Design and Data Monitoring of Clinical Trials with Co-Primary Benefit:Risk Endpoints Using Prediction

Scott Evans, Ph.D., Harvard University

Toshimitsu Hamasaki, Ph.D., Osaka University

Kenichi Hayashi, Ph.D., Osaka University

7th IASC-ARS

Σ joint 2011

Taipei Symposium

December 16-19, 2011

Special Thank You

- Organizers
- Colleagues
  - Lingling Li
  - Satrajit Roychoudhury
  - Hajime Uno
  - LJ Wei

Outline

- Predicted Interval Plots (PIPs)
- Benefits and Risks
- PEPs, POPs, or PIGs?

Practical Questions During Trial Conduct

- 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

Limitations of Many Traditional Methods
E.g., Group sequential methods, conditional power, RCIs

- Do not
  - Provide estimates of effect or associated precision
  - Evaluate "clinical relevance"
  - Information regarding the reasons for high vs. low p-value
    - E.g., high p-values:
      - Negligible effect vs. insufficient data vs. too much variation
  - 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?

- Inflexible with binding decision rules based on a single endpoint

- Desire to base decisions upon assessment of benefits AND risks
  - And potentially other factors too such as: secondary endpoints, QOL, results from other trials, availability of new alternative therapies, cost:benefit considerations

Data Monitoring in Clinical Trials Using Prediction

Scott R. Evans, PhD
Lingling Li, PhD
LJ Wei, PhD
Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts

Clinical trials (CTs) are often monitored for efficacy or futility. Several methods for interim monitoring of CTs have been developed. Although informative, few of these methods convey information regarding effect sizes (eg, treatment differences), and none use prediction to convey information regarding potential effect size estimates and associated precision, with trial continuation. We propose use of prediction and specifically "predicted intervals" (PIs) as a flexible and practical tool for quantitative monitoring of CTs. PIs provide information regarding effect sizes, are invariant to study design, and provide flexibility in the decision-making process. We outline construction of PIs for binary, continuous, and time-to-event endpoints and present examples of their use. PIs provide a valuable tool for Data Monitoring Committees.

Evans SR, Li L, Wei LJ, "Data Monitoring in Clinical Trials Using Prediction", *Drug Information Journal*, 41:733-742, 2007.

Predicted Interval Plots (PIPS): A Graphical Tool for Data Monitoring of Clinical Trials

Lingling Li, Scott R. Evans, Hajime Uno, and L.J. Wei

Li L, Evans SR, Uno H, Wei LJ, "Predicted Interval Plots: A Graphical Tool for Data Monitoring in Clinical Trials", *Statistics in Biopharmaceutical Research*, 1:4:348-355, 2009.

Predicted Intervals

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

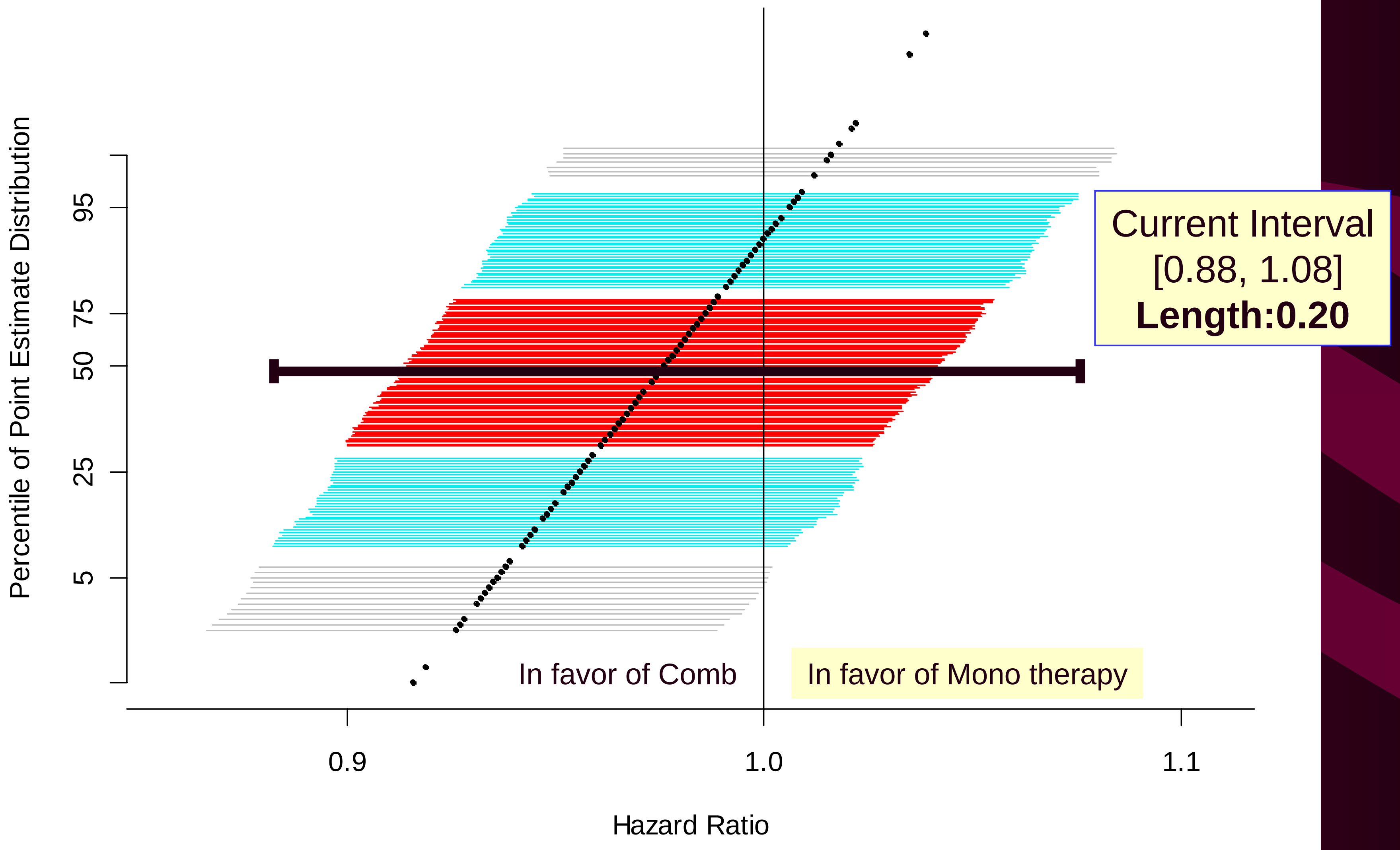
  1. Observed data
  2. Assumptions regarding future data (e.g., observed trend continues, H_A is true, H_0 is true, best/worst case scenarios, etc.)

