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

**Brian P. Hobbs**[1,*], **Bradley P. Carlin**[2], **Sumithra J. Mandrekar**[3], and **Daniel J. Sargent**[3]

[1] Department of Biostatistics, M.D. Anderson Cancer Center, Houston, TX, U.S.A

[2] Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, U.S.A

[3] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, U.S.A

## Summary

Bayesian clinical trial designs offer the possibility of a substantially reduced sample size, increased statistical power, and reductions in cost and ethical hazard. However when prior and current information conflict, Bayesian methods can lead to higher than expected Type I error, as well as the possibility of a costlier and lengthier trial. This motivates an investigation of the feasibility of hierarchical Bayesian methods for incorporating historical data that are adaptively robust to prior information that reveals itself to be inconsistent with the accumulating experimental data. In this paper, we present several models that allow for the commensurability of the information in the historical and current data to determine how much historical information is used. A primary tool is elaborating the traditional power prior approach based upon a measure of commensurability for Gaussian data. We compare the frequentist performance of several methods using simulations, and close with an example of a colon cancer trial that illustrates a linear models extension of our adaptive borrowing approach. Our proposed methods produce more precise estimates of the model parameters, in particular conferring statistical significance to the observed reduction in tumor size for the experimental regimen as compared to the control regimen.

## Keywords

Adaptive Designs; Bayesian; Colorectal Cancer; Clinical Trials; Power Priors

## 1. Introduction

Recent years have seen a dramatic increase in the use of Bayesian methods in the design, interim monitoring, and final analysis of clinical trials. By offering a formal statistical framework for incorporating all sources of knowledge (structural constraints, expert opinion, and both historical and experimental data), these methods offer the possibility of a substantially reduced sample size thanks to their more efficient use of information. This in turn typically leads to increases in statistical power and reductions in cost and ethical hazard, the latter since fewer patients need be exposed to the inferior treatment. On the other hand, the Bayesian approach carries some disadvantages when two or more sources of information conflict. In such cases, this can lead to higher than expected Type I error, as well as the possibility of a costlier and lengthier trial, since extra experimental information will be needed to resolve the conflict.

A commonly available (but often unincorporated) source of information in clinical trials is *historical data*. Such data may be available to the investigator based on previous studies in similar populations, or may simply be taken from the published summaries of other investigators. Even when no information related to a novel treatment is available, our understanding of the "standard care" group in a trial can almost always be augmented by existing information. Such borrowing of strength from historical data has long been encouraged in the case of medical device trials by the Center for Devices and Radiological Health (CDRH) at the U.S. Food and Drug Administration (FDA); (http://www.fda.gov/cdrh/osb/guidance/1601.html).

One specific challenge involves the problem of borrowing strength from a single historical study (be it in a control or a treatment group). As described by Irony (2008), simply fitting hierarchical models that borrow strength across levels in the usual way in such settings is overly sensitive to the hyperprior distribution on the variance parameters that control the amount of cross-study borrowing. That is, with only one historical study, there is no way to reliably estimate cross-study variability, and so this information must be imparted by the modeler – sometimes with drastic results for power and Type I error. This has led the FDA to *preclude* the use of historical information in such settings, even though this is clearly suboptimal in terms of cost and efficiency.

These issues motivate an investigation of the feasibility of hierarchical Bayesian borrowing of strength in such settings. The goal of such an approach is to determine a "sensible" amount of strength to borrow from the historical data that strikes a balance between increased cost-efficiency and long-run statistical integrity. Put another way, methods for incorporating historical data that are "adaptively robust" to prior knowledge that turns out to be inconsistent with the accumulating experimental data would be highly desirable. By contrast, we seek to utilize historical information given strong evidence of commensurability with the current.

In this paper, we propose various classes of *commensurate priors* as a solution to this problem. Section 2 introduces novel modifications to the traditional power prior approach which use a measure of commensurability among the historical and current data to guide the modeling. Then in Section 3 we propose several alternative hierarchical models for which borrowing depends upon some sensible measure of commensurability. Section 4 compares the frequentist performance of several of the proposed methods using simulation, while Section 5 offers an example from a colon cancer trial that illustrates the benefit of our proposed adaptive borrowing approach. Finally, Section 6 concludes and discusses our findings.

## 2. Hierarchical Power Priors

We begin with a review of power priors for general univariate models. Introduced by Ibrahim and Chen (2000), power priors offer a simple way to incorporate and downweight historical data, by raising the historical likelihood to a power $\alpha_0 \in [0, 1]$, and restandardizing the result to a proper distribution. These priors have been applied in a variety of contexts, including the sample size estimation problem by DeSantis (2007).

Let $D$ denote data from the current study and $L(\theta|D)$ the general likelihood function of the current data, where $\theta$ is the parameter of interest. Adopting the notation of Ibrahim and Chen (2000), denote the historical data by $D_0$, and the historical likelihood by $L(\theta|D_0)$. The conditional power prior for parameter $\theta$ is defined as

$$\pi(\theta|D_0, \alpha_0) \propto L(\theta|D_0)^{\alpha_0} \pi_0(\theta), \tag{1}$$

for initial prior $\pi_0(\theta)$ and *power parameter* $\alpha_0 \in [0, 1]$. (Throughout the paper we generically denote priors by $\pi$ and posteriors by $q$.) The power parameter controls the "degree of borrowing": if $\alpha_0 = 0$, (1) reduces to the initial prior (no borrowing), whereas if $\alpha_0 = 1$, equation (1) returns the usual historical posterior (full borrowing).

In the case of normal historical data, $x_{0i} \stackrel{iid}{\sim} Normal(\theta, \sigma_0^2)$, $\sigma_0^2$ known, $i = 1, \dots, n_0$, under a flat initial prior, (1) yields a $Normal(\bar{x}_0, \sigma_0^2/(\alpha_0 n_0))$ power prior distribution for $\theta$. Hence $\alpha_0$ plays the role of a relative precision parameter for the historical data. Since $0 \le \alpha_0 \le 1$, we might also think of $\alpha_0 n_0$ as the "effective" number of historical controls being incorporated into our analysis. Ibrahim and Chen (2000) introduced power priors to the broad statistical community, and illustrated their usefulness in a variety of settings; see also Ibrahim, Chen, and Sinha (2003), Chen and Ibrahim (2006), and Neelon, O'Malley, and Margolis (2008).

