Adaptive Clinical Trials: Application to Pain Studies

IMMPACT-VIII. Arlington, VA
June 13-14, 2008

Cyrus R. Mehta, Ph.D

Cytel Inc., Cambridge, MA

email: mehta@cytel.com – web: www.cytel.com – tel: 617-661-2011

Statistical Challenges of Adaptive Design

- The interest in adaptive trial design has rapidly increased, much of it driven by senior management at pharmaceutical and biotech companies
- This has opened up opportunities for statisticians to work on new types of statistical, logistical, operational problems
- Some of the statistical problems include:
  - Dose response modeling using Bayesian methods
  - Unblinded sample size re-estimation
  - Dropping arms in multi-arm trials
  - Population enrichment

Non-Statistical Challenges

- Logistical and operational challenges include:
  - Managing drug supply
  - Dynamic randomization
  - Preventing bias from premature disclosure of interim results
  - Making complex decisions affecting a company's business interests without company participation
- See PhRMA Adaptive Working Group White Paper, Drug Information Journal 40, 421-423.

Two Case Studies of Adaptive Design Involving Pain Trials

Case 1: Acute Setting. Combined POC and dose-response modeling of a medical compound for relief from acute dental pain in an exploratory phase 2 setting (8 hour endpoint)

Case 2: Chronic Setting. Sample size re-estimation in a confirmatory phase 3 trial of a medical device for relief from chronic pain (12 week endpoint)

Acknowledgements

- Case 1 is a joint project with Nitin Patel at Cytel Inc., and Jim Bologenese at Merck Pharmaceuticals
- Case 2 is joint project with Stuart Pocock at London School of Hygiene and Tropical Medicine and a small medical device company
- I thank Drs. Patel, Bologenese and Pocock for their permission to use these cases

Case 1: Bayesian Dose-Response Modeling

- Model dose-response for an analgesic for dental pain after surgical removal of molars
- Pain relief measured serially over 8 hours after single dose

  0 = none, 1 = some, 2 = a little, 3 = a lot, 4 = complete
- Primary endpoint AUC, assumed to be normal with $\sigma = 5$
- Randomize patients to 7 doses, labelled $1, 2, 3, 4, 5, 6, 7$, or placebo
- Goal: Demonstrate POC and estimate the dose-response curve in a single phase 2 trial

Options for Patient Allocation

Suppose we intend to test the drug on 120 patients; 40 on placebo, 80 on drug. How should we allocate the 80 patients?

1. **Conventional Method: Randomize equally to the seven doses**

2 **Response Adaptive. Randomize in cohorts. Use the results of previous cohort to change the randomization fraction for the next cohort by Bayesian methods**

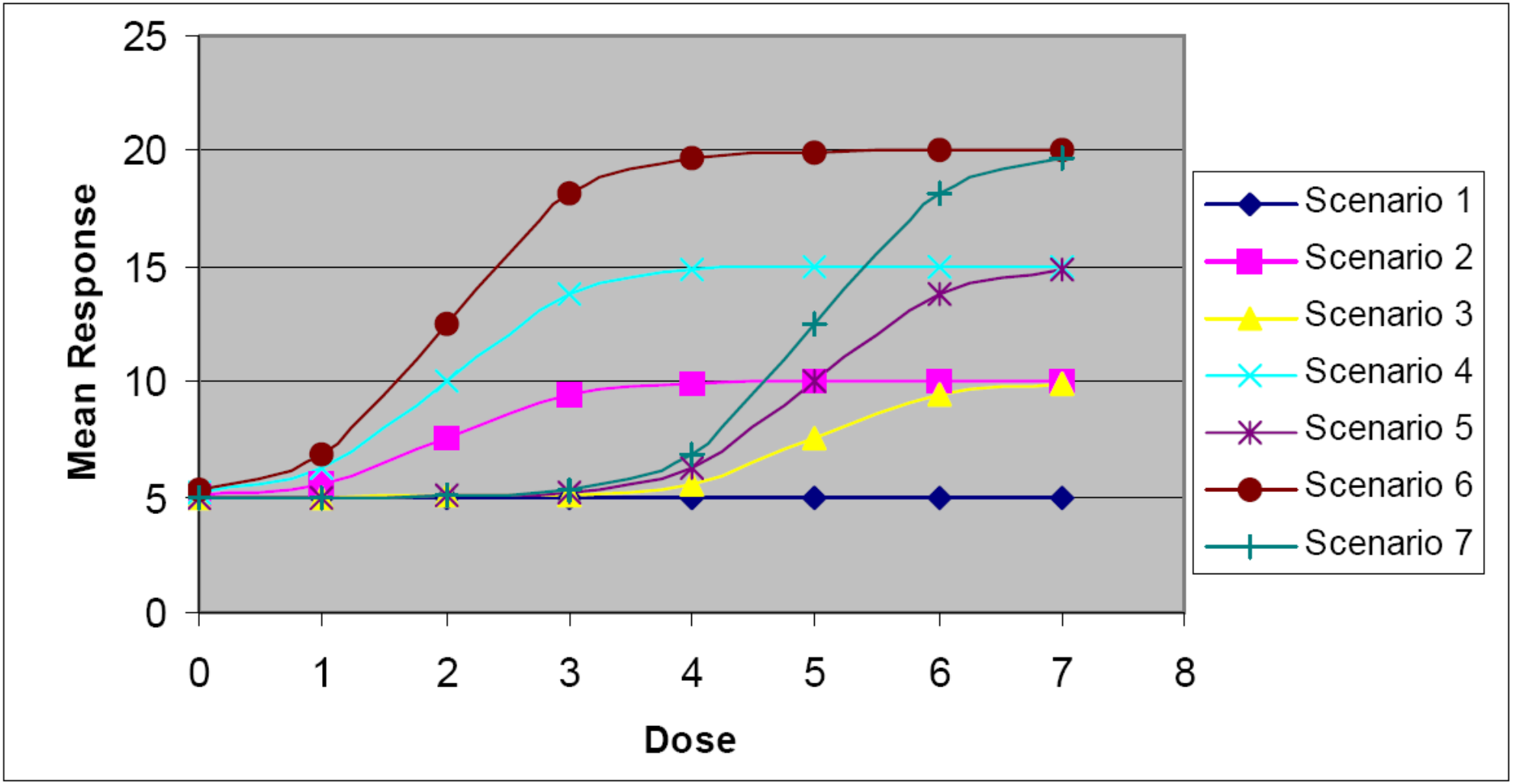
Postulate a Dose-Response Model

- 4-parameter logistic regression model for subject i, dose j

$$Y_{ij} = \beta + \frac{\delta}{(1 + e^{(\theta - d_i)/\tau})} + \epsilon_{ij}$$

- Different choices of $\beta, \delta, \theta, \tau$ yield different shapes
- We don't know the correct shape
- Simulate data from many different shapes and see how well the Bayesian allocation scheme assigns patients, regardless of shape

