Adapting Traditional Adaptive Designs: New Methods and Procedures

Scott Evans, Ph.D.

Harvard University

New Jersey Chapter of the American Statistical Association

June 3, 2011

Special Thank You

- Dirk Moore
- Acknowledgements
  - LJ Wei
  - Steve Lagakos

Outline

- Introductory Remarks on Adaptive Designs
- Predicted Intervals and Predicted Interval Plots (PIPs)
  - Extensions
- Changing Endpoints
- DMCs/DSMBs

Practical Questions at Interim

- Should the trial or trial arms be stopped?
  - For efficacy?
  - For futility?
- Should sample size be re-calculated?
  - Due to a lack of precision in estimating a parameter during trial design (e.g., variability, control group response)
- Should the duration of follow-up be modified due to unexpected event rates?

Motivation

- Answering these questions has:
  - Ethical attractiveness
    - Fewer participants generally exposed to inefficacious and potentially harmful therapies
  - Economical advantages
    - Smaller expected sample sizes and shorter expected duration than designs without interim analyses
      - Saving time, money, and other resources
  - Public health advantages
    - Answers may get to the medical community more quickly

Example: ATN 082
Adaptation based on external data

- Evaluation of Pre-exposure prophylaxis (PREP)
- Participants randomized to PREP (drug intervention) or placebo to prevent HIV transmission
  - 1st participant enrolled in August 2008
  - 11/22/2010: email notifying results from Preexposure prophylaxis intiative (iPREX) trial (Gates Foundation)
    - PREP reduced HIV acquisition in randomized double-blind trial in similar population (published in NEJM on 11/23/2010)
  - 11/23/2010: DSMB call; recommendations made

Considerations and Recommendations

- Equipoise
- Ethical to randomize and follow?
- Recommendations
  - Notify participants and IRBs of iPREX results
  - Unblind participants
  - Discontinue randomization into control arms; enrollment into PREP arm can continue as scheduled
  - Control arms offered option to roll over onto PREP

More Motivation

"I've designed >1000 clinical trials, each time having to make assumptions about variation, control-group response rate, etc. in order to calculate sample size …

More Motivation

"I've designed >1000 clinical trials, each time having to make assumptions about variation, control-group response rate, etc. in order to calculate sample size …

I have not been right yet."

Example as DSMB Member

- Trial designed to detect difference between response rates of 90% (control) and 97.5%
  - 7.5% absolute difference
  - 486 patients required to have 90% power
- Observed rate of control at interim is 80%
  - With N=486, 56% power
  - To detect 7.5% absolute improvement (80% vs. 87.5%) requires 1066 patients

It is not the strongest of the species that survives, nor the most intelligent that survives.

It is the one that is most adaptable to change.

Charles Darwin

protocols

It is not the strongest of the species that survives, nor the most intelligent that survives.

It is the one that is most adaptable to change.

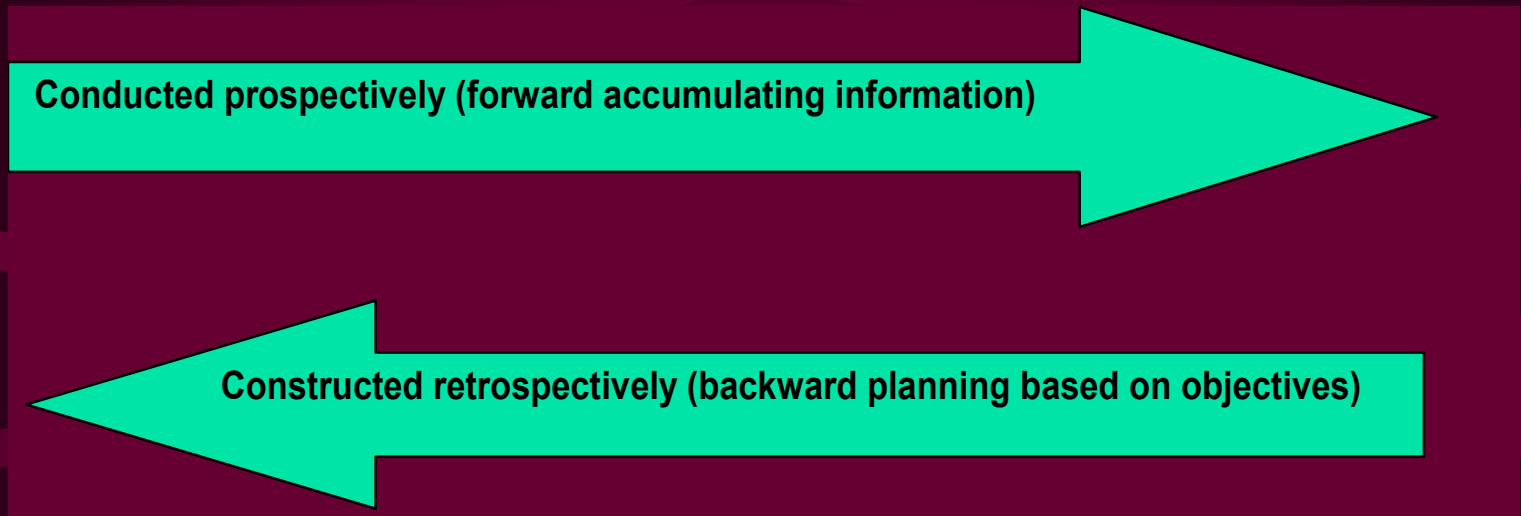
Charles Darwin

Scott Evans

Adaptive Designs

- Not a substitute for careful planning
  - Not a rescue medication (despite perception as a savior)
- Fancy adaptations and statistical methods cannot rescue poorly designed trials
- Although some adaptations are unexpected, whenever possible, it is…
- Adaptation by design
  - Adaptive designs have assumptions and limitations

Designing Clinical Trials



The key: “vision”

Major Scientific Concerns

- Statistical
  - Error control associated with multiplicity
    - Are hypotheses being tested?
    - Are adjustments based upon treatment comparisons?
- Operational bias
  - Adaptations are visible

Operational Bias

- Adaptation (or lack of adaptation) used to infer trial results
  - Affecting patient/investigator action during the remainder of the trial
    - E.g., participation, adherence, etc.
    - Objectivity threatened
- Not a statistical source of bias and thus difficult to adjust for
- Threat of leaking of results
- May cause heterogeneity of results (before vs. after adaptation)
- Significant issue for adaptive designs
  - Put details regarding the planned adaptation into a separate (limited distribution) document to reduce back-calculation for inferring effects

Addressing Concerns

- Statistical
  - Statistical methods exist
    - E.g., group sequential and modern adaptive design methods for controlling errors
- Operational bias
  - Careful and responsible application of adaptation
  - Well-constructed processes
    - Control of dissemination of adaptation
    - Interim analyses and DSMBs procedures

Complexity and Acceptability

- Some types of adaptations are well understood/accepted while others less so.

- Depends upon
  - Type of adaptation
  - The data utilized for decision-making
    - Has interim data (particularly endpoint data) been reviewed?
    - Blinded or unblinded
  - How adaptation is implemented
  - Who is reviewing data and making the recommendation/decision to adapt

Lower Threat to Trial Integrity

- Well-planned adaptations
- Adaptations prior to any data analyses
- Adaptations based on
  - Baseline data
  - External data
  - Blinded (aggregate) data
  - Nuisance parameters (e.g. variation)
  - Control group data only
  - Data unrelated to endpoints

Higher Threat to Trial Integrity

- Unplanned adaptations
- Adaptations based on observed treatment effects

Example: Adapt Sample Size based on Observed Treatment Effect?

- Issues
  - Statistical error control
    - Some newer methods exist
  - Operational bias
    - If sample size is recalculated based on observed treatment effect, quantitatively savvy researchers could back-calculate the treatment effect threatening trial integrity
  - Conceptual
    - Trials are designed to detect *relevant* effects
    - Observed effects may not be relevant

5/26/2011: Email from FDA Team Leader

- I find many sponsors re-estimate their sample based on the interim difference. I feel this is incorrect. Sample size should only be re-estimated based on mispecification of variability or control rate. The difference we are trying to detect should be based on clinical input. If we re-estimate based on observed difference, we may end up with a trial that shows a statistically significant difference but not a clinically meaningful difference. Furthermore, I think using this information to would affect the Type I error rate even if we adjust for the interim analysis. There seems to be disagreement in FDA as to whether you can re-estimate based on observed difference I may be in the minority. I would appreciate your thoughts.

2-Stage Design

Same Objectives and Endpoints

- Stage I: Evaluate preliminary evidence of effect/no effect

- ACTG 269 (Evans et.al., *JCO*, 2002)
  - Phase II single arm trial of oral etoposide for AIDS KS
  - Endpoint: tumor response rate (50% decrease in lesion number/size)
  - Stage I
    - Enroll small number of participants (N=14)
    - If response is unacceptably low (0/14), then quit for futility noting that if true response rate is 20% then <5% chance of observing 0/14
    - Otherwise continue to Stage II (not testing for efficacy)
    - Expected sample size is minimized when response is low given error constraints
  - Trial continued w/ final response rate = 36%

TheScientist News Current Issue Archive Surveys & Supplem

Volume 23 | Issue 8 | Page 55
Date: 2009-08-01
Reprints | Issue Contents

By Alla Katsnelson

Adaptive Evolution

A once-rare type of clinical trial that violates one of the sacred tenets of trial design is taking off, but is it worth the risk?

"If you're looking at trial data before it's completed, there's always the chance that you're jeopardizing the trial's integrity." —Scott Evans, Harvard University

"The reasons to do it are pretty clear," says Janet Wittes, president of the clinical trial design firm Statistics Collaborative, Inc., in Washington, DC.

Adaptive trials "hold the potential and the promise for doing trials faster and getting answers faster." Plus, assessing assumptions you made in planning a trial as it proceeds can strengthen a trial's scientific merit, says Scott Evans, a biostatistician at Harvard University.

According to Berry, a vocal champion of adaptive design, "the randomized trial moved us up to a very high level in terms of science," he says. "We're trying to preserve that level but move to even higher levels in terms of efficiency." Clinical trials, especially late-stage ones, can cost tens to hundreds of millions of dollars; done right, an adaptive design can shave 20–50% off that sum.

But like all propositions that sound too good to be true, there's a downside. Too often, critics caution, companies wrongly assume the approach is an easy fix to common clinical trial woes when, in fact, the changes can ultimately cost time and money, not save them. "There's been a lot of overselling, overmarketing, if you will," says Bob O'Neil, who heads adaptive design efforts for drugs and biologics at the US Food and Drug Administration (FDA). "This is not a panacea for all situations," he stresses. "It is not standard fare."

There's also a bigger concern, which has both US and European regulators struggling to delineate when the approach is appropriate and when it isn't. Trials are traditionally blinded in order to prevent investigators' or subjects' knowledge from influencing the outcome. "If you're looking at trial data before it's completed, there's always the chance that you're jeopardizing the trial's integrity," Evans says.

Adaptive design is a slippery term, and experts argue about its definition. Some adaptations are so straightforward that they raise no concerns whatsoever, Wittes explains. Stopping a trial early, for

Other adaptations, though, walk a finer line, and whether or not they're okayed depends on the reasons for implementing them as much as on the adaptations themselves. Suppose, for example, that interim results show that the variance in the data is larger than you'd like. "Variance is a nuisance parameter," says Wittes. "I don't think anyone has any trouble" with the idea of increasing the trial's size in this case. But say the treatment's effect appears to be smaller than expected, contradicting the prediction for efficacy that the company filed with the FDA before commencing the trial. In that case, by increasing the trial size, she believes, "you're really changing your hypothesis," which is based on a prediction of how well the drug works.

made up of people uninvolved in the study or the company. "It's a tough pill for companies to swallow, putting decisions about a trial into the hands of totally independent bodies," notes Bruce Turnbull, a biostatistician at Cornell University.

Regardless of the adaptation, unblinding the data for an interim peek invariably brings up two problems for regulators. The first one is statistical in nature. Because statistics measure the likelihood, based on probability, that a treatment is effective, repeating a statistical test multiple times increases the chances that one test along the way will mistakenly show an effect where there actually is none, and companies will submit that false-positive data to the FDA. "If you're always tweaking [the trial] to get the best result possible," says Turnbull, "then you will get the best result possible." There are statistical maneuvers to counteract that possibility, but they are far from straightforward, he explains.

The second concern is operational: When an adaptation to a trial takes place, will investigators and patients put two and two together to deduce clues about the therapy's efficacy? "Let's say we're going to increase sample size, or drop certain arms," Evans says. "Well, that adaptation could send a message to people involved in the trial that the effect isn't what you'd expect it to be. If that then changes people's actions—whether it be investigators or patients—then you've introduced a source of bias."

> "There are a lot of amateurs running around saying we should do this without understanding the statistics behind it." —Bruce Turnbull, Cornell University

By and large, adaptations in early-stage trials are not problematic. In fact, says O'Neil, "we think they've been underexplored" in that context. But regulators need much more convincing in Phase 3, or in Phase 2 trials that propose to morph directly into Phase 3 without the 6–9 months of analysis between the two steps. Proposals that include such seamless Phase 2/3 trials undergo intense regulatory scrutiny, which can offset whatever time you've gained from the adaptation.

And at any stage, the approach requires "a lot more prospective planning, a lot more complexity in design, and perhaps even more risk, in terms of will [the trial] turn out as you had planned," says O'Neil. Every possible eventuality in the trial that may result in an adaptation must be thought

When and Where?

- High levels of uncertainty/unknowns
  - Limited experience with treatments (e.g., novel interventions, new populations, etc.)
- When design characteristics (e.g., power) are very sensitive to assumptions
- Longer trials where adaptation is feasible
  - Larger studies, studies with slow recruitment, or long duration of FU
  - Accumulating medical information can influence the utility and ethics of ongoing trials of long duration
- Studies with invasive procedures, or expensive/tedious evaluations
- Studies of serious diseases, high risk treatments, vulnerable populations, and potential ethical dilemmas
- When data that serves as the basis for adaptation is available quickly

Motivating Question

How do we revise our traditional approaches (scientific and operational) to adaptive designs so to maximize trial efficiency and improve decision-making while maintaining trial integrity?

