Adaptive Group Sequential Trials with Population Enrichment: Application to Cardiology

PhRMA Adaptive Working Group
KOL Series
July 11, 2008

Cyrus R. Mehta, Ph.D
Cytel Inc., Cambridge, MA

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

Cytel STATISTICAL SOFTWARE & SERVICES

Acknowledgements

- Joint research with Ping Gao (The Medicines Company) and Jim Ware (Harvard University)
- We thank Aniruddha Deshmukh (Cytel Inc) for outstanding programming support
- We thank Stuart Pocock and Misha Salganik for helpful discussions

Population Enrichment Design for an Acute Coronory Syndrome Trial

Major challenges to trial design for treatment of acute coronary symptoms include:

- low event rates
- tiny effect sizes
- diverse patient population

Such trials require enormous sample size commitments. Therefore it is advisable to build in the possiblity for adaptive changes to the design based on interim looks

Case Study: Platelet Inhibition during Percutaneous Coronary Intervention

- Composite primary endpoint – death, MI or ischemia driven revascularization within 48 hours
- Placebo event rate between 8% and 10%
- New drug expected to reduce placebo event rate by 20%
- But actual risk reduction could be as low as 15%

Design Options Evaluated

1. **Fixed Sample Design**
2. **Group Sequential Design**
3. **Group Sequential Design with Sample Size Increase**
4. **Group Sequential Design with Sample Size Increase and Population Enrichment**

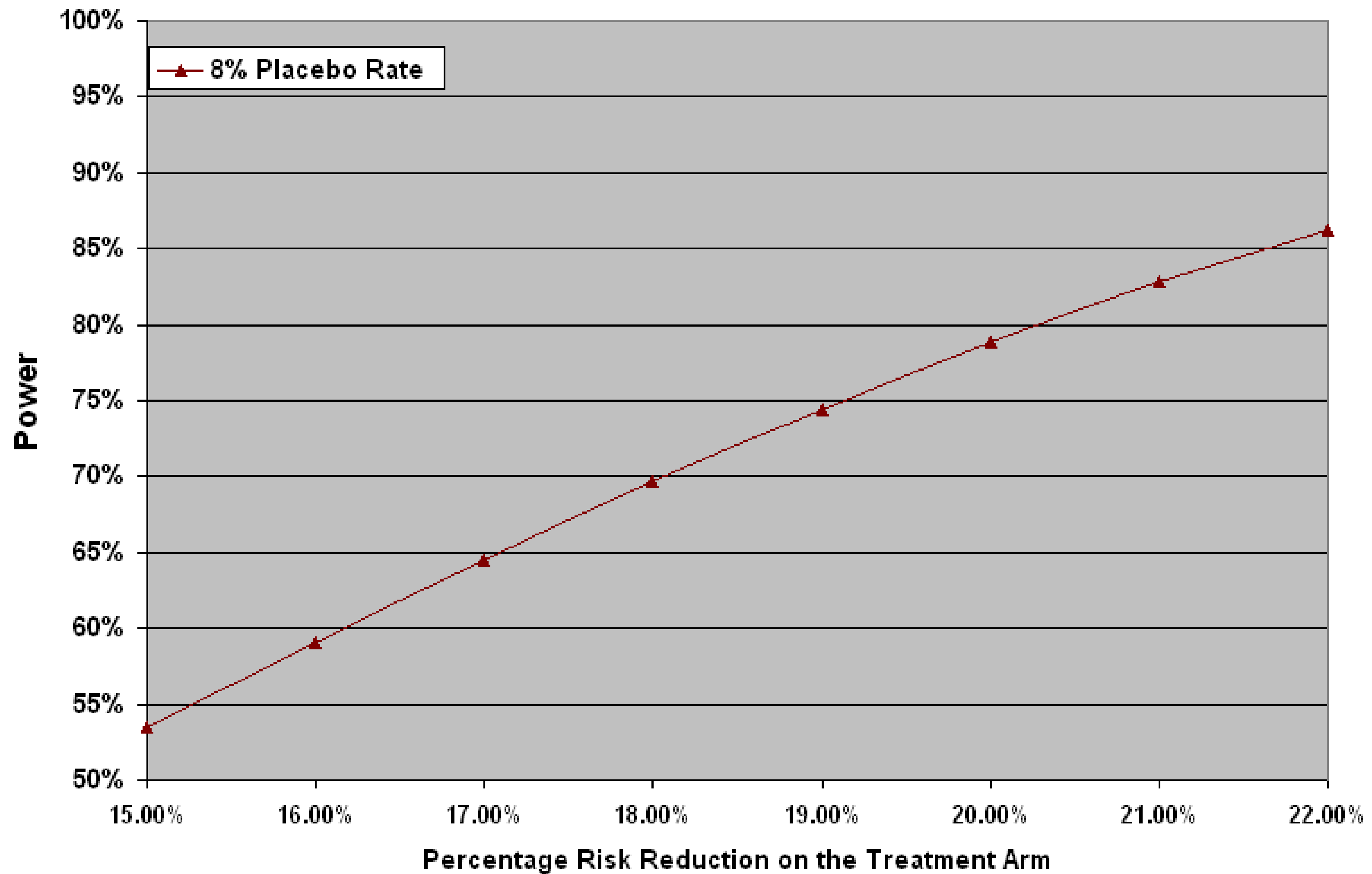
1. Fixed Sample Design

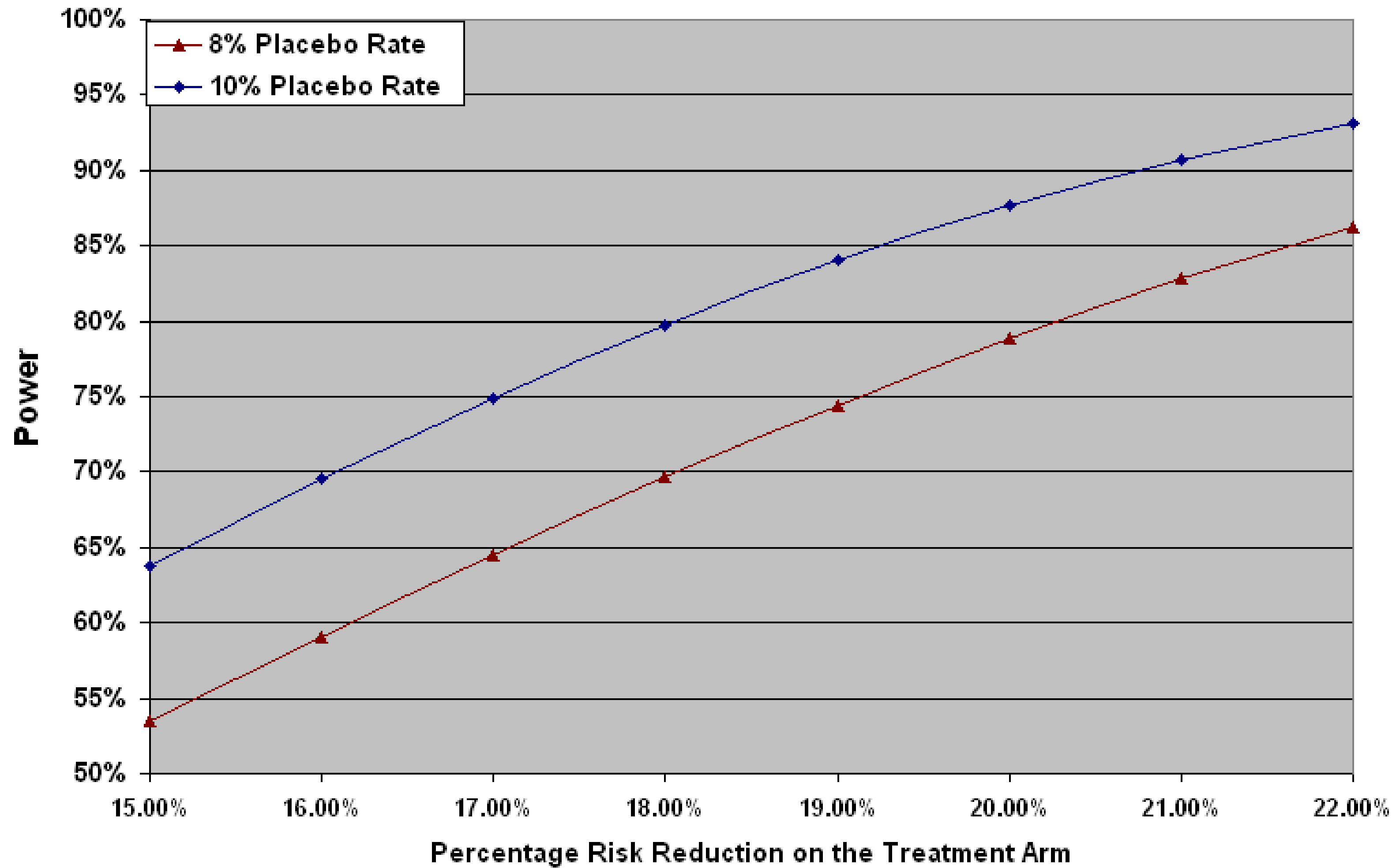
- Sponsor decides on an initial commitment of $N = 8000$
- This sample size provides more that 80% power if:

  risk reduction $\geq 18\%$ and placebo event rate $\geq 10\%$

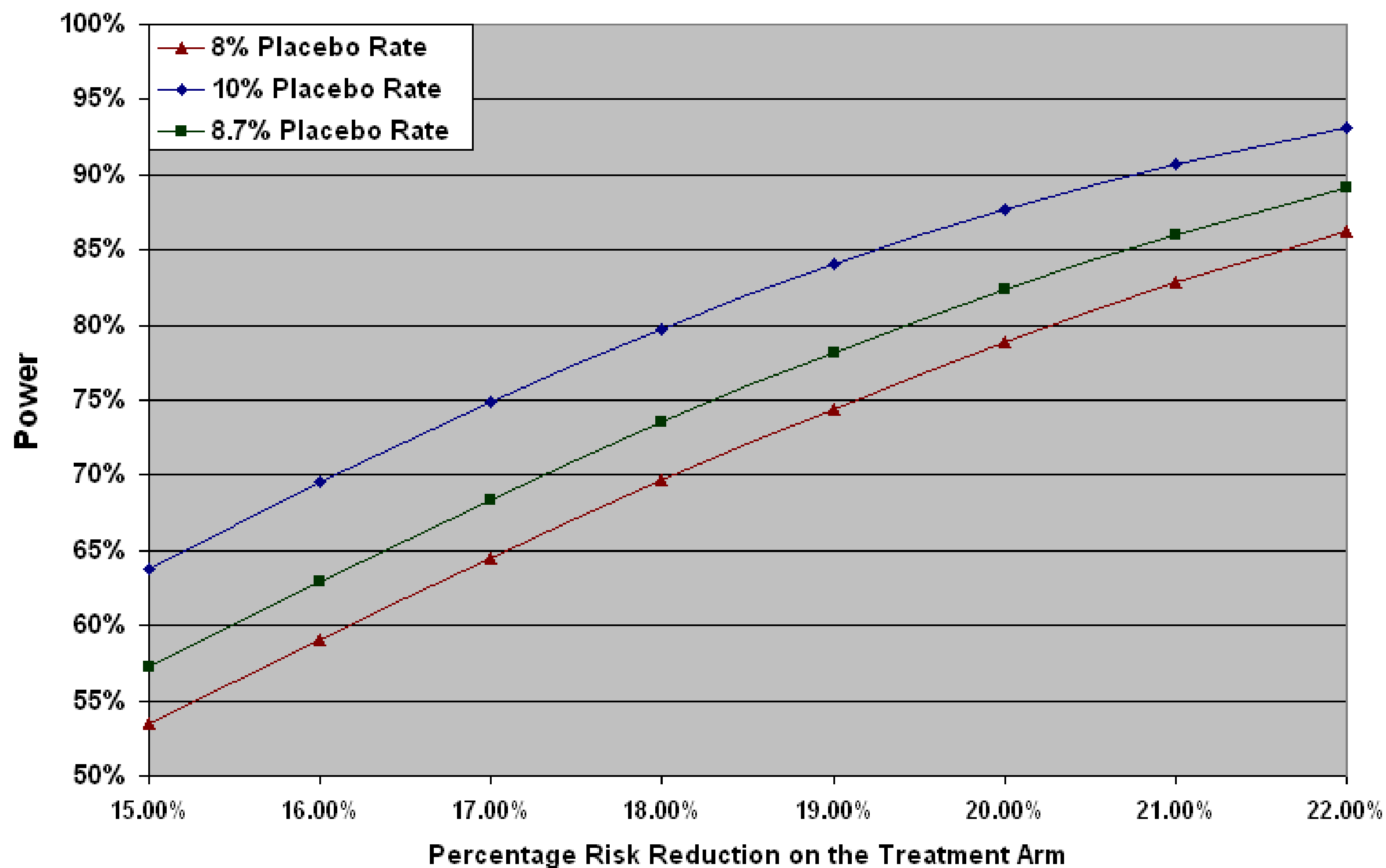
- If these parameter estimates were be off by a few percentage points, study might be underpowered

