

Hierarchical Commensurate Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials

Brian P. Hobbs¹, Bradley P. Carlin¹,
Sumithra Mandrekar,² and Daniel Sargent²

1 Division of Biostatistics, School of Public Health, University of Minnesota

2 Mayo Clinic, Rochester, Minnesota

Iowa State University – February 21, 2011

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 powerbut 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*)
- ▶ Let $D_0 = (n_0, \mathbf{x}_0)$ denote historical data, suppose θ is the parameter of interest, and let $L(D_0|\theta)$ denote the general likelihood
- ▶ Suppose $\pi_0(\theta)$ is the prior distribution on θ *before* D_0 is observed, the *initial prior*
- ▶ The *conditional power prior* on θ for the current study is the *historical likelihood*, $L(D_0|\theta)$, raised to power α_0 , where $\alpha_0 \in [0, 1]$, multiplied by the initial prior:

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

- ▶ α_0 is the *power parameter* that controls the “degree of borrowing” from D_0

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 for normal data, $x_{0i} \stackrel{iid}{\sim} N(\theta, \sigma_0^2)$, since under a flat initial prior,

$$\pi_0(\theta|D_0, \alpha_0) = N(\bar{x}_0, \sigma_0^2/(\alpha_0 n_0))$$

- ▶ Think of $\alpha_0 n_0$ as the “effective” historical sample size
- ▶ Given **current data**, $D = (n, \mathbf{x})$, the *conditional posterior*

$$q(\theta|D, D_0, \alpha_0) \propto L(D_0|\theta)^{\alpha_0} L(D|\theta)\pi_0(\theta)$$

- ▶ $\alpha_0 \rightarrow 1$, $q(\theta|D, D_0, \alpha_0) \rightarrow$ approaches *full borrowing* from D_0
- ▶ $\alpha_0 \rightarrow 0$, $q(\theta|D, D_0, \alpha_0) \rightarrow$ approaches *no borrowing* from D_0

Power Priors (cont'd)

- ▶ We could fix $\alpha_0 \in [0, 1]$ and assume consistency among D_0 and D is known, but if this is not the case, models with poor frequentist operating characteristics may result
- ▶ Choosing a **hyperprior**, $\pi(\alpha_0)$, for α_0 enables the data to help determine probable values for α_0
- ▶ Ibrahim-Chen (2000) propose **joint power priors** of form

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

- ▶ Duan et al. (2006), Neuenschwander et al. (2009), and Pericchi (2009) propose **modified joint power priors** (MPP) which respect the Likelihood Principle,

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

Location Commensurate Power Priors (LCPP)

- ▶ We propose an adaptive modification of the MPP
- ▶ Let $x_{0i} \stackrel{iid}{\sim} \text{Normal}(\mu_0, \sigma_0^2)$ and $x_i \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2)$
- ▶ *Different* parameters in historical and current group, μ_0 and μ
- ▶ Extend hierarchical model to include parameter, τ , that directly measures similarity of μ and μ_0
- ▶ Construct prior for μ dependent upon μ_0 and τ
- ▶ τ parametrizes commensurability (precision)
- ▶ Use information in τ to guide prior on α_0

$$\pi^{\text{LCPP}}(\mu, \alpha_0, \tau | \mathbf{x}_0) \propto$$

$$\int \frac{[N(\mathbf{x}_0 | \mu_0, \hat{\sigma}_0^2)]^{\alpha_0}}{\int [N(\mathbf{x}_0 | \mu_0, \hat{\sigma}_0^2)]^{\alpha_0} d\mu_0} d\mu_0 \times N(\mu | \mu_0, \frac{1}{\tau}) \times \text{Beta}(\alpha_0 | \tau^a, 1) \times p(\tau)$$

LCPP for Single Arm Trial

- ▶ Formalize commensurate as μ_0 near μ by adopting Normal prior on μ with mean μ_0 and precision τ
- ▶ $Beta(\tau^a, 1)$ prior on α_0 for some $a > 0$
- ▶ τ close to 0 corresponds to very low commensurability, while very large τ implies the two datasets may arise from similar populations
- ▶ $\tau \rightarrow \infty$, point-mass prior on α_0 at 1
- ▶ $\tau \rightarrow 0$ discourages incorporation of historical information
- ▶ Requires a fixed, known sampling historical variance $\hat{\sigma}_0^2$ (MLE)

LCPP for Single Arm Trial (cont'd)

- ▶ Both α_0 and τ inflate the “prior” variance

$$\pi^{LCPP}(\mu, \alpha_0, \tau | \mathbf{x}_0) \propto N(\mu | \bar{x}_0, \frac{1}{\tau} + \frac{\hat{\sigma}_0^2}{\alpha_0 n_0}) \times \text{Beta}(\alpha_0 | \tau^a, 1) \times p(\tau)$$

- ▶ Prior on τ : Mixture of two gammas with mixing probability $\omega = 1/2$ and with hyperparameter $a = 1/2$ provides sufficient flexibility:

$$p(\tau) \propto \left(\omega \text{Gamma}(\tau | 1, 10) + (1 - \omega) \text{Gamma}(\tau | 3/2, \frac{1}{1000}) \right)$$

- ▶ Posterior obtained after multiplying by current likelihood $N(\mathbf{x} | \mu, \sigma^2)$ and vague (say reference) prior on σ^2
- ▶ $q(\sigma^2 | \mathbf{x}) \propto IG\left(\frac{n}{2}, \frac{n}{2} [s^2 + (\bar{x} - \mu)^2]\right)$

Extension to Linear Models

Formulate linear model to borrow adaptively from the identical covariates

- ▶ Suppose that both trials identically measure $p - 1$ covariates
- ▶ Let X_0 and X be $n_0 \times p$ and $n \times p$ design matrices
- ▶ Suppose $y_0 \sim N_{n_0}(X_0\beta_0, \sigma^2)$ and $y \sim N_n(X\beta + Z\lambda, \sigma^2)$ where Z is an $n \times r$ design matrix containing variables relevant only to the current trial, and an indicator for new treatment
- ▶ Let $D_0 = (y_0, X_0, n_0, p)$, and $D = (y, X, Z, n, p, r)$
- ▶ Assume flat prior for λ
- ▶ Let $\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T y_0$

$$\pi^{LCPP}(\beta, \lambda, \sigma^2, \alpha_0, \tau^2 | D_0) \propto$$

$$\int N_{n_0}(y_0 | X_0\beta_0, \hat{\sigma}_0^2 I_{n_0}) N_p\left(\beta | \hat{\beta}_0, \frac{1}{\tau} I_p\right) d\beta_0 \times \text{Beta}(\alpha_0 | \tau^a, 1) \times \frac{1}{\sigma^2} \times p(\tau)$$

Randomized Controlled Colorectal Cancer Trials

- ▶ Two successive randomized controlled colorectal cancer trials on subjects with previously untreated metastatic colorectal cancer:

Saltz et al. (2000) trial randomized $N_0 = 683$: May 1996 and May 1998

1. Irinotecan alone (arm A)
2. Irinotecan and bolus Fluorouracil plus Leucovorin (arm B; IFL) *significantly longer progression free survival*
3. Fluorouracil and Leucovorin (arm C; 5FU/LV) *standard therapy*

Goldberg et al. (2004) trial randomized $N = 795$: May 1999 and April 2001

1. Irinotecan and bolus Fluorouracil plus Leucovorin (IFL) *regulatory standard in March 2000*
2. Oxaliplatin and infused Fluorouracil plus Leucovorin (FOLFOX) *new regimen*
3. Irinotecan and Oxaliplatin (IROX) *new regimen*

