BAYESIAN HIERARCHICAL MODELING FOR ADAPTIVE INCORPORATION OF HISTORICAL INFORMATION IN CLINICAL TRIALS

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

BRIAN PAUL HOBBS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BRADLEY P. CARLIN, Ph.D.

August, 2010

Acknowledgments

I am indebted to many people for their support and contributions to the production of this dissertation. First, to my advisor, Dr. Bradley P. Carlin, for the inspiration to pursue methods that facilitate adaptivity in the design and analysis of clinical trials. Without your vision and persistent, unrelenting support, this dissertation would not have been possible. Thank you Dr. Daniel J. Sargent, Dr. Sumithra J. Mandrekar, and Prof. James D. Neaton for providing invaluable guidance and contributing to the manuscript.

Thank you Prof. Andy Mugglin and Prof. Karen Kuntz for sharing your expertise and providing great comments and suggestions. Thank you Drs. Alan Gelfand, Telba Irony, Peter Müller, Scott Berry, Luis Raul Pericchi, Joseph Ibrahim, Brian Neelon, and James O'Malley for helpful discussions and guidance.

Thank you Drs. Xiaoxi Zhang and Laura Cisar from Pfizer for providing the Saltz et al. (2000) colorectal cancer dataset; Erin Green and Brian Bot from the Mayo Clinic, Rochester for your help with the Saltz et al. (2000) and Goldberg et al. (2004) datasets; and Grace Peng from the University of Minnesota, Minneapolis for your help with data from the MacArthur et al. (2001) study.

Most importantly, thank you Joanna for love and patience; Frankie for joy; family for support; David for inspiration.

Abstract

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. We develop several models that allow for the commensurability of the information in the historical and current data to determine how much historical information is used. First, we propose methods for univariate Gaussian data and provide an example analysis of data from two successive colon cancer trials that illustrates a linear models extension of our adaptive borrowing approach. Next, we extend the general method to linear and linear mixed models as well as generalized linear and generalized linear mixed models. We also provide two more sample analyses using the colon cancer data. Finally, we consider the effective historical sample size of our adaptive method for the case when historical data is available only for the concurrent control arm, and propose "optimal" use of new patients in the current trial using an adaptive randomization scheme that is balanced with respect to the amount of incorporated historical information. The approach is then demonstrated using data from a trial comparing antiretroviral strategies in HIV-1-infected persons. Throughout the thesis we present simulation studies that compare frequentist operating characteristics and highlight the advantages of our adaptive borrowing methods.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Bayesian statistical methods are being used increasingly in the design and 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. Clinical trials are not designed without consideration of earlier results from similar investigations. *Historical data*, data from previous studies in similar populations, can be incorporated prospectively to enhance the efficiency of an ongoing trial and provide increased precision of parameter estimates. 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 information derived from previous investigations. In the interest of safety, device trials such as the PERSEUS Workhorse study (Allocco et al., 2010), incorporate prior information to limit the number of study subjects exposed

to the investigational interventions. 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`).

On the other hand, Hobbs and Carlin (2008) show that 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. 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. Furthermore, our ability to designate patient populations as exchangeable apriori is confounded by a multitude of factors related societal and lifestyle changes over time. In order to avoid poor frequentist operating characteristics and dogmatic, subjective analyses we need prospective methods for incorporating historical data that are "adaptively robust" to prior knowledge that turns out to be inconsistent

with the accumulating experimental data. By contrast, given strong empirical data-based evidence of concurrent commensurability, historical information should be utilized to increase statistical power and reduce cost.

Adaptive trial designs often refer to interim analyses or sample size re-estimation in sequential clinical trials, an area of substantial recent development of Bayesian methods (Cheng and Shen, 2005; DeSantis 2007; Biswas et al., 2009). In this thesis we develop a prospective approach involving several models that allow for the commensurability of the information in the historical and current data to determine how much historical information should be used. The goal of our adaptive 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. Our proposed method weights the influence of prior information relative to its consistency, or *commensurability*, with data from the concurrent study, and is thus adaptive in controlling prior influence. Clinical trial designs for incorporating historical data adaptively need not be Bayesian. For example, the PERSEUS Workhorse study pre-specified a subset of the historical subjects for potential use based upon covariate characteristics, and utilized a discretized adaptive approach that incorporated historical data conditional on the observed event rate exceeding a predetermined value (Allocco et al., 2010). However, our Bayesian hierarchical commensurate prior method offers a smooth, objective approach to adaptive borrowing that is void of preconceived conditions and subjective hyperpriors. Alternative Bayesian solutions include power priors proposed by Ibrahim

and Chen (2000), which are discussed in Chapter 2, and a meta-analytic-predictive approach considered by Neuenschwander et al. (2010); see also Neelon, O'Malley, and Margolis (2008).

The remainder of the thesis evolves as follows. In Chapter 2 we propose various classes of *commensurate priors* for Gaussian data, including novel modifications to the traditional power prior approach. We illustrate, and simulate the performance our of proposed methods. Then in Chapter 3 we extend the general method to linear and linear mixed models as well as generalized linear and generalized linear mixed regression models in the context of two successive clinical trials. Chapter 4 considers the effective historical sample size of our adaptive method for the case when historical data is available only for the concurrent control arm, and proposes "optimal" use of new patients in the current trial using an adaptive randomization scheme that is balanced with respect to the amount of incorporated historical information. Finally, Chapter 5 concludes, discusses our findings, and suggests avenues for further research.

Chapter 2

Hierarchical Commensurate Prior Models

In this chapter, we propose various classes of *commensurate priors*. Section 2.1 introduces the commensurate prior approach for which borrowing depends upon a measure of commensurability among the historical and current data. Then in Section 2.2 we present novel modifications to the traditional power prior approach that use this measure of commensurability to guide the modeling. Section 2.3 compares the frequentist performance of several of the proposed methods using simulation, while Section 2.4 offers an example from a colon cancer trial that illustrates the benefit of our proposed adaptive borrowing approach.

2.1 Commensurate Priors

Let D denote data from the current study and $L(\theta|D)$ the general likelihood function of the current data, where θ 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 $L(\theta_0|D_0)$, where θ_0 represents the historical counterpart of θ. Assuming that we know how consistent D and D_0 will prove to be, we can pre-specify our prior on θ accordingly. For example, if we believed the historical and current data to be exchangeable we would assume that $\theta = \theta_0$ and pool the data from the two studies by setting the prior on θ equal to the historical likelihood. However, if the current data conflicts with the historical data, the resulting analysis will suffer from poor frequentist operating characteristics.

The dichotomous parameterization allows us to build a model that directly measures the similarity of θ and θ_0. Suppose we assume a vague *initial* prior, $p_0(\theta)$, reflecting prior knowledge about θ before D_0 is observed, but construct the prior for θ to be "centered" at θ_0 and conditional on $\tau > 0$, where τ parameterizes prior precision for θ given θ_0. Multiplying by the historical likelihood function results in a prior of the form,

$$p(\theta|D_0, \theta_0, \tau) \propto L(\theta_0|D_0)p(\theta|\theta_0, \tau)p_0(\theta). \tag{2.1}$$

We refer to τ as the *commensurability parameter*, since large τ corresponds to very high commensurability between θ and θ_0, while τ close to zero implies that the datasets do not arise from similar populations. Consequently as τ approaches zero,

$p(\theta|D_0, \theta_0, \tau) \rightarrow p_0(\theta)$, effectively ignoring the historical data. Furthermore, as $\tau \rightarrow \infty$, θ approaches θ_0 and $p(\theta|D_0, \theta_0, \tau) \rightarrow L(\theta|D_0)p_0(\theta)$, recovering the result obtained from pooling the two datasets. If the historical study favors rejecting the current null hypothesis, then decreasing τ reduces Type I error, while increasing τ increases power.

The posterior kernel is obtained by multiplying (2.1) by the current likelihood $L(\theta|D)$. Note that the full conditional posterior distribution for θ_0 would be independent of the current data, since the current likelihood would be nothing but a multiplicative constant. Therefore, θ_0 should be integrated out of the prior when the $\int L(\theta_0|D_0)p(\theta|\theta_0)d\theta_0$ is tractable. Specifying a vague prior for τ, $p(\tau)$, and adopting the joint prior $p(\theta, \tau|D_0, \theta_0) \propto p(\theta|D_0, \theta_0, \tau)p(\tau)$ allows the model to utilize data from the current trial to help estimate τ.

Now let us turn our attention to the Gaussian case for two consecutive, similar trials. Let $\mathbf{x}$ and $\mathbf{x_0}$ denote vectors of i.i.d. responses of length n and n_0, respectively, such that $x_i \overset{iid}{\sim} Normal(\mu, \sigma^2)$ and $x_{0i} \overset{iid}{\sim} Normal(\mu_0, \sigma_0^2)$, where $\mathbf{x_0}$ represents data collected during the earlier, *historical* study. Suppose we are interested in testing the point null hypothesis $H_0 : \mu = 0$ in the second trial.

First let us construct a prior on μ such that borrowing strength from the historical study depends upon the evidence in the data for commensurability among location parameters μ and μ_0. We start by assuming the conditional prior on μ and μ_0 is the product of the historical likelihood and a normal prior on μ with mean μ_0 and precision

τ, and a prior for μ_0.

$$p(\mu, \mu_0|\sigma_0^2, \tau, \mathbf{x_0}) \propto L(\mu_0|\sigma_0^2, x_0) \times N\left(\mu \,\middle|\, \mu_0, \, , \frac{1}{\tau}\right) \times p(\mu_0). \tag{2.2}$$

Note that since $L(\mu, \sigma^2|\mathbf{x})$ is free of μ_0, (2.2) provides the full conditional posterior for μ_0. If no initial information exists for μ_0, we would likely select a flat prior for μ_0, $p(\mu_0) \propto 1$ (throughout the paper we generically denote priors by p and posteriors by q). Since posterior inference on the scale of the historical data is not of direct interest, to simplify computations a bit we replace σ_0^2 in the historical likelihood with its ML or REML estimate, $\hat{\sigma}_0^2$. We also assume the noninformative reference prior for σ^2, thus $q(\sigma^2|\mathbf{x_0}, \mathbf{x}) \propto \Gamma^{-1}\left(\frac{n}{2}, \frac{n}{2}\left[s^2 + (\bar{x} - \mu)^2\right]\right)$ where Γ^{-1} denotes the inverse gamma distribution.

Adding a prior for τ, a noninformative $1/\sigma^2$ prior for the current data scale parameter σ^2, and integrating out the nuisance parameter, μ_0, and normalizing leads to the joint *location commensurate prior* (LCP),

$$p(\mu, \sigma^2, \tau \,\big|\, \mathbf{x_0}) \propto N\left(\mu \,\middle|\, \bar{x}_0 \, , \frac{1}{\tau} + \frac{\hat{\sigma}_0^2}{n_0}\right) \times \frac{1}{\sigma^2} \times p(\tau). \tag{2.3}$$

Thus, we have essentially connected the parameter of interest, μ, to the historical data likelihood through the precision parameter, τ. Alternative distributions could be used; however, the normal precision parameter offers a clear interpretation of commensurability. Furthermore, when evidence for commensurability is weak, τ is forced towards zero, increasing the conditional prior variance of μ in (2.3) by $\frac{1}{\tau}$. Therefore, we refer to τ as the *commensurability* parameter.

MCMC integration of both μ and μ_0 is problematic when the historical and concurrent data are highly commensurate, since for large values of τ (2.2) assigns most of its mass to the line $\mu = \mu_0$. Gaussian densities also offer easy marginalization of μ_0 for normal likelihoods, facilitating conditionally conjugate posterior densities for μ that can be readily Gibbs sampled. For non-Gaussian data, MCMC marginalization of μ_0 can be avoided by replacing $L(\mu_0|\mathbf{x_0})p(\mu_0)$ with a normal approximation; Chapter 3 discusses this approach in generality for generalized linear models.

Selecting a prior for τ completes the hierarchical model specification. Since τ becomes extremely large (small) for highly commensurate (conflicting) data, we work on the log-scale. In Section 2.3 we present simulations to illustrate the frequentist operating characteristics of the LCP using an agnostic $Uniform(-30, 30)$ prior on $\log(\tau)$. The simulation results suggest that $p(\log(\tau)) = Cauchy(0, 30)$ also provides for a highly flexible model. The full posterior is then proportional to the product of the joint prior in (2.3) and the current data likelihood, $L(\mu, \sigma^2|x)$. Marginal posterior inference on the commensurability parameter requires a rejection sampling algorithm, such as Metropolis-Hastings.

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, γ, that measures evidence of commensurability among σ^2 and σ_0^2 by specifying a prior on σ^2 that is "centered" at σ_0^2 and has precision γ. An obvious choice would be to assume an inverse gamma prior on

σ^2, with mean σ_0^2 and precision γ. Assuming the reference initial prior on σ_0^2, multiplying by the historical likelihood, and integrating out μ_0 results in the conditional *location-scale commensurate prior* (LSCP), $p^{LSCP}(\mu, \sigma_0^2, \sigma^2 \mid \mathbf{x_0}, \tau, \gamma)$, emerges as proportional to

$$N\left(\mu \mid \bar{x_0}, \frac{1}{\tau} + \frac{\sigma_0^2}{n_0}\right) \Gamma^{-1}\left(\sigma^2 \mid \alpha^\star, \beta^\star\right) \Gamma^{-1}\left(\sigma_0^2 \mid \frac{n_0 - 1}{2}, \frac{n_0 \hat{\sigma}_0^2}{2}\right), \tag{2.4}$$

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 μ and σ^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 τ and γ 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 $p^{LSCP}(\mu, \sigma_0^2, \sigma^2 \mid \mathbf{x_0}, \tau_j, \gamma_j)$ denote the conditional prior in (2.4) given τ and γ are fixed at τ_j and γ_j respectively. If we are also willing to specify fixed mixing proportions, $\omega_j \in (0, 1)$ where $j = 1, ..., m$ such that $\sum_{j=1}^m \omega_j = 1$, we can formulate a LSCP for μ, σ_0^2, and σ^2 that is a convex combination of the m potential relationships of interest as

$$p^{LSCMP}(\mu, \sigma_0^2, \sigma^2 \mid \mathbf{x_0}) = \sum_{j=1}^{m} \omega_j p^{LSCP}(\mu, \sigma_0^2, \sigma^2 \mid \mathbf{x_0}, \tau_j, \gamma_j). \tag{2.5}$$

We refer to this prior as the *location-scale commensurate mixture prior* (LSCMP). Section 2.3 features a LSCMP that is an equal mixture of just two pairs of (τ, γ), corresponding to very high and low commensurability among the location and scale parameters.

2.2 Hierarchical Power Priors

We begin this section 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), and offer another approach to adaptively incorporating historical data into the analysis of a new trial. The conditional power prior for parameter θ is defined as

$$p(\theta|D_0, \alpha_0) \propto L(\theta|D_0)^{\alpha_0} p_0(\theta), \tag{2.6}$$

for initial prior $p_0(\theta)$ and *power parameter* $\alpha_0 \in [0, 1]$. The power parameter controls the "degree of borrowing": if $\alpha_0 = 0$, (2.6) reduces to the initial prior (no borrowing), whereas if $\alpha_0 = 1$, equation (2.6) returns the usual historical posterior (full borrowing).

In the case of normal historical data, $x_{0i} \overset{iid}{\sim} Normal(\theta, \sigma_0^2)$, σ_0^2 known, $i = 1, \ldots, n_0$, under a flat initial prior, (2.6) yields a $Normal\left(\bar{x}_0, \sigma_0^2/(\alpha_0 n_0)\right)$ power prior distribution for θ. Hence α_0 plays the role of a relative precision parameter for the historical data. Since $0 \leq \alpha_0 \leq 1$, 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 α_0, the conditional posterior distribution for θ given D_0, D, and α_0 emerges as

$$q(\theta|D_0, D, \alpha_0) \propto p_0(\theta)L(\theta|D_0)^{\alpha_0}L(\theta|D) \; . \tag{2.7}$$

Again in the case of known-variance normal observations, $x_i \overset{iid}{\sim} N(\theta, \sigma^2)$, $i = 1, \ldots, n$, this results in another normal distribution for the posterior of θ. 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 α_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 α_0 by adopting the usual Bayesian solution of choosing a *hyperprior* $p(\alpha_0)$ for α_0.