If we are willing to specify a particular value for $\alpha_0$, the conditional posterior distribution for $\theta$ given $D_0$, $D$, and $\alpha_0$ emerges as

$$q(\theta|D_0, D, \alpha_0) \propto \pi_0(\theta)L(\theta|D_0)^{\alpha_0}L(\theta|D) \tag{2}$$

Again in the case of known-variance normal observations, $x_i \stackrel{iid}{\sim} N(\theta, \sigma^2)$, $i = 1, \dots, n$, this results in another normal distribution for the posterior of $\theta$. We may be able to use the power parameter's interpretation as "importance of each historical patient relative to each new patient" to select a value for $\alpha_0$ (say, 1/2 or 1/3) for approximately Gaussian likelihoods.

More commonly, however, we are uncertain as to the degree to which our new data will agree with the historical data, and thus somewhat reluctant to prespecify the degree of borrowing. In such cases, we can enable the data to help determine probable values for $\alpha_0$ by adopting the usual Bayesian solution of choosing a *hyperprior* $\pi(\alpha_0)$ for $\alpha_0$.

## 2.1 Modified Power Priors

Ibrahim and Chen (2000) propose joint power priors consisting of the product of the conditional power prior in (1) and an independent proper prior on $\alpha_0$. Duan, Ye, and Smith (2006, p.98) caution against this since it violates the Likelihood Principle (Birnbaum, 1962). Duan et al. (2006), Neuenschwander et al. (2009), and Pericchi (2009) modify the joint power prior to the product of the *normalized* conditional power prior (1) and an independent proper prior for $\alpha_0$, producing the *modified power prior* (MPP)

$$\pi^{MPP}(\theta, \alpha_0|D_0) \propto \frac{L(\theta|D_0)^{\alpha_0}\pi_0(\theta)}{\int L(\theta|D_0)^{\alpha_0}\pi_0(\theta)d\theta}\pi(\alpha_0). \tag{3}$$

Modified power priors obey the Likelihood Principle and produce marginal posteriors for $\alpha_0$ that are proportional to products of familiar probability distributions. If we specify $\pi(\alpha_0)$ as a *Beta(a, b)* distribution for fixed positive hyperparameters $a$ and $b$, then the likely degree of borrowing from the historical data is controlled by $a$ and $b$: ($a = 10$, $b = 1$) would strongly encourage borrowing, ($a = 1$, $b = 10$) would strongly discourage it, and ($a = b = 1$) would be agnostic on the subject, essentially letting the data determine the degree of borrowing.

## 2.2 Commensurate Power Priors

A problem with modified joint power priors is that they do not directly parametrize the commensurability of the historical and new data. For example, note that the full conditional posterior distribution for $\alpha_0$, obtained by multiplying (3) by $L(\theta|D)$, is free of the current data $D$. Furthermore, Neelon and O'Malley (2010) caution against using Ibrahim-Chen and modified power priors since they both tend to overattenuate the impact of the historical data, forcing the use of fairly large $\alpha_0$ (or in our case, fairly informative hyperpriors for $\alpha_0$) in order to deliver sufficient borrowing. In fact, under a flat $Beta(1, 1)$ prior on $\alpha_0$, the marginal posterior for $\alpha_0$ is flat for two *identical* datasets regardless of the sample sizes.

In this subsection, as a solution to the problems raised by the aforementioned authors we propose a novel adaptive modification to the basic power prior formulation that adjusts the power parameter prior conditionally through a measure of the degree to which the historical and current data are commensurate. Heretofore both the historical and current data depended on a common parameter $\theta$. Now we assume *different* parameters in the historical and current group, $\theta_0$ and $\theta$, respectively, where $\theta \in \Re$ and $\theta_0 \in \Re$ are continuous location parameters. This dichotomous parameterization allows us to extend the hierarchical model to include a parameter that measures the evidence for commensurability between $\theta$ and $\theta_0$. Suppose we pick a vague (or even flat) initial prior $\pi_0(\theta_0)$, but construct the prior for $\theta$ to be normal with mean $\theta_0$ and precision $\tau$, where $\tau$ parametrizes commensurability. We can use the information in $\tau$ to guide the prior on $\alpha_0$. Specifying a vague prior for $\tau$ or $\log(\tau)$ and normalizing with respect to $\theta_0$ results in a power prior of the form,

$$\pi^{LCPP}(\theta, \alpha_0, \tau|D_0) \propto \int \frac{[L(\theta_0|D_0)]^{\alpha_0}}{\int [L(\theta_0|D_0)]^{\alpha_0} d\theta_0} d\theta_0 \times N(\theta|\theta_0, \tfrac{1}{\tau}) \times Beta(\alpha_0|g(\tau), 1) \times \pi(\tau),$$

(4)

where $g(\tau) > 0$ is a function of the commensurability parameter that is small for $\tau$ close to zero and large for large values of $\tau$. Since inference on $\theta_0$ is not of primary interest in the current analysis, we integrate it out of the joint prior. Specifying $\tau$ as a normal precision parameter offers a clear interpretation of commensurability. When evidence for commensurability is weak, $\tau$ is forced towards zero, increasing the conditional prior variance of $\theta$ by $\frac{1}{\tau}$. Therefore, we refer to $\tau$ as the *commensurability* parameter, and to the prior in (4) as *location commensurate power prior* (LCPP), since borrowing strength from the historical study depends upon the evidence in the data for commensurability between the location parameters $\theta$ and $\theta_0$. This extended power prior model requires the estimation of more parameters from the data (notably $\tau$), but we can formulate the model such that the information gained is aimed directly at improving estimation of the crucial borrowing parameter $\alpha_0$.

