Experience with Bayesian Clinical Trial Designs for Cardiac Medical Devices

Andrew Mugglin

Statistics Manager

Cardiac Rhythm Management Clinical Research

Medtronic, Inc.

Acknowledgements

- … with thanks to the following people for their input at various stages
  - Lou Sherfesee (Medtronic)
  - Feng Tang (Medtronic)
  - Don Berry
  - Scott Berry
  - Amy Xia (formerly Medtronic)

Agenda

- Examples of Cardiac Medical Devices
- Distinctives of Medical Device Trials
- When to be Bayesian
- Case Study #1: Informative Prior
- Case Study #2: Adaptive Design
- Implications of being Bayesian

Electrical Circuitry of the Heart

Sinus Rhythm to Ventricular Fibrillation (VF) To Cardiac Arrest

A modern external defibrillator

An Implantable Defibrillator

Distinctives of Medical Device Trials

- Blinding is difficult
- Compliance to therapy is high
- Local impact, generally not systemic
- Technological Advancement vs Therapeutic Advancement
- Speed of obsolescence
- “Electricity is still electricity”
- Prior information from previous generations

When to be Bayesian

- Prior data/informative prior information
  - ➔ Can reduce size, length of a trial.
- Hierarchical modeling/random effects modeling
  - Borrow strength from prior or similar information
    - Data determine amount of borrowing
  - Can build complicated models from simple pieces (e.g. correlated data)
  - Meta-analysis
  - Multi-center trials

When to be Bayesian

- Sequential Analysis
  - Interim stopping
  - Sample size re-estimation
  - Adaptive designs
- Prediction
  - Predictive probability of future results
  - For adaptation of design, based on current outcomes
  - For modeling long-term outcomes based on short-term outcomes

Case Study 1: Next Generation Defibrillator

Case Study 1: A Next Generation Defibrillator

- Background
  - New generation of ICDs
  - Reduction in device size
  - Several enhanced algorithmic features
  - Data from two large PMA studies on predicate devices are available
  - Goals:
    - Assess overall safety of device
    - Assess enhanced detection algorithm performance

Motivation and Objectives

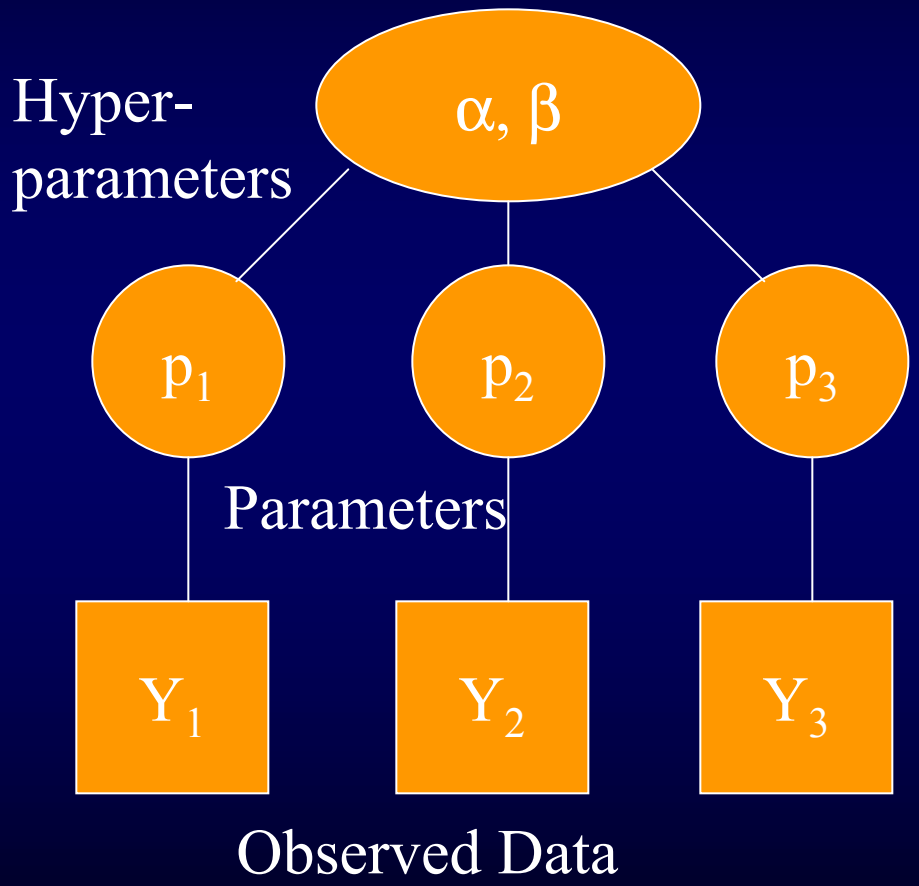
- Motivation for Bayesian Design
  - Use of prior information can increase understanding of device safety with smaller (prospective) sample size
  - Sequential analysis of data (n=105, n=200)