2.2.1 Modified Power Priors

Ibrahim and Chen (2000) propose joint power priors consisting of the product of the conditional power prior in (2.6) and an independent proper prior on α_0. Duan, Ye, and Smith (2006, p.98) caution against this since it violates the Likelihood Principle

(Birnbaum, 1962). Duan et al. (2006), Neuenschwander, Branson, and Spiegelhalter (2009), and Pericchi (2009) modify the joint power prior to the product of the *normalized* conditional power prior (2.6) and an independent proper prior for α_0, namely

$$p^{MPP}(\theta, \alpha_0|D_0) \propto \frac{L(\theta|D_0)^{\alpha_0} p_0(\theta)}{\int L(\theta|D_0)^{\alpha_0} p_0(\theta) d\theta} p(\alpha_0). \tag{2.8}$$

Modified power priors obey the Likelihood Principle and produce marginal posteriors for α_0 that are proportional to products of familiar probability distributions. If we specify $p(\alpha_0)$ 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.

2.2.2 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 α_0, obtained by multiplying (2.8) 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 α_0 (or in our case, fairly informative hyperpriors for α_0) in order to deliver sufficient borrowing. In fact, under a flat $Beta(1, 1)$ prior on α_0, the marginal posterior for α_0 is flat when fit using two

identical datasets regardless of the sample sizes.

In this subsection we propose a novel adaptive modification to the basic power prior formulation rooted in the commensurate prior approach presented in Section 2.1 as a solution to the problems raised by the aforementioned authors. Heretofore in this section both the historical and current data have depended on a common parameter θ. Now as in Section 2.1, we assume *different* parameters in the historical and current group, θ_0 and θ, respectively, where $\theta \in \Re$ and $\theta_0 \in \Re$ are continuous location parameters. We then extend the hierarchical model to include a commensurability parameter, τ, to measure the evidence the for commensurability among θ and θ_0.

Suppose we pick a vague (or even flat) initial prior $p_0(\theta_0)$, but construct the prior for θ to be normal with mean θ_0 and precision τ. We can use the information in τ to guide the prior on α_0. Specifying a vague prior for τ or $\log(\tau)$ and normalizing with respect to θ_0 results in a location commensurate power prior (LCPP) of the form

$$p(\theta, \alpha_0, \tau|D_0)^{LCPP} \propto \int \frac{[L(\theta_0|D_0)]^{\alpha_0}}{\int [L(\theta_0|D_0)]^{\alpha_0}\, d\theta_0} d\theta_0 \times N(\theta|\theta_0, \frac{1}{\tau}) Beta(\alpha_0|g(\tau), 1)p(\tau), \quad (2.9)$$

where $g(\tau) > 0$ is a function of the commensurability parameter that is small for τ close to zero and large for large values of τ. Since inference on θ_0 is not of primary interest in the current analysis, we integrate it out of the joint prior. This extended power prior model requires the estimation of more parameters from the data (notably τ), but we can formulate the model such that the information gained is aimed directly at improving estimation of the crucial borrowing parameter α_0.

2.2.3 Single Arm Trial

We now compare p^{MPP} and p^{LCPP} in the context of the simple trial for Gaussian data outlined in Section 2.1. The power prior formulation requires that σ_0^2 be fixed and known, therefore, we again replace σ_0^2 in the historical likelihood with its maximum likelihood estimate. We also assume the noninformative reference prior for σ^2, thus $q(\sigma^2|\mathbf{x_0}, \mathbf{x}) \propto \Gamma^{-1}\left(\frac{n}{2}, \frac{n}{2}\left[s^2 + (\bar{x} - \mu)^2\right]\right)$ for both power prior models. The conditional posterior for μ becomes

$$q(\mu|\mathbf{x_0}, \mathbf{x}, \alpha_0, \sigma^2) = Normal\left(\frac{\hat{\sigma}_0^2 n\bar{x} + \sigma^2\alpha_0 n_0 \bar{x_0}}{n\hat{\sigma}_0^2 + \sigma^2\alpha_0 n_0}, \frac{\sigma^2\hat{\sigma}_0^2}{n\hat{\sigma}_0^2 + \sigma^2\alpha_0 n_0}\right). \tag{2.10}$$

The conditional posterior for the full (no) borrowing model is found by fixing $\alpha_0 = 1$ ($\alpha_0 = 0$) in (2.10), and thus has mean $\frac{n\bar{x}\hat{\sigma}_0^2 + n_0\bar{x_0}\sigma^2}{n_0\sigma^2 + n\hat{\sigma}_0^2}$ ($\bar{x}$), and variance $\frac{\sigma^2\hat{\sigma}_0^2}{n_0\sigma^2 + n\hat{\sigma}_0^2}$ ($\frac{\sigma^2}{n}$). 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 α_0 under the MPP approach are

$$p^{MPP}(\mu, \sigma^2, \alpha_0|\mathbf{x_0}) \propto Normal\left(\mu \mid \bar{x_0}, \frac{\hat{\sigma}_0^2}{\alpha_0 n_0}\right) \times Beta\left(\alpha_0|a, b\right) \times \frac{1}{\sigma^2} \tag{2.11}$$

$$\text{and } \; q^{MPP}(\alpha_0|\mathbf{x_0}, \mathbf{x}) \propto Normal\left(\bar{x} - \bar{x_0} \mid 0\,, \frac{\sigma^2}{n} + \frac{\hat{\sigma}_0^2}{\alpha_0 n_0}\right) \times Beta\left(\alpha_0|a, b\right). \tag{2.12}$$

Our location commensurate power prior (LCPP) follows,

$$p^{LCPP}(\mu, \alpha_0, \tau|\mathbf{x_0}) \propto N(\mu|\,\bar{x_0}\,, \frac{1}{\tau} + \frac{\hat{\sigma}_0^2}{\alpha_0 n_0}) \times Beta(\alpha_0|g(\tau)\,, 1) \times p(\tau). \tag{2.13}$$

Notice that, in addition to adding $\frac{1}{\tau}$ to the conditional prior variance of μ, the LCPP in (2.13) also inflates the estimated posterior variance of μ_0, $\frac{\hat{\sigma}_0^2}{n_0}$, by a factor of $\frac{1}{\alpha_0}$. The posterior distribution is obtained from the product of the LCPP prior (2.13) and the

normal likelihood for $\mathbf{x}$. The full conditional posterior distribution for μ and marginal posterior distribution for α_0 and τ follow as

$$q^{LCPP}(\mu|\mathbf{x_0}, \mathbf{x}, \alpha_0, \tau, \sigma^2) \propto N\left(\mu \,\middle|\, \frac{\alpha_0 n_0 \tau \sigma^2 \bar{x_0} + nu\bar{x}}{\alpha_0 n_0 \tau \sigma^2 + nu}, \frac{u\sigma^2}{\alpha_0 n_0 \tau \sigma^2 + nu}\right), \tag{2.14}$$

where $u = \alpha_0 n_0 + \hat{\sigma}_0^2 \tau$, and

$$q^{LCPP}(\alpha_0, \tau|\mathbf{x_0}, \mathbf{x}, \sigma^2) \propto N\left(\bar{x} - \bar{x}_0 \,\middle|\, 0, \frac{\sigma^2}{n} + \frac{1}{\tau} + \frac{\hat{\sigma}_0^2}{\alpha_0 n_0}\right) Beta(\alpha_0|g(\tau), 1)p(\tau). \tag{2.15}$$

Information in the data pertaining to τ can be used to guide the $Beta(g(\tau), 1)$ prior on our power parameter, α_0. Since τ becomes extremely large (small) for highly commensurate (conflicting) data we work on the log-scale. In Section 2.3 we present simulated frequentist operating characteristics for the LCPP model using $g^\star(\log(\tau)) = \max(\log(\tau), 1)$. Thus as τ increases and $g^\star(\log(\tau))$ becomes large relative to the second beta hyperparameter (fixed at 1) the distribution of α_0 becomes increasingly peaked at 1 and the prior variance of μ tends towards $\frac{\hat{\sigma}_0^2}{n_0}$, the approximate posterior variance of μ_0 given $\mathbf{x_0}$. Thus, our model strongly encourages borrowing from the historical data when the data are commensurate. Alternatively, for $\tau \leq 1$ the distribution of α_0 becomes flat across the unit interval. Lastly, the LCPP model used in the simulation study assumes a $Cauchy(0, 30)$ prior on $\log(\tau)$. The fat tails of Cauchy facilitate very large and small values of τ, 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.

Figures 2.1 and 2.2 compare the power prior models for fixed $\bar{x_0} = 0$,

$\sigma^2 = \sigma_0^2 = \hat{\sigma}_0^2 = 1$ with values of $\bar{x}$ varying by column. Evidence for commensurability among the current and historical datasets is weakest in the first columns and strongest in the fourth columns (identical sample means and variances). Figure 2.1 compares marginal posterior distributions for α_0 derived from the modified power prior (2.12) (solid) and Ibrahim-Chen (IC) (dashed) power prior under a $Beta(1,1)$ (uniform) hyperprior. Each graph in the top row shows results for $n = 30$ and $n_0 = 60$, while $n_0 = n = 10^7$ in the bottom row. In all 8 scenarios, (2.12) places slightly more density on large values of α_0 relative to the IC marginal posterior. Notice that failing to normalize the prior with respect to μ causes the IC posterior to be skewed relative to (2.12). The figure also clearly elucidates the concern about excessive overattenuation for vague $p(\alpha_0)$. The bottom right graph contains posteriors that are *flat* despite very strong evidence for commensurability (10^7 observations with identical sufficient statistics) among the historical and current datasets.

Figure 2.2 shows marginal posterior distributions under the proposed LCPP model (2.15), assuming a $Cauchy(0, 30)$ hyperprior on $\log(\tau)$ and $g^\star(\log(\tau)) = \max(\log(\tau), 1)$. The top row contains marginal posterior (solid) and prior (dashed) distributions for α_0. The prior for α_0 is peaked at 1 and flat for values less than 0.6, which facilitates posteriors that are peaked at 1 when evidence for commensurability is strong and flat when the historical and current data conflict. The bottom row contains marginal posterior distributions for $\log(\tau)$, as well as the fat-tailed $Cauchy(0, 30)$ hyperprior. The graphs show

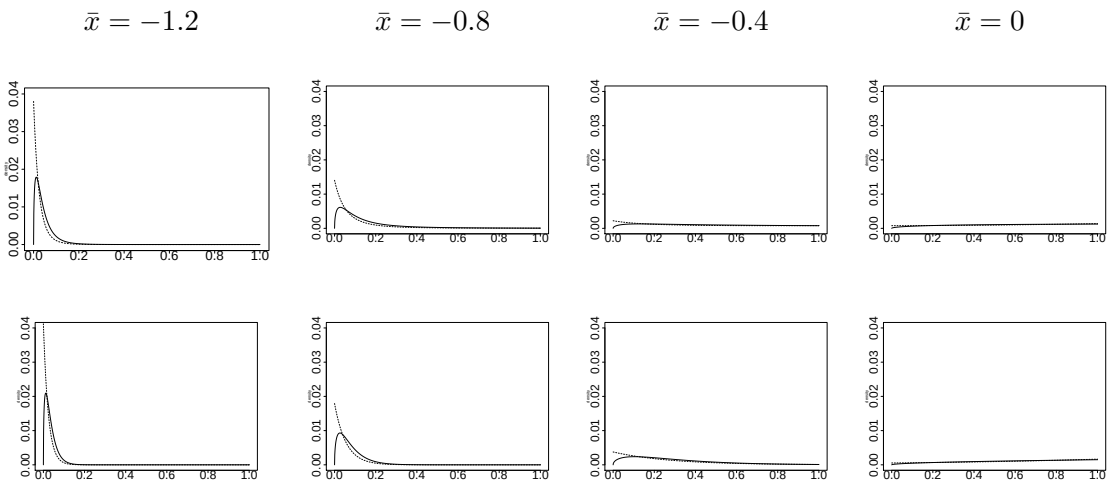

Figure 2.1: Marginal posterior distributions for α_0 from the MPP (11) (solid) and Ibrahim-Chen (dashed) power prior models under a Beta($a = 1, b = 1$) hyperprior, where $\bar{x}_0 = 0$, and fixed $\sigma^2 = \sigma_0^2 = \hat{\sigma}_0^2 = 1$. Each graph in the top row shows results for $n = 30$ and $n_0 = 60$, while $n_0 = n = 10^7$ in the bottom row. Each column corresponds to $\bar{x} = (-1.2, -0.8, -0.4, 0)$.

that the posterior for $\log(\tau)$ shrinks to values less than 0 when evidence for commensurability is weak. For commensurate data the posterior is highly right-skewed over very large, positive values of τ, which forces α_0 to be near 1 and facilitates more borrowing. This results in much less overattenutation of consistent historical data, yet facilitates sufficient variance inflation when evidence for commensurability with the current data is weak.

2.2.4 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

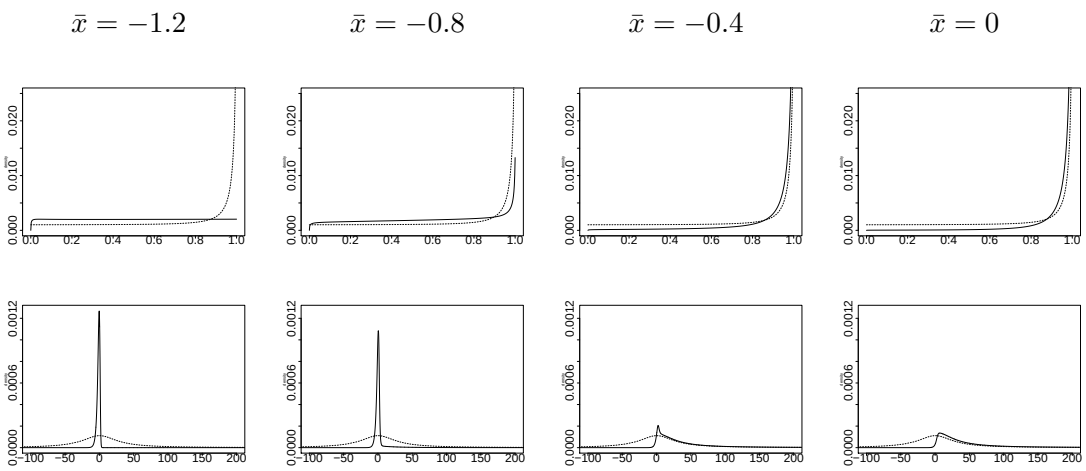

Figure 2.2: Marginal posterior (solid) and prior (dashed) distributions for α_0 and $\log(\tau)$ under the LCPP 2.15 $g^\star(\log(\tau)) = \max(\log(\tau), 1)$ and $Cauchy(0, 30)$ hyperprior on $\log(\tau)$, where $\bar{x}_0 = 0$, $n_0 = 60$, $n = 30$, and fixed $\sigma^2 = \sigma_0^2 = \hat{\sigma}_0^2 = 1$. Each graph in the top row shows results for α_0, while the bottom row shows results for $\log(\tau)$. Each column corresponds to $\bar{x} = (-1.2, -0.8, -0.4, 0)$.

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 model to adaptively borrow strength from the historical control group and identical covariates. Specifying our commensurate power prior as in (2.9) with the same priors on σ^2, α_0, and $\log(\tau)$ as in the previous subsection, a flat prior on λ, and integrating β_0 out of the joint prior leads to a full conditional prior on β that is normal with mean $V^{-1}M$, and covariance $(\tau V)^{-1}$ where $M = \left(\frac{\alpha_0 X_0^T X_0}{\hat{\sigma}_0^2} + \tau I_p\right)^{-1} \frac{\alpha_0 X_0^T X_0 \hat{\beta}_0}{\hat{\sigma}_0^2}$, $V = I_p - \tau \left(\frac{\alpha_0 X_0^T X_0}{\hat{\sigma}_0^2} + \tau I_p\right)^{-1}$, and $\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T y_0$. The joint posterior follows by multiplying the joint prior by the likelihood of y and normalizing. The full conditional posteriors for λ and σ^2 as well as the posterior for β given σ^2, α_0, and τ follow as,

$$q^{CPP}(\lambda | D_0, D, \beta, \sigma^2) \propto N_r\left(\lambda \,\middle|\, \hat{\lambda},\, \sigma^2 (Z^T Z)^{-1}\right), \tag{2.16}$$

$$q^{CPP}(\sigma^2 | D_0, D, \beta, \lambda) \propto \Gamma^{-1}\left(\sigma^2 \,\middle|\, \frac{n}{2},\, \frac{(y - X\beta - Z\lambda)^T (y - X\beta - Z\lambda)}{2}\right), \tag{2.17}$$

$$q^{CPP}(\beta | D_0, D, \sigma^2, \alpha_0, \tau) \propto N_p\left(\beta \,\middle|\, \Lambda^{-1}\left(\tau M + \frac{X^T (I_n - w) y}{\sigma^2}\right),\, \Lambda^{-1}\right), \tag{2.18}$$

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 (2.16) that the full conditional posterior mean for λ, $\hat{\lambda}$, is a function of residuals $(y - X\beta)$, whereas the conditional posterior mean of β in (2.18) is an average of the historical and concurrent data relative to the power and commensurability parameters, α_0 and τ. As τ and α_0 approach zero, the marginal posterior for β 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.