Likely Dose-Response Scenarios

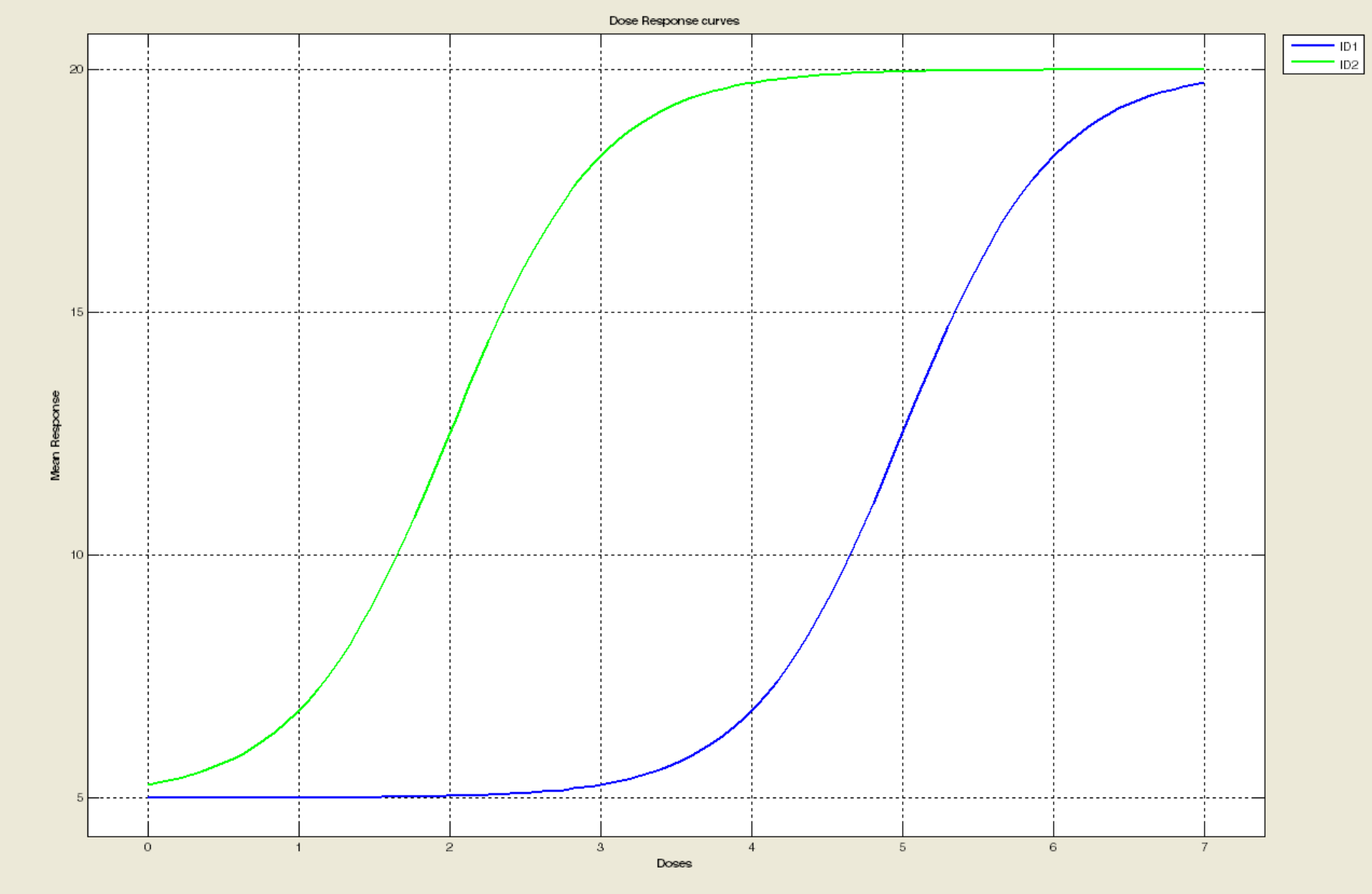


Regardless of which dose-response curve is the true one, the Bayesian algorithm should find it and allocate most patients to its sloping part

Bayesian Adaptive Randomization

1. Start with a "flat" (non-informative) prior for model parameters
2. Create 10 cohorts of size 12 (4 to placebo, 8 to doses)
3. Assign subjects to doses with equal probability for first cohort
4. Use responses of subjects to compute postrior distribution using MCMC simulation
5. Compute an "effectiveness criterion" for each possible dose, if assigned to the next subject
   - Since we are targeting sloping part of curve, effectiveness criterion is sum of expected response variances at 25th and 75th quantiles
6. Assign assign the next cohort to doses with probabilities proportional to their effectiveness criteria
7. Repeat steps 3, 4, 5 until trial sample size is reached

Simulation Demo: Scenarios 6 & 7



Criteria for Comparison

- PoC:

  Power of trend test

  (slope in linear regression).

- Dose-finding:

  Statistical efficiency at each dose

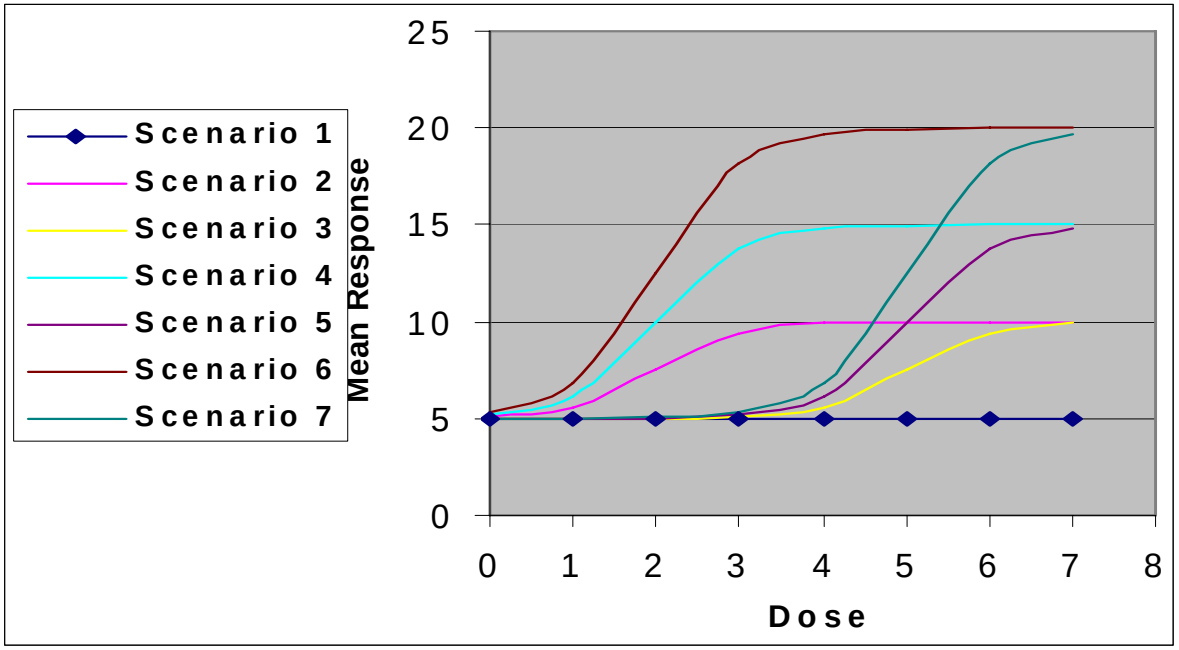
  (Mean Squared Error of Standard Design)/(MSE of Adaptive Design).

(1000 simulations used for calculations)