## 2.3 Single Arm Trial

Now let us turn our attention to the Gaussian case, where $\mu$, $\mu_0$ and $\sigma^2$, $\sigma_0^2$ parameterize the current and historical means and variances, respectively. Suppose a historical study suggests a true treatment benefit for a particular intervention, where one continuous response with mean $\mu$ can be collected for each subject. Investigators are interested in testing the point null hypothesis $H_0: \mu = 0$ in a new single arm trial. Let $\mathbf{x_0} = (x_{01}, \ldots, x_{0n0})'$ denote the independent and identically distributed historical responses, and assume $x_{0i} \overset{iid}{\sim} Normal(\mu_0, \sigma_0^2)$. If no initial information exists for $\mu_0$, we would likely select $\pi(\mu_0)$ vprop; 1 and hence the historical posterior follows as $q_0(\mu_0|\mathbf{x_0}, \sigma_0^2) = Normal(\overline{x}_0, \frac{\sigma_0^2}{n_0})$. Next let $\mathbf{x} = (x_1, \ldots, x_n)'$ denote the vector of independent and identically distributed responses from

the new study and assume $x_i \overset{iid}{\sim} Normal(\mu, \sigma^2)$ and $\pi_0(\mu)$ vprop; 1. The power prior formulation requires that $\sigma_0^2$ be fixed and known, therefore, we replace $\sigma_0^2$ in the historical likelihood with its maximum likelihood estimate, $\widehat{\sigma_0^2} = (\frac{1}{n_0}) \sum_{i=1}^{n_0} (x_{0i} - \bar{x}_0)^2$. We also assume the noninformative reference prior for $\sigma^2$, thus $q(\sigma^2|\mathbf{x_0}, \mathbf{x}) \propto \Gamma^{-1}(\frac{n}{2}, \frac{n}{2}[s^2 + (\bar{x} - \mu)^2])$ for both power prior models. The conditional posterior for $\mu$ becomes

$$q(\mu|\mathbf{x_0}, \mathbf{x}, \alpha_0, \sigma^2) = Normal\left(\frac{\widehat{\sigma_0^2} n\bar{x} + \sigma^2 \alpha_0 n_0 \bar{x}_0}{n\widehat{\sigma_0^2} + \sigma^2 \alpha_0 n_0}, \frac{\sigma^2 \widehat{\sigma_0^2}}{n\widehat{\sigma_0^2} + \sigma^2 \alpha_0 n_0}\right). \tag{5}$$

The conditional posterior for the full (no) borrowing model is found by fixing $\alpha_0 = 1$ ($\alpha_0 = 0$) in (5), and thus has mean $\frac{n\bar{x}\widehat{\sigma_0^2} + n_0 \bar{x}_0 \sigma^2}{n_0 \sigma^2 + n\widehat{\sigma_0^2}}$ ($\bar{x}$), and variance $\frac{\sigma^2 \widehat{\sigma_0^2}}{n_0 \sigma^2 + n\widehat{\sigma_0^2}}$ ($\frac{\sigma^2}{n}$). Suppose we assume $\alpha_0 \sim Beta(a, b)$, where $\alpha_0 \in [0, 1]$ and $a, b > 0$. The joint power prior and marginal posterior for $\alpha_0$ under the MPP approach are

$$\pi^{MPP}(\mu, \sigma^2, \alpha_0|\mathbf{x_0}) \propto Normal\left(m|\bar{x}_0, \frac{\widehat{\sigma_0^2}}{\alpha_0 n_0}\right) \times Beta(\alpha_0|a, b) \times \frac{1}{\sigma^2} \tag{6}$$

and

$$q^{MPP}(\alpha_0|\mathbf{x_0}, \mathbf{x}) \propto Normal\left(\bar{x} - \bar{x}_0|0, \frac{\sigma^2}{n} + \frac{\widehat{\sigma_0^2}}{\alpha_0 n_0}\right) \times Beta(\alpha_0|a, b). \tag{7}$$

Our location commensurate power prior (LCPP) follows as

$$\pi^{LCPP}(\mu, \alpha_0, \tau, \sigma^2|\mathbf{x_0}) \propto N(\mu|\bar{x}_0, \frac{1}{\tau} + \frac{\widehat{\sigma_0^2}}{\alpha_0 n_0}) \times Beta(\alpha_0|g(\tau), 1) \times \frac{1}{\sigma^2} \times \pi(\tau). \tag{8}$$

Notice that, in addition to adding $\frac{1}{\tau}$ to the conditional prior variance of $\mu$, the LCPP in (8) also inflates the estimated posterior variance of $\mu_0$, $\frac{\widehat{\sigma_0^2}}{n_0}$, by a factor of $\frac{1}{\alpha_0}$. The posterior distribution is obtained from the product of the LCPP prior (8) and the normal likelihood for $\mathbf{x}$. The full conditional posterior distribution for $\mu$ and marginal posterior distribution for $\alpha_0$ and $\tau$ follow as

$$q^{LCPP}(\mu|\mathbf{x_0}, \mathbf{x}, \alpha_0, \tau, \sigma^2) \propto N\left(\mu|\frac{\alpha_0 n_0 \tau \sigma^2 \bar{x}_0 + nu\bar{x}}{\alpha_0 n_0 \tau \sigma^2 + nu}, \frac{u\sigma^2}{\alpha_0 n_0 \tau \sigma^2 + nu}\right), \tag{9}$$

where $u = \alpha_0 n_0 + \widehat{\sigma_0^2}\tau$, and

$$q^{LCPP}(\alpha_0, \tau | \mathbf{x_0}, \mathbf{x}, \sigma^2) \propto N\left(\bar{x} - \bar{x}_0 | 0, \frac{\sigma^2}{n} + \frac{1}{\tau} + \frac{\widehat{\sigma}_0^2}{\alpha_0 n_0}\right) \times Beta(\alpha_0 | g(\tau), 1) \times \pi(\tau).$$

(10)

Information in the data pertaining to $\tau$ can be used to guide the $Beta(g(\tau), 1)$ prior on our power parameter, $\alpha_0$. Since $\tau$ becomes extremely large (small) for highly commensurate (conflicting) data we work on the log-scale. In Section 4 we present simulated frequentist operating characteristics for the LCPP model using $g^\star (\log(\tau)) = \max (\log(\tau), 1)$. Thus as $\tau$ increases and $g^\star (\log(\tau))$ becomes large relative to the second beta hyperparameter (fixed at 1) the distribution of $\alpha_0$ becomes increasingly peaked at 1 and the prior variance of $\mu$ tends towards $\frac{\widehat{\sigma}_0^2}{n_0}$, the approximate posterior variance of $\mu_0$ given $\mathbf{x_0}$. Thus, our model strongly encourages borrowing from the historical data when the data are commensurate. Alternatively, for $\tau \le 1$ the distribution of $\alpha_0$ becomes flat across the unit interval. Lastly, the LCPP model used in the simulation study below assumes a $Cauchy(0, 30)$ prior on $\log(\tau)$. The fat tails of Cauchy facilitate very large and small values of $\tau$, which provide for a highly flexible model. Fúquene, Cook, and Pericchi (2009, p.820) endorse the use of robust priors in hierarchical models to prevent unbounded and undesirable shrinkages.