2.3 Simulation Results for a Single Arm Trial

In this section we investigate the Bayesian and frequentist operating characteristics of the models described in Sections 2.1 and 2.2 via simulation. Among the commensurate priors in Section 2.1, the simulations were run using a LCP model that assumes a vague, $Uniform(-30, 30)$ prior on $\log(\tau)$ and noninformative reference prior on σ^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. For the power priors in Subsection 2.2.3, we used a MPP model with a $Beta(1, 1)$ prior on α_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 σ^2. We also ran simulations on a model that ignores the historical data completely (by assuming the noninformative Jeffreys prior on μ and σ^2) and, following Fúquene et al. (2009), a model that assumes a "robust" Cauchy prior on μ, centered at $\bar{x}_0$ with scale parameter fixed at 1. We will refer to these as the "no borrowing" and "Cauchy" models.

Figure 2.3 illustrates the adaptive movement capability of the commensurate prior methods. Each graph contains 95% posterior credible intervals for μ derived from all simulated models, generated with $n_0 = 60$ historical observations having $\bar{x}_0 = 0$ and $\hat{\sigma}_0^2 = 1$. The top interval corresponds to results from the analysis of the historical data

alone, using noninformative Jeffreys priors on μ_0 and σ_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 σ_0^2 and σ^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 $x_i \overset{iid}{\sim} N(\mu, \sigma^2)$ for some fixed true μ, σ^2, and n, and compute an equal-tail posterior credible interval for μ reflecting our decision rule. 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 2.1 contains areas under the resulting power curves. All models are compared under two approaches to hypothesis testing. The first uses 95% posterior credible intervals to test the null hypothesis, 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.

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 2.1: Area under the power curves: 95% posterior credible intervals (left); Controlled Type I error (right).

	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

Lastly, power curves for the full borrowing (dot-dashed), LCPP (solid), MPP (dashed), and no borrowing (dotted) models are shown in Figure 2.4 to augment Table 2.1. 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; c.f. Fúquene et al. (2009, p.819).

2.4 Example Using Controlled Colorectal Cancer Trial Data

We consider data from two successive randomized controlled 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 until a death or disease progression, while the trial reported by Goldberg et al. measured every 6 weeks for the first 42 weeks, or until death or disease progression. We computed the sum of the longest diameter in cm ("ld sum") for up to 9 tumors for each patient at each cycle. 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). Measurements at all cycles, including follow-up measurements recorded after disease progression, were utilized for the eligible subjects. Suppose y_0 and y are vectors of lengths n_0 and n for the historical and concurrent responses such that

$$y_0 \sim Normal(X_0\beta_0\ ,\ \sigma_0^2),\ \text{and}\ y \sim Normal(X\beta + Z\lambda\ ,\ \sigma^2) \tag{2.19}$$

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 β_0 and β parameters contain intercepts as well as regression coefficients for each of three baseline covariates, while λ represents change in average ld sum attributed to FOLFOX. Figure 2.5 contains histograms of the average change in ld tumor sum from baseline: historical for y_0 (left), $y|Z = 0$ (middle), and $y|Z = 1$ (right). The histograms suggest that assumptions of normality are acceptable. Notice that the histogram for FOLFOX (right) places more mass on smaller values. This suggests that FOLFOX achieved a greater reduction in ld sum on average than the IFL regimen.

The left and center columns of Table 2.2 summarize results from separate classical linear regression fits on the historical (y_0, X_0) and current (y, X) data alone. The center 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 significance level.

Information about β_0 appears to be relevant to β. Therefore, we implemented the commensurate power prior linear model presented in detail in Subsection 2.2.4 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 right column of Table 2.2. First, notice that the posterior for α_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 λ 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.

Table 2.2: Fits to colorectal cancer data: $n_0 = 171$ (left); $n = 270$ (right); LCPP (bottom).

	Historical data		Current data	
	estimate	95% CI	estimate	95% CI
β_1	0.88	$(-1.98, 3.74)$	-0.47	$(-2.28, 1.34)$
β_2	-0.23	$(-0.31, -0.15)$	-0.40	$(-0.45, -0.34)$
β_3	-0.02	$(-0.07, 0.02)$	0.01	$(-0.01, 0.04)$
β_4	0	$(-0.02, 0.02)$	0.01	$(-0.01, 0.02)$
λ	–	–	-0.41	$(-1.02, 0.19)$

	LCPP	
	estimate	95% Posterior CI
β_1	0.180	$(-1.11, 1.42)$
β_2	-0.39	$(-0.44, -0.33)$
β_3	0	$(-0.02, 0.02)$
β_4	0	$(-0.01, 0.01)$
λ	-0.46	$(-0.82, -0.10)$
α_0	0.86	$(0.44, 1.00)$
$\log(\tau)$	16.53	$(1.59, 82.00)$

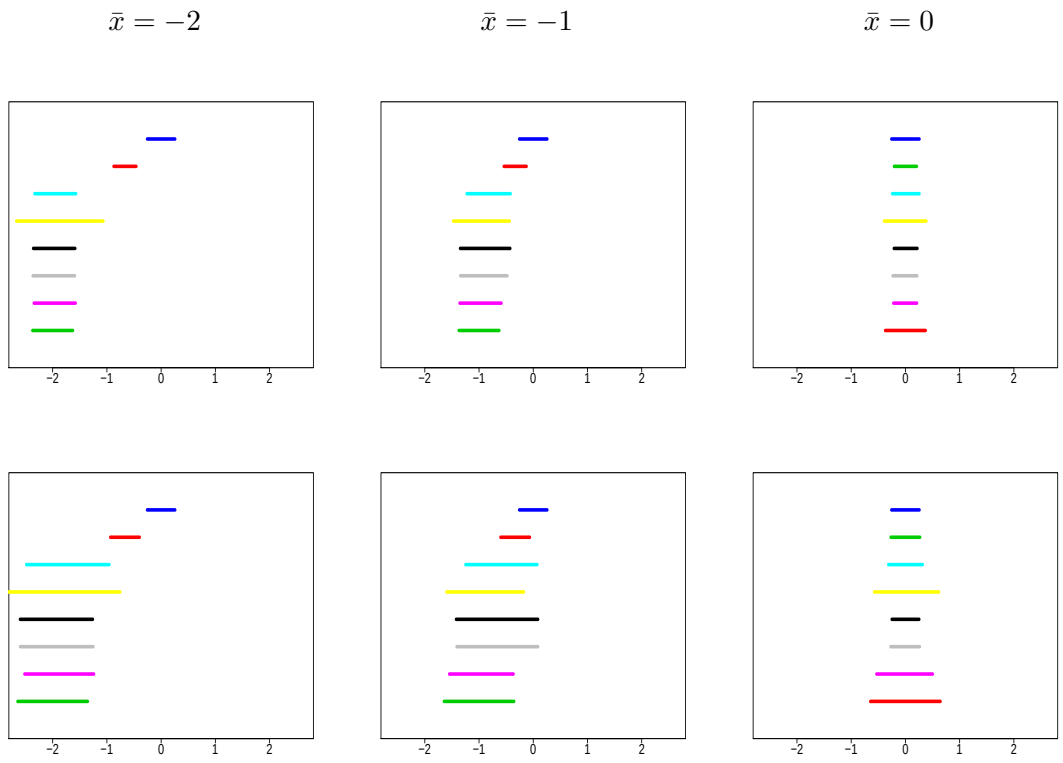

Figure 2.3: 95% posterior credible intervals for μ for all simulated models where $\bar{x}_0 = 0$, $n_0 = 60$, $n = 30$, and $\hat{\sigma}_0^2 = 1$. Each graph contains results from eight different models: historical (top), full borrowing (second), MPP (third), Cauchy (fourth), LCPP (fifth), LCP (sixth), LSCMP (seventh), and no borrowing (bottom). Columns correspond to the current sample mean, $\bar{x}$, while the top row shows results for $\hat{\sigma}^2 = 1$ and the bottom row assumes $\hat{\sigma}^2 = 3$.

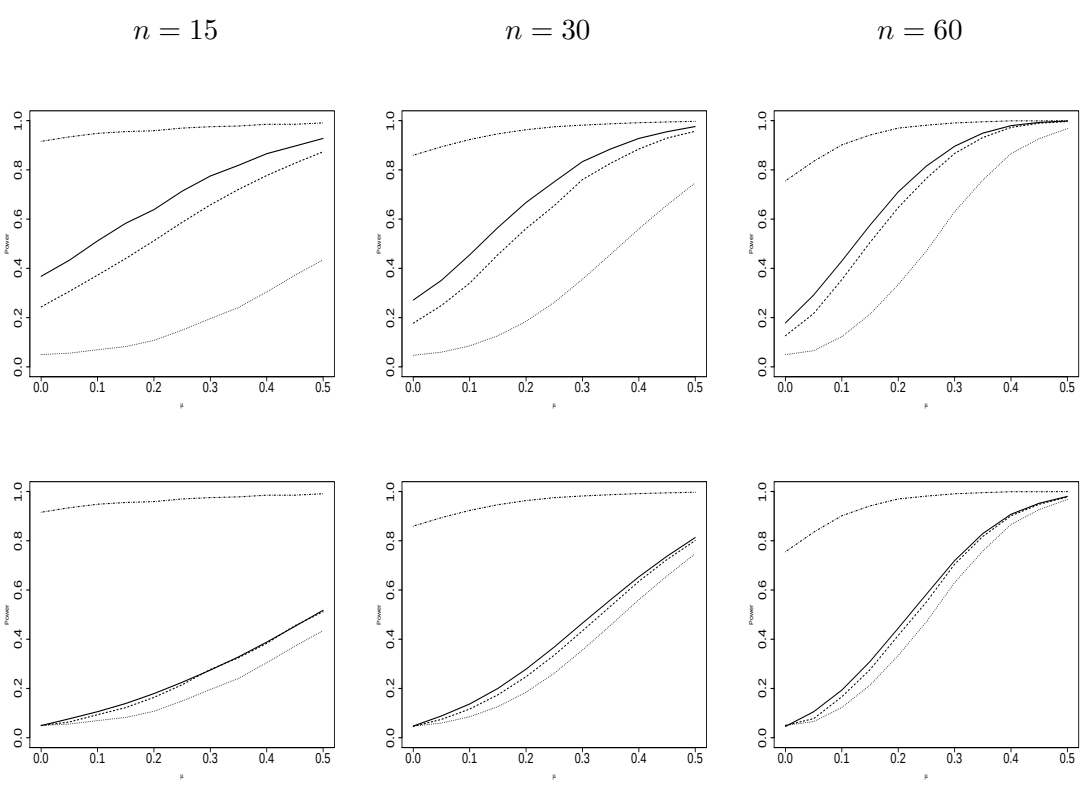

Figure 2.4: Power curves for the full borrowing, commensurate power prior, modified power prior, and no borrowing models where $\mu_0 = 0.5$, $n_0 = 60$, and $\sigma^2 = \sigma_0^2 = 1$. Each graph shows results for the full borrowing (dot-dashed), LCPP (solid), MPP (dashed), and no borrowing (dotted) 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.

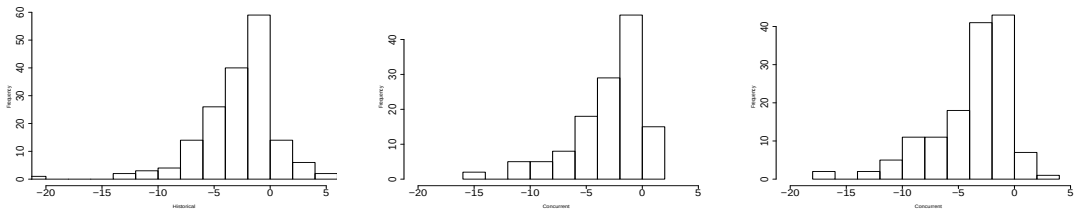

Figure 2.5: Histograms of average change in ld tumor sum from baseline for the colorectal cancer data used in Section 5: historical IFL (left), concurrent IFL (center), FOLFOX (right).

Chapter 3

General and Generalized Linear Models

In this chapter, we extend the general commensurate prior method to linear and linear mixed models as well as generalized linear and generalized linear mixed regression models in the context of two successive clinical trials. Section 3.1 introduces the commensurate prior general linear and general linear mixed models for Gaussian response data. Then in Section 3.2 we expand the commensurate linear model approach to include non-Gaussian responses for generalized linear and generalized linear mixed models. Section 3.3 offers illustrative time-to-event and longitudinal analyses that illustrate the benefit of our proposed commensurate adaptive approach, while Section 3.4 compares the frequentist performance of two proposed linear models using simulation. Throughout the chapter we assume that both trials identically measure $p-1$ covariates

representing fixed effects which are to be incorporated into the analysis. We also assume that the current trial is designed to compare a novel intervention to a previously studied control therapy, and thus historical data are available *only* for the control group. Furthermore, commensurate priors are constructed to inform about fixed regression effect parameters. Historical random effects and variance component parameters are fixed at their ML or REML estimates.

3.1 General Linear Models

3.1.1 Fixed Effect Models

Assume y_0 is a vector of n_0 responses from subjects in the historical study 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 exist. Let y be the vector of n responses from subjects in both the treatment and control arms of the current trial. 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 intercepts. Now suppose $y_0 \sim N_{n_0}(X_0\beta_0, \sigma_0^2)$ and $y \sim N_n(X\beta + d\lambda, \sigma^2)$ where λ is the (scalar) treatment effect and d is an $n \times 1$ vector of $0-1$ indicator variables for the new treatment. Let y_i, X_i, d_i represent data corresponding to the ith, subject in the current trial, $i = 1, ..., n$.

Commensurability in the linear model depends upon similarity in the intercepts and covariate effects. Therefore, we start by assuming

$$p(\beta, \beta_0|\sigma_0^2, \tau, y_0) \propto N_{n_0}(y_0|X_0\beta_0, \sigma_0^2 I_{n_0}) N_p\left(\beta|\beta_0, \frac{1}{\tau} I_p\right) p(\beta_0).$$

Therefore, β is connected to the historical data likelihood via the commensurability precision parameter τ. This construction facilitates adaptive borrowing of strength from the historical control group response data and identical covariates. Given no initial prior information about the historical location and scale parameters, we assume a flat prior on β_0 and fix the variance in the historical likelihood to its frequentist ML or REML estimate, $\hat{\sigma}_0^2$.

After integrating β_0 out of the joint prior, the normalized full conditional prior for β follows as normal with mean $V^{-1}M$ and covariance $(\tau V)^{-1}$, where $M = \left(\frac{X_0^T X_0}{\hat{\sigma}_0^2} + \tau I_p\right)^{-1} \frac{X_0^T X_0 \hat{\beta}_0}{\hat{\sigma}_0^2}$, $V = I_p - \tau \left(\frac{X_0^T X_0}{\hat{\sigma}_0^2} + \tau I_p\right)^{-1}$, and $\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T y_0$. Adopting vague priors for σ^2, $\log(\tau)$, and λ completes the model specification. The joint posterior follows by multiplying the joint prior by the likelihood of y. The full conditional posteriors for λ and σ^2 follow as

$$q(\lambda|y_0, y, \beta, \sigma^2) \propto N_r\left(\lambda \,\middle|\, \hat{\lambda}, \frac{\sigma^2}{\sum_{i=1}^n t_i}\right), \tag{3.1}$$

$$q(\sigma^2|y_0, y, \beta, \lambda) \propto \Gamma^{-1}\left(\sigma^2 \,\middle|\, \frac{n}{2}, \frac{(y - X\beta - d\lambda)^T (y - X\beta - d\lambda)}{2}\right). \tag{3.2}$$

where $\hat{\lambda} = \frac{\sum_{i=1}^n t_i (y_i - X_i\beta)}{\sum_{i=1}^n t_i}$. The full conditional posterior for β, $q(\beta|y_0, y, \lambda, \sigma^2, \tau)$,

emerges as proportional to

$$N_p\left(\beta\Big|\ \left(\frac{X^TX}{\sigma^2}+\tau V\right)^{-1}\left(\frac{X^T(y-d\lambda)}{\sigma^2}+\tau M\right),\ \left(\frac{X^TX}{\sigma^2}+\tau V\right)^{-1}\right). \tag{3.3}$$

Notice in (3.1) that the full conditional posterior mean for λ, $\hat{\lambda}$, is a function of residuals $(y - X\beta)$, whereas the full conditional posterior mean of β in (3.3) is an average of the historical and concurrent data relative to the commensurability parameter τ. As τ approaches zero, the marginal posterior for β converges to a normal density with mean $\left(\frac{X^TX}{\sigma^2}\right)^{-1}(X^Ty)$ and variance $\left(\frac{X^TX}{\sigma^2}\right)^{-1}$, recovering the standard result from 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. Moreover, as τ approaches infinity, $V^{-1}M \rightarrow \hat{\beta}_0$ and $(\tau V)^{-1} \rightarrow \hat{\sigma}_0^2\left(X_0^TX_0\right)^{-1}$, fully incorporating the historical data, recovering the results of a pooled analysis. Thus the commensurate prior is bound by the "informativeness" of the historical data or historical precision.