Cytel
STATISTICAL SOFTWARE & SERVICES

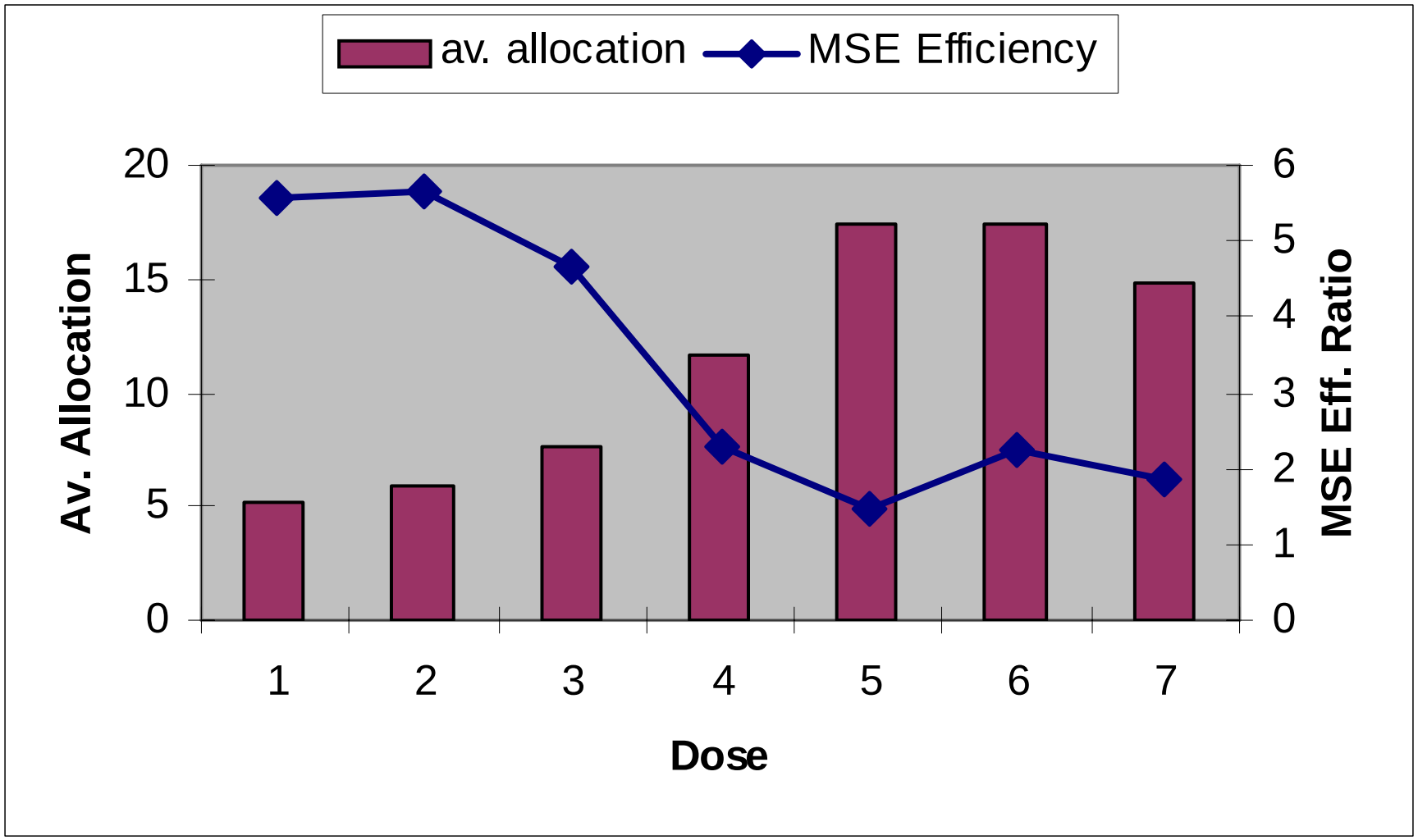
Proof-of-Concept

Scenario	Power of Standard Design	Power of Adaptive Design
1	5.3	5.4
2	49.4	84.0
3	49.4	67.2
4	95.8	100
5	95.8	99.2
6	99.9	100
7	99.9	100



Cytel
STATISTICAL SOFTWARE & SERVICES

Efficiency of Adaptive



Cytel
STATISTICAL
SOFTWARE & SERVICES

Comparative dose-finding efficiency of adaptive design

Scenarios	Doses 1	2	3	4	5	6	7
1	5.8	6.4	6.5	5.8	5.1	3.6	2.0
2	2.7	2.0	2.2	3.4	3.6	3.6	2.6
3	5.7	4.9	3.6	2.5	2.1	2.0	1.7
4	2.7	1.2	2.0	3.6	4.0	3.5	2.8
5	5.6	5.6	4.2	2.2	1.4	2.0	1.8
6	2.4	1.6	2.5	3.3	3.1	2.6	2.1
7	5.6	5.7	4.6	2.3	1.5	2.3	1.8

Interesting part of the Dose Response curve

The adaptive design is more efficient (>1) for all scenarios and all doses.

The average efficiency is 3.3 over all doses and scenarios.

For the interesting part of the dose response curve it is 2.1.

Cytel STATISTICAL SOFTWARE & SERVICES

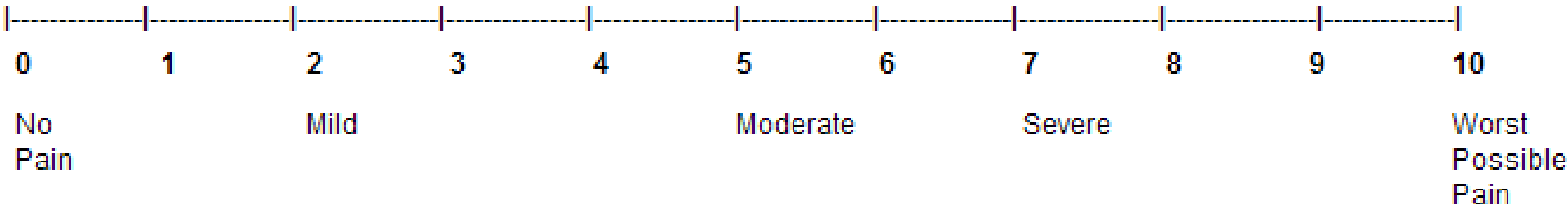
Conclusions

- Bayesian algorithm assigns more patients to the interesting part of the dose-response curve
- Hence can estimate responses more accurately with fewer patients
- Can demonstrate proof of concept with greater power
- Dominates over fixed allocation with respect to efficiency of estimating response

Case 2: Sample Size Re-Estimation

- **Two-arm clinical trial of relief for fibromyalgia**
- **Primary endpoint: decrease in Numeric Rating Scale (NRS) pain score at week-12**

Figure 1. Numeric Rating Scale (NRS)



- **Sponsor has limited experience with the device and endpoint. Unable to determine sample size**

Sample Size Formula

- What sample size is needed to detect an average NRS improvement of δ in this patient population with $(1-\beta)$ power, using a one-sided level-α test?
- Sample size depends on both δ and σ^2 (the variance in NRS between subjects)

$$N = \sigma^2 \left[\frac{z_\alpha + z_\beta}{\delta} \right]^2$$

- Sponsor needs to specify δ and σ but can only guess at their values

Option 1: Two Separate Trials

- Run a small exploratory trial to get an idea of δ and σ
- Run a second confirmatory trial, adequately powered to detect the estimated δ
- Practical Problems:
  - White space between the two trials (due to closing and opening sites, IRB approval, management buy-in, etc.)
  - If the first trial does not show statistical significance, often difficult to get funding for the second trial
  - Small companies often wish to show significance on a single trial and then license the technology to large pharma

Option 2: A Single Two-Stage Trial

- Plan initially for a sample size of 200 patients
- After 100 patients have completed, calculate conditional power (CP) using unblinded estimates of δ and σ. Then:
  - complete the trial with no change if CP is high
  - increase the sample size if CP is medium
  - terminate for futility if CP is low
- Statistical and Practical Problems
  - Choosing the initial sample size
  - Criteria for sample size increase and futility termination
  - Preventing inflation of type-1 error
  - Avoiding possibility of bias due to premature unblinding

Initial Sample Size Calculation

- Assume average pain reduction of 1 point for placebo arm and 2 points for treatment arm
- Estimate from other NRS data that $\sigma \approx 2.15$
- Powered trial to detect $\delta = 1$ (with $\sigma = 2.15$) on a one-sided level-0.025 test
- Enroll 200 subjects (100/arm) to achieve 90.5% power
- Great uncertainty about δ and σ

Futility Stopping and Sample Size Increase at Interim

- Interim analysis, after 100 subjects have completed 12 weeks of follow-up
- Compute conditional power at interim analysis:

  CP_{200} = conditional power if final sample size is 200

  CP_{400} = conditional power if final sample size is 400
- If $CP_{200} > 90\%$ continue with no sample size change
- If $CP_{400} < 50\%$ terminate for futility
- Else continue with sample size increase that achieves 90% conditional power, up to maximum 400 subjects