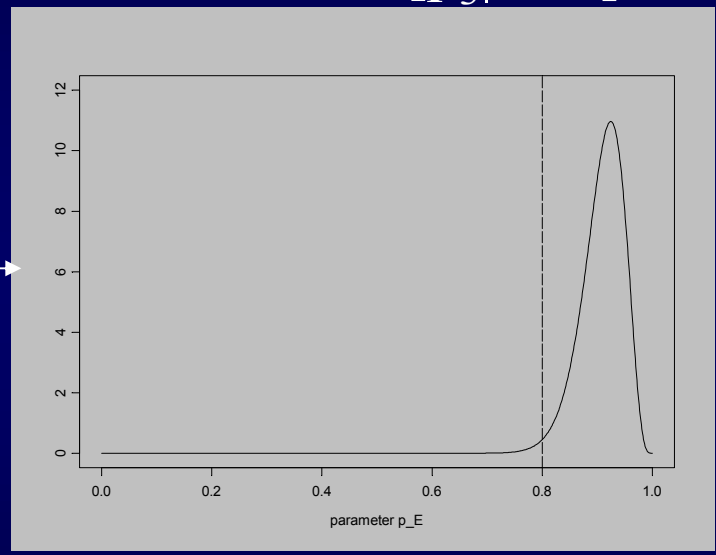
- Primary objectives
  - AE-free rate at 3 Months
  - Sensitivity of detection algorithm to ventricular arrhythmias

Hierarchical Model (AE rates objective)

Goal:

Posterior: $[p_3|data]$





Bayesian Random Effects Hierarchical Models

Likelihood: $Y_i \sim \text{Binomial}(n_i, p_i)$

$i = 1,2,3$ (trials, for AE-free rate)

or $i = 1, \ldots, n$ (patients within trial, for sensitivity)

Parameter of interest: p_3 (AE-free)

$\theta = \alpha/(\alpha+\beta)$ (sensitivity)

Prior: $p_i \sim \text{Beta}(\alpha, \beta)$

Hyperpriors: $\alpha \sim$ Vague, $\alpha \geq 1$

$\beta \sim$ Vague, $\beta \geq 1$

Posterior doesn't have closed form, resort to MCMC (WinBUGS)

To have a successful trial, must show:

- AE-free rate at 3 Months:

$$\Pr(P_3 \geq K_1 \mid \text{data}) \geq 0.99$$

(Prior is informative, based on historical data from 2 previous studies)

- Sensitivity

$$\Pr(\theta \geq K_2 \mid \text{data}) \geq 0.95$$

(Prior is essentially non-informative)

Design Considerations

- A logit formulation: $\text{logit}(p_i) \sim N(\mu,\sigma^2)$
  - Vague priors on μ, σ^2
  - Inference drawn on p_3 (AEs), μ (sensitivity)
  - More computationally stable except when $p \approx 1$ (or 0)
  - … which is exactly the case for sensitivity
  - This model not practical for sensitivity
- We opted to use the same model for each objective.
- No censoring mechanism (Binomial outcomes)
  - Required careful prespecification of how to include/exclude patients based on followup

Design Considerations (2)

- This is an uncontrolled trial.
- Standard is OPC (Objective Performance Criterion) K1, K2
- Posterior thresholds (e.g, 0.99, 0.95) are design choices
  - No consensus on value
  - Generally need to be high
  - Adjustable to influence frequentist operating characteristics
  - We opted to adjust these values (ad-hoc) rather than downweight prior information (ad-hoc)
- Watch out for “unethical trials”: Prior may be so informative that the criterion is met prior to trial start.