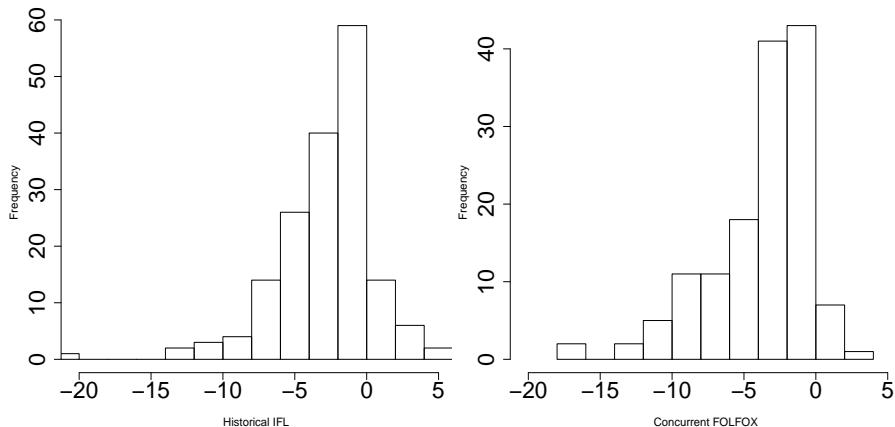
Randomized Controlled Colorectal Cancer Trials (cont'd)

- ▶ Longest diameter (ld) in cm of 1 to 9 tumors measured every 6 weeks for the first 42 weeks or until a response
- ▶ Compare **FOLFOX** to **IFL** for average reduction in ld sum from BL
- ▶ Covariate adjustments for baseline ld sum, age, and AST in units/L
- ▶ Historical: arm B (**IFL**) from the Saltz trial, $n_0 = 171$
- ▶ Current: **IFL**, $n = 129$, and **FOLFOX** in the Goldberg trial, $n = 141$

Ordinary linear regression fits to colorectal cancer data:

	Historical data		Current data	
	estimate	95% CI	estimate	95% CI
Intercept	0.880	(-1.977, 3.738)	-0.467	(-2.275, 1.341)
BL Tumor Sum	-0.232	(-0.310, -0.154)	-0.397	(-0.453, -0.340)
Age	-0.022	(-0.067, 0.022)	0.014	(-0.014, 0.041)
AST	-0.001	(-0.017, 0.015)	0.005	(-0.007, 0.017)
FOLFOX	-	-	-0.413	(-1.017, 0.190)

Randomized Controlled Colorectal Cancer Trials (cont'd)



Histograms of average change in Id tumor sum from baseline: historical IFL (left), FOLFOX (right); note FOLFOX results are slightly better (more negative).

Randomized Controlled Colorectal Cancer Trial (cont'd)

LCPP fit to colorectal cancer data:

	estimate	95% BCI
β (Intercept)	0.180	(-1.11, 1.42)
β (BL Tumor Sum)	-0.39	(-0.44, -0.33)
β (Age)	0	(-0.02, 0.02)
β (AST)	0	(-0.01, 0.01)
λ (FOLFOX)	-0.46	(-0.82, -0.10)
α_0	0.86	(0.44, 1.00)

- ▶ High posterior estimate for $\alpha_0 \Rightarrow$ LCPP analysis incorporates virtually all of the historical data
- ▶ Conclude FOLFOX resulted in a significant reduction in average Id sum when compared to the IFL
- ▶ Consistent with those of Goldberg et al. (2004), who determined FOLFOX to have better times to progression and response rates

Competing Non-Power Prior Approaches

Let $x_{0i} \stackrel{iid}{\sim} \text{Normal}(\mu_0, \sigma_0^2)$ and $x_i \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2)$

1. **Cauchy** prior on μ centered at historical sample mean \bar{x}_0
 - ▶ $\pi^{\text{cau}} \propto \text{Cauchy}(\mu | \text{median} = \bar{x}_0, \text{scale} = \gamma)$
2. **Location Commensurate Prior (LCP)**
 - ▶ Adaptive mechanism based solely on the commensurability parameter τ :

$$\pi^{\text{LCP}}(\mu, \sigma^2, \tau | \mathbf{x}_0) \propto N(\mu | \bar{x}_0, \frac{1}{\tau} + \frac{\hat{\sigma}_0^2}{n_0}) \times \frac{1}{\sigma^2} \times p(\tau)$$

- ▶ We use a mixture prior on τ , often with $\omega = 1/2$:

$$p(\tau) \propto \left(\omega \text{Gamma}(\tau | 1, 10) + (1 - \omega) \text{Gamma}(\tau | 3/2, \frac{1}{1000}) \right)$$

Non-Power Prior Approaches (cont'd)

3. Location-Scale Commensurate Mixture Prior (LSCMP)

- ▶ Borrowing depends upon commensurability of **both** the location and scale parameters
- ▶ $q_0(\mu_0, \sigma_0^2 | \mathbf{x}_0) \propto \text{Normal}(\mathbf{x}_0 | \mu_0, \sigma_0^2) \times \left(\frac{1}{\sigma_0^2}\right)$
- ▶ Commensurability Priors
 - ▶ $N(\mu | \mu_0, \frac{1}{\nu})$
 - ▶ $IG(\sigma^2 | A = \gamma\sigma_0^4 + 2, B = \sigma_0^2(\gamma\sigma_0^4 + 1))$

$$\pi^{\text{LSCMP}}(\mu, \sigma^2, \sigma_0^2 | \mathbf{x}_0, \nu, \gamma) \propto$$

$$\int q_0(\mu_0, \sigma_0^2 | \mathbf{x}_0) \times N\left(\mu | \mu_0, \frac{1}{\nu}\right) \times IG(\sigma^2 | A(\sigma_0, \gamma), B(\sigma_0, \gamma)) d\mu_0$$

$$\propto N\left(\mu | \bar{x}_0, \frac{n_0 + \nu\sigma_0^2}{\nu n_0}\right) \times IG(\sigma^2 | A, B) \times IG\left(\sigma_0^2 | \frac{n_0 - 1}{2}, \frac{n_0 \hat{\sigma}_0^2}{2}\right)$$

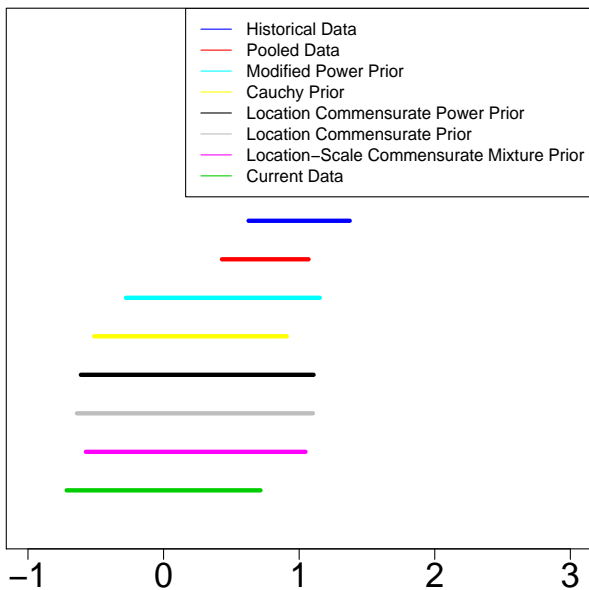
Non-Power Prior Approaches (cont'd)