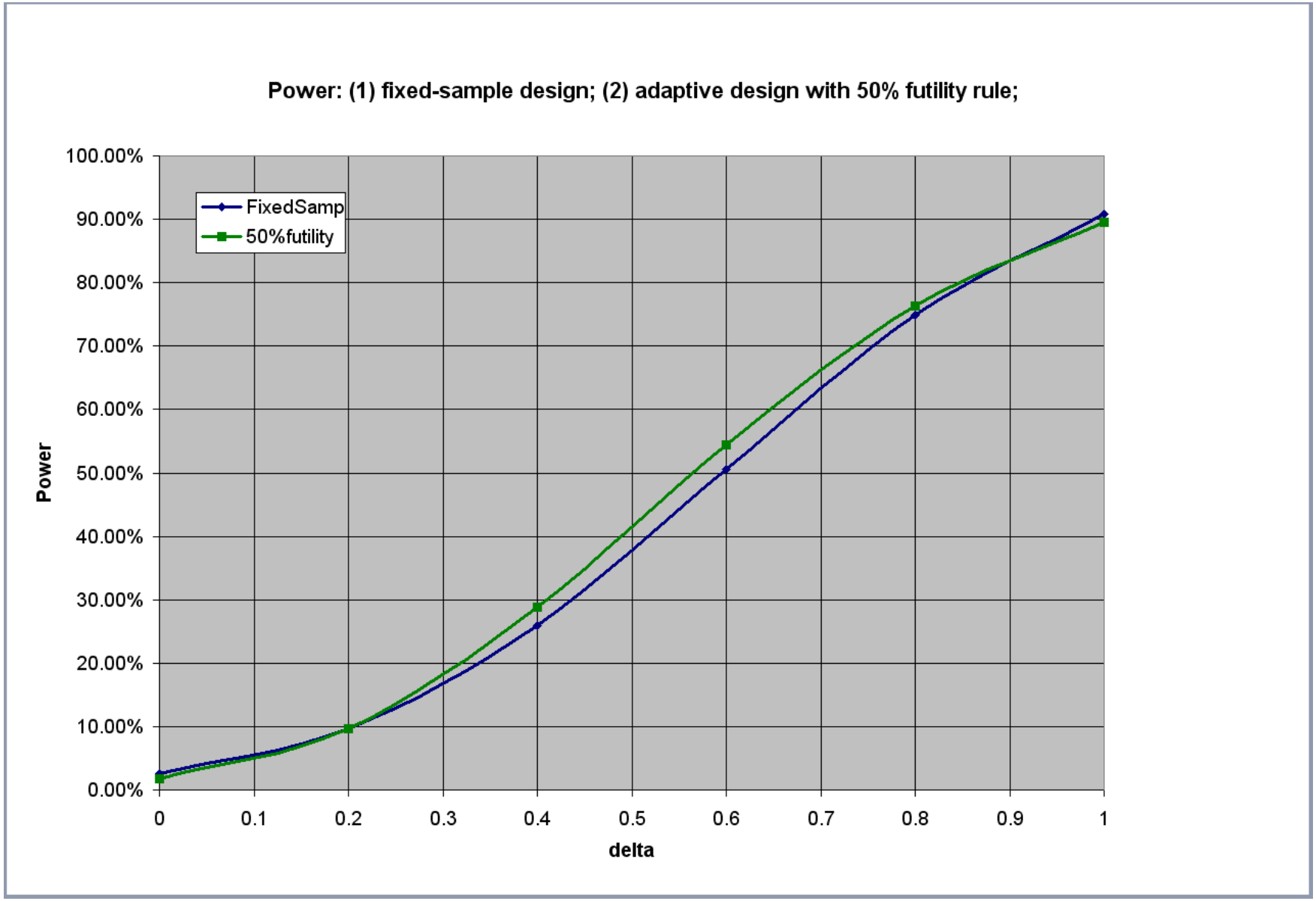
The Conditional Power Formula

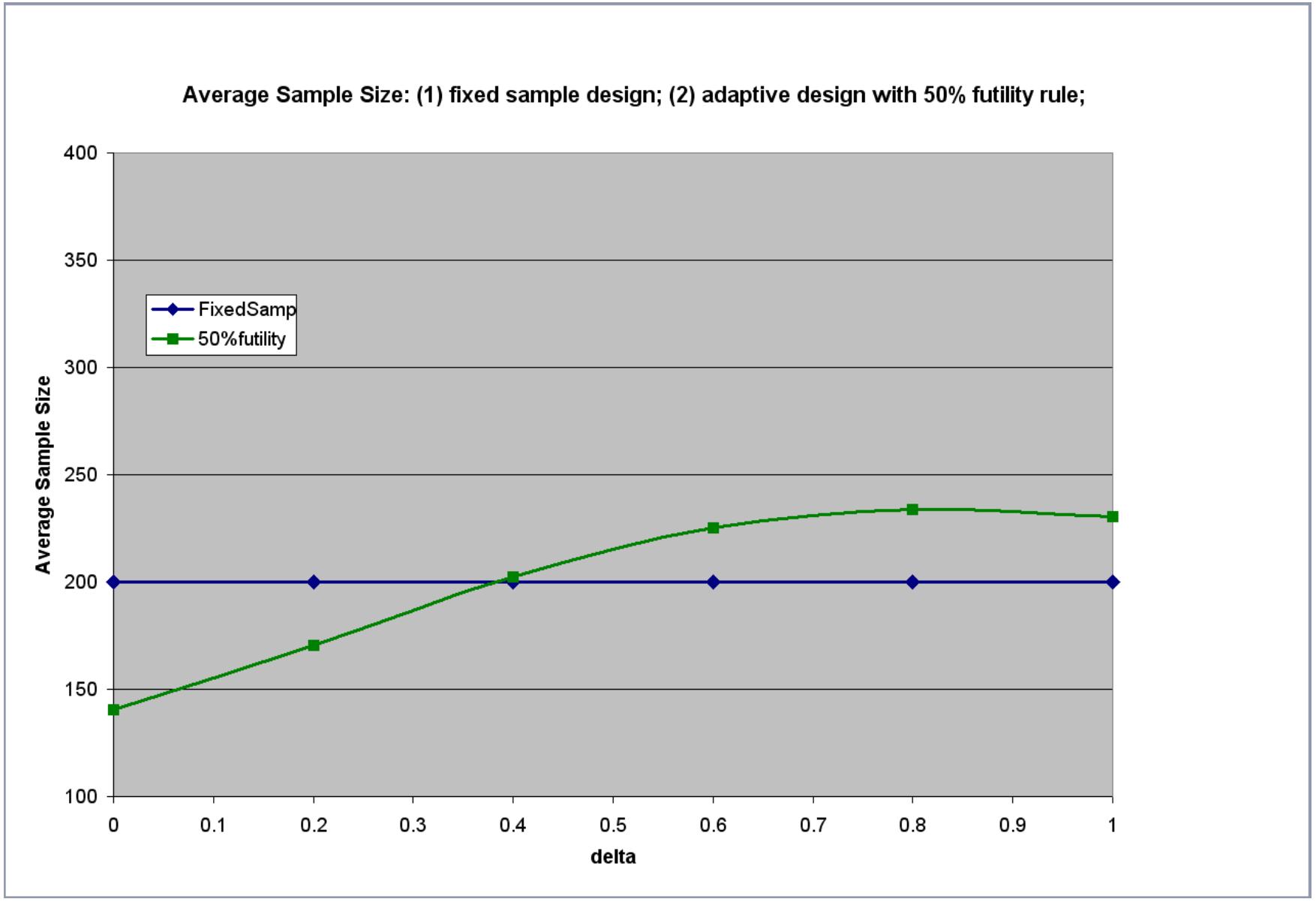
- Let $N =$ total sample size; $n_1 =$ sample size at interim
- Define the information fraction $t_1 = (n_1/N)$
- Suppose we obtain estimates $\hat{\delta}_1$ and $\hat{\sigma}_1$ at the interim look
- Then the conditional power is calculated as

$$\text{CP}(\hat{\delta}_1) = 1 - \Phi\left\{\frac{c_\alpha}{\sqrt{1-t_1}} - \frac{z_1\sqrt{t_1}}{\sqrt{1-t_1}} - 2\left[\frac{\hat{\delta}_1}{\hat{\sigma}_1}\right]\sqrt{(N-n_1)}\right\}$$