Sampling Plan

- When 105 patients have completed 3 mo:
  - Analyze AE rate (n=105)
  - Analyze sensitivity (n=200?)
- If both objectives are met, stop and submit
  - (frequentist power = 80% to stop at this point)
- If either objective is not met, continue until 200 patients have completed 3 mo
  - Repeat both analyses
- Frequentist OC were computed via simulation
  - Note: Definitions of α, β are tricky
  - Have to be acceptable to sponsor, FDA
- (overall power = 95% to reach successful conclusion)

If we had been frequentist… (fixed sample size design)

- N=135 for 80% power, α = 0.05
- N=282 for 95% power, α = 0.05
- N=263 for 80% power, α = 0.01
- N=380 for 95% power, α = 0.01

(Non-sequential designs. Trial is over at first analysis.)

Case Study #2: Novel Pacing Technique

Case Study #2: Novel Pacing Technique

Case Study #2: Novel Pacing Technique

- Goal: Compare Experimental (E) pacing technique with Control (C)
- Patient is blinded
- Endpoint: Composite of
  - Clinical Event (e.g., mortality)
    - Can happen anytime
  - Lab measure (e.g., ECG)
    - Detected at office visits
- Time-to-event analysis
- Little information on effect size or timing of effect
- ➔ Sample size?

Case Study #2: Novel Pacing Technique

- When you don't know what to do, let the data guide you.
- Motivation
  - Interim analysis
  - Sample size re-estimation
  - Predictive probability of eventual success
- Use predictive probability to guide sample size re-estimation and possible early stopping
- Primary objective: time-to-event analysis
- Many secondary objectives

Possible Outcomes and Desirable Corresponding Actions

- E clinically superior to C

  → Stop for success, when established or highly likely

- E clinically inferior to C

  → Stop for harm, when established

- E reasonably likely to be superior

  → Stop enrollment but continue followup

- E superior, but difference is not clinically meaningful

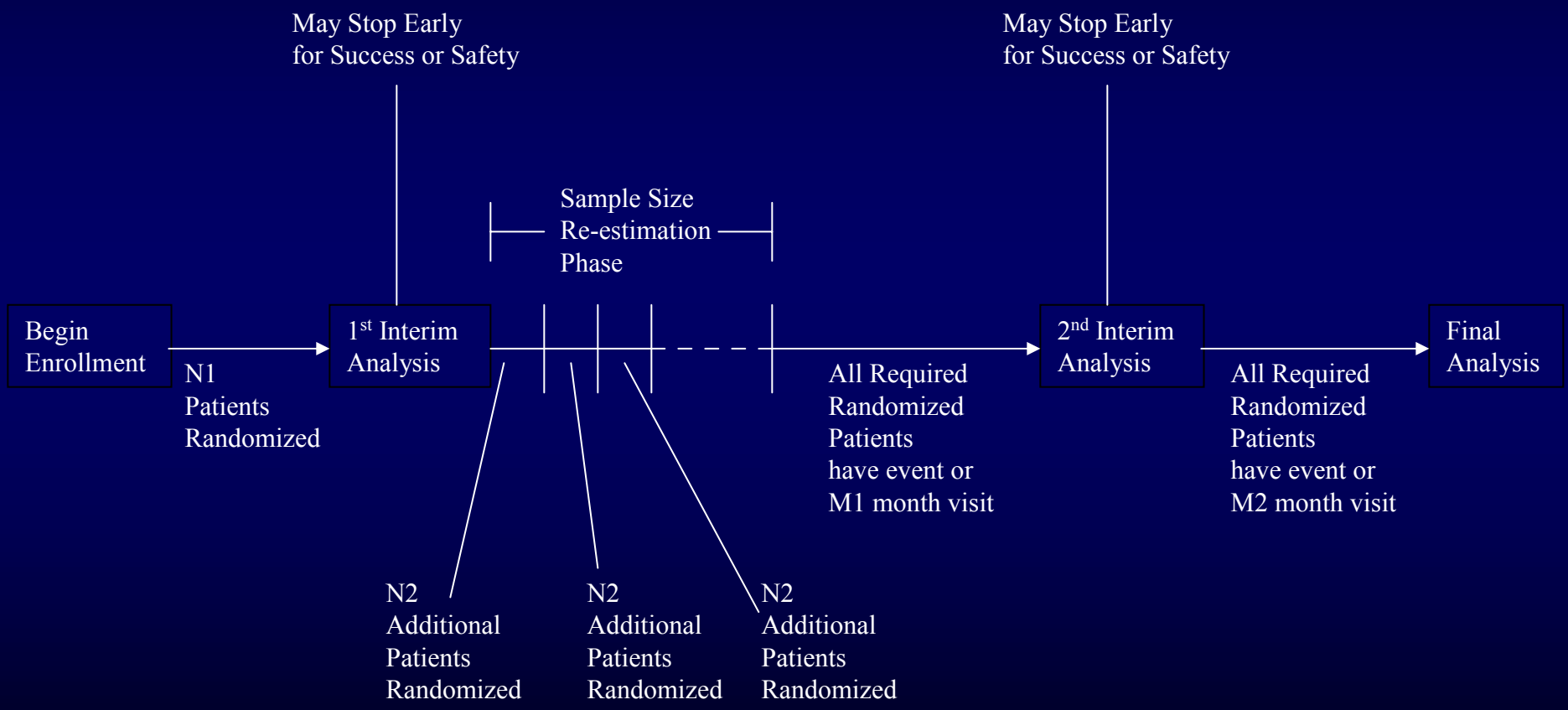
  → Stop enrollment but not the trial (when established)