Interim Analyses Methods

- Group sequential methods
  - Control α spending
  - E.g., Slud and Wei, O'Brien-Fleming, Lan-DeMets, Pocock
  - Generally boundary driven with test statistics
- Conditional power/futility index

Limitations of Many Traditional Methods

- Do not
  - Provide estimates of effect or associated precision (only test statistics, p-values, and decision rules)
  - Evaluate "clinical relevance"
    - Statistical significance is not the only consideration
  - Information regarding the reasons for:
    - High p-values (or test statistics):
      - Negligible effect vs. insufficient data vs. too much variation
    - Low p-values (or test statistics):
      - Clinical significance?

Limitations of Many Traditional Methods

- Inflexible with binding decision rules based usually on a single (primary) endpoint
  - Desire to base decisions upon simultaneous assessment of many factors, such as:
    - Safety data
    - Secondary endpoints
    - Quality of life
    - Benefit:risk assessment
    - Results of other trials
    - Scientific relevance
    - Availability of new alternative therapies
    - Cost:benefit considerations

Repeated Confidence Intervals (RCIs)

- Sequential CIs
  - Simultaneous coverage control
    - Uses principles of group sequential methods
- Provides estimates of effects sizes
- Allows for flexibility in decision making

- Jennison & Turnbull, *Controlled Clinical Trials*, 1984.
- Mehta et.al., *Statistics in Medicine*, 2000.

Limitations of Repeated CIs

- At the interim, we wish to weigh the options of stopping vs. continuing
- Repeated CIs do not:
  - Provide formal evaluation of the ramifications of continuing
    - What effect size estimates and associated precision will be observed at the end of the trial? At the next interim?
- Thus how do we weigh the options?

Need for Methods that:

- Control error rates
- Are flexible to allow for expert DSMB judgment
  - Allow incorporation of other information into decision
- Provide effect size estimates and associated precision
  - Assess clinical relevance and statistical significance
- Provide information about decision alternatives

Predicted Intervals and Predicted Interval Plots (PIPs)

Predicted Intervals

- Predict CI at future timepoint (e.g., end of trial or next interim analysis time) conditional upon:

  1. Observed data
  2. Assumptions regarding future data (e.g., observed trend continues, H_A is true, H_0 is true, best/worst case scenarios, etc.)

- Evans SR, Li L, Wei LJ, *Drug Information Journal*, 41:733-742, 2007.

NARC 009

- Randomized, double-blind, placebo-controlled, multicenter, dose-ranging study of prosaptide (PRO) for the treatment of HIV-associated neuropathic pain
- Participants were randomized to 2, 4, 8, 16 mg/d PRO or placebo administered via subcutaneous injection
- Primary endpoint:
  - 6 week change from baseline in weekly average of random daily Gracely pain scale prompts using an electronic diary

NARC 009

- Designed for 390 participants equally allocated between groups
- Interim analysis conducted after 167 participants completed the 6-week double-blind treatment period
- Computed PIs

Interim Analysis Results: NARC 009

Treatment	N	95% CI for Mean Change	95% CI for Diff[1]	95% PI for Diff[2]	95% PI for Diff[3]	Required Diff[4]
Placebo	31	(-0.35, -0.11)				
2 mg	34	(-0.21, -0.04)	(-0.04, 0.25)	(-0.01, 0.21)	(-0.16, 0.06)	-0.54
4 mg	34	(-0.38, -0.12)	(-0.19, 0.16)	(-0.14, 0.10)	(-0.23, 0.01)	-0.45
8 mg	32	(-0.18, -0.02)	(-0.01, 0.28)	(0.03, 0.23)	(-0.15, 0.05)	-0.56
16 mg	36	(-0.34, -0.09)	(-0.16, 0.19)	(-0.11, 0.14)	(-0.21, 0.04)	-0.54

1: 95% CI for the difference in mean changes vs. placebo
2: 95% PI for the difference in mean changes vs. placebo assuming full enrollment, assuming current trend
3: 95% PI for the difference in mean changes vs. placebo assuming full enrollment, assuming per protocol, $\mu_{placebo}$ = -0.17 and μ_{drug} = -0.34
4: Difference in mean changes needed in the remaining participants for the CI for the difference in mean changes to exclude zero (in favor of active treatment) at the end of the trial

NARC 009

- Sensitivity analyses shows that the futility assessment is robust
- Trial was discontinued by NARC DSMB for futility
  - Evans et. al., *PLoS ONE*, 2007.

Predicted Intervals

- Intuitive
- Use with repeated confidence interval theory to control error rates
- Design and monitor trials
- Advantages
  - Flexible decision making
    - Considering all data (all endpoints, external data, etc.)
  - Effect sizes and associated precision
    - Clinical relevance and statistical significance
  - Evaluation of trial with continuation
  - Can be used for all types of endpoints (e.g., binary. Continuous, event-time) and hypotheses (e.g., superiority or noninferiority)

Issues

- Sensitivity analyses necessary to assess the robustness of results to varying assumptions
  - Strategic assessment of assumptions to be employed
- Need to assess impact of sampling variability
- Need concise and intuitive summaries for DSMBs
  - Too much information is not digestible
    - Graphics are helpful
  - Too little information is not informative
  - Remember most DSMB members are not statisticians

Predicted Interval Plots (PIPs)

- Evaluates the sampling variability associated with the assumed model using simulation
- Plots the simulated PIs under the model assumption
- Conditional power is readily available
- Li L, Evans SR, Uno H, Wei LJ (*Statistics in Biopharmaceutical Research*)

PIP Construction

- Impose parametric assumptions for the unobserved data
  - Estimate or specify the values of the unknown parameters under reasonable and strategic assumptions
- Simulate future data
- Combine the observed data with the simulated data
- Construct PIs using standard methods
- Repeat to obtain many simulated PIs

PIPs: Construction Schema

DSMB

Observed dataset at interim point

\+ Assumption about future data

Simulate future outcome

A **simulated complete** dataset at the end of the trial or later interim

Calculate the "final" result

Adjusted PI

PIPs: Construction Schema

DSMB

Observed dataset at interim point

\+ Assumption about future data

Simulate future outcome

A **simulated complete** dataset at the end of the trial or later interim

Calculate the "final" result

Repeat

Adjusted PI

PIPs: Construction Schema

DSMB

Observed dataset at interim point

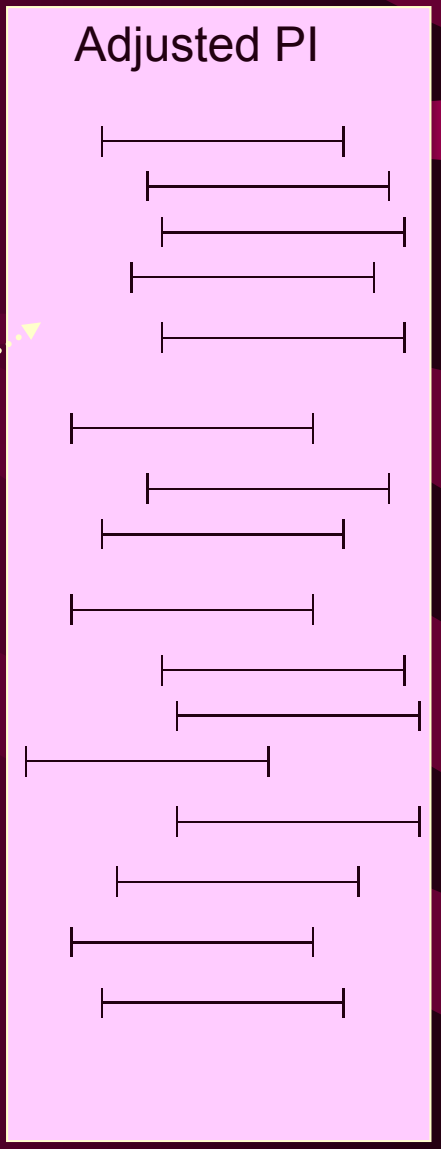
\+ Assumption about future data

Simulate future outcome

A **simulated complete** dataset at the end of the trial or later interim

Calculate the "final" result

Adjusted PI



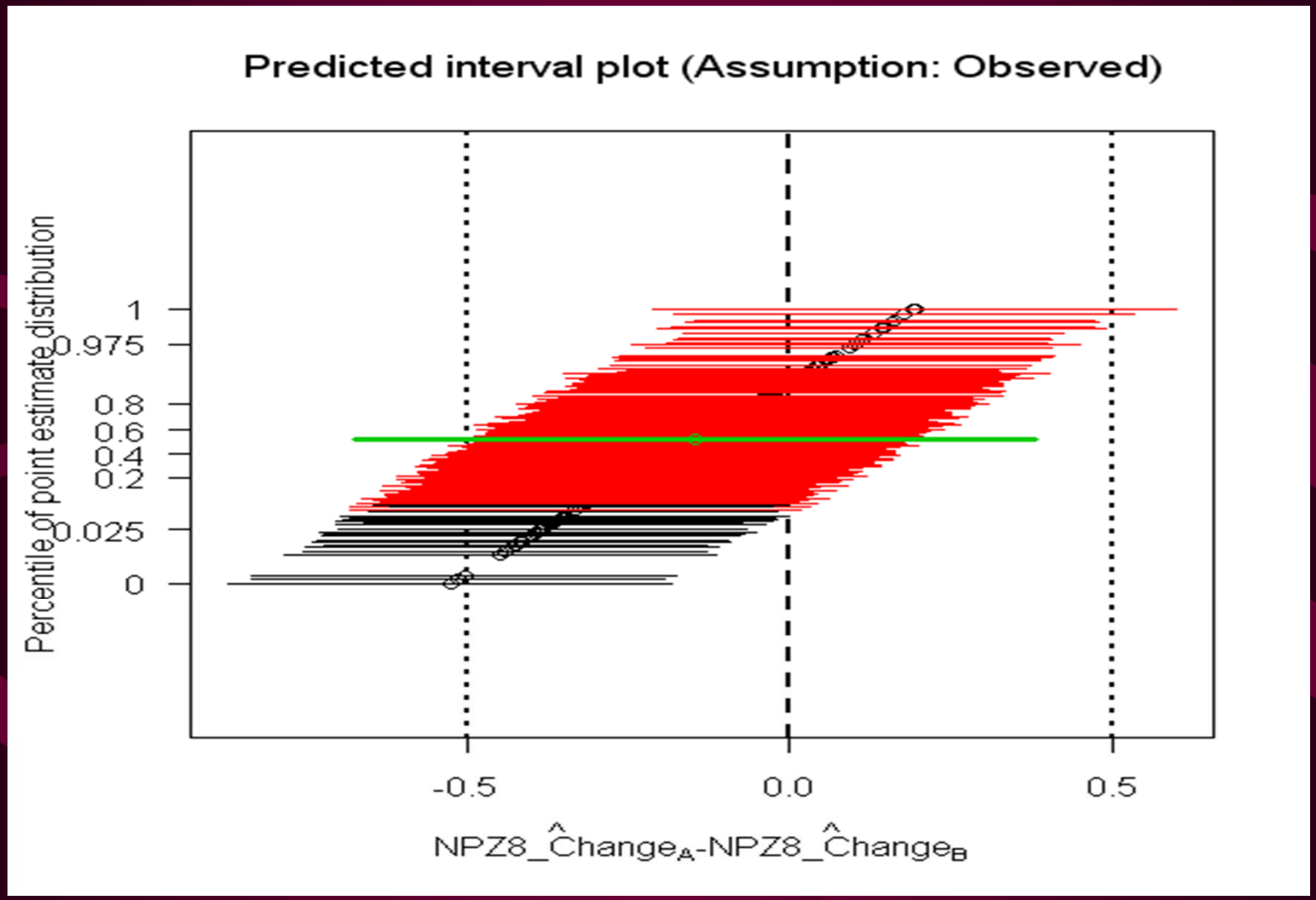
Simulate many times and get many adjusted PIs

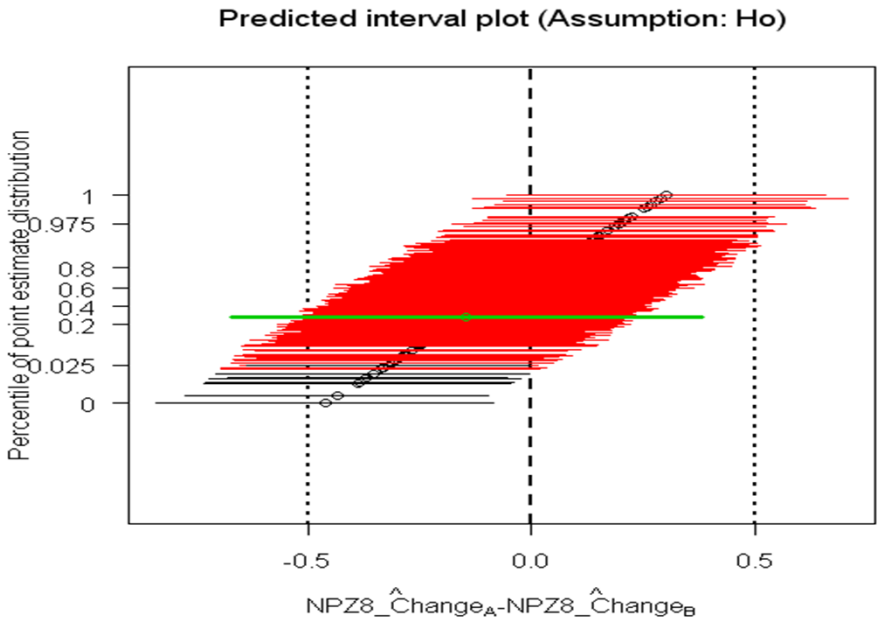
Minocycline for Cognitive Impairment in HIV

