Bayesian Adaptive Methods for Clinical Trial Design and Analysis

or

What I Did On My Fall Semester Leave

Bradley P. Carlin

Division of Biostatistics
School of Public Health
University of Minnesota

March 25, 2009

Broad Overview

- Biostatisticians in the drug and medical device industries are increasingly faced with data that are:
  - highly multivariate, with many important predictors and response variables
  - temporally correlated (longitudinal, survival studies)
  - costly and difficult to obtain, but often with historical data on previous but similar drugs or devices
- Recently, the FDA Center for Devices has encouraged hierarchical Bayesian statistical approaches –
  - Methods are not terribly novel: Bayes (1763)!
  - But their practical application has only become feasible in the last decade or so due to advances in computing via Markov chain Monte Carlo (MCMC) methods and related WinBUGS and BRugs (BUGS within R) software

Bayesian advantages in clinical trials research

- Probabilities of parameters: all unknowns have both prior (pre-data) and posterior (post-data) probability distributions, permitting direct statements about the probability of efficacy, toxicity, and so on at any time during a trial.
- Using all available evidence: from previous data, expert opinion, known structural relationships, etc. Also, via hierarchical modeling it is easy to borrow estimative power across similar but independent experiments (metaanalysis)
- Flexibility: Strictly speaking, frequentist measures require a complete experiment, carried out according to a prespecified design. BUT:
  - Bayesian inferences can be updated continually as data accumulate, and are not tied to the design chosen; in particular, the sample size need not be chosen in advance. Deviations from the original plan are possible, and we can stop for any reason we like (safety, futility, efficacy, etc.)

Bayesian advantages in clinical trials research

- Role of randomization: it minimizes the possibility of selection bias, and it tends to balance the treatment groups over covariates, both known and unknown. BUT:
  - Frequentist: also serves as the basis for inference
  - Bayesian: randomization not essential for inference!
- Predictive probabilities:
  - Frequentist: probabilities of future observations are possible only by conditioning on particular values of the parameters
  - Bayesian: average these probabilities over unknown parameters (unconditional probability is the expected value of conditional probabilities)
- Decision making: Bayesian approach is readily tailored to decision problems, e.g.,
  - designing the trial, or drawing a conclusion from it
  - allocating resources among R&D projects
  - when to stop device or drug development

  Each has costs and benefits, naturally weighted by Bayes. But the frequentist approach is poorly suited here.

Specific examples of Bayes in drug/device settings

- Safety/efficacy studies: Historical data and/or information from published literature can be used to reduce sample size, reducing time and expense. Unlimited looks at accumulating data are also permitted (due to different framework for testing).
- Equivalence studies: Bayes allows one to make direct statements about the probability that one drug is equivalent to another, rather than merely "failing to reject" the hypothesis of no difference.
- Meta-analysis: Bayes facilitates combining disparate but similar studies of a common drug or device.
- Hierarchical models: Realistic models can be fit to complicated, multilevel data (e.g., multiple observations per patient, or multiple patients per clinical site), accounting for all sources of uncertainty.

Brad's Fall Semester Leave Site

The University of Texas M.D. Anderson Cancer Center (MDACC)!

- Created by the Texas state legislature in 1941
- Largest component of the UT system
- One of the first 3 NCI Comprehensive Cancer Centers
- more than 25 buildings and 17,000 employees; nearly 1,400 faculty and more than 1,600 volunteers
- ranks first in number of NCI grants and total NCI grant dollars

MDACC is a tremendous resource for clinical research:

- Hundreds of clinical trials for every cancer
- Nearly 800,000 cancer patients registered since 1944
- statistics for most recent reporting year:
  - More than 79,000 persons with cancer received care
  - approximately 27,000 new patients
  - more than 11,500 patients in therapeutic clinical studies

Largest such program in the nation

Brad's Fall Semester Leave Site (more specific)

Division of Quantitative Sciences, MDACC

- 36 full-time and part-time faculty in two departments:
  - Department of Biostatistics (23 faculty)
  - Department of Bioinformatics and Computational Biology (13)

  Lots of prominent Bayesians: Don Berry (chair since 1999), Peter Thall, Peter Müller, J. Jack Lee, Val Johnson, Gary Rosner, Nebi Bekele, Guosheng Yin, Jeff Morris, Ying Yuan, Veera Baladandayuthapani, Yuan (formerly Steven) Ji (MS 1999, Biostatistics, U of Minnesota!)

The Department of Biostatistics runs a lot of Bayesian studies: During March 2007 – Feb 2009, of 677 protocols reviewed:

- 244/677 (36%) were designed by department faculty
- 89/244 (36%) were Bayesian

So while most trials are still run traditionally, MDACC has the highest number and concentration of Bayesian trials on earth!

Bayesian design of experiments

- In traditional sample size formulae, one often plugs in a "best guess" or "smallest clinically significant difference" for $\theta \Rightarrow$ "Everyone is a Bayesian at the design stage." – T.A. Louis
- In practice, frequentist and Bayesian outlooks arise:
  - Applicants may have a more Bayesian outlook:
    - to take advantage of historical data or expert opinion (and possibly stop the trial sooner), or
    - to "peek" at the accumulating data without affecting their ability to analyze it later
  - Regulatory agencies may appreciate this, but also retain many elements of frequentist thinking:
    - to ensure that in the long run they will only rarely approve a useless or harmful product, or expose patients to unacceptable levels of risk

---

Bayesian applicants must thus design their trials accordingly; i.e., ensure their designs have good frequentist operating characteristics!

Bayesian clinical trial design

Example: an N-patient safety study, in which we must show θ, the probability of freedom from severe device-related adverse events at 3 months, has 95% lower confidence bound at least 0.85.

- To find optimal Bayesian design in R:
  - For $j = 1, \ldots, Nrep$, draw θ_j from the design prior, followed by X_j from the corresponding $Bin(N, \theta_j)$ likelihood
  - Estimate the posterior under the analysis (or fitting) prior, perhaps by calling BUGS from within R if necessary
  - Check to see if the 2.5% point of the estimated posterior is in fact greater than 0.85.
  - The observed proportion of times this happens is the "Bayesian power"!
- Repeat this over several possible sample sizes N, and several priors. This then produces the "Bayesian sample size table"!
- Note there is no need to use the same prior at the design and analysis stages; the latter is typically more conservative.