See Web Appendix A for further discussion on the power and commensurability parameters under the Ibrahim-Chen, MPP, and LCPP approaches. Web Figure 1 compares marginal posterior distributions for $\alpha_0$ under the MPP and Ibrahim-Chen power prior models, while Web Figure 2 compares the marginal posterior and prior distributions for $\alpha_0$ and $\log(\tau)$ under the LCPP model.

## 2.4 Extension to Linear Models

Commensurate power priors are vastly more useful in clinical trials if they can be used in association with linear models. Ibrahim and Chen (2000) propose a framework for using power priors in GLMs. We formulate our own commensurate power prior linear model. Let us assume $y_0$ is a vector of $n_0$ responses from subjects in a previous investigation of an intervention that is to be used as a control in a current trial testing a newly developed intervention for which no reliable prior data exists. Let $y$ be the vector of $n$ responses from subjects in the current trial in both treatment and control arms. Suppose that both trials are designed to identically measure p−1 covariates of interest. Let $X_0$ be an $n_0 \times p$ design matrix and $X$ be an $n \times p$ design matrix, both of full column rank $p$, such that the first columns of $X_0$ and $X$ are vectors of 1s corresponding to the intercept. Now suppose $y_0 \sim N_{n_0}(X_0\beta_0, \sigma_0^2)$ 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, as well as an indicator for the new treatment. Let $D_0 = (y_0, X_0, n_0, p)$ and $D = (y, X, Z, n, p, r)$.

We can design a commensurate power prior (CPP) model to adaptively borrow strength from the historical control group and identical covariates. Specifying our commensurate power prior as in (4) with the same priors on $\sigma^2$, $\alpha_0$, and $\log(\tau)$ as in the previous sub-section, a flat prior on $\lambda$, and integrating $\beta_0$ out of the joint prior leads to a full conditional prior on $\beta$ that is normal with mean $V^{-1}M$ and covariance $(\tau V)^{-1}$, where

$M = \left(\frac{\alpha_0 X_0^T X_0}{\widehat{\sigma}_0^2} + \tau I_p\right)^{-1} \frac{\alpha_0 X_0^T X_0 \widehat{\beta}_0}{\widehat{\sigma}_0^2}$, $V = I_p - \tau \left(\frac{\alpha_0 X_0^T X_0}{\widehat{\sigma}_0^2} + \tau I_p\right)^{-1}$, and $\widehat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T y_0$. The joint posterior follows by multiplying the joint prior by the likelihood of $y$ and normalizing. The full conditional posteriors for $\lambda$ and $\sigma^2$ and posterior for $\beta$ given $\sigma^2$, $\alpha_0$, and $\tau$ follow as,

$$q^{CPP}(\lambda|D_0, D, \beta, \sigma^2) \propto N_r\left(\lambda|\widehat{\lambda}, \sigma^2(Z^T Z)^{-1}\right),$$

(11)

$$q^{CPP}(\sigma^2|D_0, D, \beta, \lambda) \propto \Gamma^{-1}\left(\sigma^2|\frac{n}{2}, \frac{(y - X\beta - Z\lambda)^T(y - X\beta - Z\lambda)}{2}\right),$$

(12)

$$q^{CPP}(\beta|D_0, D, \sigma^2, \alpha_0, \tau) \propto N_p\left(\beta|\Lambda^{-1}\left(\tau M + \frac{X^T(I_n - w)y}{\sigma^2}\right), \Lambda^{-1}\right),$$

(13)

where $\hat{\lambda} = (Z^T Z)^{-1} Z^T (y - X\beta)$, $\Lambda = \tau V + \frac{X^T(I_n - w)X}{\sigma^2}$, and $w = Z(Z^T Z)^{-1} Z^T$. Notice in (11) that the full conditional posterior mean for $\lambda$, $\hat{\lambda}$, is a function of residuals $(y - X\beta)$, whereas the conditional posterior mean of $\beta$ in (13) is an average of the historical and concurrent data relative to the power and commensurability parameters, $\alpha_0$ and $\tau$. As $\tau$ and $\alpha_0$ approach zero, the marginal posterior for $\beta$ converges to a normal density with mean

$\left(\frac{X^T X - X^T Z(Z^T Z)^{-1} Z^T X}{\sigma^2}\right)^{-1}(X^T y - X^T Z(Z^T Z)^{-1} Z^T y)$ and variance $\left(\frac{X^T X - X^T Z(Z^T Z)^{-1} Z^T X}{\sigma^2}\right)^{-1}$, recovering the result from a linear regression that ignores all of the historical data. In this case, $\hat{\lambda}$ also converges to the no borrowing estimate of the treatment difference.

## 3. Commensurate Priors

Heretofore, we have considered posterior estimation of commensurability to be adjuvant for the purpose of facilitating more borrowing of strength in conjunction with a hierarchical power prior model. Yet, for Gaussian data, (8) reveals that both power and commensurability parameters inflate the conditional prior variance of $\mu$ given weak evidence for commensurability. In this section we consider hierarchical models that incorporate commensurate priors as the primary mechanism for weighting the influence of prior information relative to its consistency with data from the concurrent study.

Again let $D_0$ and $D$ denote data from the historical and current studies and $L(\theta_0|D_0)$ and $L(\theta|D)$ the general likelihood functions, respectively, where $\theta$ is the parameter of interest. The dichotomous parameterization facilitates estimation of commensurability among the historical and current data in a hierarchical model by specifying the prior for $\theta$ to be "centered" at $\theta_0$ and conditional on $\tau > 0$, where $\tau$ parameterizes prior precision for $\theta$ given $\theta_0$. Multiplying by the historical likelihood function results in a prior of the form,

$$\pi(\theta|D_0, \theta, \tau) \propto L(\theta_0|D_0)\pi(\theta|\theta_0, \tau)\pi_0(\theta).$$

(14)

As $\tau$ approaches zero, $\pi(\theta|D_0, \theta_0, \tau) \to \pi_0(\theta)$, effectively ignoring the historical data. On the other hand, as $\tau \to \infty$, $\theta$ approaches $\theta_0$ and $\pi(\theta|D_0, \theta_0, \tau) \to L(\theta|D_0)\pi_0(\theta)$, recovering the result obtained from pooling the two datasets. If the historical study favors rejecting the current null, then decreasing $\tau$ reduces Type I error, while increasing $\tau$ increases power.