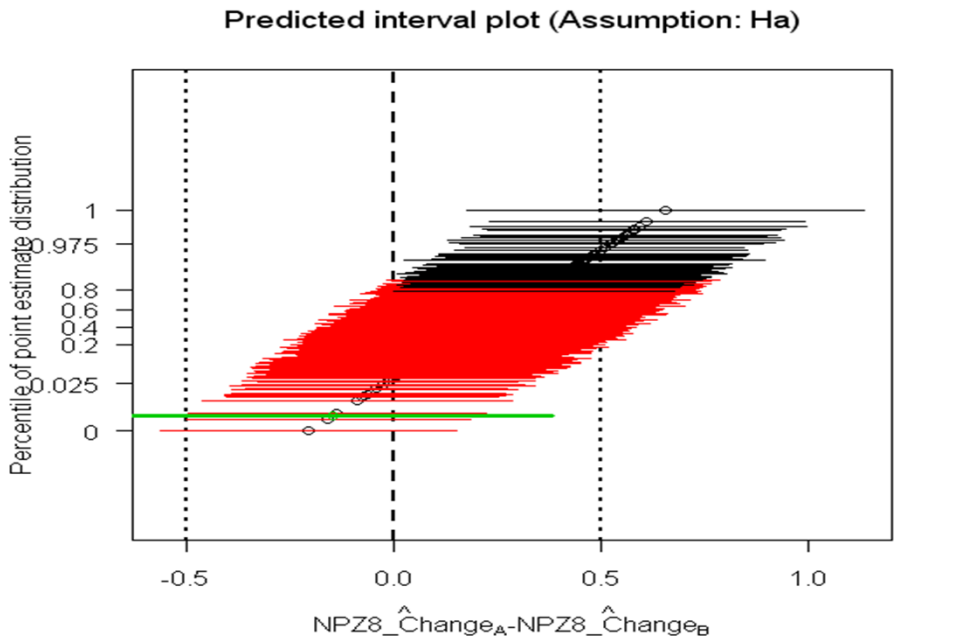
- Design
  - Randomized, blinded, placebo-controlled single site study
  - Primary endpoint: 24 week change in composite of standardized neuropsychological testing battery
  - N=100
- DSMB reviewed results after ~40% information

24-Week Change in NP Summary

	Total	Treatment Arm		p-value
		A	B	
U NP Sum Change				0.593
N	41	20	21	
# missing	24	13	11	
Mean (SD)	0.41 (0.80)	0.34 (0.84)	0.49 (0.78)	
Min, Max	-1.43, 2.06	-1.43, 2.06	-1.00, 1.66	
Median	0.58	0.60	0.52	
Q1, Q3	-0.17, 0.84	-0.30, 0.81	0.17, 0.93	


Predicted interval plot (Assumption: Observed)
Percentile of point estimate distribution
1
0.975
0.8
0.6
0.4
0.2
0.025
0
-0.5
0.0
0.5
NPZ8_^Change_A-NPZ8_^Change_B


Predicted interval plot (Assumption: Ho)
Percentile of point estimate distribution
1
0.975
0.8
0.6
0.4
0.2
0.025
0
-0.5
0.0
0.5
NPZ8_^Change_A-NPZ8_^Change_B


Predicted interval plot (Assumption: Ha)
Percentile of point estimate distribution
1
0.975
0.8
0.6
0.4
0.2
0.025
0
-0.5
0.0
0.5
1.0
NPZ8_^Change_A-NPZ8_^Change_B

Summary

- Futility analyses suggest low probability of a positive trial/superiority result with respect to primary endpoint
- DSMB evaluated if relevant effects could reasonably be ruled out
- DSMB evaluated if other reasons to continue (e.g., secondary endpoints)
- DSMB recommended early termination of the study

Extension: Bayesian Analog

R2WINBUGS

- Prior, E.g., continuous outcome $\mu_k \sim \mathrm{N}(\mu, \sigma^2);$ $\sigma_k{}^2 \sim \mathrm{IG}(c,d); \quad k = 1,2$
- Posterior using normal inverse gamma $\mu_j | y, \sigma_j \sim Normal - InverseGamma, \quad j = 1,2$
- Derive credible interval using MCMC
- Using beta-binomial, predicted posterior density $y^j_{new} | y \sim \text{Scaled T-distribution}, \quad j = 1,2$
- Draw samples from predicted posterior density and combine with interim data
- Use normal inverse gamma and obtain estimate from corresponding posterior
- Repeat many times
- BPIPs obtained from percentiles of many estimates

Return to NARC 009 Example

Table 1: Interim Analysis Results for NARC 009

				Bayesian interval estimate	
Treatment	N	95% CI for Mean	95% CI for Diff *	95% PI for Diff †	95% PI for Diff ‡
Placebo	31	(-0.35, -0.12)			
2 mg	34	(-0.24, -0.02)	(-0.05, 0.26)	(0.03, 0.19)	(-0.13, 0.03)
4 mg	34	(-0.36, -0.14)	(-0.18, 0.14)	(-0.10,0.06)	(-0.19, -0.02)
8 mg	32	(-0.21, 0.01)	(-0.03, 0.29)	(0.05, 0.21)	(-0.13, 0.04)
16 mg	36	(-0.32, -0.11)	(-0.14, 0.17)	(-0.06, 0.10)	(-0.17, -0.01)

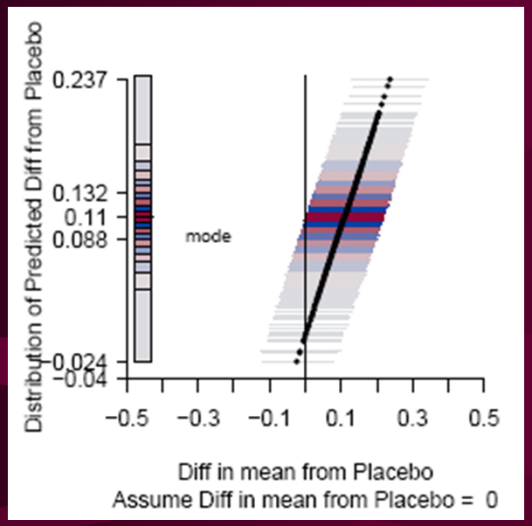
*: 95% CI for the difference in mean changes versus placebo.

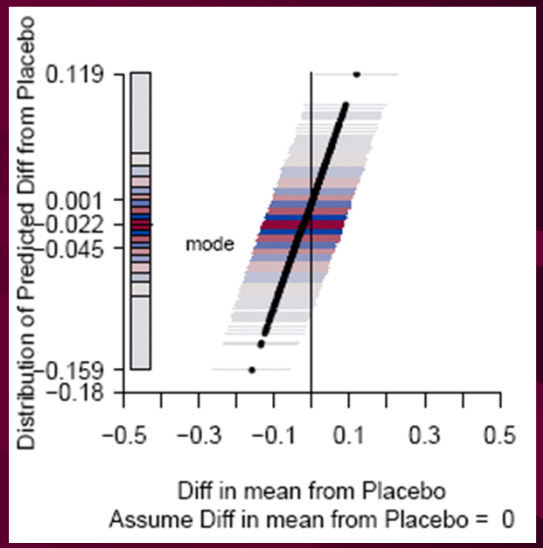
†: 95% PI for the difference in mean changes versus placebo assuming full enrollment, assuming current trend

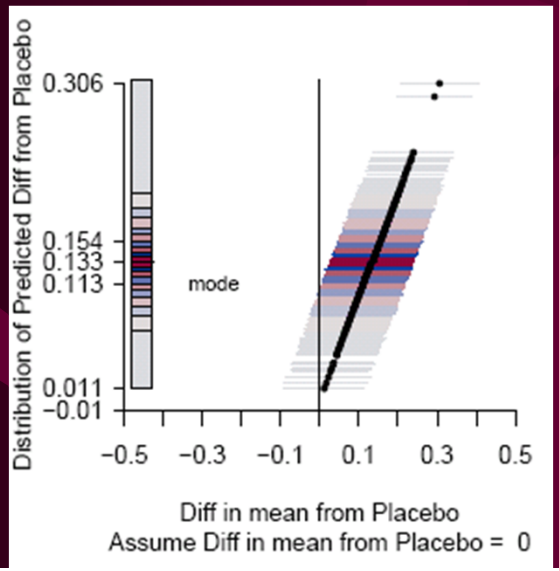
‡: 95% PI for the difference in mean changes versus placebo assuming full enrollment, assuming per protocol, $\mu_{placebo} = 0.17$ and $\mu_{drug} = 0.34$.

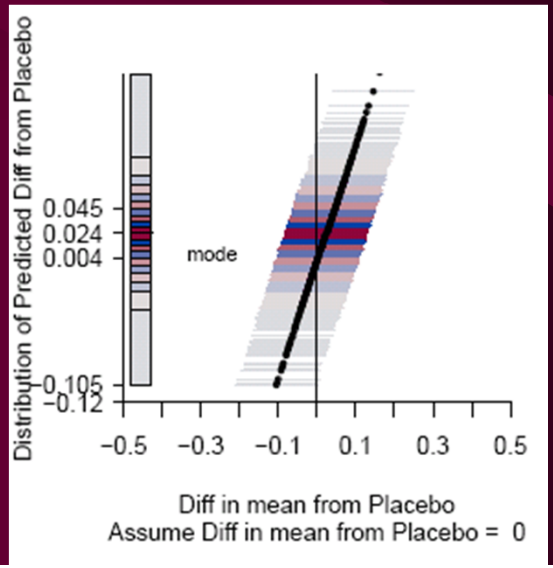
§: Difference in mean changes needed in the remaining patients for the CI for the difference in mean changes to exclude zero (in favor of active treatment) at the end of the trial.

Bayesian Predicted (Credible) Intervals (BPIPs)

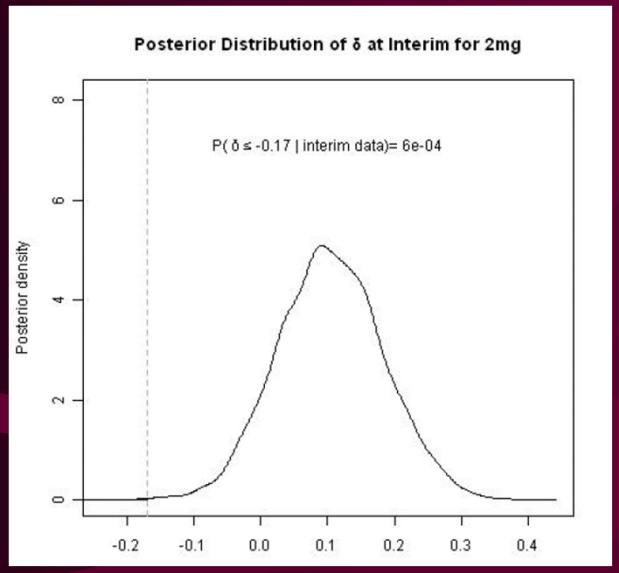


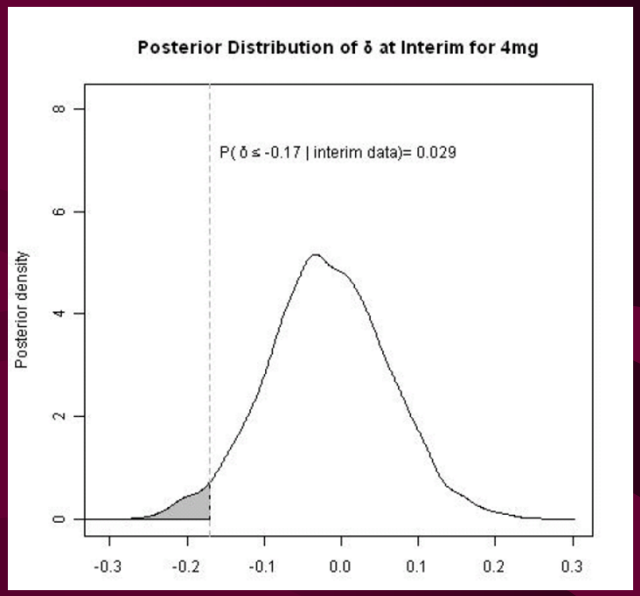


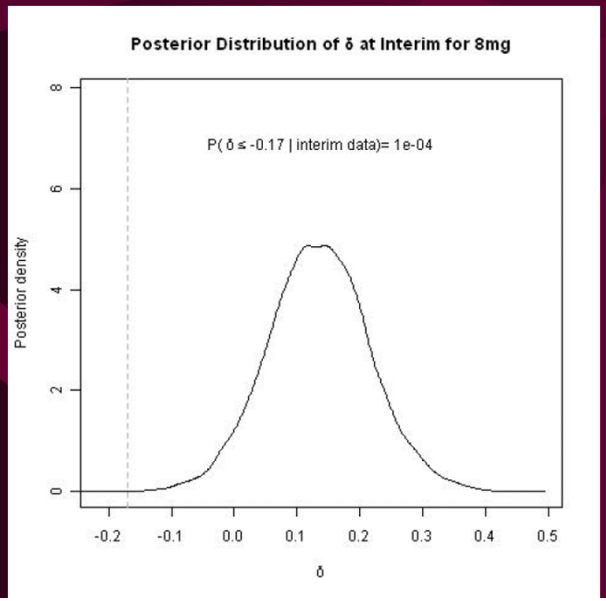


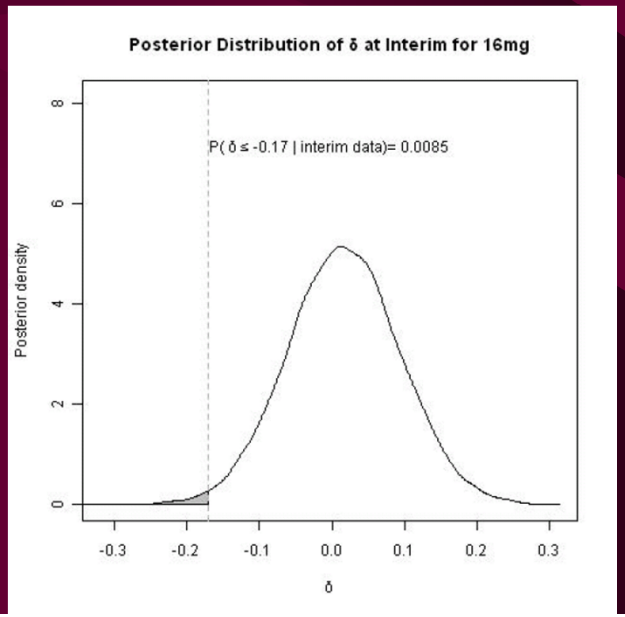


Posterior Distributions at Interim









Vision: Improving DMC Decisions
A More Informative and Flexible Approach

- PIPs (Frequentist Confidence Intervals)
  - Assumed Trend Continues
  - Null Hypothesis
  - Alternative Hypothesis
- BPIPs (Bayesian Credible Intervals)
  - Assumed trend continues
  - Null Hypothesis
  - Alternative Hypothesis
- Probability Density at Interim
- Predicted Probability Densities
  - Assumed Trend Continues
  - Null Hypothesis
  - Alternative Hypothesis

