Bayesian Statistics & Clinical Trial Disruption

Gary L. Rosner

Chenguang Wang & Marianna Zahurak

27 April 2021

JOHNS HOPKINS MEDICINE | THE SIDNEY KIMMEL COMPREHENSIVE CANCER CENTER

What Might a Bayesian Approach Help?

- Interim futility analysis, regardless of type of design
  - Predictive probability calculations
- Augmenting study data with external data
  - Propensity scores (Dr. Lilly Yue's presentation)
  - Marc Vandemeulebroecke
- Change in patient population after disruption
  - Drift in patient characteristics over time
- Sensitivity analyses

Questions in Light of a Disruption

- What might happen going forward?
  - Prediction ➡ Bayesian predictive dist'n:

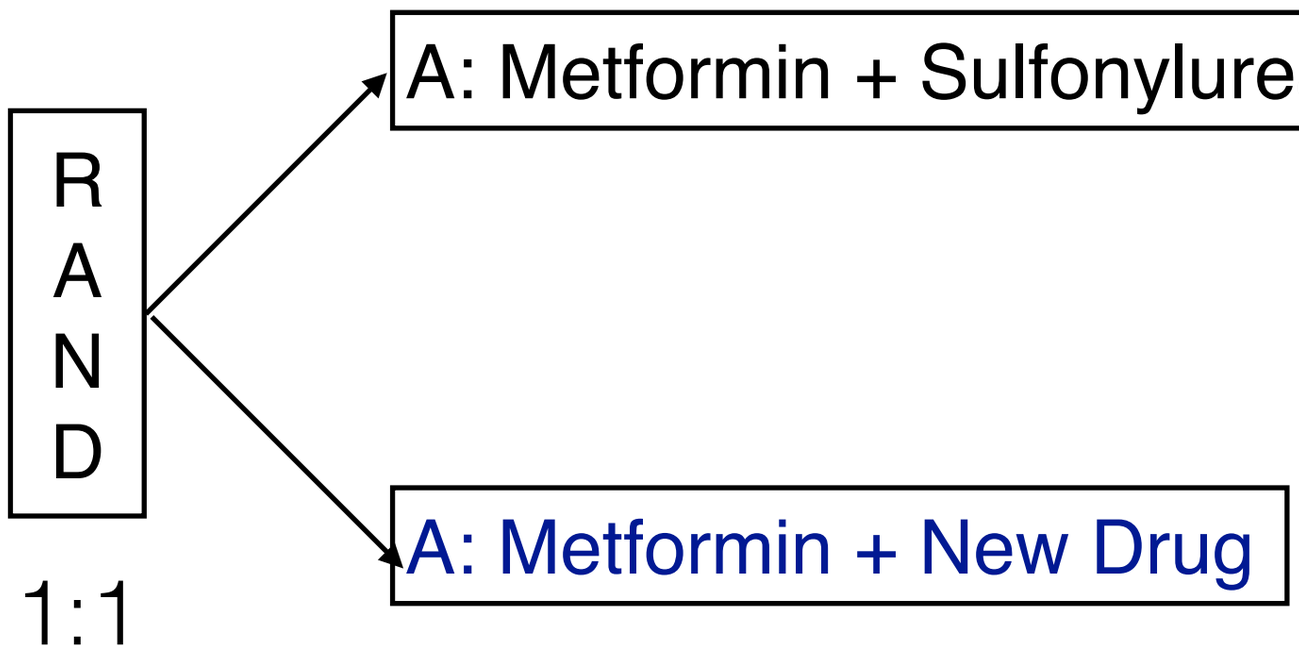
$$p(Y_{\text{new}} \mid Y_{\text{current}}) = \int p(Y_{\text{new}} \mid \theta) p(\theta \mid Y_{\text{current}}) d\theta$$

- What might have happened had there not been a disruption?
  - If condition on actual data, Bayesian predictive dist'n
    - Can allow for perturbation in treatment effect (sensitivity)
  - If condition on hypothesized parameter, frequentist

Unplanned Interim Analysis Regardless of Type of Design

- Questions
  - Should we continue the study?
    - What is the chance that the study will show a difference?
  - Even if a standard design based on Neyman-Pearson hypothesis test, can use Bayesian inference at interim
  - Examples for Bayesian inference at interim
    - Binary outcome (beta-binomial)
    - Continuous outcome (normal-normal)
    - Time-to-event (parametric or nonparametric)