Cytel STATISTICAL SOFTWARE & SERVICES


Power Curves with Three Different Placebo Event Rates
Fixed Sample Designs (N = 8000)
8% Placebo Rate
Power
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
15.00%
16.00%
17.00%
18.00%
19.00%
20.00%
21.00%
22.00%
Percentage Risk Reduction on the Treatment Arm



Cytel STATISTICAL SOFTWARE & SERVICES



Cytel STATISTICAL SOFTWARE & SERVICES

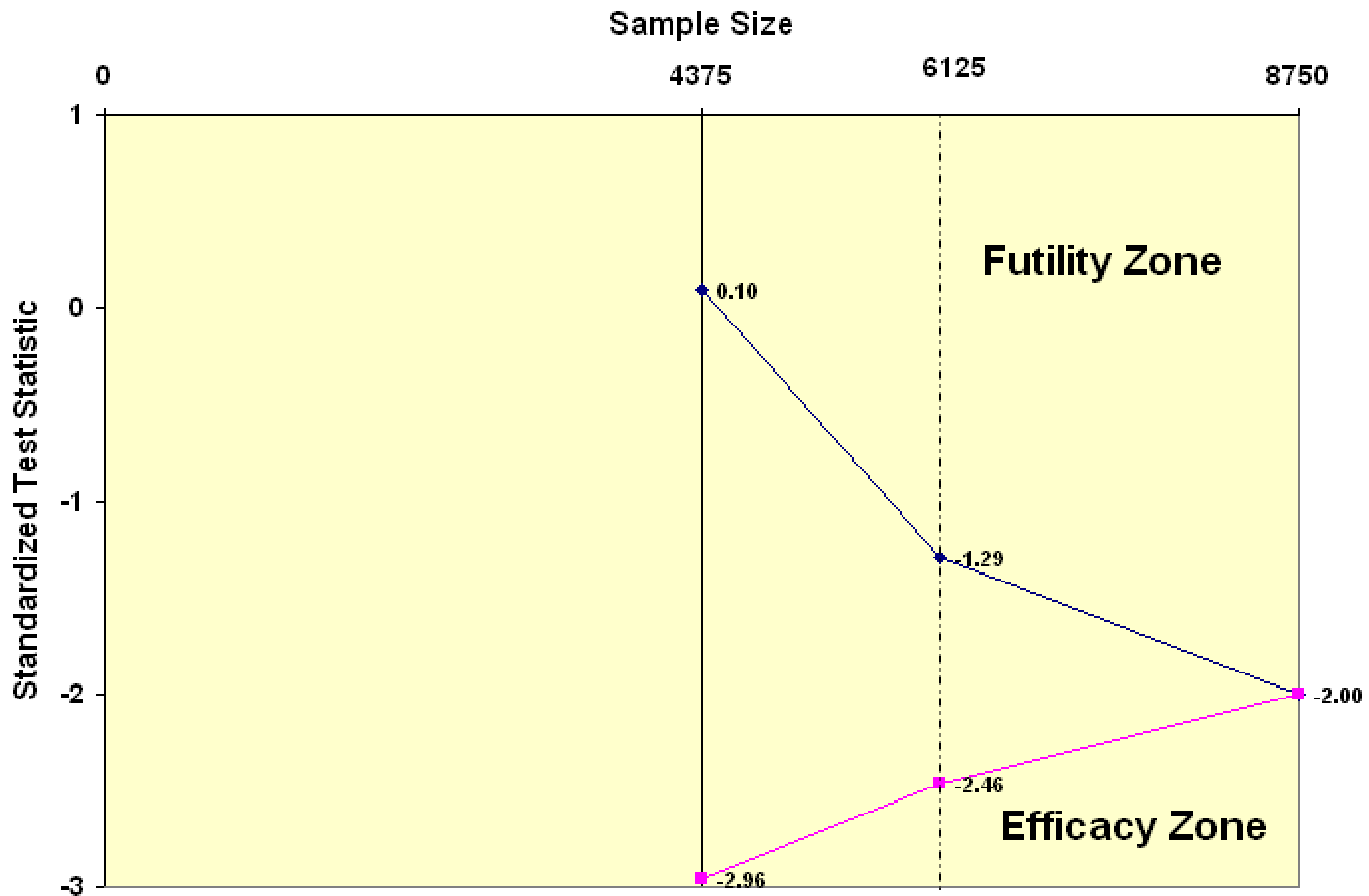
Can we do better?

- 8000 patients is a very large fixed commitment
- We can do better with a group sequential design having early stopping boundaries for efficacy or futility
- Such a design would have a larger up-front commitment but lower expected sample size

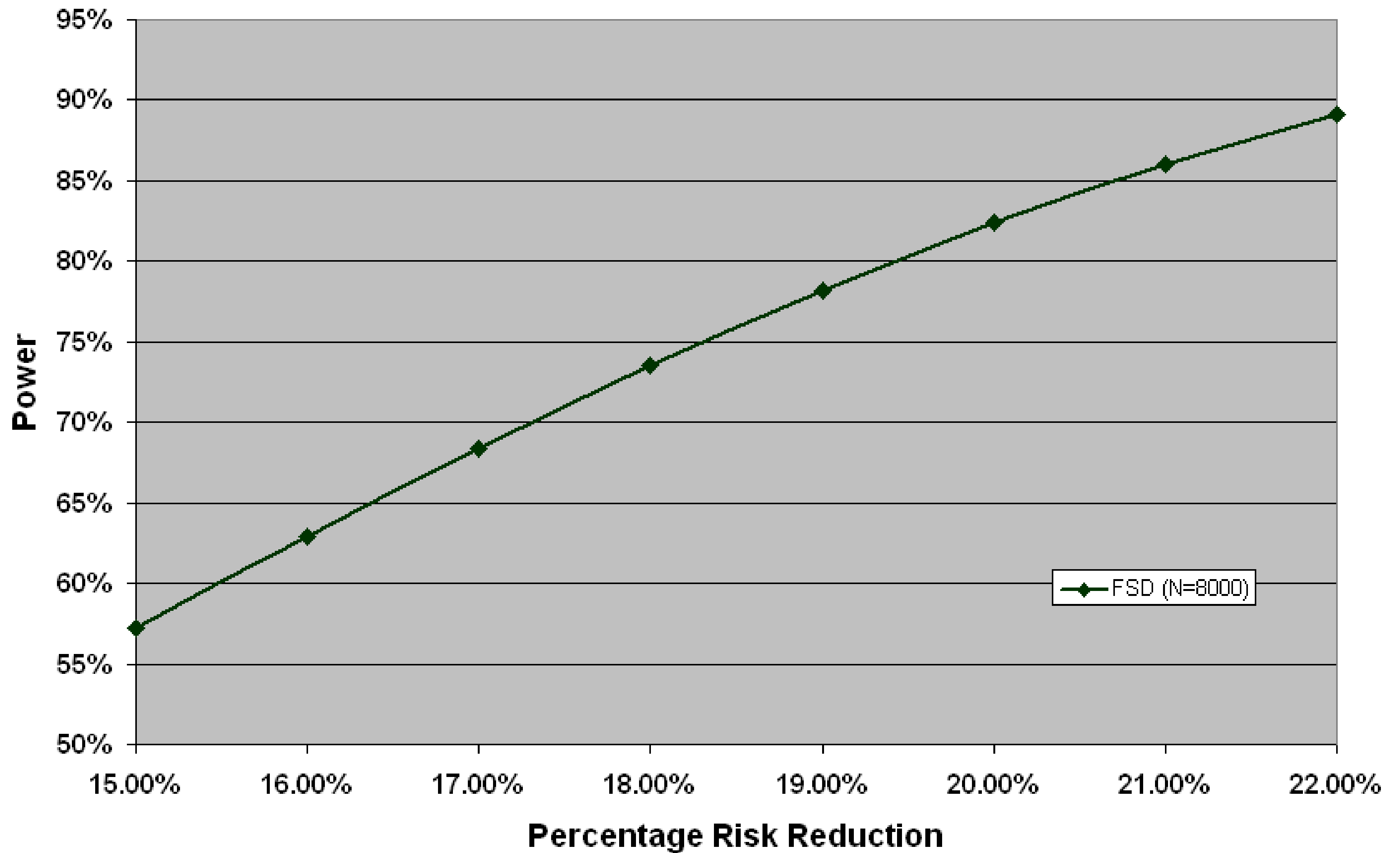
2. Group Sequential Design (GSD)

- Enroll up to $N_{\max} = 8750$ subjects
- First Interim Analysis (IA1): at 50% information
  - Stop for efficacy if $Z \leq -2.9626$ (OBF bdry)
  - Stop for futility if 1% change in event rate in the wrong direction
- Second Interim Analysis (IA2): at 70% information
  - Stop for efficacy if $Z \leq -2.4623$ (OBF bdry)
  - Stop for futility if conditional power $\leq 20\%$