- Use with repeated CI theory to control error rates

- PIP: Uses simulation to account for the sampling variability associated with the assumed model

PIP: Predicted Interval Plot

Assumed HR = 0.975



Benefits and Risks

Need for Systematic and Transparent Assessment of Benefits and Risks

- September, 2006
  - Congressionally mandated Institutes of Medicine study recommended that FDA develop and continually improve a *systematic* approach to benefit:risk

- December, 2006
  - European Committee for Proprietary Medicinal Products (CPMP) called for improved methodology leading to a more *systematic* approach to benefit:risk analysis

- April, 2009
  - EMEA Leaders Call for Regulator Refinement of Methods to Assess Benefit:Risk
    - Qualitative → Quantitative description of "net health benefit"
    - Ensuring safety → ensuring a positive benefit:risk profile
    - Communication of risk → communication of benefit:risk

Examples: Trial Endpoints
Benefits and Harms

- HIV
  - ACTG A5257 ARDENT Trial
    - Compares 3 nNRTI-based regimens for treatment of naïve HIV+
  - Efficacy endpoint: time to virologic failure
  - Safety endpoint: time to discontinuation due to toxicity

- Oncology
  - Efficacy endpoint: tumor response
  - Safety endpoint: dose-limiting toxicity

Benefits and Harms

- Suppose benefits and harms are measured in 2 dimensions

- Consider a trial with two primary objectives (composite hypotheses)
  - Demonstrate noninferiority with respect to efficacy
    - Show that between-arm difference is less than a selected noninferiority margin M, and
  - Demonstrate superiority with respect to safety

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

Design: Sample Size
Trials with Co-primary Endpoints

- Hamasaki T, Evans SR, Power and Sample Size Determination in Clinical Trials with Two-Correlated Binary Relative Risks *International Conference on Applied Statistics,* Taiwan, 2011

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

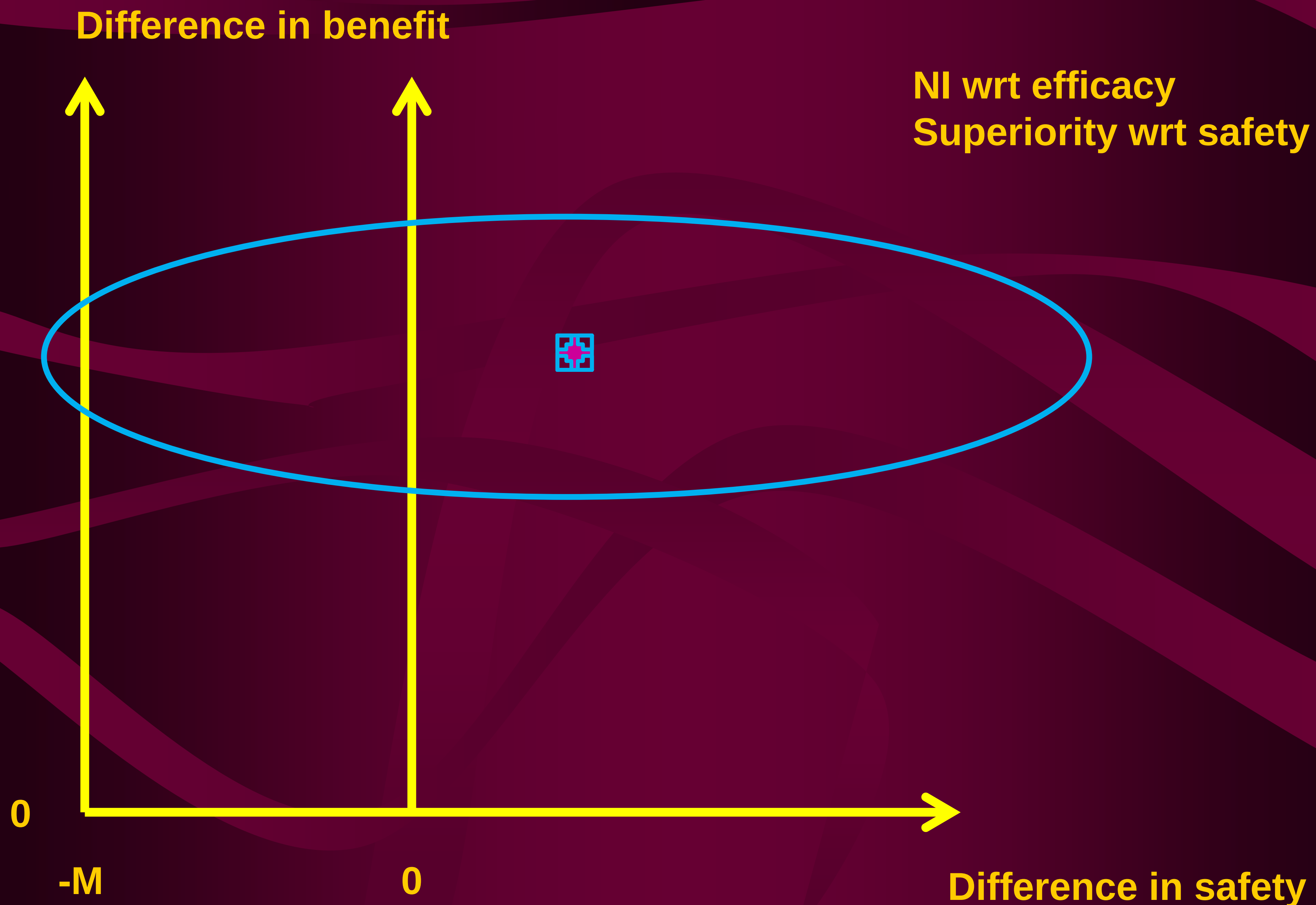
Sugimoto, Hamasaki, Sozu (2011, presented at MPC)

