.

32nd Annual Symposium, NJ ASA Chapter of ASA, 2011

Bayesian and Frequentist Adaptive Designs in Clinical Trials

Scott Evans, Ph.D.
Senior Research Scientist, Department of Biostatistics, Harvard School of Public Health, MA

Scott Berry, Ph.D.
President and Senior Statistical Scientist
Berry Consultants, TX

Gene Pennello, Ph.D.
Team Leader, Diagnostics Devices Branch
FDA/CDRH/OSB/Division of Biostatistics, MD

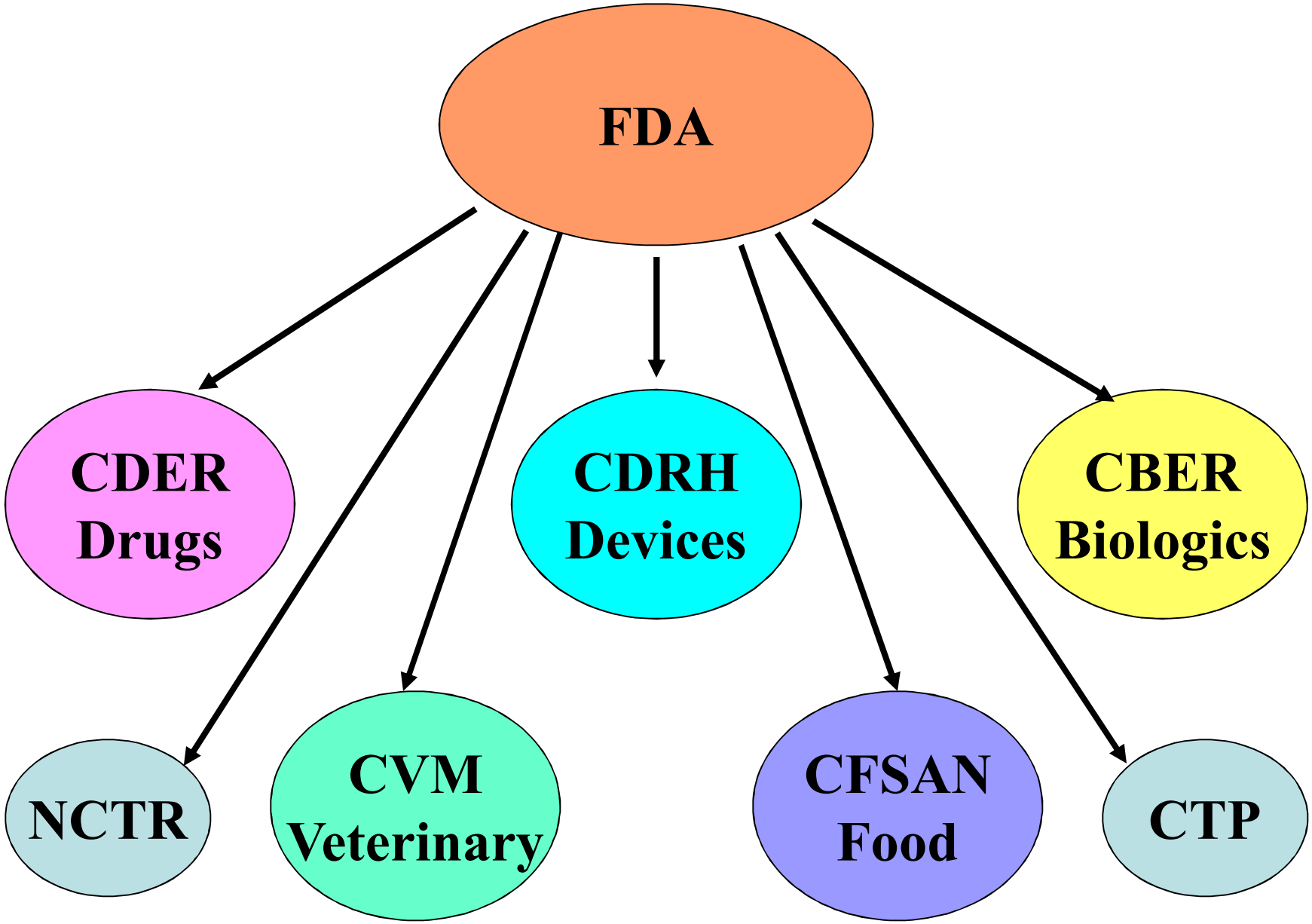
Qing Liu, Ph.D.
Senior Research Fellow, Johnson and Johnson, NJ

.

32nd Annual Symposium, NJ ASA Chapter of ASA, 2011

Bayesian and Frequentist Adaptive Designs for Studies of Medical Devices

Gene Pennello, Laura Thompson, and Tom Gwise
Division of Biostatistics
Center for Devices and Radiological Health


FDA
CDER
Drugs
CDRH
Devices
CBER
Biologics
NCTR
CVM
Veterinary
CFSAN
Food
CTP

Therapeutic & Aesthetic Devices

Relatively Simple Devices
tongue depressors
latex gloves
simple surgical instruments

Ophthalmic devices
intraocular lenses
PRK lasers

Radiological devices
selective internal radiation (SIR)-sphere cancer therapy

Dental, Ear, Nose, and Throat Devices
hearing aids

Cardiovascular Devices
pacemakers
defibrillators
heart valves
coronary stents
artificial hearts

General, Surgical, and Restorative Devices
artificial hips
spinal fixation devices
artificial skin
prosthetics
breast implants

Diagnostic Devices

Relatively Simple Devices
thermometers

Obstetrics/Gynecology
fetal heart rate monitor
cervical imaging CAD

Radiological devices
digital mammography
MRI machines
CT scanners
computer aided detection (CAD)
FDG-PET scan

ENT Devices
autofluorescent bronchoscopy

Cardiovascular Devices
EKG
cardiac monitoring tools

Monitoring Devices
glucometers
bone densitometers
pulse oximeters

In Vitro Diagnostic Devices
diagnostic test kit for HIV
prostate-specific antigen (PSA)
human papillomavirus (HPV)
prognostic biomarkers
treatment selection biomarkers

FDA Regulatory Science Initiative

- ñThe science of developing new tools, standards and approaches to assess the safety, efficacy, quality and performance of FDA-regulated products.ò

Å Margaret Hamburg, Advancing Regulatory Science*, Science.* 2011.

- ñeffectively translating scientific discoveries into therapies é fully applying knowledge to ensure the safety of food and medical productsò

www.fda.gov/ScienceResearch/SpecialTopics/RegulatoryScience/ucm228131.htm

FDA Regulatory Science Initiative

I. Accelerating the Delivery of New Medical Treatments to Patients

- *"**I-SPY 2 TRIAL**", a groundbreaking new clinical trial model*
- *new statistical approaches to detect changes in process or product quality.*

II. Improving Pediatric and Child Health

- *combine data from multiple clinical studies and to use analytical tools and advanced statistics to assess these data*

III. Protecting Against Emerging Infectious diseases and Terrorism (Medical Countermeasures or MCM)

- *New statistical approaches to assess efficacy when data is limited*
- *advanced, biostatistical approaches are rapidly evolving.*

IV. Enhancing safety and health through informatics

- *Apply appropriate statistical analysis of genomic studies.*

FDA Guidance on Adaptive Design

1. Adaptive Design Clinical Trials for Drugs and Biologics, 2010 (Draft)
   http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm201790.pdf
2. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials, 2010 (Final).
   http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf

Outline

- Adaptive Designs in Medical Device Studies

1. Response-Adaptive Randomization in Trials of a Drug and its Companion Diagnostic
2. Bayesian Sequential Monitoring
   a. Incorporating prior information
   b. Monitoring by maintaining Bayes Risk

- Concluding Remarks

Adaptive Designs in Medical Device Studies

Reasons for Adaptive Designs in Medical Device Studies

- Sample size re-estimation
  - Initial sample size based on highly uncertain effectiveness of device or control (e.g., from feasibility or historical studies).
- Stop early for effectiveness, futility, or safety
  - Bayesian interim monitoring has become an accepted practice for FDA / CDRH (many PMA submissions have used it).
  - Enroll a few patients, evaluate for safety, before deciding to enroll more patients.
- Technological improvements within a class of devices
  - Pressure to reduce sample size for device under study if next generation device is already under development (coronary stent).
  - As new technology becomes available during a trial, randomization to the control may become infeasible (embolic protection).
- Adaptive selection of a subgroup
  - Promising subgroup for a therapeutic device
  - Sensitive subgroup defined by a biomarker.

Examples of Medical Device Adaptive Designs

- Add / drop arms
  - BATTLE, I-SPY futility arm dropping
  - Embolic protection for coronary stenting *(next)*
- Response-Adaptive randomization
  - Extracorporeal membrane oxygenation (ECMO) vs. conventional treatment
  - BATTLE, I-SPY: drugs / putative biomarkers
  - CARA randomization *(Gwise, next)*
- Diagnostic Devices
  - For predicting treatment effect, adaptive biomarker subgroup (Freidlin, Simon, 2005; Zhao, Dmitrienko, Tamura, 2010) and biomarker threshold (Jiang, Freidlin, Simon, 2007) designs.
  - For paired comparisons of AUC, group sequential (Tang, Emerson, Zhou, 2008), sample size re-estimation (Tang, Liu, 2010) designs.

EX. Hybrid Randomized Trials of Embolic Protection Systems

- Embolic protection device used during coronary stenting in saphenous vein bypass grafts.
- Hybrid adaptive design
  - Initial control is no treatment
  - Sites can make a one-time, irrevocable switch of the control arm from no treatment to PercuSurge, the first embolic protection device cleared.
- Multiple hypothesis testing of (30 day) MACE rate
  - superiority to no treatment
  - non-inferiority to PercuSurge
  - May include an interim look for each test

Hybrid Randomized Trials of Embolic Protection Systems

- Embolic Protection Systems with Hybrid Trials
  - Boston Scientific FilterWire
  - Cordis AngioGuard
  - MedNova CardioShield
  - http://www.summitmd.com/pdf/pdf/5case2_slide%281%29.pdf
- Boston Scientific FilterWire
  - http://www.accessdata.fda.gov/cdrh_docs/pdf3/k032884.pdf

Covariate-Adjusted Response-Adaptive (CARA) Randomization for Trials of a Drug with a Companion Diagnostic Biomarker

Tom Gwise and Gene Pennello, Ph.D.
US FDA Center for Devices and Radiological Health
Division of Biostatistics / Diagnostics Branch

Tests that Tailor Therapies

- Safety
  - CYP2D6 genotypesôeffect on metabolic rate of drugs
  - HLA allele B*1502 as a marker for carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis
  - UGT1A1 genotype for risk of neutropenia in CRC patients taking irinotecan
  - KRAS mutation for likely absence of cetuximab, panitumumab efficacy in CRC patients
- Effectiveness
  - HER2 positive breast cancer patient selection for trastuzumab
  - EGFR to select CRC patients for cetuximab, panitumumab.
- Dosing
  - VKORC1 and CYP2C9 genotype to predict warfarin dose.