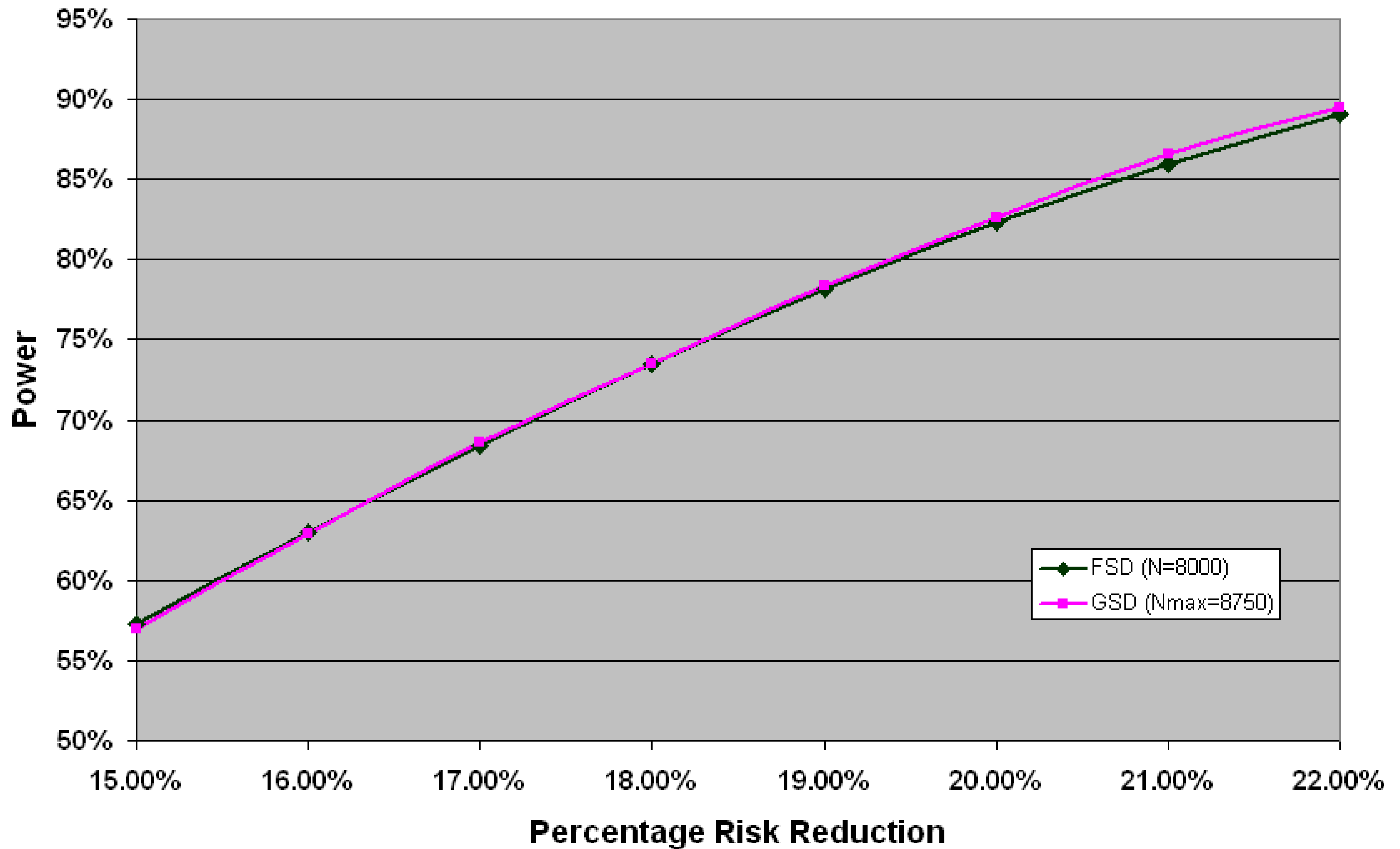
Cytel STATISTICAL SOFTWARE & SERVICES



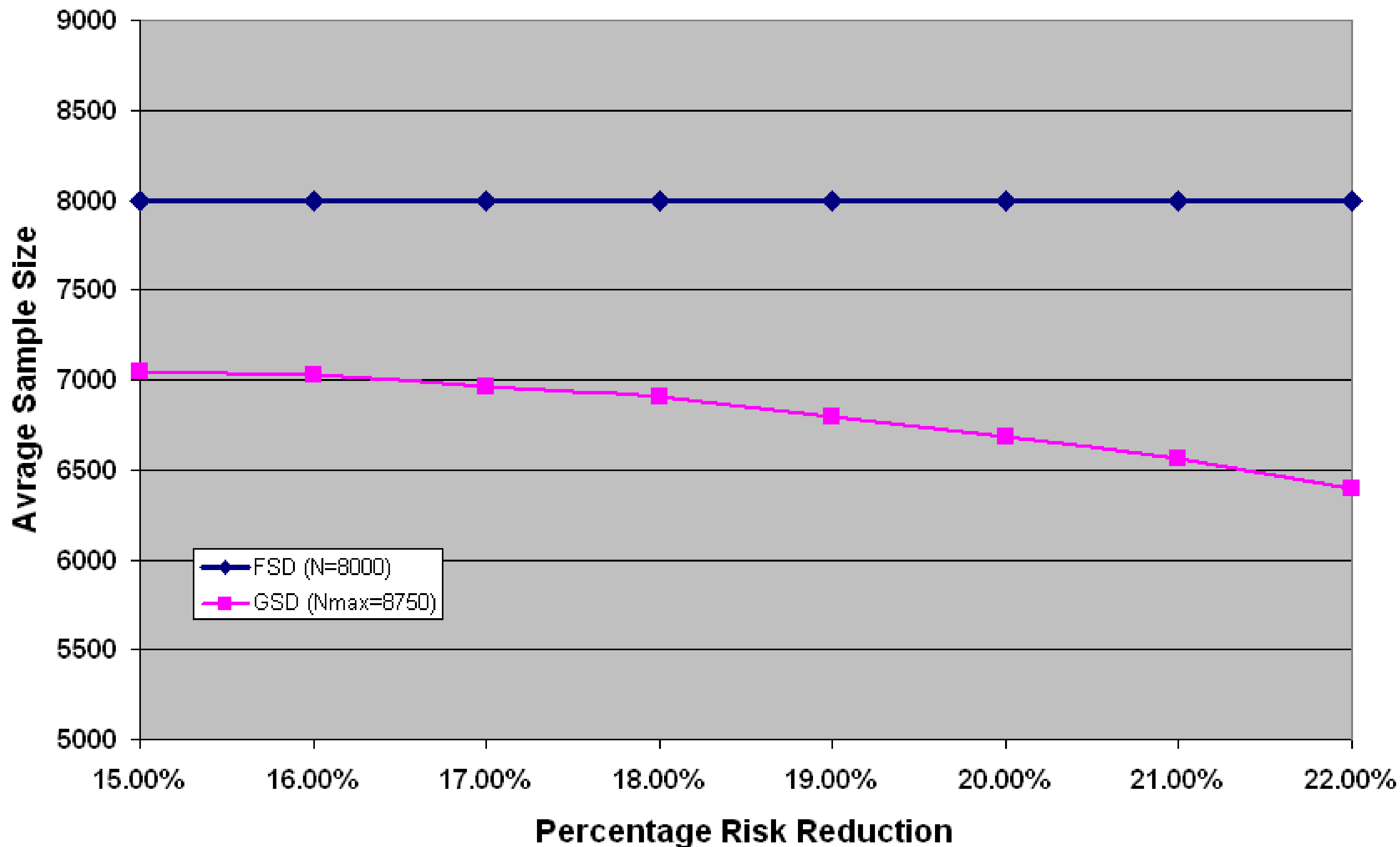
Cytel STATISTICAL SOFTWARE & SERVICES



Cytel STATISTICAL SOFTWARE & SERVICES



Power Curves for Fixed-Sample and Group Sequential Designs
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
Power
15.00%
16.00%
17.00%
18.00%
19.00%
20.00%
21.00%
22.00%
Percentage Risk Reduction
FSD (N=8000)
GSD (Nmax=8750)



Cytel STATISTICAL SOFTWARE & SERVICES

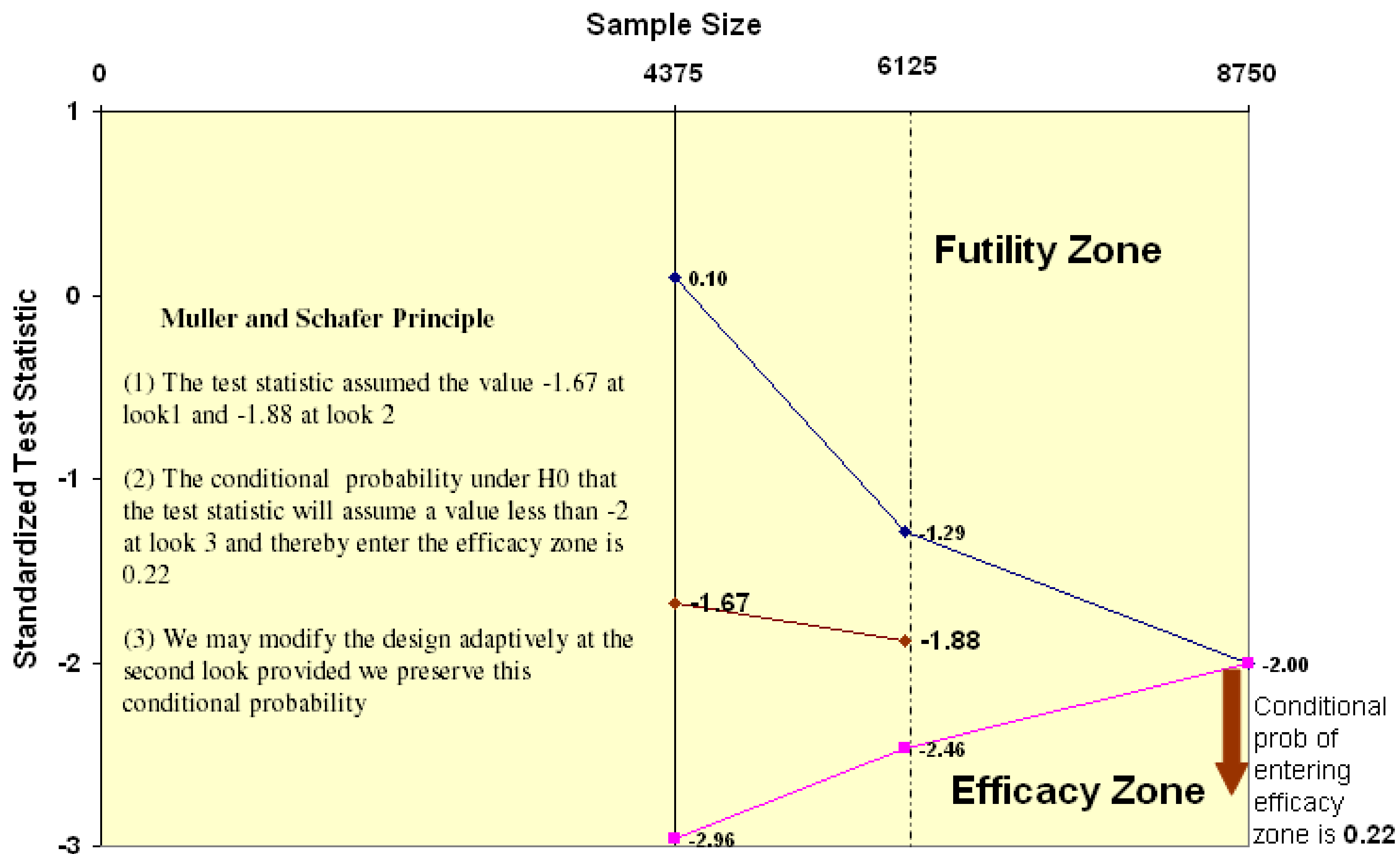
3. GSD with Adaptive Sample Size Increase at IA2