3. Location-Scale Commensurate Mixture Prior (cont'd)

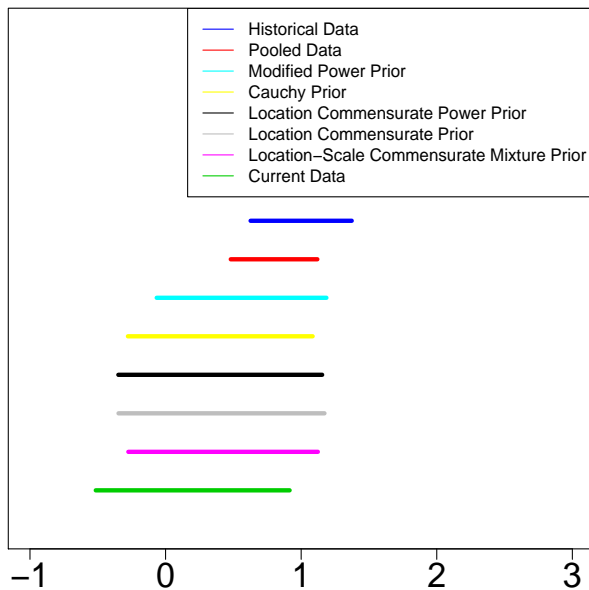
- ▶ Fix values of $\nu = (\nu_1, \nu_0)$ and $\gamma = (\gamma_1, \gamma_0)$ corresponding to high and low precision
 - ▶ $\nu = (10^{12}, (10\hat{\sigma}_0^2)^{-1})$
 - ▶ $\gamma = (10^2, 5^{-1})$
- ▶ Formulate mixture prior with fixed mixing probability θ such that $\pi^*(\mu, \sigma | \mathbf{x}_0, \nu, \gamma, \theta)$ is proportional to

$$\theta\pi(\mu, \sigma^2, \sigma_0^2 | \mathbf{x}_0, \nu_1, \gamma_1) + (1 - \theta)\pi(\mu, \sigma^2, \sigma_0^2 | \mathbf{x}_0, \nu_0, \gamma_0),$$

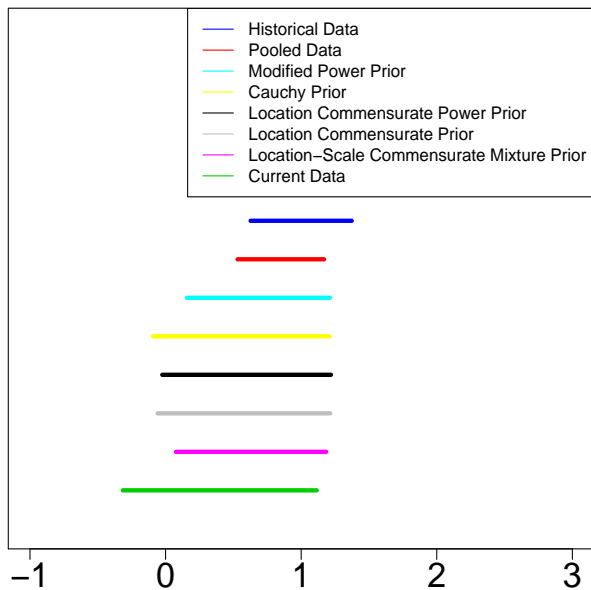
where fixing $\theta = 0.5$ seems to work well.



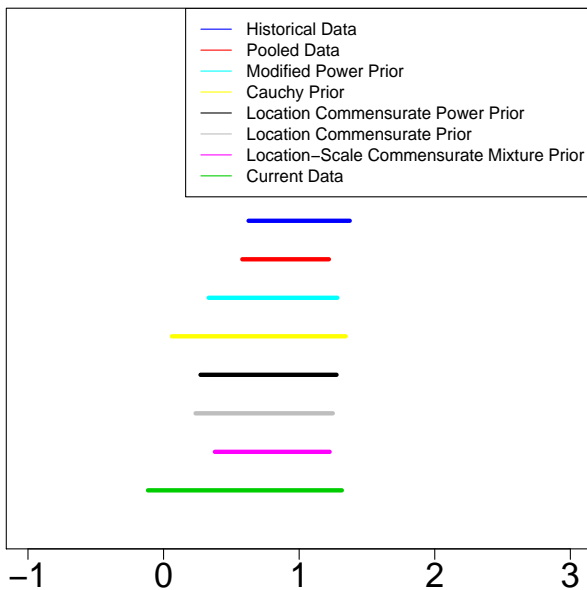
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 0$, $\bar{x}_0 = 1$



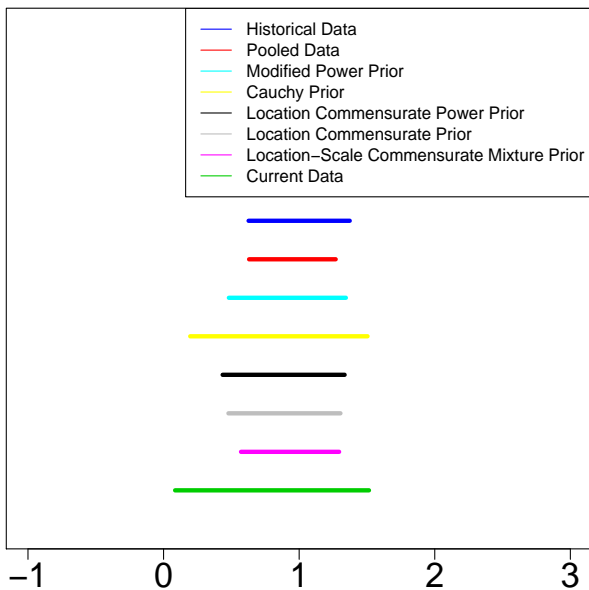
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 0.2$, $\bar{x}_0 = 1$



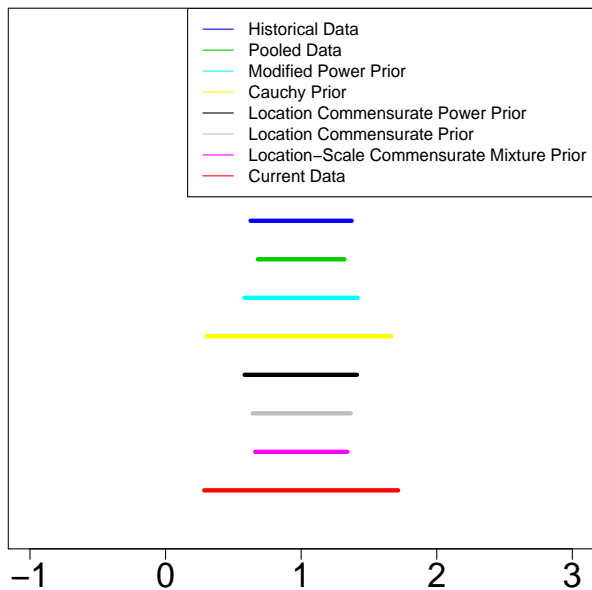
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 0.4$, $\bar{x}_0 = 1$



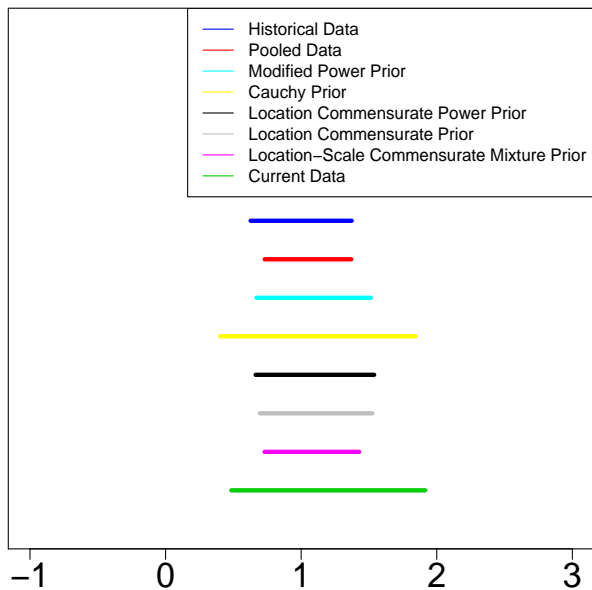
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 0.6$, $\bar{x}_0 = 1$