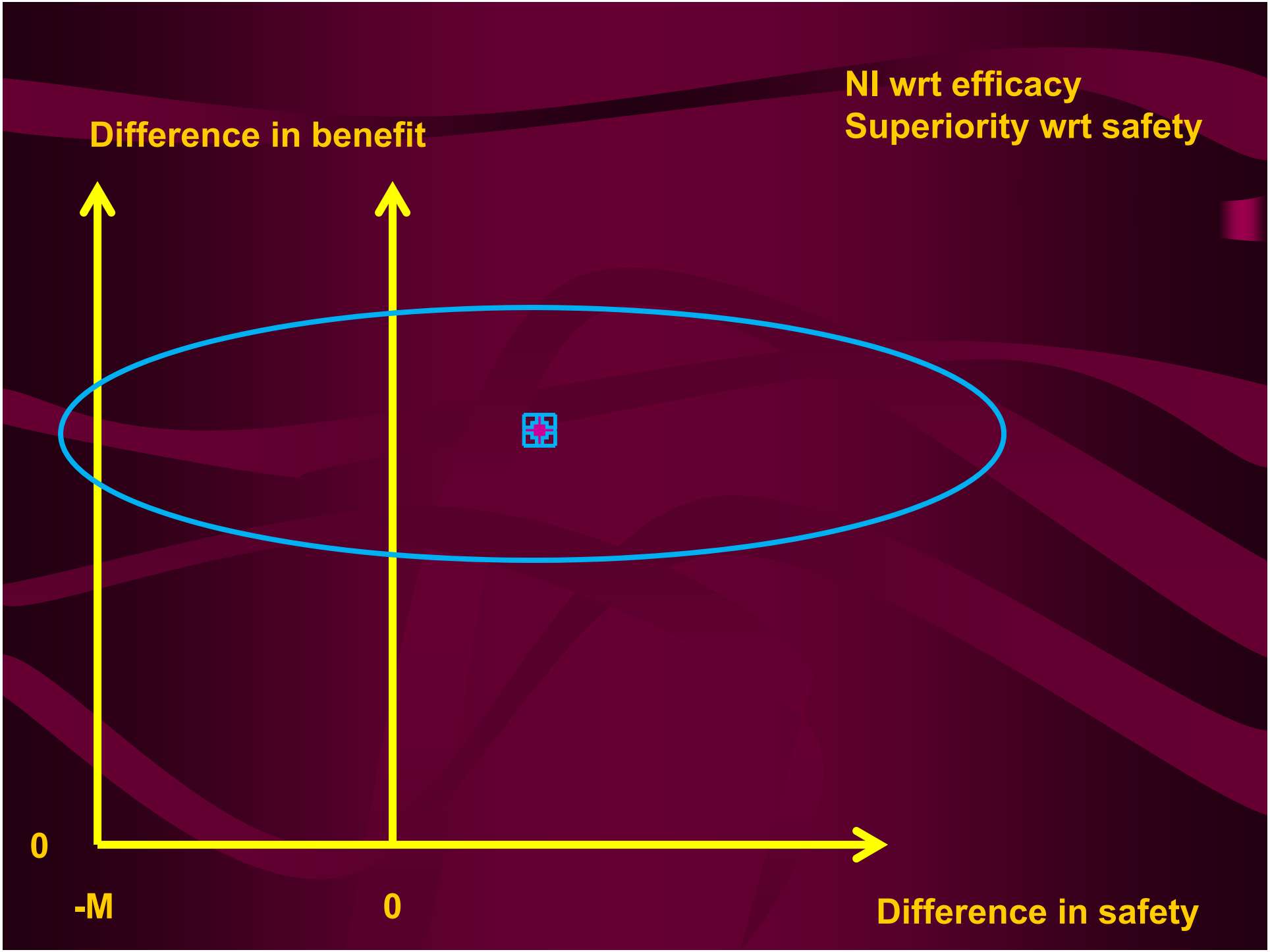
Extension to 2 Dimensions: Benefits and Risks

Benefits and Harms

- Suppose you measure benefit and harm in 2 dimension (e.g., a result can be plotted in 2-dimensional space and confidence rings can be constructed

- For example, consider a trial with two primary objectives
  - Demonstrate noninferiority with respect to efficacy
    - Show that between-arm difference is less than a selected noninferiority margin M
  - Demonstrate superiority with respect to safety

- Joint results can be plotted in 2 dimensions
  - Point estimate and associated 95% confidence ring


NI wrt efficacy
Superiority wrt safety
Difference in benefit
0
-M
0
Difference in safety

Design: Sample Size
Trials with Co-primary Endpoints

All continuous co-primary endpoints

Xiong *et al* (2005), Sozu *et al* (2006), Eaton, Muirhead (2007), Senn S, Bretz F (2007), Hung, Wang (2009), Sozu, Sugimoto, Hamasaki (2010, 2011), Sugimoto, Sozu, Hamasaki (2011), Kordzakhia, Siddiqui, Huque(2010), Asakura *et al*. (2011, presented at JJSM2011)

All binary co-primary endpoints

Song (2009), Sozu, Sugimoto, Hamasaki (2010, 2011)

All time to event co-primary endpoints

Sugimoto, Hamasaki, Sozu (2011, presented at MPC)

Mixed co-primary endpoints

Sozu, Sugimoto, Hamasaki (2010, presented at IBC2010)

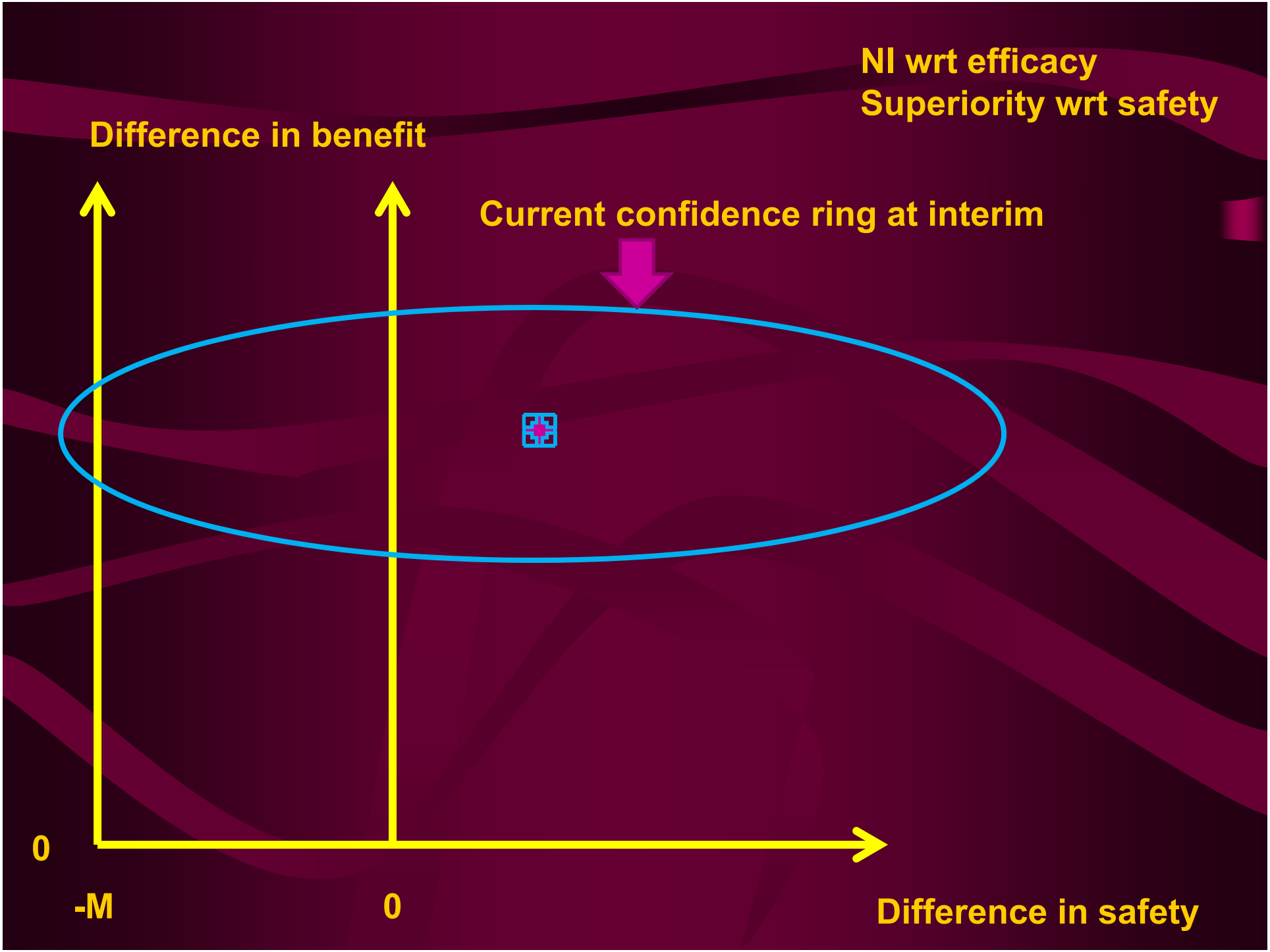
Sugimoto, Sozu, Hamasaki (2011, presented at MPC2011)