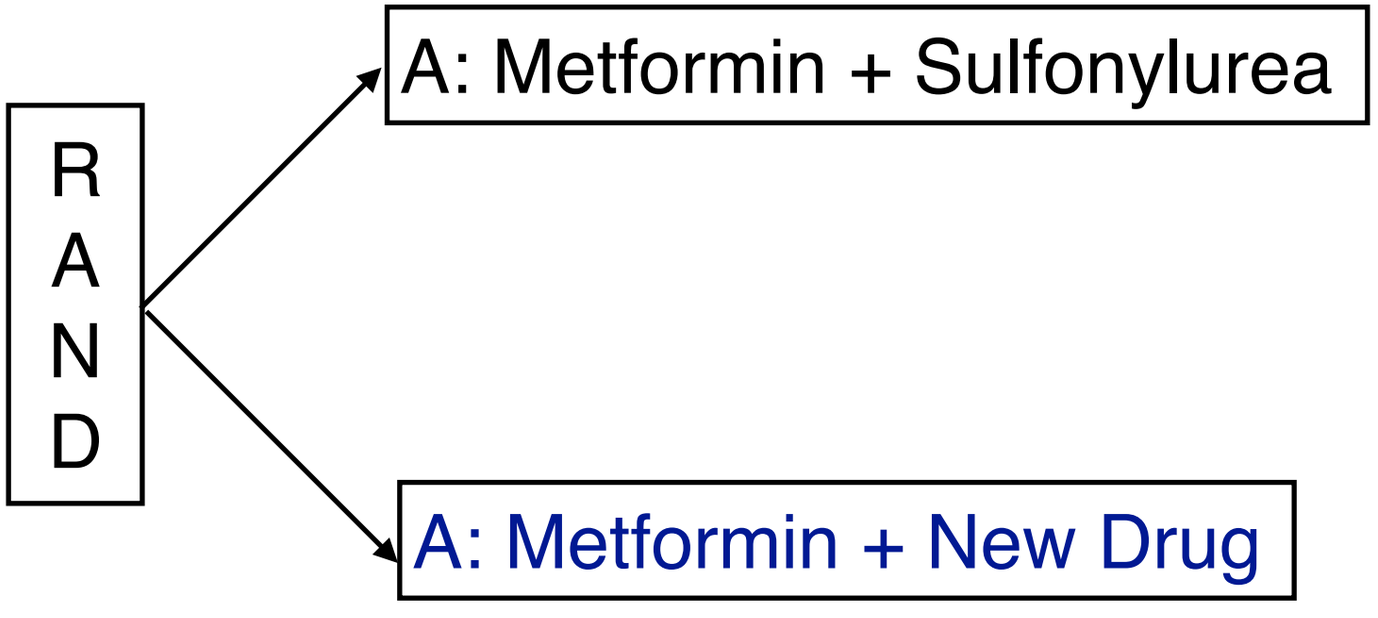
Example: RCT of 2nd-Line Therapy for Type 2 Diabetes



- Endpoint: $HbA1c_6$ - $HbA1c_0$
- Type I error prob (1-sided) = 0.025
- Power in protocol = 0.9
- Effect size = 0.2
- Assumed variance = 0.95
- Sample size = 500 / arm

Disruption

- Recruitment started October 2019
- Trial disrupted March 2020
  - Reached 300 patients / arm
  - No patients admitted since but follow-up has continued
  - 6-month endpoint recorded for all 300 patients (per arm)



Decisions for Disrupted Diabetes RCT

- Analyze current data: determine efficacy of adding New Drug?
- Re-start recruitment?
  - Perhaps change the target final sample size.
    - It is not clear when recruitment will re-commence and the time taken to reach the final decision is an important consideration.
      - Decision analysis (utility-based ⇒ Bayesian analysis)
- If data currently blinded
  - Develop procedure to allow “stopping now” or “continuing,” depending on results when current data unblinded

Choices for Interim Analyses

- Use Bayesian predictive distribution:

$$p(\text{Data}_{\text{post}} \mid \text{Data}_{\text{pre}}) = \int p(\text{Data}_{\text{post}} \mid \theta) p(\theta \mid \text{Data}_{\text{pre}}) d\theta$$

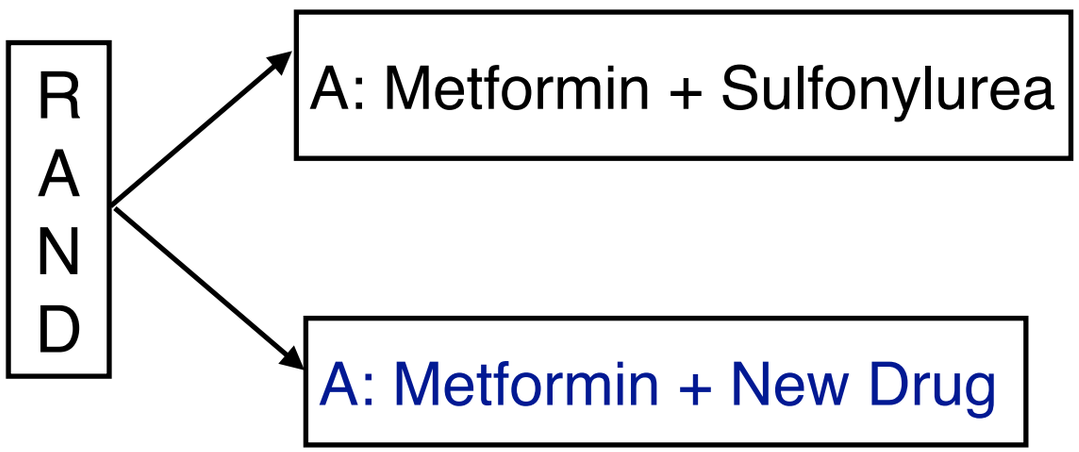
  - Set up enthusiastic and skeptical priors
    - Stop if enthusiast loses almost all enthusiasm
    - Continue if skeptic gains considerable enthusiasm
  - Zellner's g-prior ("objective" Bayesian approach)
    - Stop if low predictive probability ultimately reject
    - Continue if high predictive probability ultimately reject