Mixed co-primary endpoints

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

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

Analysis Vision



Predicted Confidence Rings

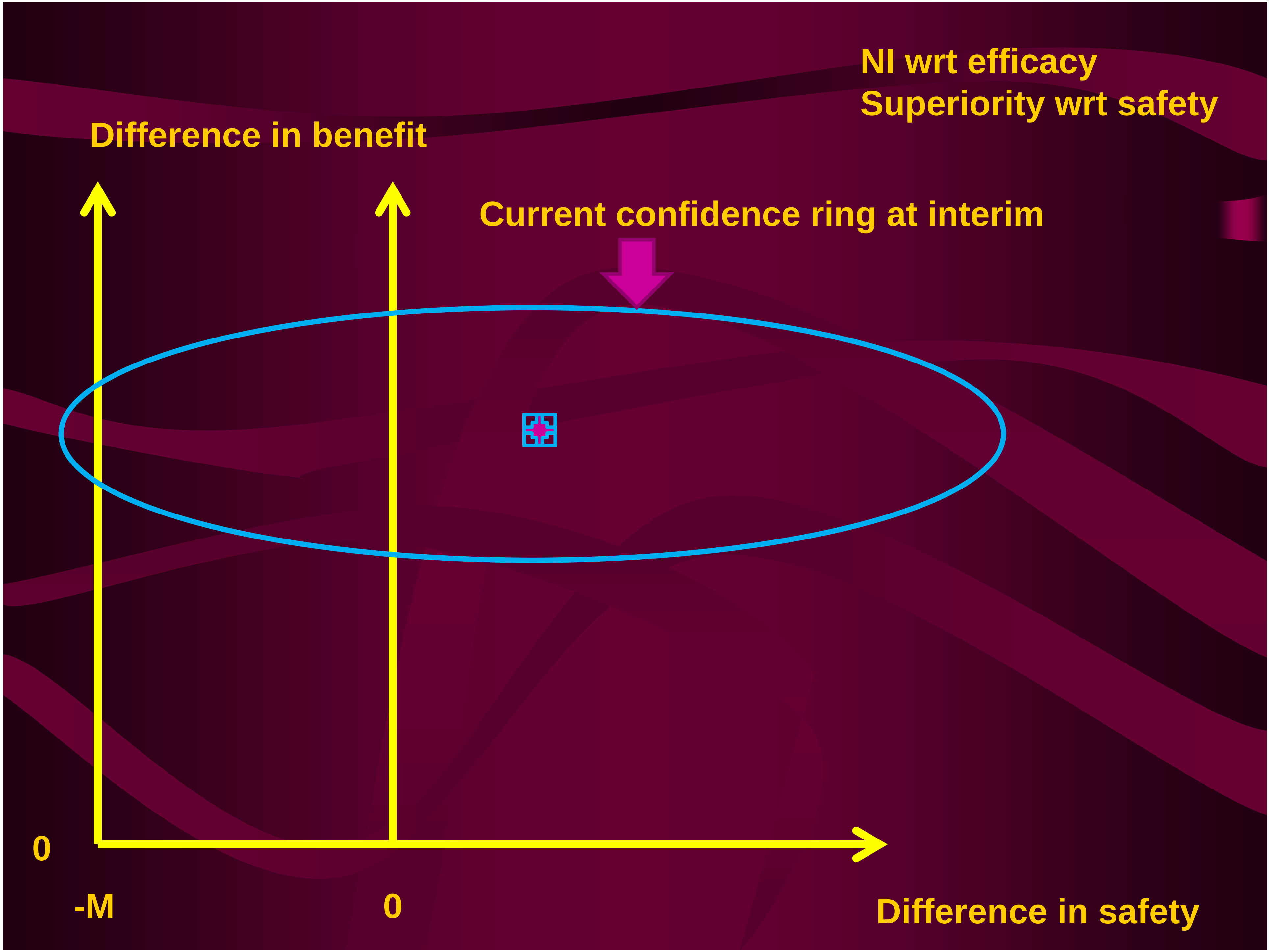
- Extend PIPs strategy to 2 dimensions

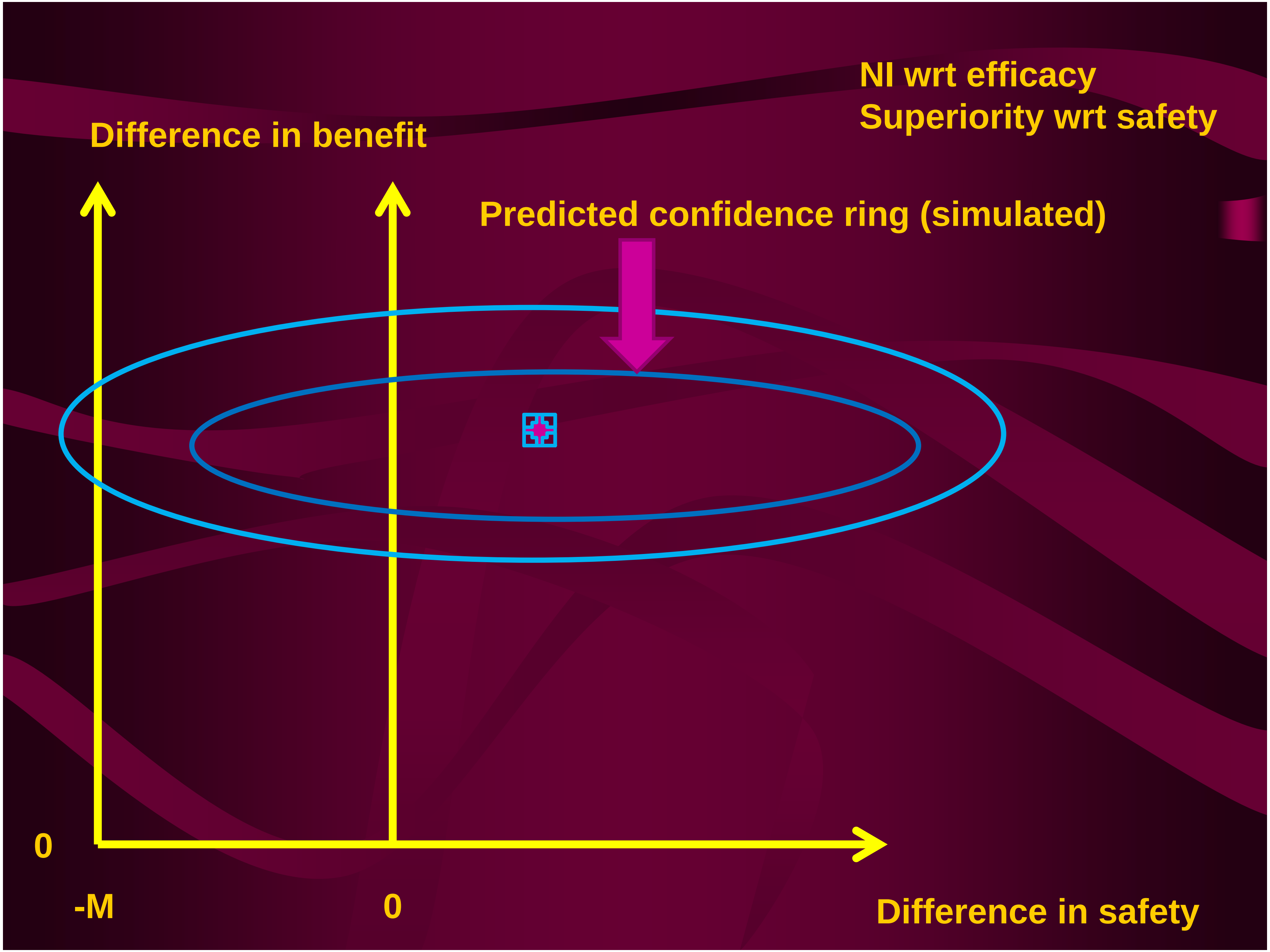
- Predict confidence ring at future timepoint (e.g., end of trial) conditional upon:
  1. Observed data
  2. Assumptions regarding future data (e.g., joint distribution: observed trend continues, H_A is true, H_0 is true, best/worst case scenarios, etc.)
  3. Simulation is used to account for random variation