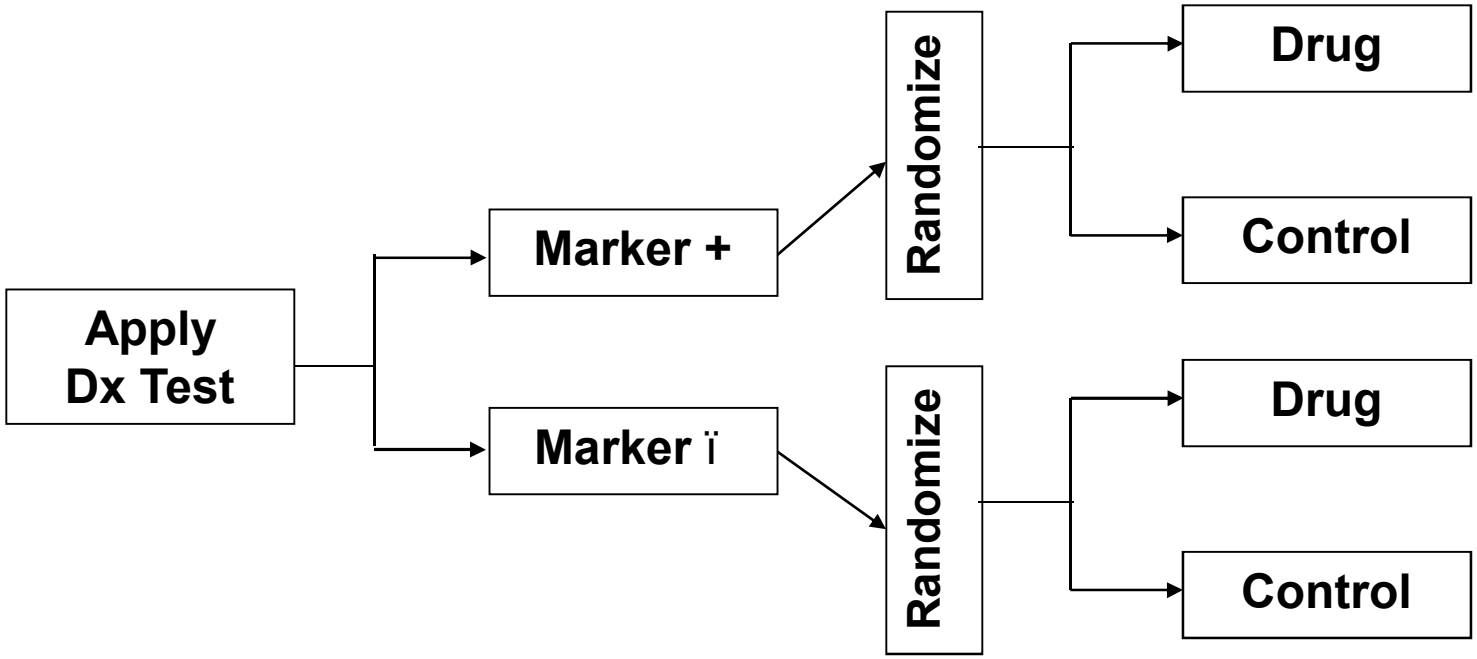
Scenarios for Drug-Dx Combination Products

	Companion Diagnostic	
Drug	Old	New
Old	warfarin:2C9+VKORC1	irinotecan:UGT1A1 cetuximab:KRAS panitumumab:KRAS
New	I-SPY Phase II trials panitumumab:EGFR	trastuzumab:HER2 cetuximab:EGFR

Static Trial Designs

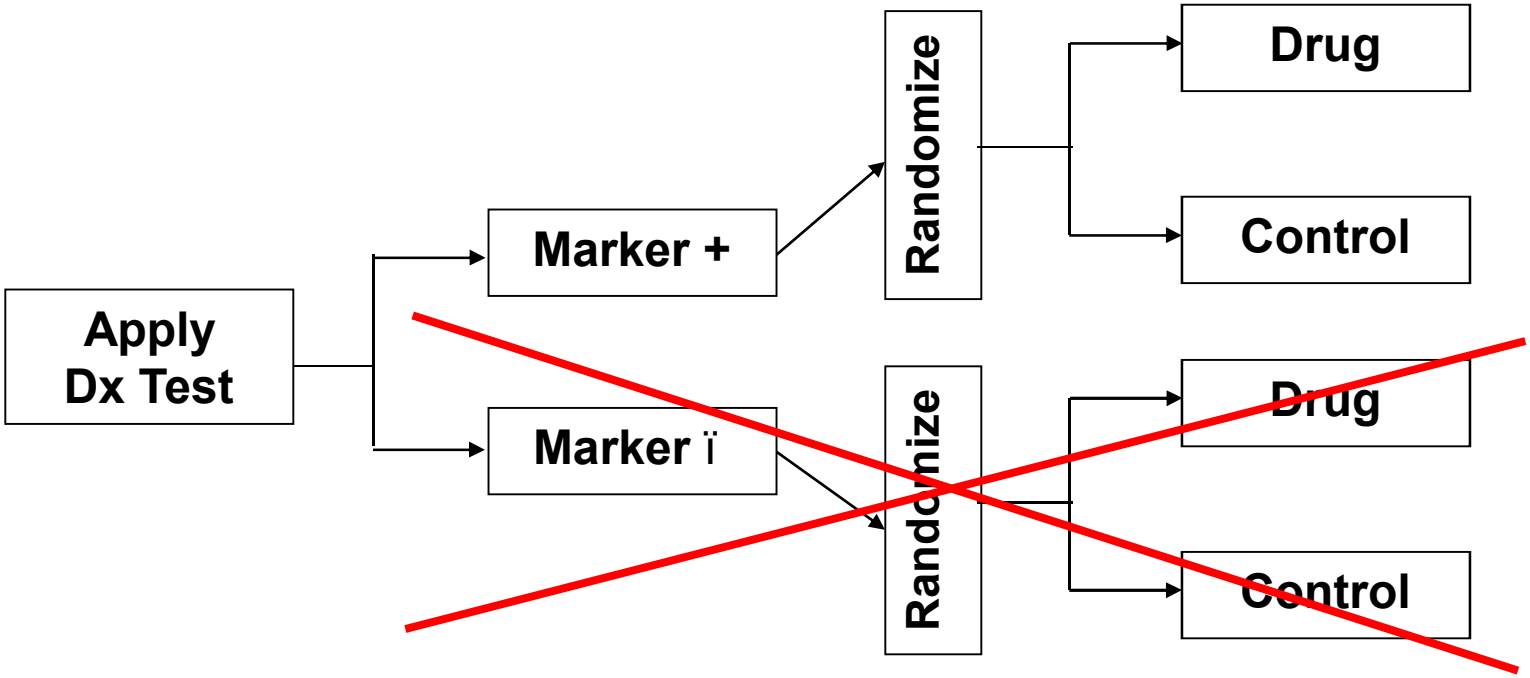
1. Marker by Treatment Design
   - ï Randomization is stratified by marker
2. Targeted Design
   - ï Enroll a subgroup defined by marker.
3. Marker by Treatment Design with Response Adaptive Randomization
   - ï Response adaptive randomization within strata of marker.

1. Marker by Treatment Design



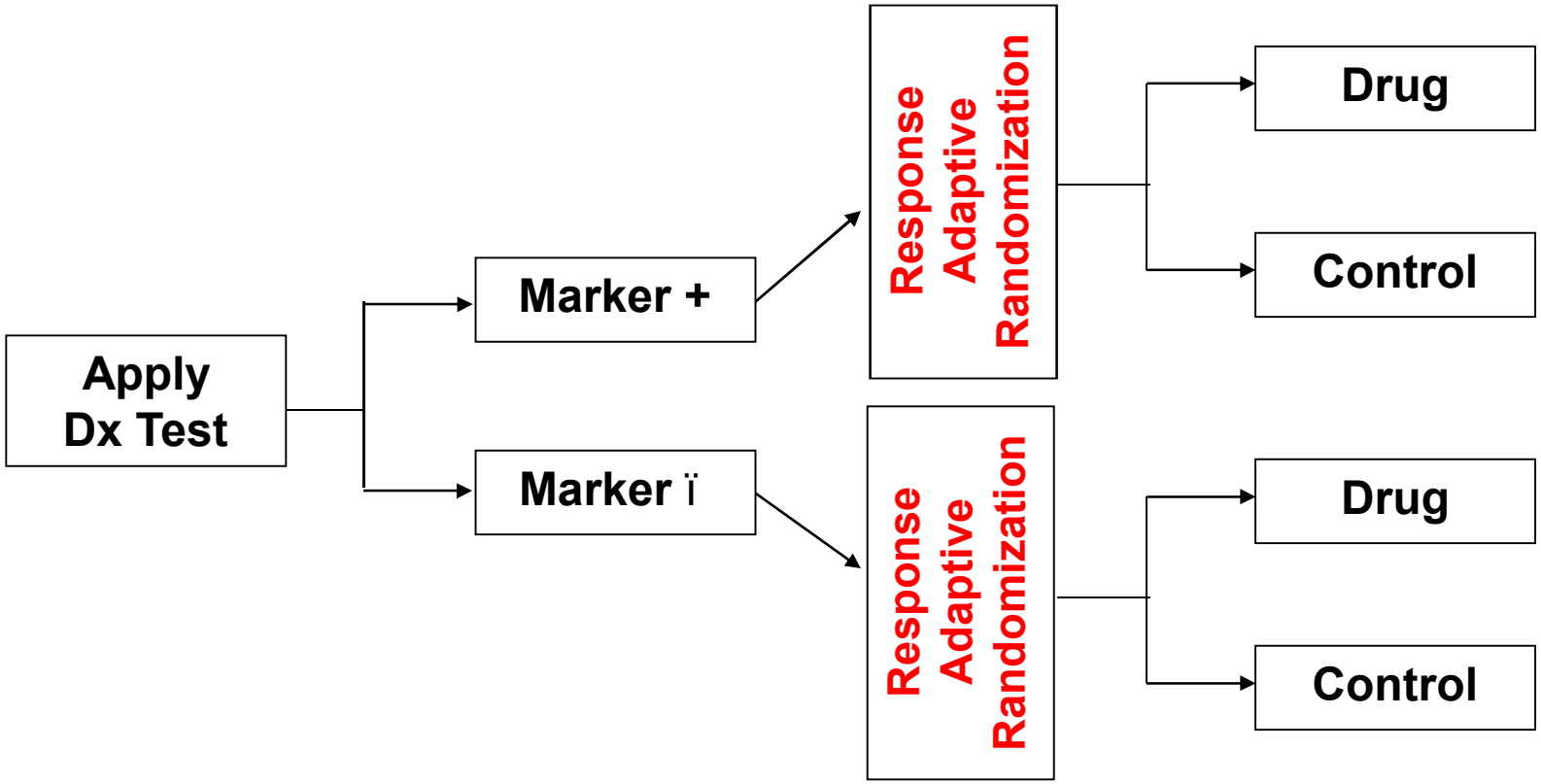
- Dx test must be available at time of study.
- Opportunity to enrich low prevalence subset (including adaptive enrollment of the subset).

