Using Prediction and Simulation to Guide Clinical Trial Design

J. Kyle Wathen
Scientific Director
Model Based Drug Development
kwathen@its.jnj.com

janssen
PHARMACEUTICAL COMPANIES
OF Johnson&Johnson

BASS 2014

Outline

- Simulation guided design of two different trials
- Consideration of all available information
- Example for calculating the likelihood of success
- Conclusions

Simple Idea

- Using historical database we have today on similar treatments/populations to plan 1 or 2 studies to gather information for planning phase III
- Use information at the end to plan phase III
- Want the best chance of identifying the “best” dose(s) for phase III
- How should we proceed?

Trial Planning Stage
Phase II b

- Preliminary data is available from POC
- Historical database on similar treatments
  - Utilize to identify population for phase II b
  - Estimate safety and efficacy
  - The target population is different
- Three doses to maximize chance of identifying safe and effective dose for further development

Safety Outcomes

- Two safety outcomes are negatively correlated
- Safety outcome 1 (SO1)
  - Continuous outcome
  - Increase is good, decrease may harmful
  - Measured for 48H after treatment
- Safety outcome 2 (SO2)
  - Continuous outcome
  - Decrease is good, increase is harmful
  - Measured for 48H after treatment

Efficacy Outcome

- Binary outcome
  - Treatment failure rate (TFR) measured at day 30
- Goal is to decrease TFR by 25%

Trial Design

- Resources for treating approximately **450** patients
- Need sufficient information for planning a phase III study
- Option 1
  - Dose finding study to identify MTD
  - Run a two arm randomized study of MTD vs Placebo monitoring efficacy
- Option 2
  - Randomized study monitoring both safety outcomes and efficacy outcome

Option 1

- Phase I
  - Typically a small number of patient (30-60) are utilized to identify MTD
  - Collect safety data but not efficacy data
- Phase II
  - MTD is used for a randomized phase II study
  - Monitor efficacy outcome and positive run phase III

Option 1 – Pros/Cons

- Pros
  - Easy to do
- Cons
  - What if a lower dose could also be effective?
  - Cannot use efficacy information on patients enrolled in phase I
  - What if the MTD must be stopped in phase II for safety considerations (not formerly monitoring but DSMB is monitoring)
  - Must consider the likelihood of getting to phase 3 with a dose that is unsafe or not effective

Option 2

- Combined phase II b
  - Based on historical data all doses are safe in similar populations
- Initial patients are randomized between three doses and placebo
  - Compare mean SO1 and SO2 to placebo for each dose
  - Drop a dose if it appears to be harmful (SO1/SO2)
  - Collect efficacy outcome on all patients
  - Drop dose(s) for lack of efficacy

Option 2 – Pros/Cons

- Pros
  - Formally monitor efficacy and safety
  - If a higher dose is found to be unsafe at any point lower doses still have information
  - Could identify multiple doses (safe and effective) at the end
- Cons
  - More complicated
  - No “off-the-shelf” solution
  - Need Operating Characteristics (OCs)

Initial Design – Option 1

- Use the first 100 patients to collect safety information
- Select the highest dose that appears to be safe
- Use 350 patients to randomize between selected dose and placebo comparing the treatment failure rate (TFR)
  - No formal safety rules in the evaluation of OCs
- Simple Idea- We need to evaluate likelihood of getting to phase III with the “right” dose
  - Simulation!

Initial Design

- Only had a 40% chance of selecting the best/good dose
- Now what?
  - Evaluated design option 2 under the assumption of 450 patient max.
  - Could increase likelihood of selecting best/good dose to around 60%
  - Team proposed increasing sample size to around 1000 to see if we could increase this to 80%

Standard Design – 2 Stage

- Stage 1 – Enroll 100 patients, select the best dose and enroll up to the total sample size to get the various powers
- BIG simplify assumption – we ALWAYS select the "best" dose for stage 2
- Required Sample size for overall one-sided $\alpha = 0.05$

Placebo Rate	Active Rate	Percent Reduction	Total Sample Size 80% Power	Total Sample Size 70% Power	Total Sample Size 60% Power
40%	30%	25%	802	650	536
30%	22.5%	25%	1182	952	778
20%	15%	25%	1940	1554	1260

Option 2 - Details

- After 40 patients are enrolled begin monitoring safety (SO1 and SO2)
  - If at any point is very likely that a dose causes safety concerns on SO1 AND SO2 drop the dose from the trial
- After 100 patients per arm begin monitoring efficacy and compare each dose to placebo
  - If it is unlikely that a dose lowers the TFR when compared to placebo drop the dose
- Patients are equally randomized between all open doses

Bayesian Modeling

For treatment t,

We assume

(SO1, SO2) ~ MVN(μ_t, Σ)

Allows us to capture correlation between SO1 and SO2

Denote the TFR on t by π_t ~ Beta(0.6, 1.4)

- Very BASIC model for TFR – at this point

Design - Simulation V1.0

Stopping Rules V1.0

- Safety rules based on SO1 (decrease by 5.5) and SO2 (Increase by more than 10%)
  - SO1 and SO2 are modeled jointly to account for correlation
- Stop the trial if the probability that P has the lowest TFR is greater than 90%.
- Drop a dose if it is unlikely that it has a lower TFR than placebo
- Drop a dose if it is unlikely the dose with the lowest TFR.

Simulated Scenarios

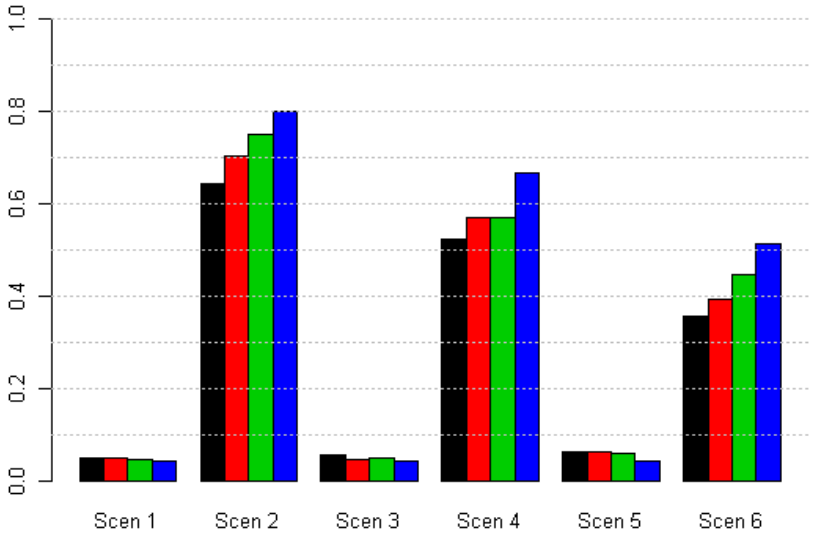
- In these scenarios all doses are safe –
- TFR = Treatment failure rate
- Results that are presented – Probability of selecting the best dost, average number of patients that received the best treatment and average total sample size.
- Randomization to placebo fixed at 20%

Scenario	Dose	True TFR
1 – Null 1	P	0.4
	D1, D2, D3	0.4
2 – Alt 1	P	0.4
	D1,2	0.4
	D3	**0.3**
3 – Null 2	P	0.3
	D1, D2, D3	0.3
4 – Alt 2	P	0.3
	D1, D2	0.3
	D3	**0.225**
5 – Null 3	P	0.2
	D1, D2, D3	0.2
6 – Alt 3	P	0.2
	D1, D2	0.2
	D3	**0.15**

