (2:1 allocation)
- 360 total events
- Group Sequential Design
- Gam(2) Futility interim analysis after 30% events
- Gam(-2) Efficacy interim analysis after 50% events

Power Optimized

Power: **96% - 83%**
Avg. Sample Size: **509-542 pts**
Avg. Duration: **28 Months**
Avg. Cost: **$114-116M**

- 600 Patients (2:1 allocation)
- 390 total events
- Group Sequential Design
- Gam(-3.3) Futility interim analysis after 10% events
- Gam(-2) Efficacy interim analysis after 50% events

Cost Optimized

Power: **91% - 74%**
Avg. Sample Size: **437-445 pts**
Avg. Duration: **26-27 Months**
Avg. Cost: **$96-97M**

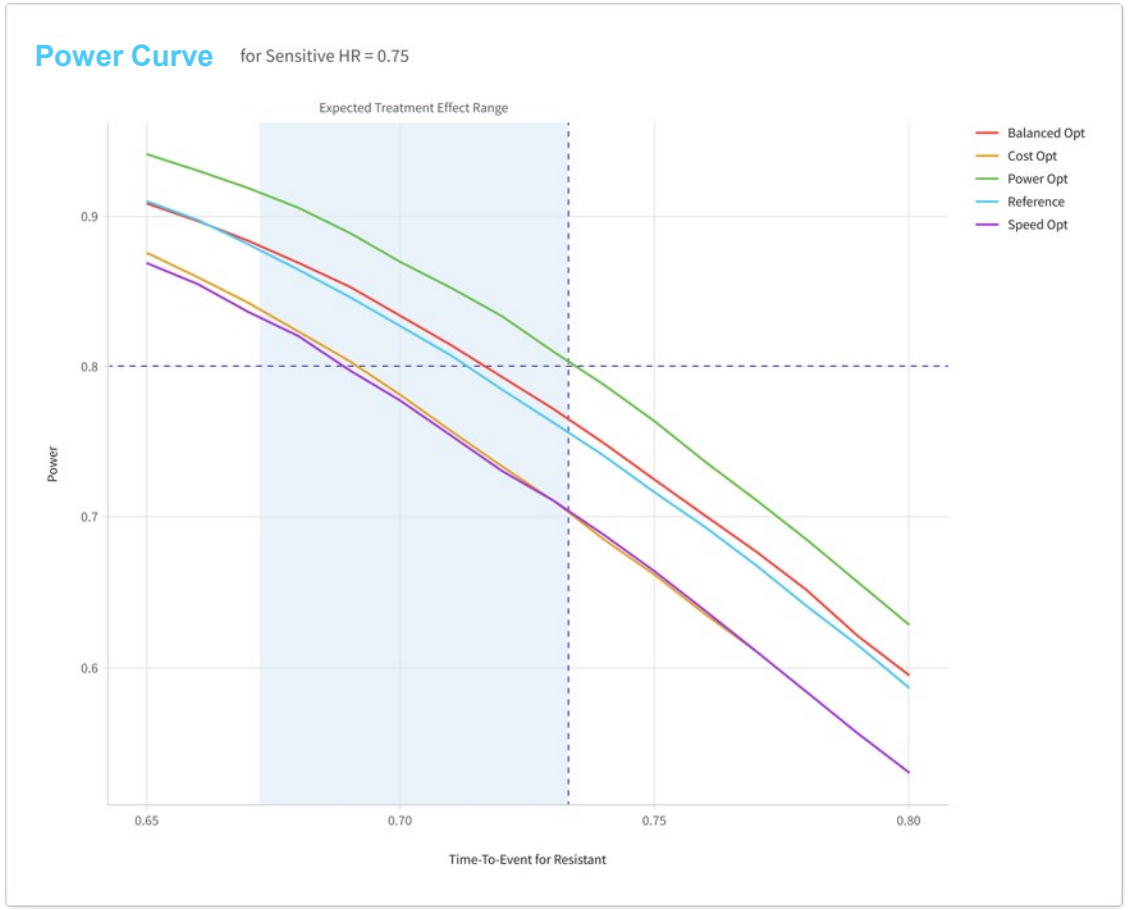
- 500 Patients (2:1 allocation)
- 360 total events
- Group Sequential Design
- Gam(2) Futility interim analysis after 30% events
- Gam(-2) Efficacy interim analysis after 40% events