Calibrate Prior for Δ to Achieve Desired Type I and Type II Error Probabilities

- Calibrate prior normal dist'n of Δ via alternative, $\Delta_{\text{alt}} = 0.2$
  - Skeptic's $p_{\text{skep}}(\Delta)$ & enthusiast's $p_{\text{enth}}(\Delta)$
    - Skeptic: $\Delta \sim N\left(0,\, \sigma^2_{\text{skep}}\right)$
      - $\Pr(\Delta > \Delta_{\text{alt}}) = 0.025 \;\Rightarrow \sigma^2_{\text{skep}} = (\Delta_{\text{alt}}/Z_{1-\alpha})^2$
    - Enthusiast: $\Delta \sim N\left(\Delta_{\text{alt}},\, \sigma^2_{\text{enth}}\right)$
      - $\Pr(\Delta < 0) = 0.1 \;\Rightarrow \sigma^2_{\text{enth}} = (-\Delta_{\text{alt}}/Z_{\beta})^2$
  - If set variances for frequentist level-α test and power $1-\beta$

Data at Disruption (Simulated with $p < 0.025$ with full 1000)

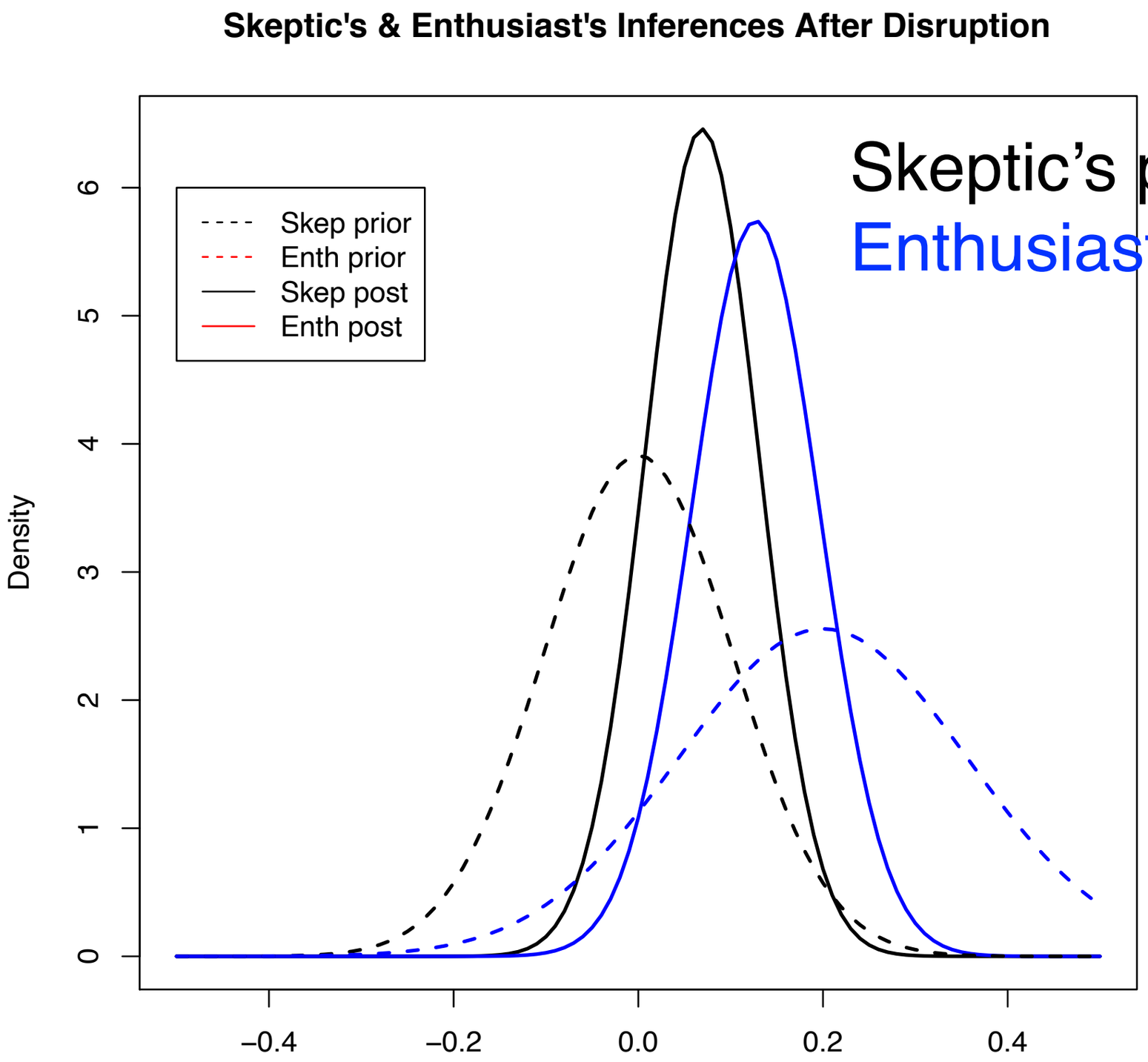
- Reached 300 patients / arm
  - 6-month endpoint recorded for all 300 patients (per arm)



	M + SU	M + ND
Mean	1.5112	1.6201
St. Error	0.0578	0.0518
n	300	300

$\hat{\Delta} = 0.109$ (st. error = 0.0776)
p-value = 0.081 (1-sided)