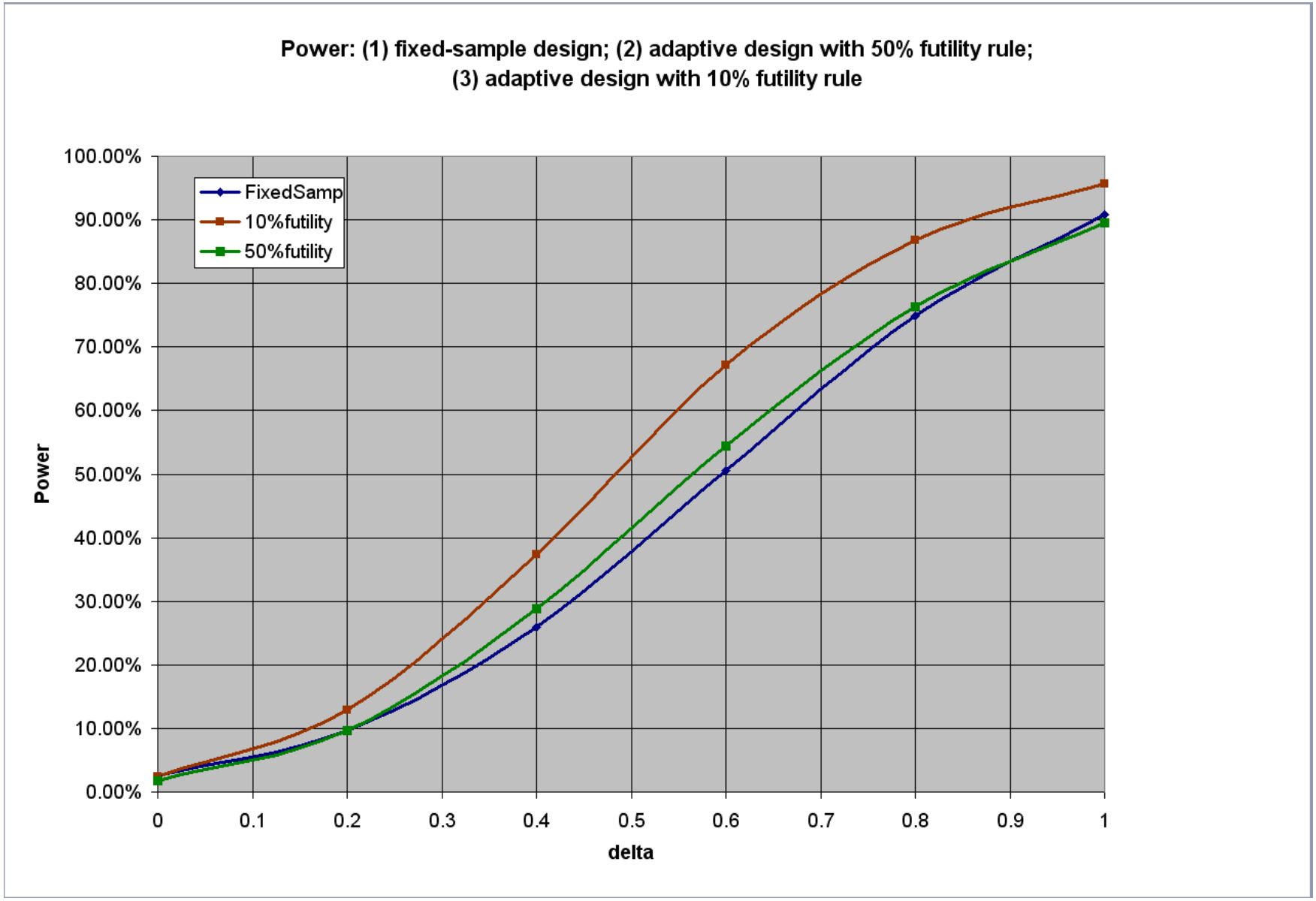
where c_α is the critical cut-off and $z_1 = \hat{\delta}_1/\text{se}(\hat{\delta}_1)$

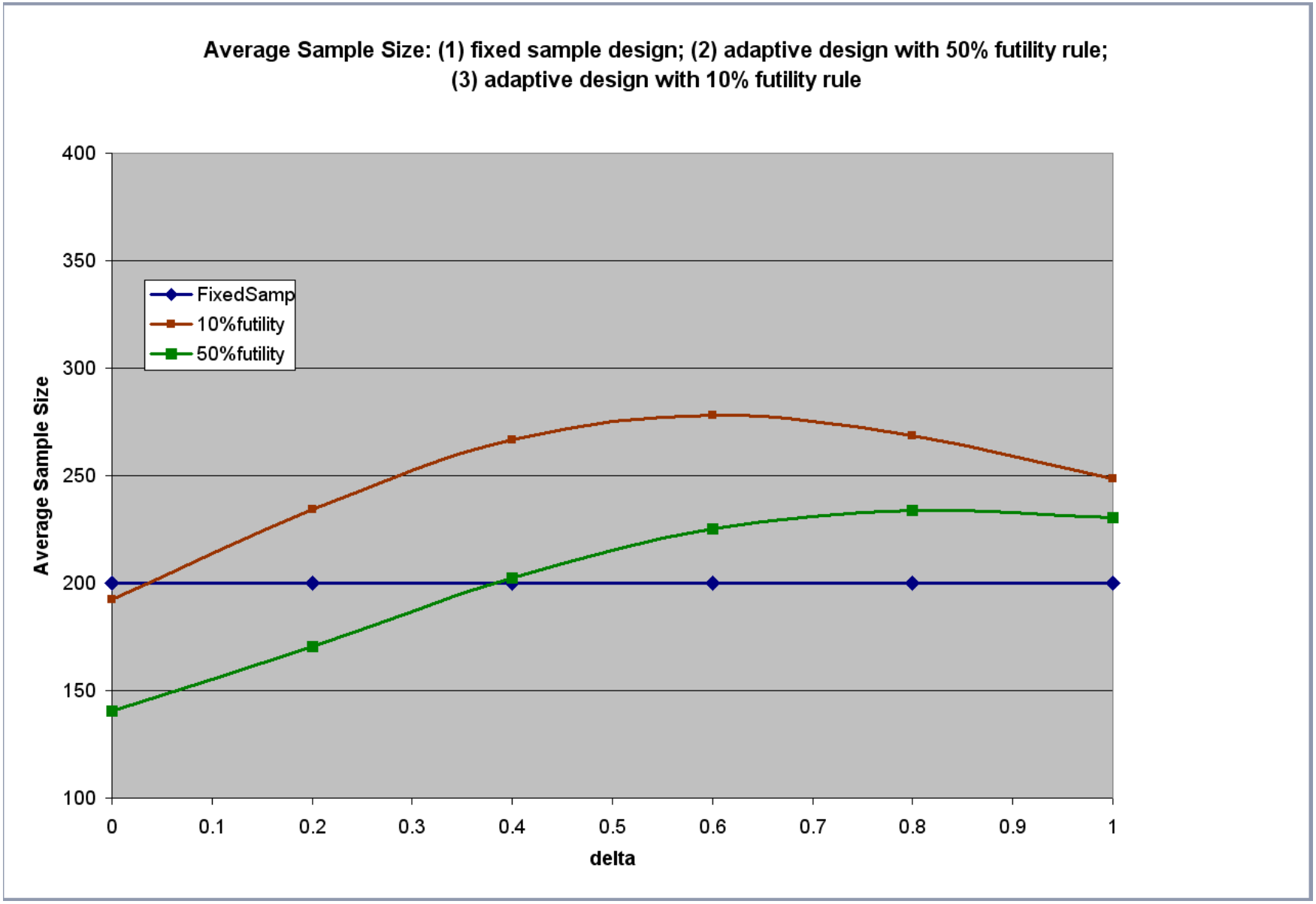
How Good is this Design?