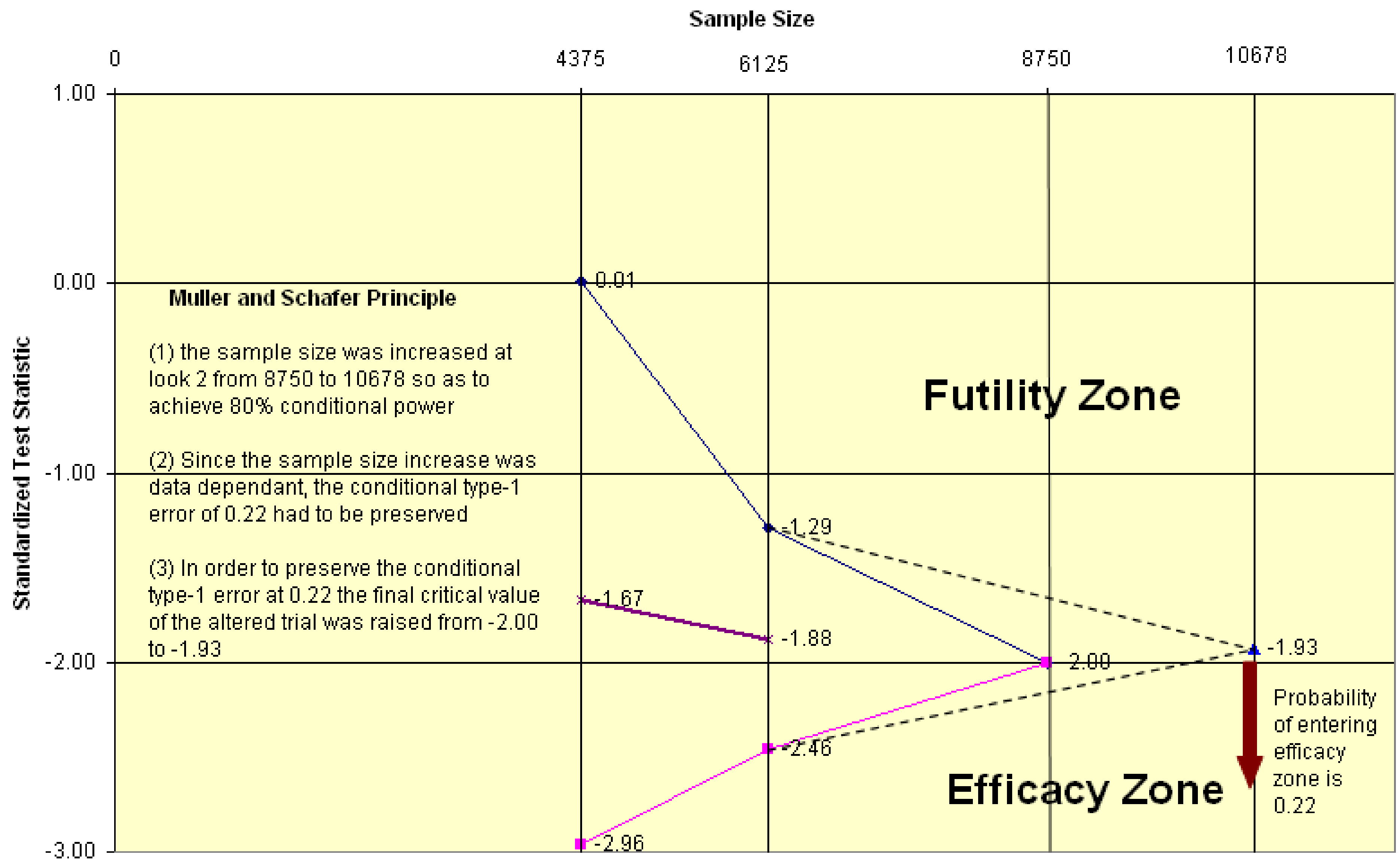
- Suppose neither efficacy nor futility boundaries are crossed at IA2
- Is conditional power less than 80%?
- If so, increase sample size such that conditional power equals 80%
- Subject to upper limit of 15,000 subjects

Cytel
STATISTICAL SOFTWARE & SERVICES

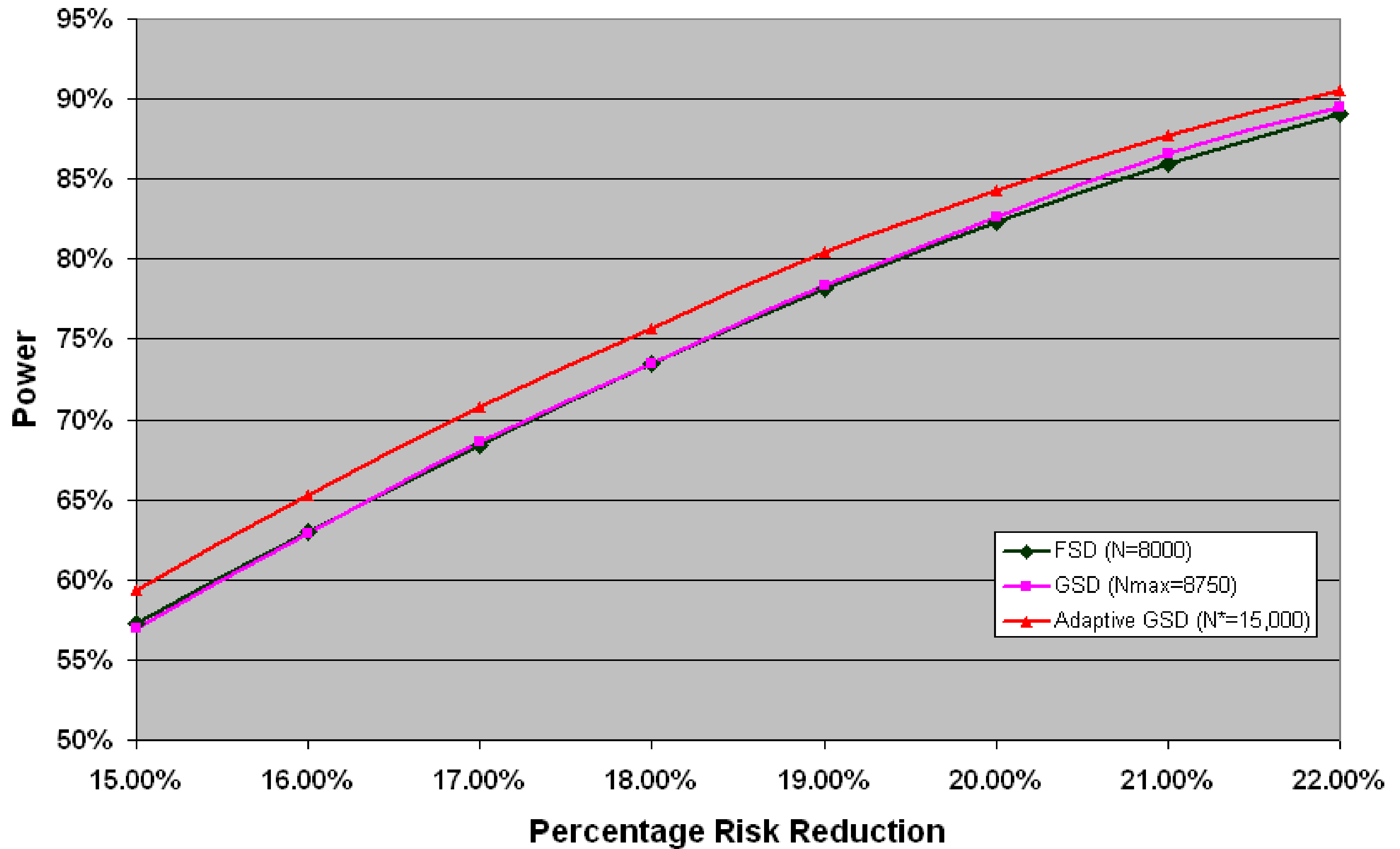
Preservation of Type-1 Error after an Adaptive Design Change

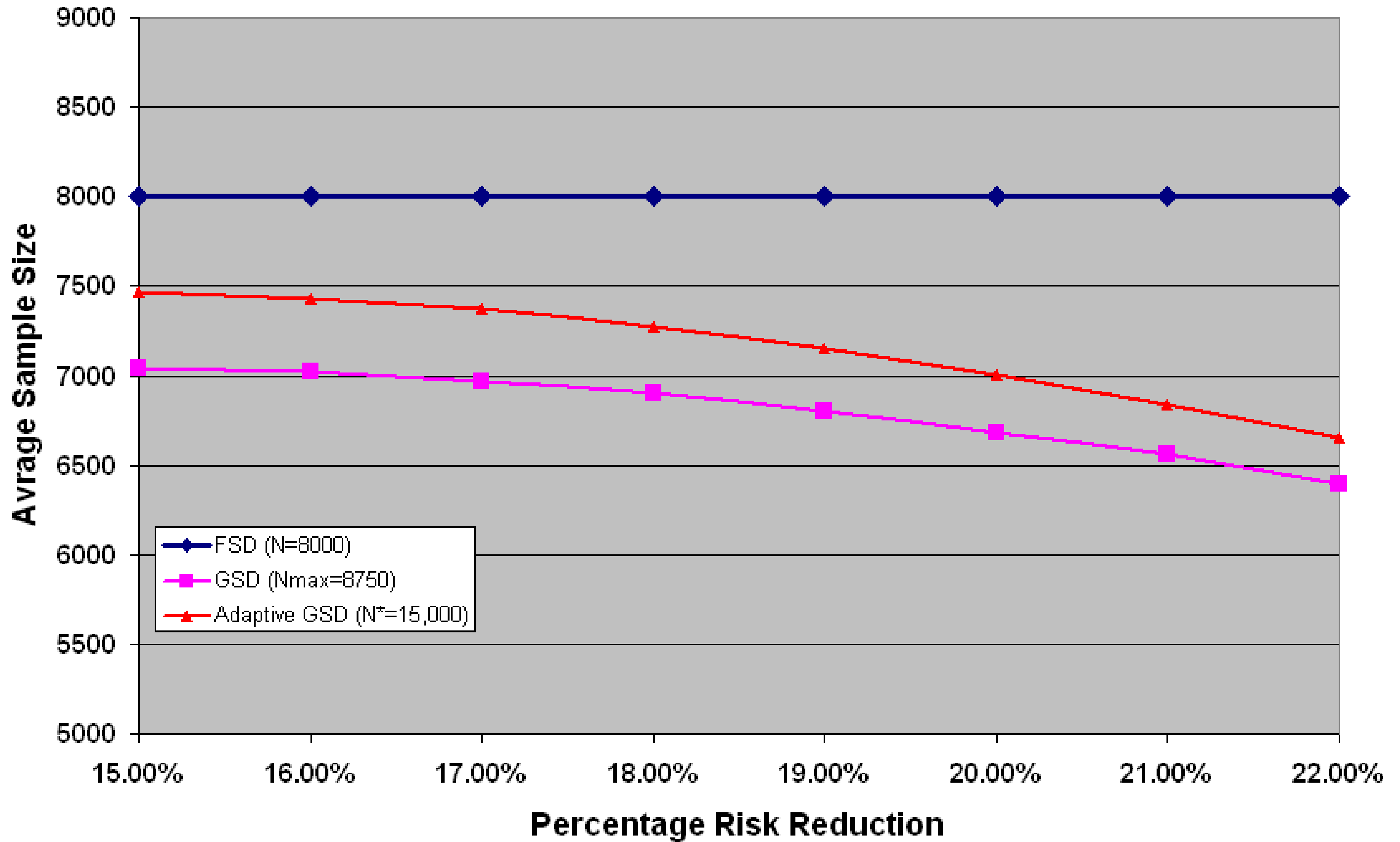
- Unblinded sample size increase can inflate α
- To control this you must preserve the conditional type-1 error at the time of the adaptive change of design
- If you pre-specify that you will do this anytime you make an adaptive change, the overall type-1 error will still be α (Muller and Schafer, 2001)
- Allowable adaptive changes include: sample size change, change in number of future looks, change in future stopping boundaries, change in eligibility criteria


Group Sequential Design; 8750 Patients; Conditional Type-1 Error is 0.22
Sample Size
0
4375
6125
8750
1
0
-1
-2
-3
Standardized Test Statistic
Futility Zone
0.10
Muller and Schafer Principle
(1) The test statistic assumed the value -1.67 at look1 and -1.88 at look 2
(2) The conditional probability under H0 that the test statistic will assume a value less than -2 at look 3 and thereby enter the efficacy zone is 0.22
(3) We may modify the design adaptively at the second look provided we preserve this conditional probability
-1.29
-1.67
-1.88
-2.00
Conditional prob of entering efficacy zone is 0.22
-2.46
Efficacy Zone
-2.96



Adaptive Sample Size Increase Keeping Conditional Type 1 Error Unchanged
Sample Size
0
4375
6125
8750
10678
1.00
0.00
-1.00
-2.00
-3.00
Standardized Test Statistic
Muller and Schafer Principle
(1) the sample size was increased at look 2 from 8750 to 10678 so as to achieve 80% conditional power
(2) Since the sample size increase was data dependant, the conditional type-1 error of 0.22 had to be preserved
(3) In order to preserve the conditional type-1 error at 0.22 the final critical value of the altered trial was raised from -2.00 to -1.93
0.01
-1.29
-1.67
-1.88
-2.00
-1.93
-2.46
-2.96
Futility Zone
Efficacy Zone
Probability of entering efficacy zone is 0.22