Skeptic's and Enthusiast's Beliefs Before and After Disruption



Skeptic's pred power = 0.30

Enthusiast's pred power = 0.53

Consider Usual Linear Regression

$$Y = X\beta + \varepsilon,\ \ \varepsilon \sim N(0, \sigma^2)$$

- A common objective prior: Unit Information Prior
  - Prior uses covariate info in data to specify prior $var(\beta)$
  - Ordinary least squares: $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$, with
    - $\hat{\beta}_{OLS} \sim N\left(\beta, \sigma^2(X'X)^{-1}\right)$ in frequentist presentation
  - With n obs'ns, information in sample is $(X'X)/\sigma^2$ (i.e., 1/var)
    - with respect to precision of $\hat{\beta}_{OLS}$
  - Divide by n gives information in 1 observation
  - Unit Information Prior: $\beta \mid \sigma^2 \sim N\left(b_0,\ n\sigma^2(X'X)^{-1}\right)$

Zellner's g-Prior

- Instead of dividing information by n, use arbitrary constant *g*
- Posterior dist'n of β (conditional on σ^2) has particularly simple form with *g*-prior (prior mean: $b_0 = 0$)

$$\beta \mid Y, X, \sigma^2 \sim N(m, V)$$

$$m = \frac{g}{g+1}(X'X)^{-1}X'Y = \frac{g}{g+1}\hat{\beta}_{OLS}$$

$$V = \frac{g}{g+1}\sigma^2(X'X)^{-1} = \frac{g}{g+1}var(\hat{\beta}_{OLS})$$

Variance with Zellner's *g*-Prior

- Complete prob. model with $\sigma^2 \sim$ Inv-gamma($\nu_0/2$, $\nu_0\sigma_0^2/2$)
- Marginal posterior is

$$\sigma^2 \mid Y, X \sim \text{Inv-gamma}\left(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + SSR_g}{2}\right),$$

- where SSR_g = sum of squared residuals with the *g*-prior

$$SSR_g = Y'\left[I - \frac{g}{g+1}X(X'X)^{-1}X'\right]Y$$

- Since using data pre-disruption, might set $\sigma_0^2 = \sigma^2_{\text{OLS}}$

Simulation to Predict Final Result: Generate Remaining Obs'ns & Covars

- Generate remaining observations and covariates at each iter
  - $\sigma^{2(j)} \sim p(\sigma^2 \mid \text{Data}_1)$
  - $\beta^{(j)} \sim p(\beta \mid \text{Data}_1, \sigma^{2(j)})$
  - $X_i^{(j)} \sim p(X \mid \text{Data}_1),\ i = n_1 + 1, \ldots, N$, possibly with other information
  - $Y_i^{(j)} \sim N\left(X_i^{(j)'}\beta^{(j)}, \sigma^{2(j)}\right),\ i = n_1 + 1, \ldots, N$
  - Combine $(Y_i^{(j)}, X_i^{(j)}),\ i = n_1 + 1, \ldots, N$ with Data_1
  - Regression w/ full data; Reject H_0 : trt effect $= 0$?

Model for Example Trial: Generate Study Data

- X_1 = Baseline HbA1c ($HbA1c_0$)
  - $X_1 \sim N(7.8,\ 0.4^2)$
- X_2 = 6-month HbA1c ($HbA1c_6$)
  - $X_2 = 7.0 + 0.5 * (X_1 - 7.0) - 1 + \epsilon_2$ if t=0 (metform+SU)
  - $X_2 = 7.0 + 0.5 * (X_1 - 7.0) - 1.2 + \epsilon_2$ if t=1 (metform+N)
    - $\epsilon_2 \sim N(0,\ 0.95)$
- $Y = X1 - X2$ $\quad E[Y] = 1.4 \text{ or } 1.6,\ var[Y] = 0.95,$ $cov[Y, X_1] = 0.08,\ corr[Y, X_1] = 0.205$

Operating Characteristics (Interim at 1/2): g=500, nSim = 2000, ν_0=4, σ_0^2=0.95

- Under null hypothesis (Stop if pred prob reject < 1/3 at interim)

```
Would stop early for futility 1847 times (92.3 %).
Rejected, 53 times (2.6 %).
```

- Under alternative, β_{trt}=0.2 (Stop if pred prob reject < 1/3)

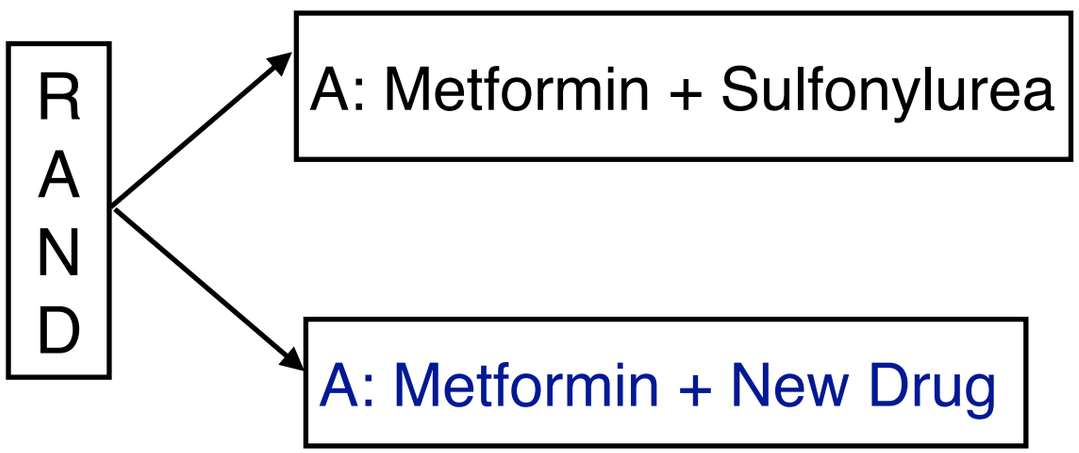
```
Would stop early for futility 231 times (11.6 %).
Rejected, 1712 times (85.6 %).
```