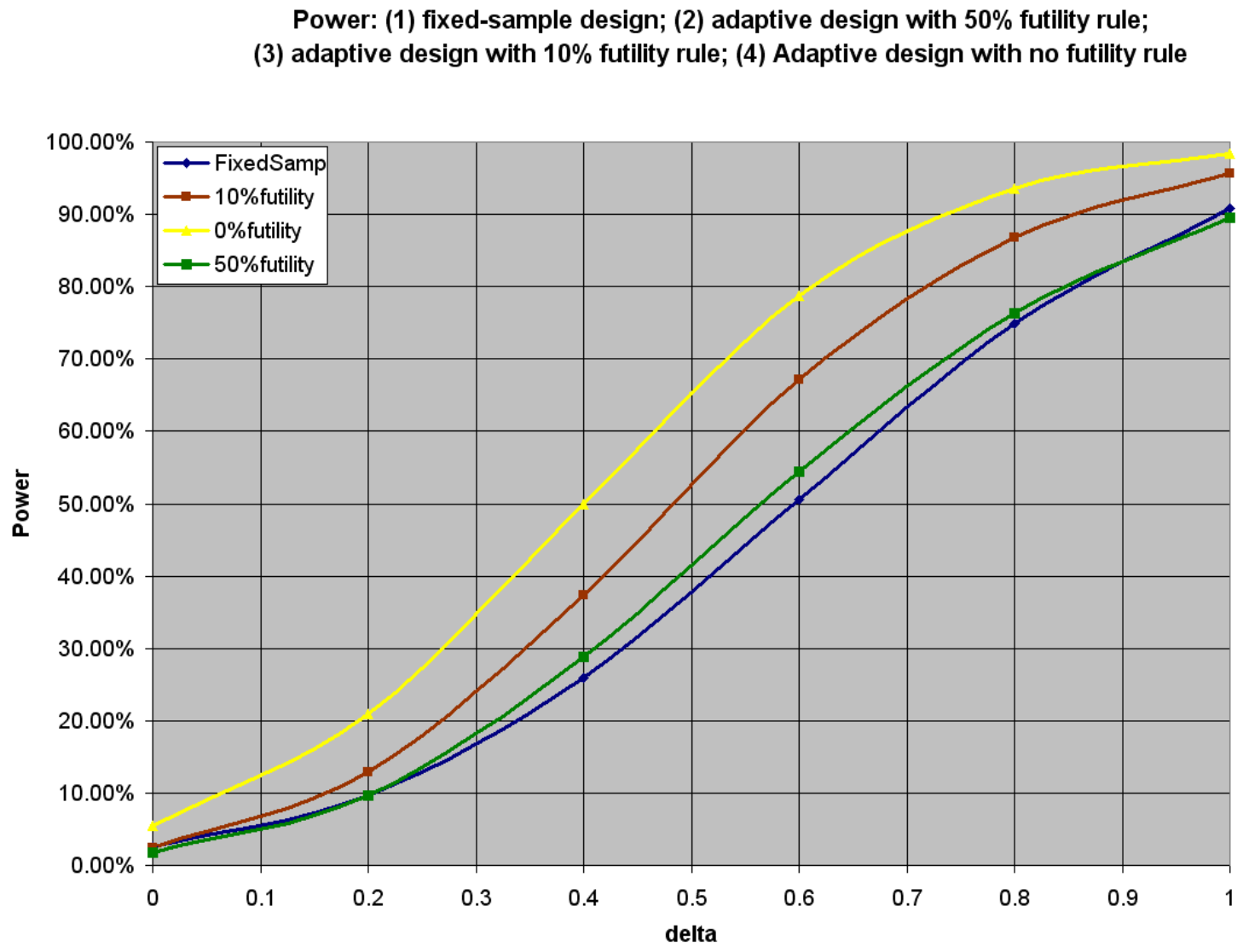

Average Sample Size: (1) fixed sample design; (2) adaptive design with 50% futility rule;
Average Sample Size
400
350
300
250
200
150
100
FixedSamp
50%futility
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
delta