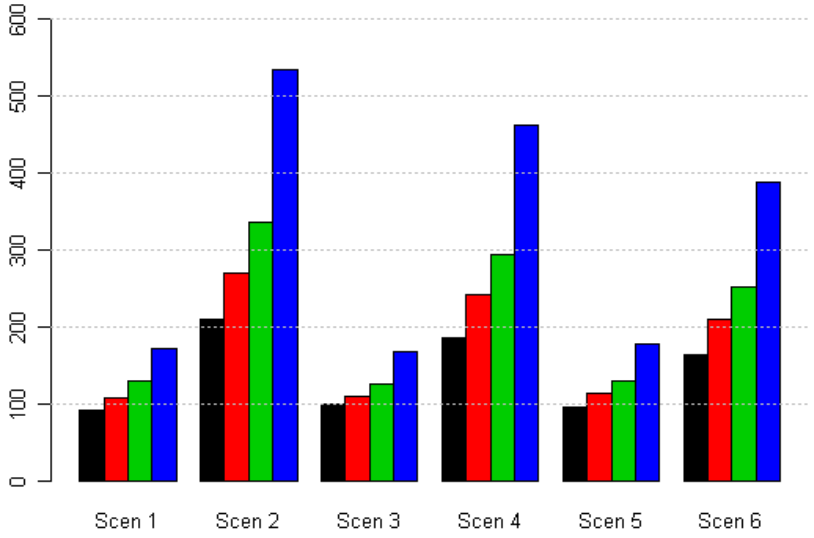
Adjustments?

- Given the power using a standard design should we consider a larger trial?
- Early dropping of a dose that is unlikely to be selected at the end of the trial?
- Early superiority for a dose?

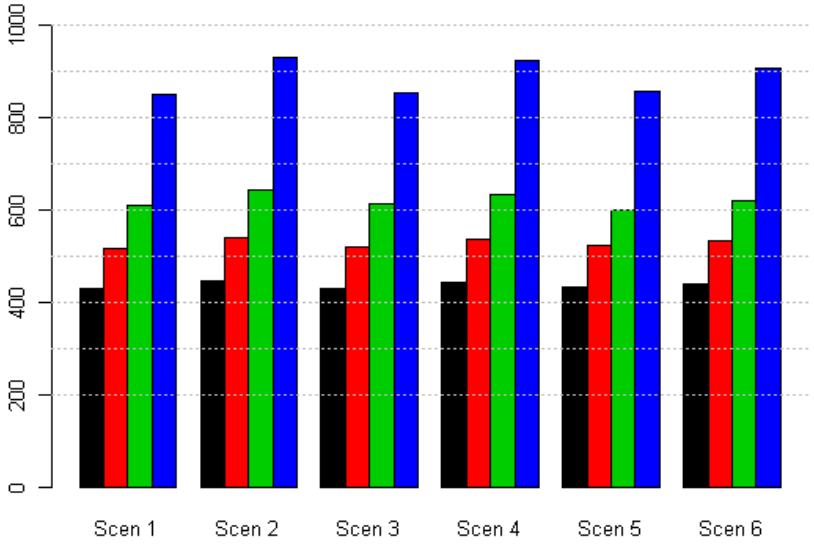
Power



Ave. Num. Pats On Best Dose

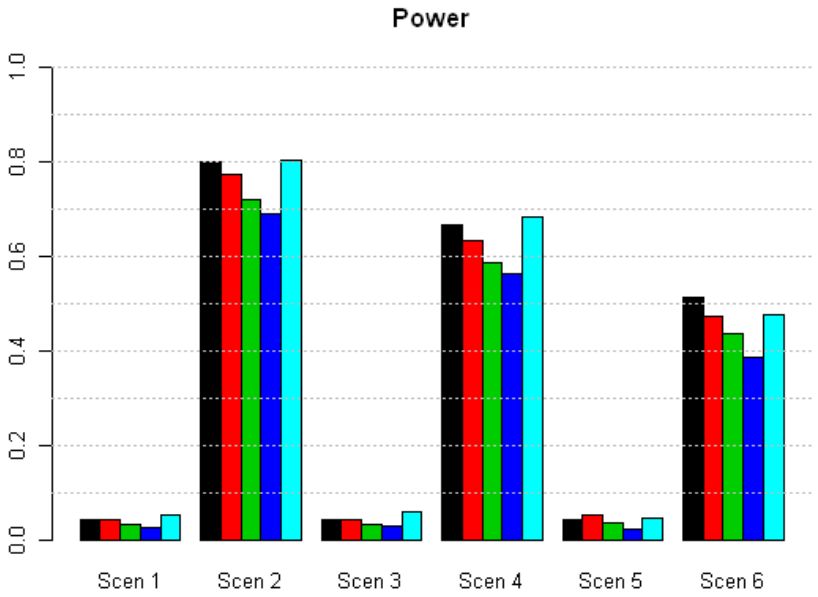


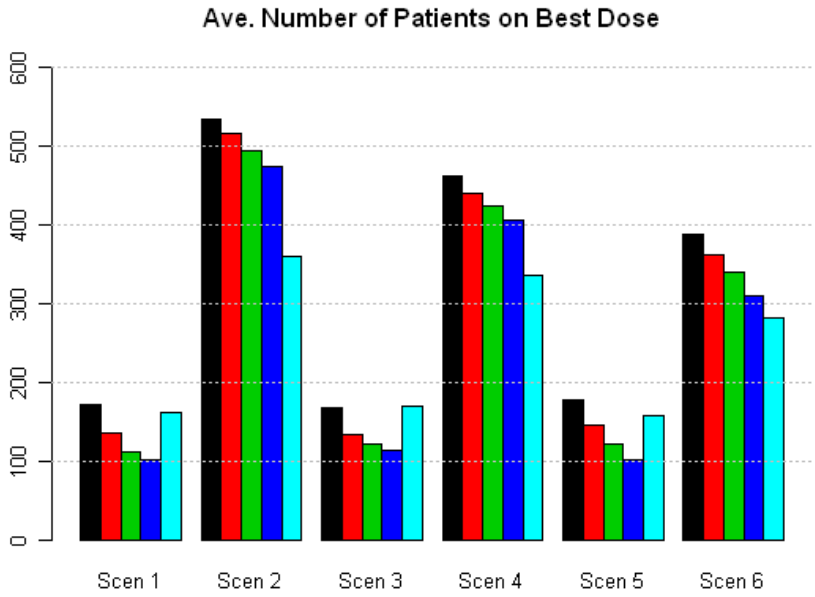
Ave Total Sample Size

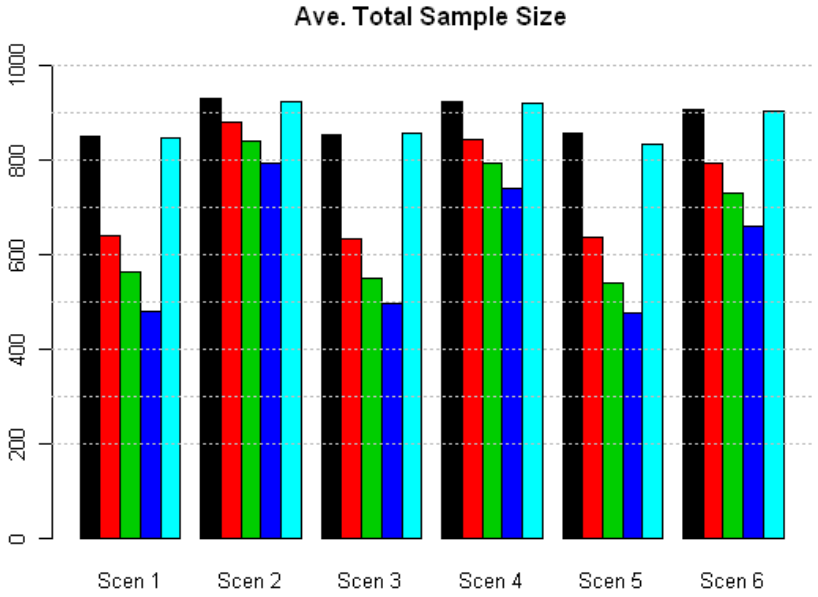


Stopping Rules V1.1

- Safety rules based on SO1 (decrease by 5.5) and SO2 (Increase by more than 10%)
  - SBP and HR are modeled jointly to account for correlation
- Stop the trial if the probability that P has the lowest TFR is greater than 90%.
- Drop a dose if it is unlikely that it has a lower TFR than placebo
- Drop a dose if it is unlikely the dose with the lowest TFR.
- Drop a dose if the Bayesian predictive probability of selecting it at the end of the study is less than 10%