- None of the above? (i.e., not enough information?)

  → Enroll more patients

  (up to a preset limit)

Study Design Timeline



The Model

- Proportional Hazards model:

  $\lambda_E(t) = \lambda_C(t)\, exp(\theta)$ for all $t > 0$

- $\lambda_C(t)$ piece-wise constant: (distribution of survival time is piece-wise exponential)

  $$\lambda_C(t) = \lambda_i, \; t \in I_i$$

- Pieces are chosen to reflect visit windows
  - (sensitivity analysis is planned)

Prior, Likelihood, Posterior

- Prior:

  $\lambda_{i,}$ ~ gamma(α,β)

  θ ~ N(0, σ^2)

- Likelihood:

  $L(\theta,\lambda_{1,},\ldots,\lambda_{k,} \mid data) \propto \prod_{i:ti=E} \prod_{j} (\lambda_j)^{dij} \exp(-\lambda_j x_{ij})$

  $\prod_{i:ti=C} \prod_{j} (\lambda_j \exp(\theta))^{dij} \exp(-\lambda_j x_{ij} \exp(\theta))$

  d_{ij} = 0: censored failure time for patient i in interval j

  d_{ij} = 1: observed failure time for patient i in interval j

- Posterior:

$$[\theta \mid \lambda_1,\ldots,\lambda_k, data] \propto \exp\left(-\frac{\theta^2}{2\sigma^2} + \sum_{i:t_i=2}^{k}\sum_{j=1}\left(d_{ij}\theta - \lambda_j x_{ij}\exp(\theta)\right)\right)$$ Metropolis-Hastings

$$[\lambda_j \mid \theta, data] \sim gamma\left(\alpha + \sum_{i=1}^{n} d_{ij}, \left[\beta^{-1} + \sum_{i:t_i=1} x_{ij} + \sum_{i:t_i=2} x_{ij}\exp(\theta)\right]^{-1}\right)$$ Closed form

Decision Quantities

- Posterior Probability:
  - P(θ > 0 | *current data*) = P(Harm)
  - P(θ < 0 | *current data)* = P(Benefit)
  - PBS = P(e^{θ} > 0.9 | *current data*) = evidence that Benefit is Small