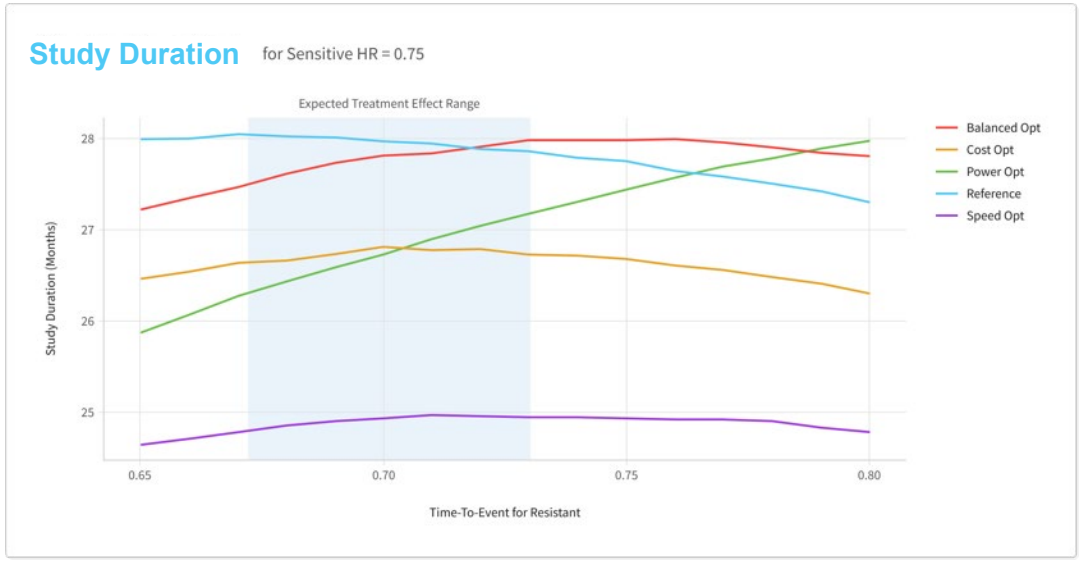
Balanced Optimized

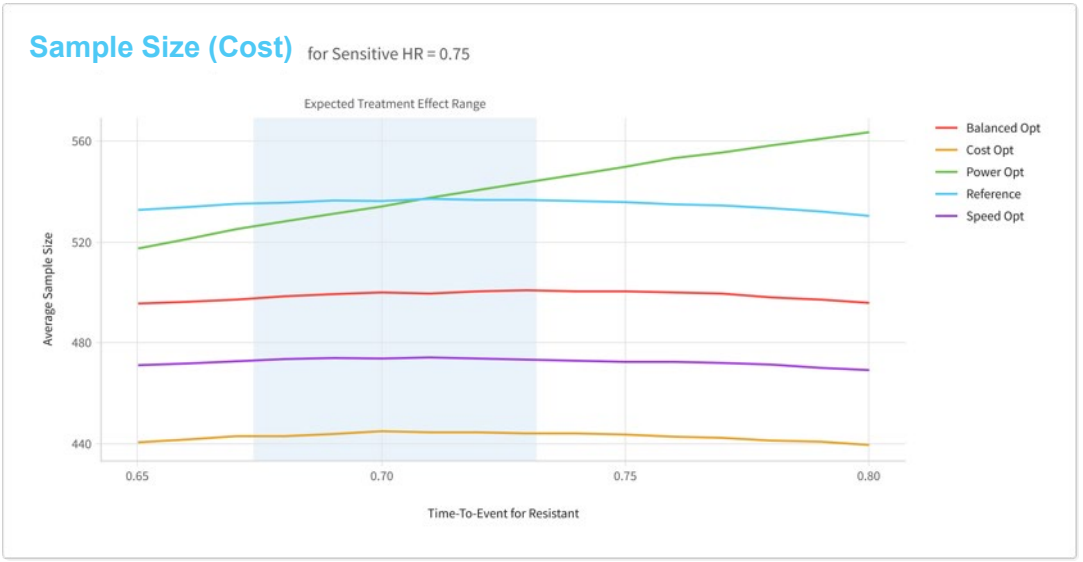
Power: **94% - 82%**
Avg. Sample Size: **492-496 pts**
Avg. Duration: **26-27 Months**
Avg. Cost: **$107-108M**