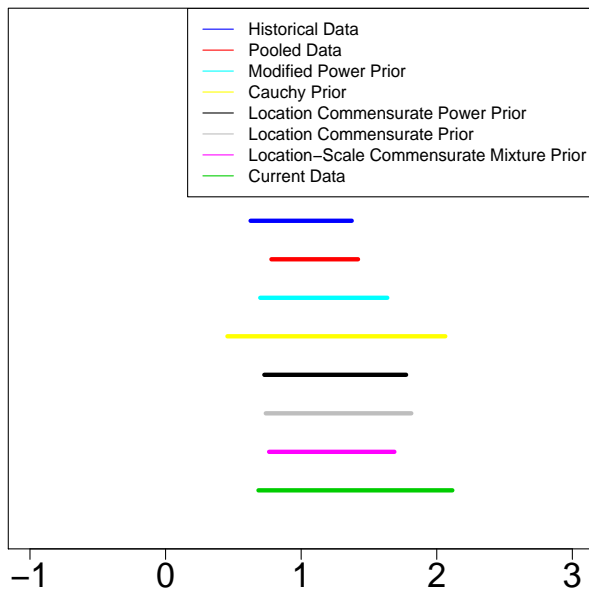


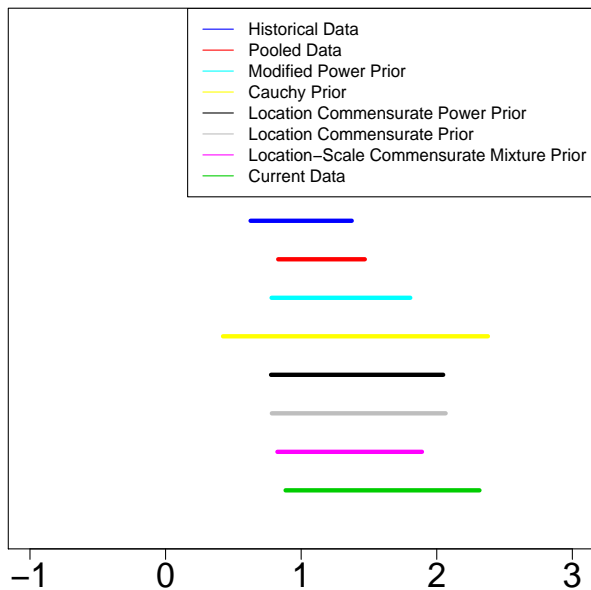
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 0.8$, $\bar{x}_0 = 1$



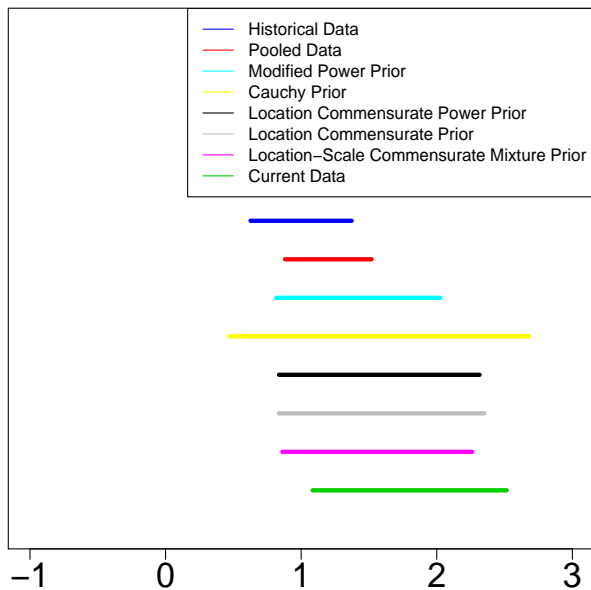
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 1$, $\bar{x}_0 = 1$



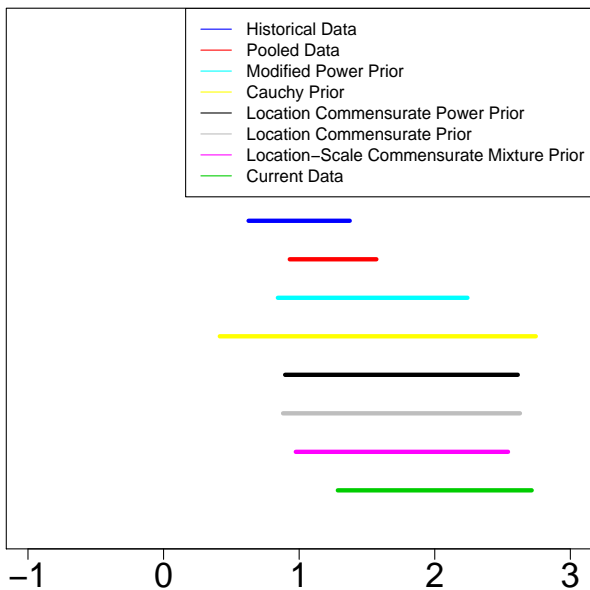




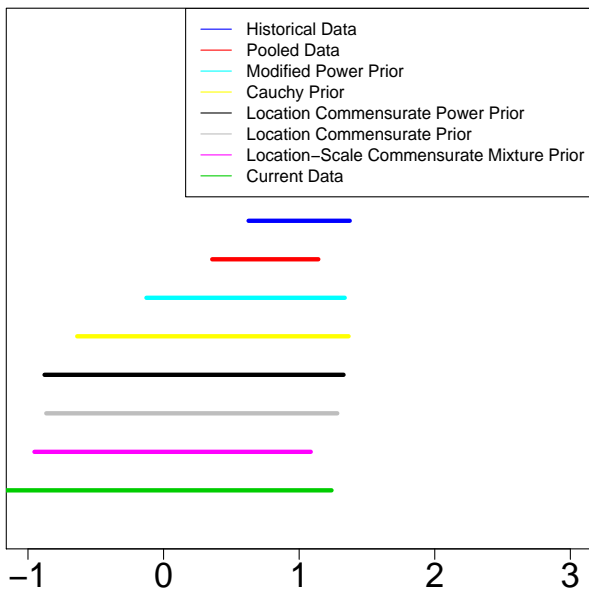
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 1.6$, $\bar{x}_0 = 1$



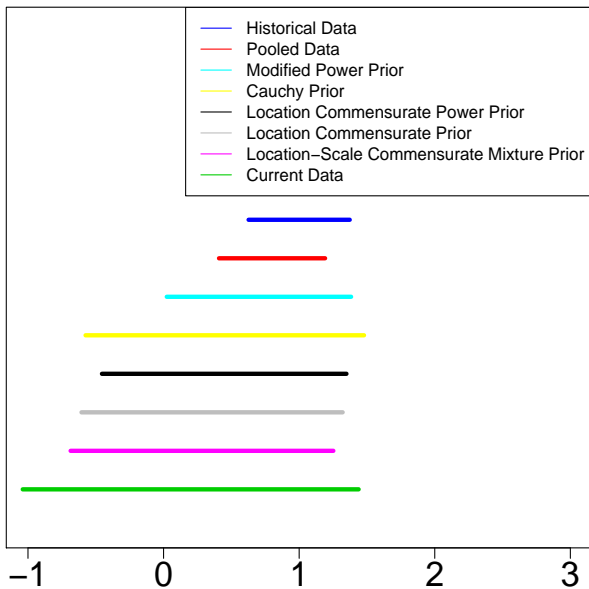
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 1.8$, $\bar{x}_0 = 1$