2. Targeted Design



- Assumes biology is well-understood
- Can be efficient for studying drug efficacy.
- Test is not studied for effectiveness.
  - ï Test approval, reimbursement problematic.

3. Marker by Treatment CARA Design



Å Covariate-Adjusted Response-Adaptive Randomization (CARA).

Response Adaptive Randomization

- Response information used in setting assignment probabilities for incoming patients
- Examples
  - Randomized play the winner, other urn models
  - Treatment mapping/biased coins
  - Bayesian designs
- Target allocations
  - Minimize treatment failures
  - Maximize power
  - Optimize allocation
- Advantages
  - Ethical
  - Recruitment incentive

Description of Data Source

- Simulations based loosely on Shephard et al, NEJM, 2005.
- Randomized, double-blind, placebo-controlled trial of Erlotinib in previously treated non-small-cell lung cancer.
- Results
  - HR 0.7 (0.6, 0.9) for OS
  - HR 0.7 (0.5-0.9) for patients with EGFR + tumors,
  - HR 0.9 (0.6-1.4) for patients with EGFR tumors
- EGFR expression defined as 10% of cells stained (any intensity) using Dako IHC kit.

Response Adaptive Procedure

- Doubly Adaptive Biased Coin (Eisele, Woodroofe, 1995)
- Assign treatment A with probability defined by a function of the response data (Hu, Zhang, 2004)

$$f(x,y) = \frac{y^{\lambda+1}}{x^{\lambda}}\left(\frac{y^{\lambda+1}}{x^{\lambda}} + \frac{(1-y)^{\lambda+1}}{(1-x)^{\lambda}}\right)^{-1}$$

$x =$ current fraction of patients on treatment A

$y =$ target allocation ratio w.r.t. treatment A

$\lambda =$ controls adaptation variability

$$f(0,y) = 1, \quad f(1,y) = 0, \quad f(x,y) \overset{\lambda=0}{=} y, \overset{\lambda=\infty}{=} 0.5$$

Simulated Clinical Trial Parameters

- Marker by treatment interaction design
  - 1:1 randomization stratified by marker result
  - Response adaptive randomization by marker result
- Binary endpoint: 90-day progression free survival
- P(PFS≥90 days):

	Marker	
Treatment	−	+
A	0.35	0.45
B	0.35	0.35

Simulated Clinical Trial Parameters

- Accrual Rate: Arrival time has exponential distribution with mean of 1 subject per day
- Time to progression (failure) has exponential distribution.
- Test for treatment effects uses normal approx.
- Ď=0.05, ß=0.20, ρ=0.1, ∂=1
- Diagnostic performance:
  - Sensitivity and specificity of marker,
  - Marker by treatment interaction

Response Adaptive Procedure

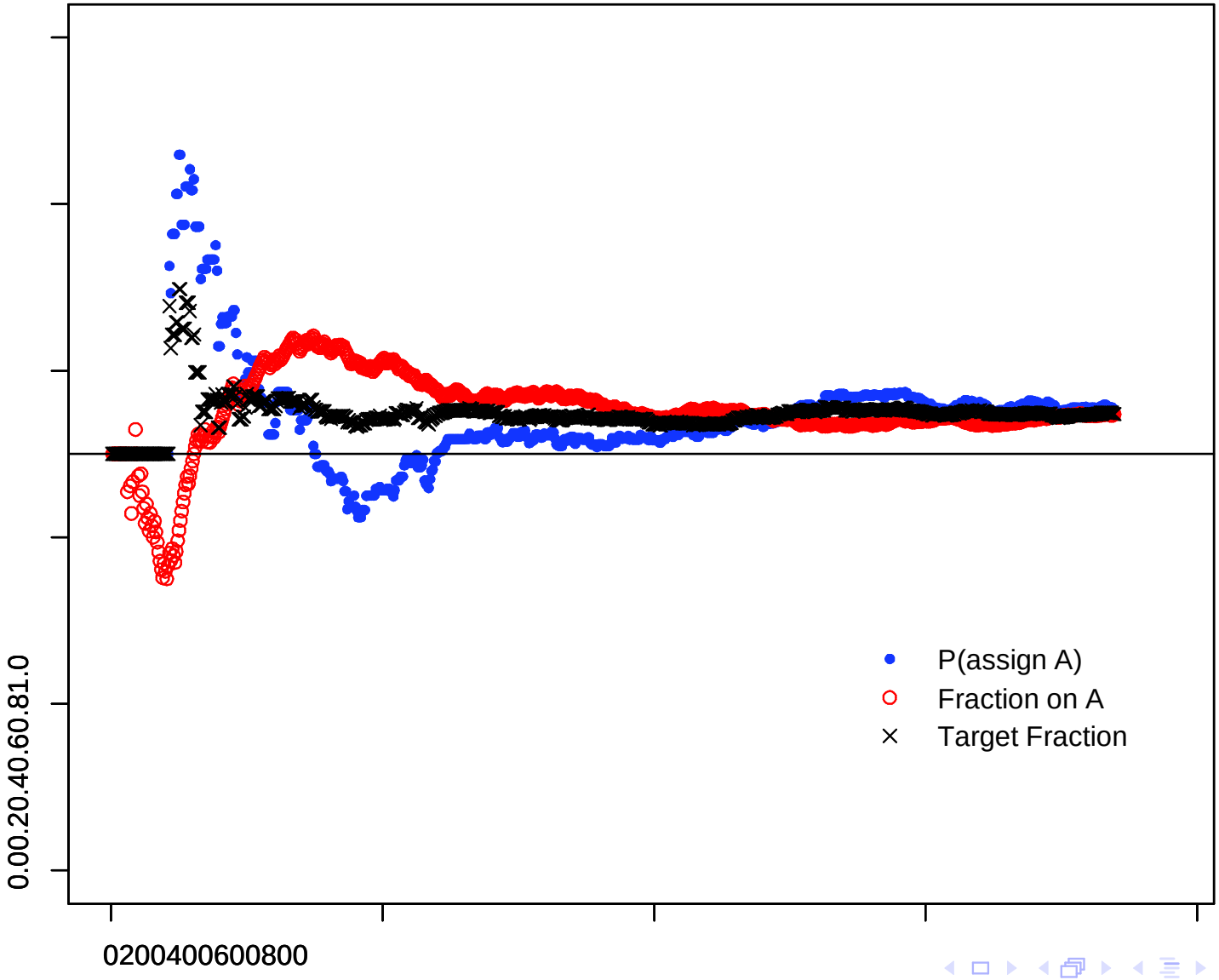
- Target $y = \frac{N_A}{N} = \frac{q_B}{q_A + q_B}$ (minimizes number of treatment failures)
- Assign m_0 patients to both treatments (assume immediate response)
- Calculate $\hat{p}_{m,t}, t = A, B$
- Assign patient $m+1$ using $f_{A_{m+1}}$, $m \geq 2m_0$, $\lambda = 1$

$$f_{A_{m+1}} = \hat{p}_{m,B}^2 \left(\frac{N_m}{N_{m,A}} \right) \left(\hat{p}_{m,B}^2 \left(\frac{N_m}{N_{m,A}} \right) + \hat{p}_{m,A}^2 \left(\frac{N_m}{N_{m,B}} \right) \right)^{-1}$$

Sample Size

- Total sample = 1520
- Assume prevalence of marker = 0.5.
- Non-adaptive balanced design
  - nå380 per treatment arm per marker stratum
- Adaptive design
  - nå760 in each marker stratum

Assignment History of One Simulated Trial



Simulation

- 1000 trials simulated
- Power (treatment) = fraction of H:$p_A=p_B$ rejected in marker + stratum
- Type 1 error = fraction of H:$p_A=p_B$ rejected in marker ï stratum.
- Power (interaction) = fraction of H:$p_{A(+)}=p_{A(ï)}$ rejected

Simulation

Design	Power (Treatment)	Type I Error	Power (Interaction)
Marker by Treatment	0.809	0.054	0.823
Adaptive	0.802	0.059	0.809

Simulation

-

	Marker (+)		
	mean % on A	mean # on A	(5^{th} , 95^{th} %-tiles)
Non-adaptive	0.5	380	-
Adaptive	0.542	412	(387 , 437)

-

	Marker (-)		
	mean % on A	mean # on A	(5^{th} , 95^{th} %-tiles)
Non-adaptive	0.5	380	-
Adaptive	0.5	380	(356, 404)

- On average 32 more subjects receive better treatment.
- $\sim$ 15 subjects expected to experience better PFS.