- Posterior Predictive Probability of Eventual Study Success: PPES

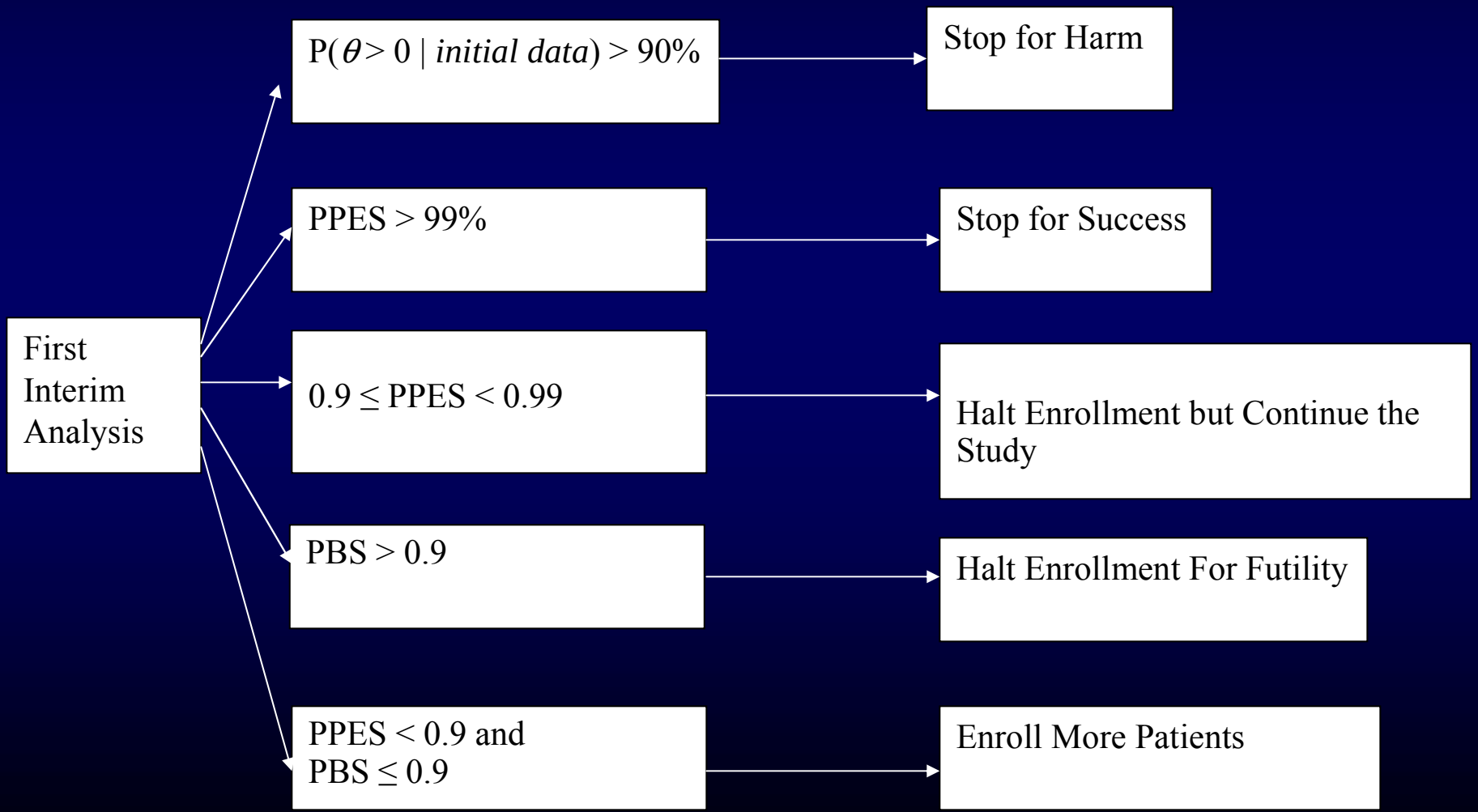
Denote eventual success as

γ ={ P(θ < 0| "data") > 0.981 },