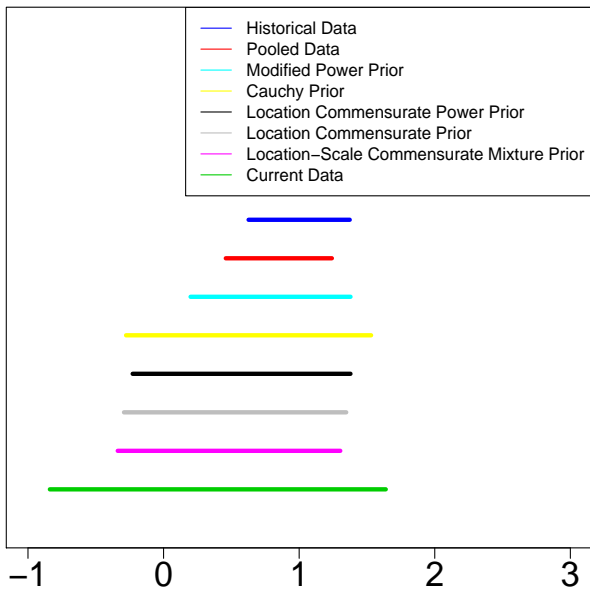
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = s_0^2 = 1$; $\bar{x} = 2$, $\bar{x}_0 = 1$



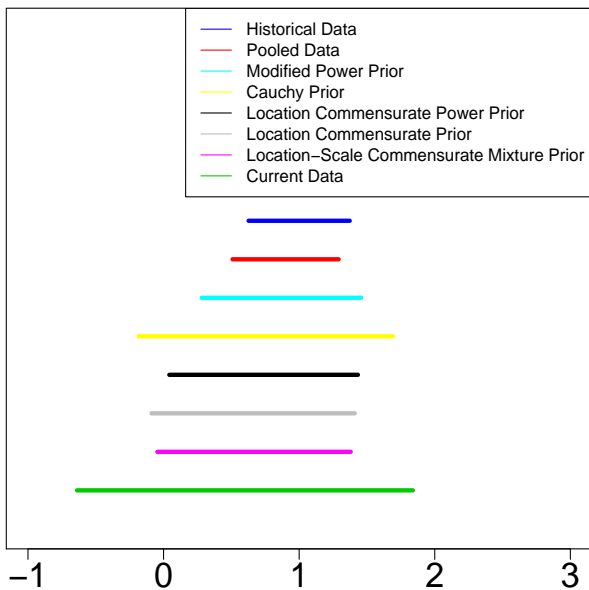
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 0$, $\bar{x}_0 = 1$



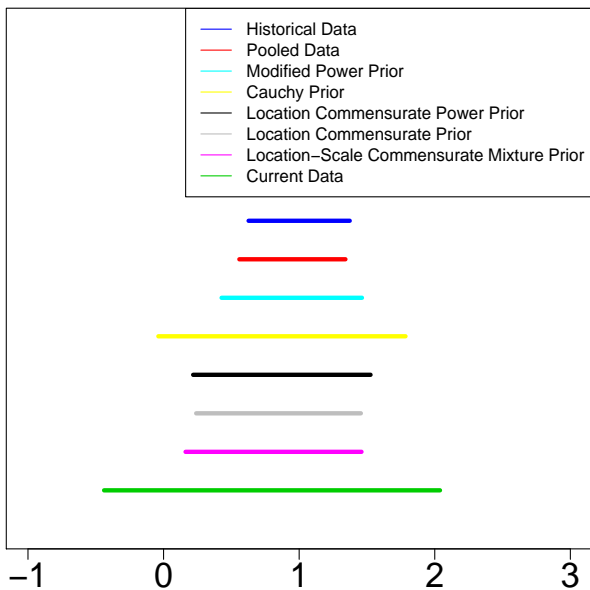
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 0.2$, $\bar{x}_0 = 1$



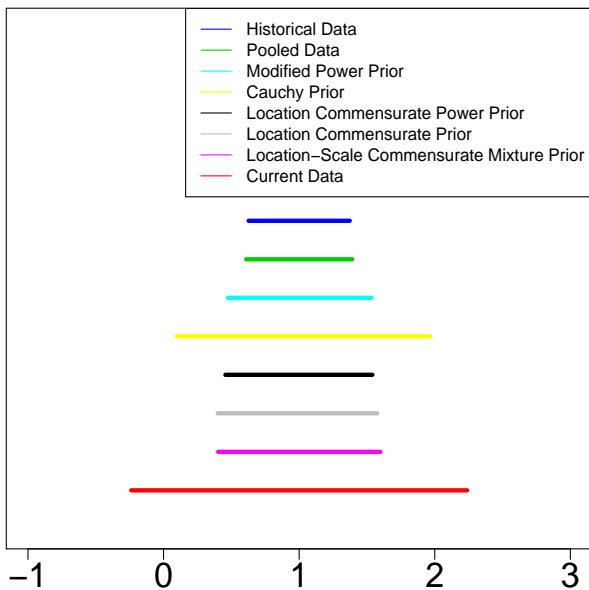
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 0.4$, $\bar{x}_0 = 1$



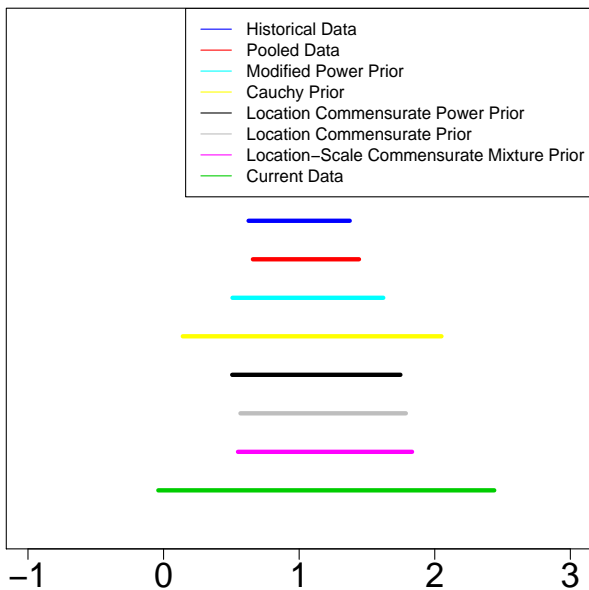
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 0.6$, $\bar{x}_0 = 1$



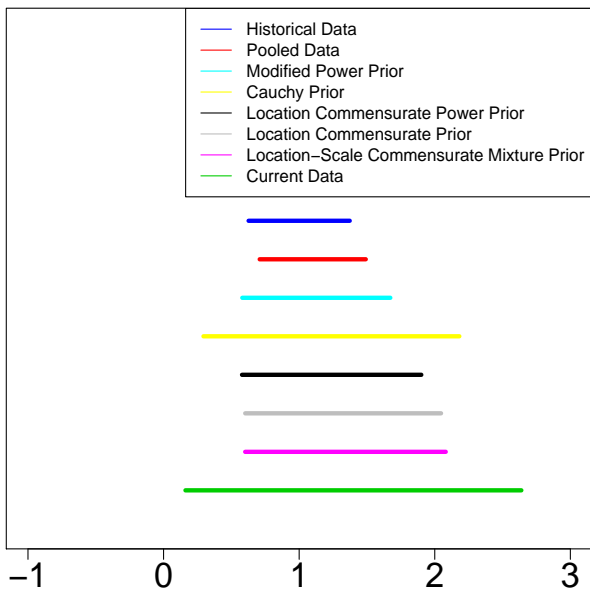
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 0.8$, $\bar{x}_0 = 1$