Power
0.0
0.2
0.4
0.6
0.8
1.0
Scen 1
Scen 2
Scen 3
Scen 4
Scen 5
Scen 6


Ave. Number of Patients on Best Dose
0
100
200
300
400
500
600
Scen 1
Scen 2
Scen 3
Scen 4
Scen 5
Scen 6

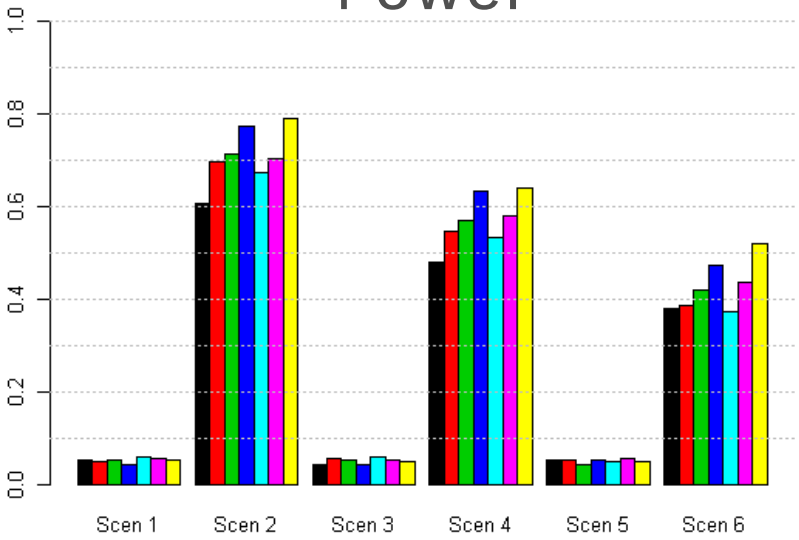

Ave. Total Sample Size
0
200
400
600
800
1000
Scen 1
Scen 2
Scen 3
Scen 4
Scen 5
Scen 6


Design 6.0.950
Design 6.1.950
Design 6.2.950
Design 6.3.950
Design 6.ER.950

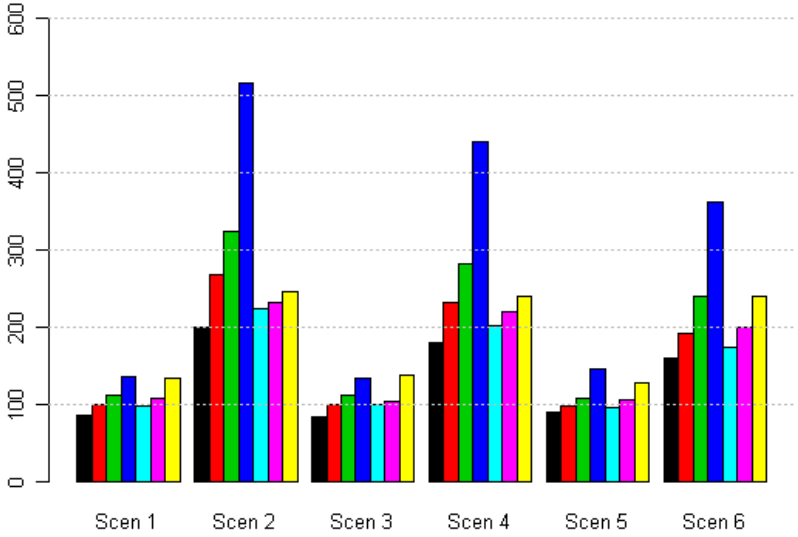
Stopping Rules V1.2

- Safety rules based on SO1 (decrease by 5.5) and SO2 (Increase by more than 10%)
- Stop the trial if the probability that P has the lowest TFR is greater than 90%.
- Drop a dose if it is unlikely that it has a lower TFR than placebo
- Drop a dose if it is unlikely the dose with the lowest TFR.
- Drop a dose if the Bayesian predictive probability of selecting it at the end of the study is less than 10%
- Select a dose early if the probability that it is the best dose (and safe) is greater than 90%

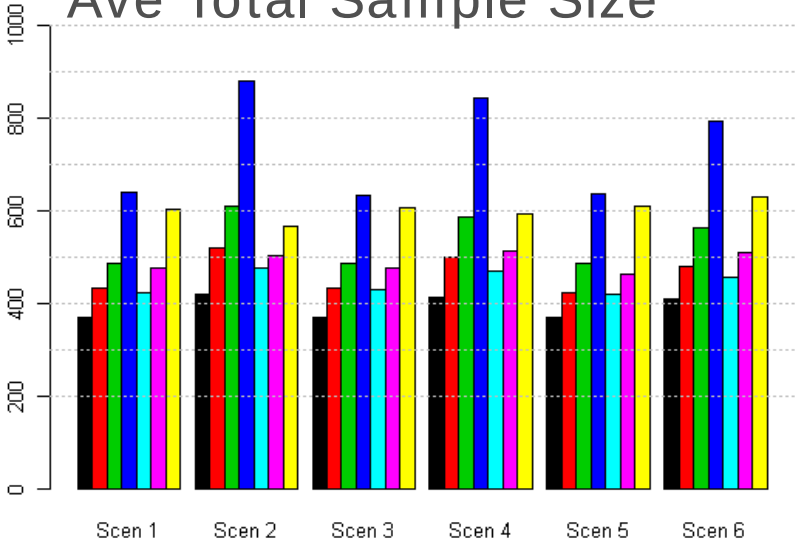
Power



Ave. Num. Pats On Best Dose



Ave Total Sample Size



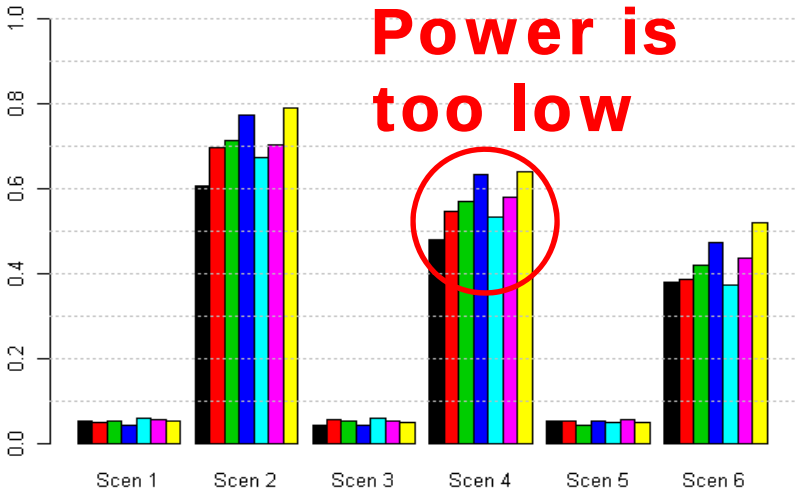
- Design 6.1.450
- Design 6.1.550
- Design 6.1.650
- Design 6.1.950
- Design 7.1.550
- Design 7.1.650
- Design 7.1.950