3.1.2 Mixed Models

In this subsection, we extend our model to include random effects. We begin with a familiar and useful one-way ANOVA model. Then following McCulloch and Searle (2001, p.156), give the linear mixed model for general variance-covariance structures between and within levels of the random components.

3.1.2.1 One-way random effects model

Following the notation of Browne and Draper (2006, p.474), suppose $y_{0ij} = \mu_0 + u_{0j} + \epsilon_{0ij}$, for $i = 1, ..., n_{0j}$, $j = 1, ..., m_0$, and $n_0 = \sum_{i=1}^{m_0} n_{0j}$. We assume $\epsilon_{0ij} \overset{iid}{\sim} N(0, \sigma^2_{\epsilon 0})$, and that the random effects are distributed $u_{0j} \overset{iid}{\sim} N(0, \sigma^2_{u0})$. Let y_{0j} be an n_{0j}-vector of responses from subject j, and let y_0 denote the collection of responses for all subjects in all groups. Therefore, $\sigma^2_{\epsilon 0}$ represents the conditional variance of $y_{0ij}|u_j$, while the marginal variance of y_{0ij} follows as $\sigma^2_{u0} + \sigma^2_{\epsilon 0}$. Note that observations from different subjects are assumed to be uncorrelated. We can write the historical likelihood as a product of multivariate normals with mean μ_0 and correlated errors,

$$L(\mu_0, \sigma^2_{\epsilon 0}, \sigma^2_{u0}|y_0) = \prod_{j=1}^{m_0} N_{n_{0j}} \left(y_{0j} \,\middle|\, \mu_0 1_{n_{0j}},\ \sigma^2_{u0} J + \sigma^2_{\epsilon 0} I_{n_{0j}} \right), \tag{3.4}$$

where $1_{n_{0j}}$ is a column vector of 1s, J is a $n_{0j} \times n_{0j}$ matrix of 1s, and $I_{n_{0j}}$ is a $n_{0j} \times n_{0j}$ identity matrix. As in the previous subsection, given no prior information we assume a flat prior on μ_0 and fix the variance components in the historical likelihood, σ^2_{u0}, and $\sigma^2_{\epsilon 0}$, to their respective frequentist REML estimates, $\hat{\sigma}^2_{\epsilon 0}$ and $\hat{\sigma}^2_{u0}$. For a discussion of the advantages of using REML to estimate variance components, see McCulloch and Searle (2001, p.177) or Browne and Draper (2006). Comparisons of frequentist and objective Bayesian approaches can be found in Chaloner (1987) and Severini (1994).

Similarly, the model for responses in the current trial follows $y_{ij} = \mu + u_j + d_j\lambda + \epsilon_{ij}$, for $i = 1, ..., n_j$, $j = 1, ..., m$, and $n = \sum_{i=1}^{m} n_j$, where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2_{\epsilon})$ and $d_j = 1$ indicates treatment and $d_j = 0$ corresponds to the standard of care for the jth subject

in the concurrent trial. Therefore, fixed effects μ and λ represent the intercept and response effect for a patient receiving the new intervention.

In the one-way random effects model, borrowing from the historical study depends upon commensurability of the intercepts. Therefore, we specify a normal prior on μ with mean μ_0 and precision τ. Given that outcomes are measured identically in both trials, small $\mu - \mu_0$ relative to the data's informativeness implies that the response experience was similar for subjects receiving control in both trials. At the current stage we treat the random components as latent normal variables, $u_j \overset{iid}{\sim} N(0, \sigma_u^2)$, and proceed with inference on fixed effects and concurrent variance components. As in the previous subsection, we adopt vague priors for σ_ϵ^2, σ_u^2, $\log(\tau)$, and λ. Therefore, the location commensurate prior for the one-way random effects model, $p(\mu, \lambda, \sigma_\epsilon^2, \sigma_u^2, \tau | y_0)$, follows as proportional to the joint prior for $\lambda, \sigma_\epsilon^2, \sigma_u^2$, and τ, $p\left(\lambda, \sigma_\epsilon^2, \sigma_u^2, \tau\right)$, multiplied by,

$$\int \left[\prod_{j=1}^{m_0} N_{n_{0j}} \left(y_{0j} \,\middle|\, \mu_0,\, \hat{\sigma}_{u0}^2 J + \hat{\sigma}_{\epsilon 0}^2 I_{n_{0j}} \right) \right] \times N\left(\mu \,\middle|\, \mu_0,\, \frac{1}{\tau} \right) d\mu_0 \times N_m\left(0, \sigma_u^2 I_m\right). \quad (3.5)$$

To ease the subsequent algebra, it is useful to note that $\hat{\sigma}_{\epsilon 0}^2$ times the inverted marginal estimated historical covariance for all observations in the jth subject, $\hat{\sigma}_{\epsilon 0}^2 \left(\hat{\sigma}_{u0}^2 J + \hat{\sigma}_{\epsilon 0}^2 I_{n_{0j}} \right)^{-1}$, is equal to $\left(I_{n_{0j}} - \frac{\hat{\sigma}_{u0}^2}{\hat{\sigma}_{\epsilon 0}^2 + n_{0j} \hat{\sigma}_{u0}^2} J_{n_{0j}} \right)$, which has $n_{0j} - 1$ eigenvalues equaling 1 and one non-unit eigenvalue equal to $1 - r_j^{-1}$, where $r_j = \frac{\hat{\sigma}_{\epsilon 0}^2}{n_{0j} \hat{\sigma}_{u0}^2} + 1$. Following the recommendations of Gelman (2006, p.527), we use independent noninformative uniform priors on the two standard deviation parameters, σ_u and σ_ϵ, for large m (≥ 5), which is equivalent to a product of inverse-χ^2 densities with -1 degrees of freedom, $p(\sigma_u^2, \sigma_\epsilon) \propto \frac{1}{\sigma_u \sigma_\epsilon}$. For small m (< 5), we use the half-t (the absolute value of

a Student-t distribution centered at zero) prior for σ_u, which becomes a half-Cauchy density for 1 degree of freedom (Gelman, 2006, p.520). For $m \geq 5$, the joint prior, $p(\mu, \lambda, \sigma_\epsilon^2, \sigma_u^2, \tau | y_0)$, emerges as proportional to

$$N\left(\mu \Big| \frac{\sum_{j=1}^{m_0}\sum_{i=1}^{n_{0j}} y_{0ij}(1-r_j^{-1})}{\sum_{j=1}^{m_0} n_{0j}(1-r_j^{-1})}, \frac{1}{\tau} + \frac{\hat{\sigma}_{\epsilon 0}^2}{\sum_{j=1}^{m_0} n_{0j}(1-r_j^{-1})}\right) N_m\left(0, \sigma_u^2 I_m\right) \frac{p(\tau)}{\sigma_u \sigma_\epsilon}. \quad (3.6)$$

For $m < 5$, the joint prior is proportional to

$$\begin{aligned} &N\left(\mu \Big| \frac{\sum_{j=1}^{m_0}\sum_{i=1}^{n_{0j}} y_{0ij}(1-r_j^{-1})}{\sum_{j=1}^{m_0} n_{0j}(1-r_j^{-1})}, \frac{1}{\tau} + \frac{\hat{\sigma}_{\epsilon 0}^2}{\sum_{j=1}^{m_0} n_{0j}(1-r_j^{-1})}\right) N_m\left(0, \sigma_u^2 I_m\right) \\ &\times \left(1 + \frac{1}{\nu}\left(\frac{\sigma_u}{\kappa}\right)^2\right)^{-(1+\nu)/2} \times \frac{p(\tau)}{\sigma_\epsilon}, \end{aligned} \quad (3.7)$$

where the half-t prior distribution for σ_u is parameterized in terms of scale κ and degrees of freedom ν. Let A and B denote the normalized full conditional prior mean and variance of μ in (3.6). The joint posterior follows by multiplying the joint prior by the likelihood of y and normalizing. Full conditional posteriors for λ, u, μ, σ_ϵ^2, and σ_u^2 follow as

$$\begin{aligned} q(\lambda | u, \mu, \sigma_\epsilon^2, \sigma_u^2, y_0, y) &\propto N\left(\frac{\sum_{j=1}^{m} d_j \sum_{i=1}^{n_j}\left[y_{ij} - n_j(\mu + u_j)\right]}{\sum_{j=1}^{m} n_j d_j}, \frac{\sigma_\epsilon^2}{\sum_{j=1}^{m} n_j d_j}\right), \\ q(u | \lambda, \mu, \sigma_\epsilon^2, \sigma_u^2, y_0, y) &\propto \prod_{j=1}^{m} N\left(\frac{\sigma_u^2 \sum_{i=1}^{n_j}\left[y_{ij} - (\mu + d_j\lambda)\right]}{n_j\sigma_u^2 + \sigma_\epsilon^2}, \left(\frac{n_j}{\sigma_\epsilon^2} + \frac{1}{\sigma_u^2}\right)^{-1}\right), \\ q(\mu | \lambda, u, \sigma_\epsilon^2, \sigma_u^2, \tau, y_0, y) &\propto N\left(\frac{V\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left[y_{ij} - (u_j + d_j\lambda)\right] + A\sigma_\epsilon^2}{nB + \sigma_\epsilon^2}, \left(\frac{n}{\sigma_\epsilon^2} + \frac{1}{B}\right)^{-1}\right), \\ q(\sigma_\epsilon^2 | \lambda, u, \mu, y_0, y) &\propto \Gamma^{-1}\left(\frac{n-1}{2}, \frac{\sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{ij} - \mu - u_j - d_j\lambda)^2}{2}\right), \\ q(\sigma_u^2 | u, y_0, y) &\propto \Gamma^{-1}\left(\frac{m-1}{2}, \frac{\sum_{j=1}^{m} u_j^2}{2}\right). \end{aligned}$$

Note that $\frac{\sigma_u^2}{\sigma_u^2+\sigma_\epsilon^2}$ represents the conditional correlation for all y_{ij} and $y_{i'j}$ within the same subject j. Several alternative prior specifications of the correlation structure may be more natural for incorporating prior information for the variance components. For example, we could formulate our prior opinion about the model smoothness by specifying a prior on the variance ratio, $\frac{\sigma_\epsilon^2}{\sigma_u^2}$ or the degrees of freedom it induces (Hodges and Sargent, 2001). For discussion about degrees of freedom and how it can be used to sensibly determine variance component priors, as well as the general marginal posterior for the variance ratio, see Reich and Hodges (2008), and Cui et al. (2010).

3.1.2.2 Linear mixed model

The one-way random effects model presented in Subsection 3.1.2.1 is a special case of a linear mixed model for which between-subject observations are independent, all within subject observations have identical covariance (compound symmetry within groups), and no fixed regression effects (only intercepts). In this subsection we present details for the general location commensurate prior linear mixed model.

Let y_{0ij} denote the ith historical response in subject j, for $1 \leq i \leq n_{0j}$, $1 \leq j \leq m_0$, and y_{kl} denote the kth concurrent response in subject l, for $1 \leq k \leq n_l$, $1 \leq l \leq m$, where $n_0 = \sum_{j=1}^{m} n_{0j}$ and $n = \sum_{l=1}^{m} n_l$. Suppose X_0 and X are $n_0 \times p$ and $n \times p$ design matrices such that the first columns contain vectors of 1s corresponding to the intercepts, β and β_0 are vectors of identically measured regression coefficients of length p representing fixed covariate effects, and d is an $n \times 1$ new intervention indicator.

Furthermore, let u_0, u and Z_0, Z denote $m_0 \times 1$ and $m \times 1$ random effects vectors and their respective $n_0 \times m_0$ and $n \times m$ design matrices for the historical and concurrent data.

Adopting the notation of McCulloch and Searle (2001, p.156), we formulate the general linear mixed model by first assuming normally distributed random effects with means 0 and covariances D_0 and D, $u_0 \sim N(0, D_0)$ and $u \sim N(0, D)$. Models for the historical and concurrent responses are,

$$y_0 = X_0\beta_0 + Z_0u_0 + \epsilon_0,$$

$$\text{and } y = X\beta + Zu + d\lambda + \epsilon,$$

where $\epsilon_0 \sim N_{n_0}(0, R_0)$ and $\epsilon \sim N_n(0, R)$; R_0 and R represent the conditional covariances of $y_0|u_0$ and $y|u$. The marginal covariance for y_0 and y emerge as $\Sigma_0 = cov(y_0) = Z_0D_0Z_0^T + R_0$ and $\Sigma = cov(y) = ZDZ^T + R$.

Commensurability in the linear mixed model now depends upon similarity in the intercepts and covariate effects. Therefore, we start by assuming $p(\beta, \beta_0|\Sigma_0, \tau, y_0) \propto N_{n_0}(y_0|X_0\beta_0, \Sigma_0)N_p\left(\beta|\beta_0, \frac{1}{\tau}I_p\right)p(\beta_0)$. Given no initial prior information about the historical location and covariance parameters, we assume a flat prior on β_0 and fix the marginal covariance in the historical likelihood to its frequentist REML estimate, $\hat{\Sigma}_0$. Assuming that concurrent fixed effects parameters, β, λ, and variance components within D and R are of interest, β_0 is a nuisance parameter that can be integrated out of the joint prior. Then the normalized full conditional prior for β emerges as

$p(\beta|\tau, y_0) \propto N_p\left(\beta \,\middle|\, V^{-1}M,\ \frac{1}{\tau}V^{-1}\right)$, where $M = \left(X_0^T\hat{\Sigma}_0^{-1}X_0 + \tau I_p\right)^{-1} X_0^T\hat{\Sigma}_0^{-1}X_0\tilde{\beta}_0$, $V = I_p - \tau\left(X_0^T\hat{\Sigma}_0^{-1}X_0 + \tau I_p\right)^{-1}$, and $\tilde{\beta}_0 = \left(X_0^T\hat{\Sigma}_0^{-1}X_0\right)^{-1} X_0^T\hat{\Sigma}_0^{-1}y_0$, the usual integrated least squares estimate based on the historical data. The joint posterior follows by multiplying the joint prior by the current likelihood and normalizing.