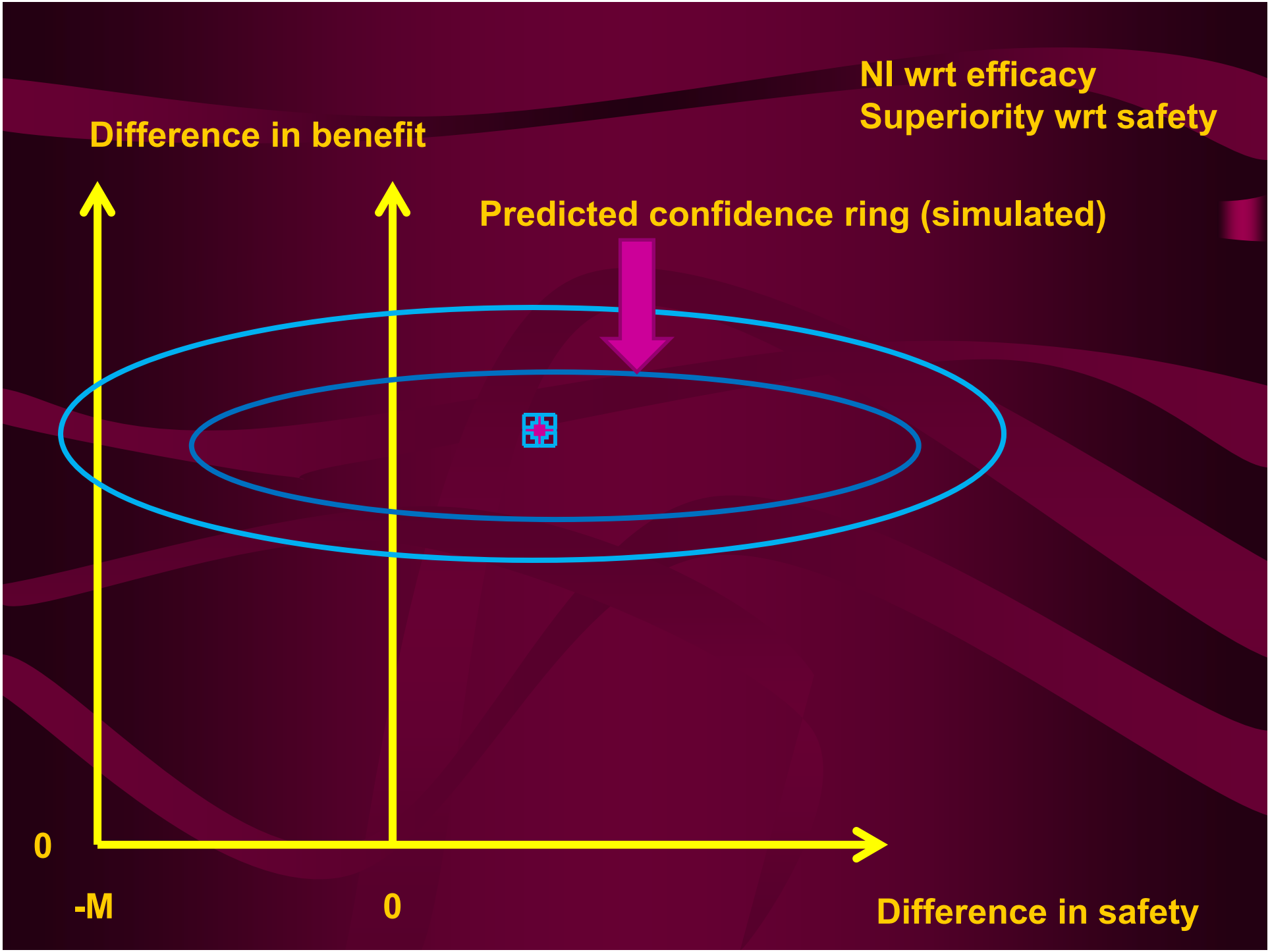
Predicted Confidence Rings

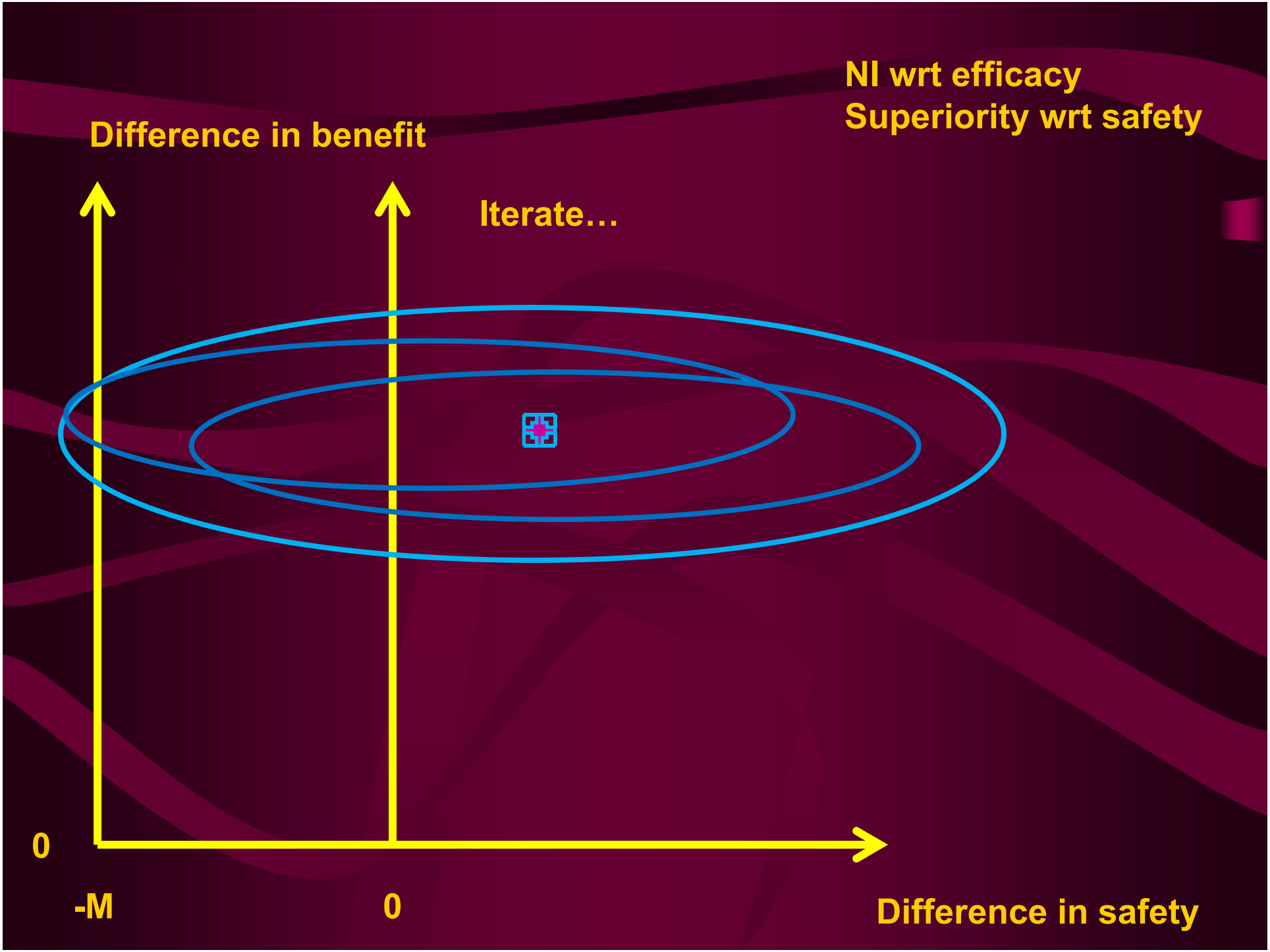
- Extend predicted interval idea to 2 dimensions (predicted confidence rings)
- Predict confidence ring at future timepoint (e.g., end of trial) conditional upon:
  1. Observed data
  2. Assumptions regarding future data (e.g., joint distribution: observed trend continues, H_A is true, H_0 is true, best/worst case scenarios, etc.)
  3. Simulation is used to account for random variation
- Use repeated confidence interval theory to control error rates when conducting multiple analyses

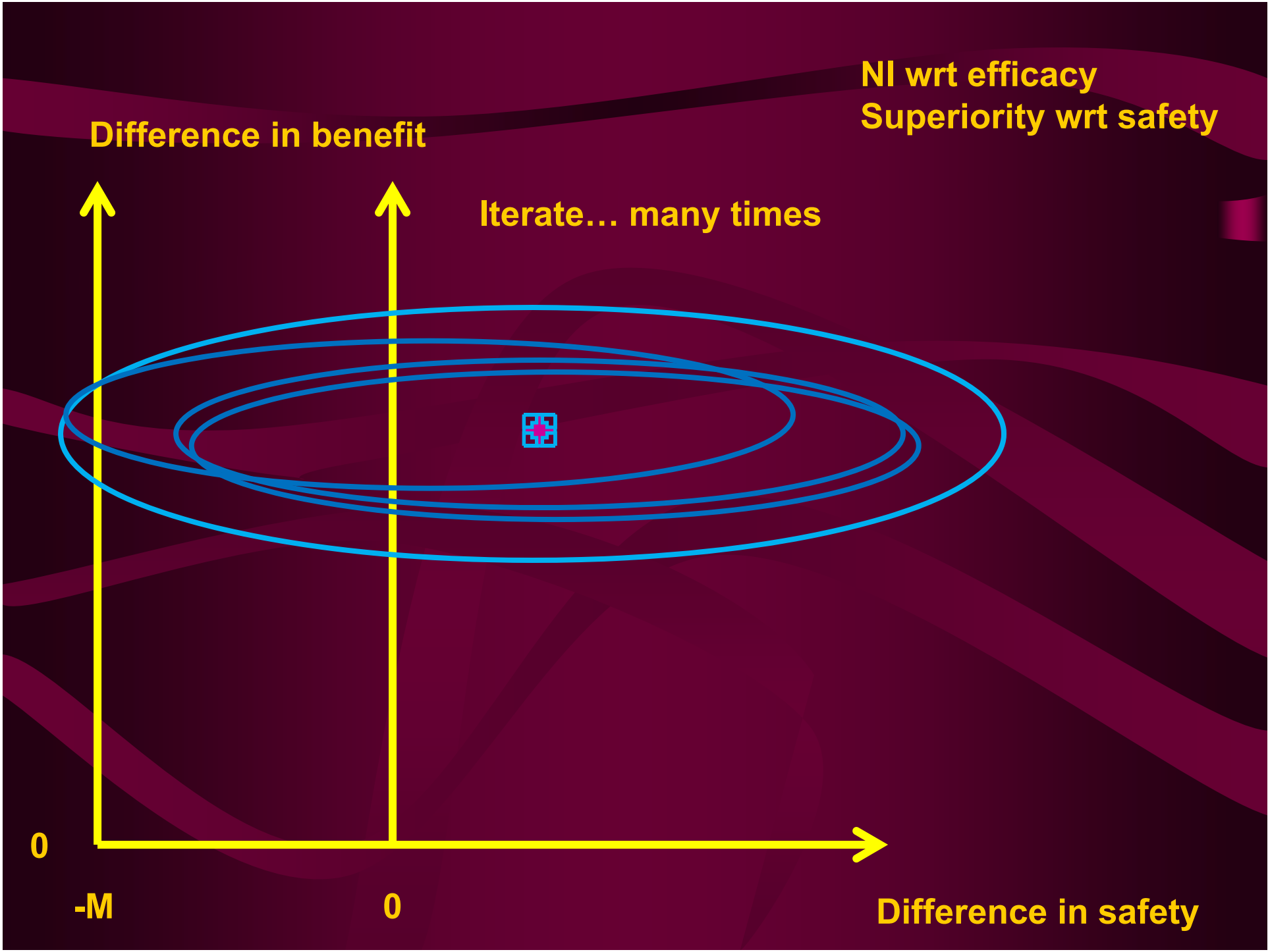
Predicted Confidence Ring Simulation

- Impose parametric assumption for joint distribution of unobserved data
  - Estimate or specify values of unknown parameters under reasonable and strategic assumptions
- Simulate future data
- Combine observed data with simulated data
- Construct predicted confidence ring
- Iterate


NI wrt efficacy
Superiority wrt safety
Difference in benefit
Current confidence ring at interim
0
-M
0
Difference in safety


NI wrt efficacy
Superiority wrt safety
Difference in benefit
Predicted confidence ring (simulated)
0
-M
0
Difference in safety


NI wrt efficacy
Superiority wrt safety
Difference in benefit
Iterate…
0
-M
0
Difference in safety


NI wrt efficacy
Superiority wrt safety
Difference in benefit
Iterate… many times
0
-M
0
Difference in safety

Summary

- Table
  - Summary of gain in precision with continuation
    - Reduction in ring area
    - Reduction of maximum width in benefit dimension
    - Reduction of maximum width in harms dimension
  - Probability of rejection joint null
  - Probability of rejecting in each of 2 dimensions
  - Vary assumptions as sensitivity analyses

- Tornado plot (with contours)

Changing Enpoints

Changing Endpoints

ClinPage: Pssst! Need To Change An End Point?

TODAY'S NEWS. TOMORROW'S TRIALS.

ClinPage

HARVARD PROF ON STATS SHIFTS

Pssst! Need To Change An End Point?

FEBRUARY 07, 2008

Comments Tell a Friend Print

Tags: biostatistics, irb, vioxx

Not long ago, on a bleak winter morning when the U.S. Congress and the New York governor and the Connecticut attorney general and the second most important medical journal were trashing the industry, we banged out an email to Scott Evans.

He's a senior research scientist in **Harvard University's School of Public Health**. He teaches biostatistics there, is an advisor to the FDA, and the lead investigator on a few large NIH projects.

More than a year ago, Evans authored **this cogent peer-reviewed article** (**it's also in PDF format**) about when it's appropriate to change primary end points in clinical trials. His analysis remains timely and cogent today, with some sponsors appearing to struggle with the question.

A sample quotation from the article: "Revisions to end points (particularly primary end points) should be uncommon. If not appropriately evaluated, such revisions lead to misguided research and suboptimal patient care."

So end points can *never* be changed? No, says Evans: "There are circumstances where it can be done and is acceptable."

ENHANCE Trial: Vytorin Story

- Compared Vytorin to simvastatin in preventing atherosclerosis (hardening of the arteries)
- Completed in 2006
- Registered in clinicaltrials.gov October 31, 2007
  - Endpoints entered differed from original design
  - Endpoint changed?
- Result: failure to improve heart disease
- Release of results delayed ~2 years
- Questions regarding stock sales of company executive
- Congress became involved
- Rationale: endpoint change expidicted analyses
- Difficult to determine if endpoint change is driven by science or business (perception issue regardless)

OPEN ACCESS Freely available online

PLOS CLINICAL TRIALS

Essay

When and How Can Endpoints Be Changed after Initiation of a Randomized Clinical Trial?

Scott Evans

- Changes to endpoints can compromise the scientific integrity of a trial
- Thus it is not generally recommended (concern for "cherry-picking" and inflation of type I error) except for some 2-stage protocols

Changing Endpoints is not Uncommon

- Chan et al. (*JAMA*, 2004) compared published articles with protocols for 102 randomized trials approved by the Scientific-Ethical Committees for Copenhagen and Frederiksberg, Denmark in 1994-1995, and reported that 62% of the trials had at least one primary endpoint that had been changed, introduced, or omitted.

- Chan et al. (*CMA*, 2004) compared published articles with protocols for 48 randomized trials approved for funding by the Canadian Institutes of Health Research in 1990-1998, and reported that primary endpoints differed between protocols and publications in 40% of the trials.

- Importance of precise definitions

Changing Endpoints

- New information could merit endpoint changes
  - Particularly in long-term trials with evolving medical knowledge
  - E.g., Results from other trials or identification of better biomarkers/surrogates
- Allow incorporation of up-to-date knowledge into design
  - Can theoretically be okay if you can show that the decision to change endpoints is “independent” of trial data (e.g., external data)
  - Practical demonstration of independence is difficult
  - DSMBs may not be appropriate decision-maker if they have seen the data

Changing Endpoints

- Could theoretically be okay if you can show that the decision to change endpoints is "independent" of trial data
  - Wittes, 2002 (*SiM*): "may consider changes to the primary endpoint when the trial has airtight procedures to gaurantee separation of the people making the decision from data providing insight into the treatment effect"
- Practical demonstration of independence is difficult
- DMCs may not be appropriate in this case because they have seen the data

Legit Examples

- Post CABG (post coronary artery bypass graft)
  - Compared 2 lipid-lowering drugs
  - No endpoint specified at protocol stage other than stating that it would be some measures of lipid disposition
  - Lipid disposition results would not available for 5 years
  - Team used this time to define endpoint and analyses
- Virologic trial
  - Lack of reliable data on virologic endpoints
  - 2 stages proposed with primary endpoint not specified
  - Stage 1 is a pilot to evaluate reproducibility and variability of virologic assays (candidates for the primary endpoint) with results used to determine the primary endpoint in stage 2 (not based on treatment effects)
  - Data from stage 1 not used in stage 2
  - Decision made by independent committee

Handling Endpoint Adaptations

Box 1. Proposal for Handling Changes in Endpoints in Clinical Trials

Questions to Ask:

- What is the source of the new information that triggers consideration of a change in endpoints?
- Have interim data on the endpoint (or related data) been reviewed?
- Who is making the decision to change endpoints? Are trial sponsors involved, or is there an independent external advisory committee?

Documenting the Endpoint Changes:

- Update the protocol in a formal protocol amendment.
- Update the clinical trial registry record.
- Revise the statistical analysis plan.

Reporting the Trial Results:

- Include a clear statement describing the changes in endpoints, and the information obtained after the start of the trial that led to these changes.
- Include a description of the reasons for these changes and the decision-making procedure.
- Consider the potential biases that may have come about as a result of the change in endpoints.
- Consider including a disclaimer that the results should be interpreted with caution and may need to be confirmed in future trials.
- Report the reasons for excluding endpoints from the analyses and whether this was independent of trial data.

- Perhaps use external committee other than DMC

Reporting Trials of Gabapentin

NEJM; November 2009

- 12 trials reported
  - 8 had a primary endpoint in manuscript different from protocol
  - New primary outcome in 6 trials
  - 5 trials failed to report protocol-defined primary outcomes
  - 8 positive trials of which 5 had a different endpoint

> When investigators find it necessary to make legitimate changes in a trial's protocol while studies are ongoing, these changes should be clearly stated in protocol amendments, clinical trial registry records, and the statistical analysis plan.[2] Furthermore, the published report should include a clear statement explaining the reason for the changes and their timing, potential biases that might have been introduced as a result, and the reasons for excluding outcomes from the analysis (if this was done).