Type I error rate calculation

- The design calculations on the previous slide are Bayesian because θ_j is being sampled from a design prior
- Frequentist operating characteristics arise by fixing θ at some "true" value. For example, to find Type I error, fix $\theta = 0$, and generate only the X_j for each of the *Nrep* iterations.
- Note that while Bayesians are free to look at their data at any time without affecting the inference, multiple looks will alter the frequentist Type I error behavior of the procedure. If this is of interest, the algorithm must be modified to explicitly include these multiple looks, checking for early stopping after each look.
- Note also that making the stopping rule adaptive (say, increasing allocation to the treatment that is winning) or multi-purpose (say, for futility or efficacy) do not materially complicate these calculations

Using Historical Data

- Email 8 Sep 2008 from Dr. Telba Irony, FDA: "When we try to borrow strength from only one historical study (be it a control group or a treatment group) ... [the results] become VERY sensitive to the hyperprior [on the variance parameters that control the amount of borrowing]."
- Borrowing from historical data offers advantages:
  - reduced sample size (at least in control group) hence lower cost and ethical hazard, plus higher power

  but also disadvantages:
  - higher Type I error, plus a possibly lengthier trial if the informative prior turns out to be wrong
- Thus what is needed is a recipe for how much strength to borrow from the historical data
  - One possibility: "back out" this amount based on Type I error and power considerations. This is often done, but tends to defeat the historical data's original purpose!

Proposed solution: Power Priors

Introduced by Ibrahim and Chen (2000, *Statistical Science*)

- IC (2000) define *historical data* as "data arising from previous similar studies"
- Let $D_0 = (n_0, \mathbf{x_0})$ denote historical data and suppose θ is the inference parameter
- Suppose $\pi_0(\theta)$ is the prior distribution on θ *before* D_0 is observed, the *initial prior*
- The *power prior* on θ for the current study is proportional to the initial prior times the historical data likelihood *raised to power* α_0, where $\alpha_0 \in [0, 1]$:

$$\pi(\theta|\alpha_0, D_0) \quad \propto \quad \pi_0(\theta)L(\theta|D_0)^{\alpha_0} \ ,$$

or just $L(\theta|D_0)^{\alpha_0}$ for flat initial priors (which are all we use).

Power Priors (cont'd)

- The *power parameter*, α_0, "can be interpreted as a relative precision parameter for the historical data" (IC, 2000, p.48)
  - certainly apparent if $x_{0i} \overset{iid}{\sim} N(\theta, \sigma_0^2), i = 1, \ldots, n_0$, since then under a flat initial prior we get a $N\left(\bar{x}_0, \sigma_0^2/(\alpha_0 n_0)\right)$ power prior for θ, with $\alpha_0 n_0$ "effective historical controls"
- As $\alpha_0 \to 1$, $q(\theta|D, D_0, \alpha_0)$ approaches *full borrowing*
- As $\alpha_0 \to 0$, $q(\theta|D, D_0, \alpha_0) \to q(\theta|D)$ hence *no borrowing*
- As such, we might use a *Beta*(a, b) hyperprior on α_0, to try to learn about α_0 rather than just fix it
- For given n, the amount of borrowing from the historical data depends on how consistent with the current data we think it is, with the "degree of consistency" controlled by the Beta hyperparameters, a and b

Hierarchical Power Priors

- Under the $Beta(a, b)$ hyperprior on the power parameter α_0, we obtain the joint posterior for θ and α_0 given D and D_0,

$$q(\theta, \alpha_0 | D_0, D) \propto \alpha_0^{a-1}(1-\alpha_0)^{b-1}\pi_0(\theta)L(\theta|D_0)^{\alpha_0}L(\theta|D) \; .$$

- Problem: Only D_0 (not D) can inform about α_0, and it can't say much: marginal posterior can be flat or multimodal
- Potential solution: Assume instead that $\alpha_0 \sim Bernoulli(p)$ and $p \sim Beta(c, d)$, so that now $\alpha_0 \in \{0, 1\}$ has a two-point mixture prior with mixing probability $p \in [0, 1]$. Now

$$q(\theta, \alpha_0, p | D_0, D) \propto p^{\alpha_0+c-1}(1-p)^{d-\alpha_0}\pi_0(\theta)L(\theta|D_0)^{\alpha_0}L(\theta|D) \; .$$

- Now borrowing is conceptually an "all or nothing" proposition, and could be made even more so by marginalizing out p as follows: If $\hat{p} = argmax_p \, q(p|D_0, D) > 1/2$, fix $\alpha_0 = 1$ and proceed with the full borrowing design; otherwise proceed with no borrowing ($\alpha_0 = 0$) — quasi-empirical Bayes!

Example: Power Prior Model for a One-Arm Trial

Suppose a pilot study suggests that a true treatment effect for a particular drug exists and is indicated by $\mu \neq 0$

Historical Data

- Suppose $\mathbf{x_0} = (x_{01}, ..., x_{0n_0}) \sim Normal(\mu, \sigma_0^2)$ i.i.d. where $D_0 = (\mathbf{x_0}, n_0, \sigma_0^2)$ and σ_0^2 is known
- $\pi_0(\mu) \propto 1$
- $q(\mu|D_0) = Normal(\bar{x_0}, \frac{\sigma_0^2}{n_0})$

Power Prior Model

- Suppose $\mathbf{x} = (x_1, ..., x_n) \sim Normal(\mu, \sigma^2)$ i.i.d. where $D = (\mathbf{x}, n, \sigma^2)$ and σ^2 is known
- $\pi(\mu|\alpha_0, D_0) = [q(\mu|D_0)]^{\alpha_0}$, where $\alpha_0 \in [0, 1]$
- $\pi(\alpha_0) = Beta(a, b)$

MCMC: Gibbs-Metropolis steps for μ and α_0

Power Prior Model Operating Characteristics

Compare frequentist properties of power prior models to *no borrowing* ($\alpha_0 = 0$) and *full borrowing* ($\alpha_0 = 1$) for testing the null hypothesis $H_0 : \mu = 0$ (no treatment effect).

- Define μ_0 be the true mean of the *historical data*
- Fix $\mu_0 = 2$ (trt historically effective), $n_0 = 30, \sigma_0 = \sigma = 1$
- Compute empirical probability of rejecting null hypothesis and covering μ for $\mu = (0, 0.25, 0.5, 1, 2)$ and $n = (1, 5, 10, 20, 30)$

Simulation procedure:

1. Given true μ, draw $x_{0i} \overset{iid}{\sim} N(\mu_0, \sigma_0^2)$ and $x_i \overset{iid}{\sim} N(\mu, \sigma^2)$
2. Run MCMC and compute the 95% marginal posterior credible interval for μ
3. Repeat $Nrep = 500$ times, recording the number of times 0 is excluded and μ is covered to obtain empirical probabilities

Power Prior Model Operating Characteristics

Power Parameter Priors

1. Uniform power parameter prior: $\pi(\alpha_0) = Beta(1, 1)$
2. Optimistic power parameter prior: $\pi(\alpha_0) = Beta(4, 1)$