- Under alternative, β_{trt}=0.2 (No interim analysis)

```
Would stop early for futility 0 times (0 %).
Rejected, 1722 times (86.1 %).
```

Data at Disruption
Predicted Power with *g*-Prior

- Reached 300 patients / arm
  - 6-month endpoint recorded for all 300 patients (per arm)



	M + SU	M + ND
Mean	1.5112	1.6201
St. Error	0.0578	0.0518
n	300	300

Predicted power with *g*-prior = 0.38

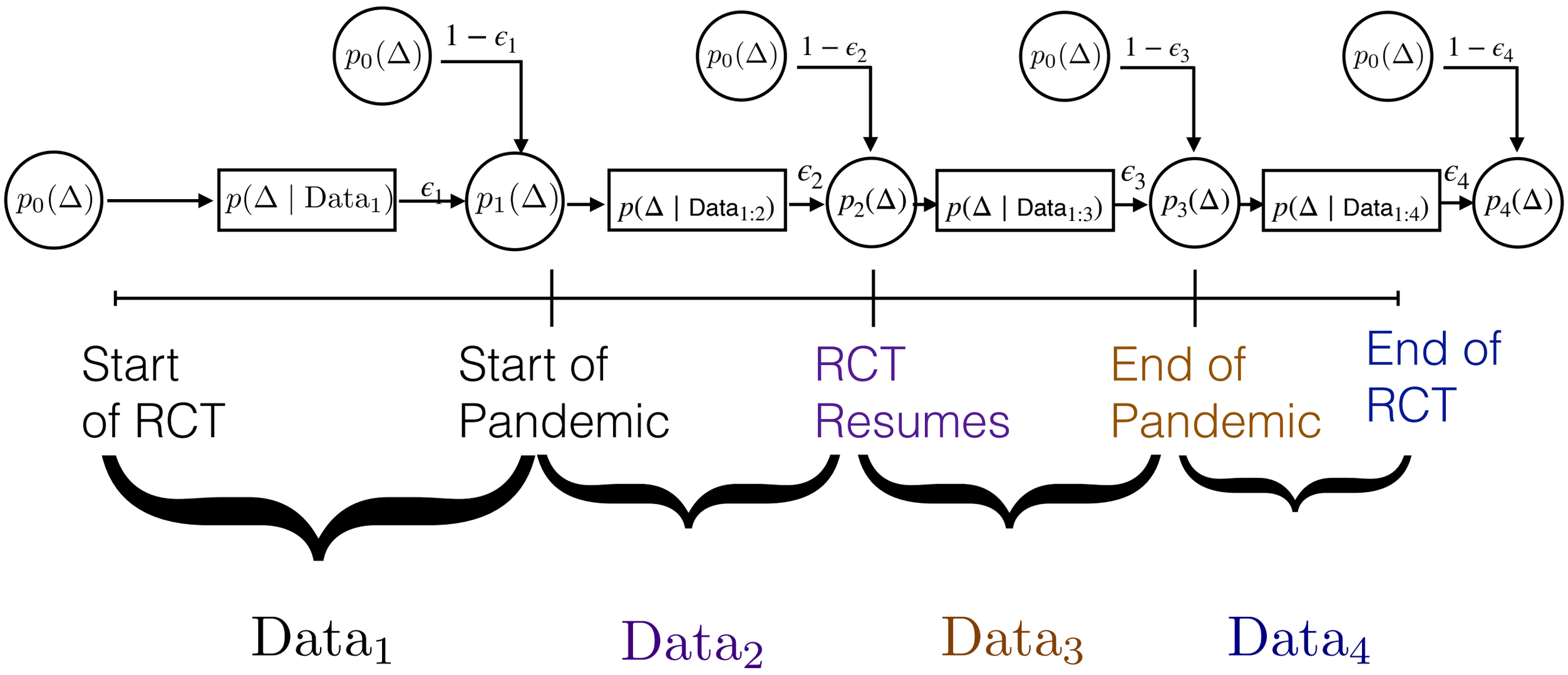
At interim: $\hat{\beta}_{trt}$ =0.099 (0.077), p-value = 0.099 (1-sided)

At final: $\hat{\beta}_{trt}$ =0.142 (0.060), p-value = 0.019 (1-sided)

Potential Lack of Exchangeability: Sensitivity Analysis for Change in Pt Pop

- Disruption might lead to sampling dist'n post-disruption ≠ pre
  - Might "adjust" via covariate (time interval)
  - Consider adapting robust meta-analysis predictive prior (RMAP) approach Schmidli et al. (2014)
    - $p_{\text{RMAP}}(\theta) = \epsilon p(\theta \mid \text{Data}) + (1-\epsilon)p_0(\theta)$
      - $\epsilon \in [0,1]$
      - $p_0(\theta)$ is a *vague* prior
    - Prior post-disruption is mixture of posterior from pre-disruption & $p_0(\theta)$