The posterior kernel is obtained by multiplying (14) by the current likelihood $L(\theta|D)$. Note that the full conditional posterior distribution for $\theta_0$ would be independent of the current data, since the current likelihood would be nothing but a multiplicative constant. Therefore, $\theta_0$ should be integrated out of the prior when the $\int L(\theta_0|D_0) \pi(\theta|\theta_0)d\theta_0$ is tractable.

Specifying a vague prior for $\tau$, $\pi(\tau)$, and adopting the joint prior $\pi(\theta, \tau|D_0, \theta_0) \propto \pi(\theta|D_0, \theta_0, \tau)\pi(\tau)$ allows the model to utilize data from the current trial to help estimate $\tau$.

Now let us return to the Gaussian case in the context of the simple trial outlined in Subsection 2.3. We again assume the noninformative reference prior for $\sigma^2$, thus $q(\sigma^2|\mathbf{x_0}, \mathbf{x}) \propto \Gamma^{-1}(\frac{n}{2}, \frac{n}{2}[s^2+(\overline{x}-\mu)^2])$. First let us construct a prior on $\mu$ such that borrowing strength from the historical study depends upon the evidence in the data for commensurability among location parameters $\mu$ and $\mu_0$. We start by assuming the joint prior on $\mu$ and $\mu_0$ is the product of the historical likelihood and a normal prior on $\mu$ with mean $\mu_0$ and precision $\tau$. Since posterior inference on the scale of the historical data is not of direct interest we again replace $\sigma_0^2$ in the historical likelihood with its maximum likelihood estimate. Integrating out the nuisance parameter, $\mu_0$, leads to the *location commensurate prior* (LCP),

$$\pi^{LCP}(\mu, \sigma^2, \tau|\mathbf{x_0}) \propto N(\mu|\overline{x}_0, \frac{1}{\tau}+\frac{\widehat{\sigma}_0^2}{n_0}) \times \frac{1}{\sigma^2} \times \pi(\tau).$$

(15)

If evidence for commensurability is weak (ie, $\tau$ is close zero), the conditional prior variance of $\mu$ in (15) is increased by $\frac{1}{\tau}$. Assuming a vague prior on $\tau$ completes the prior specification. The posterior is proportional to the product of the joint prior in (15) and the current data likelihood. In Section 4 we present simulations to illustrate the frequentist operating characteristics of the LCP using an agnostic *Uniform*$(-30, 30)$ prior on $\log(\tau)$.

Suppose instead we want borrowing from the historical data to depend upon evidence for commensurability among both the location *and* scale parameters. We must now extend the hierarchical model to include a parameter, $\gamma$, that measures evidence of commensurability among $\sigma^2$ and $\sigma_0^2$ by specifying a prior on $\sigma^2$ that is "centered" at $\sigma_0^2$ and has precision $\gamma$. An obvious choice would be to assume an inverse gamma prior on $\sigma^2$, with mean $\sigma_0^2$ and precision $\gamma$. Assuming the reference initial prior on $\sigma_0^2$, multiplying by the historical likelihood, and integrating out $\mu_0$ results in the conditional *location-scale commensurate prior* (LSCP),

$$\pi^{LSCP}(\mu, \sigma_0^2, \sigma^2|\mathbf{x_0}, \tau, \gamma) \propto N\left(\mu|\overline{x}_0, \frac{1}{\tau}+\frac{\sigma_0^2}{n_0}\right)\Gamma^{-1}\left(\sigma^2|\alpha^\star, \beta^\star\right)\Gamma^{-1}\left(\sigma_0^2|\frac{n_0-1}{2}, \frac{n_0\widehat{\sigma}_0^2}{2}\right),$$

(16)

where $\alpha^\star=\gamma\sigma_0^4+2$ and $\beta^\star=\sigma_0^2(\gamma\sigma_0^4+1)$. One can assume vague priors for the commensurability parameters, and proceed with posterior inference on $\mu$ and $\sigma^2$. Borrowing of strength that requires commensurate scales in addition to commensurate locations is more cautious and perhaps more appealing to skeptics.

Another option involves fixing values of $\tau$ and $\gamma$ and using a *mixture* prior. Suppose we specify $m$ distinct relationships of interest among the locations and scales of the historical and current data represented by fixed pairs of the commensurability parameters, $(\tau_1, \gamma_1)$, …, $(\tau_m, \gamma_m)$. Let $\pi^{LCSP}(\mu, \sigma_0^2, \sigma^2|\mathbf{x_0}, \tau_j, \gamma_j)$ denote the conditional prior in (16) given $\tau$ and $\gamma$ are fixed at $\tau_j$ and $\gamma_j$ respectively. If we are also willing to specify fixed mixing proportions, $\omega_j$

$\in (0, 1)$ where $j = 1, \ldots, m$ such that $\sum_{j=1}^{m} \omega_j = 1$, we can formulate a LSCP for $\mu$, $\sigma_0^2$, and $\sigma^2$ that is a convex combination of the $m$ potential relationships of interest as

$$\pi^{LSCMP}(\mu, \sigma_0^2, \sigma^2 | \mathbf{x_0}) = \sum_{j=1}^{m} \omega_j \pi^{LSCP}(\mu, \sigma_0^2, \sigma^2 | \mathbf{x_0}, \tau_j, \gamma_j).$$

(17)

We refer to this prior as the *location-scale commensurate mixture prior* (LSCMP). In contrast to the non-mixture commensurate and power priors, the LSCMP does not permit posterior estimation of commensurability, but instead facilitates measured, "partial borrowing". The LSCMP featured in Section 4, an equal mixture of just two pairs of $(\tau, \gamma)$, corresponding to very high and low commensurability among the location and scale parameters, is shown to work rather well in our simulation study.

## 4. Simulation Results for a Single Arm Trial

In this section we investigate the Bayesian and frequentist operating characteristics of the models described in Sections 2 and 3 via simulation. For the power priors in Subsection 2.3, we used a MPP model with a *Beta*(1, 1) prior on $\alpha_0$ and a LCPP assuming a Cauchy prior on $\log(\tau)$ centered at zero with scale fixed at 30. Both power prior models use our noninformative reference prior on $\sigma^2$. Among the commensurate priors in Section 3, the simulations were run using a LCP model that assumes a vague, *Uniform*(−30, 30) prior on $\log(\tau)$ and noninformative reference prior on $\sigma^2$, and a LSCMP model that is a mixture of $(\tau_1, \gamma_1) = (10^6, 10)$ and $(\tau_2, \gamma_2) = (1/2, 1/2)$ with the mixing proportion fixed at 1/2. We also ran simulations on a model that ignores the historical data completely (by assuming the noninformative Jeffreys prior on $\mu$ and $\sigma^2$) and, following Fúquene et al. (2009), a model that assumes a "robust" Cauchy prior on $\mu$, centered at $\bar{x}_0$ with scale parameter fixed at 1. We will refer to these as the "no borrowing" and "Cauchy" models.