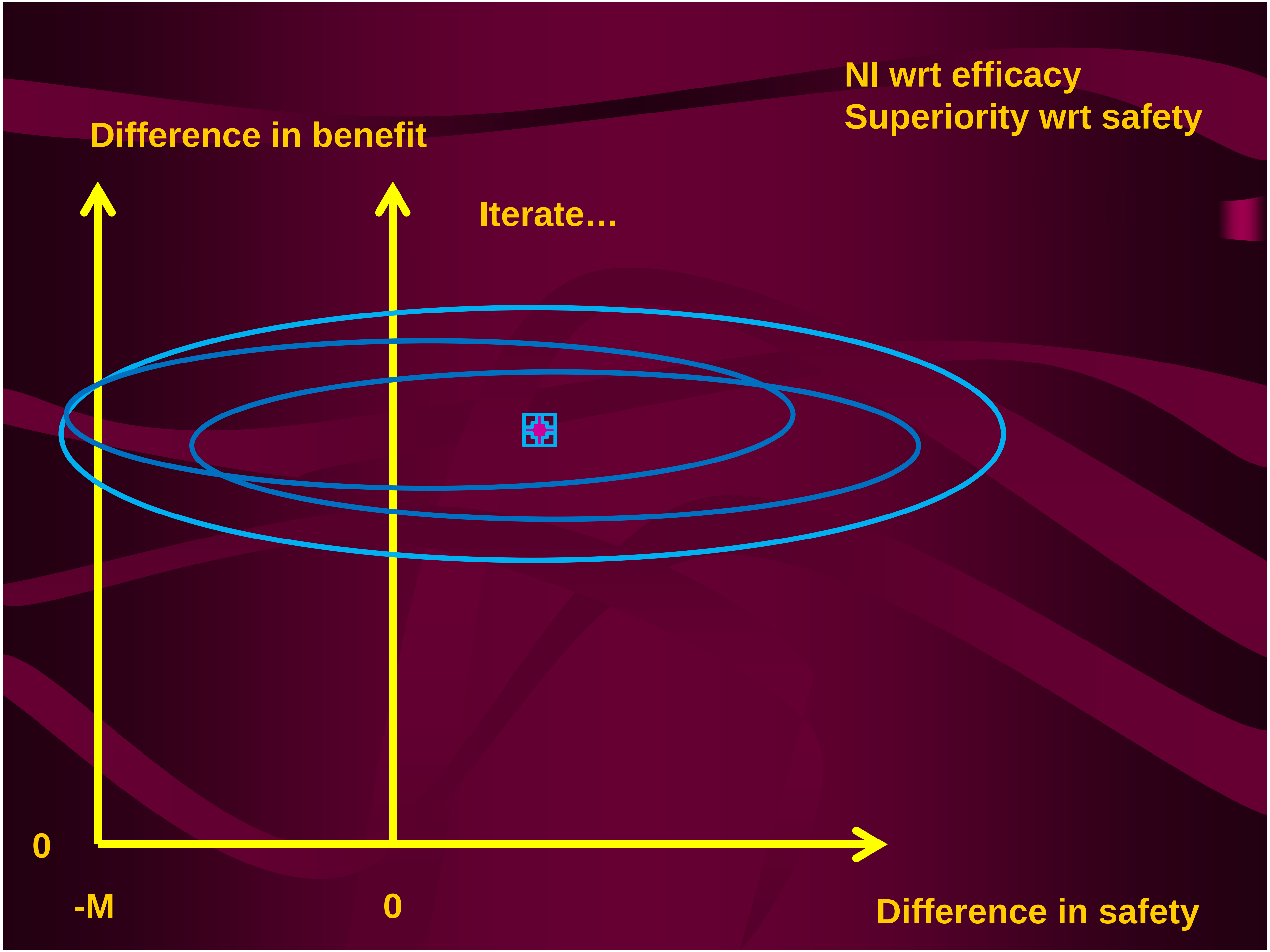
- Use repeated confidence interval theory to control error rates when conducting multiple analyses