Consideration of Revisions to DMC/DSMB Processes

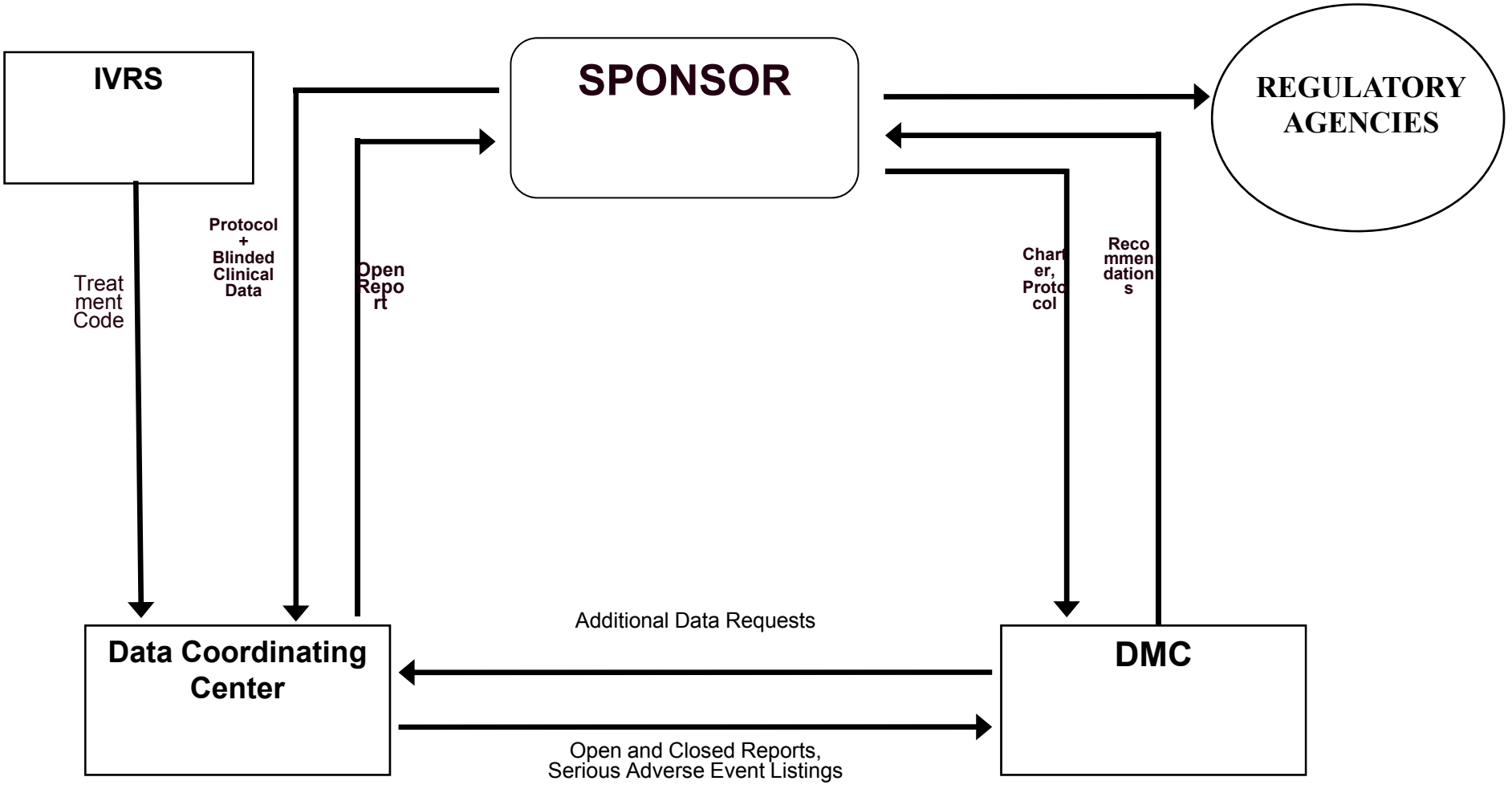
DMC Issues

- DMC should be independent, impartial, and unbiased
- DMC should understand key issues from sponsor perspective
- More modern charters necessary
  - Adaptive design issues

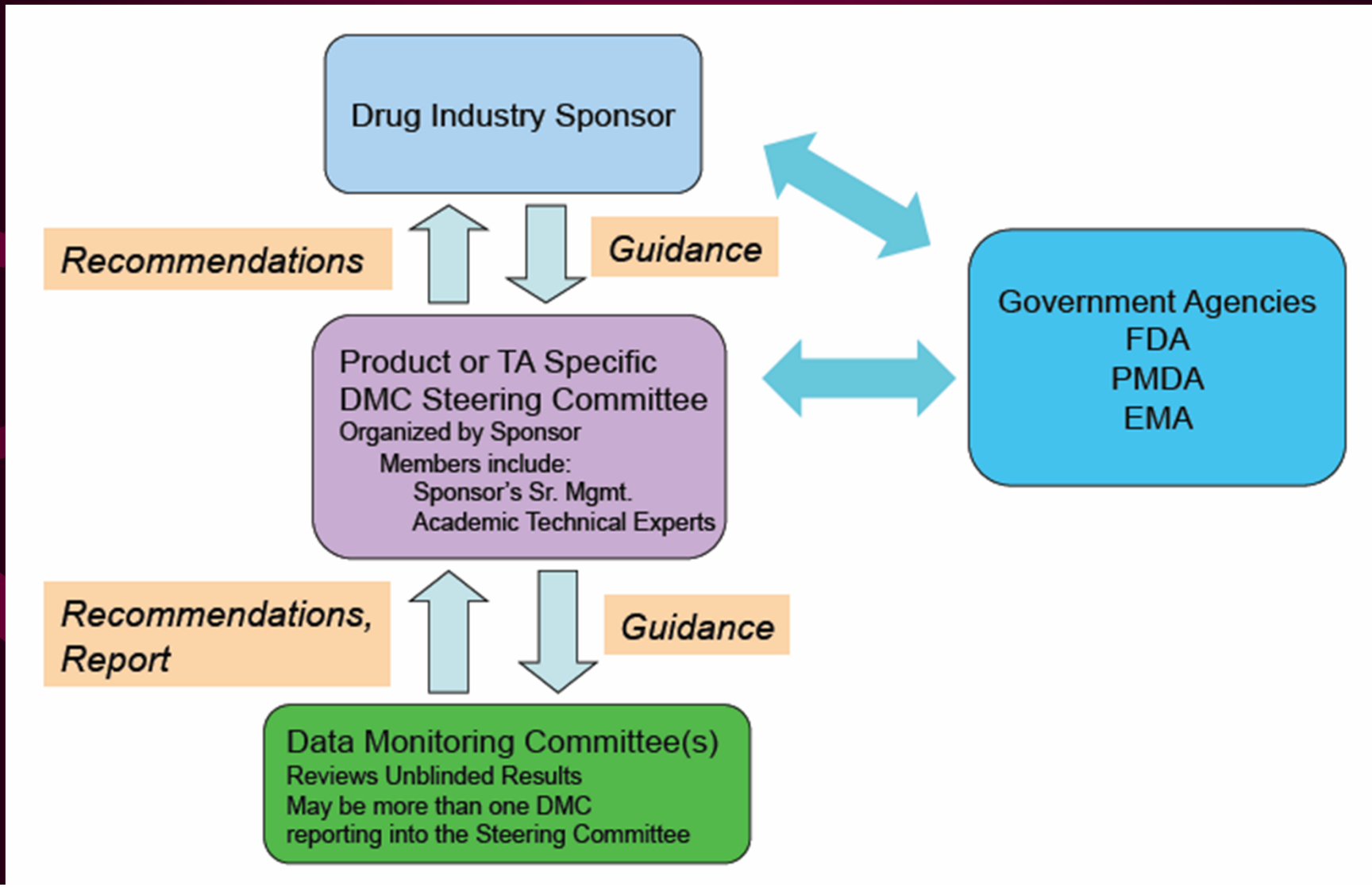
Sub-optimal process?

- DMCs have a lot of responsibility but sometimes have insufficient DMC experience, lack overall picture of development program, and an understanding of key benefit:risk issues
- Difficult to implement data-dependent adaptation via DMC without sponsor and regulatory input

Typical Organization Flow



A More Modern Proposal from Industry

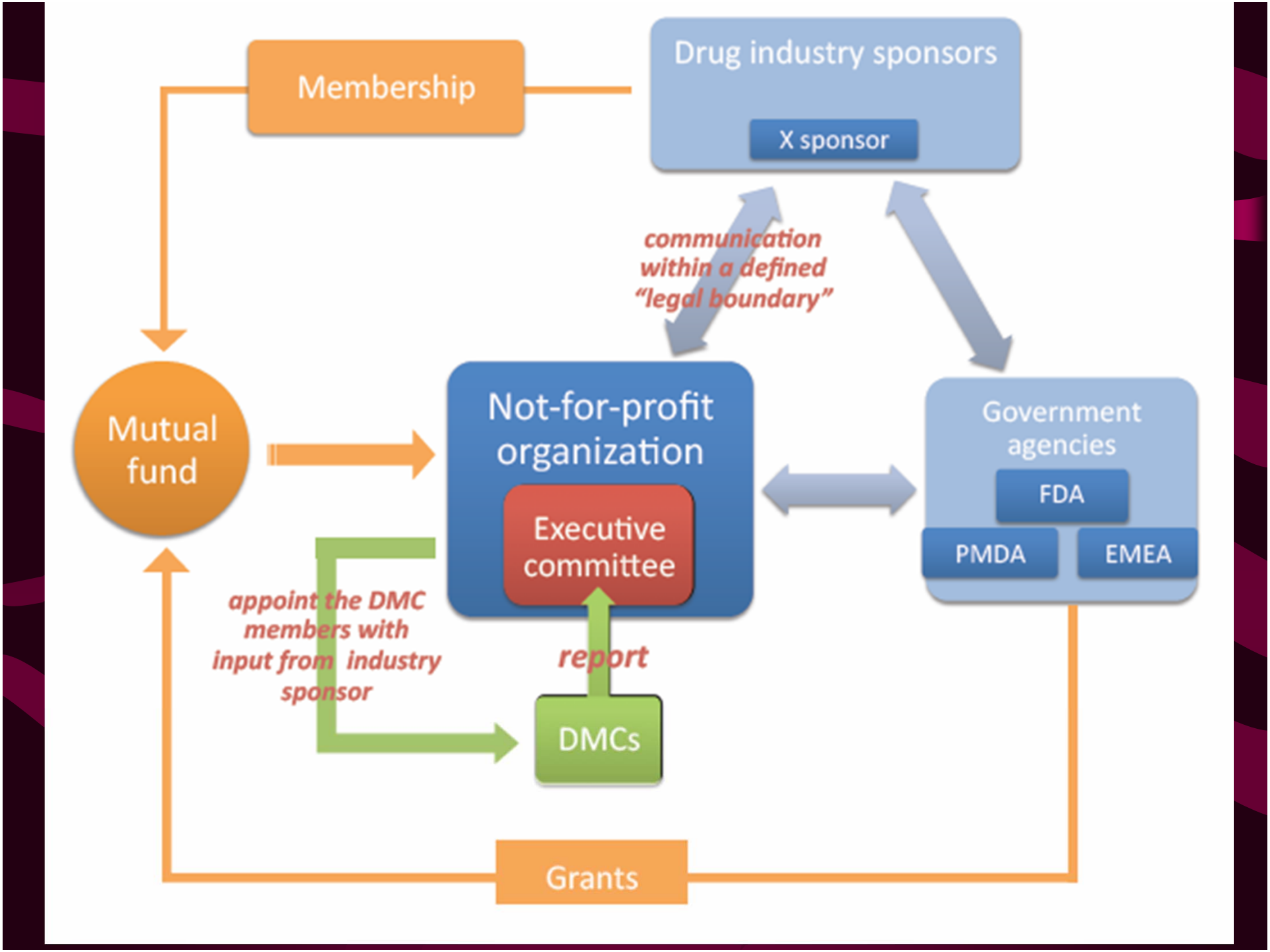


A Proposal from Academia

Use of a not-for-profit organization for coordinating data monitoring

Uses a “mutual fund” approach to support the organization

(from industry, government and foundation grants)


Membership
Drug industry sponsors
X sponsor
communication within a defined "legal boundary"
Mutual fund
Not-for-profit organization
Executive committee
Government agencies
FDA
PMDA
EMEA
appoint the DMC members with input from industry sponsor
report
DMCs
Grants

Not-for-Profit Organization

- Employs a flexible approach and includes sponsor input
- Appoints DMC and the independent statistical analysis team
- Keeps all confidential records of data monitoring
- Has a uniform, state-of-the-art reporting system
- Has an educational program for training DMC members
- Funds research projects / conferences on data monitoring related topics
- Handles insurance/indemnification issues

Not-for-Profit Organization
Executive Committee

- The Organization has a permanent Executive Committee (EC) with special members from academia, sponsor and DMC for each trial
- The EC has a broad view on the development program of a specific intervention.
- The EC makes the final decision regarding the DMC recommendation
- The EC communicates with the DMC
- The EC communicates with regulatory agencies for unexpected and critical issues

DMC Issues

- Poor reports
  - Need more knowledgeable independent statisticians
  - Need more graphics
  - Use rehearsal meetings
- Poor recommendation processes
- Lack of DMC member experience
  - Training/certification needed?
  - Need for statistician chairs?