Power Curves for Fixed-Sample, Group Sequential and Adaptive Group Sequential Designs
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
Power
15.00%
16.00%
17.00%
18.00%
19.00%
20.00%
21.00%
22.00%
Percentage Risk Reduction
FSD (N=8000)
GSD (Nmax=8750)
Adaptive GSD (N*=15,000)



Cytel STATISTICAL SOFTWARE & SERVICES

4. GSD with Sample Size Increase and Population Enrichment

- There is scientific evidence that some subgroups might benefit more than others from the new drug
- Sponsor wishes to take advantage of this possibility
- Design must allow drilling down into subgroups if overall benefit cannot be established

Qualifications for Subgroups

- Likely to influence the event rates
- More likely to influence experimental group than control group
- Easily identifiable in a clinical setting
- Consistently collected in the trial

Prior Selection of Subgroup

1. Using clinical judgement, create four non-overlapping partitions of the overall population

Non-Overlapping Partition	% of Population
high-risk and clopidogrel-naive	30%
high-risk and pre-treated with clopidogrel	30%
low-risk and clopiddogrel-naive	20%
low-risk and pre-treated	20%

high risk patients are those with diabetes and a troponin + marker

Cytel STATISTICAL SOFTWARE & SERVICES

2. **Form three nested subgroups from these four non-overlapping partitions:**
   - G_0 **= full population (100%)**
   - G_1 **= high-risk subgroup (60% of** G_0**)**
   - G_2 **= high-risk, clopidogrel-naive subgroup (50% of** G_1**)**
3. **The goal is to try and win with** G_0**; if not, then with** G_1**; if not, then with** G_2

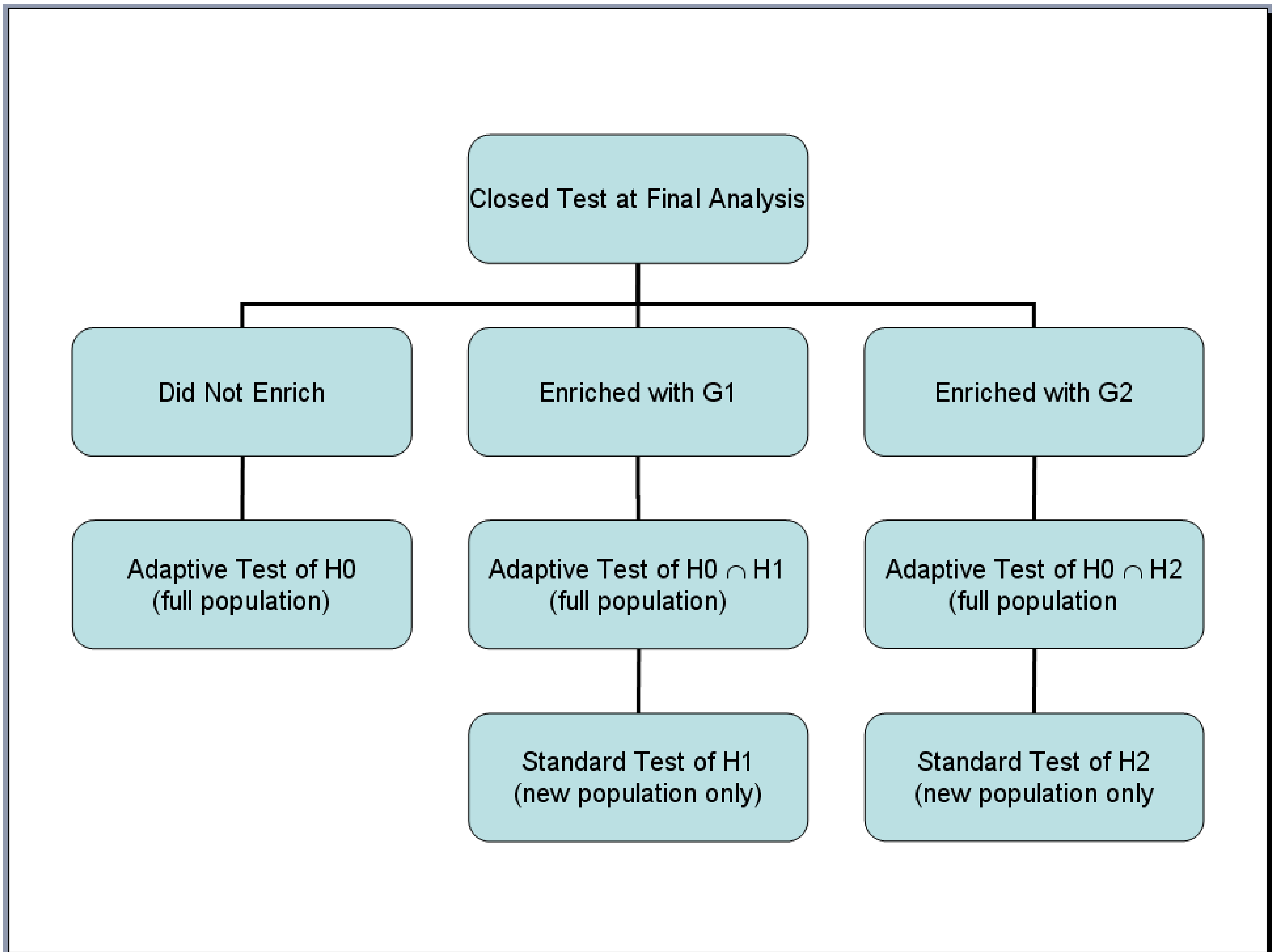
The Enrichment Strategy

Suppose the group sequential boundary was not crossed at IA1 or IA2. Then:

- If CP $\geq 80\%$ at IA2, carry on with no change
- If CP $< 80\%$ at IA2:
  - Try for 80% with G_0 and sample size increase
  - If unable, try for 80% CP by enriching with only G_1 patients and sample size increase
  - If unable, try for 80% CP by enriching with only G_2 patients and sample size increase
- Terminate for futility if CP $< 20\%$ despite enrichment and sample size increase

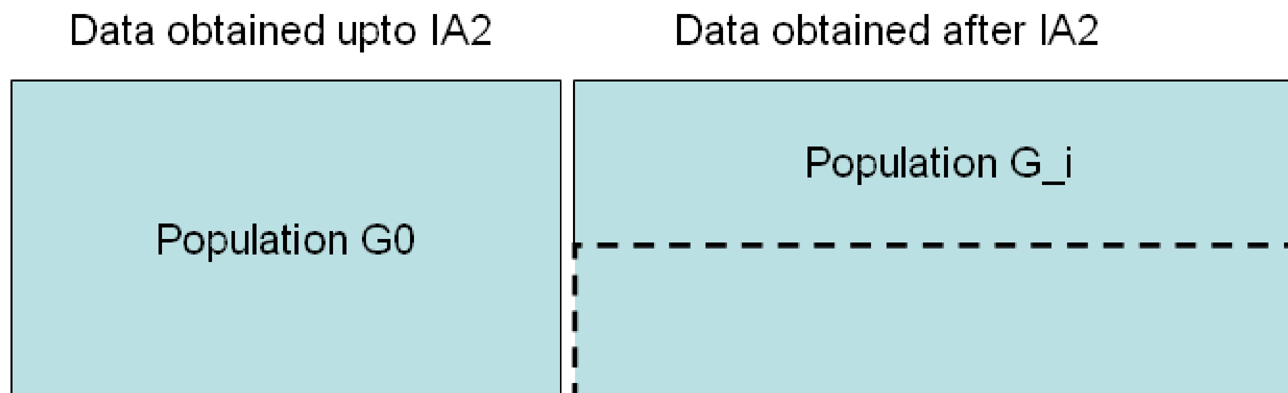
Hypothesis Testing Strategy

- For populations $\{G_1, G_2, G_3\}$, let $\{H_1, H_2, H_3\}$ be corresponding null hypotheses that treatment and control event rates are equal
- Consider three cases:
  - Case 1 if population was not enriched, test H_0 at final analysis
  - Case 2: if population was enriched by G_1, test H_1 at final analysis
  - Case 3: population was enriched by G_2, test H_2 at final analysis
- The testing strategy must ensure strong control of type-1 error. This means that no matter which hypothesis is tested, the false positive rate must not exceed α



Closed Test at Final Analysis
Did Not Enrich
Enriched with G1
Enriched with G2
Adaptive Test of H0
(full population)
Adaptive Test of H0 ∩ H1
(full population)
Adaptive Test of H0 ∩ H2
(full population
Standard Test of H1
(new population only)
Standard Test of H2
(new population only

Graphical Depiction of Hierarchical Testing



- **Test 1** is performed with combined data from G0 and G_i
- If G_i is also G0 (i.e., no enrichment), no further testing
- If G_i is either G1 or G2 (i.e., enrichment) perform **Test 2**
  - **Test 2** is only performed with the new data obtained after IA2