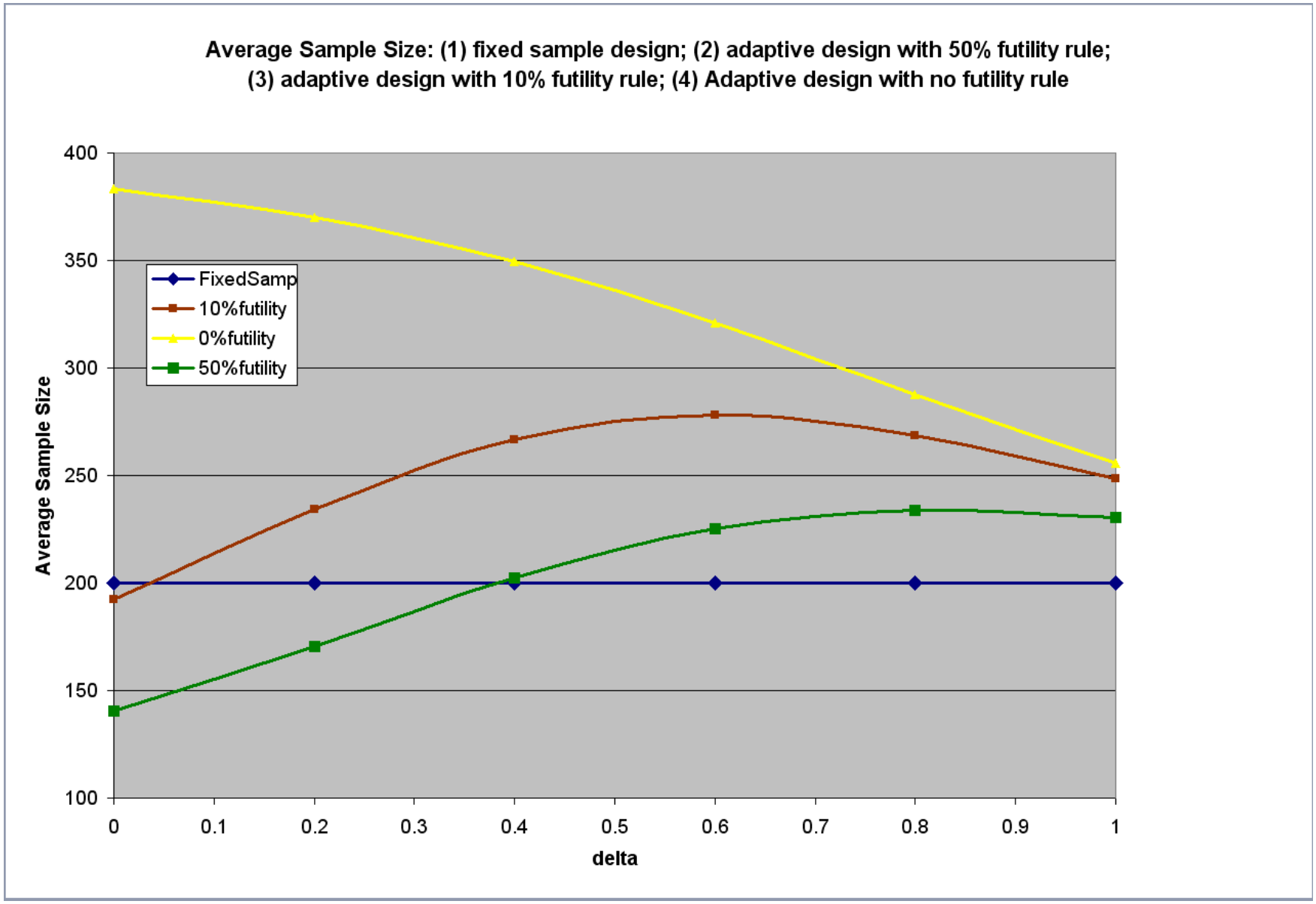
Relax the Futility Criterion



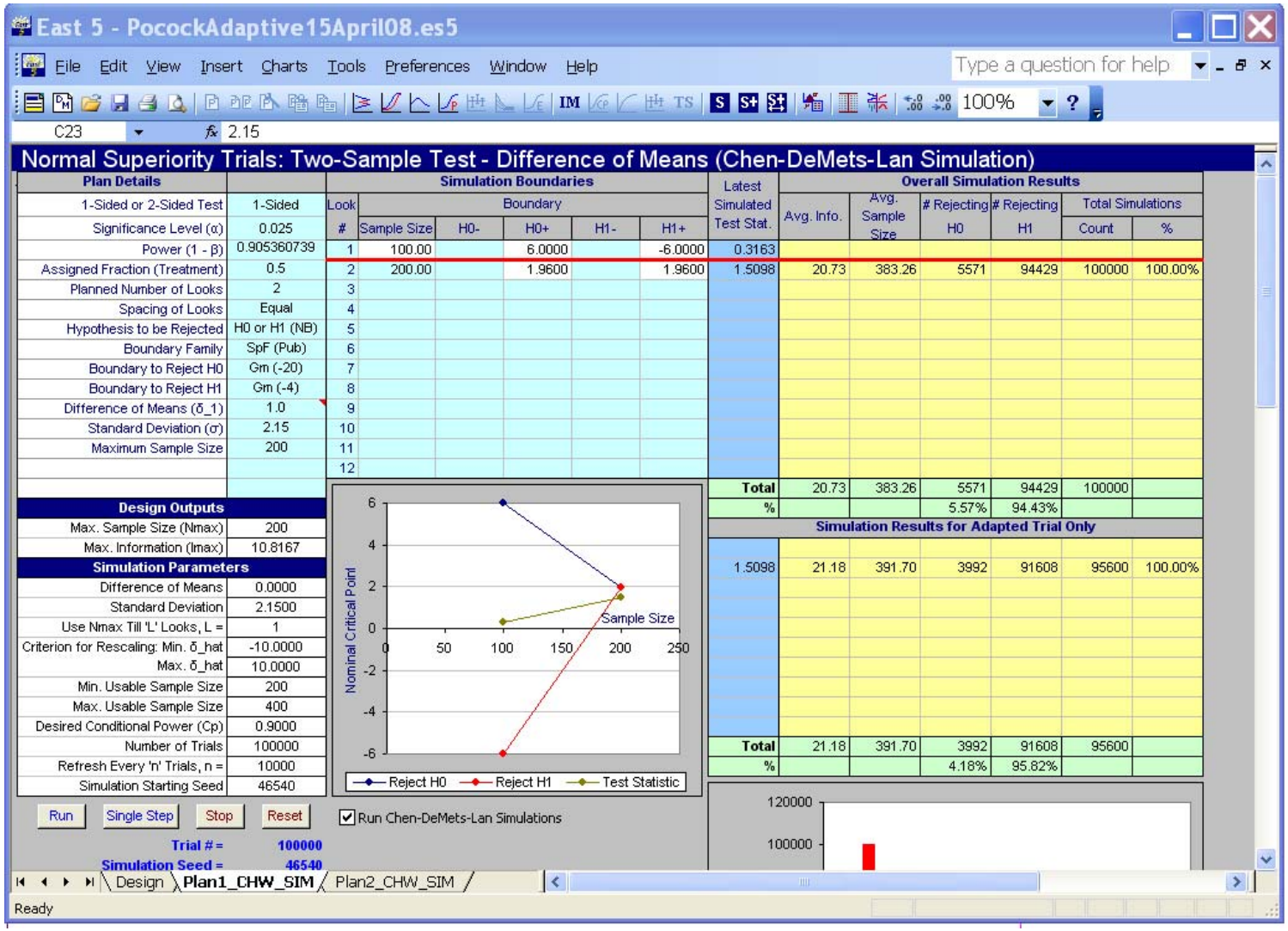

Average Sample Size: (1) fixed sample design; (2) adaptive design with 50% futility rule;
(3) adaptive design with 10% futility rule
Average Sample Size
400
350
300
250
200
150
100
FixedSamp
10%futility
50%futility
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
delta

Eliminate the Futility Criterion




Average Sample Size: (1) fixed sample design; (2) adaptive design with 50% futility rule;
(3) adaptive design with 10% futility rule; (4) Adaptive design with no futility rule
Average Sample Size
400
350
300
250
200
150
100
FixedSamp
10%futility
0%futility
50%futility
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
delta

Without the Futility Boundary, Type-1 Error is Inflated



Preserving the Type-1 Error

The following are the stage-1 and stage-2 data:

- (n_1, n_2) are initially planned sample sizes for the two stages
- $(\hat{\delta}_1$ is the stage-1 estimate of δ from stage-1 data
- $n_2^* \geq n_2$ is the adaptively increased stage-2 sample size
- $\hat{\delta}_2^*$ is the estimate of δ from independent stage-2 data
- How can we construct a level-α test of H_0: $\delta = 0$ after a data dependent sample size increase?

Constructing a Level-α Test

1. **CHW Test:** Cui, Hung and Wang (1999)

$$\sqrt{\frac{n_1}{n_1+n_2}} \times \frac{\hat{\delta}_1}{\text{se}(\hat{\delta}_1)} + \sqrt{\frac{n_2}{n_1+n_2}} \times \frac{\hat{\delta}_2^*}{\text{se}(\hat{\delta}_1)} \geq C_\alpha$$

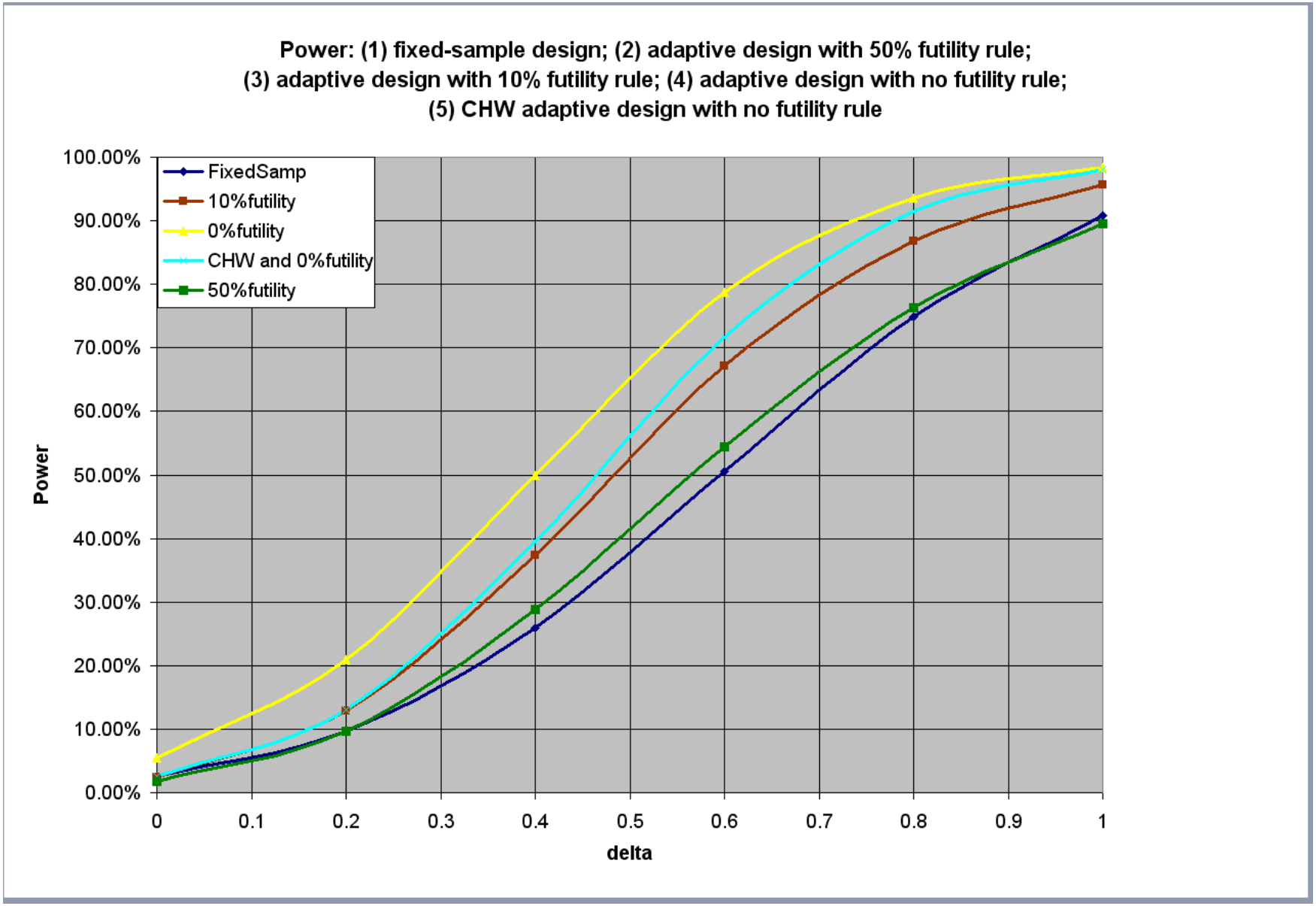
   This test is guaranteed to preserve the type-1 error