- 550 Patients (2:1 allocation)
- 390 total events
- Group Sequential Design
- Gam(2) Futility interim analysis after 20% events
- Gam(-4) Efficacy interim analysis after 50% events

Understanding the robustness of designs to different risks



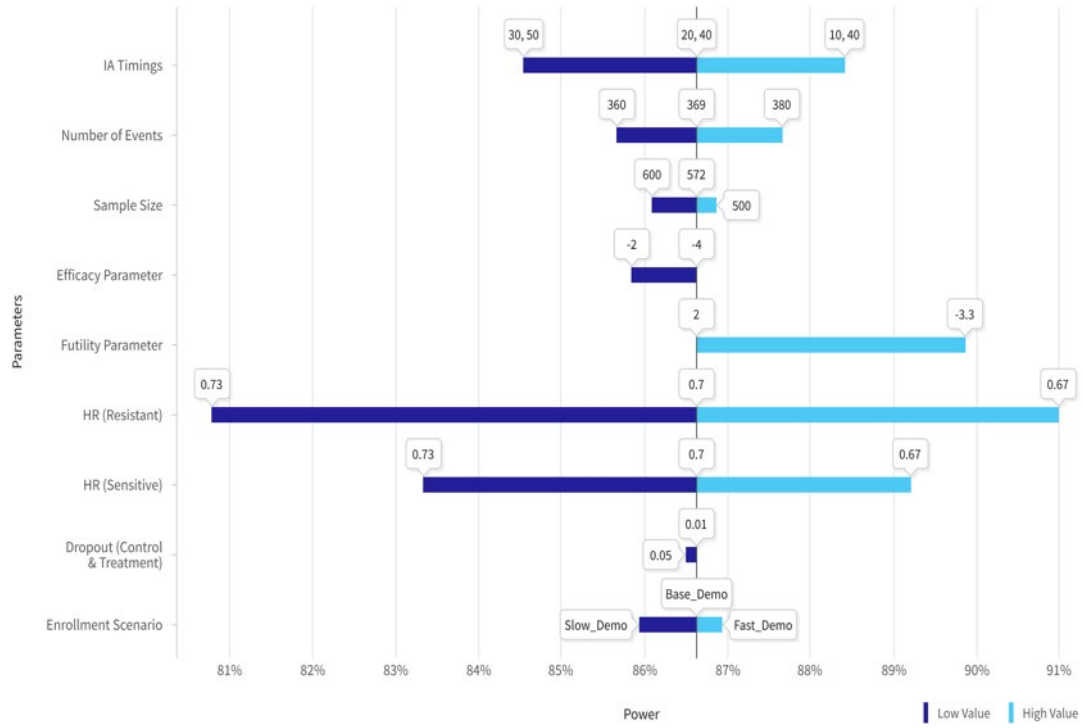




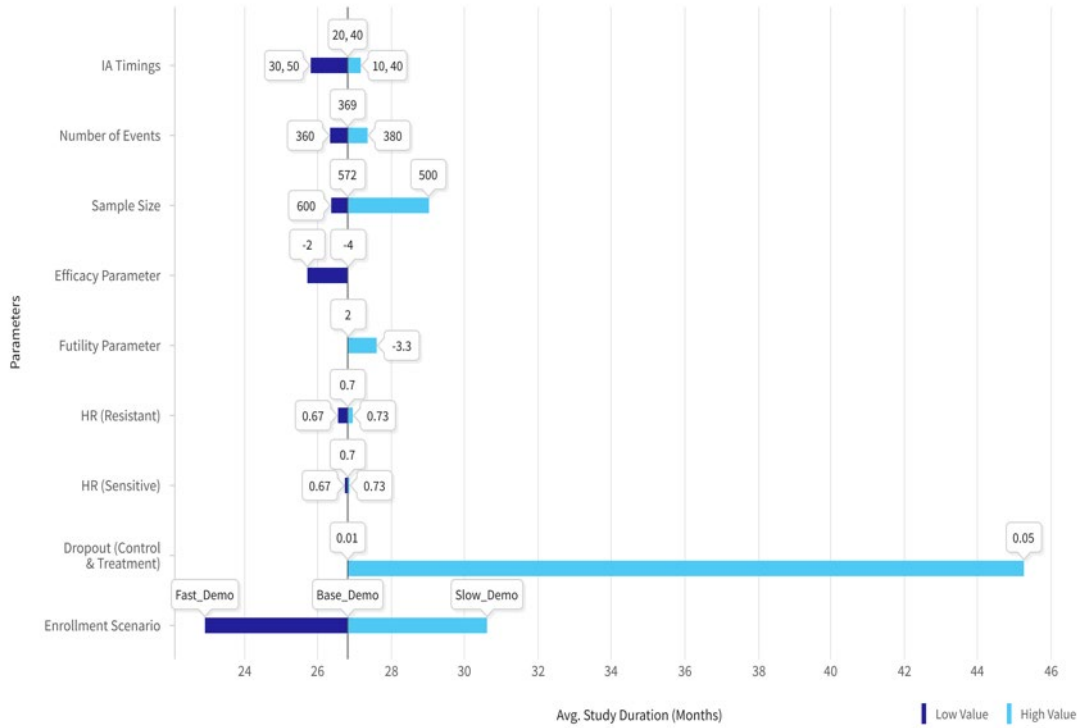
To provide 80%+ power in HR =.73 for the reference, speed and cost optimized designs would require 20-30 more events (requiring 50-100 more subjects to retain timelines)