Proof that Test Procedure is Closed

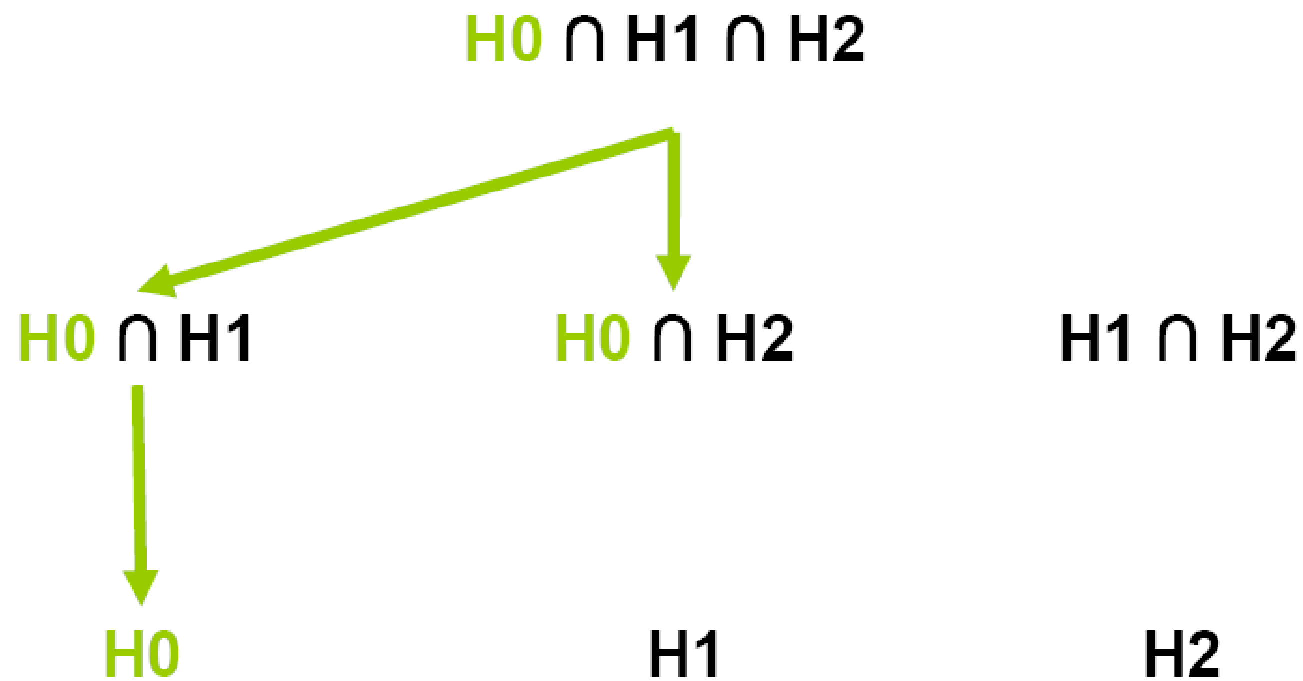
$H0 \cap H1 \cap H2$

$H0 \cap H1$ $H0 \cap H2$ $H1 \cap H2$

H0 H1 H2

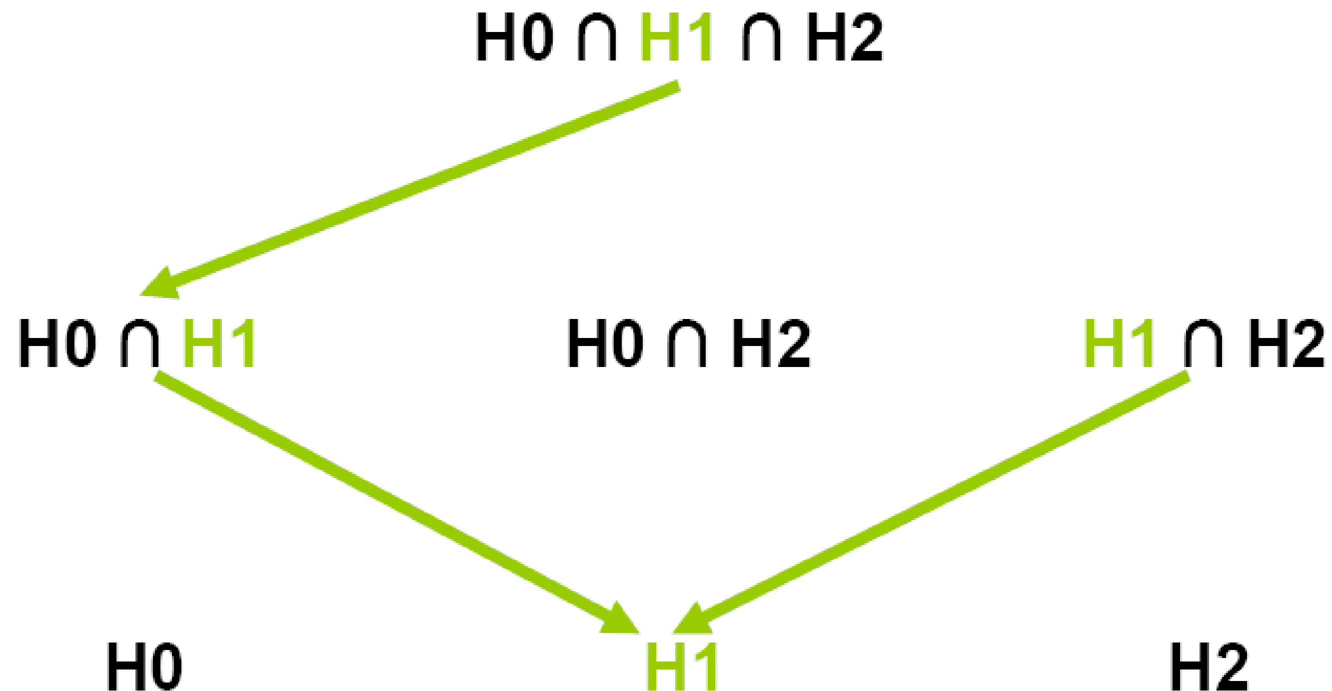
To reject any hypothesis you must also reject all intersections of that hypothesis with all other hypotheses

Case 1: If you don't enrich,all these tests must be rejected in order to claim G0



If you don't enrich, you must reject H0 at level-alpha in order to win with G0. Test 1 (on the total population) is an adaptive level-alpha test of H0. If it rejects, all the other tests indicated automatically reject also at level-alpha.

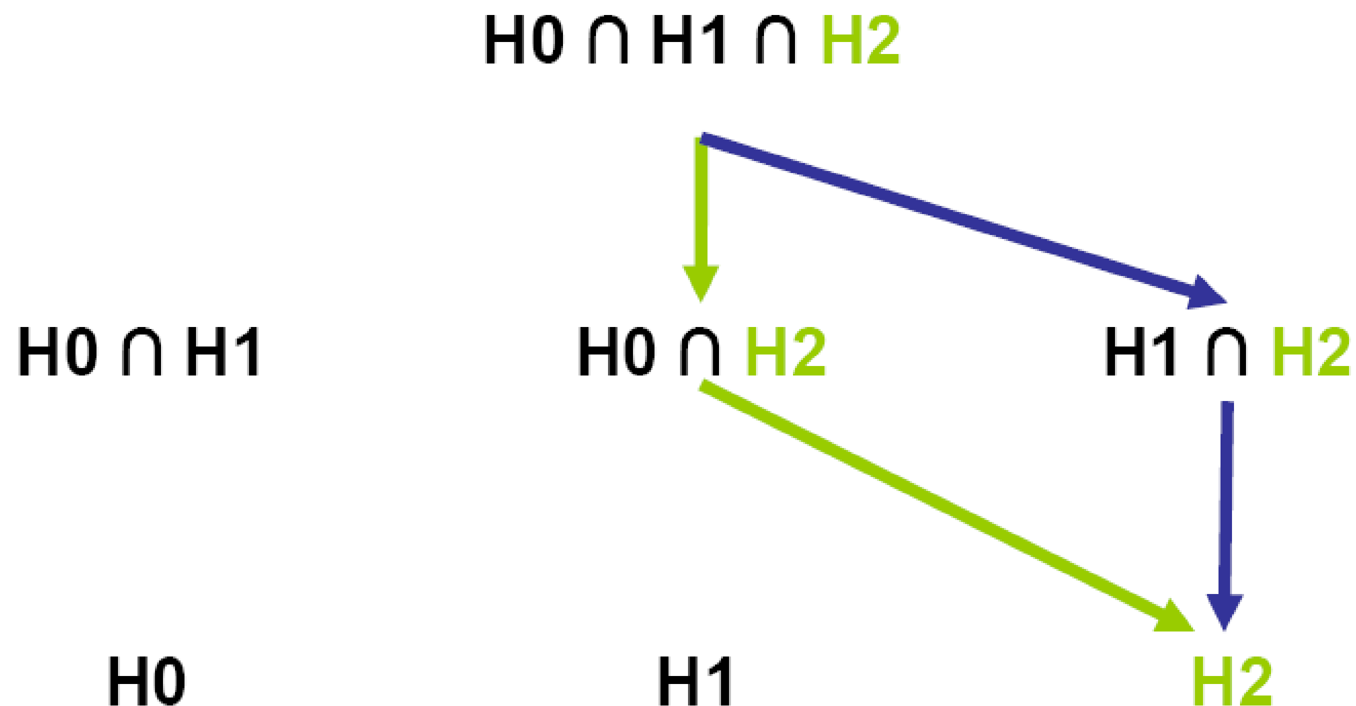
Case 2: If you enrich with G1, all these tests must be rejected to claim G1



If you enrich with G1 you must first reject (H0∩H1) with <u>Test 1</u> , the "gatekeeper". This is a level-alpha adaptive test. Then you must reject H1 with <u>Test 2</u> using only the new G1 data . This is a conventional level-alpha test. If it rejects, all other tests indicated reject automatically at level-alpha

Cytel STATISTICAL SOFTWARE & SERVICES

Case 3: If you enrich with G2, all these tests must be rejected to claim G2



If you enrich with G2 you must first reject (H0∩H2) with <u>Test 1</u> , the "gatekeeper". This is a level-alpha adaptive test. Then you must reject H2 with <u>Test 2</u> using only the new G2 data . This is a conventional level-alpha test. If it rejects all other tests indicated automatically reject at level-alpha

Properties of Testing Strategy

- Closed test!. Guarantees strong control of type-1 error; i.e., chance of false rejection of either H_0 or H_1 or H_2 cannot exceed α
- Boosts power by 2-5% if new treatment benefits all subgoups homogeneously
- Boosts power by 7-18% if new treatment benefits G_1 or G_2 differentially relative to G_0

Cytel STATISTICAL SOFTWARE & SERVICES

Verify by Simulation

Simulations for TheXYZ Company

Population

Parameters of Atomic Subgroups

Parameter	HighR + Naïve	HighR + PreTrt	LowR + Naïve	LowR + PreTrt	
Subgr. Fr.	0.30	0.30	0.20	0.20	
π_c	0.120	0.090	0.070	0.050	0.087
π_t	0.084	0.072	0.063	0.048	
% Improvement	30.0%	20.0%	10.0%	4.0%	

Subpopulations of interest — Refresh

Subpop.	Subpop. Fraction	Event Rates π_c	Event Rates π_t	% Improv.	Difference δ	Discount Factor
G_0	1.00	0.087	0.069	20.7%	0.018	1.00
G_1	0.60	0.105	0.078	25.7%	0.027	1.00
G_2	0.30	0.120	0.084	30.0%	0.036	1.00

G_0: All Patients, G_1: HighR Patients, G_2: HighR & Plavix Naïve Patients

Test

Type-I Error (α)	0.025
Initial Total Sample Size (Both Arms; n_{max})	8000

Look #	SS (Both Arms)	Stopping Rule: Efficacy		Stopping Rule: Futility	
1	4000	Z ≥	2.9626	% Impr. <	-1%
2	5600	Z ≥	2.4623	CP <	0.2
3	8000	Z ≥	2.0018		

Adaptation

Max. Sample Size With Adaptation (Both Arms; n^*_{max})	15000
Adaptation Criterion: CP ≥ Upper Threshold	0.8

Approach to Sample Size Re-estimation
- (•) Single Trial
- ( ) Separate Trials

☐ Disable Enrichment

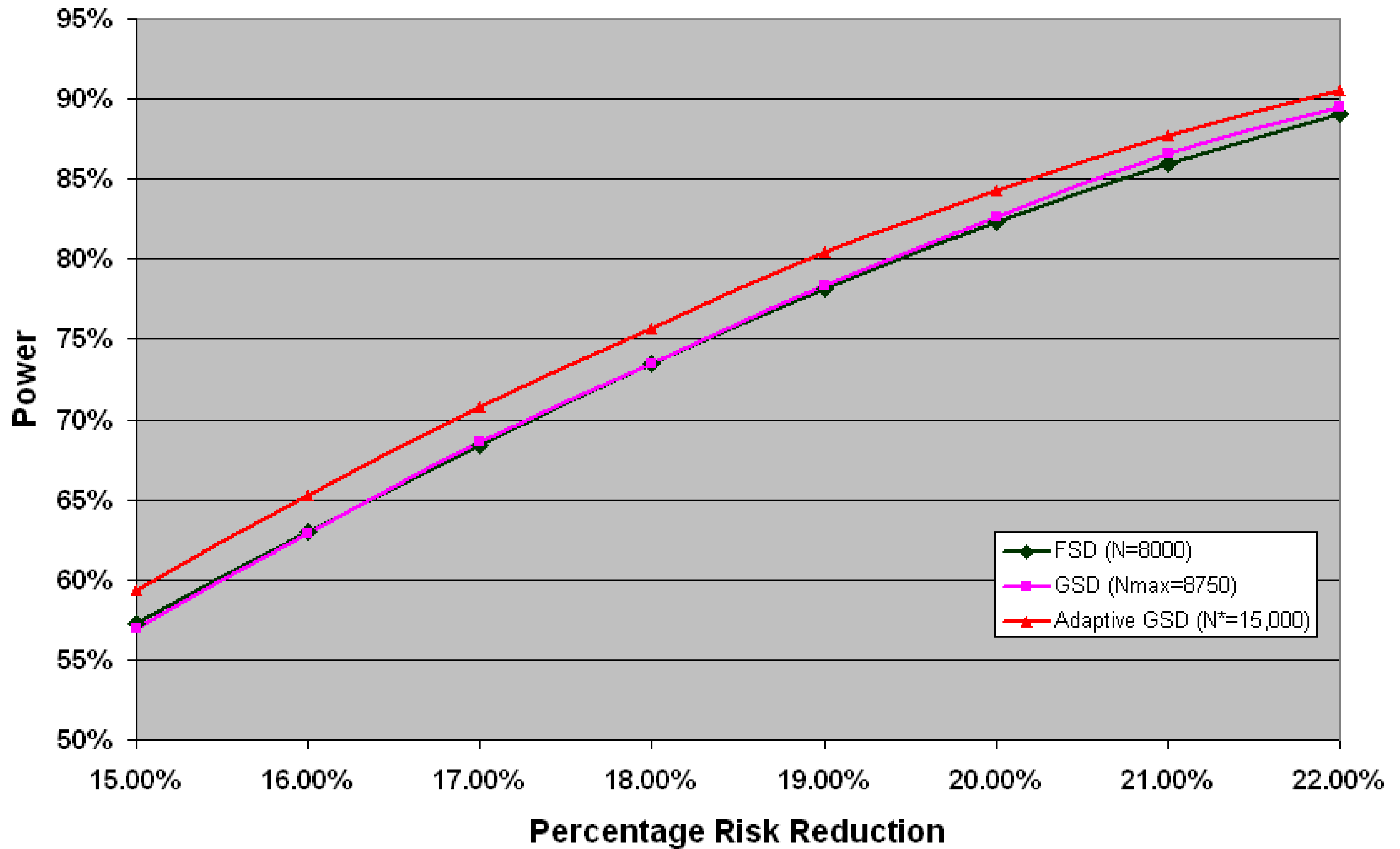
Total Number of Simulations	10000
Initial Random Number Seed	clock