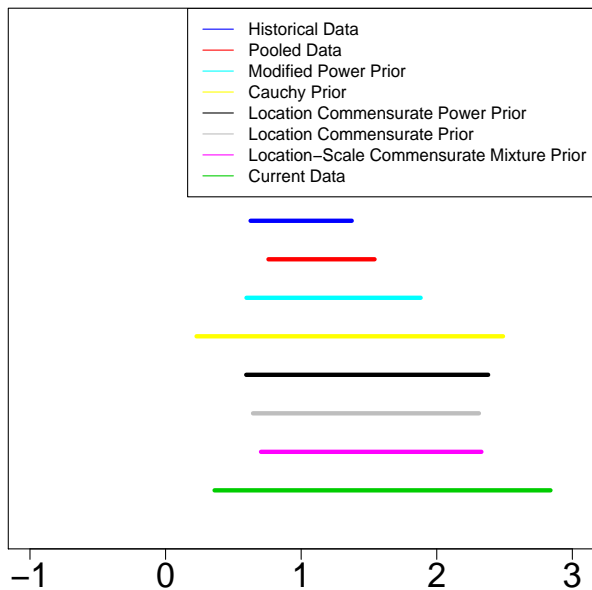
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 1$, $\bar{x}_0 = 1$



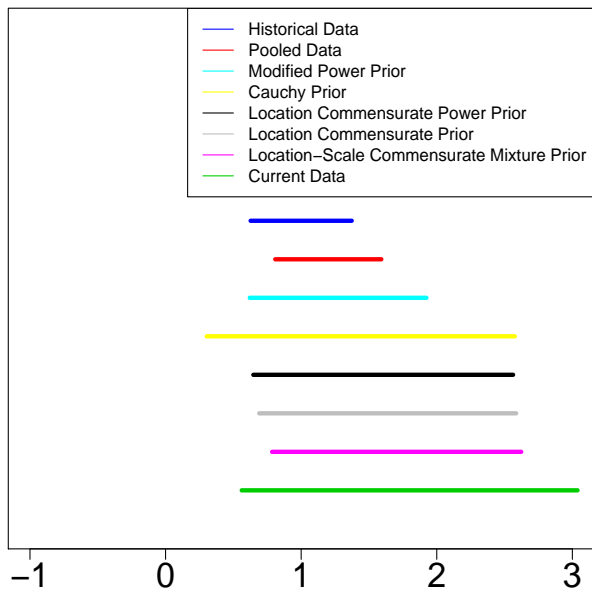
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 1.2$, $\bar{x}_0 = 1$



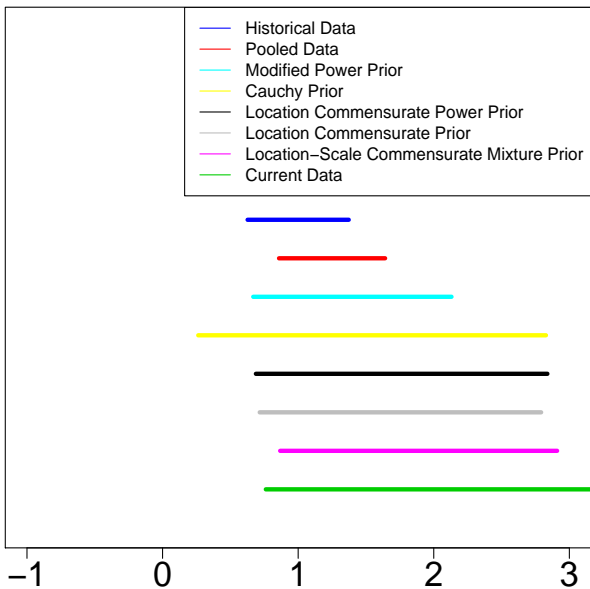
95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 1.4$, $\bar{x}_0 = 1$



95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 1.6$, $\bar{x}_0 = 1$



95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 1.8$, $\bar{x}_0 = 1$



95% BCI for μ given $n = 10$, $n_0 = 30$; $s^2 = 3$, $s_0^2 = 1$; $\bar{x} = 2$, $\bar{x}_0 = 1$

Extension to Binary Regression

Formulate commensurate power prior for binary outcomes

- ▶ Historical data $y_{0i} \sim \text{Ber}[\pi_0(X_{0i})]$; Current data $y_i \sim \text{Ber}[\pi(X_i, d_i)]$ where d_i is the treatment indicator
- ▶ Location commensurate prior for binary regression follows proportional to

$$\prod_{i=1}^{n_0} \text{Ber}(y_{0i} | \pi(x_{0i})) \times N\left(\beta | \beta_0, \frac{1}{\tau}\right) \times p(\tau),$$

where we use $p(\tau)$ to bound τ away from 0 or ∞ .

- ▶ Can't integrate out β_0 analytically, so instead just multiply by current data likelihood and normalize via MCMC...
- ▶ Can choose **probit** link, $\pi_0(X_0) = \Phi(X_0\beta_0)$ and $\pi(X, d) = \Phi(X\beta + d\lambda) \Rightarrow$ closed form full conditionals!
- ▶ Can instead pick **logit** link, $\pi(x_0) = (1 + e^{-x_0\beta_0})^{-1} \Rightarrow$ requires Metropolis-Hastings
- ▶ **Example:** Adaptively incorporating nonrandomized HIV trial arms!

Extension to Generalized Linear Models

- ▶ Assume an initial flat prior on β_0 , and use the “Bayesian Central Limit Theorem” with the historical likelihood to obtain

$$\beta_0 \sim N(\hat{\beta}_0, \hat{\Sigma}_0),$$

where $\hat{\beta}_0$ is the historical MLE for β_0 , and $\hat{\Sigma}_0$ is the inverse of the historical observed Fisher information matrix

- ▶ Specifying a flat prior for λ , a $N(\beta_0, \frac{1}{\tau} I_p)$ prior for β , and integrating out β_0 again leads to joint location commensurate prior of

$$p(\beta, \lambda, \tau | y_0) \propto N_p\left(V^{-1}M, \frac{1}{\tau}V^{-1}\right) p(\tau)$$

- ▶ The joint posterior follows by multiplying by the current data likelihood, which yields intractable full conditionals. Here we use **Metropolis-Hastings** (NOT the BCLT again, which would require fixed β and λ in the Fisher information matrix)
- ▶ **Primary target application:** formulation for **survival** outcomes...

Time-to-Event Response

Historical, concurrent data are triples $(t_{0j}, \delta_{0j}, X_{0j})$ for $j = 1, \dots, n_0$ and (t_i, δ_i, X_i) for $i = 1, \dots, n$; where t s are the observed (possibly censored) failure times, δ s are noncensoring indicators, and X_{0j} and X_i are row vectors of p covariates associated with historical subject j and concurrent subject i .

- ▶ Let $f(t_{0j})$ and $f(t_i)$ denote survival time densities with survival functions $F(t)$ and $F(t_0)$
- ▶ Adopt a **log-linear** model: $y_0 = \log(t_0) = X_{0j}\beta_0 + \sigma_0 e_0$ and $y = \log(t) = X\beta + d\lambda + \sigma e$, where $e_0 = (y_0 - X_{0j}\beta_0)/\sigma_0$ and $e = (y - X\beta - d\lambda)/\sigma$. The likelihoods follow as

$$L_0(\beta_0, \sigma_0 | y_0) \propto \prod_{j=1}^{n_0} \left[\frac{1}{\sigma_0} f(e_{0j}) \right]^{\delta_{0j}} F(e_{0j})^{1-\delta_{0j}}$$

and

$$L(\beta, \sigma | y) \propto \prod_{i=1}^n \left[\frac{1}{\sigma} f(e_i) \right]^{\delta_i} F(e_i)^{1-\delta_i}$$

Time-to-Event Response (cont'd)

- ▶ Weibull regression occurs when e_0 and e follow the **extreme value** distribution, $f(u) = \exp[u - \exp(u)]$.
- ▶ We consider a **location-scale commensurate** approach, so borrowing depends upon commensurability among σ_0 and σ , as well as β_0 and β .
- ▶ LSCP follows from general theory (2 slides back) as

$$p(\theta, \lambda, \tau | y_0) \propto N_{p+1} \left(\Lambda^{-1} Q, \frac{1}{\tau} \Lambda^{-1} \right) p(\tau),$$

where $Q = \left(\hat{\Psi}_0 + \tau I_{p+1} \right)^{-1} \hat{\Psi}_0 \hat{\theta}_0$, $\Lambda = I_{p+1} - \tau \left(\hat{\Psi}_0 + \tau I_{p+1} \right)^{-1}$, and $\hat{\Psi}_0 = \hat{\Psi}_0(\hat{\theta}_0)$ is the observed Fisher information matrix. The posterior is then proportional to the product of this LCPP and the concurrent data likelihood, as usual.