Treating u as a latent variable and adopting conjugate Wishart priors for D^{-1} and R^{-1}, namely $D^{-1} \sim W([\phi D^\star, \phi])$ and $R^{-1} \sim W([\rho R^\star, \rho])$, posterior inference may again proceed via the Gibbs sampler for all but the commensurability parameter, τ, which again requires Metropolis sampling. The remaining full conditional posteriors follow as

$q(u|\lambda, \beta, D, R, y_0, y) \propto$

$$N_m\left(\left(D^{-1} + Z^TR^{-1}Z\right)^{-1} Z^TR^{-1}(y - X\beta - d\lambda),\ \left(D^{-1} + Z^TR^{-1}Z\right)^{-1}\right),$$

$q(\beta|\lambda, u, R, \tau, y_0, y) \propto$

$$N_p\left(\left(X^TR^{-1}X + \tau V\right)^{-1}\left(X^TR^{-1}(y - d\lambda - Zu) + \tau M\right),\ \left(X^TR^{-1}X + \tau V\right)^{-1}\right),$$

$q(\lambda|u, \beta, R, y_0, y) \propto N\left(\left(d^TR^{-1}d\right)^{-1} d^TR^{-1}(y - X\beta - Zu),\ \left(d^TR^{-1}d\right)^{-1}\right),$

$q(D^{-1}|\beta, \lambda, u, y_0, y) \propto W\left(\left[uu^T + \phi D^\star\right], m + \phi\right),$

and $q(R^{-1}|\beta, \lambda, u, y_0, y) \propto$

$$W\left(\left[(y - X\beta - d\lambda - Zu)^T(y - X\beta - d\lambda - Zu) + \rho R^\star\right] n + \rho\right).$$

If we assume that observations are uncorrelated across subjects, this simplifies the covariance structures to $D_0 = \sigma_{u0}^2 I_{m_0}$, $D = \sigma_u^2 I_m$, $R_0 = \sigma_{\epsilon 0}^2 I_{n_0}$, and $R = \sigma_\epsilon^2 I_n$. This

essentially extends the one-way random effects model in Subsection 3.1.2.1 to incorporate fixed covariate effects. Using the same priors as in Subsection 3.1.2.1, the full conditional posteriors for and σ_u^2 and σ_ϵ^2 follow as
$q(\sigma_\epsilon^2|\lambda, u, \beta, y_0, y) \propto \Gamma^{-1}\left(\frac{n-1}{2}, \frac{(y-X\beta-d\lambda-Zu)^T(y-X\beta-d\lambda-Zu)}{2}\right)$ and
$q(\sigma_u^2|u, y_0, y) \propto \Gamma^{-1}\left(\frac{m-1}{2}, \frac{\sum_{j=1}^m u_j^2}{2}\right)$. See Kass and Natarajan (2006) for an empirical Bayes approach using an inverted Wishart prior on Σ for general covariance structure and design matrices. For discussion on the convergence properties, see Wang and Titterington (2006).

3.2 Generalized Linear Models

In this section we extend the commensurate prior method to incorporate generalized linear models for non-Gaussian error distributions. We first present the general method for fixed effects models, and then discuss a location commensurate logistic regression model for binary outcomes followed by a location-shape commensurate log-linear Weibull regression model. We then extend the general method to incorporate random effects, and illustrate in two important specific cases: a probit regression model for binary outcomes, and a Poisson regression model for count data.

3.2.1 Fixed Effect Models

Let y_0 and y denote column vectors of length n_0 and n consisting of independent measurements from a distribution that is a member of the exponential family, $f_Y(y)$.

That is, we suppose $y_{0j} \overset{indep}{\sim} f_{Y_{0j}}(y_{0j})$ and $y_i \overset{indep}{\sim} f_{Y_i}(y_i)$ such that the log-likelihoods are of form

$$log f_{Y_{0j}}(y_{0j}) = \sum_{j=1}^{n_0} \left[y_{0j}\gamma_{0j} - b_0(\gamma_{0j}) \right] / \sigma_0^2 - \sum_{j=1}^{n_0} c_0(y_{0j}, \sigma_0) \tag{3.8}$$

$$\text{and } log f_{Y_i}(y_i) = \sum_{i=1}^{n} \left[y_i\gamma_i - b(\gamma_i) \right] / \sigma^2 - \sum_{i=1}^{n} c(y_i, \sigma), \tag{3.9}$$

for $j = 1, ..., n_0$ and $i = 1, ..., n$ (McCulloch and Searle, 2001, p.139). Suppose X_0 and X are $n_0 \times p$ and $n \times p$ design matrices such that the first columns contain vectors of 1s corresponding to the intercept, β and β_0 are vectors of identically measured regression coefficients of length p, and d is an $n \times 1$ new intervention indicator. Assuming $E[y_0] = \mu_0$ and $E[y] = \mu$, let $g(\mu_0) = X_0\beta_0$ and $g(\mu) = X\beta + d\lambda$, for known link function g.

To construct the location commensurate prior, we first must tie β to an asymptotic normal approximation to $L(\beta_0|y_0)p(\beta_0)$ using the so-called Bayesian Central Limit Theorem (see e.g. Carlin and Louis, 2009, p.108). Given no initial prior information, the approximation takes the mean equal to the historical MLE (computed numerically via Newton-Raphson or Fisher scoring) and variance equal to the inverted observed Fisher information matrix. Thus we have $L(\beta_0|y_0) \approx N_p\left(\hat{\beta}_0,\ \left(X_0^T \hat{W}_0 X_0\right)^{-1}\right)$, where $\hat{W}_0 = W_0(\hat{\mu}_0)$ is an $n_0 \times n_0$ diagonal matrix having $jj-$element

$$W_{0jj}(\mu_0) = \left[\sigma^2 v(\mu_0) \left(\frac{\partial g(\mu_0)}{\partial \mu_0} \right)^2 \right]^{-1}, \tag{3.10}$$

where $v(\mu_{0j}) = var(y_{0j})/\sigma^2 = \partial^2 b(\gamma_{0j})/\partial \gamma_{0j}^2$, for $j = 1, ..., n_0$ (McCulloch and Searle, 2001, p.141). Specifying a conditional prior on β that is normal with mean β_0 and

covariance $\frac{1}{\tau}I_p$, and integrating out β_0 again leads to a normalized full conditional prior on β that is normal with mean $V^{-1}M$ and covariance $(\tau V)^{-1}$, where $M = \left(X_0^T\hat{W}_0X_0 + \tau I_p\right)^{-1} X_0^T\hat{W}_0X_0\hat{\beta}_0$ and $V = I_p - \tau\left(X_0^T\hat{W}_0X_0 + \tau I_p\right)^{-1}$. Adopting a flat prior for λ, the joint location commensurate prior follows as,

$$p(\beta, \lambda, \tau|y_0) \propto N_p\left(V^{-1}M,\ \frac{1}{\tau}V^{-1}\right)p(\tau). \tag{3.11}$$

The joint posterior is then proportional to the product of this prior and the current data likelihood, which now likely results in intractable non-conjugate full conditional distributions for *all* parameters. Posterior inference then proceeds using the Metropolis algorithm. Alternatively, one could use a second asymptotic normal approximation for the current likelihood, as in Spiegelhalter, Abrams, and Myles (2004, p.25). However, since the variance of y is a function of β and λ, the asymptotic covariance or expected Fisher information matrix required depends upon β and λ. Therefore, inference would require fixed values of the parameters of interest in the observed Fisher information matrix, which is not desirable.

3.2.1.1 Binary Response

Perhaps the most important special case of non-Gaussian data arising in clinical trials is that of binary responses. Let y_0 and y denote the historical and concurrent data such that $y_{0j} \sim Ber\left[\pi_0(X_{0j})\right]$, $\pi_0(X_{0j}) \in [0,1]$, for $j = 1, ..., n_0$, and $y_i \sim Ber\left[\pi(X_i, d_i)\right]$, $\pi(X_i, d_i) \in [0,1]$, for $i = 1, ..., n$. The logistic link function transforms the expectations

of y_0 and y such that, $\log\left(\frac{\pi_0(X_0)}{1-\pi_0(X_0)}\right) = X_0\beta_0$ and $\log\left(\frac{\pi(X,d)}{1-\pi(X,d)}\right) = X\beta + d\lambda$. The diagonal elements of $\hat{W}_0$ in (3.10) now contain the estimated historical sampling variance, $\hat{W}_{0jj} = \hat{\pi}_0(X_{0j})\left(1 - \hat{\pi}_0(X_{0j})\right)$, where $\hat{\pi}_0(X_{0j}) = \left(1 + e^{-X_{0j}\hat{\beta}_0}\right)^{-1}$.

Adopting a flat prior for λ, the location commensurate posterior is proportional to the product of (3.11) and the current data likelihood, $\prod_{j=1}^{n} \pi\left(X_j\right)^{y_j}\left[1 - \pi\left(X_j\right)\right]^{1-y_j}$. Sampling proceeds by Metropolis, though switching to a probit link function can lead to closed form full conditionals (see Carlin and Louis, 2009, p.364). Fúquene, Cook, and Pericchi (2009) propose an adaptive approach using a robust Cauchy prior for univariate logistic transformations.

3.2.1.2 Time-to-Event Response

Following the notation of Kalbfleisch and Prentice (2002, p.52), data for the historical and concurrent trials consist of triples $(t_{0j}, \delta_{0j}, X_{0j})$ for $j = 1, ..., n_0$ and (t_i, δ_i, X_i) for $i = 1, ..., n$. Here, $t_{0j}, t_i > 0$ are the observed, possibly censored, failure times; δ_{0j}, δ_i are censoring indicators (0 if censored, 1 if failure); and X_{0j} and X_i are row vectors of p covariates associated with historical subject j and concurrent subject i. Let $\tilde{t}_{0j}$ and $\tilde{t}_i$ be the underlying uncensored failure times, with corresponding densities $f(t_{0j})$ and $f(t_i)$. Denote the survivor functions for the jth historical and ith concurrent individuals as $P(\tilde{t}_i > t) = F\left(t\right)$ and $P(\tilde{t}_{0j} > t_0) = F\left(t_0\right)$.

Log-linear models are commonly used for analyzing time-to-event data. Suppose $y_0 = log(t_0) = X_0\beta_0 + \sigma_0 e_0$ and $y = log(t) = X\beta + d\lambda + \sigma e$ where $e_0 = (y_0 - X_0\beta_0)/\sigma_0$

and $e = (y - X\beta - d\lambda)/\sigma$. Assuming that censoring times are conditionally independent of each other and of the independent failure times given X_0 and X (*noninformative censoring*), the historical and concurrent data likelihoods follow as $L_0(\beta_0, \sigma_0|y_0) = \prod_{i=j}^{n_0} \left[\frac{1}{\sigma_0} f(e_{0j})\right]^{\delta_{0j}} F(e_{0j})^{1-\delta_{0j}}$ and $L(\beta, \sigma|y) = \prod_{i=j}^{n} \left[\frac{1}{\sigma} f(e_i)\right]^{\delta_i} F(e_i)^{1-\delta_i}$. The location commensurate prior follows (3.11) where ML estimates for β_0 and σ_0 are computed numerically via Newton-Raphson or Fisher-Scoring; see Kalbfleisch and Prentice (2002, p.66–69). For discussion about censoring mechanisms see Kalbfleisch and Prentice (2002, p.193) or Klein and Moeschberger (2003, p.63). Following Section 3.1 we can fix the historical shape parameter to its ML estimate $\hat{\sigma}_0$, and we might also estimate the current shape σ using only data from the concurrent study. Another option involves expanding (3.11) to include $\log(\sigma)$ and then use location-shape commensurate approach; this method is demonstrated with a Weibull regression model in Section 3.3.1 below.

Weibull regression arises when e_0 and e follow the extreme value distribution, $f(u) = exp[u - exp(u)]$. This results in a parametric regression model which has both a proportional hazards and an accelerated failure-time representation. Commensurability among the regression coefficients alone may be inadequate since the hazard function, $h(t) = -d \log F(t)/dt$, is monotone decreasing for shape parameter $\sigma > 1$, increasing for $\sigma < 1$, constant for $\sigma = 1$. Therefore, we consider a location-shape commensurate approach so that borrowing depends upon commensurability among σ_0 and σ as well as β_0 and β. Simulations results in Section 3.4 highlight the importance of commensurability among the shape parameters when borrowing information from historical data.

Let $\theta_0 = \left(\beta_0^T, \log(\sigma_0)\right)^T$ and $\theta = \left(\beta^T, \log(\sigma)\right)^T$ denote column vectors of length $p+1$. The location-shape commensurate prior on θ follows from (3.11), extended to include the historical ML estimate of σ_0, and extended observed Fisher information matrix for θ_0. Let e_0 and δ_0 denote vectors of length n_0 such that the jth element is equal to e_{0j} and δ_{0j}, $j = 1, ..., n_0$. ML equations for the historical coefficients follow as $\frac{\partial \log L_0}{\partial \beta_0} = \frac{1}{\sigma_0} X_0^T \left(\exp(e_0) - \delta_0\right)$ and $\frac{\partial \log L_0}{\partial \log(\sigma_0)} = \sum_{j=1}^{n_0} \left[e_{0j} \exp(e_{0j}) - \delta_{0j}(e_{0j} + 1)\right]$. Defining $\hat{e}_{0j} = (y_{0j} - X_{0j}\hat{\beta}_0)/\hat{\sigma}_0$, the observed Fisher information matrix $\hat{\Psi}_0 = \hat{\Psi}_0(\hat{\theta}_0)$ follows as,

$$\hat{\Psi}_0(\hat{\theta}_0) = \begin{bmatrix} X_0^T \hat{W}_0 X_0 & X_0^T \hat{E}_0 \\ \hat{E}_0^T X_0 & \sum_{j=1}^{n_0} \hat{e}_{0j} \left[\delta_{0j} - (\hat{e}_{0j} + 1) \exp(\hat{e}_{0j})\right] \end{bmatrix}, \tag{3.12}$$

where $\hat{E}_0$ is the vector of length n_0 containing elements $\hat{E}_{0j} = (\hat{e}_{0j} + 1) \exp\{\hat{e}_{0j} - \log(\hat{\sigma}_0)\} - \delta_{0j}/\hat{\sigma}_0$, and the diagonal elements of $\hat{W}_0$ are $\hat{W}_{0jj} = \frac{1}{\hat{\sigma}_0^2} \exp(\hat{e}_{0j})$, for $j = 1, ..., n_0$ (see Breslow and Clayton, 1993). Specifying a flat prior on λ, the location-shape commensurate prior follows as

$$p(\theta, \lambda, \tau | y_0) \propto N_{p+1}\left(\Lambda^{-1} Q, \frac{1}{\tau} \Lambda^{-1}\right) p(\tau). \tag{3.13}$$

where $Q = \left(\hat{\Psi}_0 + \tau I_{p+1}\right)^{-1} \hat{\Psi}_0 \hat{\theta}_0$ and $\Lambda = I_{p+1} - \tau \left(\hat{\Psi}_0 + \tau I_{p+1}\right)^{-1}$. The posterior is proportional to the product of (3.13) and the concurrent data likelihood. Note that the exponential model is a special case where $\sigma_0 = \sigma = 1$, since this leads to probability distribution functions for t_0 and t following $f(t_0) = \frac{1}{\mu_0} \exp(t_0/\mu_0)$ and $f(t) = \frac{1}{\mu} \exp(t/\mu)$, where $\mu_0 = \exp(X_0 \beta_0)$ and $\mu = \exp(X\beta + d\lambda)$.

3.2.2 Mixed Models

As with the general linear models in Section 3.1, we now extend our generalized linear model. Let y_0 and y denote response vectors of length n_0 and n consisting of conditionally independent measurements given the random effects u_0 and u vectors of length m_0 and m. For generalized linear mixed models, we assume the conditional distributions of y_0 given u_0 and y given u have distributions with densities from the exponential family, $y_{0j}|u_0 \overset{indep}{\sim} f_{Y_{0j}}(y_{0j}|u_0)$ and $y_i|u \overset{indep}{\sim} f_{Y_i}(y_i|u)$, where $log f_{Y_{0j}}(y_{0j}|u_0)$ and $log f_{Y_i}(y_i|u)$ follow (3.8) and (3.9) for $j = 1, ..., n_0$ and $i = 1, ..., n$. In addition, let the random effects be distributed as $u_0 \sim f_{U_0}(u_0)$ and $u \sim f_U(u)$. The marginal likelihoods follow as $f_{Y_{0j}}(y_{0j}) = \int \prod_{j=1}^{n_0} f_{Y_{0j}}(y_{0j}|u_0) f_{U_0}(u_0) d_{u_0}$ and $f_{Y_i}(y_i) = \int \prod_{i=1}^{n} f_{Y_i}(y_i|u) f_U(u) d_u$. Let Z_0 and Z denote $n_0 \times m_0$ and $n \times m$ design matrices corresponding to the random effect vectors, X_0 and X denote $n_0 \times p$ and $n \times p$ design matrices such that the first columns contain vectors of 1s corresponding to the intercept, β and β_0 denote vectors of identically measured regression coefficients of length p, and d again denotes an $n \times 1$ indicator vector for the new intervention. Assuming that the conditional means are $E[y_0|u_0] = \mu_0$ and $E[y|u] = \mu$, let $g(\mu_0) = X_0\beta_0 + Z_0u_0$ and $g(\mu) = X\beta + d\lambda + Zu$, where g is a known link function. Following Subsection 3.2.1 with normally distributed random effects, $u_0 \sim N(0, D_0)$ and $u \sim N(0, D)$, the generalized linear mixed location commensurate prior follows from (3.11), with $\hat{W}_0$ replaced by $(\hat{W}_0^{-1} + Z_0\hat{D}_0Z_0^T)^{-1}$. See McCulloch (1997), McCulloch and Searle (2001, p.263), or Breslow and Clayton (1993) for algorithms for computing ML estimates for fixed effects and prediction of random

effects.