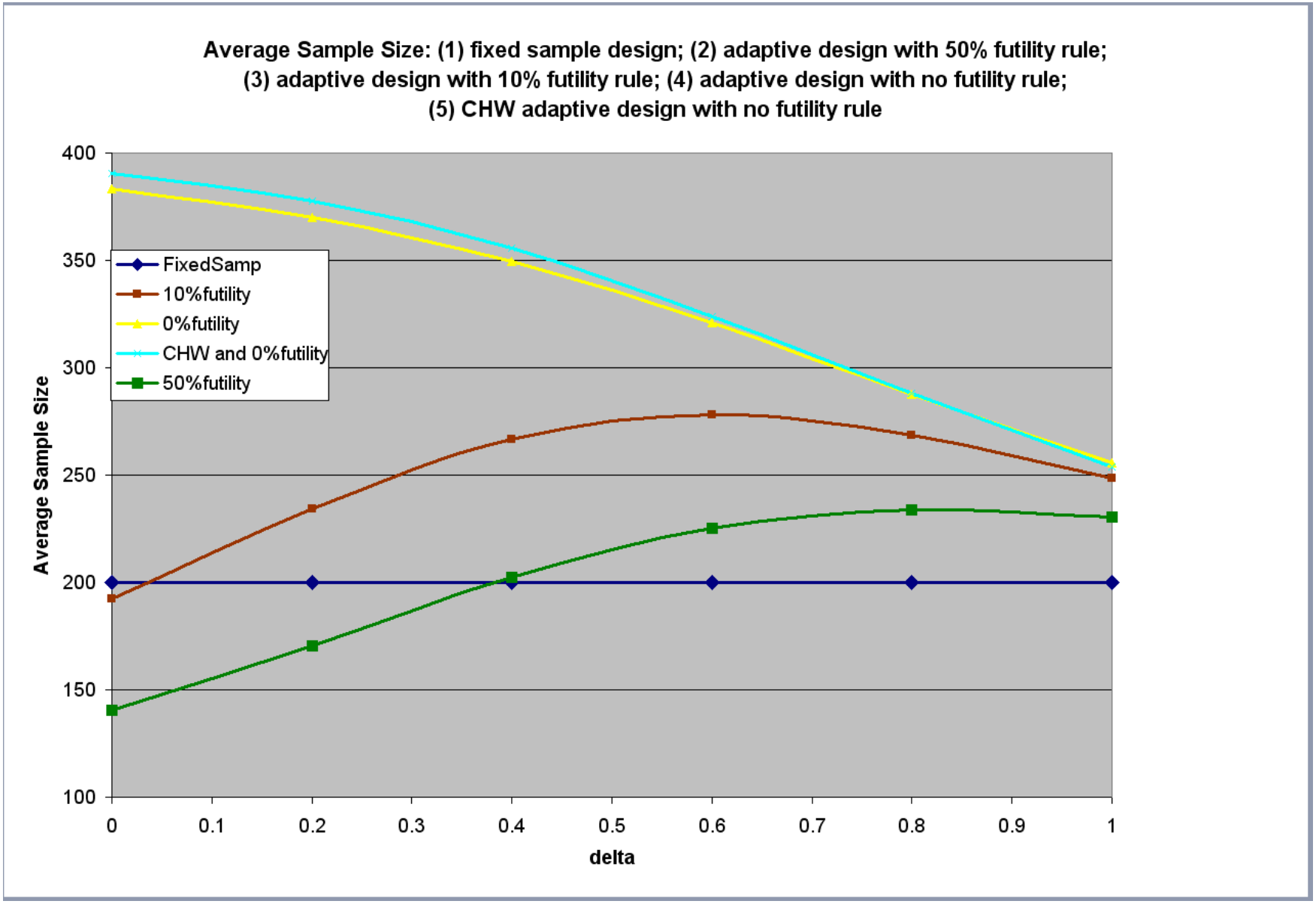
2. **Conventional Test:** Wald (195?)

$$\sqrt{\frac{n_1}{n_1+n_2^*}} \times \frac{\hat{\delta}_1}{\text{se}(\hat{\delta}_1)} + \sqrt{\frac{n_2^*}{n_1+n_2^*}} \times \frac{\hat{\delta}_2^*}{\text{se}(\hat{\delta}_1)} \geq C_\alpha$$

   This test may or may not preserve the type-1 error if there is a sample size change

CHW Test Preserves the α




Average Sample Size: (1) fixed sample design; (2) adaptive design with 50% futility rule;
(3) adaptive design with 10% futility rule; (4) adaptive design with no futility rule;
(5) CHW adaptive design with no futility rule
Average Sample Size
400
350
300
250
200
150
100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
delta
FixedSamp
10%futility
0%futility
CHW and 0%futility
50%futility

When Does Conventional Test Preserve the α?

The type-1 error of the conventional test is preserved despite a sample size change, in the following situations:

1. **If sample size increase is unrelated to the observed $\hat{\delta}$**
2. **If there is a suitable futility boundary to act as a brake on the posible error inflation**
3. **If the conditional power at the interim analysis exceeds 50% (Chen, DeMets and Lan, *Statistics in Medicine*, 2004)**

Attractiveness of Conventional Method

- Utilizes classical Wald statistic; no down-weighting of second cohort!
- standard methods for hypothesis testing
- standard methods for parameter estimation
- nevertheless permits data dependent sample size increase in certain circumstances

Recommendation to Sponsor

- Build in the option for a sample size increase at the interim
- Second chance to get get the right sample size
- Could boost conditional power to 90%
- Dominates over fixed sample design in terms of unconditional power
- Use the 10% futility boundary and conventional test

A Final Take-Away Message

Adaptive trials require a considerable amount of planning up-front. One of the most versatile tools for the planning phase is simulation

- **The simulations clarify the risks and benefits of the proposed approach by putting probabilities on each possible outcome**
- **The simulations facilitate better communication with the FDA**
- **The simulations facilitate greater communication between the various members of the study team by showing how different design options and trial outcomes will have different implications for:**
  - **patient recruitment**
  - **drug supply**
  - **economic analyses**
  - **clinical outcomes**
  - **statistical power**
  - **regulatory concerns**