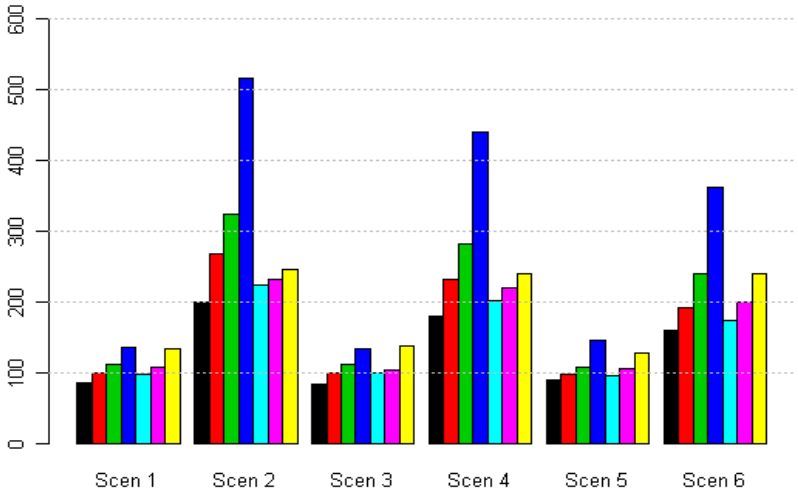
Design 8 – Changes from 7

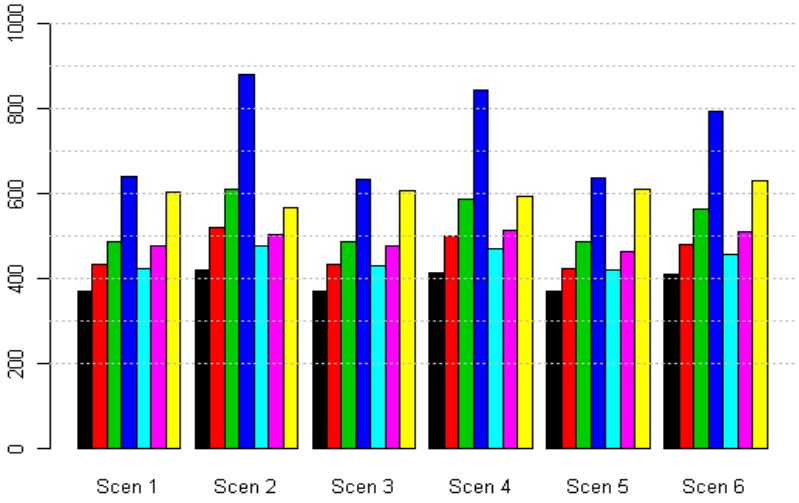
- False positive increase from 5% to 10%
- No “aggressive” dropping rule.
- TFR – Dropping rules are based on predictive probability of success at the end of the study or Pr(placebo has the lowest TFR)

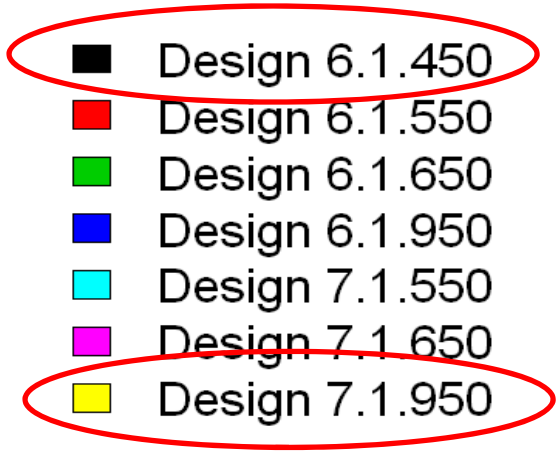
Stopping Rules V1.3

- Safety rules based on SO1 (decrease by 5.5) and SO2 (Increase by more than 10%)
- Stop the trial if the probability that P has the lowest TFR is greater than 75%.
- Drop a dose if it is unlikely that it has a lower TFR than placebo
- ~~Drop a dose if it is unlikely the dose with the lowest TFR.~~
- Drop a dose if the Bayesian predictive probability of selecting it at the end of the study is less than 30%
- Select a dose early if the probability that it is the best dose (and safe) is greater than 90%
- At the end of the study select a dose if the probability that it is the best dose is greater than 60%

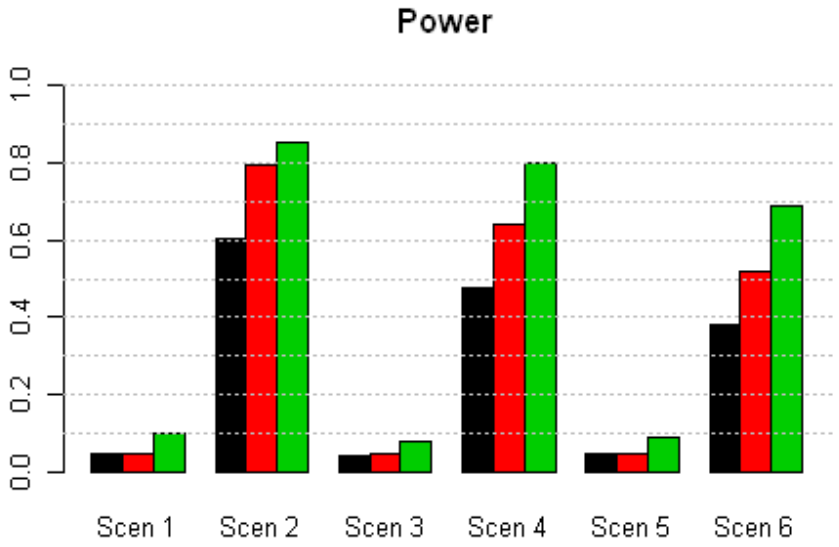

Power is too low
0.0
0.2
0.4
0.6
0.8
1.0
Scen 1
Scen 2
Scen 3
Scen 4
Scen 5
Scen 6