Expected Sensitivity and Specificity (*assuming balanced allocation*)

	Treatment A Success	Treatment A Failure	Totals
Marker (+)	171	209	380
Marker (-)	133	247	380
Totals	304	456	760

- Se= $\frac{171}{171+133} = 0.563$
- Sp= $\frac{247}{209+247} = 0.542$

Expected Sensitivity, Specificity

Treatment A

	Probabilities given Marker		Unconditional Probabilities		
Prev	*Marker*	*R–* *R+*	*Marker*	*R–*	*R+*
0.5	ï	*0.65* *0.35*	ï	*0.325*	*0.175*
0.5	+	*0.55* *0.45*	+	*0.275*	*0.225*

- *Specificity = 0.325 / (0.325 + 0.275) = .542*
- *Sensitivity= 0.225 / (0.175 + 0.225) = .563*

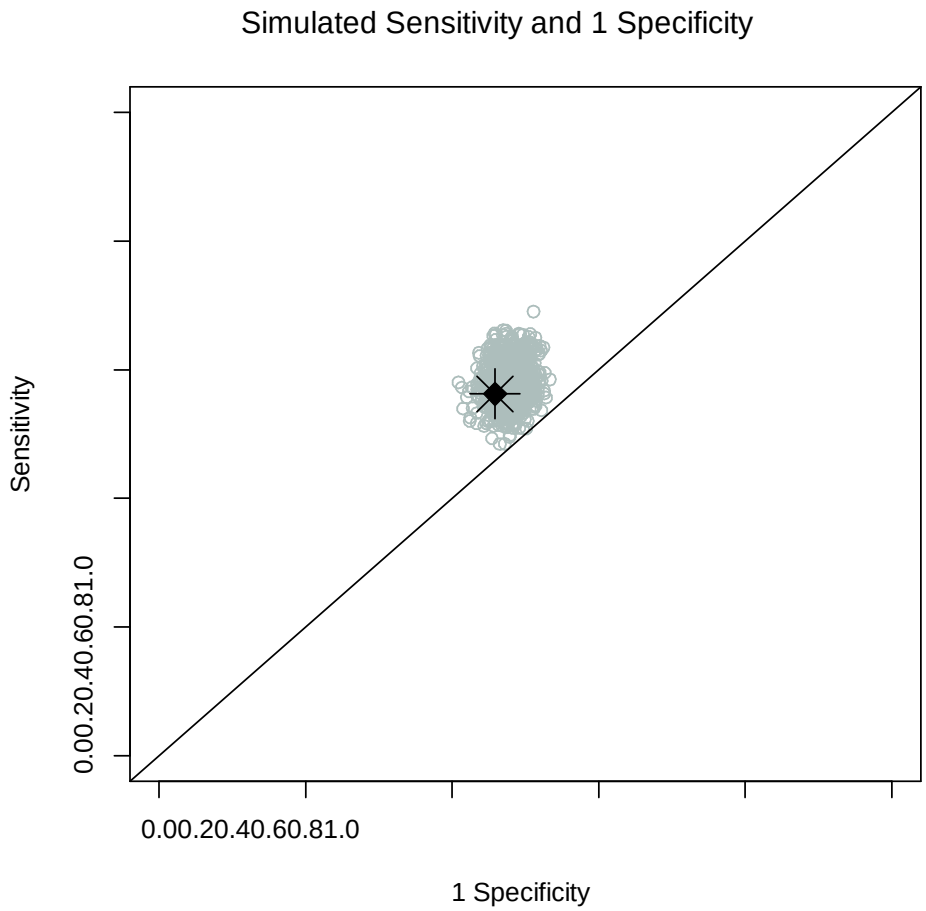
Adaptive Sensitivity, Specificity

Treatment A

	Probabilities given Marker		Unconditional Probabilities		
Prev	*Marker*	*R–* *R+*	*Marker*	*R–*	*R+*
0.458	ï	*0.65* *0.35*	ï	*0.298*	*0.160*
0.542	+	*0.55* *0.45*	+	*0.298*	*0.244*

- *Specificity = 0.298 / (0.298 + 0.298) = .500*
- *Sensitivity= 0.244 / (0.160 + 0.244) = .603*

Sensitivity and Specificity (Naive)



- Biased, star is expected (1-Sp, Se): (1-Sp= 1-0.542, Se= 0.563)

Verification Bias

- Favoring the better treatment stratified on marker results is analogous to verifying more marker positives than marker negatives
- Assume missing at random holds...Assignment based solely on marker status
  - marker (+) assigned to A with P_{A1}
  - marker (-) assigned to A with P_{A0}
- We can adjust for the bias using the allocation fractions
  - Inverse probability weighting (See Pepe, 2003)
  - For each simulation, multiplied table entries by reciprocal of realized allocation

Inverse Probability Weighting

Let $A =$ assignment status (to A)

$R =$ response status ($R-, R+$)

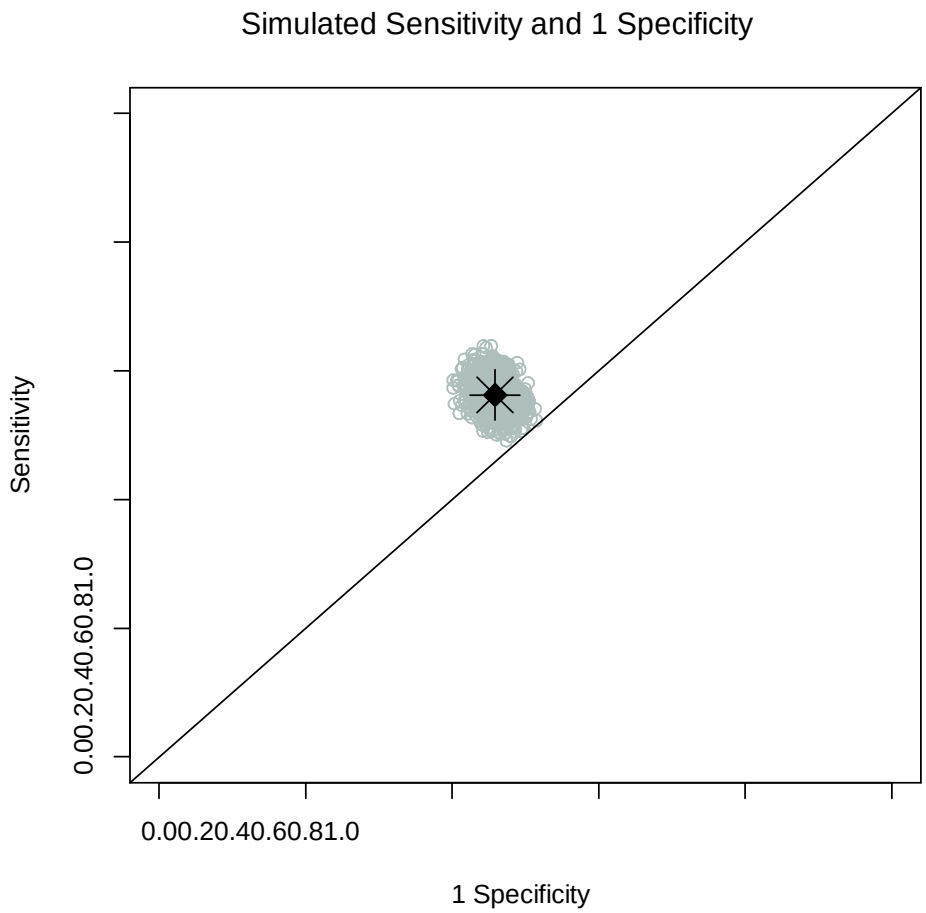
$T =$ marker status ($-, +$)

$$\Pr(A \mid R, T) = \Pr(A \mid T)$$

i.e., $$\Pr(R \mid T, A) = \Pr(R \mid T)$$

$$\begin{aligned} \Pr(T, R) &= \Pr(R \mid T)\Pr(T) \\ &= \Pr(R \mid T, A)\Pr(T) \\ &= \Pr(R, T, A) / \Pr(A \mid T) \end{aligned}$$

Sensitivity and Specificity (Adjusted)



- Confidence intervals for Se and Sp can be calculated using method of Begg and Greenes (1983)

Incorporating Prior Information in Bayesian Sequential Monitoring

With Laura Thompson

FDA Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

- Adherence to likelihood principle offers flexibility:
  - Reason for stopping or adapting may not matter to the Bayesian inference.
- Adaptive design changes mentioned:
  - Adaptive sample size
  - Stop early for success, futility, or harm
  - Dropping unfavorable arm
  - Change to randomization scheme
  - Change hypothesis from non-inferiority to superiority

Models for Borrowing Prior Information from Historical Studies

- Hierarchical modeling
  - Assumes study-specific treatment effects are random with a common distribution.
  - A prior may be placed on the distribution (e.g., a Dirichlet process prior)
- Power prior (Ibrahim, Chen, 2000)
- Commensurate prior (Hobbs, Carlin, Mandrekar, Sargent, *Biometrics,* 2010)
- Uniform Shrinkage Prior (Strawderman, 1971)
- Fiducial for variance components (T,P)

Bayesian Interim Analysis with Prior Information

EX. Consider new study of a coronary stent and 3 prior studies $i = 1, 2, 3$ of similar stents.

For study $i = 1, 2, 3, new$, patient $j = 1, 2, \ldots n_i$, assume percent diameter stenosis (%DS)

$$y_{ij} \sim N(\mu_i, \sigma_i^2)$$

Null Hypothesis: $H : \mu_{new} > 17\%$

Bayesian Interim Analysis, Coronary Stent %DS

Prior Study Results (3 Studies):

Study	N	Mean	SD	Z	Decision
1	200	10.0	25.0	3.96	Approval
2	200	13.0	19.0	2.98	Approval
3	200	15.0	14.0	2.02	Approval
Total	600	12.7	19.8		

Bayesian Interim Analysis, Coronary Stent %DS

Prior Study Results (3 Studies):

Study	N	Mean	SD	Z	Decision
1	200	10.0	25.0	3.96	Approval
2	200	13.0	19.0	2.98	Approval
3	200	15.0	14.0	2.02	Approval
New	200	NA	NA	NA	NA

Bayesian Interim Analysis with Prior Information

Data: $\bar{y}_{i\cdot} \sim N(\mu_i, \sigma_i^2 / n_i)$

$$f_i s_i^2 / \sigma_i^2 \sim \chi^2(f_i) = \Gamma(f_i / 2, 1/2)$$

Prior: $\mu_i \sim N(\mu_0, \sigma_\mu^2),\ i = 1, 2, 3, new$

$$\mu_0 \sim N(0, 1000),\ \sigma_\mu^{-2} \sim \Gamma(.001, .001)$$

Because study means are exchangeable, μ_{new} borrows strength from 3 prior studies.