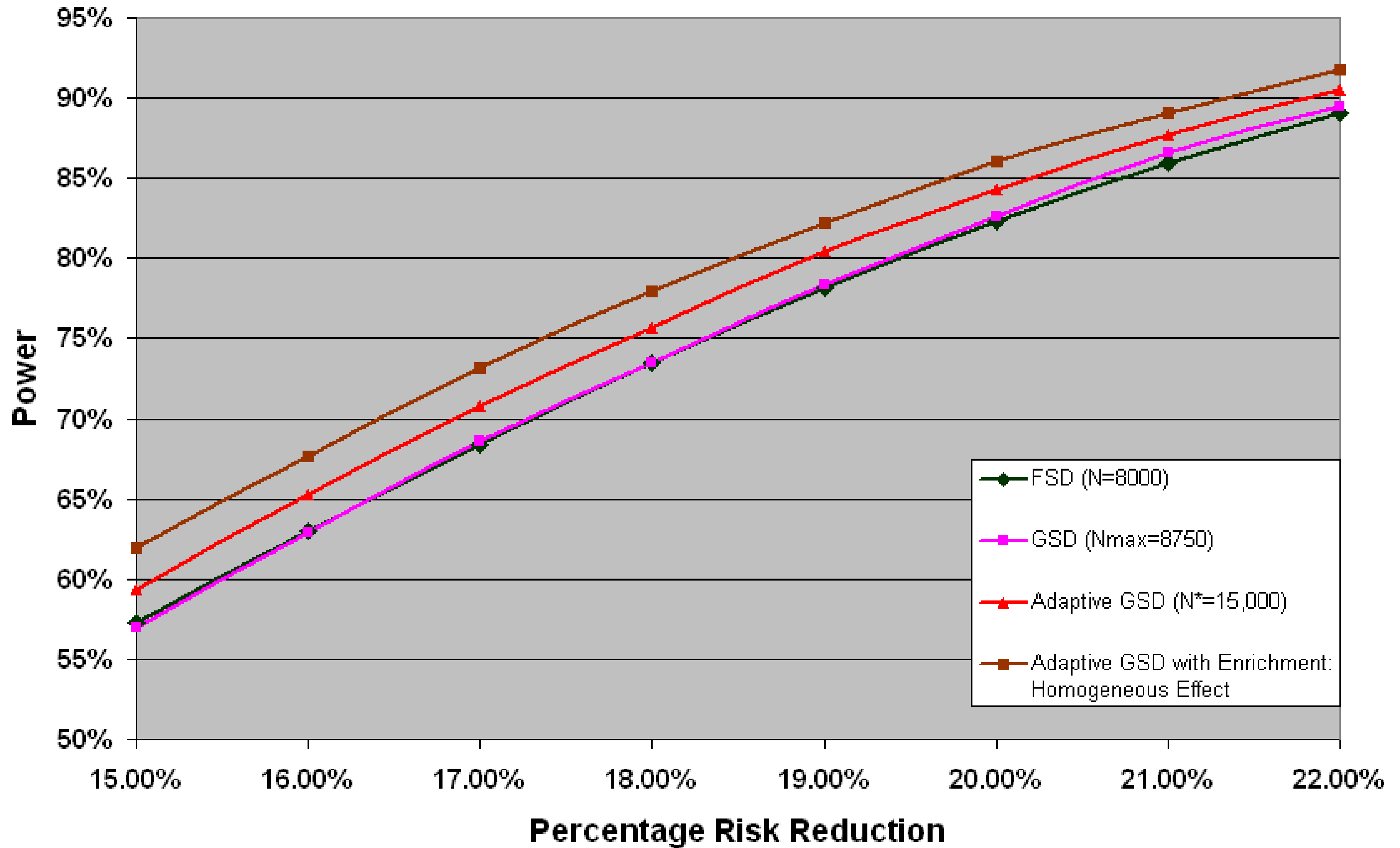
Run — Stop — Reset

☐ Compute Power with Fixed Sample Size (Both Arms)