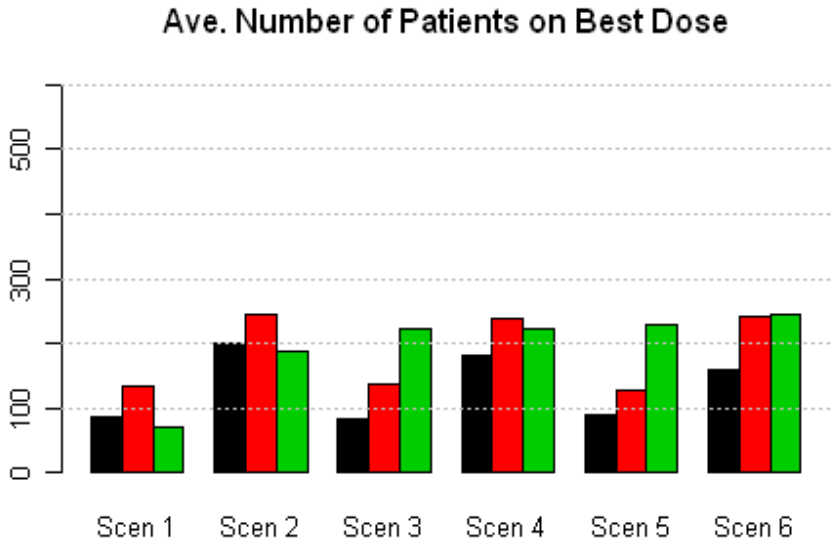

0
100
200
300
400
500
600
Scen 1
Scen 2
Scen 3
Scen 4
Scen 5
Scen 6

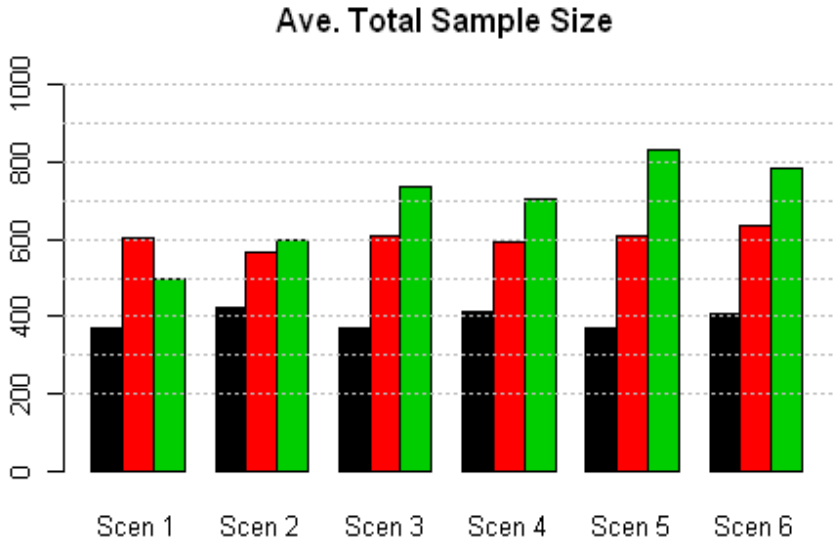

0
200
400
600
800
1000
Scen 1
Scen 2
Scen 3
Scen 4
Scen 5
Scen 6


Design 6.1.450
Design 6.1.550
Design 6.1.650
Design 6.1.950
Design 7.1.550
Design 7.1.650
Design 7.1.950

Comparing Designs

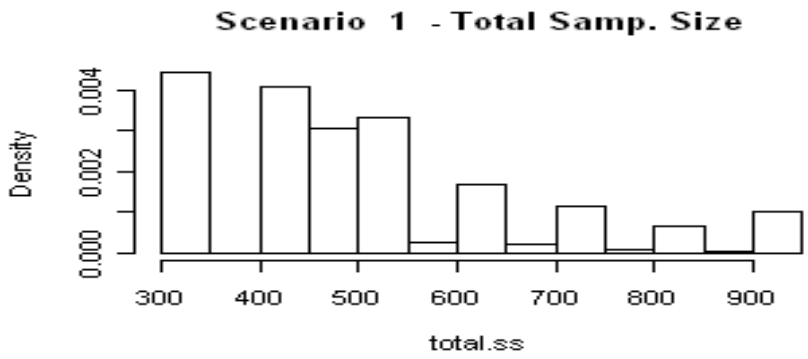


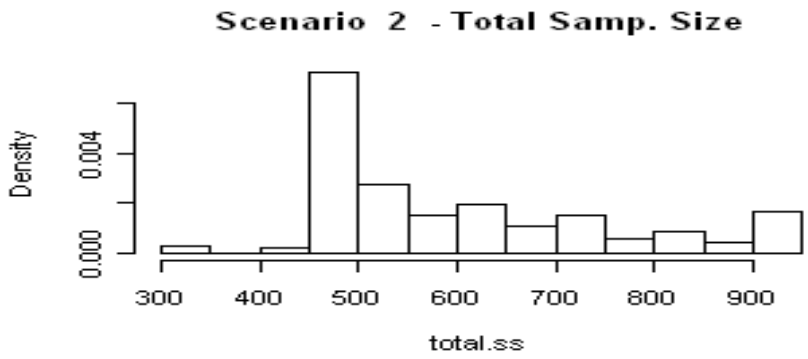


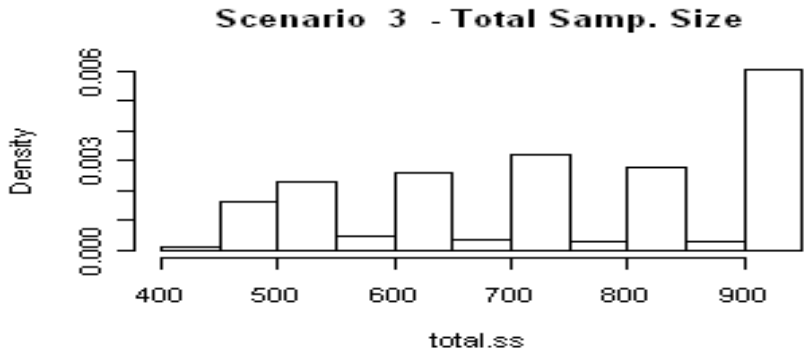


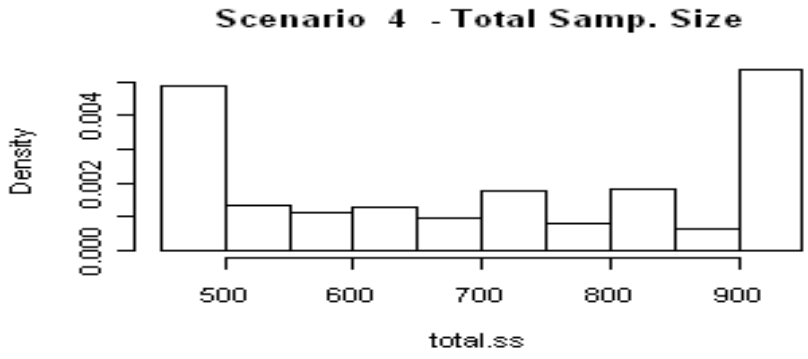
Design 6.1.450
Design 7.1.950
Design 8.3.950