Bayesian Interim Analysis with Prior Information

Data: $\bar{y}_{i\cdot} \sim N(\mu_i, \sigma_i^2 / n_i)$

$$f_i s_i^2 / \sigma_i^2 \sim \chi^2(f_i) = \Gamma(f_i / 2, 1/2)$$

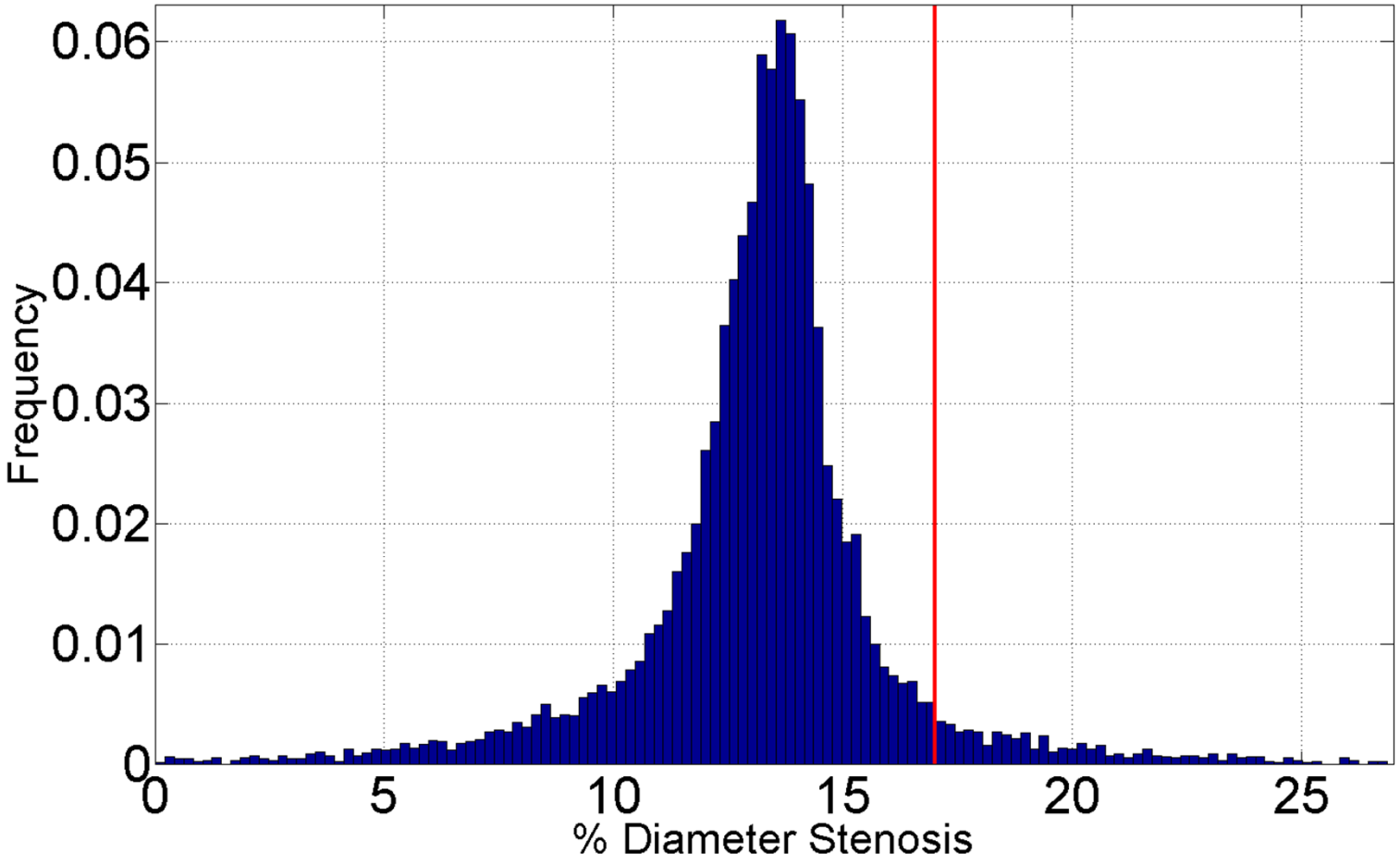
Prior: $\mu_i \sim N(\mu_0, \sigma_\mu^2),\ i = 1, 2, 3, new$

$$\mu_0 \sim N(0, 1000),\ \sigma_\mu^{-2} \sim \Gamma(.001, .001)$$

Model I: $\sigma_i^{-2} \sim \Gamma(.001, .001),\ i = 1, 2, 3$

$$\sigma_{new}^{-2} \sim \Gamma(f_* / 2,\ f_* s_*^2 / 2),\ s_*^2 = \max_{i=1.2,3} s_i^2$$

Prior Distribution for μ_{new}, Model 1



Bayesian Interim Analysis with Prior Information

Data: $\bar{y}_{i\cdot} \sim N(\mu_i, \sigma_i^2 / n_i)$

$$f_i s_i^2 / \sigma_i^2 \sim \chi^2(f_i) = \Gamma(f_i / 2, 1/2)$$

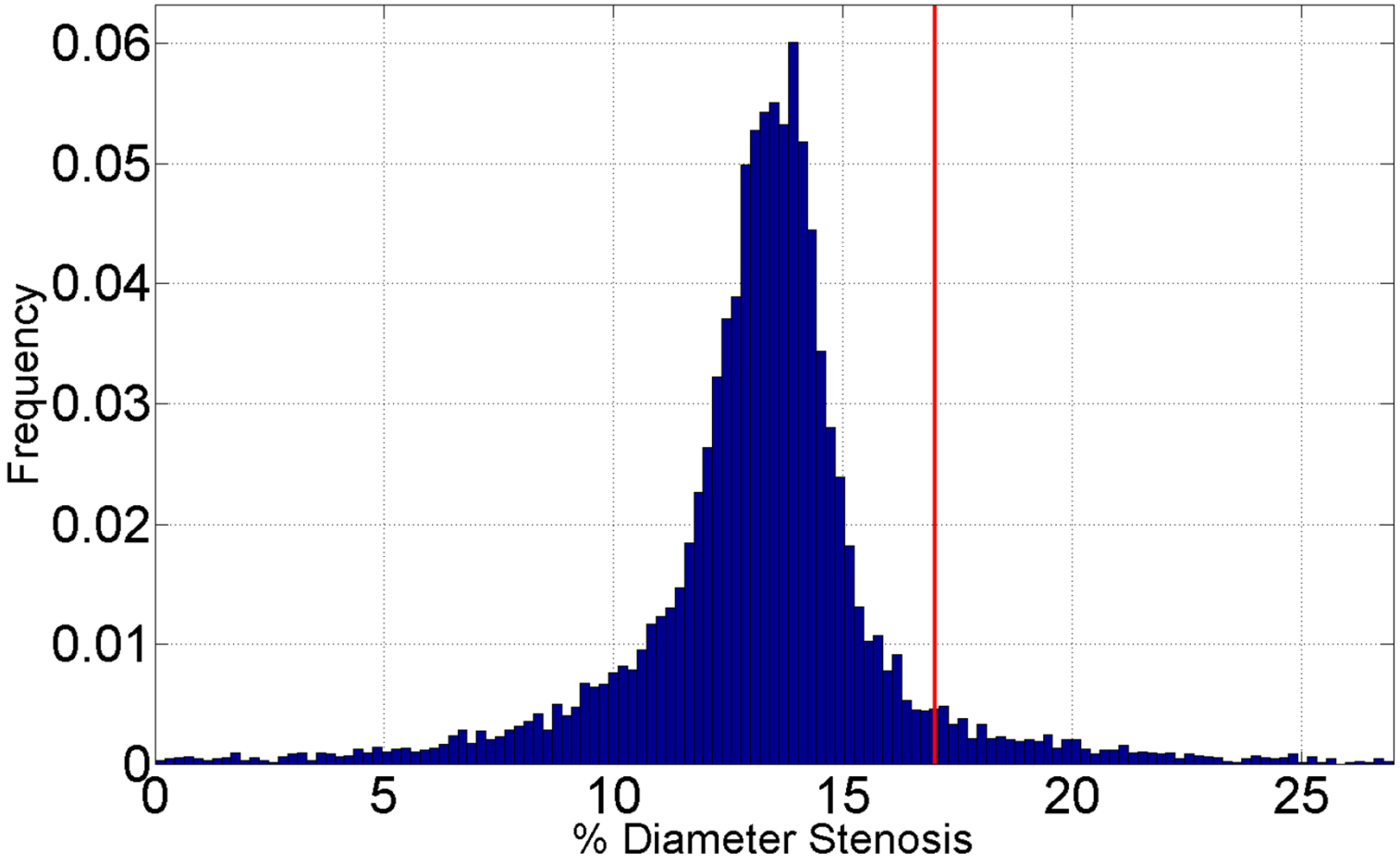
Prior: $\mu_i \sim N(\mu_0, \sigma_\mu^2),\ i = 1,2,3,4$

$$\mu_0 \sim N(0,1000),\ \sigma_\mu^{-2} \sim \Gamma(.001,.001)$$

Model II: $\ln \sigma_i^{-2} \sim \Gamma(\eta,\ \sigma_\eta^2),\ i = 1,2,3,4$

$$\eta \sim N(0,1000),\ \sigma_\eta^{-2} \sim \Gamma(.001,.001)$$

Prior Distribution for μ_{new}, Model 2



Interim Monitoring of *P(H|y)*, No Prior Information

Frequentist plan

Stages	n_1	n_2	n_3
Ŭ-level	.01	.02	.02

Bayesian plan, no prior information

Stages	n_1	n_2	n_3
P(*H*|*y*)	.01	.02	.02

NOTE: For normal data with non-informative prior, Bayesian = frequentist critical region .

Interim Monitoring of *P(H|y)*, With Prior Information

Proposal I:

Stages	n_1 ï n_0	n_2 ï n_0	n_3 ï n_0
P(*H*|*y*)	.01	.02	.02

n_0 = prior *effective* sample size (PESS)

NOTE: We simply follow the same interim plan as before, but start with n_0 observations before new study begins.

Bayesian Interim Analysis, Coronary Stent %DS

Prior Distribution, Model I:

		Prior			
Study	N	Mean	SD	2.5%, 97.5%	Pr<17%
1	200	12.1	1.78	8.1, 14.9	1.000
2	200	13.2	1.08	10.9, 15.2	1.000
3	200	14.2	1.00	12.4, 16.3	0.995
4	200	13.2	4.59	5.5, 20.2	0.941

$$\overline{PESS_i^1} = 16.9 \quad \overline{PESS_i^2} = 18.7 \quad PESS^3 = 18.7$$

Bayesian Interim Analysis, Coronary Stent %DS

Prior Distribution, Model II:

		Prior			
Study	N	Mean	SD	2.5%, 97.5%	Pr<17%
1	200	12.1	1.78	8.1, 14.7	1.000
2	200	13.2	1.09	10.9, 15.2	1.000
3	200	14.2	1.03	12.3, 16.3	0.994
4	200	13.2	5.96	4.9, 20.6	0.937

$$\overline{PESS_i^1} = 10.1 \quad \overline{PESS_i^2} = 11.1 \quad PESS^3 = 11.1$$

Interim Monitoring of *P(H|y)*, With Prior Information

Proposal I:

Stages	150 ï n_0	200 ï n_0	250 ï n_0
P(*H*|*y*)	.01	.02	.02

n_0 = PESS = 18 or 11 (Model 1,2)

Interim Monitoring of P(H|y), With Prior Information

Average Sample Size, by true mean μ_{new}

Mo-del	Mon-itor	Average n				
		13%	14%	15%	16%	17%
0†	n	179	201	224	238	248
1	$n-n_0$	147	166	183	210	223
2	$n-n_0$	151	160	188	214	231

†model with non-informative prior

Interim Monitoring of P(H|y), With Prior Information

Power, by True Mean %DS μ_{new}

Mo-del	Mon-itor	Power				
		13%	14%	15%	16%	17%
0†	n	0.92	0.78	0.41	0.21	0.06
1	$n–n_0$	0.97	0.87	0.66	0.37	0.15
2	$n–n_0$	0.98	0.88	0.64	0.36	0.14

†model with non-informative prior

In models 1-2, Type 1 error rate is inflated, but I contend it is matter of interpretation.