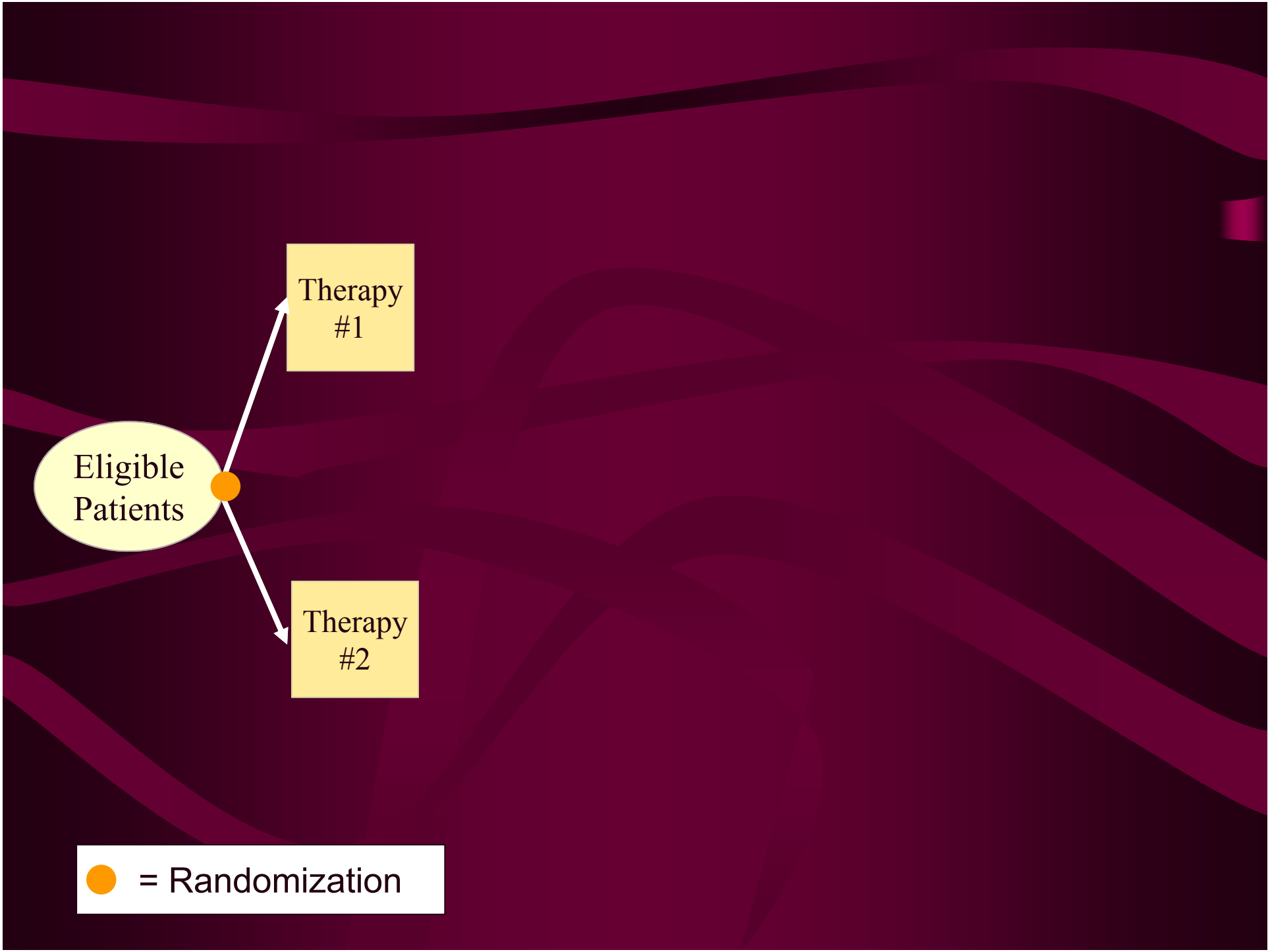
Example: One month ago…

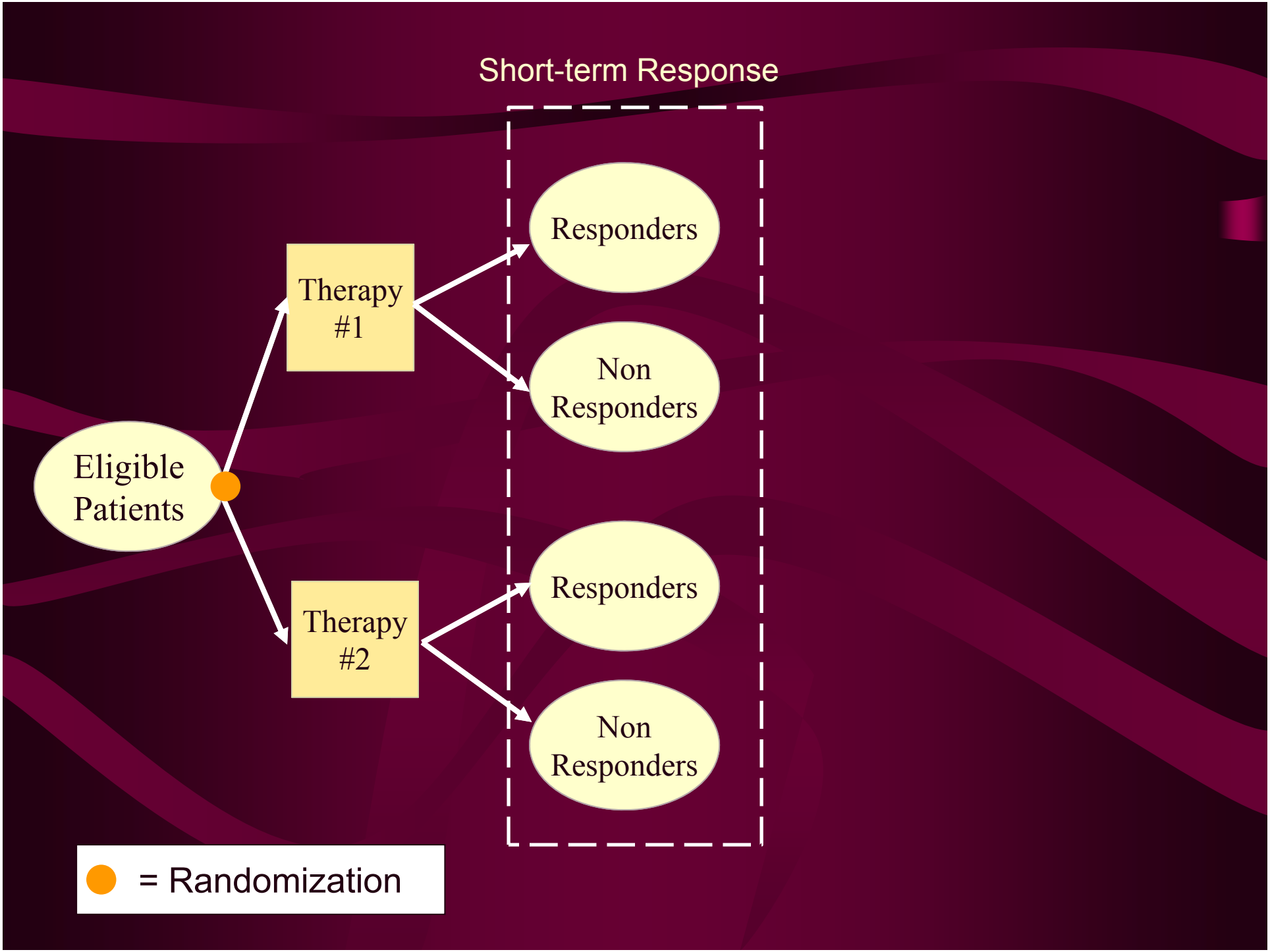
- Sitting on DMC for major pharma company
- 1-day prior to DMC meeting, I received a disk with the "report"
- 49 files on disk
- Opened 1st file
  - 1600 pages
  - No text

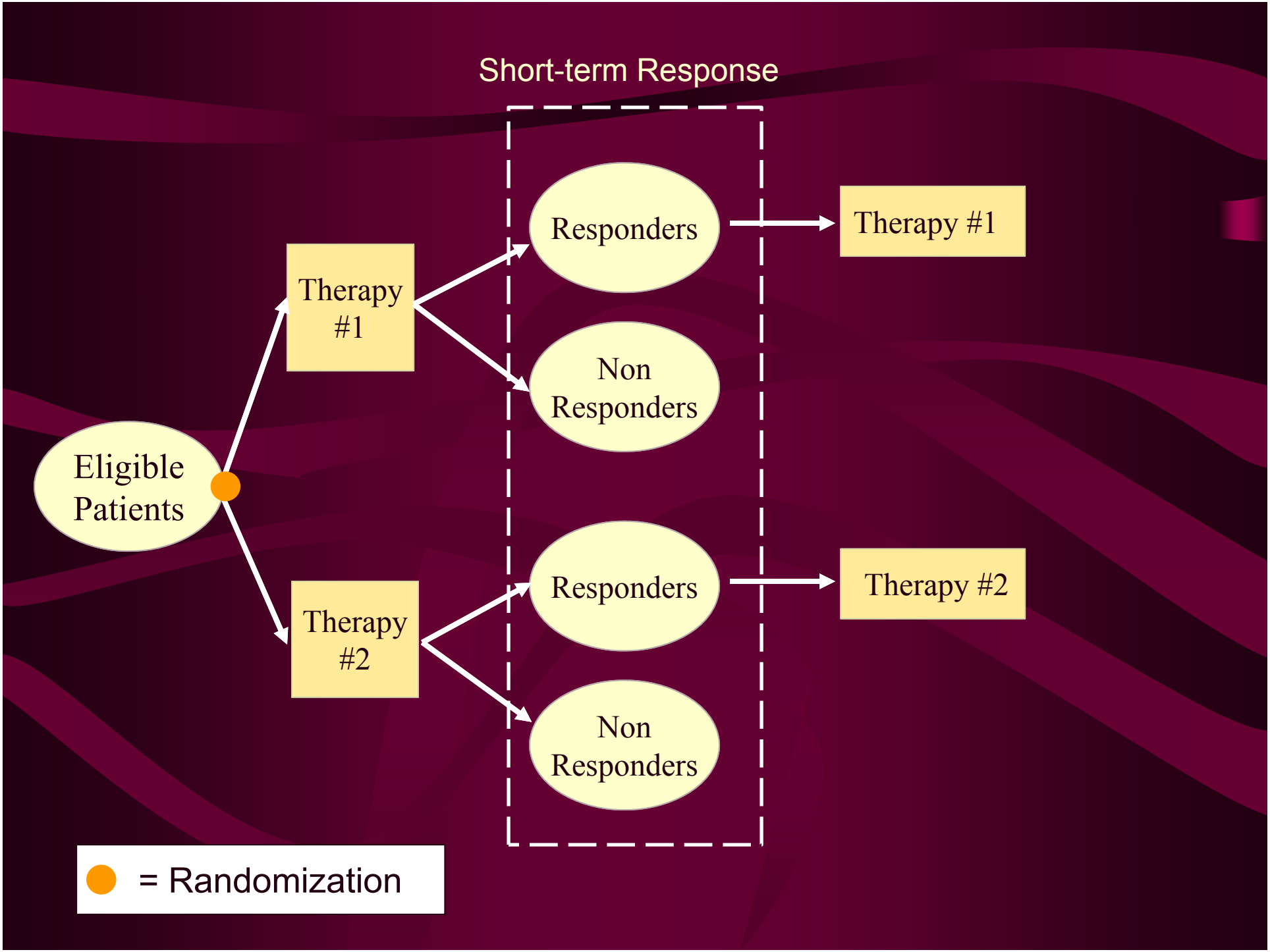
Response-Adaptive Treatment Regimes
Motivation

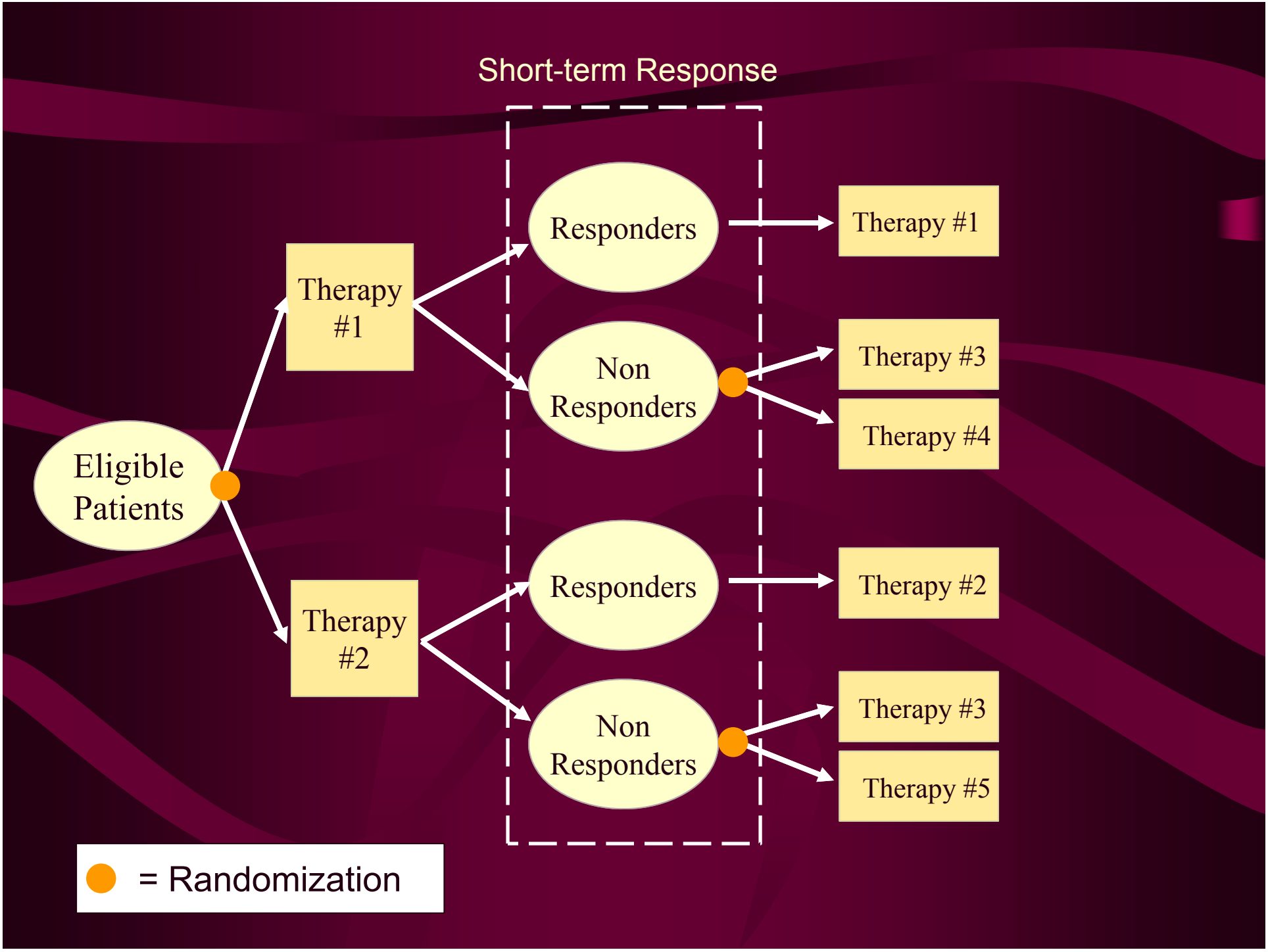
- Patient management
  - Not based on a single decision but sequential treatment decisions (adjustments of therapy over time) based on transitions of health states based on efficacy, toxicity, adherence, QOL, etc.
    - Tailored decisions based on individual patient response
  - Mixture of treatment of many diseases/coinfections which have short-term and long-term outcomes