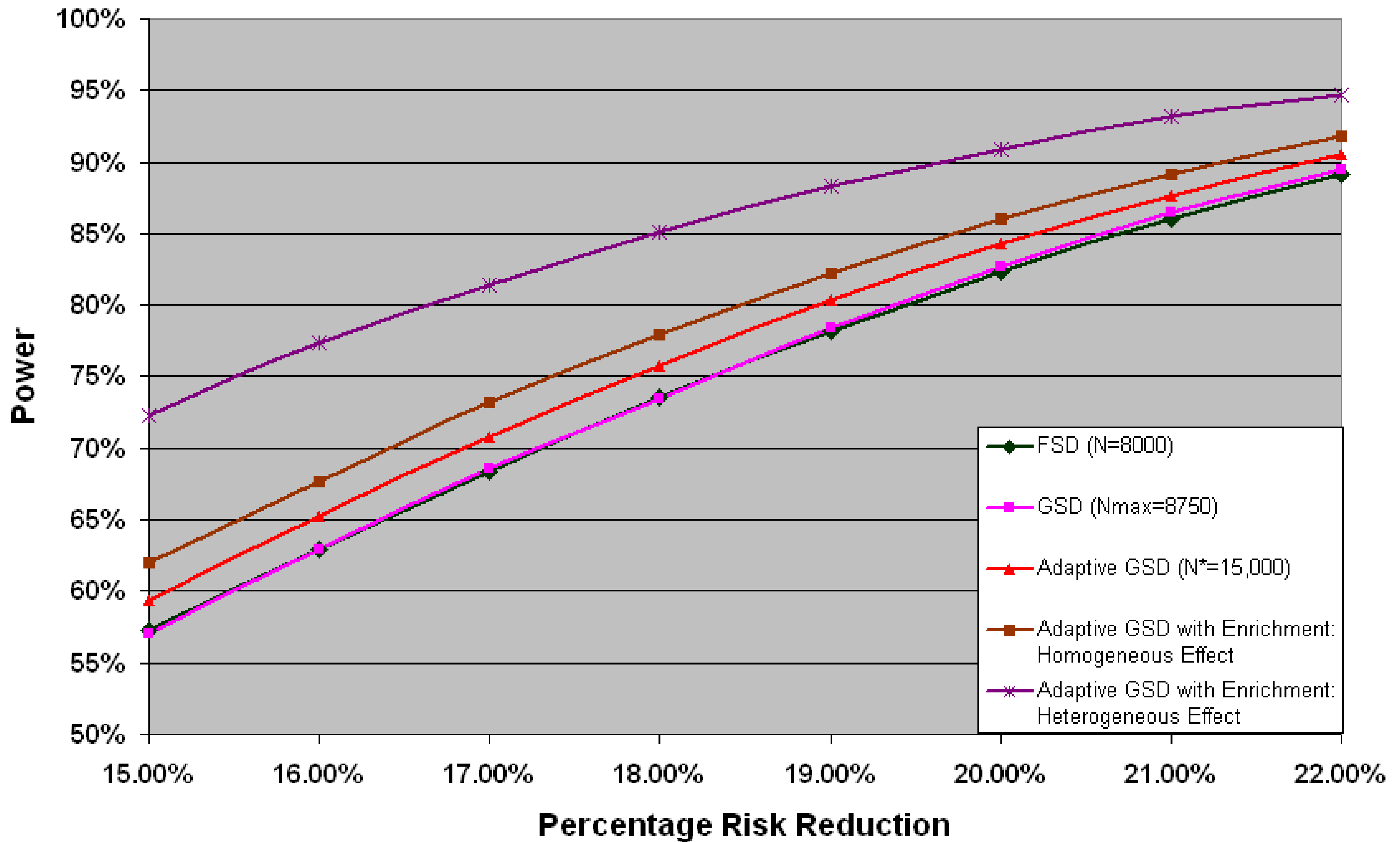
n_{max}		n^*_{max}	

Cytel STATISTICAL SOFTWARE & SERVICES

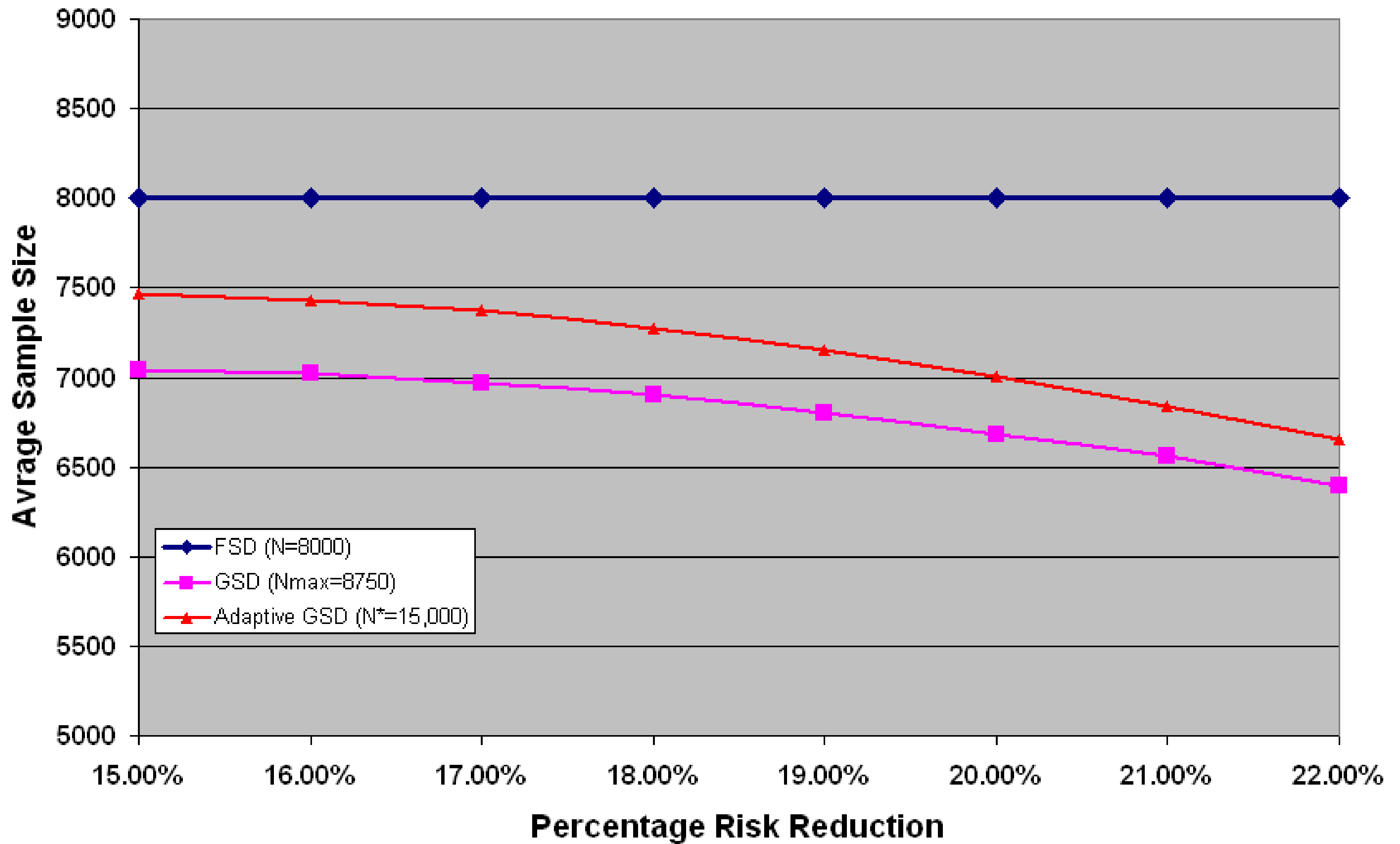

Power Curves for Fixed-Sample, Group Sequential and Adaptive Group Sequential Designs
Power
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
15.00%
16.00%
17.00%
18.00%
19.00%
20.00%
21.00%
22.00%
Percentage Risk Reduction
FSD (N=8000)
GSD (Nmax=8750)
Adaptive GSD (N*=15,000)



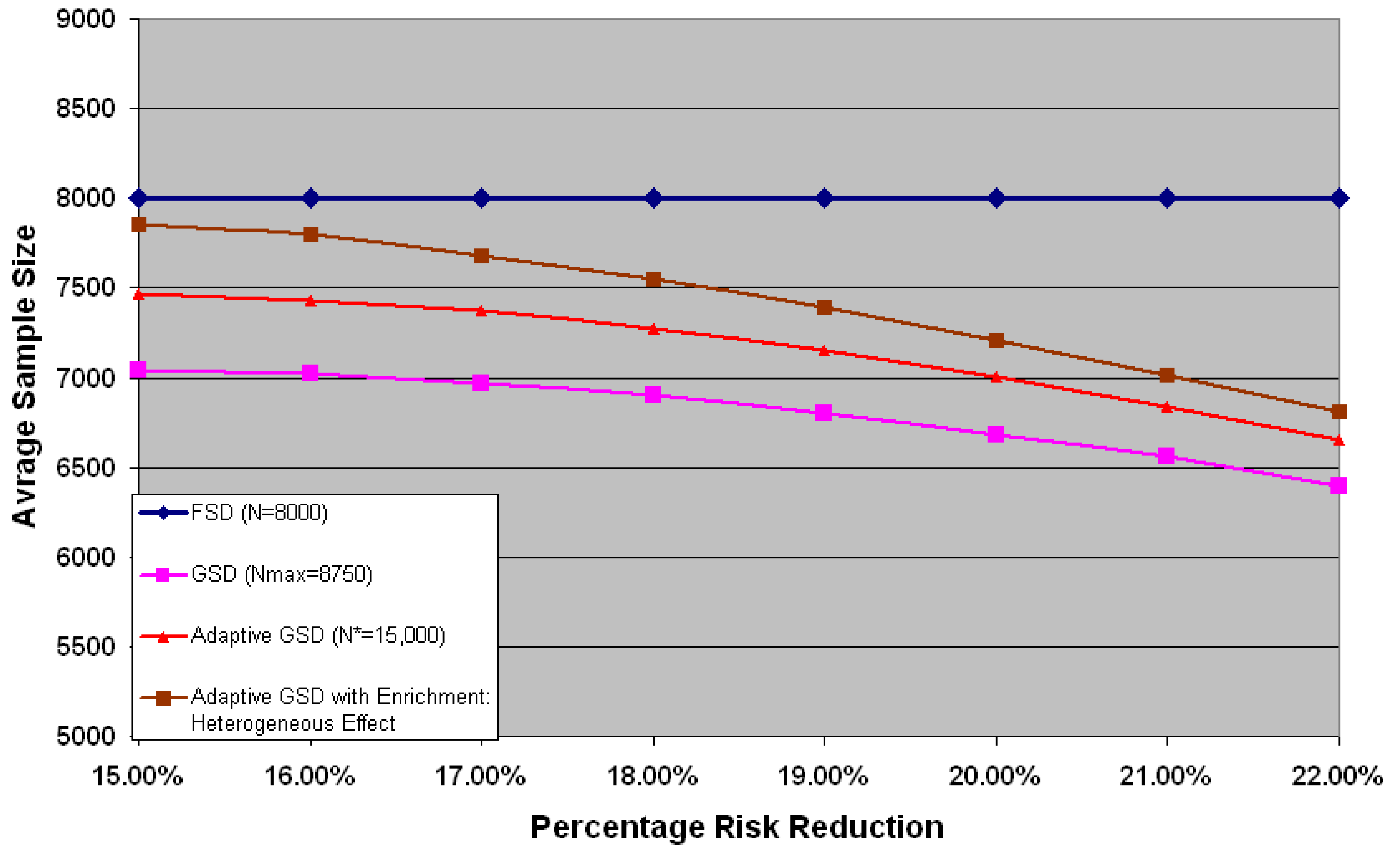
Cytel STATISTICAL SOFTWARE & SERVICES

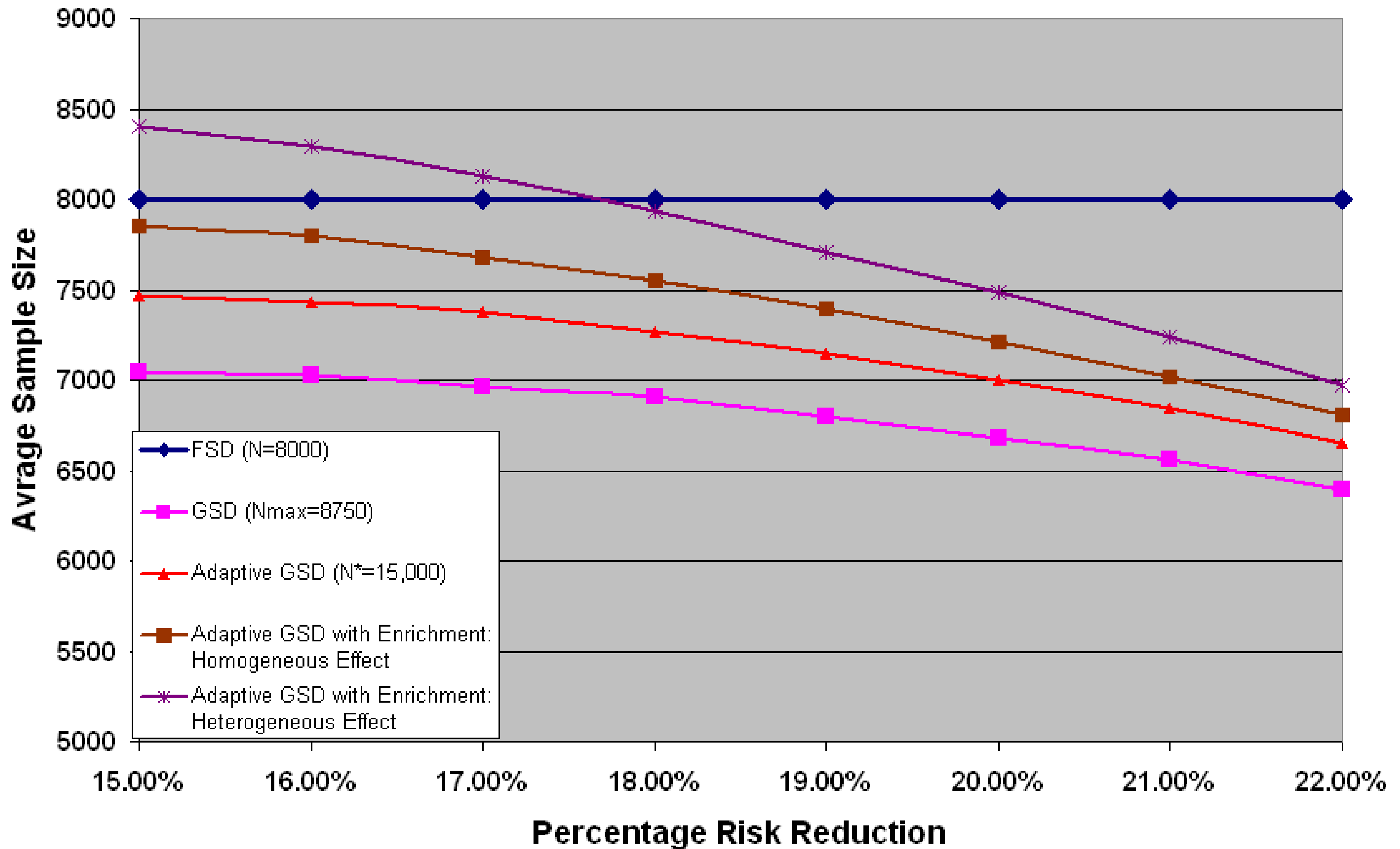


Cytel STATISTICAL SOFTWARE & SERVICES



Average Sample Sizes for Fixed Sample, Group Sequential and Adaptive Group Sequential
Avrage Sample Size
9000
8500
8000
7500
7000
6500
6000
5500
5000
FSD (N=8000)
GSD (Nmax=8750)
Adaptive GSD (N*=15,000)
15.00%
16.00%
17.00%
18.00%
19.00%
20.00%
21.00%
22.00%
Percentage Risk Reduction


Average Sample Sizes for Fixed Sample, Group Sequential Adaptive Group Sequential and Enrichment under Homogeneous Treatment Effect
9000
8500
8000
7500
7000
6500
6000
5500
5000
Avrage Sample Size
FSD (N=8000)
GSD (Nmax=8750)
Adaptive GSD (N*=15,000)
Adaptive GSD with Enrichment: Heterogeneous Effect
15.00%
16.00%
17.00%
18.00%
19.00%
20.00%
21.00%
22.00%
Percentage Risk Reduction


Average Sample Sizes for Fixed Sample, Group Sequential, Adaptive Group Sequential, and Enrichment Designs under Homogeneous and Heterogeneous Treatment Effect
Avrage Sample Size
9000
8500
8000
7500
7000
6500
6000
5500
5000
15.00%
16.00%
17.00%
18.00%
19.00%
20.00%
21.00%
22.00%
Percentage Risk Reduction
FSD (N=8000)
GSD (Nmax=8750)
Adaptive GSD (N*=15,000)
Adaptive GSD with Enrichment: Homogeneous Effect
Adaptive GSD with Enrichment: Heterogeneous Effect

Concluding Remarks

- Why not run two independent trials; exploratory and confirmatory?
  - exploratory trial won't be adequately powered for subgroups
  - if it fails, there won't be any support for the confirmatory trial
  - current design contains an insurance policy for handling subgroups
- Pre-specification of subgroups not a statistical requirement. But important safeguard against chasing random noise
- Flexiblity to change sample size and enrich population greatly increases complexity of trial management
  - Logistics of drug are complicated if sample size is increased at IA2
  - Recruitment might slow down if population is enriched at IA2
- Sponsor cannot be involved in interim decision making. Must appoint independent committee with clinical and statistical expertise
- Simulate the design thoroughly before committing to it