Figure 1 illustrates the adaptive movement capability of the commensurate prior methods. Each graph contains 95% posterior credible intervals for $\mu$ derived from all simulated models, generated with $n_0 = 60$ historical observations having $\bar{x}_0 = 0$ and $\widehat{\sigma}_0^2 = 1$. The top interval corresponds to results from the analysis of the historical data alone, using noninformative Jeffreys priors on $\mu_0$ and $\sigma_0^2$. The interval directly beneath it represents a pooled analysis using the full borrowing prior, and the bottom interval corresponds to the no borrowing analysis of the current data that ignores the historical data. Intervals in between from top to bottom correspond to the MPP, Cauchy, LCPP, LCP, and LSCMP.

Looking at graphs in the left column where the current data is the most inconsistent with the historical data, we see that intervals for the posteriors using commensurate priors and modified power priors are virtually identical to that for no borrowing. Conversely, the full borrowing prior, which contains no mechanism for acknowledging the obvious conflict, leads to a much tighter interval around the weighted average of the two sample means, $-\frac{2}{3}$. In fact, for Normal-Normal conjugate priors, increasing the sample sizes, $n$ and $n_0$, always decreases the length of an equal-tail credible interval, leading to high Type I error. The Cauchy prior interval is properly centered at −2, but much wider, suggesting the procedure may be somewhat conservative. The center graphs demonstrates the adaptive shrinkage capabilities of the commensurate, MPP, and Cauchy models for intermediately commensurate datasets which lead to good Type I error behavior. The current and historical datasets have identical sufficient statistics in the third column. Intervals for the MPP, LCPP, LCP models have narrowed to mirror the pooled result. Notice that the LCPP has narrowed

slightly more than the MPP, suggesting that the LCPP obtains more power. Intervals for the LSMCP are more reluctant to shrink towards the full borrowing result in the bottom right graph given the evidence for incommensurability among $\sigma_0^2$ and $\sigma^2$.

Next, to mimic the case of a single arm trial being run after a promising pilot study to test the null hypothesis that $\mu = 0$, let $\mu_0 = 0.5$ be the true mean of the historical data, and fix $n_0 = 60$ and $\sigma_0^2 = 1$. We can sample $x_{0i} \overset{iid}{\sim} N(\mu_0, \sigma_0^2)$ and $x_i \overset{iid}{\sim} N(\mu_0, \sigma^2)$ for some fixed true $\mu$, $\sigma^2$, and $n$, and conduct a two-sided test of the null hypothesis $\mu = 0$ using an equal-tail posterior credible interval. We repeated this entire process $Nrep = 5000$ times for current sample sizes of $n = 15$, 30, and 60 to compare frequentist Type I error and power properties across the various approaches.

Table 1 contains areas under the resulting power curves, where power was computed for true $\mu = 0, 0.05, 0.1, \ldots, 0.5$. All models are compared under two approaches to hypothesis testing. The first uses 95% posterior credible intervals to test the null, which for the adaptive methods maintain Type I error probability of 0.05 for the considered sample sizes only when $\mu_0 = \mu = 0$. The other approach is a "calibrated" one that controls Type I error at 0.05 for *all* considered sample sizes by using equal-tail posterior credible intervals with varying tail probabilities to test the null hypothesis. For discussion about controlling Type I error for Bayesians designs from a regulatory perspective see Pennello and Thompson (2008).

Notice that the LSCMP, MPP, LCP, and LCPP models result in higher area than the no borrowing model for all cases. Therefore, the adaptive approaches always facilitate more power than an analysis that ignores the historical data, even when Type I error is controlled. Furthermore, the table suggests that the LCPP approach is always more powerful than the MPP, although the difference is slight for controlled Type I error. Furthermore, the LCPP is most powerful for the analysis using 95% credible intervals, while the LCP emerges as slightly better when Type I error is controlled. Our simulations also suggest that the calibrated analysis for the Cauchy model provides very slight gains in power over an analysis that ignores the historical data. The reader should note that the calibrated (controlled Type I error) analysis for the Cauchy model used equal-tail credible intervals corresponding to tail probabilities larger than 0.025. Type I error results for the full borrowing model were extremely poor and hence excluded.

Table 2 contains Type I error probabilities for all models and sample sizes given in Table 1. Comparing results in the left columns of the two tables for all models reveals that larger area under the power curve corresponds to higher Type I error probability when using 95% posterior credible intervals to test the null hypothesis. Results in the right column of Table 2 show that Type I error is controlled at 0.05 for all models and sample sizes under the "calibrated" approach.

Lastly, power curves for the full borrowing (dot-dashed), LCPP (solid), MPP (dashed), and no borrowing (dotted) models are shown in Figure 2 to augment Table 1. The top row of plots corresponds to testing the null hypothesis that $\mu = 0$ using the 95% posterior credible intervals, while the bottom row contains power curves for the analysis with controlled Type I error. The power curve for the LCPP is clearly above the MPP across the top row of plots, and either overlapping or slightly larger for the bottom row. This suggests that the LCPP approach obtains more power than the MPP approach. Notice the atrocious Type I error resulting from pooling the two datasets in the top row, which achieves a *minimum* of approximately 0.8 when $n = 60$. The ill-fated full borrowing prior has "unbounded influence" resulting in a dogmatic analysis of the current trial; Fúquene et al. (c.f. 2009, p. 819).

## 5. Example Using Controlled Colorectal Cancer Trial Data

We consider data from two successive randomized colorectal cancer clinical trials originally reported by Saltz et al. (2000) and Goldberg et al. (2004). The initial trial randomized $N_0 = 683$ patients with previously untreated metastatic colorectal cancer between May 1996 and May 1998 to one of three regimens: Irinotecan alone; Irinotecan and bolus Fluorouracil plus Leucovorin (IFL); or a regimen of Fluorouracil and Leucovorin (5FU/LV) "standard therapy". In an intent-to-treat analysis, IFL resulted in significantly longer progression free survival and overall survival than Irinotecan alone and 5FU/LV (Saltz et al., 2000).