Remarks

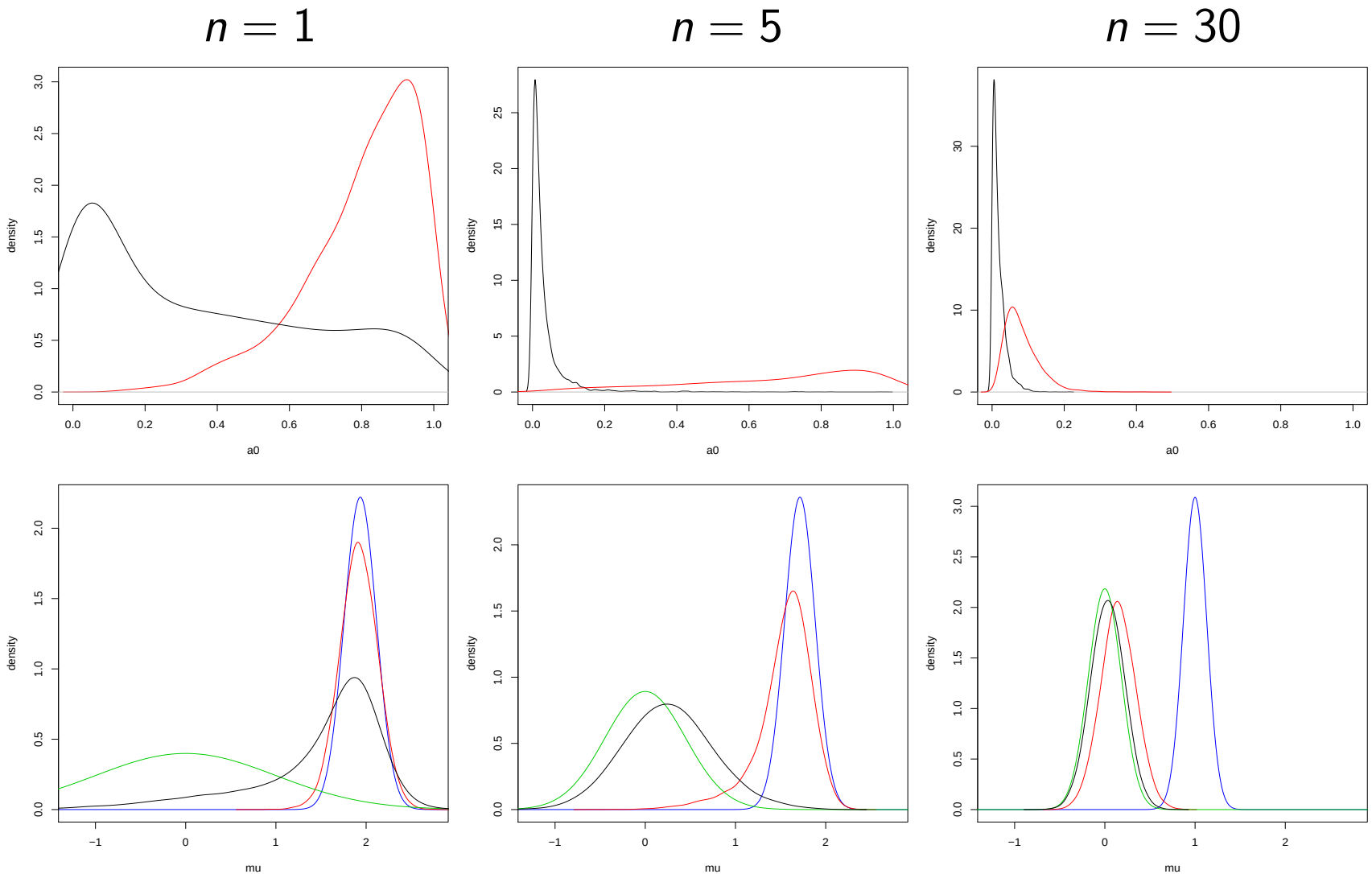
- Optimism with respect to exchangeability of historical and current data
- For $Beta(a, 1)$ power parameter prior, increasing a increases optimism
- Choose a relative to n and desired type I error rate. We can "afford" larger a (more power) by increasing n and/or decreasing the Type I error rate.

	n				
μ	1	5	10	20	30
0	(1.00,0.34)	(0.78,0.11)	(0.40,0.07)	(0.19,0.06)	(0.14,0.05)
	(0.05,1.00)	(0.05,1.00)	(0.05,1.00)	(0.05,1.00)	(0.05,1.00)
0.25	(1.00,0.45)	(0.90,0.23)	(0.67,0.20)	(0.60,0.31)	(0.59,0.37)
	(0.06,1.00)	(0.09,1.00)	(0.12,1.00)	(0.20,1.00)	(0.28,1.00)
0.50	(1.00,0.56)	(0.97,0.45)	(0.91,0.54)	(0.92,0.71)	(0.95,0.85)
	(0.08,1.00)	(0.20,1.00)	(0.35,1.00)	(0.61,1.00)	(0.78,1.00)
1	(1.00,0.74)	(1.00,0.82)	(1.00,0.95)	(1.00,1.00)	(1.00,1.00)
	(0.17,1.00)	(0.61,1.00)	(0.89,1.00)	(0.99,1.00)	(1.00,1.00)
2	(1.00,0.94)	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)
	(0.52,1.00)	(0.99,1.00)	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)