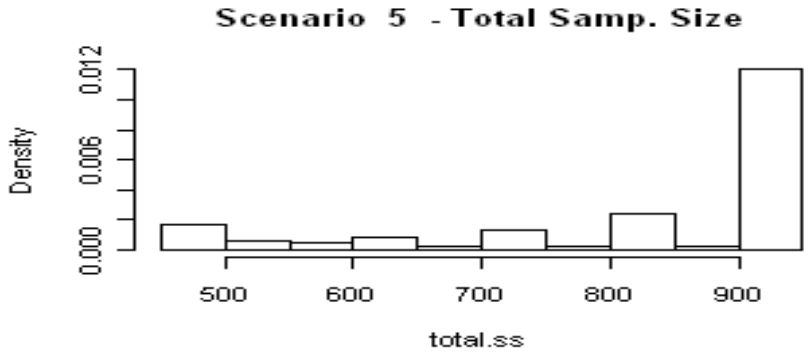
Sample Size – Max 950

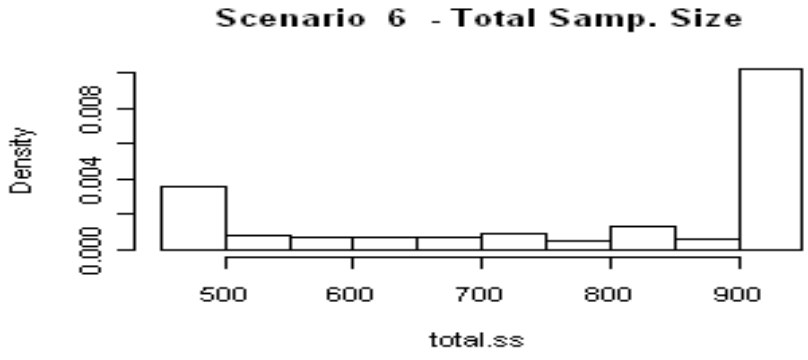




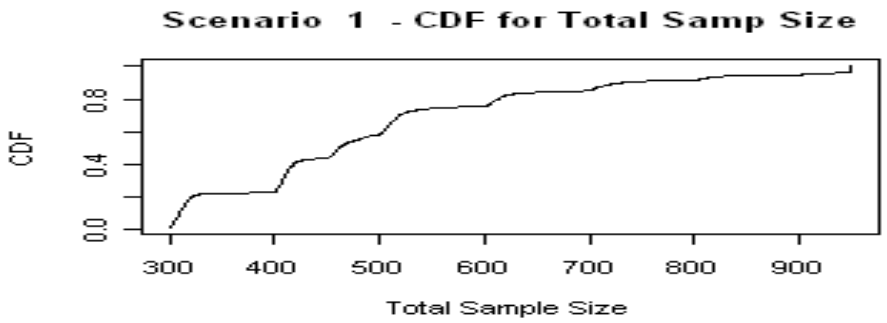


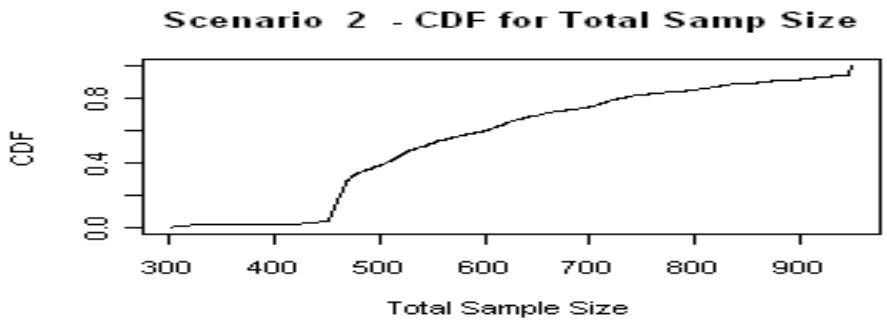


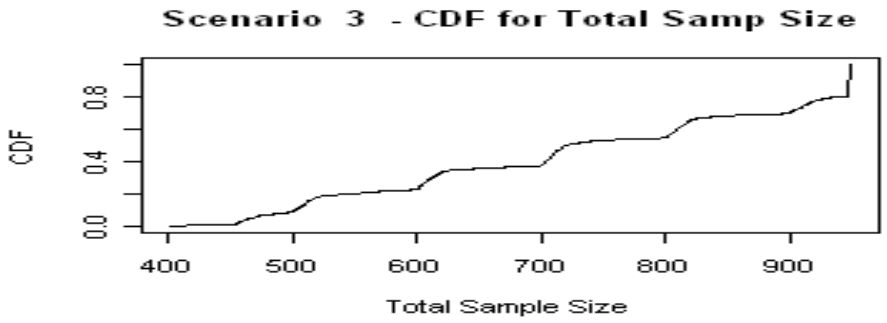


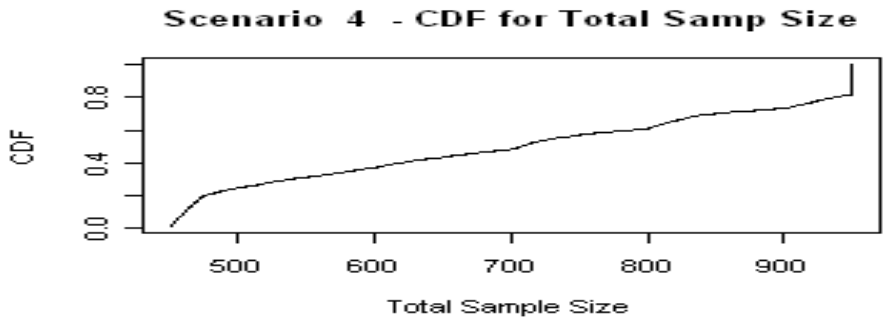


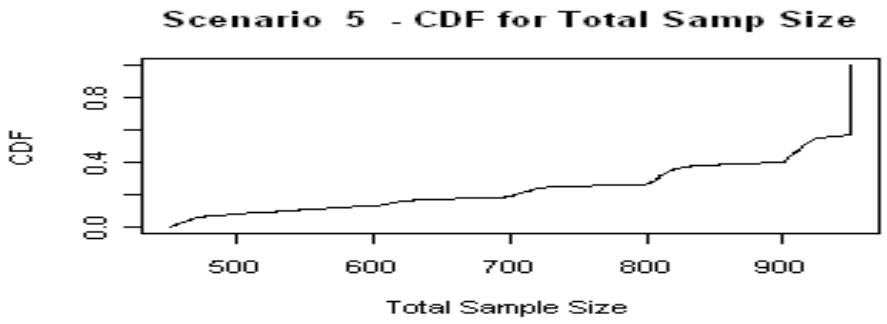
Sample Size CDF

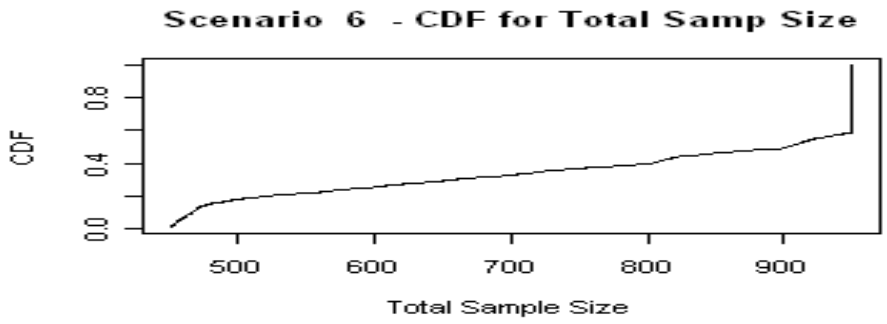












Design 9

Stopping Rules V1.4

- Safety rules based on SO1 (decrease by 5.5) and SO2 (Increase by more than 10%)
  - SO1 and SO2 are modeled jointly to account for correlation
- Drop a dose if it is unlikely that it will be selected at the end of the study. This decision is based on a **Bayesian predictive probability**
  - If Pr[**Pr(TFR on Dose < TFR on Placebo | Data at the end) >0.875)** < 0.15 – Drop the dose
- At the end of the study "select" a dose if Pr(TFR on Dose < TFR on Placebo | Data) > 0.875
- Randomization will be equal among all open doses

Conclusion – Example 1

- By monitoring by safety and efficacy in a trial with multiple doses we can improve the likelihood of getting the best/good dose(s) for phase III
- Using simulation to guide the design process can greatly increase the likelihood of success

Example 2

Motivating Trial

- Outcomes – Progression Free Survival (PFS) and Overall Survival (OS)
  - Standard initial design was approved – simulations would not provide much additional information
  - N=760, 80% power, 5% false-positive
- New information available
  - Correlation between OS and PFS was higher than originally believed
  - Good/Poor prognosis patients responded very differently to SOC
  - Delay of treatment effect - 6-24 months
  - Would need positive result in PFS with a “positive” result in OS