Continued: understanding the robustness of designs to different risks

Parameters vs. Power

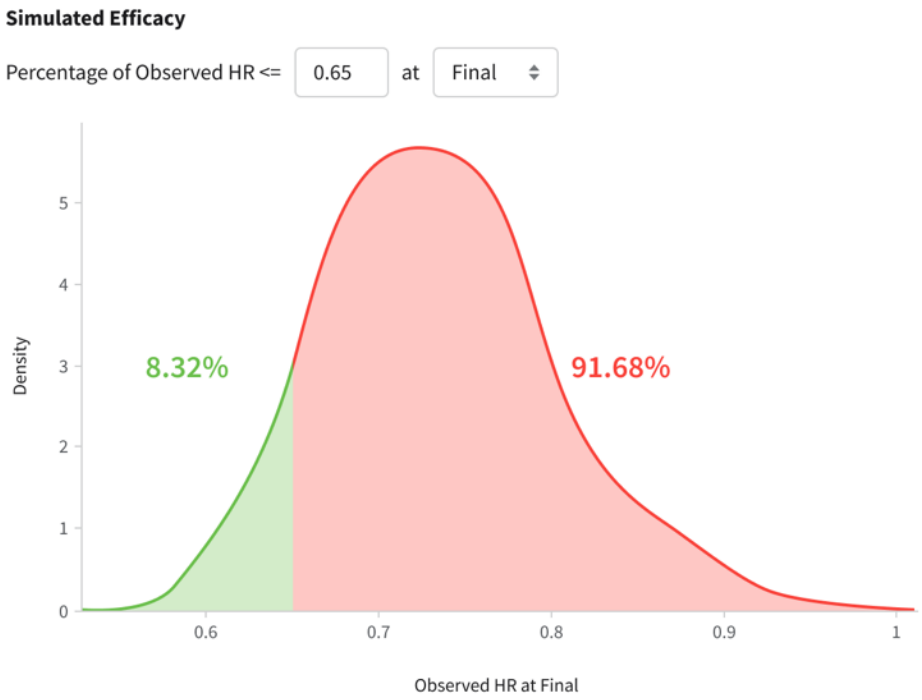


Parameters vs. Average Study Duration

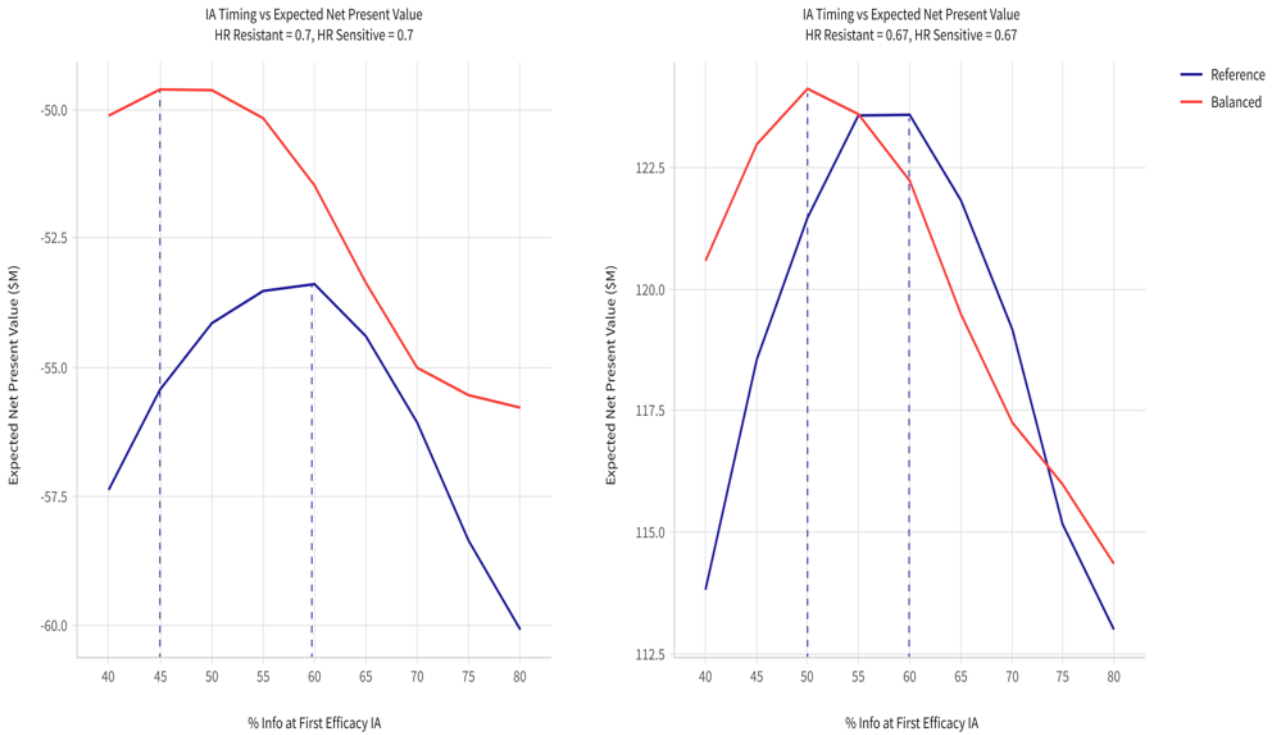


Making decisions with confidence based upon an understanding of the entire relevant design space