Schematic of Disruption



Might set $\epsilon_1 = 1$

Analysis at End: Control for Drift

- Patients after start of pandemic may be different from enrollees during (and after) the pandemic
  - Patients no longer fully exchangeable
  - Notation
    - $p_1(\Delta \mid \text{Data}_1)$: posterior of data pre-disruption,
    - $p_1(\Delta \mid \text{Data}_2)$: posterior with just data between start of disruption and restart of RCT
    - $p_1(\Delta \mid \text{Data}_3)$: posterior with just data between restart of RCT and end of pandemic
    - $p_1(\Delta \mid \text{Data}_4)$: posterior with data after end of pandemic

Summary

- Bayesian analyses at disruption useful for decision making
- No inflation of Type 1 error probability with futility analysis
- Can use Bayesian methods for sensitivity analyses
  - At interim and final analyses

Extra Slides

Bayesian Interim Futility Analysis: Binary Outcome

- Suppose RCT
  - $H_0 : \Delta = 0$ vs. $H_A : \Delta \neq 0$, where $\Delta = p_T - p_C$
    - $\alpha = 0.05$ (1-sided), $1 - \beta = 0.9$ for $\Delta = \Delta_a$ $(\Delta_a > 0)$
  - Consider *skeptical* and *enthusiastic* priors a la Spiegelhalter & Freedman
    - Enthusiast's prior: $p_{\text{enth}}(\Delta \geq \Delta_a) = 0.05$
    - Skeptic: $p_{\text{skep}}(\Delta \geq \Delta_a) = 0.9$ & $p_{\text{skep}}(\Delta \leq 0) = 0.1$
  - Stop for futility if $p_{\text{skep}}(\Delta \geq \Delta_a \mid \text{Data}) \leq \gamma$, for small γ

Binary Outcome Example

- Cancer trial of an anti-angiogenic agent vs. placebo
  - Target sample size = 100 patients, 1:1 randomization
  - 80% power for alternative hypothesis:
    - Risk of progression decreases from 94% (placebo) to 72% (new agent) by 16 weeks after randomization