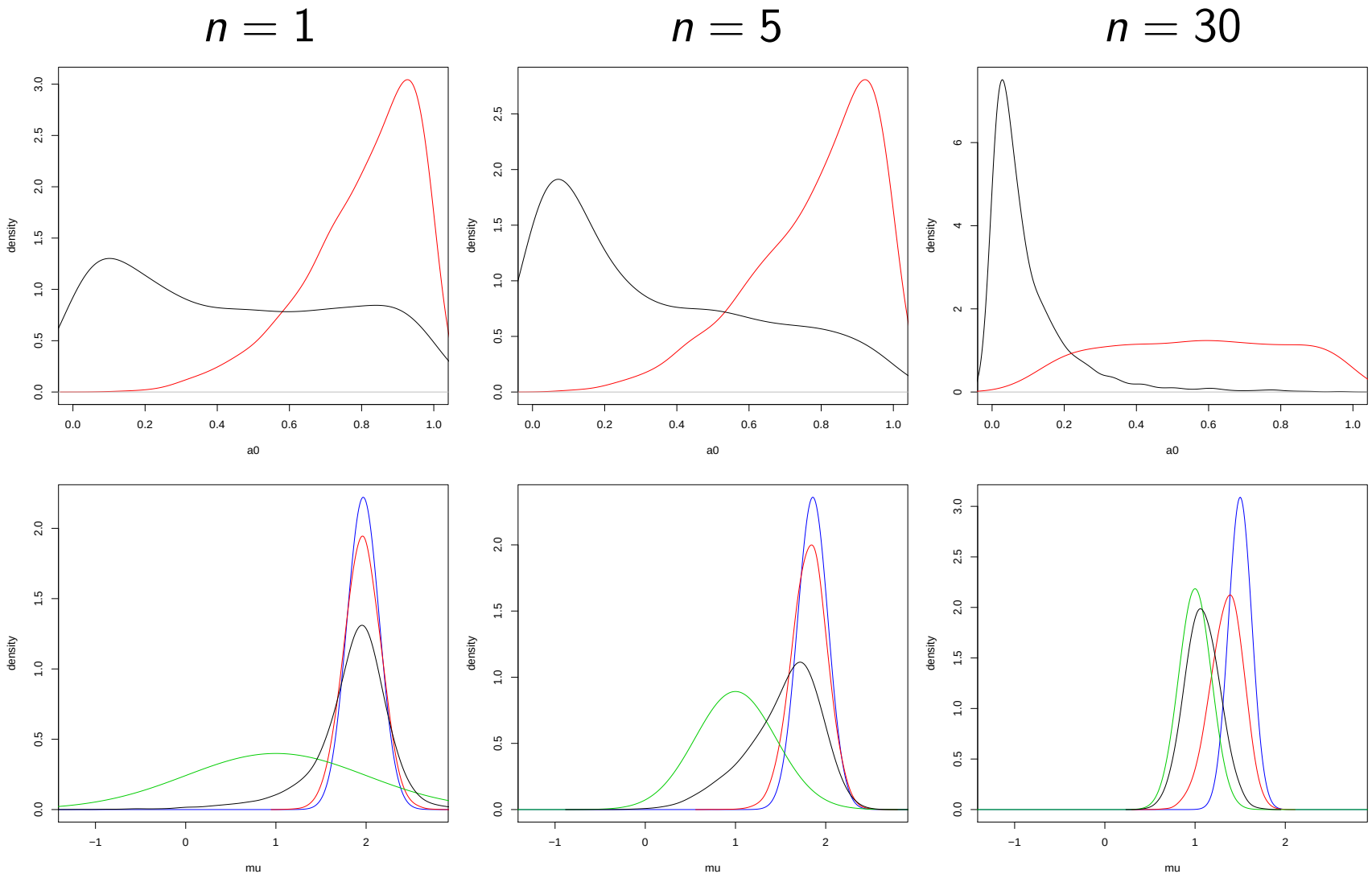
Empirical probabilities of rejecting the null hypothesis $\mu = 0$ for $\mu_0 = 2$, $n_0 = 30$, and $\sigma_0 = \sigma = 1$; color key: optimistic power parameter prior, uniform power parameter prior, no borrowing, and full borrowing.

	n				
μ	1	5	10	20	30
0	(0.00,0.66) (0.95,0.00)	(0.23,0.89) (0.95,0.00)	(0.60,0.94) (0.95,0.00)	(0.81,0.93) (0.95,0.00)	(0.87,0.93) (0.95,0.00)
0.25	(0.00,0.64) (0.95,0.00)	(0.15,0.88) (0.95,0.00)	(0.51,0.93) (0.95,0.00)	(0.75,0.94) (0.95,0.00)	(0.85,0.93) (0.95,0.00)
0.50	(0.00,0.65) (0.95,0.00)	(0.08,0.83) (0.95,0.00)	(0.35,0.90) (0.95,0.00)	(0.63,0.94) (0.95,0.00)	(0.77,0.94) (0.95,0.00)
1	(0.00,0.68) (0.95,0.00)	(0.04,0.80) (0.95,0.00)	(0.18,0.83) (0.95,0.00)	(0.45,0.89) (0.95,0.01)	(0.59,0.94) (0.95,0.03)
2	(0.95,0.95) (0.95,0.95)	(0.95,0.95) (0.95,0.95)	(0.95,0.95) (0.95,0.95)	(0.95,0.95) (0.95,0.95)	(0.95,0.95) (0.95,0.95)

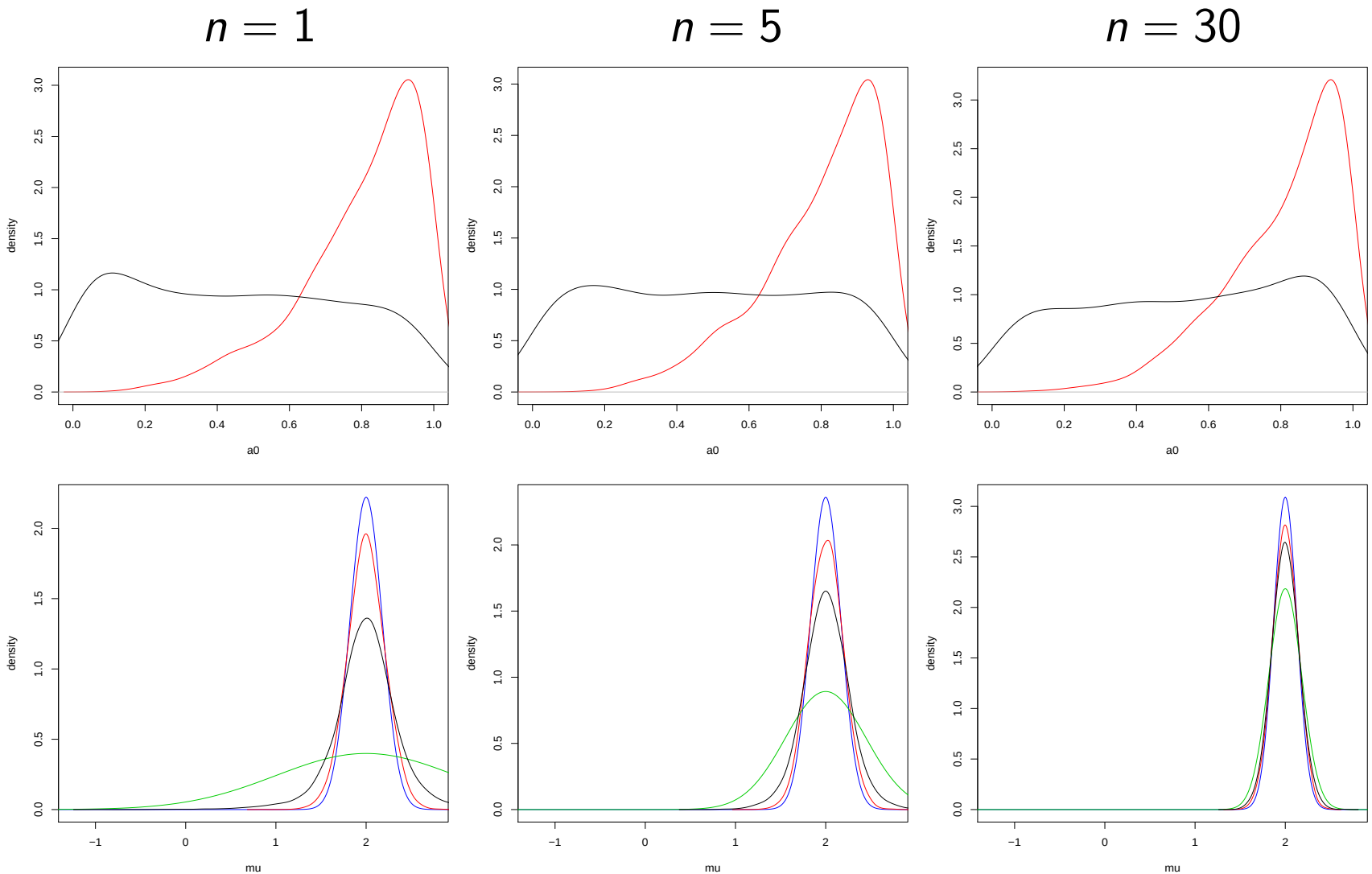
Empirical probabilities of 95% posterior credible interval covering μ for $\mu_0 = 2$, $n_0 = 30$, and $\sigma_0 = \sigma = 1$; color key: optimistic power parameter prior, uniform power parameter prior, no borrowing, and full borrowing.