Interim Monitoring of *P(H|y)*, With Prior Information

Proposal II:

Stages	ESS=n_1	ESS=n_2	ESS=n_3
P(*H*|*y*)	.01	.02	.02

Take interim look when *effective* sample size (ESS) of the new study reaches n_i.

NOTE: Requires frequent monitoring (but you need not break the blind).

Interim Monitoring of *P(H|y)*, With Prior Information

Proposal II:

Stages	ESS=150	ESS=200	ESS=250
P(*H*|*y*)	.01	.02	.02

Interim Monitoring of P(H|y), With Prior Information

Average Sample Size, by true mean μ_{new}

Mo-del	Mon-itor	Average n				
		13%	14%	15%	16%	17%
0†	n	179	201	224	238	248
1	ESS	110	121	140	176	218
2	ESS	106	117	124	164	180

†model with non-informative prior

Interim Monitoring of P(H|y), With Prior Information

Power, by True Mean %DS μ_{new}

Mo-del	Mon-itor	Power				
		13%	14%	15%	16%	17%
0†	n	0.92	0.78	0.41	0.21	0.06
1	ESS	0.94	0.78	0.60	0.40	0.11
2	ESS	0.90	0.70	0.58	0.42	0.29

†model with non-informative prior

Interim Monitoring of P(H|y), With Prior Information

Average Sample Size, by true mean μ_{new}

Mo-del	Mon-itor	Average n				
		13%	14%	15%	16%	17%
0†	n	179	201	224	238	248
1	PESS	147	166	183	210	223
2	PESS	151	160	188	214	231
1	ESS	110	121	140	176	218
2	ESS	106	117	124	164	180

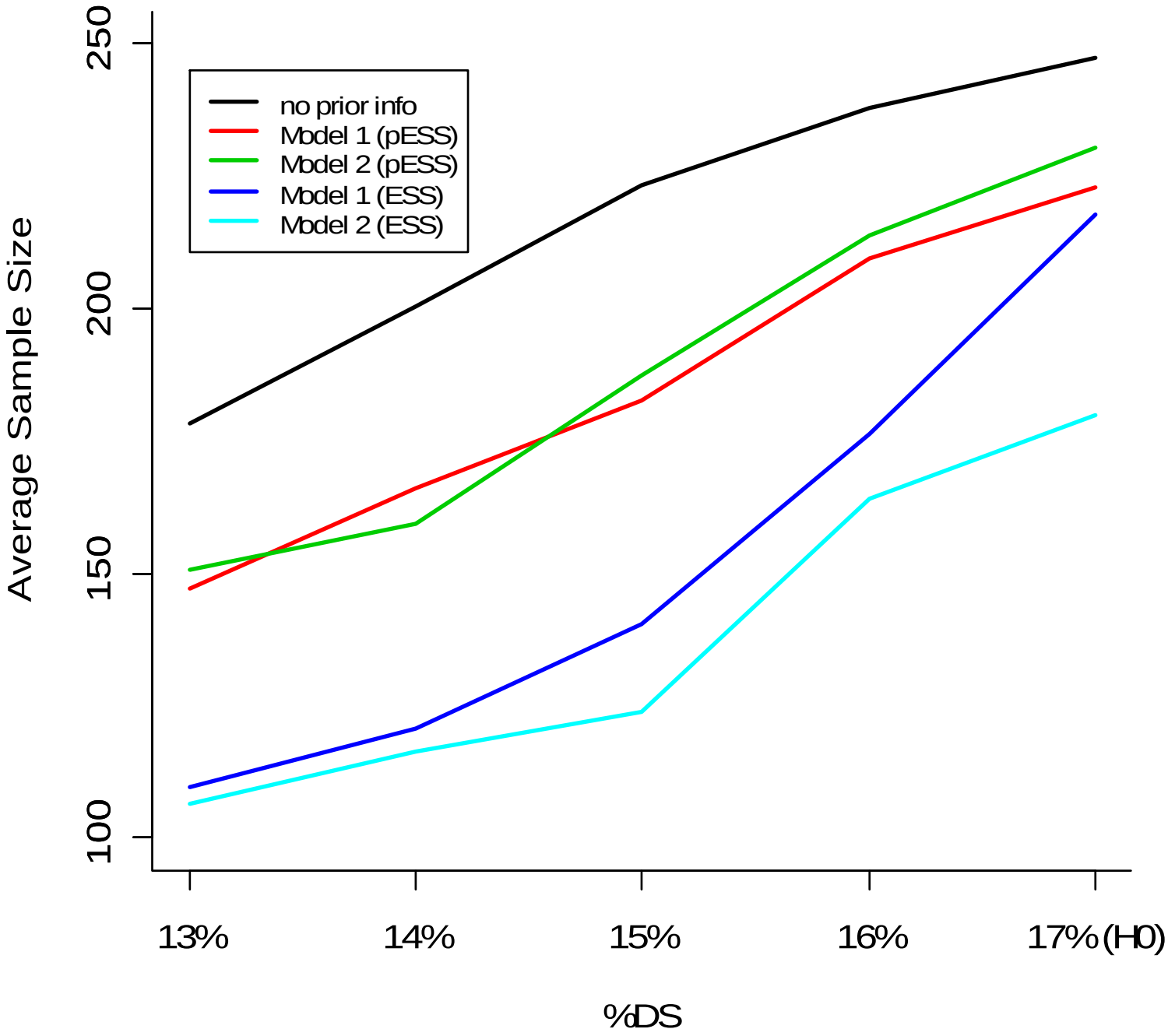
†model with non-informative prior

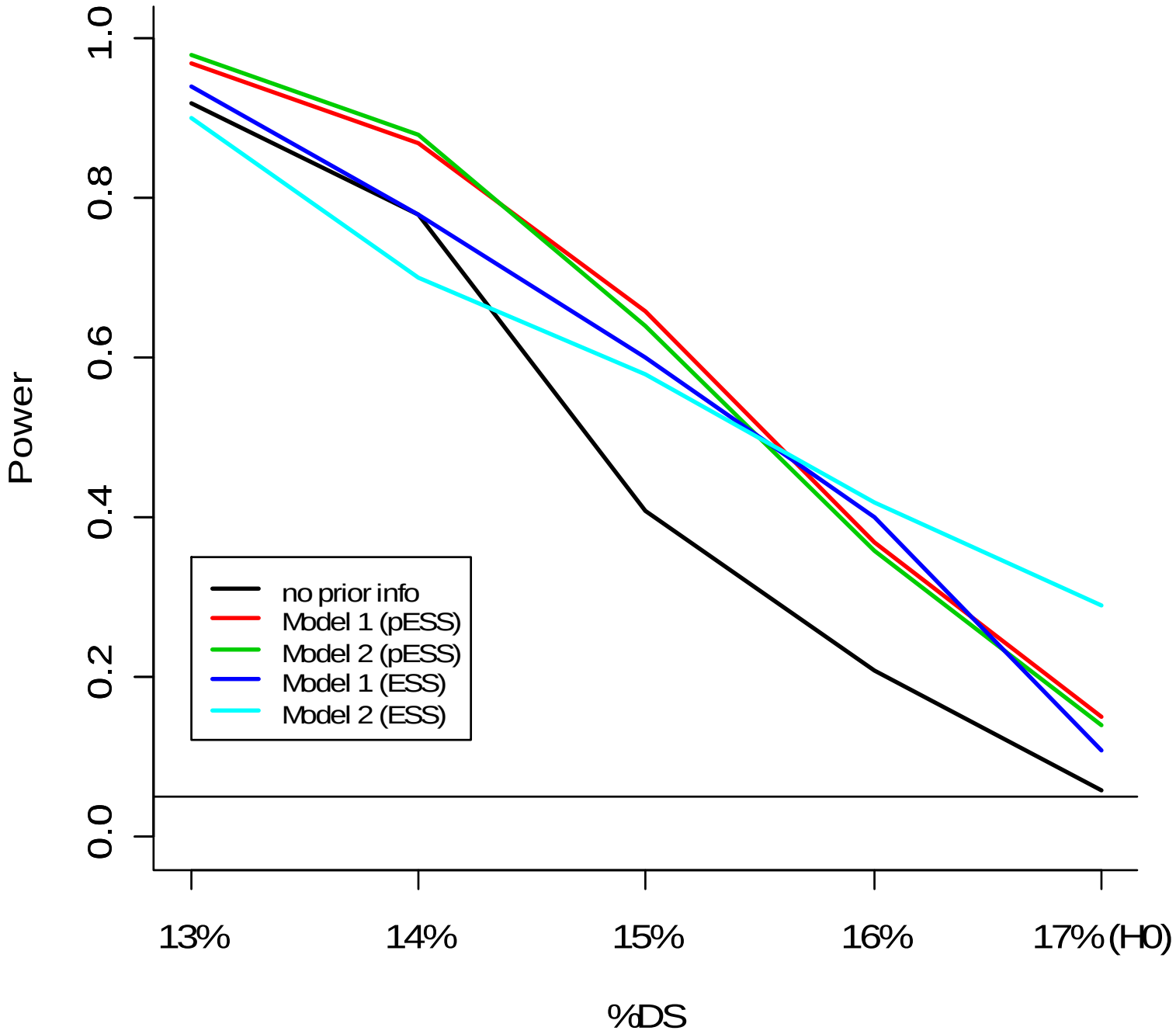
Interim Monitoring of P(H|y), With Prior Information

Power, by True Mean %DS μ_{new}

Mo-del	Mon-itor	Power				
		13%	14%	15%	16%	17%
0†	n	0.92	0.78	0.41	0.21	0.06
1	PESS	0.97	0.87	0.66	0.37	0.15
2	PESS	0.98	0.88	0.64	0.36	0.14
1	ESS	0.94	0.78	0.60	0.40	0.11
2	ESS	0.90	0.70	0.58	0.42	0.29

†model with non-informative prior


Average Sample Size
100
150
200
250
no prior info
Model 1 (pESS)
Model 2 (pESS)
Model 1 (ESS)
Model 2 (ESS)
13%
14%
15%
16%
17% (H0)
%DS


Power
1.0
0.8
0.6
0.4
0.2
0.0
no prior info
Model 1 (pESS)
Model 2 (pESS)
Model 1 (ESS)
Model 2 (ESS)
13%
14%
15%
16%
17% (H0)
%DS

Effective Sample Size (ESS)

Because variance is proportional to sample size (roughly), define

$$ESS(\theta) = \frac{n_{new} Var(\theta \mid data,\ no\ borrowing)}{Var(\theta \mid data,\ borrowing)}$$

Malec, D. (2001). A closer look at combining data among a small number of binomial experiments. Stat. Med. 20:1811ï 1824.