- ▶ Note that the **exponential** model is a special case where $\sigma = 1$.

Extension to General/Generalized Mixed Models

Consider the Gaussian case for now (non-Gaussian also feasible):

- ▶ **historical data:** y_{0ij} , $i = 1, \dots, n_{0j}$, $j = 1, \dots, m_0$; arrange in a vector y_0 and model as $y_0 = X_0\beta_0 + Z_0u_0 + \epsilon_0$, where

$$E\begin{pmatrix} u_0 \\ \epsilon_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \text{Cov}\begin{pmatrix} u_0 \\ \epsilon_0 \end{pmatrix} = \begin{pmatrix} G_0 & 0 \\ 0 & R_0 \end{pmatrix}$$

- ▶ **current data:** uses similar notation: $y = X\beta + d\lambda + Zu + \epsilon$, where d is a $N \times 1$ indicator of treatment.
- ▶ The location commensurate prior emerges as proportional to

$$N_p\left(\beta \mid V^{-1}M, \frac{1}{\tau}V^{-1}\right) \times N_m\left(u \mid 0, \sigma_u^2 I_m\right) \times p(\lambda, \sigma_\epsilon, \sigma_u, \tau),$$

where $M = \left(X_0^T \hat{\Sigma}^{-1} X_0 + \tau I_p\right)^{-1} X_0^T \hat{\Sigma}^{-1} X_0 \tilde{\beta}_0$,

$V = I_p - \tau \left(X_0^T \hat{\Sigma}^{-1} X_0 + \tau I_p\right)^{-1}$, and $\tilde{\beta}_0$ is the BLUE.

- ▶ As usual, multiply by current data likelihood and normalize; full conditionals are normal and inverse Wishart; non-Gaussian case requires Metropolis-Hastings

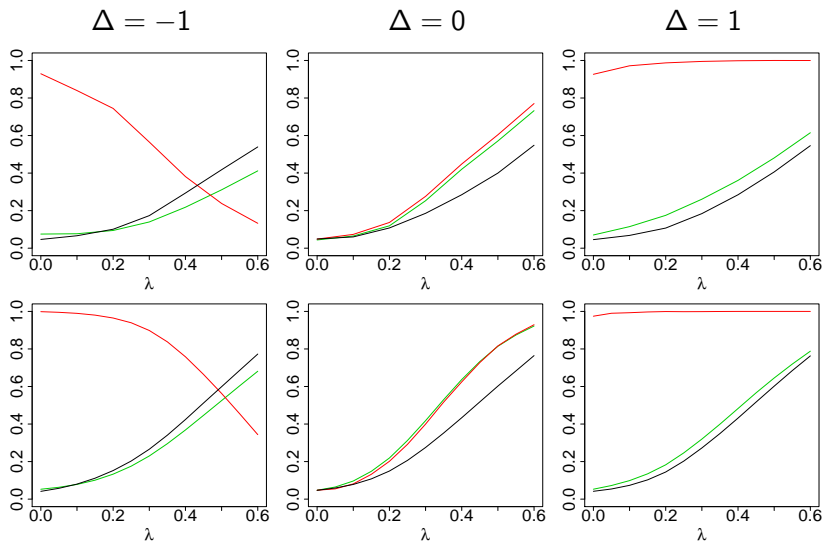
Regression Model Simulations

Check frequentist OCs of the linear, exponential, & Weibull models:

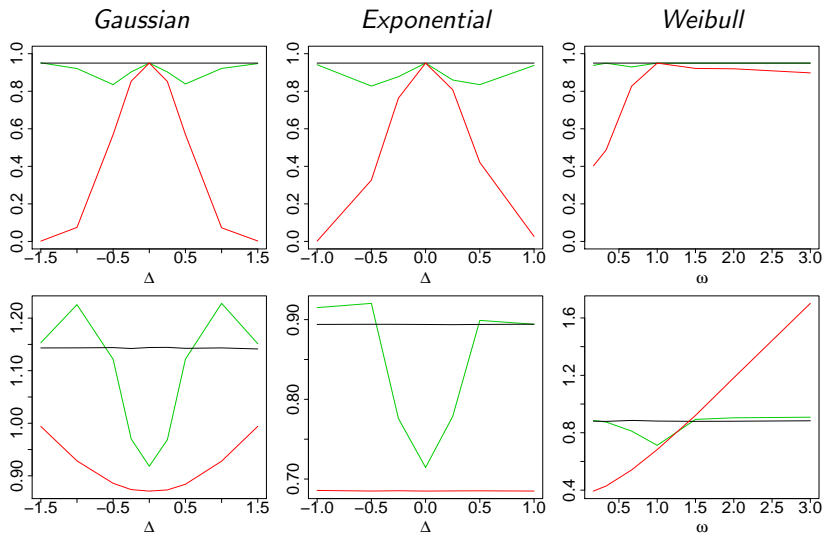
$$E[y_0] = \mu_0 \text{ and } E[y] = \mu_0 + d\lambda,$$

where d is treatment indicator and μ_0 & μ are intercepts for controls

- ▶ Let $\Delta = \mu_0 - \mu$, & spz $\lambda > 0$ indicates positive treatment effect
 - ▶ $\Delta > 0$ implies hist. performed better than current controls
 - ▶ $\Delta < 0$ implies hist. performed worse than current controls
- ▶ Type I error ($\lambda = 0$), power, & 95% post. CI interval coverage were computed by sampling y_0 and y for true fixed values of Δ and λ
- ▶ Use 95% posterior CI to test the null hypothesis $H_0 : \lambda = 0$
- ▶ Below, Weibull sims compare commensurability among **shape** parameters, through ratio $\omega = \sigma_0/\sigma$, when we set $\Delta = 0$
- ▶ Compare **commensurate priors** to models that ignore (no borrowing) & **fully incorporate (full borrowing)** the historical data



Power curves for the **full borrowing**, **commensurate prior**, and “no borrowing” models based on the 95% posterior CIs for λ . The top row shows results for the Gaussian linear model, where $\mu_0 = 1$, $\sigma_0^2 = \sigma^2 = 1$, $n_0 = 100$, and $n = 50$; bottom row the exponential model, for $\mu_0 = 2$, $n_0 = 200$, and $n = 100$.



Sim. results for **full borrowing**, **commensurate prior**, and “no borrowing” models. Top row plots average 95% posterior coverage probabilities for λ . Bottom row plots average length of 95% posterior CIs for λ . X-axes correspond to intercepts, $\Delta = \mu_0 - \mu$ (left & center), or the scale ratio, $\omega = \sigma_0/\sigma$ where $\Delta = 0$ (right).