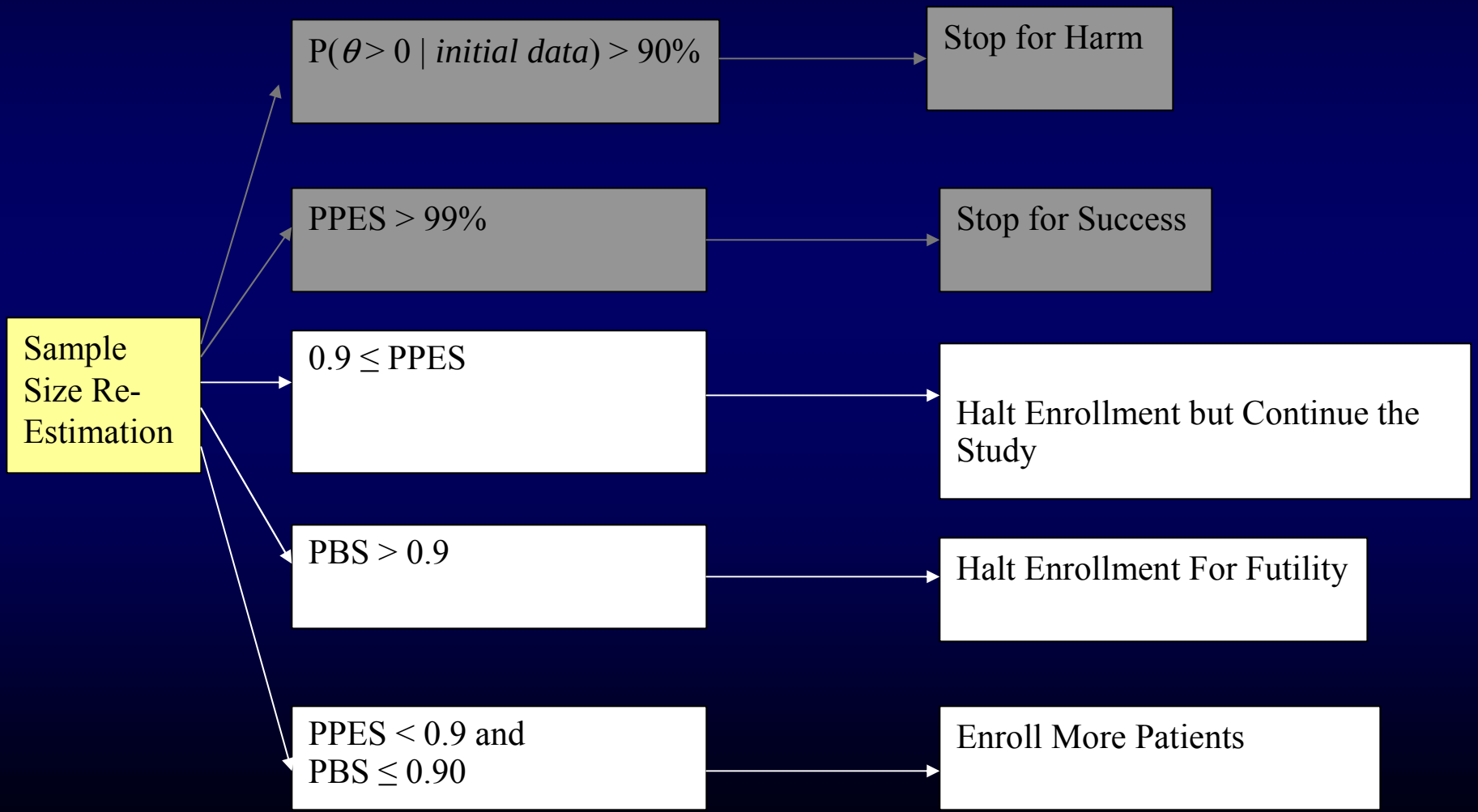
where "data" constructed by extrapolating, based on current evidence, to the time at which the last patient randomized has 12 months follow-up

PPES = P(γ)