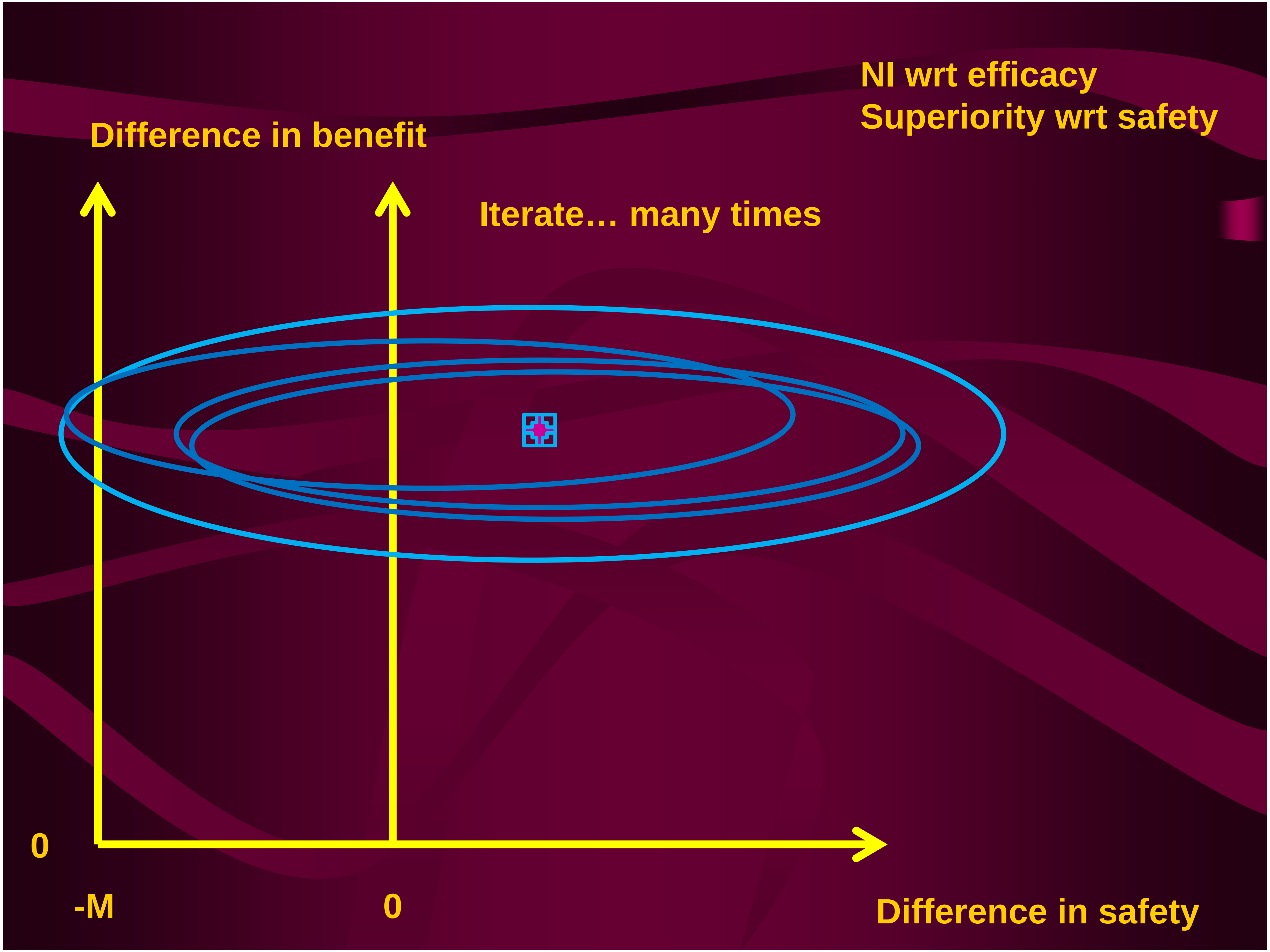
Predicted Confidence Ring Simulation

- 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 many times


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

DESIGN

TOTAL SAMPLE SIZE REQUIRED FOR 80% POWER FOR JOINT HYPOTHESES: 714 (357 per group)
TYPE I ERROR 0.025
Total Sample Size required for Endpoint #1: 186
Total Sample Size required for Endpoint #2: 712