Posteriors for α_0 (top) and μ for true $\mu = 0$, $\mu_0 = 2$, $n_0 = 30$, and $\sigma_0 = \sigma = 1$; color key: optimistic power parameter prior, uniform power parameter prior, no borrowing, and full borrowing.



Posteriors for α_0 (top) and μ for true $\mu = 1$, $\mu_0 = 2$, $n_0 = 30$, and $\sigma_0 = \sigma = 1$; color key: optimistic power parameter prior, uniform power parameter prior, no borrowing, and full borrowing.



Posteriors for α_0 (top) and μ for true $\mu = 2$, $\mu_0 = 2$, $n_0 = 30$, and $\sigma_0 = \sigma = 1$; color key: optimistic power parameter prior, uniform power parameter prior, no borrowing, and full borrowing.

Time Sensitive Power Prior Model

Model borrows disproportionately from historical data by imposing high influence on recent observations, but allowing influence to decrease with time. The power prior parameter, $\alpha_0(t_j)$, is now a function of t_j, the time passed since observing the *jth* outcome.

General Time Sensitive Power Prior Model

- Suppose $\mathbf{x} = (x_1, ..., x_n) \sim Normal(\mu, \sigma^2)$ i.i.d. where $D = (\mathbf{x}, n, \sigma^2)$ and σ^2 is known
- We now have n_0 distinct power parameters, $\alpha_0(t_1), ..., \alpha_0(t_{n_0})$
- The power prior in our Gaussian case is thus

$$\pi(\mu|D_0, \alpha_0(\mathbf{t})) = \left(\prod_{j=1}^{n_0} (2\pi\sigma_0^2)^{-\frac{\alpha_0(t_j)}{2}} \right) e^{-\frac{1}{2\sigma_0^2} \sum_{j=1}^{n_0} \alpha_0(t_j)(\mu - x_{0j})^2}$$

Time Sensitive Power Prior Model

Logistic Time Sensitive Power Prior Model

- Let $\mathit{logit}\left[\alpha_0(t_j)\right] = \beta_0 - \beta_1 t_j$
- Fix $\beta_0 = 3$ so $\alpha_0(0) = \frac{1}{1+e-3} \cong 1$
- Assume $\pi(\beta_1) = \mathit{Gamma}(c, d)$ (influence non-increasing in t)
- Choose c and d to achieve acceptable Type I error by enabling large β_1 when the historical and current data are inconsistent

Full conditional posterior distribution for $\mu|D, D_0, \beta_1$:

$$q(\mu|D, D_0, \beta_1) \propto e^{-\frac{1}{2\sigma_0^2}\sum_{j=1}^{n_0}\left(\frac{1}{1+e^{-(3-\beta_1 t_j)}}\right)(\mu - x_{0j})^2 + \frac{n}{\sigma^2}(\bar{x}-\mu)^2}$$

MCMC: Again, Gibbs-Metropolis steps

Adaptive Power Prior for Controlled Trial

Power prior models for controlled trials naturally advocate a randomization scheme that is "optimal" with respect to α_0 by defining the allocation ratio as function of the number of the "effective historical controls", $n_0\alpha_0$

- Let s_j and r_j denote the number of subjects randomized to new and control devices in the current trial after the *jth* enrollment
- Define η_j to be the proportion of "effective historical controls" and current controls after the *jth* enrollment

$$\eta_j = \frac{r_j + n_0\alpha_0}{s_j + r_j + n_0\alpha_0}$$

- Use median $q(\eta_j|D_j, D_{0j}, G_j, \alpha_{0j})$ as the probability that the $j+1^{st}$ subject is assigned to the new device, to encourage balance by imposing optimal use of new subjects relative to amount of incorporated prior information
- Update posterior in blocks after initial period using η_j=1/2

Randomized Controlled Colorectal Cancer Trial

Patients randomized (N=795) May 1999 to April 2001

1. irinotecan and bolus fluorouracil plus leucovorin (IFL) ($n = 264$) regulatory standard in March 2000
2. oxaliplatin and infused fluorouracil plus leucovorin (FOLFOX) ($n = 267$) new regimen
3. irinotecan and oxaliplatin (IROX) ($n = 264$) new regimen

Analysis with Power Prior Model

- Longest diameter (cm) of 1 to 9 tumors measured every 6 weeks for the first 42 weeks or until a response (death or disease progression)
- Compare IFL, IROX, and FOLFOX for median change in tumor sum from BL in patients with measurable tumors and at least two cycles ($N = 590$)
- IFL historical controls: patients enrolled before Oct. 2000

Analysis of Colorectal Cancer Data with Power Priors

Suppose (v_0, v, y_1, y_2) are responses for (historical IFL, current IFL, FOLFOX, IROX) and

- $v_{0i} \overset{iid}{\sim} Normal(\mu, \sigma^2), i = 1, \ldots, n_0 = 82$
- $v_i \overset{iid}{\sim} Normal(\mu, \sigma^2), i = 1, \ldots, n = 120$
- $y_{1i} \overset{iid}{\sim} Normal(\lambda_1, \tau_1^2), i = 1, \ldots, m_1 = 200$
- $y_{2i} \overset{iid}{\sim} Normal(\lambda_2, \tau_2^2), i = 1, \ldots, m_2 = 188$