Prior Effective Sample Size (PESS)

Some possible ways to define:

$$PESS_i^1(\mu_{new}) = \frac{n_i Var(\mu_i \mid borrowing)}{Var(\mu_{new} \mid borrowing)}, i = 1,2,3$$

$$PESS_i^2(\mu_{new}) = \frac{E(\sigma_i^2 \mid borrowing)}{Var(\mu_{new} \mid borrowing)}, \; i = 1,2,3$$

$$PESS^3(\mu_{new}) = \frac{\hat{\sigma}^2_{pooled}}{Var(\mu_{new} \mid borrowing)}$$

Monitoring by Bayes Risk

with Professor David Duncan

FDA Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

- "Other potential uses [of Bayesian statistics] include é optimal decision making (Bayesian decision theory)ò
- "A decision analysis method might in principle be used to develop an interim analysis plan. Carlin, Kadane, & Gelfand (1998) propose a method to approximate a decision analysis approach in interim analyses.ò

CDRH Task Force on Utilization of Science in Regulatory Decision Making

- The Task Force was convened in 2009 to
  - review how CDRH uses science in its regulatory decision making process
  - make recommendations on how the Center can quickly incorporate new science – including evolving information, novel technologies, and new scientific methods – into its decision making, while also maintaining as much predictability as possible.
- Preliminary Internal Evaluations, August 2010
- "The Task Force recommends that CDRH "support evidence synthesis and quantitative decision making as a long-term goal."

http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDRH/CDRHReports/UCM220783.pdf

Interim Monitoring by Bayes Risk

Proposal III:

Stages	n_1	n_2	N
d_n	$a_{n1} < b_{n1}$	$a_{n2} < b_{n2}$	c^*

At interim stage (sample size n), choose $a_n < b_n$ to maintain Bayes risk at same level as at end stage (sample size N).

if $d_n > b_n$, stop, reject *H,*
if $d_n < a_n$, stop, do not reject *H,*
if $a_n < d_n < b_n$, continue at no cost

Bayesian Decision Rule: SvS

Test $H:\delta=\delta_1 \qquad A:\delta=\delta_2$

$$\delta=\mu_{Trt}-\mu_{Ctl},\ \delta_2>\delta_1$$

Sample Data (*n*): $d_n \mid \delta \sim f_n(d \mid \delta)$

Prior: $\pi(\delta_i)=\pi_i,\ i=1,2$

Decision Rule at *n*: $Reject\ H\ if\ d_n > c$

Bayesian Decision Rule: SvS

Test $H : \delta = \delta_1 \qquad A : \delta = \delta_2$

Decision Loss:

$$L_n(\delta, c) = kI(d_n > c)I(H) + I(d_n \leq c)I(A),$$

k = loss of Type I error relative to Type II error

Bayesian Decision Rule: SvS

Test $H : \delta = \delta_1 \qquad A : \delta = \delta_2$

Bayes Risk: $B_n(c) = k\pi_1\alpha_n(c) + \pi_2\beta_n(c),$

$$\alpha_n(c) = \int_c^{\infty} f_n(d \mid \delta_1)\partial d \; = \text{Type I error rate,}$$

$$\beta_n(c) = \int_{-\infty}^{c} f_n(d \mid \delta_2)\partial d = \text{Type II error rate}$$

Bayesian Decision Rule: SvS

Test $H : \delta = \delta_1 \qquad A : \delta = \delta_2$

Bayes Rule: $Reject\ H\ if\ d_n > c^*,$

$$B_n(c^*) = B_n^* = \min_c B_n(c)$$

Generally, $B_n(c^*) \downarrow$ with $\uparrow$ n.

because cost of sampling is 0.

Bayesian Decision Rule: SvS

Test $H : \delta = \delta_1 \qquad A : \delta = \delta_2$

c satisfies likelihood ratio* $$\frac{f_n(c^* \mid \delta_2)}{f_n(c^* \mid \delta_1)} = k\pi,\ \pi = \frac{\pi_1}{\pi_2},$$

posterior odds $$\frac{\pi_n(\delta_2 \mid c^*)}{\pi_n(\delta_1 \mid c^*)} = k,$$

or posterior probability $$\pi_n(\delta_2 \mid c^*) = \frac{k}{k+1}.$$

Bayesian Decision Rule: SvS

1. Decide upon maximum sample size N to limit minimum Bayes risk to B_N^*.
2. At interim stage n, introduce two critical values a_n and b_n for d_n, $a_n < b_n$, such that

   if $d_n > b_n$, stop, reject H,

   if $d_n < a_n$, stop, do not reject H,

   if $a_n < d_n < b_n$, continue at no cost.

Bayesian Decision Rule: SvS

With one critical value *c*,

$$B_n(c) = k\pi_1\alpha_n(c) + \pi_2\beta_n(c),$$

With two critical values a_n, b_n

$$B_n(a_n, b_n) = k\pi_1\alpha_n(b_n) + \pi_2\beta_n(a_n)$$

Principle: Maintain Bayes risk to that at end stage, i.e., set $B_n(a_n, b_n) = B_N^*$.

Rationale: At interim stage $n \le N$, should be willing to incur same risk as at end stage N.

Choosing *a* and *b*

Minimize probability of continuing

$$(1)\quad \min_{a_n, b_n} \Pr(a_n < d_n < b_n) = \int_{a_n}^{b_n} m_n(d)\partial d,$$

$$m_n(d) = \text{ marginal distribution of } d_n$$

subject to

$$(2)\qquad B_n(a_n, b_n) = B_N^*$$

II. Minimize Pr(continuing)

$$\frac{\partial}{\partial b}\left[\int_a^b m_n(d)\partial d\right] = m_n(b) - m_n(a)\frac{\partial a}{\partial b} \overset{\text{set}}{=} 0$$

$$\frac{\partial}{\partial b} B_n(a,b) = -k\pi_1 f_n(b \mid \delta_1) + \pi_2 f_n(a \mid \delta_2)\frac{\partial a}{\partial b}$$

$$= \frac{\partial}{\partial b} B_N^* = 0$$

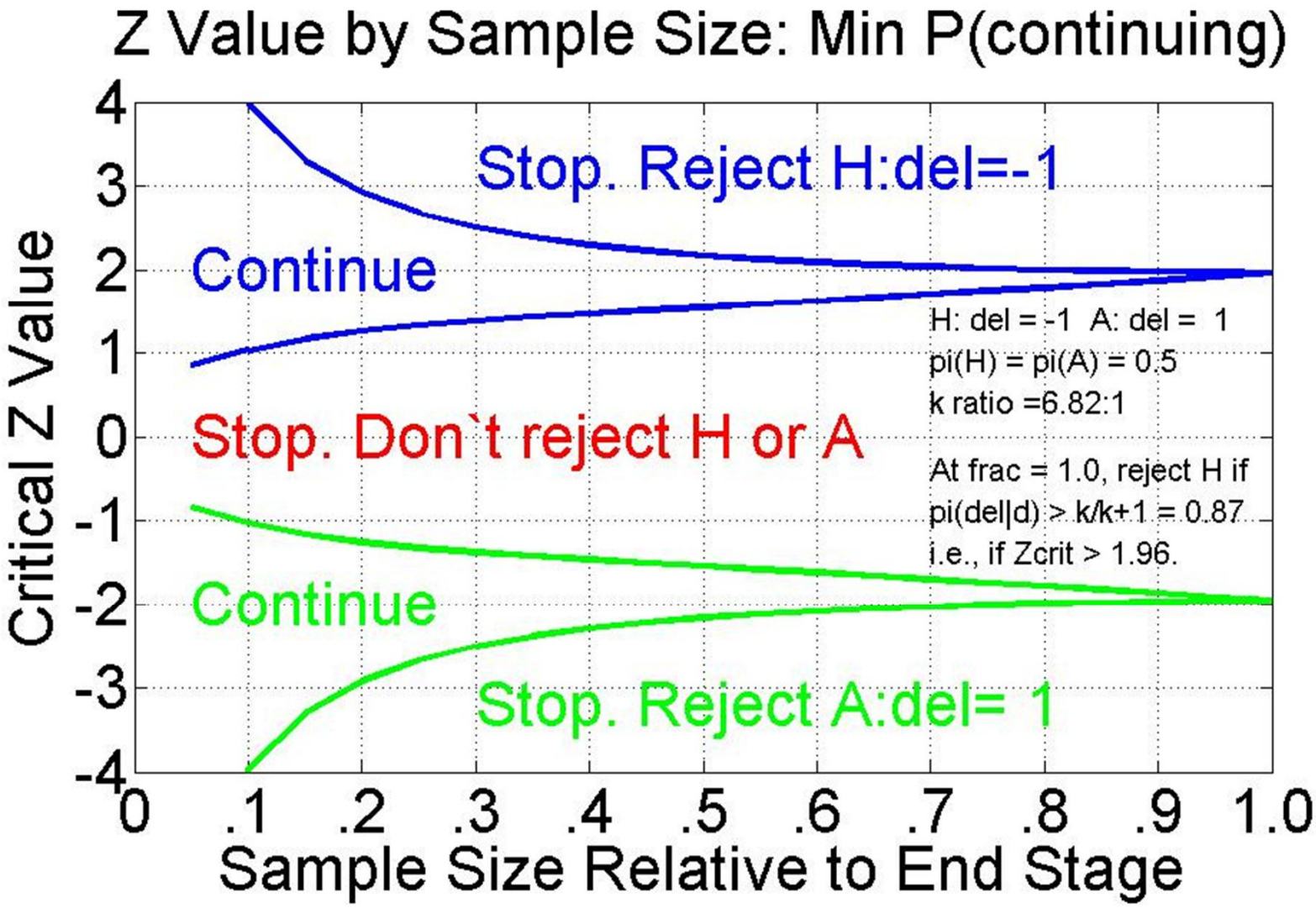
II. Minimize Pr(continuing)