	E1	E2
TEST	50.0%	40.0%
CONTROL	30.0%	30.0%
COMMON CORRELATION	0.20	

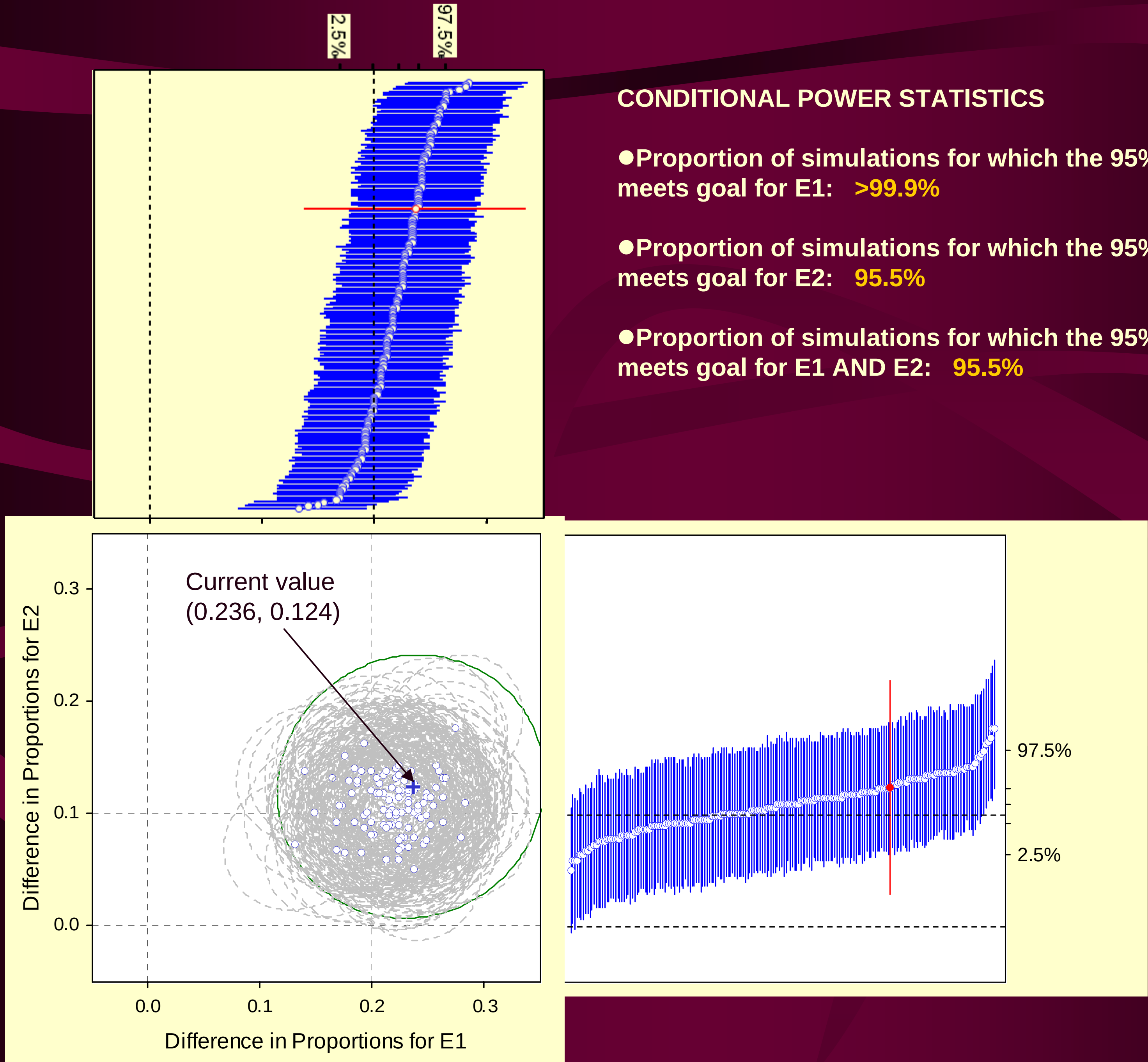
OBSERVED INTERIM VALUES

ACUMULATED SAMPLE SIZE: 357 (178 per group)

	E1	E2
TEST	51.1%	38.2%
CONTROL	27.5%	25.8%
COMMON CORRELATION	0.04	

ASSUMPTION: FUTURE VALUES
E.g., under alternative hypothesis

	E1	E2
TEST	50.0%	40.0%
CONTROL	30.0%	30.0%
COMMON CORRELATION	0.04	



CONDITIONAL POWER STATISTICS

- **Proportion of simulations for which the 95% confidence ring meets goal for E1: >99.9%**
- **Proportion of simulations for which the 95% confidence ring meets goal for E2: 95.5%**
- **Proportion of simulations for which the 95% confidence ring meets goal for E1 AND E2: 95.5%**

Summaries

- Gain in precision with continuation (reduction in ring area)

- Plots
  - Tornado plot stacking predicted rings like pancakes
  - Contour plot
    - Get 1000 predicted rings
    - For each point in 2 dimensions, calculate the proportion of predicted rings that contain the point (note each point is not associated with a proportion)
    - Create contours of similar proportions
    - Superimpose current confidence ring