3.2.2.1 Binary Response

As a first specific illustration of generalized linear mixed modeling using commensurate priors, consider again the case of binary response. We specify the distribution of the historical responses, y_0, conditionally on random effects, u_0, as $y_{0jk}|u_{0k} \sim Ber\left[\pi_0(X_{0j})\right]$, $\pi_0(X_{0jk}) \in [0,1]$, where y_{0jk} denotes observation j for the kth subject, $j = 1, ..., n_0$, $k = 1, ..., m_0$. Similarly, the concurrent responses are distributed independently as $y_{il}|u_l \sim Ber\left[\pi(X_{il}, d_{il})\right]$, $\pi(X_{il}, d_{il}) \in [0,1]$, where y_{il} denotes observation i for the lth subject, $i = 1, ..., n$, $l = 1, ..., m$. Using matrix notation, let X_0, X, Z_0, Z denote $n_0 \times p$, $n \times p$, $n_0 \times m_0$, and $n \times m$ design matrices, where $n_0 = \sum_{k=1}^{m_0} n_{0k}$ and $n = \sum_{l=1}^{m} n_l$.

Since we have already illustrated the logit link in Section 3.2.1.1, let us instead consider the probit link. Given $u_0 \sim N(0, D_0)$, the probit uses the standard normal c.d.f., $\Phi(\cdot)$, to transform the means of y_0 and y such that $\pi_0(X_0, Z_0) = \Phi(\beta_0 X_0 + Z_0 u_0)$ and $\pi(X, d, Z) = \Phi(X\beta + d\lambda + Zu)$. Using matrix notation, the location commensurate prior for the probit random effects model follows from (3.11), where the diagonal elements of $\hat{W}_0$, now an $n_0 \times n_0$ matrix, contain $\hat{W}_{0aa} = \frac{\phi(X_0\hat{\beta}_0 + Z_0\hat{u}_0)^2_{aa}}{\Phi(X_0\hat{\beta}_0 + Z_0\hat{u}_0)_{aa}\left(1 - \Phi(X_0\hat{\beta}_0 + Z_0\hat{u}_0)_{aa}\right)}$, where $\phi(.)$ is the standard normal p.d.f. and $a = 1, ..., n_0$, McCulloch and Searle (2001, p.136). The observed Fisher information matrix follows as $\left(X_0^T(\hat{W}_0^{-1} + Z_0\hat{D}_0Z_0^T)^{-1}X_0\right)^{-1}$.

3.2.2.2 Count Response

Consider next the common case of count responses. Suppose that the historical responses follow $y_{0jk}|u_{0k} \overset{indep}{\sim} Poisson(\mu_{0jk})$, where $log(\mu_{0jk}) = X_{0jk}\beta_0 + u_{0k}$ and $u_{0k} \overset{i.i.d.}{\sim} N(0, \sigma^2_{u_0})$. Here, y_{0jk} denotes the jth count observed for the kth subject, where $j = 1, ..., n_{0k}$ and $k = 1, ..., m_0$. Similarly, let y_{il} denotes the ith count observed for the lth subject, where $y_{il}|u_l \overset{indep}{\sim} Poisson(\mu_{il})$, $log(\mu_{il}) = X_{il}\beta + u_l$, and $u_l \overset{i.i.d.}{\sim} N(0, \sigma^2_u)$, for $i = 1, ..., n_l$ and $l = 1, ..., m$. Using matrix notation, where X_0 is an $n_0 \times p$ design matrix, for $n_0 = \sum_{k=1}^{m_0} n_{0k}$, the matrix, $\hat{W}_0$, now contains diagonal elements $\hat{W}_{0aa} = \exp(X_0\hat{\beta}_0 + \hat{u}_0)_{aa}$, where $a = 1, ...n_0$. The location commensurate prior once again follows from (3.11), with observed Fisher information matrix equal to $\left(X_0^T(\hat{W}_0^{-1} + \hat{\sigma}^2_{u0}Z_0Z_0^T)^{-1}X_0\right)^{-1}$.

3.3 Examples

In this section, we illustrate our commensurate prior approach using data from the successive randomized controlled colorectal cancer clinical trials described in Section 2.4. Recall that in the initial trial, originally reported by Saltz et al. (2000), the IFL regimen resulted in significantly longer progression free survival, and that the subsequent trial, originally reported by Goldberg et al. (2004), compared two new regimens, FOLFOX and IROX, to the IFL regimen identical to that used in the Saltz et al. study. In both trials, disease progression required a 25% or greater increase in measurable tumor or

the appearance of new lesions.

Given the relatively short time lag between trials and uniform outcome ascertainment and covariate measurement, information about IFL in the Saltz et al. trial appears to worthy of potential incorporation for comparing IFL to FOLFOX in the Goldberg et al. trial. Therefore, we implemented our adaptive commensurate prior method to test for significant differences between the two regimens for two outcomes. Section 3.3.1 below offers a fixed effects time-to-event analysis using the Weibull regression model presented in Subsection 3.2.1.2 to compare disease progression rates among the FOLFOX and IFL regimens. In addition, Section 3.3.2 presents a longitudinal analysis using the linear mixed effects model described in Subsection 3.1.2.2 that concentrates on the trial's measurements of tumor size. In both illustrations, 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. For simplicity, we omit data from the Irinotecan alone and 5FU/LV arms in the Saltz study, and the IROX arm in the Goldberg study. Both analyses incorporate baseline ld sum as a predictor. The linear mixed model uses the cycle-by-cycle tumor measurements to test for a significant treatment difference in ld sum reduction between the FOLFOX and control regimens.

3.3.1 Time-to-event model

We restricted our Weibull regression analysis to patients that had measurable tumors, nonzero ld sum at baseline, and nonzero progression times, bringing the total

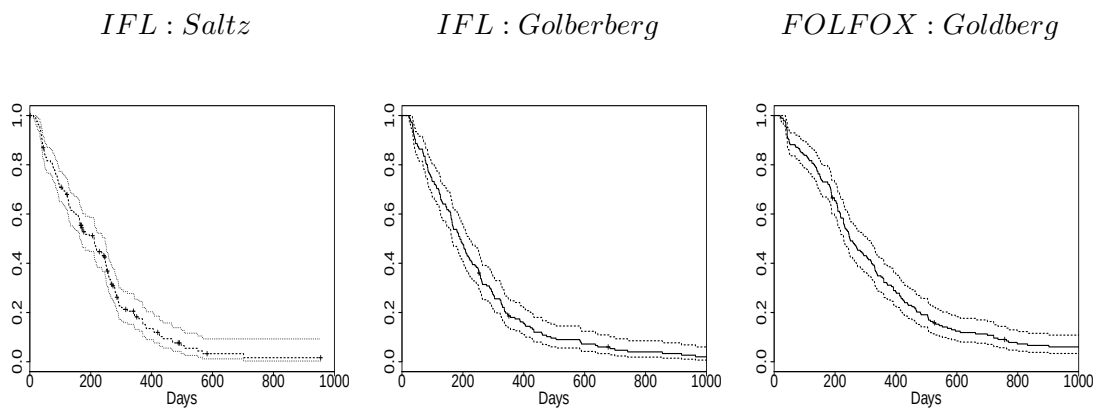

Figure 3.1: Separate Kaplan-Meier survival curves corresponding subjects on IFL in the Saltz et al. trial (left), IFL in the Goldberg et al. trial (center), and FOLFOX in the Goldberg et al. trial.

sample size to 586: 224 historical and 362 current observations. Among the current patients, 176 are controls (IFL) and 186 are patients treated with the new regimen (FOLFOX). Figure 3.1 contains Kaplan-Meier estimated survival curves for subjects on each treatment regimen in both trials. The plots suggest that the survival experience for subjects on IFL was similar in both the Saltz et al. (left panel) and Goldberg et al. trials (center), and that FOLFOX (right) is associated with somewhat prolonged time-to-progression.

Following the log-linear model notation of Subsection 3.2.1.2 for progression times t_0 and t, let $y_0 = X_0\beta_0 + \sigma_0 e_0$ and $y = X\beta + d\lambda + \sigma e$, where $y_0 = \log(t_0)$ and $y = \log(t)$; here, X_0 and X are $n_0 \times 3$ and $n \times 3$ design matrices with columns corresponding to (1, ld sum at baseline, $I[age < 65]$), and d is the FOLFOX indicator function. Thus, the β_0 and β parameters contain intercepts as well as regression coefficients for each of baseline

Table 3.1: Weibull model fits to colorectal cancer data: $n_0 = 224$ (left); $n = 362$ (right); LSCP (bottom).

	Historical data		Current data	
	estimate	95% CI	estimate	95% CI
Intercept	5.54	(5.29, 5.79)	5.68	(5.47, 5.90)
BL ldsum	-0.01	$(-0.02, 0.01)$	-0.02	$(-0.03, -0.004)$
$age < 65$	0.04	$(-0.19, 0.27)$	0.08	$(-0.11, 0.26)$
FOLFOX	–	–	0.42	(0.24, 0.60)
σ	0.75	(0.66, 0.84)	0.86	(0.79, 0.93)

	LSCP	
	estimate	95% Posterior CI
Intercept	5.63	(5.47, 5.79)
BL ldsum	-0.01	$(-0.02, -0.003)$
$age < 65$	0.06	$(-0.08, 0.20)$
FOLFOX	0.46	(0.31, 0.61)
σ	0.83	(0.78, 0.88)
$\log(\tau)$	23.30	(6.82, 38.35)

covariates, while $\exp(\lambda)$ represents the acceleration factor associated with FOLFOX. Note that since $F(t|d = 1) = F\left(te^{\lambda}|d = 0\right)$ for all t, a negative value is indicative of *decreased* survival. Exploratory data analysis on the covariates strongly suggested that the trials enrolled patients from comparable populations. The first, second, and third quartiles for baseline tumor sum are $5, 8.5, 12.8$ in the Saltz trial and $4.7, 7.9, 12.7$ in the Goldberg trial; for age they are $54, 62, 69$ in the Saltz trial and $53, 61, 69$ in the Goldberg trial.

The left and center columns of Table 3.1 summarize results from separate classical linear regression fits on the historical (t_0, X_0) and current (t, X, d) data alone. The center values constitute a "no borrowing" analysis. Results from the current dataset suggest

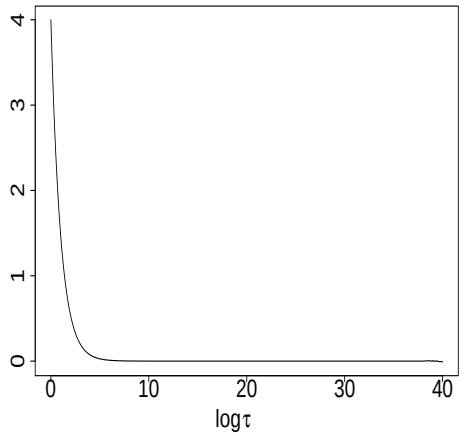

Figure 3.2: Plot of the difference of the traces of the inverted observed fisher information matrix $\hat{\Psi}_0(\hat{\theta}_0)$ obtained using the IFL arm of the Saltz et al. trial and the commensurate prior covariance $\frac{1}{\tau}\Lambda^{-1}$ (y-axis) as a function of $\log(\tau)$ (x-axis).

that ld sum at baseline is significant, while in both datasets age is not. Moreover, results from the current data alone suggest that the estimated acceleration factor corresponding to FOLFOX is highly significant at the 0.05 level. Figure 3.2 plots of the difference of the trace of the inverted observed Fisher information matrix $\hat{\Psi}_0(\hat{\theta}_0)$ and the trace of the commensurate prior covariance $\frac{1}{\tau}\Lambda^{-1}$. The plot shows that the commensurate prior model essentially borrows fully from the Saltz et al. data for all $\log(\tau) > 5$.

Point estimates (posterior medians) and 95% equal-tail Bayesian credible intervals corresponding to our location-shape commensurate model are given in the right column of Table 3.1. This portion of the table reveals that our our model considers the historical and current data to be quite commensurate, since the lower bound of the equal-tail 95% posterior credible interval for $\log(\tau)$ is estimated to be approximately 6.82. The

commensurate method results in credible intervals of lengths 0.32, 0.02, 0.28, and 0.30 for the four regression coefficients of interest; analysis of the Goldberg et al. data produces confidence intervals of lengths 0.43, 0.03, 0.37, and 0.36. Therefore, incorporating the Saltz et al. data into the analysis of the Goldberg et al. trial using our adaptive approach leads to increased precision of the parameter estimates. Note that our model estimates progression times to be exp(0.46) times longer on average in the FOLFOX group, and since the posterior median of σ is less than 1, suggests that the hazard rate is increasingly slightly over time. 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.

3.3.2 Longitudinal model with Gaussian response

Our second analysis concentrates on the trial's measurements of tumor size, and how the FOLFOX regimen compared to the IFL regimen for reducing them. Responses consist of the within subject cycle-by-cycle tumor measurements at cycles $3, 6, 9, 12$, and 15 minus baseline measurements, a total $n_0 = 146$ observations on $m_0 = 111$ subjects in the Saltz et al. study and $n = 390$ observations on $m = 231$ subjects in the Goldberg et al. study. Plots of the within subject changes by cycle for both studies are shown in Figure 3.3. If FOLFOX outperforms IFL for reducing tumor size over time, then the

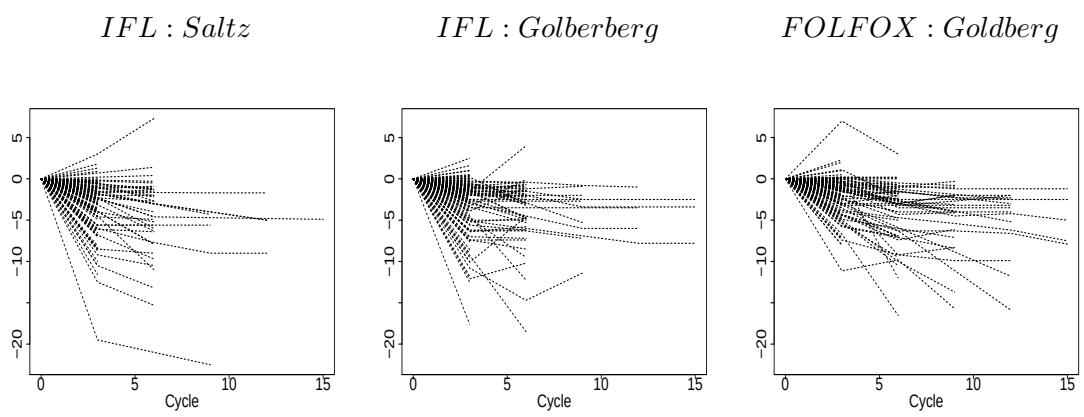

Figure 3.3: Plots of within subject changes from baseline in ld sum by cycle for IFL in the Saltz et al. trial (left), IFL in the Goldberg et al. trial (center), and FOLFOX in the Goldberg et al. trial. Each line represent results for one patient.

data may provide evidence for a significant, negative FOLFOX intercept or FOLFOX-by-cycle interaction. We present results for a longitudinal analysis using the linear mixed model presented in Subsection 3.1.2.2, with $D_0 = \sigma^2_{u0} I_{m_0}$, $R_0 = \sigma^2_{\epsilon 0} I_{n_0}$, $D = \sigma^2_u I_m$, and $R = \sigma^2_\epsilon I_n$. That is, the correlation structure within subject is assumed to be compound symmetric, while observations between subjects are assumed to be independent. More complex correlation structures, such as first order autoregressive could be considered. Note that random slopes over time are not used since we include cycle as a single fixed effect variable in the analysis.

Design matrices X_0 and X now contain columns corresponding to (1, ld sum at baseline, cycle), d is now a $n \times 2$ matrix consisting of (FOLFOX indicator function, FOLFOX$\times$ cycle), and Z_0 and Z are $n_0 \times m_0$ and $n \times m$ design matrices corresponding to the random intercepts. Thus, parameters β_0 and β contain the grand intercepts and

Table 3.2: Linear mixed model fits to colorectal cancer data: $n_0 = 146$ (left); $n = 390$ (center); LCP (right).