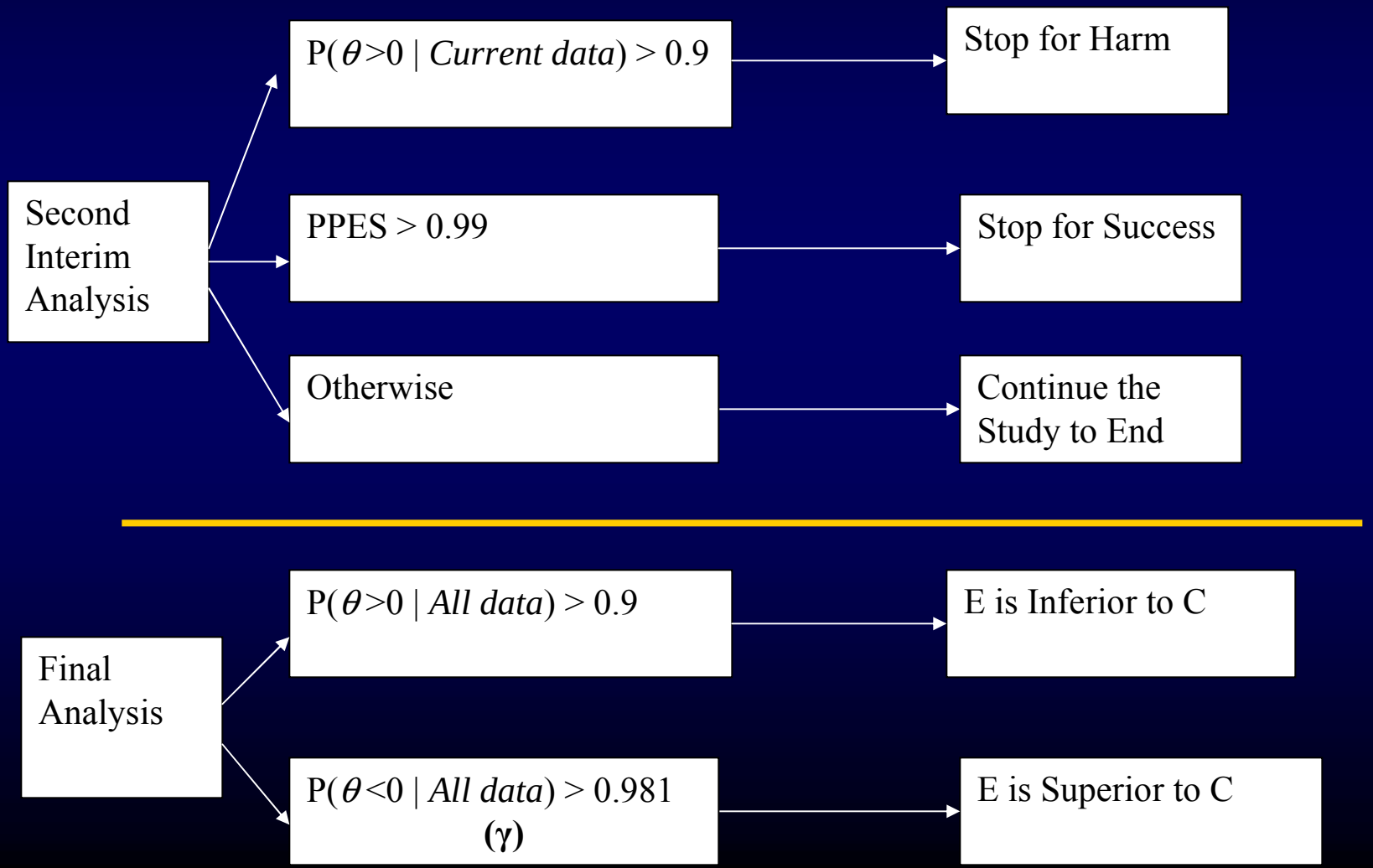
Criteria Used in First Interim Analysis



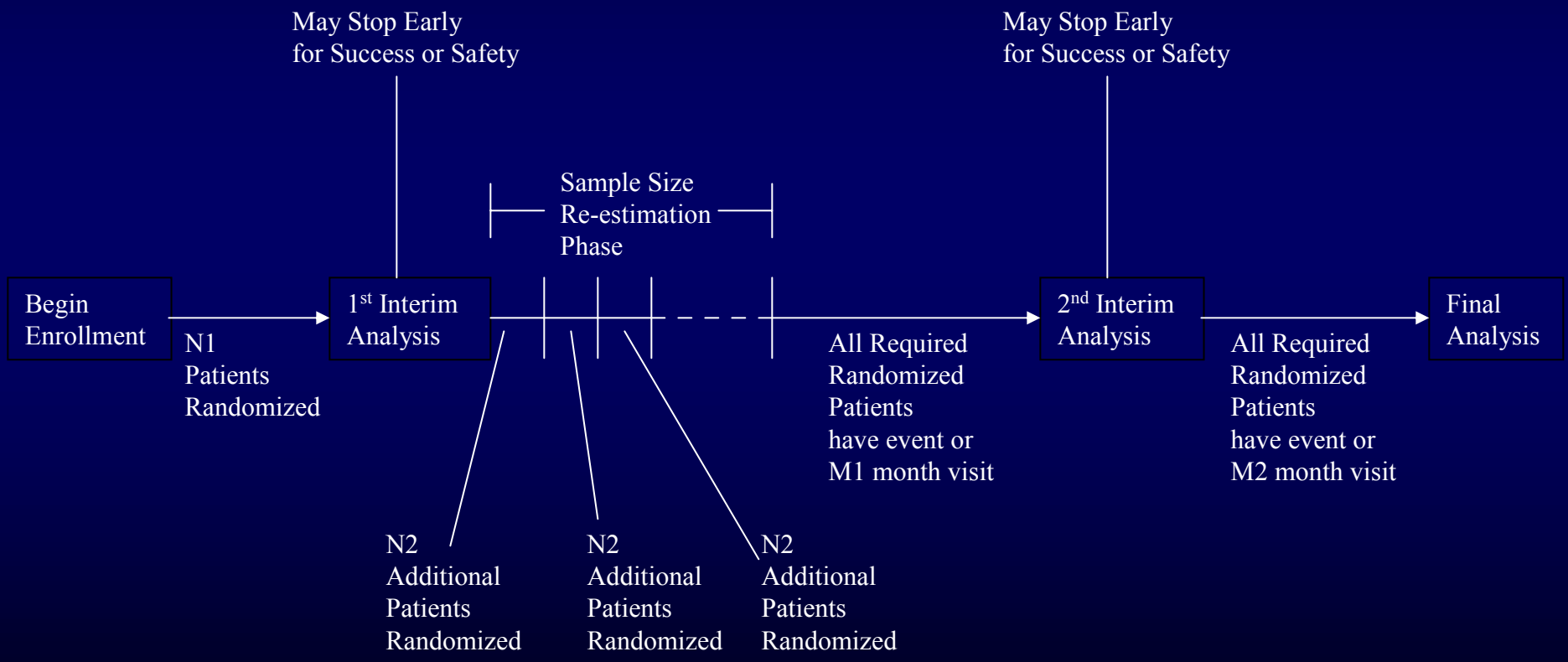
Criteria Used in First Interim Analysis, Sample Size Re-estimation Phase



Criteria Used in Second Interim And Final Analyses



Reminder: Study Design Timeline



Evaluating Frequentist Operating Characteristics

- Simulating different scenarios:
  - Harm: $\theta > 0$ (e.g., $\theta = 0.3$)
  - Null: $\theta = 0$
  - Benefit: $\theta < 0$ (e.g., $\theta = -0.2$)
  - Vary enrollment rates
- Calculating the probability of reaching “Harmful”, “Futile”, “Benefit” to evaluate the Frequentist OC
  - P(Conclude Benefit | $\theta < 0$) high → good power
  - P(Conclude benefit | $\theta = 0$) low → low type I error rate (< 0.025)
- Very computationally intensive
  - MCMC
  - Prediction of future during trial
  - Iterations to get long-run frequencies
  - Different scenarios
- ($\alpha < 0.025$)

Summary

- Examples of Cardiac Devices
- Distinctives of device trials
- When to be Bayesian
- Case Study #1: Informative Priors
- Case Study #2: Adaptive Design
- Implications of being Bayesian