Trt arm	CR/PR	SD	PD
Placebo	0%	6%	94%
CAI	0%	28%	72%

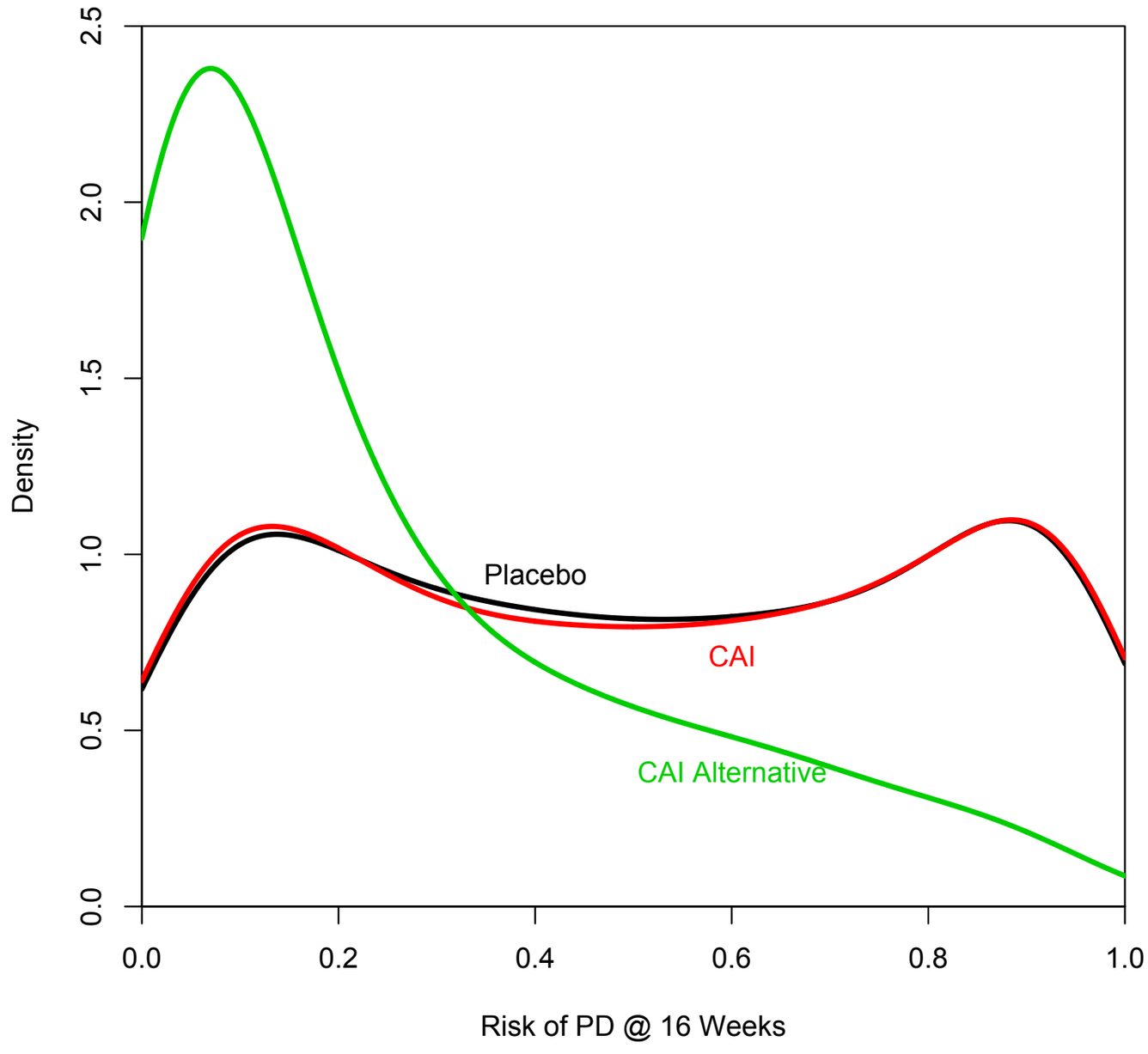
Alternative in protocol

Priors for Cancer Trial

- Accrual slowed after 49 patients. Continue?
  - Odds ratio of PD (placebo vs CAI) ≈ 6
    - Odds of PD w/ placebo (94/6 = 15.7)
    - Odds of PD w/CAI (72/28 = 2.6)
  - Calibrated prior variance of trt effect parameter (log OR) to give around 80% predictive power *a priori*

Priors

- Normal distributions for log odds (placebo & CAI)
- Roughly uniform prior for placebo risk of PD
- Alternative risk for CAI has approx 80% power



Marginal Dist'n (a.k.a. Prior Predictive Dist'n)

	Reject & Conclude	Mean	St Dev	2.5 %'ile	50 %'ile	97.5 %'ile
Noninformative Prior: *Equivalent Risks*	Placebo PD* Risk > CAI PD Risk	0.071	0.258	0.00	0.00	1.00
	CAI PD Risk > Placebo PD Risk	0.074	0.261	0.00	0.00	1.00
Enthusiastic Prior: *Placebo Odds 6 > CAI*	Placebo PD Risk > CAI PD Risk	0.788	0.409	0.00	1.00	1.00
	CAI PD Risk > Placebo PD Risk	0.024	0.153	0.00	0.00	0.00

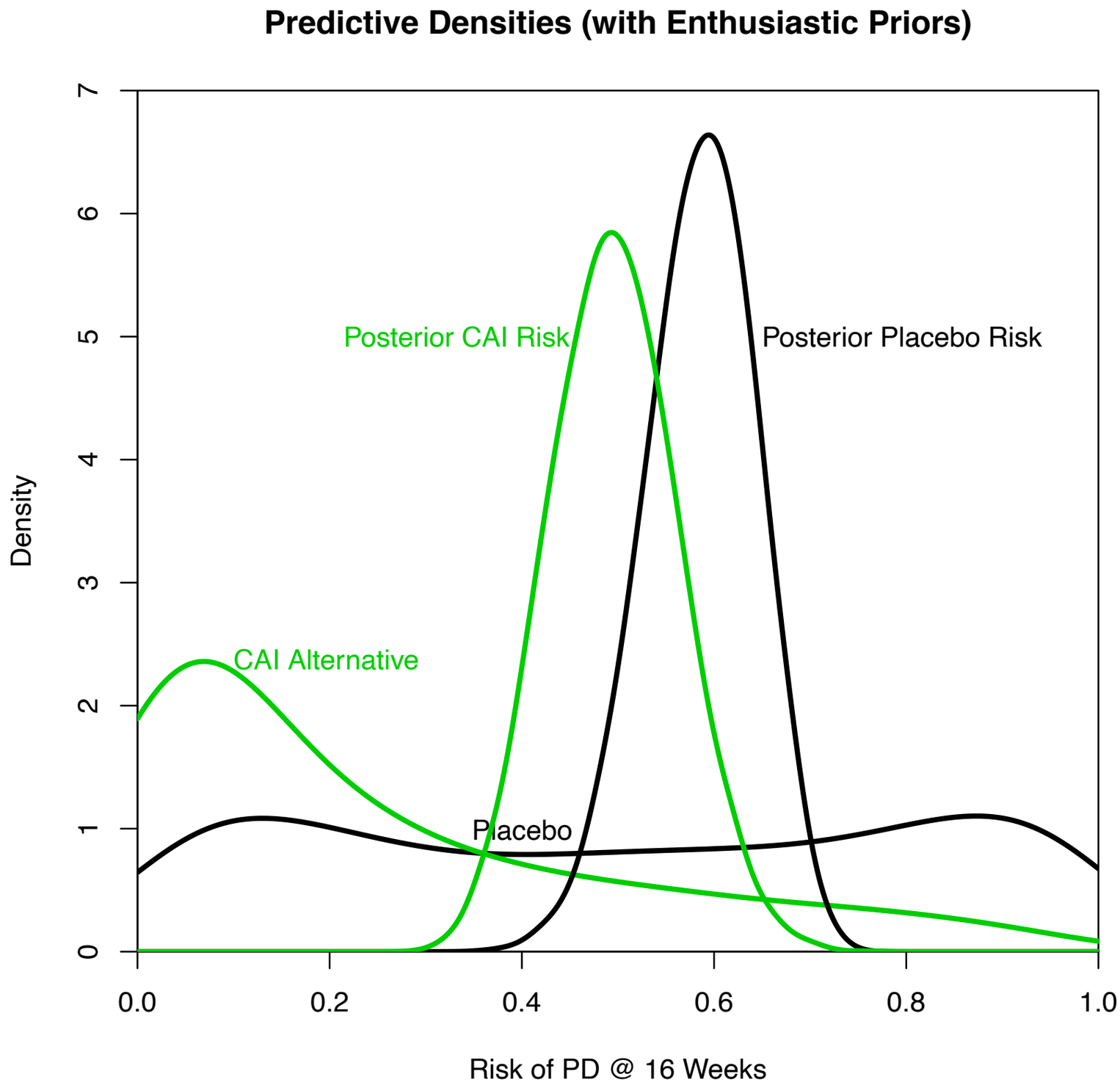
A priori probability final test stat in the rejection region after randomizing 100 patients: Results of 5000 simulations