	Historical data		Current data	
	estimate	95% CI	estimate	95% CI
Intercept	−0.75	(−1.97, 0.48)	0.36	(−0.44, 1.16)
BL ldsum	−0.22	(−0.32, −0.13)	−0.35	(−0.40, −0.30)
Cycle	−0.25	(−0.38, −0.12)	−0.17	(−0.28, −0.07)
FOLFOX	–	–	1.75	(0.86, 2.63)
FOLFOX×Cycle	–	–	−0.15	(−0.29, −0.02)
σ_u	3.20	–	2.09	–
σ_ϵ	1.22	–	1.54	–

	LCP	
	estimate	95% Posterior CI
Intercept	0.12	(−0.49, 0.80)
BL ldsum	−0.32	(−0.37, −0.28)
Cycle	−0.20	(−0.28, −0.11)
FOLFOX	1.71	(0.91, 2.53)
FOLFOX×Cycle	−0.13	(−0.25, −0.01)
σ_u	2.09	–
σ_ϵ	1.55	–
$\log(\tau)$	23.45	(4.43, 43.11)

regression coefficients for baseline ld sum and cycle, while λ consists of the effect of FOLFOX on tumor sum and a regression coefficient corresponding to the interaction of FOLFOX with cycle.

The left and center columns of Table 3.2 summarize results from separate classical linear regression fits on the historical (y_0, X_0, Z_0) and current (y, X, d, Z) data alone. Again the center values constitute the "no borrowing" analysis. Results from both datasets suggest that ld sum at baseline and cycle are negative and highly significant,

while both intercepts are not significantly different from zero. As such, our model considers β and β_0 to be highly commensurate, and thus fully incorporates the historical data to provide more precise estimates of the parameters shown in the right column of Table 3.2. Note that results for both the "no borrowing" and location commensurate prior models suggest that the additive effect of FOLFOX is significantly positive, while the FOLFOX-by-cycle interactions are significantly negative, at the 0.05 level respectively. This suggests that IFL is associated with more rapid reductions that dissipate over time, while FOLFOX leads to more gradual reductions in tumor size.

3.4 Simulations

In this section we use simulation to investigate the Bayesian and frequentist operating characteristics of the linear model and both exponential and Weibull versions of the time-to-event models described in Sections 3.1 and 3.2. For the linear and exponential models, let y_0 and y denote responses for the historical and concurrent data and assume models $E[y_0] = \mu_0$ and $E[y] = \mu_0 + d\lambda$, where d is a indicator for new treatment and μ_0 and μ are intercepts corresponding to the historical and current controls. Our simulations compare the commensurate prior approach to models that ignore (no borrowing) and fully incorporate (full borrowing) the historical data for testing the null hypothesis $H_0 : \lambda = 0$. Let $\Delta = \mu_0 - \mu$, and suppose that positive values of λ indicate positive

treatment effects. Type I error, power, and interval coverage were computed by sampling y_0 and y for true fixed values of Δ and λ, and then computing the 95% posterior credible intervals for λ. The entire process was repeated several thousand times. Note that $\Delta > 0$ (< 0) represents the case where the historical controls performed better (worse) than the concurrent controls. In addition, simulations of the Weibull model compared varying degrees of commensurability among the shape parameters, σ_0 and σ, through the ratio $\omega = \sigma_0/\sigma$, for true $\Delta = 0$. All commensurate prior models used in the simulations assume a vague, $Uniform(-50, 50)$ prior on $\log(\tau)$. The linear model uses the location commensurate prior discussed in Subsection 3.1 with a noninformative prior on σ^2, while the Weibull model uses the location-shape commensurate prior presented in Subsection 3.2.1.2. Operating characteristics of the linear model were simulated for true $\mu_0 = 1$, $\sigma_0^2 = \sigma^2 = 1$, $n_0 = 100$, and $n = 50$, and $\lambda = (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$. Note that results for $\lambda = 0$ correspond to Type I errors since the null hypothesis is true.

Power curves for the full borrowing (dashed), location commensurate (solid), and no borrowing (dotted) models are shown in the top row of Figure 3.4 for $\Delta = (-1, 0, 1)$. The graphs reveal that full borrowing from the historical data provides the best results when $\Delta = 0$, but extremely poor operating characteristics otherwise, including Type I error probabilities near 1. Furthermore, when the current controls perform worse than their historical counterparts, ($\Delta < 0$), pooling the two datasets reduces power, since the positive treatment effect vanishes when combined with these historical controls. Conversely, the location commensurate approach, while essentially recovering the full

borrowing result when $\Delta = 0$, is able to detect the conflict and diminish the historic control's influence. The bottom row of Figure 3.4 reveals analogous results for time-to-event analysis using the exponential model, for $\mu_0 = 2$, $n_0 = 200$, and $n = 100$.

We also compared approximate 95% posterior coverage and average interval length for inference on λ. The top row of Figure 3.5 plots the average coverage probability of the equal-tail 95% posterior credible intervals, averaged across seven values of λ, $(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$, for each of the three error distributions. The bottom row plots the average length of the 95% posterior credible intervals used for inference on λ. Results for full borrowing (dashed), commensurate (solid), and no borrowing (dotted) models are shown.

The left and center graphs plot results for varying levels of commensurability among the intercepts, indicated by Δ on the x-axis, for the Gaussian linear and exponential time-to-event models. Note that the no borrowing model results in a constant coverage probability of 0.95, and results in the widest intervals for $\Delta \in (-0.6, 0.6)$. The plots reveal that fully borrowing obtains the shortest intervals, which results in poor coverage for conflicting data, the outcome of a dogmatic analysis of the current trial. Conversely, given conflicting results, the adaptive location commensurate models discount the historical data to improve coverage. The plots suggest that the point of maximum ambiguity occurs near $\Delta = \pm 0.6$.

Let us turn our attention to the bottom row plots in Figure 3.5, corresponding to Gaussian (left) and exponential (center) errors. The graphs show that as Δ approaches

zero, intervals for the location commensurate models *tighten* toward those of their full borrowing counterparts, increasing the precision of the parameter estimate and leading to more powerful procedures than "no borrowing". For Gaussian errors (left), interval length behaves symmetrically with respect to $|\Delta| \neq 0$. Moving away from $\Delta = 0$, we see that interval length for the location commensurate model eventually surpasses "no borrowing" interval length when $|\Delta|$ is approximately 0.6, then peaks at $|\Delta| \approx \sigma$ (σ is fixed at 1), and finally decreases, tending toward "no borrowing" length for $|\Delta| > 1$. The general trend is similar for exponential errors (center), however here, interval length is *larger* for $\Delta < 0$ than $\Delta > 0$, although the difference is slight for $|\Delta| < 0.6$. Therefore, when the data conflict, incorporating historical controls that "progress" sooner on average than the current controls, $\Delta < -c$ for $c > 0$, actually decreases the precision of our parameter estimate when compared to the inverted case when historical controls survive longer on average, $\Delta > c$.

The right graphs, consider commensurability among the shape parameters for the Weibull time-to-event model, where the x-axis corresponds to ω, $\sigma_0 = 1$, and $\Delta = 0$. Recall that for $\sigma < 1$, the hazard rate becomes ever more increasing with time for $\sigma \to 0$. Therefore, small ω represents decreased survival for the historical controls when compared to concurrent controls. The graphs show that full borrowing from the historical data when ω is small, results in the shortest intervals and poorest coverage; and the largest intervals yet poorer coverage than "no borrowing" and the location commensurate models when ω is large. Coverage for the location commensurate model

is virtually identical to "no borrowing" coverage for all ω, yet when the shapes are commensurate, e.g. $\omega \in (0.5, 1.5)$, the adaptive approach results in shorter intervals and hence more precise parameter estimates.

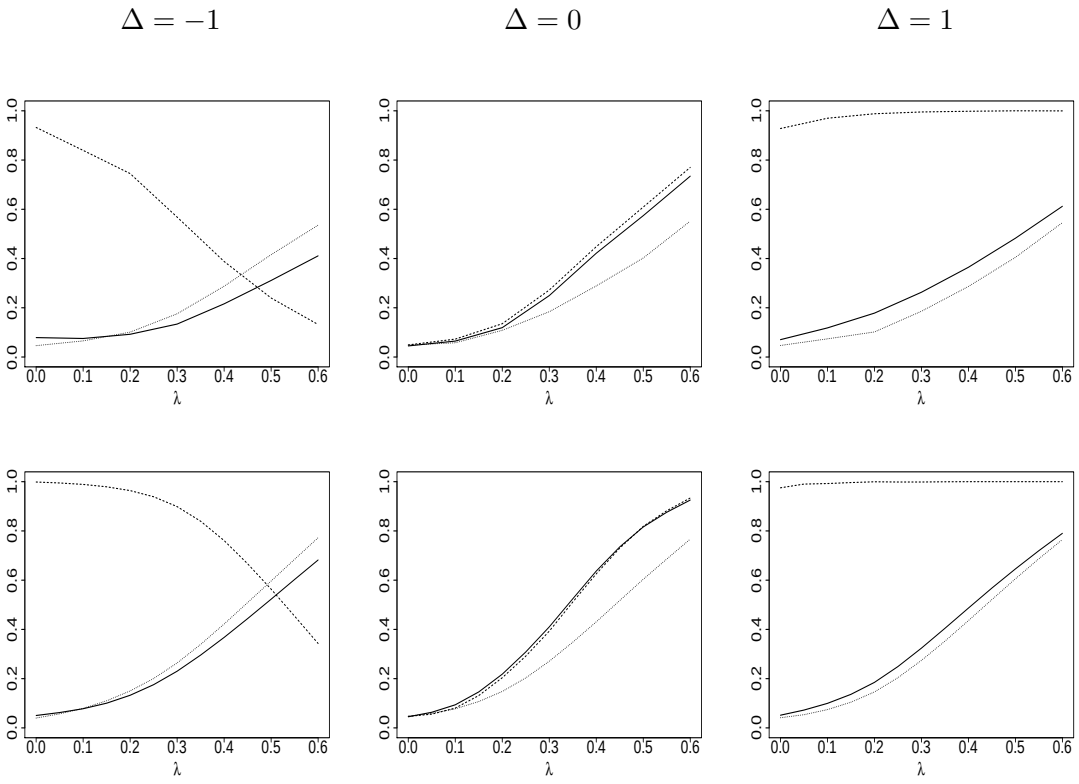

Figure 3.4: Power curves for the full borrowing (dashed), commensurate prior (solid), and "no borrowing" (dotted) models corresponding to hypothesis tests for λ based on the 95% posterior credible intervals. The top row graphs show results for the Gaussian linear model, where $\mu_0 = 1$, $\sigma_0^2 = \sigma^2 = 1$, $n_0 = 100$, and $n = 50$. The bottom row graphs show results for a time-to-event analysis using the exponential model, for $\mu_0 = 2$, $n_0 = 200$, and $n = 100$. Columns correspond to differences in the intercepts, $\Delta = \mu_0 - \mu$.

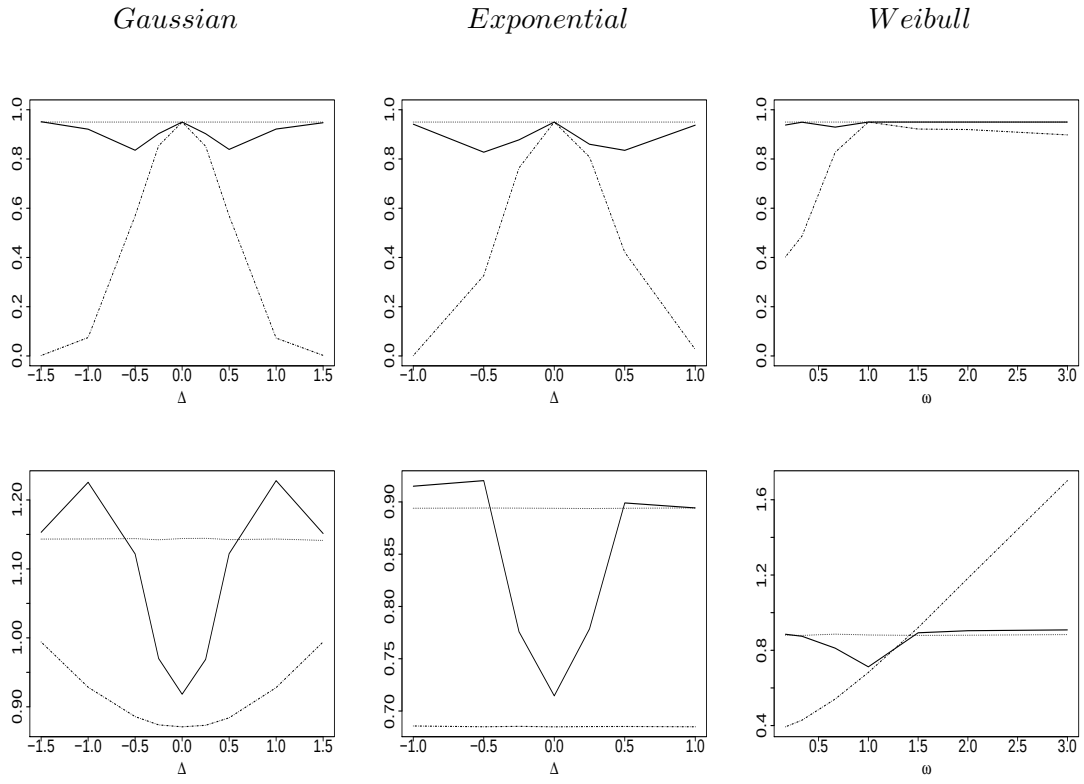

Figure 3.5: Plots of simulation results for varying degrees of commensurability among the historical and concurrent controls. The top row of plots show the average approximate probability that the equal-tail 95% posterior credible interval covers λ. The bottom row plots show average length of the 95% posterior credible intervals used for inference on λ. All graphs plot results for the full borrowing (dashed), commensurate prior (solid), and "no borrowing" (dotted) models. The left and center graphs show results for the Gaussian linear and time-to-event exponential models, where the x-axis corresponds to differences in the intercepts, $\Delta = \mu_0 - \mu$. The right graphs shows results for the Weibull model for $\Delta = 0$, where the x-axis corresponds to values of the shape ratio, $\omega = \sigma_0/\sigma$.

Chapter 4

Adaptive Randomization using Historical Controls

In this chapter, we investigate adaptive randomization for controlled, non-sequential clinical trials comparing a novel intervention to a previously studied control intervention, for which historical data is available to the investigators. Specifically, we propose a blocked randomization scheme rooted in our commensurate prior approach that balances the allocation ratio with respect to the amount of incorporated historical information. We refer to our approach as "optimal-balance" randomization, since we use current patients "optimally," with respect to the amount of information existing on the control therapy. This approach depends upon a reliable measure of the historical data's influence, namely the effective sample size of the historical controls, that is re-evaluated at

each block. The method is illustrated using binary response data in context of two successive clinical trials, where both trials identically measure $p-1$ covariates representing fixed effects that are to be incorporated into the analysis.

Section 4.1 introduces a useful latent variable representation of the location commensurate probit regression model. Then in Section 4.2 we define prior effective sample size for our adaptive commensurate prior method, and a balance function for our adaptive allocation ratio. Finally, Section 4.3 provides simulated realizations of our adaptive randomization scheme based upon the design of the FIRST Study, a trial comparing antiretroviral strategies in HIV-1-infected persons, described in MacArthur et al. (2001) and Berg-Wolf et al. (2006).

4.1 Probit Regression

Generalized linear regression models for binary response data are essential to the design and analysis of clinical trials. In this section we introduce the latent variable representation of the location commensurate probit regression model, which will be later used to analyze the data from FIRST Study and simulate our proposed adaptive randomization scheme. This formulation leads to closed form full conditionals, which allows posterior inference to proceed using the Gibbs Sampler (see Carlin and Louis, 2009, p.364).

Let $\mathbf{y_0}$ and $\mathbf{y}$ denote the historical and concurrent data such that $y_{0j} \sim Ber\left[\pi_0(X_{0j})\right]$, $\pi_0(X_{0j}) \in [0,1]$, for $j = 1, ..., n_0$, and $y_i \sim Ber\left[\pi(X_i, d_i)\right]$, $\pi(X_i, d_i) \in [0,1]$, for

$i = 1, ..., n$. Switching to the linear model notation used in Chapter 3, suppose y_0 and y are response vectors of length n_0 and n, X_0 and X are $n_0 \times p$ and $n \times p$ design matrices such that the first columns contains a vector of 1s corresponding to the intercept, β and β_0 are vectors of identically measured regression coefficients of length p, and d is an $n \times 1$ new intervention indicator. As mentioned in Section 3.2.1.1, probit regression transforms the means of y_0 and y through the inverse standard normal c.d.f. function, such that $\Phi^{-1}[\pi_0(X_0)] = \beta_0 X_0$ and $\Phi^{-1}[\pi(X)] = X\beta + d\lambda$.