Analysis of historical IFL, $\mathbf{v_0}$

- $\pi_0(\mu) \propto 1$
- $\pi_0(\sigma) = \text{Log-Normal}(-5, \sqrt{10})$
- $q(\mu|\mathbf{v_0}) = Normal(\bar{v}_0, \frac{\sigma^2}{n_0})$ (historical posterior)
- $q^*(\sigma|\mathbf{v_0}) \propto \frac{1}{\sigma}^{n_0+1} \exp\left(-\frac{1}{2}\frac{(v_0-\mu)^T(v_0-\mu)}{\sigma^2} + \frac{(\log\sigma+5)^2}{\sqrt{10}}\right)$

Analysis of Colorectal Cancer Data with Power Priors

Since FOLFOX and IROX are new regimens, let $\pi(\lambda_j) \propto 1$, $j = 1, 2$
Power Priors for current IFL, $(\mathbf{v})$, FOLFOX, $(\mathbf{y_1})$, and IROX, $(\mathbf{y_2})$

- $\pi(\mu|\alpha_0, \mathbf{v_0}) = [q(\mu|\mathbf{v_0})]^{\alpha_0}$
- $\pi(\sigma|\phi_0, \mathbf{v_0}) = [q^*(\sigma|\mathbf{v_0})]^{\phi_0}$
- $\pi(\tau_j|\omega_{0j}, \mathbf{v_0}) = [q^*(\tau_j|\mathbf{v_0})]^{\omega_{0j}}$, $j = 1, 2$

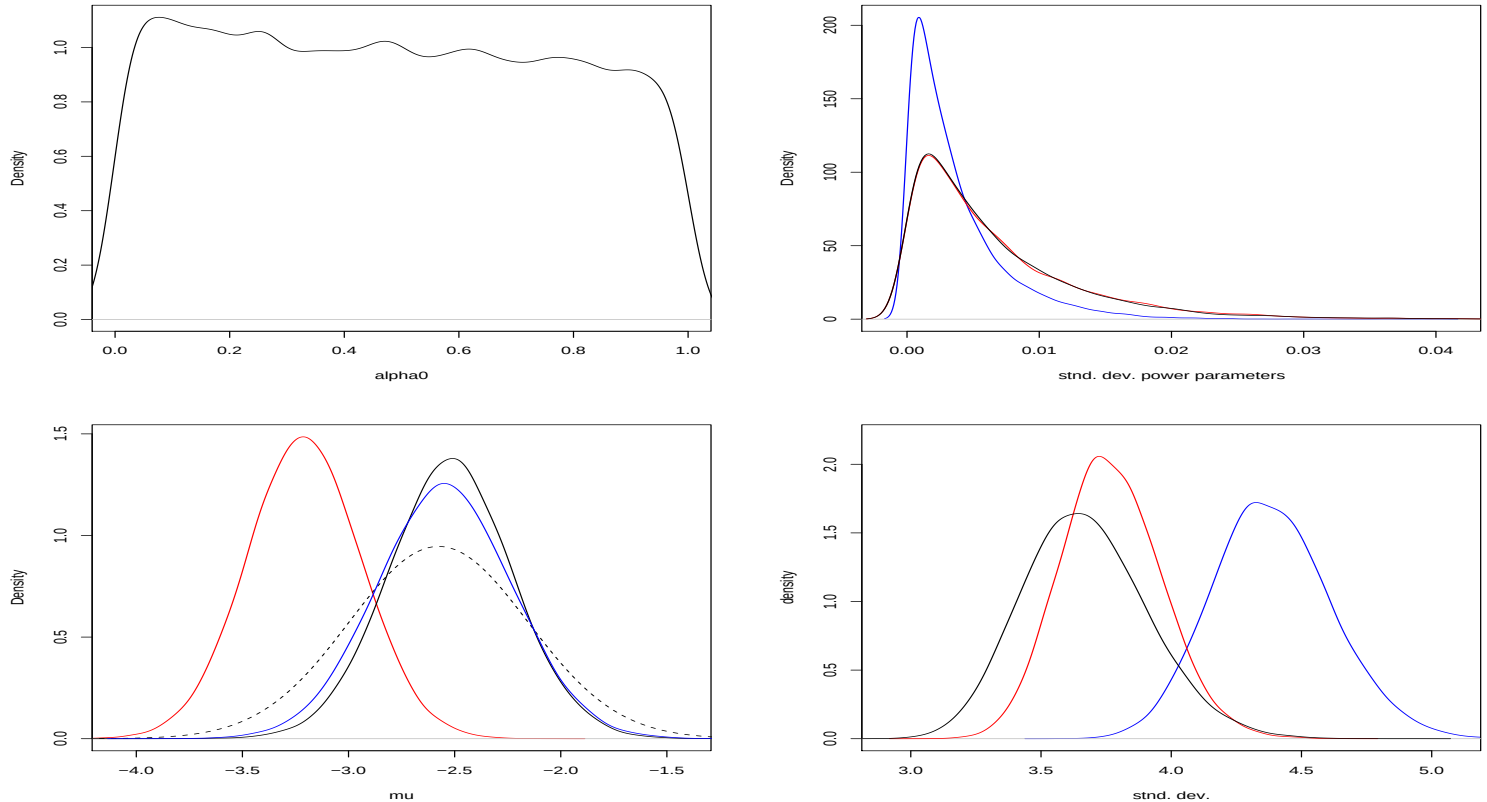
Power Parameter Priors

- $\pi(\alpha_0) = Beta(1, 1)$, $\pi(\phi_0) = Beta(1, 1)$, and $\pi(\omega_{0j}) = Beta(1, 1)$, $j = 1, 2$

Full conditional distributions for mean parameters:

- $f(\lambda_j|rest) = N(\bar{y}_j, \frac{\tau_j^2}{m_j})$, $\quad j = 1, 2$
- $f(\mu|rest) = N\left(\frac{\tau_1^2\tau_2^2(n\bar{v}+n_0\bar{v}_0(\alpha_0+\phi_0))+\sigma^2 n_0\bar{v}_0(\tau_2^2\omega_{01}+\tau_1^2\omega_{02})}{\sigma^2\tau_1^2\tau_2^2} V\ ,\ V\right)$
  where $V = \frac{\sigma^2\tau_1^2\tau_2^2}{\tau_1^2\tau_2^2(n+n_0(\alpha_0+\phi_0))+\sigma^2 n_0(\tau_2^2\omega_{01}+\tau_1^2\omega_{02})}$

Analysis of Colorectal Cancer Data with Power Priors



Top row: posteriors for α_0 (left) and $\phi_0, \omega_{01}, \omega_{02}$ (right). Bottom row, posteriors for $\mu, \lambda_1, \lambda_2$ (left; dashed line is historical posterior for μ) and for σ, τ_1, τ_2 (right).