- Sensitivity analyses: vary data-generating assumptions

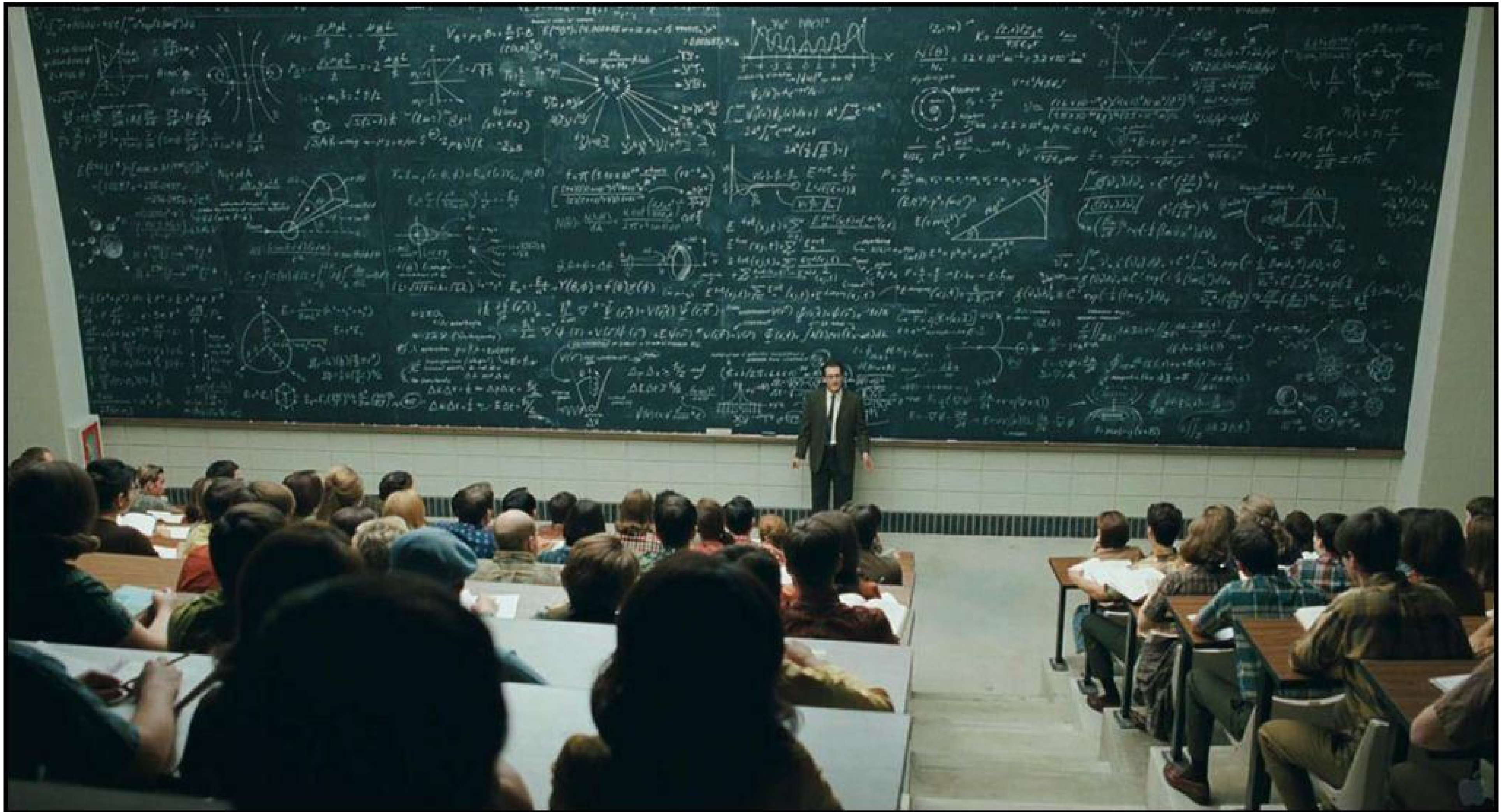
Other Applications

- Infectious disease trial endpoints
  – Clinical Cure
  – Microbiological Cure

- Oncology trial endpoints
  – Overall survival
  – Disease-free survival

- Coinfection / comorbidity trial endpoints
  – HIV-1 RNA
  – Kaposi's sarcoma progression

- Cardiovascular trial endpoints
  – Stroke or MI
  – Death

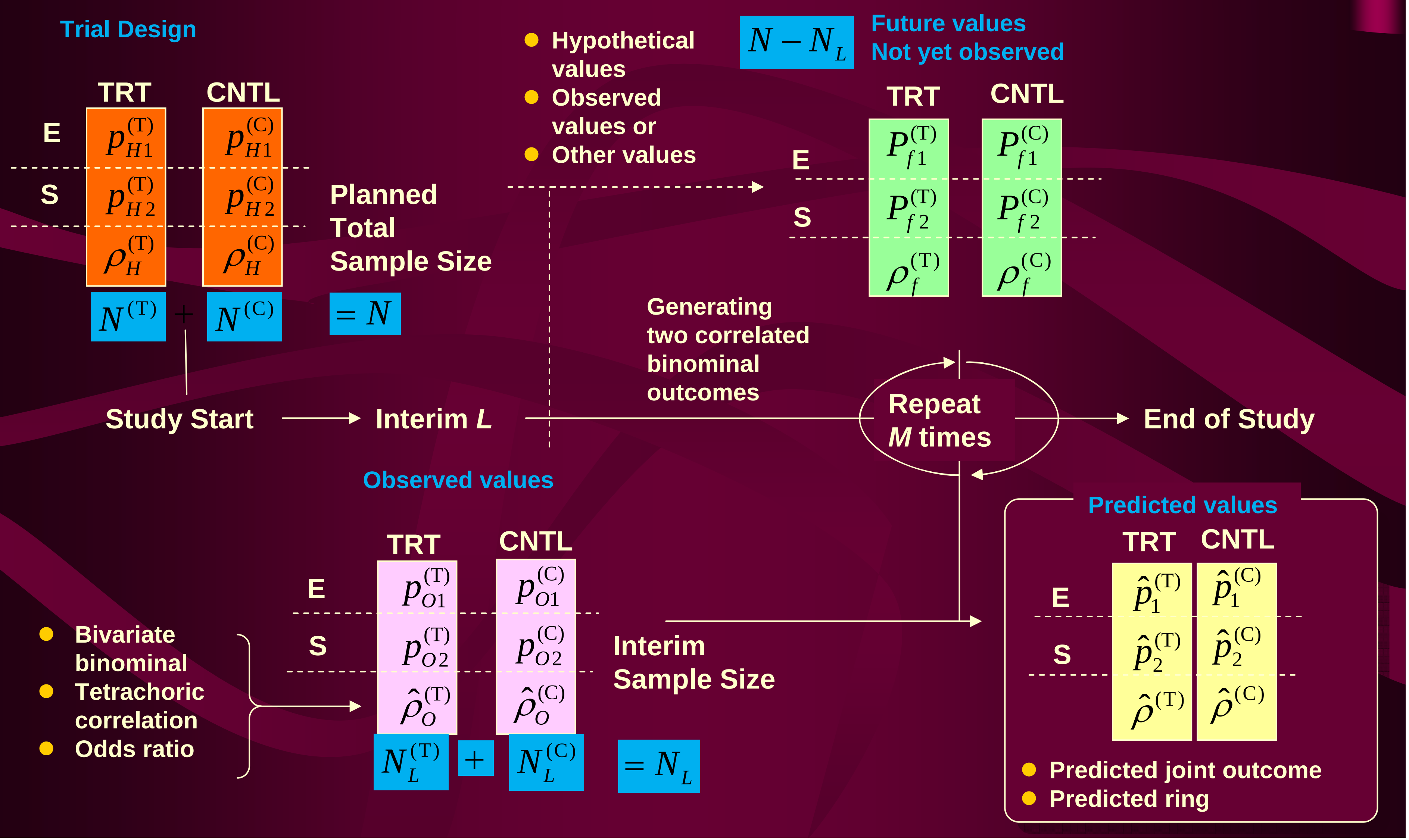


…and thus dear colleagues,
after this very elementary presentation,
I need your help…

What Should We Name This?

- Recall PIPs = Predicted Interval Plots
- Options
  - POPs = Predicted Oval Plots
  - PEPs = Predicted Ellipse Plots
  - PIGs = Predicted Interval Graphs

Predicted rings for binary risk differences/relative risks



Motivating Question

How do we revise our traditional approaches to design, monitoring, analyses, and reporting of trials to address the challenges of benefit:risk evaluation?