Adaptive Randomization

Friedman, Furberg, and DeMets (1998, p.69) broadly refer to **randomization** procedures that *adjust* the **allocation ratio** as the study progresses as *adaptive*. Commonly used to

- ▶ balance prognostic factors among the intervention arms
- ▶ assign patients to the better performing regimen in early phase trials

Commensurate prior models for controlled trials naturally advocate a randomization scheme that is “**optimal**” with respect to the amount **strength borrowed from the historical controls**

- ▶ Use more of the *new* patients to learn about the efficacy and safety profile of the new intervention
- ▶ Adjust prob. of allocating next patient to **new intervention according to the commensurability of the historical and current controls**
- ▶ Evaluate sequentially using repeated fitting

Optimal-Balanced Randomization

- ▶ Define “allocation probability balance” as a function of the **effective sample size** of the historical controls
- ▶ Morita, Thall, and Müller (2008) define prior effective sample size (ESS) of a parametric prior for non-adaptive models
- ▶ We expand their method for computing the **prior effective sample size of the intercepts & identically measured covariates** (previously β)
- ▶ Referred to as the *effective number of historical controls*, EHC
- ▶ First find the **posterior distribution of the commensurability parameter**, τ ; then *EHC* is essentially *ESS* is computed for

$$p(\beta, \lambda, \hat{\tau} | y_0) \propto N_p \left(V^{-1} M, \frac{1}{\hat{\tau}} V^{-1} \right) p(\hat{\tau}),$$

where τ replaced with its **posterior median**, $\hat{\tau}$

Optimal-Balanced Randomization (cont'd)

- ▶ The new treatment **allocation probability**, or balance function, for the j th+1 new patient follows as

$$\delta_j = \frac{C_j + EHC_j}{T_j + C_j + EHC_j},$$

where T_j and C_j denote the **number of subjects randomized** to new treatment and control following the j th enrollment

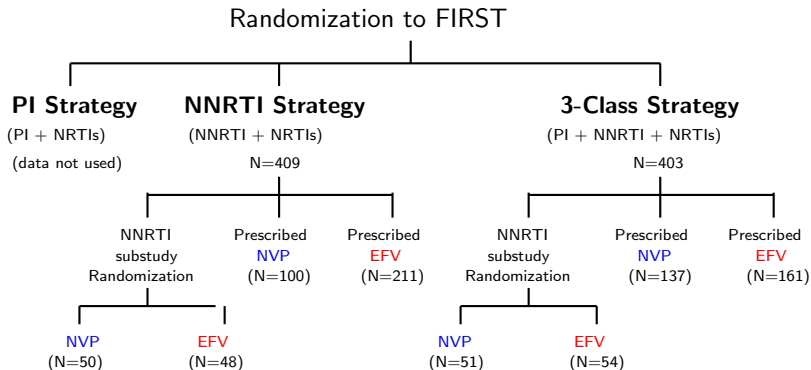
- ▶ This imposes **information balance** by encouraging optimal use of new patients relative to amount of incorporated prior information
- ▶ We suggest adjusting the allocation probability in blocks after an initial period where δ_j is fixed at 1/2

Adaptive Randomization Example

Flexible Initial Retrovirus Suppressive Therapies (FIRST) trial:

- ▶ This was a large, long-term, randomized, prospective comparison of three different antiretroviral strategies in highly active antiretroviral therapy-naive, HIV-1-infected persons (MacArthur et al., 2001)
- ▶ Patients within all three strategies were also assigned one or two nucleoside reverse transcriptase inhibitors (NRTIs)
- ▶ The three strategies for initial treatment were compared for long-term virological and immunological durability and safety, for the development of drug resistance, and for clinical disease progression
- ▶ Before randomization, patients within the two NNRTIs arms were given the option of preselecting the NNRTI drug, either nevirapine (NVP) or efavirenz (EFV), or allowing an additional randomization to NVP or EFV

Randomization Design



Outline of FIRST design and randomization for eligible subjects (Berg-Wolf et al., 2006).

Adaptive Randomization Example (cont'd)

- ▶ Current trial consists of data from the randomized arms of FIRST,
 1. **EFV** ($n = 99$) is the novel intervention and
 2. **NVP** ($n = 104$) the control therapy
- ▶ Historical controls consist of the **NVP** obs. arm ($n = 237$)
- ▶ Compare the prob. of virological suppression (HIV RNA < 50 copies/mL) under **EFV** and **NVP** at 32 weeks
- ▶ Use the commensurate prior probit regression model

Technical details:

$$\Phi^{-1}[\pi_0] = \mu_0 \text{ and } \Phi^{-1}[\pi(d)] = \mu + d\lambda$$

$$\hat{W}_{0jj} = \frac{\phi(X_0 \hat{\beta}_0)_{jj}^2}{\Phi(X_0 \hat{\beta}_0)_{jj} (1 - \Phi(X_0 \hat{\beta}_0)_{jj})},$$

where $\Phi()$ is the standard normal c.d.f. & $\phi(\cdot)$ is the standard normal p.d.f.

Adaptive Randomization Example (cont'd)

Simulating the adaptive randomization method proceeds iteratively:

- ▶ assign the j th+1 new patient to **NVP** or **EFV** with probability, δ_j
- ▶ generate response for several fixed values of $\Delta = \mu_0 - \mu$
- ▶ λ is fixed at 0.23 (MLE obtained from fitting the real current data from FIRST)
- ▶ equal allocation to **NVP** & **EFV** for first 80 assignments
- ▶ after randomizations 80, 100, 120, and 160, *EHC* is computed and δ_j is updated
- ▶ record *EHC* and δ_j after each randomization block, & total assigned to each group after all $n = 203$ patients are randomized
- ▶ $\Delta = -0.221$ is MLE obtained from fitting the real current data from FIRST

Optimal-Balanced Randomization Simulation

Simulation Averages for adaptive randomization to EFV, for $n = 203$, where $n_0 = 237$, Ratio= $\delta/(1 - \delta)$:1.

Rz	$\Delta = -2.25$ <u>EHC,Ratio</u>	$\Delta = -1.5$ <u>EHC,Ratio</u>	$\Delta = -0.75$ <u>EHC,Ratio</u>	$\Delta = -0.221$ <u>EHC,Ratio</u>	$\Delta = 0$ <u>EHC,Ratio</u>
80	1, 1.03:1	18, 1.42:1	177, 5.43:1	235, 6.86:1	237, 6.92:1
100	1, 1.02:1	15, 1.23:1	168, 3.74:1	235, 4.82:1	237, 4.85:1
120	1, 1.01:1	11, 1.11:1	164, 2.95:1	235, 3.80:1	237, 3.82:1
160	1, 1.01:1	5, 1.01:1	144, 2.02:1	235, 2.74:1	237, 2.75:1
203	<u>NVP, EFV</u> 101, 102	<u>NVP, EFV</u> 99, 104	<u>NVP, EFV</u> 77, 126	<u>NVP, EFV</u> 66, 137	<u>NVP, EFV</u> 65, 138

Is all of this “legal”?

- ▶ X-L Meng: **NO**: Agreement between the historical and concurrent data does **not** necessarily imply they are studying the same thing! External validation/information is **always** required!
- ▶ (Fairly hardcore) frequentist referee: **Not yet**: One needs to “evaluate theoretical properties, such as Bernstein-von Mises type results [showing asymptotic equivalence between Bayes and frequentist results] to provide some kind of universal assurances and/or insight beyond simulations into when gains can be expected.”
- ▶ BP Carlin: **Yes**: Point taken, but our goal here was never to “match” frequentist results, and moreover the frequentist simulations we have carried out speak for themselves and are “Bayesianly justified” in the sense of Rubin (1984)!

Conclusion

- ▶ *Adaptive* borrowing is consistent with recent arguments on behalf of the ease and desirability of adaptivity in Bayesian clinical trials generally; c.f. [new Chapman and Hall textbook by Berry, Carlin, Lee, and Müller \(2010\)](#)!

Thanks

- ▶ Mr. Brian Hobbs, Van Ryzin Award-winning University of Minnesota PhD, now postdoc at MD Anderson Cancer Center
- ▶ Drs. Telba Irony, Peter Müller, Scott Berry, Luis Pericchi, Joseph Ibrahim, Brian Neelon, and James O'Malley for helpful discussions
- ▶ Drs. Daniel Sargent and Sumithra Mandrekar for permission to analyze the Mayo colorectal dataset
- ▶ Drs. Xiaoxi Zhang and Laura Cisar for permission to use the Pfizer colorectal dataset