I.e., a_n and b_n satisfy

$$\frac{\pi_n(\delta_2 \mid a_n)}{\pi_n(\delta_1 \mid b_n)} = k \left(= \frac{\pi_N(\delta_2 \mid c^*)}{\pi_N(\delta_1 \mid c^*)} \right)$$

and

$$B_n(a_n, b_n) = B_N^*$$


Z Value by Sample Size: Min P(continuing)
Critical Z Value
4
3
2
1
0
-1
-2
-3
-4
Stop. Reject H:del=-1
Continue
Stop. Don`t reject H or A
Continue
Stop. Reject A:del= 1
H: del = -1 A: del = 1
pi(H) = pi(A) = 0.5
k ratio =6.82:1
At frac = 1.0, reject H if
pi(del|d) > k/k+1 = 0.87
i.e., if Zcrit > 1.96.
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
Sample Size Relative to End Stage

Error Rate Control

Robbins (1970):

$$\Pr\left(\frac{f_n(x|\delta_2)}{f_n(x|\delta_1)} > k, \text{ for some } n = 1, 2, \ldots\right) \leq \frac{1}{k}$$

Note $$k = \frac{\pi_n(\delta_2 \mid a_n)}{\pi_n(\delta_1 \mid b_n)} < \frac{\pi_n(\delta_2 \mid b_n)}{\pi_n(\delta_1 \mid b_n)} = \frac{f_n(b_n \mid \delta_2)\pi_2}{f_n(b_n \mid \delta_1)\pi_1}$$

Thus $\Pr(\text{Type I error}) < 1/k\pi$

Similarly $\Pr(\text{Type II error}) < \pi/k$

Since can't make both,

$$\Pr(\text{Type I or II error}) < \max(1/k\pi, \pi/k)$$

Bayesian Decision Rule: lvl

Test $H : \delta \leq 0 \quad A : \delta > 0$

$$\delta = \mu_{Trt} - \mu_{Ctl}$$

Sample Data (*n*): $d_n \mid \delta \sim f_n(d \mid \delta)$

Prior: $\pi(\delta)$

Decision Rule at *n*: $Reject\ H\ if\ d_n > c$

Bayesian Decision Rule: lvl

Test $H : \delta \leq 0 \quad A : \delta > 0$

Linear Decision Loss:

$$L_n(\delta, c) = k\left|\delta\right| I(d_n > c) I(H) + \delta I(d_n \leq c) I(A),$$

k = loss of Type I error relative to Type II error

Bayesian Decision Rule: SvS

Test $H:\delta \le \delta_0 \qquad A:\delta > \delta_0$

At *N*, *c*=*c satisfies *weighted posterior odds***

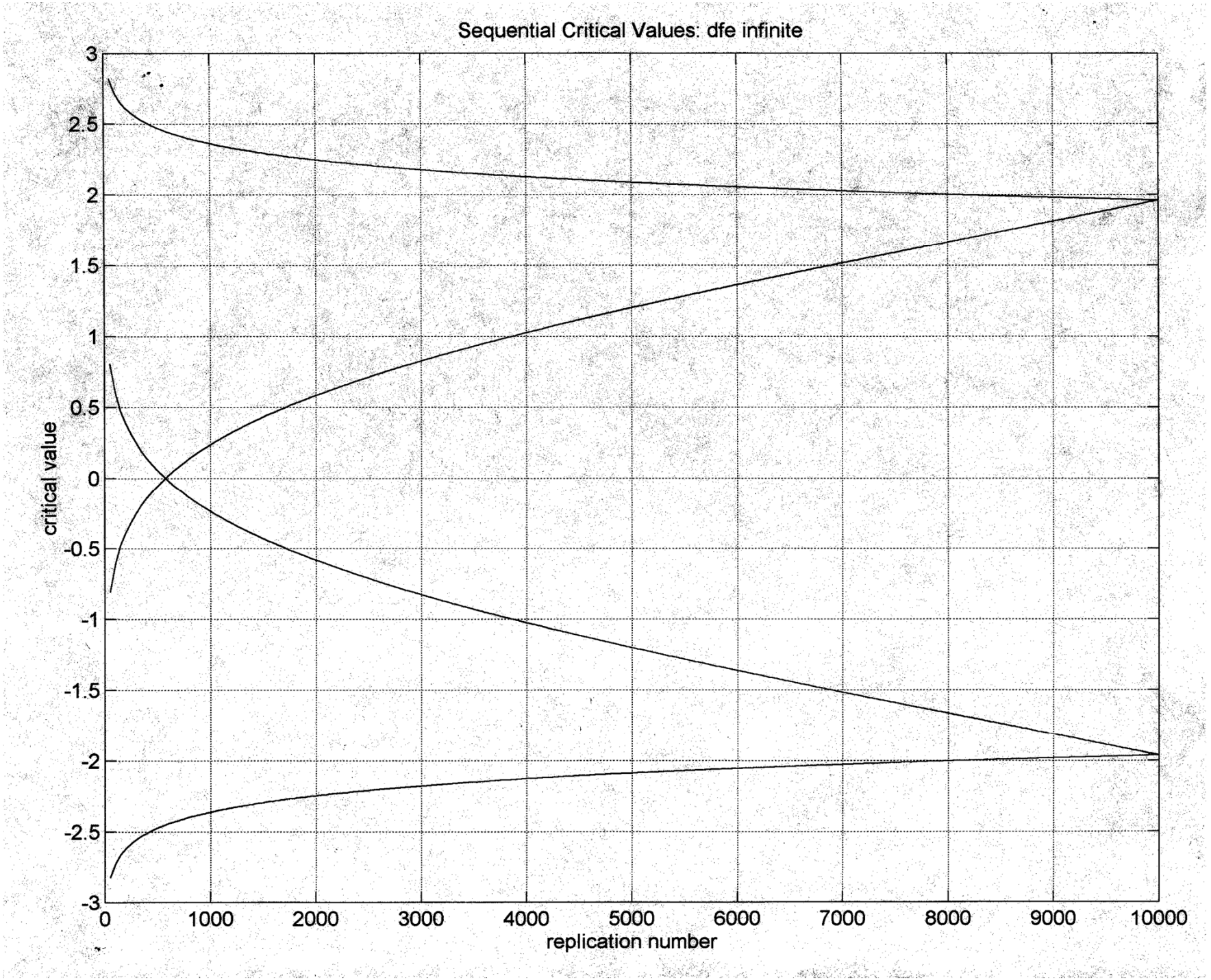
$$\frac{\int_0^\infty \delta\pi_N(\delta|c)\partial\delta}{\int_{-\infty}^0 |\delta|\pi_N(\delta|c)\partial\delta} = k$$

Likewise, at *n* < *N*, a_n and b_n satisfy

$$\frac{\int_0^\infty \delta\pi_n(\delta|a_n)\partial\delta}{\int_{-\infty}^0 |\delta|\pi_n(\delta|b_n)\partial\delta} = k$$

subject to

$$B_n(a_n,b_n) = B_N^*$$


Sequential Critical Values: dfe infinite
critical value
replication number
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

Challenges with Adaptive Design

- Implementation requires more work
  - planning
  - conduct
- Information on safety may be compromised.
- Simulation of Type I error rate should exhaust all scenarios:
  - Type 1 error rate may not be monotonic in parameters.
  - Accrual rate may be uncertain.
  - Predicting delayed response from early measure-ments requires assumption on their correlation.

Challenges with Adaptive Design

- Estimation bias
  - A challenge in labeling medical products
  - E.g., in our adaptive randomization trial, naïve sensitivity and specificity were biased.
  - Rosenbaum, Rubin (1984) show coverage of Bayesian intervals is sensitive to correctness of prior when a data-dependent stopping rule is used.
- In Bayesian adaptive studies, how to incorporate prior information and utilities (losses) are open questions.

References

Anscombe FJ (1963). Sequential Medical Trials, *JASA*, 58, 365-383.

Bonangelino, P., Irony, T., Liang, S., Li, X., Mukhi, V., Ruan, S., Xu, Y., Yang, Y. & Wang, C.G. (2011). Bayesian approaches in medical device clinical trials: A discussion with examples in the regulatory setting. *Journal of Biopharmaceutical Statistics* **21** (to appear).

Carlin BP, Kadane JB, Gelfand AE (1998). Approaches for Optimal Sequential Decision Analysis in Clinical Trials, *Biometrics*, 54, 964-975.

Freidlin, Simon (2005). Adaptive signature design: an adaptive clinicaltrial design for generating and prospectively testing a gene expression signature for sensitive patients, *Clin Cancer Res* 11(21)

Hobbs, Carlin, Mandrekar, Sargent (2011). Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials. *Biometrics*. (online)

Hu, Zhang (2004). Asymptotic properties of doubly adaptive biased coin deisngs for multi treatment clinical trials. *Ann Stat*, 32, 268-301.

Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science*, 15, 46ï 60.

References

Jiang, Freidlin, Simon (2007). Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect, *JNCI*, 99: 1036ï 43

Lewis RJ, Berry DA. (1994). Group Sequential Clinical Trials: A Classical Evaluation of Bayesian Decision-Theoretic Designs, *JASA*, 89, 1528-1534

Malec, D. (2001). A closer look at combining data among a small number of binomial experiments. Stat. Med. 20:1811ï 1824.

Pennello, G. & Thompson, L. (2008). Experience with reviewing Bayesian medical device trials. *J Biopharm Stat*, **18**: 81ï 115.

Rosenbaum PR, Rubin DB (1984). Sensitivity of Bayes Inference with Data Dependent Stopping Rules., *Amer Statistician,* 38, 106-109.

Tang L, Emerson, Zhou (2008), Nonparametric and Semiparametric Group Sequential Methods for Comparing Accuracy of Diagnostic Tests, *Biometrics,* 64, 1137ï 1145

Tang L, Liu A (2010). Sample size recalculation in sequential diagnostic trials, *Biostatistics*, **11**, 151ï 163

Zhao, Dmitrienko, Tamura (2010). Design and Analysis Considerations in Clinical Trials With a Sensitive Subpopulation, *Stat Biopharm Res*, 2(1), 72-83.


Center for Devices and Radiological Health
CDRH