- FOLFOX doing marginally better than IROX ($\lambda_1 < \lambda_2$)
- *But* α_0 posterior $\approx$ flat $\Rightarrow$ try a less overparametrized model?

Hierarchical Model for a Doubly Controlled Trial

Now suppose the full "Telba model" where historical data exists for both the treatment and control groups

To address this for future FDA applicant audiences we step back from power priors to a straight hierarchical modeling framework using standard components:
Let $g = 0, 1$ indicate group (historical or current), and let $i = 1, \ldots, n_g$ index the patients in each group

Alternative hierarchical model:

$$\textbf{Likelihood:} \quad Y_{gi} \overset{ind}{\sim} N(\theta_g + \beta_g x_{gi}\ ,\ \sigma_g^2)$$

$$\text{where } x_{gi} = \begin{cases} 0 & \text{if patient } gi \text{ received control} \\ 1 & \text{if patient } gi \text{ received treatment} \end{cases}$$

$$\textbf{Prior:} \quad \theta_g \overset{iid}{\sim} N(\mu_\theta, \tau_\theta^2)$$

$$\text{and } \beta_g \overset{iid}{\sim} N(\mu_\beta, \tau_\beta^2)$$

Hierarchical Model for a Doubly Controlled Trial

Prior shrinkage

- If $\tau_\theta^2 = 0$, then $\theta_g = \mu_\theta$ for all g, and we have no borrowing among the control groups
- If $\tau_\beta^2 = 0$, then $\beta_g = \mu_\theta$ for all g, and we have no borrowing among the treatment groups

Hyperprior

- Take flat hyperpriors on the mean parameters μ_θ and μ_β.
- For the variances, work (like `WinBUGS`) on the precision scale, assuming $\eta \sim G(a_\theta, b_\theta)$ and $\eta_\beta \sim G(a_\beta, b_\beta)$ where $\eta = 1/\tau^2$
  - So if $a_\beta = 1000$ and $b_\beta = 10 \Rightarrow \eta_\beta \approx 100 \Rightarrow \tau_\beta \approx 0.1$, a high-shrinkage hyperprior
  - If $a_\beta = 40$ and $b_\beta = 4$, this is a vaguer prior having $\eta_\beta \approx 10 \Rightarrow \tau_\beta \approx 0.3$, a moderate-shrinkage hyperprior
  - If $a_\beta = b_\beta = \epsilon = 0.1$, the hyperprior is vague and the data must do all the work a low-shrinkage hyperprior

  Differing levels of shrinkage would also be assigned to the control group via a_θ and b_θ.

Simulation of Operating Characteristics

Suppose we take $n_0 = n_1 = 20$, and without loss of generality take $\sigma_g^2 = 1$. Simulate frequentist power under a couple different scenarios:

- $\theta_0 = \theta_1 = 0$, $\beta_0 = \beta_1 = 0$: complete homogeneity; no reason not to borrow
- $\theta_0 = \theta_1 = 0$, $\beta_0 = 0, \beta_1 = 2$: slight heterogeneity across treatment groups; borrowing suspect
- $\theta_0 = 0$, $\theta_1 = 30$, $\beta_0 = 0$, $\beta_1 = 2$: Enormous heterogeneity across control groups, slight heterogeneity across treatment groups; borrowing very suspect

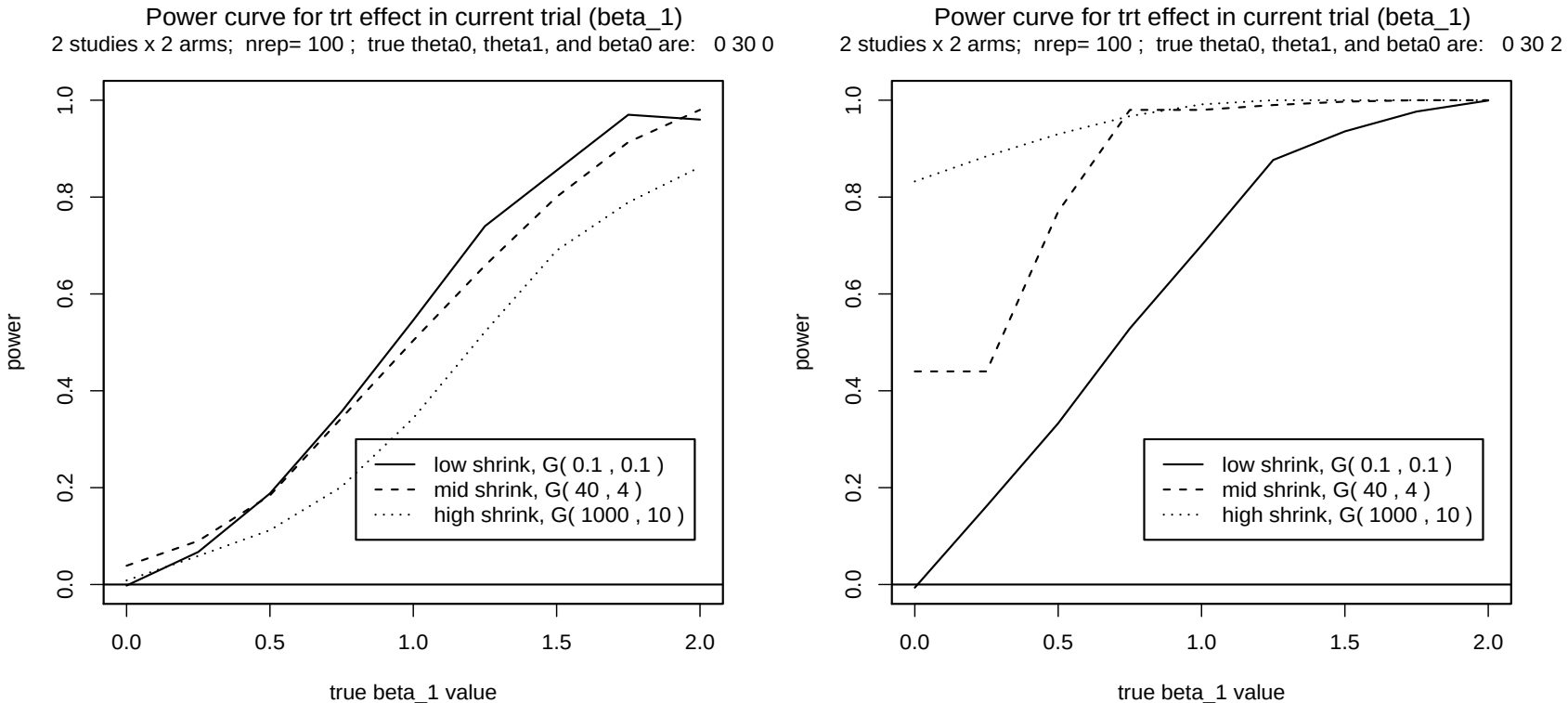
In each case, we would lay out a grid of "true" β_1 values, choose the shrinkage level in the treatment and control group hyperpriors, and simulate frequentist power under various hypotheses....