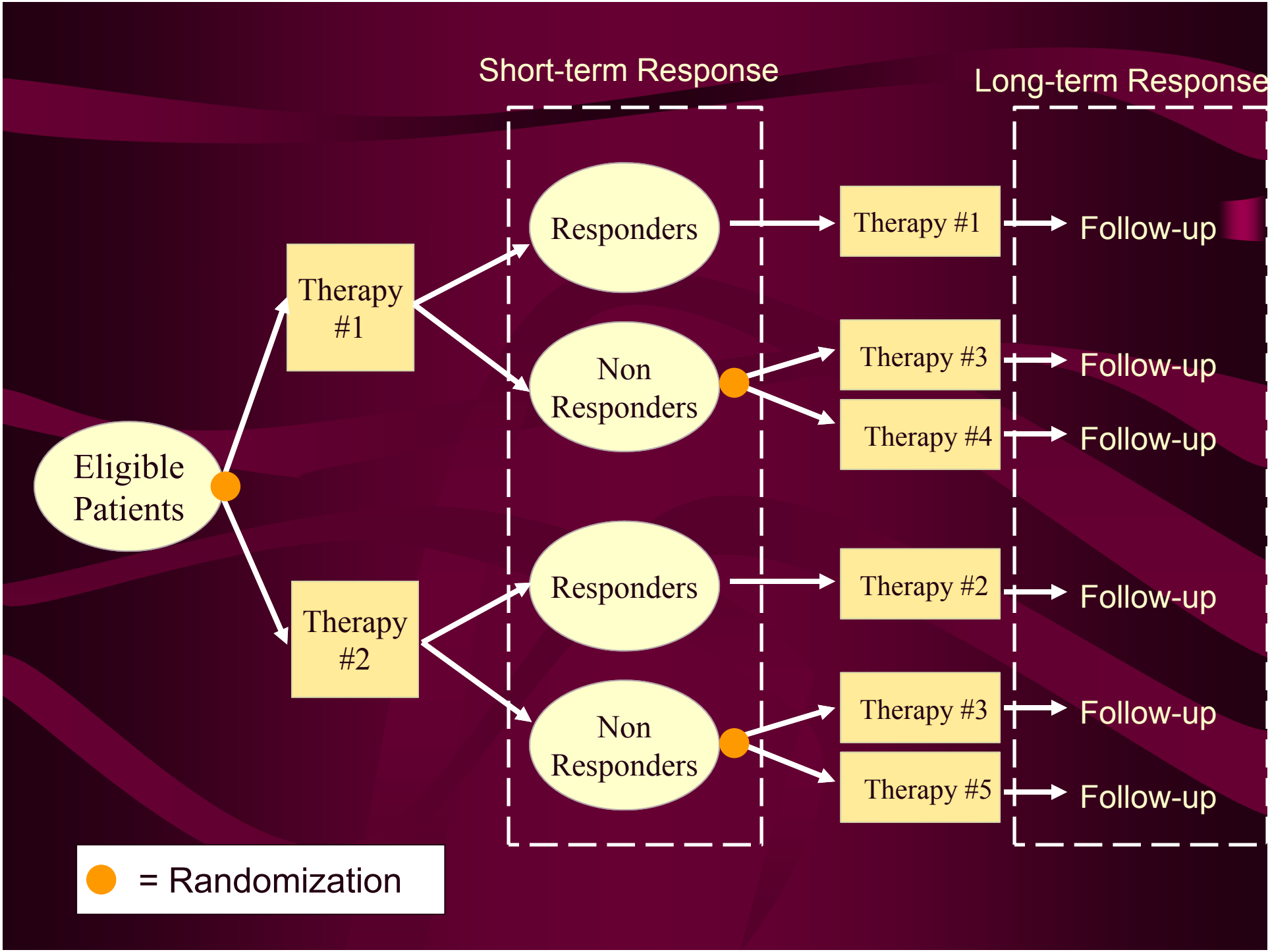
- Adaptive treatment regime designs
  - Compares sequential treatment strategies that are consistent with clinical practice
  - Uses sequential randomization


Therapy #1
Eligible Patients
Therapy #2
= Randomization


Short-term Response
Responders
Non Responders
Therapy #1
Eligible Patients
Therapy #2
Responders
Non Responders
= Randomization

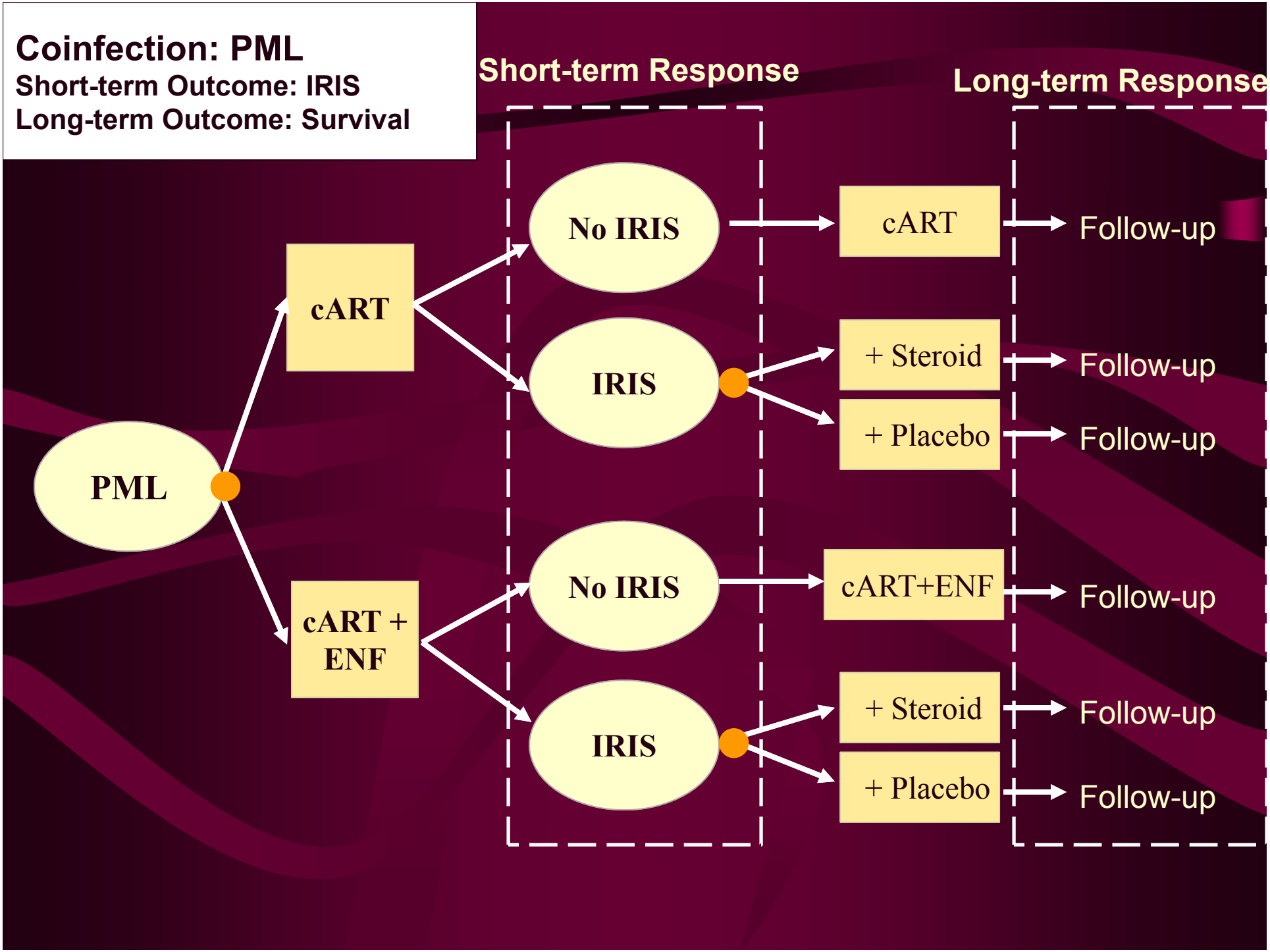

Short-term Response
Eligible Patients
Therapy #1
Therapy #2
Responders
Non Responders
Responders
Non Responders
Therapy #1
Therapy #2
= Randomization


Short-term Response
Eligible Patients
Therapy #1
Therapy #2
Responders
Non Responders
Responders
Non Responders
Therapy #1
Therapy #3
Therapy #4
Therapy #2
Therapy #3
Therapy #5
= Randomization


Short-term Response
Long-term Response
Eligible Patients
Therapy #1
Therapy #2
Responders
Non Responders
Responders
Non Responders
Therapy #1
Therapy #3
Therapy #4
Therapy #2
Therapy #3
Therapy #5
Follow-up
Follow-up
Follow-up
Follow-up
Follow-up
Follow-up
= Randomization

Example: HIV-Associated PML

- Design compares 4 treatment STRATEGIES
  - cART + steroids if IRIS is observed
  - cART without steroids
  - Enhanced-cART + steroids if IRIS is observed
  - Enhanced-cART without steroids
- Step 1
  - Randomized to cART or enhanced-cART (cART + enfuvirtide)
  - Observe patient response, particularly for IRIS
- Step 2
  - If no IRIS then patient continues with therapy
  - If IRIS, then randomize to steroids or placebo


Coinfection: PML
Short-term Outcome: IRIS
Long-term Outcome: Survival
Short-term Response
Long-term Response
PML
cART
cART + ENF
No IRIS
IRIS
No IRIS
IRIS
cART
+ Steroid
+ Placebo
cART+ENF
+ Steroid
+ Placebo
Follow-up
Follow-up
Follow-up
Follow-up
Follow-up
Follow-up

Adaptive Treatment Regimes

- Distinction between the regime (strategy dictating patient treatment) vs. realized experiences
  - Data from individual patients can contribute to multiple strategies
  - Patients on the same regime can have different treatment experiences
- Potential uses?
  - Coinfections (e.g., TB)
  - Personalizing treatments (e.g., HCV treatment duration)
- ITT complexity
  - Assigning treatment at later stages for patients LFU in early stages
  - Should consent patients to agree to ALL sequential randomizations (not by step)

2-Stage Designs

- Stage ≠ step: patients can enter multiple steps but only 1 stage
- "Internal pilot": Stage 1 vs. Stage II: learn vs. confirm
  - Hypothesis generation vs. hypothesis testing
- Efficiency advantage
  - Single trial addresses objectives traditionally addressed in two trials
  - Eliminates down-time between separate trials (but less thinking time)
  - IRB advantage (vs. approval of two trials)
- Classify by whether objectives or endpoints changes across stages
  - E.g., Similar objectives and different endpoints
    - Stage I: biomarker endpoint (e.g., tumor shrinkage)
    - Stage II: clinical endpoint (e.g., survival)

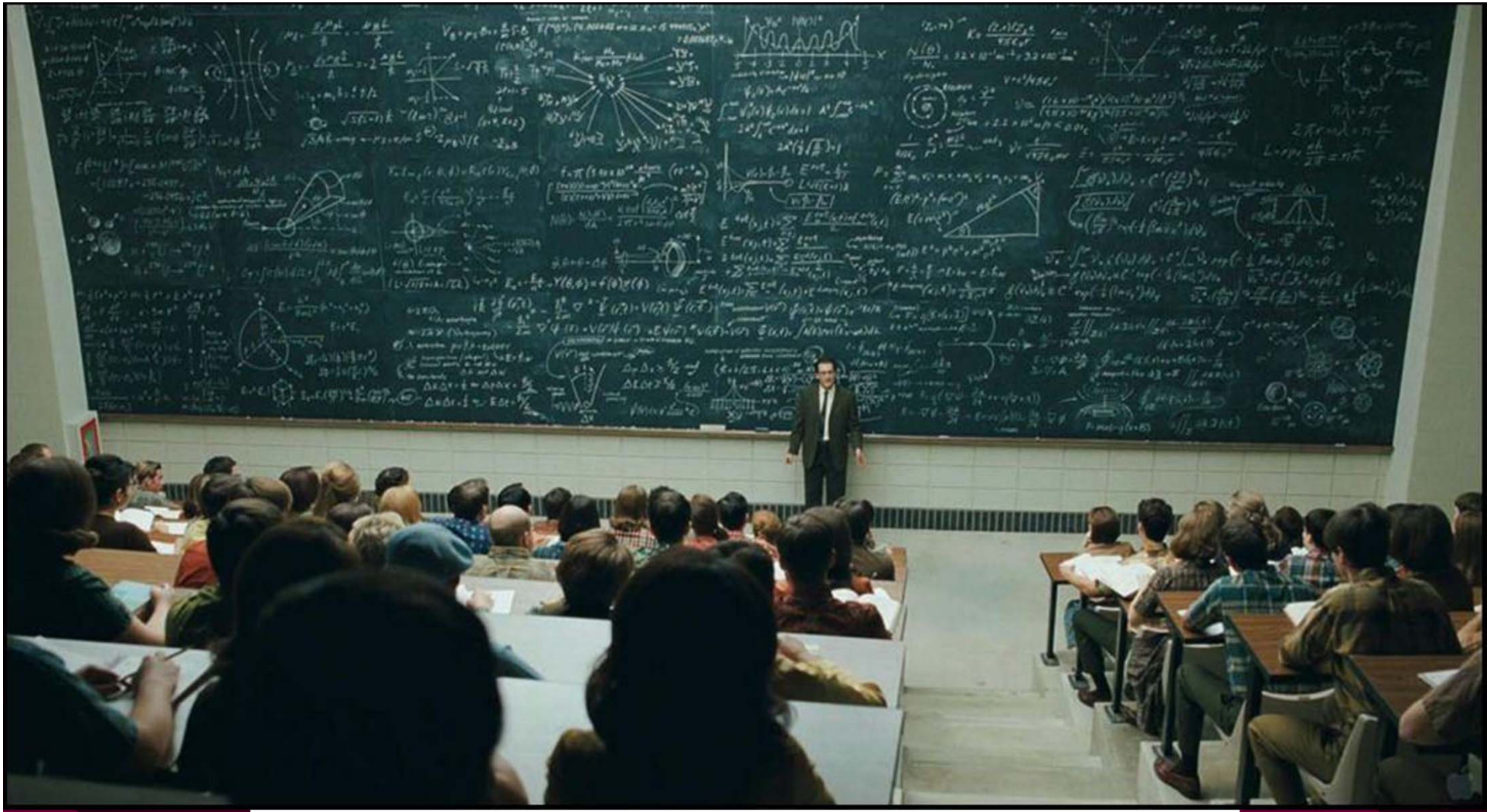
Adaptive Statistician

Many collaborating clinicians ask:

Adaptive Statistician

Many collaborating clinicians ask:

"Can we change statisticians? I'm tired of listening to Evans explain all of the mistakes we are making."



…as you can see, adaptive designs are intuitively obvious even to the most casual observer.