The subsequent trial compared three drug combinations in $N = 795$ patients with previously untreated metastatic colorectal cancer, randomized between May 1999 and April 2001. Patients in the first drug group received the current "standard therapy," the IFL regimen identical to that used in the historical study. The second group received Oxaliplatin and infused Fluorouracil plus Leucovorin (abbreviated FOLFOX), while the third group received Irinotecan and Oxaliplatin (abbreviated IROX); both of these latter two regimens were new as of the beginning of the second trial.

While both trials recorded many different patient characteristics and outcomes, in our analysis we concentrate on the trial's measurements of tumor size, and how the FOLFOX regimen compared to the IFL regimen. Therefore, the historical dataset will consist of the IFL treatment arm from the initial study, while the current data will consist of patients randomized to IFL or FOLFOX in the subsequent trial. We omit data from the Irinotecan alone and 5FU/LV arms in the Saltz study and the IROX arm in the Goldberg study.

Both trials recorded two bi-dimensional measurements on each tumor for each patient at regular cycles. The trial reported by Saltz et al. measured patients every 6 weeks for the first 24 weeks and every 12 thereafter weeks after, while the trial reported by Goldberg et al. measured every 6 weeks for the first 42 weeks. We computed the sum of the longest diameter in cm ("ld sum") for up to 9 tumors for each patient at each cycle, until the patient's final follow-up visit; for 92% of the patients in our very ill study population, this was the final observation prior to the patient's death. We then used the average change in ld sum from baseline to test for a significant treatment difference in ld sum reduction between the FOLFOX and control regimens. Our analysis will also incorporate baseline ld sum as a predictor as well as two important covariates identically measured at baseline: age in years, and aspartate aminotransferase (AST) in units/L.

We restricted our analysis to patients that had measurable tumors, at least two cycles of followup, and a nonzero ld sum at baseline, bringing the total sample size to 441: 171 historical and 270 current observations. Among the current patients, there are 129 controls (IFL) and 141 patients treated with the new regimen (FOLFOX). Suppose $y_0$ and $y$ are vectors of lengths $n_0$ and $n$ for the historical and concurrent responses such that

$$y_0 \sim Normal(X_0\beta_0, \sigma_0^2), \text{ and } y \sim Normal(X\beta+Z\lambda, \sigma^2) \tag{18}$$

where $X_0$ and $X$ are $n_0 \times 4$ and $n \times 4$ design matrices with columns corresponding to (1, ld sum at baseline, age, AST), and $Z$ is the FOLFOX indicator function. Thus the $\beta_0$ and $\beta$ parameters contain intercepts as well as regression coefficients for each of three baseline covariates, while $\lambda$ represents change in average ld sum attributed to FOLFOX. Web Figure 3 contains histograms of the average change in ld tumor sum from baseline.

The "historical data" and "current data" columns of Table 3 summarize results from separate classical linear regression fits on the historical ($y_0$, $X_0$) and current ($y$, $X$) data alone. The "current data" values constitute the "no borrowing" analysis. Results from both datasets suggest that ld sum at baseline is highly significant while age and AST are not. Furthermore, while the estimated intercept corresponding to FOLFOX in the current data is negative, −0.413, the estimate is not precise enough to conclude a significant treatment difference at the 0.05 level.

Information about $\beta_0$ appears to be relevant to $\beta$. Therefore, we implemented the commensurate power prior linear model presented in detail in Subsection 2.4 which borrows strength adaptively relative to the degree to which ($y_0$, $X_0$) is commensurate with ($y$, $X$). Point estimates (posterior medians) and 95% equal-tail Bayesian credible intervals are given in the bottom portion of Table 3. First, notice that the posterior for $\alpha_0$ is peaked at 1. Therefore, our power prior linear model considers the historical and current data to be quite commensurate, increasing the precision of the parameter estimates. As a result, the 95% credible interval upper bound for $\lambda$ is now less than zero, and so we can now conclude that FOLFOX resulted in a significant reduction in average ld sum when compared to the IFL regimen. This finding is consistent with those of Goldberg et al. (2004), who determined FOLFOX to have superior time to progression and response rate compared to IFL.

## 6. Discussion

In this paper, we have presented classes of hierarchical models using priors that facilitate adaptive borrowing from historical data when this is justified by its commensurability with the accumulating current data. Such *adaptive* borrowing is consistent with recent arguments on behalf of the ease and desirability of adaptivity in Bayesian clinical trials generally (Berry, 1993, 2006; Berry et al., 2010). The approach was shown to work well both with simulated and actual data, the latter based on two recent studies in colorectal cancer.

Before using the proposed linear model in the context of a new clinical trial, investigators must consider carefully the design (ie, randomized versus single-arm) of the historical study. Differences in patient populations between the historical and new study and other known/ unknown confounding factors can be potential sources of bias when borrowing from the historical data. Furthermore, commensurate priors require extra care if the sampling distributions differ; estimating nuisance parameters like $\sigma^2$ becomes ever more challenging if, say, the two distributions were normal and Student's t, respectively.

Future work looks toward extending our approach to non-Gaussian settings, especially those involving categorical and time-to-event data. We are also currently pursuing the use of commensurate priors for adaptive borrowing that allows the sample size or allocation ratio in the ongoing trial to be altered if this is warranted. For example, if historical and concurrent controls emerge as commensurate, we might randomize fewer patients to the control group, thus enhancing the efficiency of the ongoing trial.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

# References

Berry DA. A case for Bayesianism in clinical trials (with discussion). Statistics in Medicine. 1993; 12:1377–1404. [PubMed: 8248653]

Berry DA. Bayesian clinical trials. Nature Reviews Drug Discovery. 2006; 5:27–36.

Berry, SM.; Carlin, BP.; Lee, JJ.; Müller, P. Bayesian Adaptive Methods for Clinical Trials. Chapman and Hall/CRC Press; Boca Raton, FL: 2010.

Birnbaum A. On the foundations of statistical inference (with discussion). Journal of the American Statistical Association. 1962; 57:269–326.

Chen MH, Ibrahim JG. The relationship between the power prior and hierarchical models. Bayesian Analysis. 2006; 1:554–571.

DeSantis F. Using historical data for Bayesian sample size determination. Journal of the Royal Statistical Society, Series A. 2007; 170:95–113.

Duan Y, Ye K, Smith EP. Evaluating water quality using power priors to incorporate historical information. Environmetrics. 2006; 17:95–106.

Fúquene JA, Cook JD, Pericchi LR. A case for robust Bayesian priors with applications to clinical trials. Bayesian Analysis. 2009; 4:817–846.