Probability of Observing Target Value



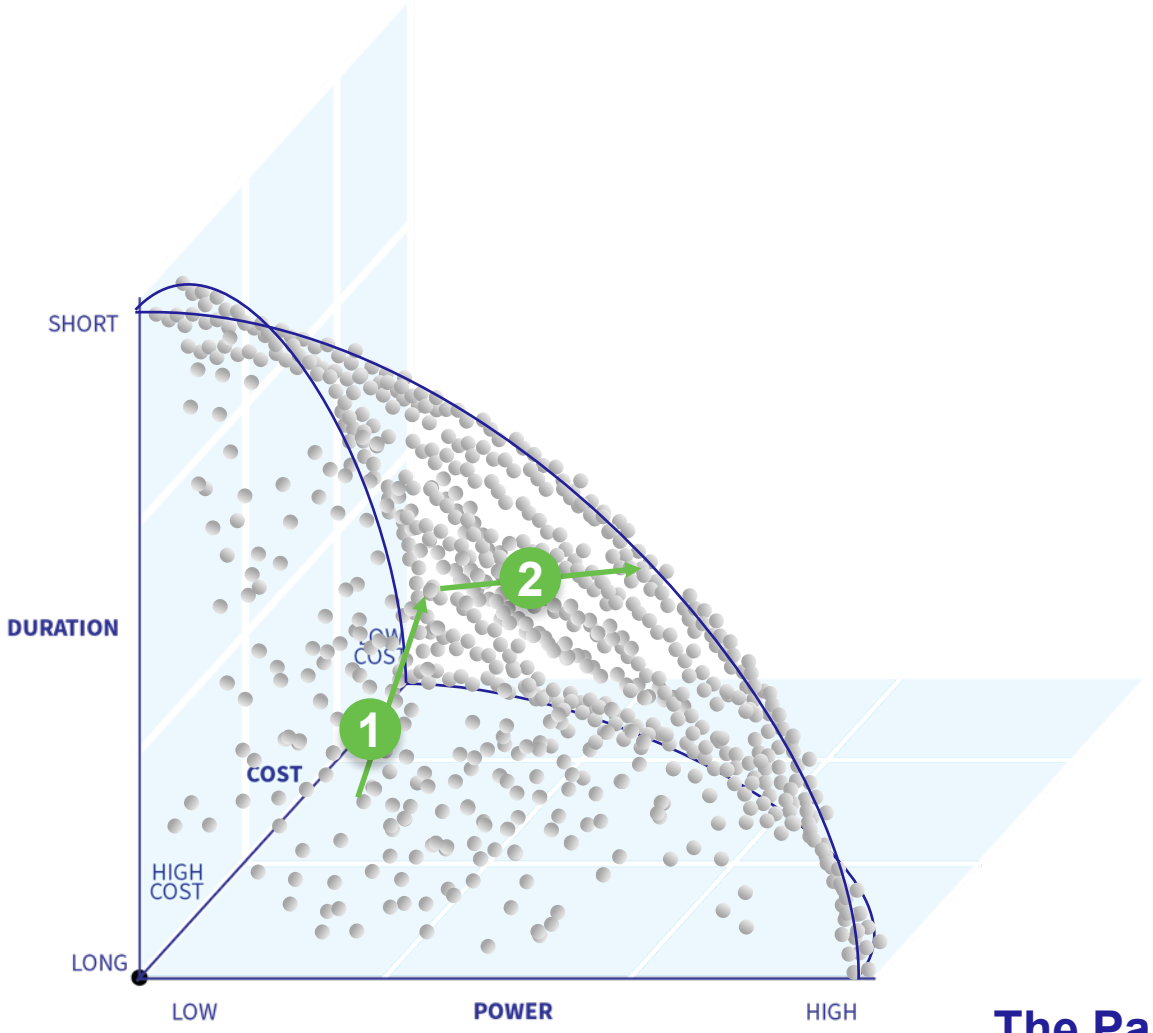
Optimizing for an Early Readout



Balancing timing of interim analysis and probability of declaring success can be optimized through the lens of eNPV

Pareto optimizing a trial design

A data-driven approach to ensuring trial optimization: thousands of design variations can be quickly and easily analyzed



Objective 1

Reach the Pareto Frontier: Improve power, cost or duration without sacrificing anything

Objective 2

Balance power and duration

Further optimize [balance] by discussing tradeoffs with cross-functional teams

- Better alignment with sponsor strategy
- Include concerns of Clin. Ops.
- Be valuable adviser for sponsor

The Pareto Frontier

Challenges and limitations to simulations

Accuracy

Garbage in, garbage out – importance of quality data

Overreliance

Simulations are tools, not replacements for real-world trials

Complexity

Advanced simulations require sophisticated models and expertise

Overcoming Challenges

Embed into organization processes

Educate stakeholders

Standardize technology

“All models are wrong, but some are useful”

- George E. P. Box

Advantages of large-scale simulation in clinical development

1. **Enhanced Efficiency:** Faster hypothesis testing and strategy formulation
2. **Risk Reduction:** Identifying potential pitfalls or difficulties before they happen
3. **Informed Decisions:** Supporting evidence-based choices for trial outcomes.
4. **Cost-effectiveness:** Less resources wasted on ineffective strategies or unexpected problems
5. **Ethical Considerations:** Ensuring patient safety and well-being

Cytel

Thank You!