Design Options

- **OPTION 1:** All comers ignore risk group and only test overall

- **OPTION 2:** Good prognosis only
  - Smaller market, slower accrual

- **OPTION 3:** Poor prognosis only
  - Smaller market, slower accrual

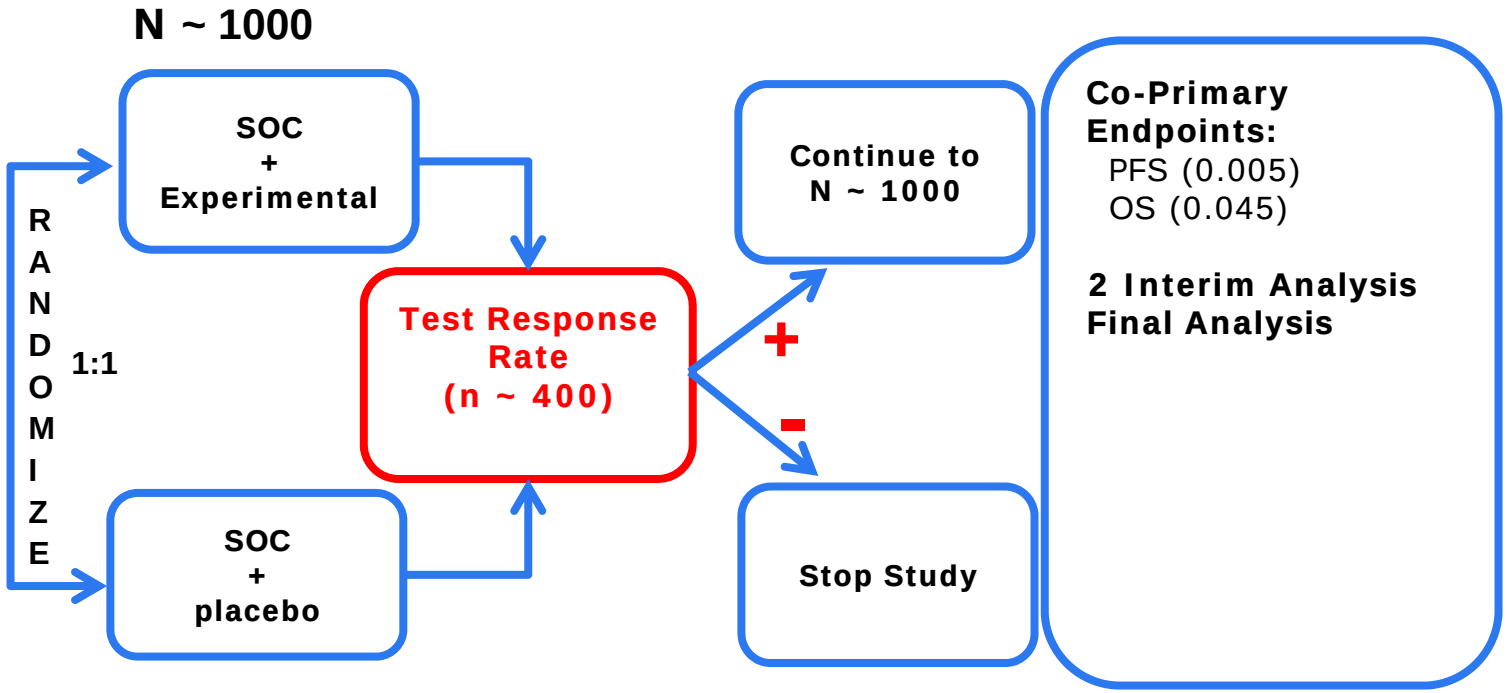
- **OPTION 4:** All comers with goal to test each prognosis group
  - Testing procedure for controlling overall false-positive

Design Options
Concerns and Evaluation of Risk

- What happens when treatment only provides improvement in one prognosis group?
- Things to consider
  - correlation, different OS/PFS in each group, lag of curve separation
  - Potential to use futility rule based on response outcome?
- Impact on likelihood of success, timelines, value
- What is the “cost” to use design 1 vs 2,3 vs 4
- No standard software to account for each of these and understand the risks of each approach

Proposed Design 3- Poor Prognosis Risk

Randomized, DB, Placebo Control



Proposed Design 3

- For PFS
  - Alpha=0.005, power ~85%, HR=0.67
  - Total events 350 (observed at ~30mo)
- For OS
  - Alpha=0.045, power~80%, HR=0.75
  - Total events=400 (n~1000)
  - Enrollment: 25 months
  - Two OS efficacy interim analyses (IA) planned at 60% and 80% of total events
- One futility analysis for % of patients with Response at month 15
- Study analysis timing
  - Response rate futility analysis at 16mo.
  - IA at 60% of OS events, PFS final analysis – month 43
  - IA at 80% of OS events – month 56
  - Final OS Analysis at month 70

Simulations Results
Poor Prognosis (Design 3)

Simulation Results: Design 3, n=1000

Design and Scenario Definitions:

- Design 3.1: **No Response futility** with (Resp, PFS, OS) correlation
  - Scenario 1: Under Null
  - Scenario 2: Under H_1 with no lag in OS curve separation
  - Scenario 3: Under H1 with lag in OS curve separation (1>HR>0.75 for 0-12m, then HR 0.75 thereafter)
  - Scenario 4: Under H_1 with lag in OS curve separation (HR=1 for 0-12m, then HR=0.75 thereafter)
- Design 3.2: **With Response futility** with (Resp, PFS, OS) correlation
  - Same Scenarios as Design 1

Simulation Results:
NON-HIGH VOLUME DESIGN, n=1000

Design / Scenario	Pr (Fut)	Power PFS	Power OS	Ave SS	Mean Time
3.1/1	0.0	0.005	0.045	1000	60
			0.80	1000	55
			0.75	1000	57
			0.49	1000	63

Sample Size Savings of 241 patients and 31

Issues Simulation Uncovered

- Decrease in power due to evaluating design under likely settings (eg delayed treatment effects)
- Are IAs at a fixed time or event driven?
- Average sample size/trial time?
- Evaluating impact of how the trial will actually be run- eg PFS analysis at IA 1
- Using historical data to help design the patient simulation portion can give great insight into the risks
- Increased likelihood of success – delaying IAs

The Simple Question

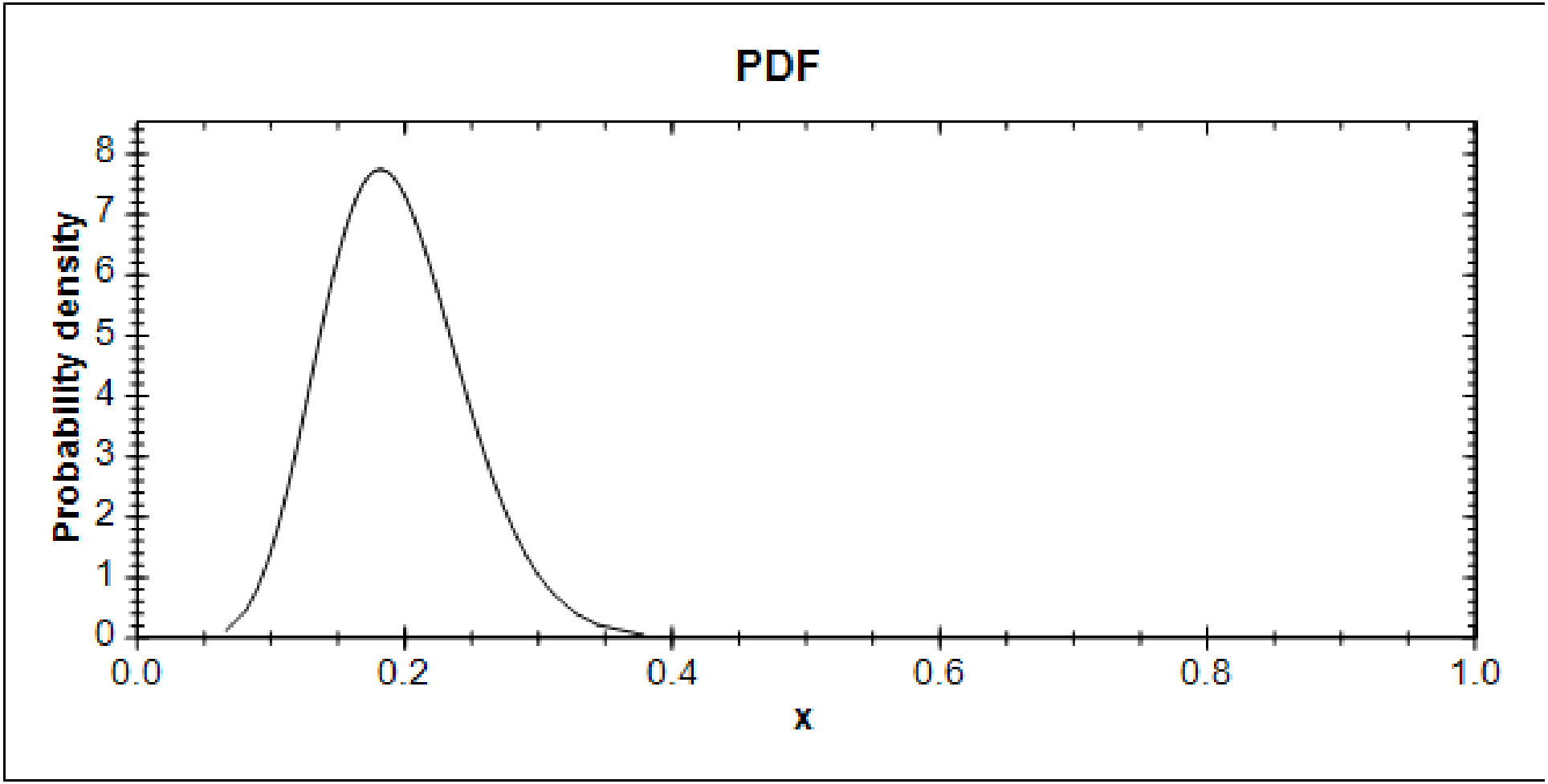
- Two very different examples
  - Different outcome types
  - Different statistical frameworks
  - Different TAs