Goldberg RM, Sargent DJ, Morton RF, Fuchs CS, Ramanathan RK, Williamson SK, Findlay BP, Pitot HC, Alberts SR. A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. Journal of Clinical Oncology. 2004; 22:23–30. [PubMed: 14665611]

Ibrahim JG, Chen MH. Power prior distributions for regression models. Statistical Science. 2000; 15:46–60.

Ibrahim JG, Chen MH, Sinha D. On optimality properties of the power prior. Journal of the American Statistical Association. 2003; 98:204–213.

Irony, T. Personal communication. 2008.

Neelon, B.; O'Malley, AJ. CEHI Working Paper 2010–01. Duke University; Durham, NC: 2010. The use of power prior distributions for incorporating historical data into a Bayesian analysis.

Neelon, B.; O'Malley, AJ.; Margolis, PA. Bayesian analysis using historical data with application to pediatric quality of care. Proceedings of the 2008 Joint Statistical Meetings; 2008. p. 2960-2967.

Neuenschwander B, Branson M, Spiegelhalter DJ. A note of the power prior. Statistics in Medicine. 2009; 28:3562–3566. [PubMed: 19735071]

Pennello G, Thompson L. Experience with reviewing Bayesian medical device trials. Journal of Biopharmaceutical Statistics. 2008; 18:81–115. [PubMed: 18161543]

Pericchi, LR. Personal communication. 2009.

Saltz LB, Cox JV, Blanke C, et al. the Irinoteca Study Group. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. The New England Journal of Medicine. 2000; 343:905–914. [PubMed: 11006366]

**Figure 1.**
95% posterior credible intervals for $\mu$ for all simulated models where $\bar{x}_0 = 0$, $n_0 = 60$, $n = 30$, and $\widehat{\sigma}_0^2 = 1$. Each graph contains results from eight different models: historical (top), full borrowing (second), MPP (third), Cauchy (fourth), LCPP (fifth), LCP (sixth), LSCMP (seventh), and no borrowing (bottom). Columns correspond to the current sample mean, $\bar{x}$, while the top row shows results for $\widehat{\sigma}^2 = 1$ and the bottom row assumes $\widehat{\sigma}^2 = 3$.
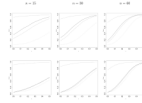
**Figure 2.**
Power curves for the full borrowing, commensurate power prior, modified power prior, and
no borrowing models where $\mu_0 = 0.5$, $n_0 = 60$, and $\sigma^2 = \sigma_0^2 = 1$. Each graph shows results for
the full borrowing (dot-dashed), LCPP (solid), MPP (dashed), and no borrowing (dotted)
models. Columns correspond to the indicated current sample size $n$. The top row
corresponds to hypothesis testing based on the respective 95% posterior credible intervals,
while the bottom row contains results for hypothesis tests that use varying equal-tail
posterior credible intervals such that the Type I Error is controlled at 0.05, except for the full
borrowing model for which controlling Type I error at 0.05 is impossible for the considered
scenarios.

**Table 1**

Area under the power curves: 95% posterior credible intervals (left); Controlled Type I error (right), for true $\mu = 0, 0.05, 0.1, \ldots, 0.5$ and $\mu_0 = 0.5$.

| | 95% Posterior CI | | | Controlled Type I Error | | |
|---|---|---|---|---|---|---|
| | $n = 15$ | $n = 30$ | $n = 60$ | $n = 15$ | $n = 30$ | $n = 60$ |
| No borrowing | 0.090 | 0.158 | 0.244 | 0.090 | 0.158 | 0.244 |
| *Cauchy* | 0.079 | 0.161 | 0.249 | 0.109 | 0.175 | 0.261 |
| *LSCMP* | 0.250 | 0.294 | 0.326 | 0.133 | 0.190 | 0.263 |
| *MPP* | 0.288 | 0.312 | 0.341 | 0.119 | 0.185 | 0.269 |
| *LCP* | 0.309 | 0.329 | 0.346 | 0.137 | 0.208 | 0.276 |
| *LCPP* | 0.344 | 0.351 | 0.362 | 0.122 | 0.196 | 0.278 |

**Table 2**

Type I error probabilities: 95% posterior credible intervals (left); Controlled Type I error (right), for true $\mu = 0$ and $\mu_0 = 0.5$.

| | 95% Posterior CI | | | Controlled Type I Error | | |
|---|---|---|---|---|---|---|
| | $n = 15$ | $n = 30$ | $n = 60$ | $n = 15$ | $n = 30$ | $n = 60$ |
| No borrowing | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| *Cauchy* | 0.025 | 0.035 | 0.046 | 0.050 | 0.050 | 0.050 |
| *LSCMP* | 0.175 | 0.168 | 0.126 | 0.050 | 0.050 | 0.050 |
| *MPP* | 0.247 | 0.177 | 0.130 | 0.050 | 0.050 | 0.050 |
| *LCP* | 0.266 | 0.209 | 0.148 | 0.050 | 0.050 | 0.050 |
| *LCPP* | 0.368 | 0.271 | 0.184 | 0.050 | 0.050 | 0.050 |

**Table 3**

Fits to colorectal cancer data: $n_0 = 171$ (left); $n = 270$ (right); LCPP (bottom).

| | Historical data | | Current data | |
|---|---|---|---|---|
| | estimate | 95% CI | estimate | 95% CI |
| Intercept | 0.88 | (−1.98, 3.74) | −0.47 | (−2.28, 1.34) |
| BL ld sum | −0.23 | (−0.31, −0.15) | −0.40 | (−0.45, −0.34) |
| Age | −0.02 | (−0.07, 0.02) | 0.01 | (−0.01, 0.04) |
| AST | 0 | (−0.02, 0.02) | 0.01 | (−0.01, 0.02) |
| FOLFOX | – | – | −0.41 | (−1.02, 0.19) |

| | LCPP | |
|---|---|---|
| | estimate | 95% Posterior CI |
| Intercept | 0.180 | (−1.11, 1.42) |
| BL ld sum | −0.39 | (−0.44, −0.33) |
| Age | 0 | (−0.02, 0.02) |
| AST | 0 | (−0.01, 0.01) |
| FOLFOX | −0.46 | (−0.82, −0.10) |
| $\alpha_0$ | 0.86 | (0.44, 1.00) |
| $\log(\tau)$ | 16.53 | (1.59, 82.00) |