Test for treatment effect in current trial

Hypotheses: $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$

Rule: Reject H_0 if the central 95% credible interval excludes 0

Results for $\theta_0 = 0$, $\theta_1 = 30$, and $\beta_0 = 0$ (left) vs. $\beta_0 = 2$ (right) under "moderate shrinkage" prior for the θ's:



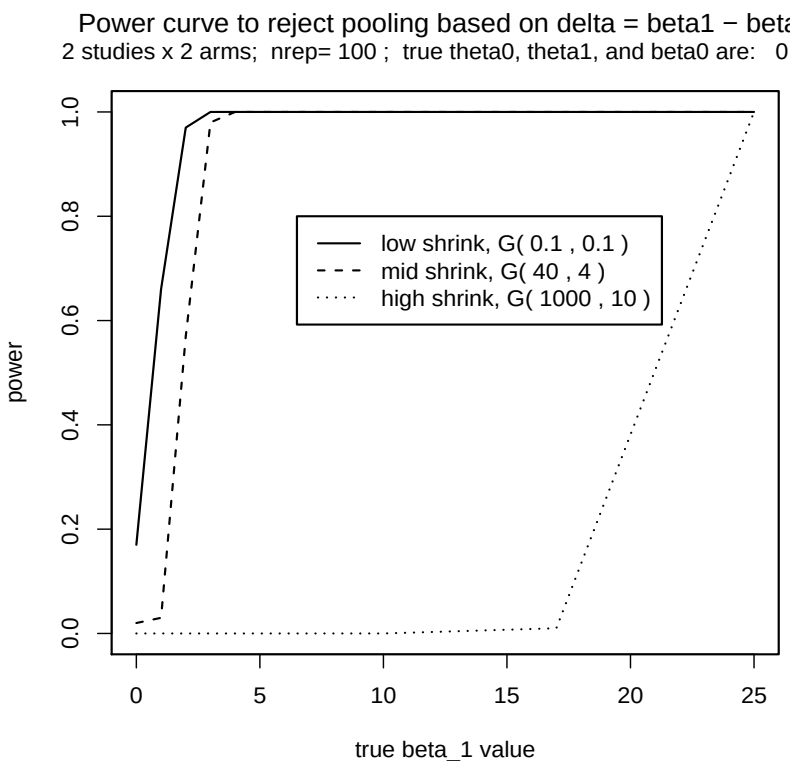
Note high-shrinkage does well when $\beta_0 = 0$ (left panel) even though $\theta_1 \neq 0$, but has high Type I error when $\beta_0 \neq 0$ (right panel)

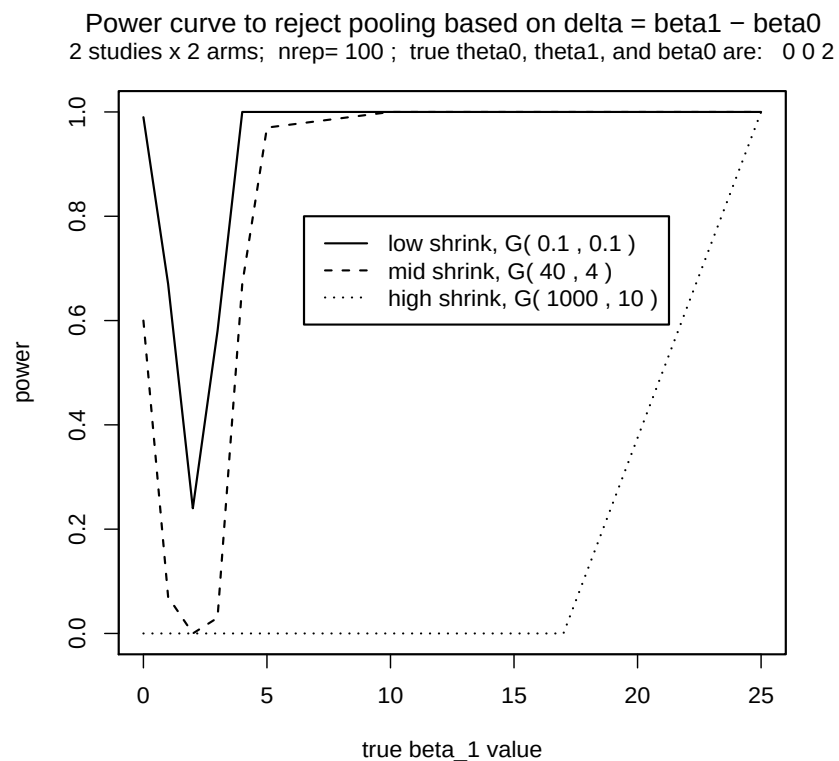
Test whether "To pool or not to pool"

Hypotheses: Define $\Delta = \beta_1 - \beta_0$, and test $H_0 : \Delta \in (-c, c)$ for some $c > 0$, vs. $H_a : \Delta \notin (-c, c)$

Rule: Reject H_0 if $P(\Delta \in (-1, 1)|\mathbf{y}) < 0.80$

Results for $\theta_0 = \theta_1 = 0$, and $\beta_0 = 0$ (left) vs. $\beta_0 = 2$ (right) under "low shrinkage" prior for the θ's:





Conclusions

- We've investigated several frameworks for sensible use of Bayesian adaptive methods in clinical trial design and analysis
- Power priors emerge as effective tools for borrowing strength from historical data
  - nagging question: Data-based estimation of α_0?
- For the doubly controlled trial, the alternative hierarchical models offer a viable power prior-free approach, and also answer a company's question, "May we pool these two data sets in our FDA application?"
- Future work:
  - re: the "nagging question,"
    - Try empirical Bayes (EB) methods to estimate model variances, then use power priors only on the mean parameters
    - Investigate identifiability of power parameters, and extend to models that "parametrize commensurability" (similarity between D and D_0) so this can be used to help
  - Extending our basic power prior models to the binary and time-to-event data settings

Thanks

I owe a debt of gratitude to many collaborators:

- Mr. Brian Hobbs, PhD candidate, University of Minnesota
- Dr. Peter Müller, University of Texas M.D. Anderson Cancer Center
- Drs. Dan Sargent and Sumithra Mandrekar, Mayo Clinic
- myriad other Bayes/clinical trials collaborators over the years who helped shape the work in a variety of ways