- **What is the likelihood of success?**

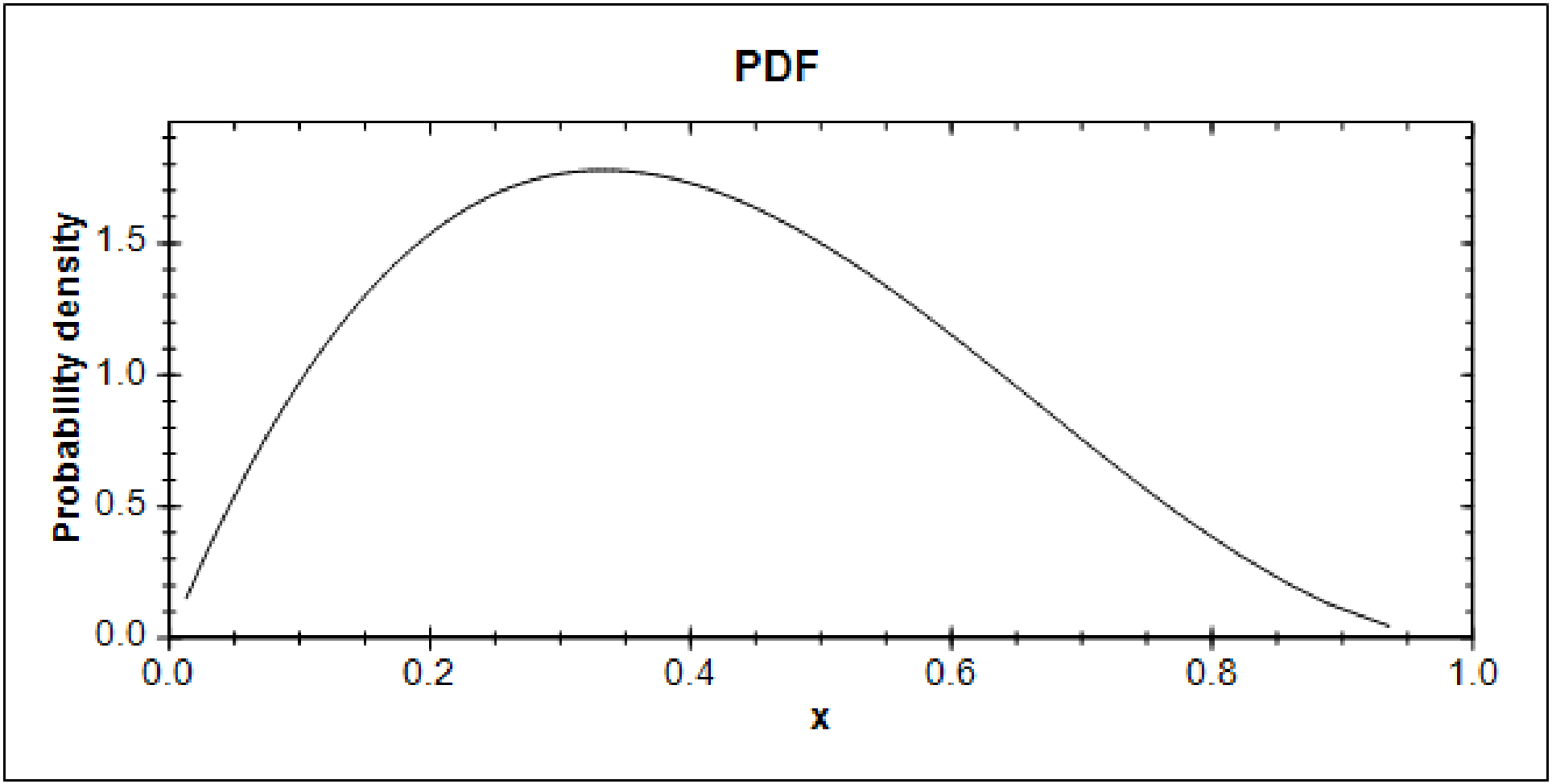
Simple Phase II Example

- Control (C) vs Experimental (E)
- Binary outcome
- Lot of historical data on C
- Historical data suggests the response rate, P_C is 20% and very likely between 10% and 30%
- Very little info based about the response rate of E, P_E, but 2/5 patients responded
- Enroll 100 patients (50 on each arm) and if $P_E - P_C > 0.10$ then success and we run a phase 3
- What is the likelihood of success?

Prior for P_C ~ Beta(11, 46) 95% CI (.1, .3)



Prior for P_E ~ Beta(2, 3) 95% CI (.05,.65)



What is the likelihood of success?

- Use the available info to create the "Prediction Priors"
- P_C ~ Beta(11, 46)
- P_E ~ Beta(2, 3)
- We are NOT doing a Bayesian analysis
  - $P_E - P_C > 0.10$ → Defines success
- Sample P_C, P_E from the prediction priors.
- For each sampled value simulate the trial
- Evaluate how often $P_E - P_C > 0.10$
- In this case Probability of success is 64%

Scenario	Dose	True TFR
1 – Null 1	P	0.4
	D1, D2, D3	0.4
2 – Alt 1	P	0.4
	D1,2	0.4
	D3	**0.3**
3 – Null 2	P	0.3
	D1, D2, D3	0.3
4 – Alt 2	P	0.3
	D1, D2	0.3
	D3	**0.225**
5 – Null 3	P	0.2
	D1, D2, D3	0.2
6 – Alt 3	P	0.2
	D1, D2	0.2
	D3	**0.15**

Example 2
Design 3, n=1000

Design and Scenario Definitions:

- Design 3.1: **No Response futility** with (Resp, PFS, OS) correlation
  - Scenario 1: Under Null
  - Scenario 2: Under H_1 with no lag in OS curve separation
  - Scenario 3: Under H1 with lag in OS curve separation (1>HR>0.75 for 0-12m, then HR 0.75 thereafter)
  - Scenario 4: Under H_1 with lag in OS curve separation (HR=1 for 0-12m, then HR=0.75 thereafter)

Conclusion

- Simulation guided design improved the likelihood of success in both of the examples.
- Evaluating the options under realistic conditions can identify potential issues
- Using a Bayesian framework to obtain the likelihood of success – useful for decision making

Thank you!

J. Kyle Wathen, PhD
kwathen@its.jnj.com