Predictive Dist'n at Interim Analysis

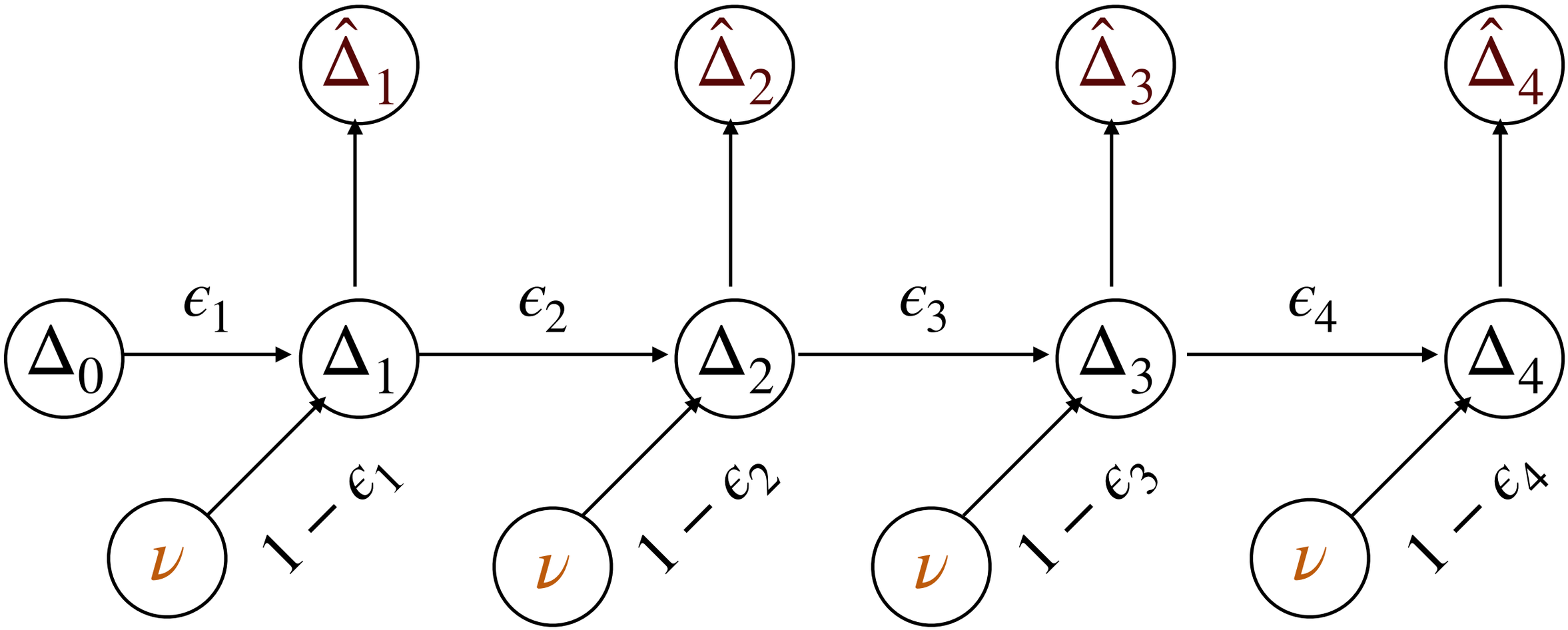
	Reject & Conclude	Mean	St Dev	2.5 %'ile	50 %'ile	97.5 %'ile
Noninformative Prior: *Equivalent Risks*	Placebo PD* Risk > CAI PD Risk	0.000	0.010	0.00	0.00	0.00
	CAI PD Risk > Placebo PD Risk	0.065	0.247	0.00	0.00	1.00
Enthusiastic Prior: *Placebo Odds 6 > CAI*	Placebo PD Risk > CAI PD Risk	0.087	0.282	0.00	0.00	1.00
	CAI PD Risk > Placebo PD Risk	0.000	0.020	0.00	0.00	0.00

A posteriori probability of being in the rejection region after randomizing 51 new patients, given data from 49 current patients (2 parallel MCMC chains, 5000 iterations, 25000 burn-in)

Any Reason to Continue for Enthusiast?



Graphical Model for HMM (ignoring nuisance parameters)



$\nu \sim p_0(\nu)$, vague prior

Might set $\epsilon_1 = 1$