The latent variable representation of the current data model arises by defining latent variables $y_i^\star \sim N(X_i\beta + d_i\lambda,\ 1)$ and assigning $y_i = 0$ if $y_i^\star \leq 0$, and $y_i = 1$ if $y_i^\star > 0$, such that $\pi(x_i) = P(y_i^\star > 0) = \Phi(X_i\beta + d_i\lambda)$. The joint posterior, $q(\beta, \lambda, y^\star, \tau | y_0, y)$, then follows as proportional to

$$\prod_{i=1}^{n} \Phi(x_i\beta + d_i\lambda)^{y_i} \left[1 - \Phi(x_i\beta + d_i\lambda)\right]^{1-y_i}$$
$$\times N_n\left(y^\star \,\middle|\, X\beta + d\lambda,\ I_n\right) \times N_p\left(\beta \,\middle|\, V^{-1}M, \frac{1}{\tau}V^{-1}\right) p(\tau), \tag{4.1}$$
$$\propto \prod_{i=1}^{n} \left[1(y_i = 1)1(y_i^\star > 0) + 1(y_i = 0)1(y_i^\star \leq 0)\right] N\left(y_i^\star \,\middle|\, X_i\beta + d_i\lambda,\ 1\right)$$
$$\times N_p\left(\beta \,\middle|\, V^{-1}M, \frac{1}{\tau}V^{-1}\right) p(\tau). \tag{4.2}$$

The full conditional posterior for $y_i^\star$ follows from (4.1) as a truncated normal,

$$q(y_i^\star | y_0, y, \beta, \lambda) \propto \prod_{i=1}^{n} 1(y_i^\star > 0)^{y_i} 1(y_i^\star \leq 0)^{1-y_i} \times N\left(y_i^\star \,\middle|\, X_i\beta + d_i\lambda,\ 1\right). \tag{4.3}$$

Finally, the full conditional posteriors for λ and β derived from (4.2) are normal, namely

$$q(\lambda | y_0, y, y^\star, \beta) \propto N\left(\frac{\sum_{i=1}^n d_i(y_i^\star - X_i\beta)}{\sum_{i=1}^n d_i},\ \left(\sum_{i=1}^{n} d_i\right)^{-1}\right),$$

and

$$q(\beta|y_0, y, y^\star, \lambda, \tau) \propto N_p\left(\left(X^TX + \tau V\right)^{-1}\left(X^T(y^\star - d\lambda) + \tau M\right),\ \left(X^TX + \tau V\right)^{-1}\right).$$

Marginal posterior inference on the commensurability parameter, τ, requires a nonconjugate mcmc sampling algorithm, such as Metropolis-Hastings. Recall that for probit regression, $\hat{W}_0$ in (3.11) is an $n_0 \times n_0$ diagonal matrix containing diagonal elements,

$$\hat{W}_{0jj} = \frac{\phi(X_0\hat{\beta}_0)^2_{jj}}{\Phi(X_0\hat{\beta}_0)_{jj}\left(1 - \Phi(X_0\hat{\beta}_0)_{jj}\right)}, \tag{4.4}$$

where $j = 1, ..., n_0$ and $\phi(.)$ is the standard normal p.d.f. McCulloch and Searle (2001, p.136).

4.2 Optimal Balanced-Randomization

Friedman, Furberg, and DeMets (1998, p.69) broadly refer to randomization procedures that adjust the allocation ratio as the study progresses as *adaptive*. These methods are commonly used to balance prognostic factors among the intervention arms, or if the primary response variable can be determined quickly, to assign patients to the better performing regimen in early phase trials. Algorithms for implementing adaptive randomization schemes for frequentist designs can be found in Friedman, Furberg, and DeMets (1998, p.69–78).

Commensurate prior models for controlled trials naturally advocate a randomization scheme that is "optimal" with respect to the amount of strength borrowed from the historical controls. Suppose investigators would like to learn as much as possible about

the efficacy and safety profile of the new intervention, and so they wish to incorporate the historical controls, in order to use more of the new patients to learn about the new treatment. We propose adjusting the probability of allocating the next patient, or block of patients, to the new intervention according to the commensurability of the historical and current controls, evaluated sequentially using repeated fitting of the commensurate prior model.

This can be achieved by defining "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 as the value that minimizes the distance, defined in terms of the curvature of the logarithm of each density, between the parametric prior and a sequence of posteriors corresponding to a range of potential sample sizes. Following their notation, let $f(Y|\theta)$ denote the probability distribution function of an $s-$dimensional random vector Y, and let $p(\theta|\tilde{\theta})$ be the prior on the parameter vector $\theta = (\theta_1, ..., \theta_d)$, where $\tilde{\theta}$ denotes the vector of hyperparameters. Suppose we have an independent and identically distributed sample, $\mathbf{Y}_m = (Y_1, ..., Y_m)$ and let $q_m(\theta|\tilde{\theta}, \mathbf{Y}_m) \propto p_0(\theta|\tilde{\theta}_0) \prod_{i=1}^{m} f(Y_i|\theta)$ denote the posterior obtained from a sample of size m, where $p_0(\theta|\tilde{\theta}_0)$ denotes an "$\epsilon-$information prior" which has the same mean and correlations as prior, $p(\theta|\tilde{\theta})$, yet inflated variance. Morita et al. (2008) define ESS as the sample size m, that would minimize the distance between the normal approximations of the prior $p(\theta|\tilde{\theta})$ and the posterior $q_m(\theta|\tilde{\theta}, \mathbf{Y}_m)$. Specifically,

let

$$D_{p,j} = -\frac{\partial^2 \log\{p(\theta|\tilde{\theta})\}}{\partial \theta_j^2}, \tag{4.5}$$

and

$$D_{q,j}(m, \theta, \mathbf{Y}_m) = -\frac{\partial^2 \log\{q_m(\theta|\tilde{\theta}_0, \mathbf{Y}_m)\}}{\partial \theta_j^2}, \tag{4.6}$$

where $j = 1, ..., d$. For sample size m, the distance between $p(\theta|\tilde{\theta})$ and $q_m(\theta|\tilde{\theta}, \mathbf{Y}_m)$ is defined as the difference of the trace of the Fisher information matrix of the prior and the expected Fisher information matrix of the posterior,

$$\delta(m, \bar{\theta}, p, p_0) = \Big| \sum_{j=1}^{d} D_{p,j}(\bar{\theta}) - \sum_{j=1}^{d} \int D_{q,j}(m, \bar{\theta}, \mathbf{Y}_m) f_m(\mathbf{Y}_m) d\mathbf{Y}_m \Big|, \tag{4.7}$$

where expectation is with respect to marginal the distribution of $\mathbf{Y}_m$, $f_m(\mathbf{Y}_m) = \int \prod_{i=1}^{m} f(Y_i|\theta) p(\theta|\tilde{\theta}) d\theta$, and $\bar{\theta}$ denotes the prior mean under $p(\theta|\tilde{\theta})$. Then the ESS of $p(\theta|\tilde{\theta})$ with respect to the likelihood $\prod_{i=1}^{m} f(Y_i|\theta)$ is the integer m that minimizes $\delta(m, \bar{\theta}, p, p_0)$ (Morita et al., 2008).

We expand their method to include our adaptive commensurate prior approach for computing the prior effective sample size of β, which we refer to as the *effective number of historical controls*, (EHC). To compute EHC we must first find the posterior distribution of τ. Then we implement the Morita et al. method with respect to (3.11), with τ replaced by its posterior median. The new treatment allocation probability, or balance function, for the $(j+1)$st new patient follows as

$$\delta_j = \frac{C_j + EHC_j}{T_j + C_j + EHC_j}, \tag{4.8}$$

where T_j and C_j denote the number of subjects randomized to new treatment and control, respectively, in the current trial following the jth 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. For the univariate Gaussian power prior models presented in Section 2.2, EHC is equal to the posterior median of α_0 multiplied by the historical sample size, n_0.

This method is considered here only for non-sequential settings; that is, designs involving a single decision at the end of the trial. The method we present here is sequential in assessing commensurability, just not in deciding whether to stop early for efficacy or futility. However, sequential analysis and monitoring of clinical trials using commensurate priors is an obvious area for future work. Seminal papers in this area from the frequentist perspective include Pocock (1977), O'Brien and Fleming (1979), and Lan and DeMets (1983); see Freedman and Spiegelhalter (1989), Spiegelhalter, Freedman, and Parmar (1994), and Carlin, Kadane, and Gelfand (1998) for work using the Bayesian paradigm.

4.3 Example using HIV Antiretroviral Strategies Trial Data

We now present a numerical example based on Community Programs for Clinical Research on AIDS (CPCRA) protocol 058, the Flexible Initial Retrovirus Suppressive

Therapies (FIRST) trial. As explained in MacArthur et al. (2001), this was a large, long-term, randomized, prospective comparison of three different antiretroviral strategies (protease inhibitor (PI) only, non-nucleoside reverse transcriptase inhibitor (NNRTI), and PI plus NNRTI-containing regimens) in highly active antiretroviral therapy-naive, HIV-1-infected persons. Patients within all three strategies were also assigned one or two nucleoside reverse transcriptase inhibitors (NRTIs), most commonly abacavir plus lamivudine (ABC+3TC) and didanosine plus stavudine (ddI+d4T). The three strategies (and specific antiretroviral regimens) 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. A novel aspect of the trial was that, before randomization, patients and their clinicians within the two strategy arms involving NNRTIs were given the option of preselecting the NNRTI drug, either nevirapine (NVP) or efavirenz (EFV), or allowing an additional randomization to NVP or EFV. Patients in the latter group were randomized to study-specified drugs at the time that they were randomized to a strategy arm.

We illustrate our adaptive method assuming that current trial consists of data from the randomized arms of FIRST, where EFV ($n = 99$) is the novel intervention that is compared to control therapy, NVP ($n = 104$). We then consider patients who chose the NVP ($n = 237$) in the observation arm as historical controls. Our goal is to compare the probabilities of virological suppression (HIV RNA < 50 copies/mL) under EFV and NVP at 32 weeks, using the commensurate prior probit regression model described

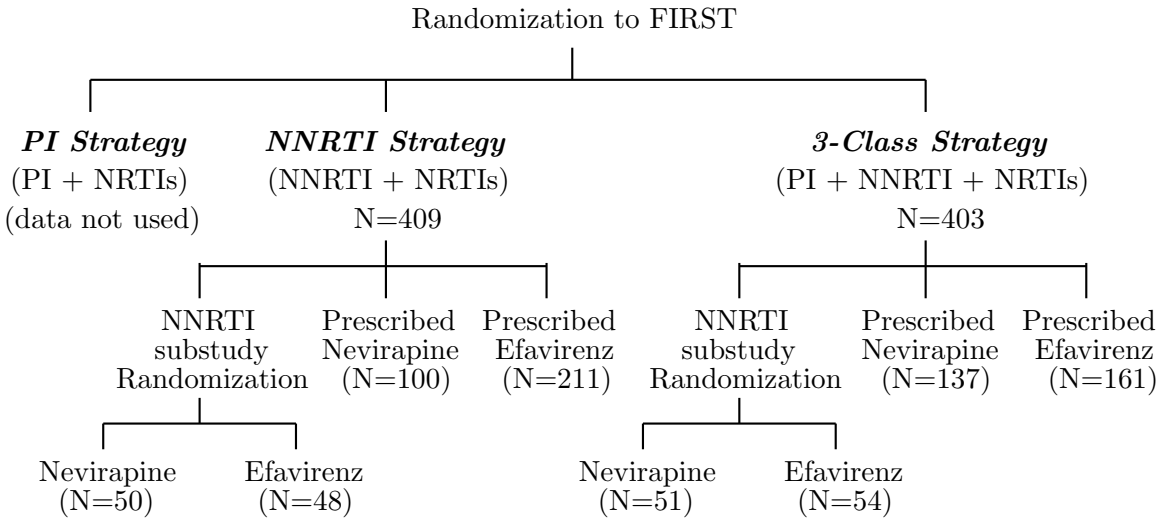

Figure 4.1: Outline of FIRST design and randomization for eligible subjects (Berg-Wolf et al., 2006).

in Section 4.1. Note that 54 of the patients randomized to EFV, 51 of the patients randomized to NVP, 161 of the patients whose physician choose EFV, and 137 of the patients whose physician chose NVP had a regimen that also included one PI. The data set excludes patients missing an 8-month plasma HIV RNA measurement. Figure 4.1 offers an pictorial representation of the study design. Note the PI strategy did not include NNRTI regimens and therefore is not used in our illustration.

Let y_0 and y denote responses for the historical and concurrent data and assume models $E[y_0] = \mu_0$ and $E[y] = \mu + d\lambda$, where d is a indicator for new treatment and μ_0 and μ are intercepts corresponding to the historical and current NVP. Simulating the adaptive randomization method proceeds iteratively by assigning each patient to NVP or EFV and generating responses for several fixed values of $\Delta = \mu_0 - \mu$. Note that $\Delta = -0.221$ represents the posterior median obtain from fitting the real data from

Table 4.1: Proportion of virological suppression for FIRST patients under NVP and EFV at 32 weeks.

	NVP	EFV
Nonrandomized	136/237	249/372
Randomized	54/104	57/99

FIRST. We also fixed λ to its FIRST posterior median, 0.23. Patients are allocated equally to NVP and EFV for the first 80 assignments. Then after randomizations $80, 100, 120$, and 160, EHC is computed and the allocation probability, δ, is updated accordingly. Note that the analog to (4.5) for μ is $\left(\frac{1}{\hat{\tau}} + \sum_{j=1}^{n_0} \hat{W}_{0jj}\right)^{-1}$, where $\hat{W}_{0jj}$ follows from (4.4) with $X_0\hat{\beta}_0$ replaced by the historical MLE of μ_0, $\hat{\mu}_0$. The expected Fisher information matrix for μ requires numerical approximation. We use the general method described in Morita et al. (2008, p.600). Table 4.1 contains proportions of the number of patients who experienced virological suppression at 32 weeks under NVP and EFV for both nonrandomized (historical) and randomized (current) data. Note that the nonrandomized subjects experienced better for virological suppression within both treatment groups. Fitting the usual frequentist probit regression model with two predictors, one $0-1$ indicator variable for randomization status (nonrandomized vs randomized) and another $0-1$ indicator variable for treatment (NVP vs EFV), results in a positive treatment effect for EFV that is significant at the 0.05 significance level.

Table 4.2 summarizes the results for the average EHC and treatment allocation ratio, $\delta/(1-\delta)$, as well as the average total number of subjects assigned to the NVP and EFV arms, for $n = 203$ total assignments. Results after randomization $80, 100, 120, 160$ and

the final analysis are shown by row. The last row contains total the number of subjects randomized to each arm on average at the end of the trial. The columns vary by Δ representing degrees of commensurability among the intercepts. The table clearly shows that for $\Delta = 0$, (fifth column), our method successfully incorporates all of the historical controls ($n_0 = 237$) and adjusts the allocation ratio to assign many more new patients to EFV, culminating on average in 65 new patients assigned to NVP versus 138 new patients assigned to EFV. Furthermore, for large $|\Delta|$, we see that the method is able to properly identify the need for more concurrent information about NVP. For example, in the first column where $\Delta = -2.25$, EHC is only 1, and the allocation to treatment is balanced on average, since after $n = 203$ assignments, 101 are assigned to NVP and 102 are assigned to EFV. Notice that for $\Delta = -0.221$, $EHC = 235$ (very close to the upper bound of 237) for all rows. Therefore, our commensurate model considers the actual observational and randomized NVP data from FIRST to be quite commensurate.

Figure 4.2 augments Table 4.2 by plotting the average value of the balance function, δ, after each re-assessment block: $80, 100, 120$, and 160. Each line represents a different value of Δ. Here we see that the probability of allocation to EFV is large when the historical and current NVP data are most commensurate (small Δ). Then, the imbalance begins to subside as many more new patients are assigned to EFV. Conversely, when the data conflict, on average, δ remains relatively flat near 0.5.

Table 4.2: Results for adaptive randomization to EFV, for $n = 203$ total randomizations, where $n_0 = 237$, Ratio=$\delta/(1-\delta)$:1.

block	$\Delta = -2.25$	$\Delta = -1.5$	$\Delta = -0.75$	$\Delta = -0.221$	$\Delta = 0$
	EHC, Ratio	EHC, Ratio	EHC, Ratio	EHC, Ratio	EHC, Ratio
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
	NVP, EFV	NVP, EFV	NVP, EFV	NVP, EFV	NVP, EFV
Total	101, 102	99, 104	77, 126	66, 137	65, 138

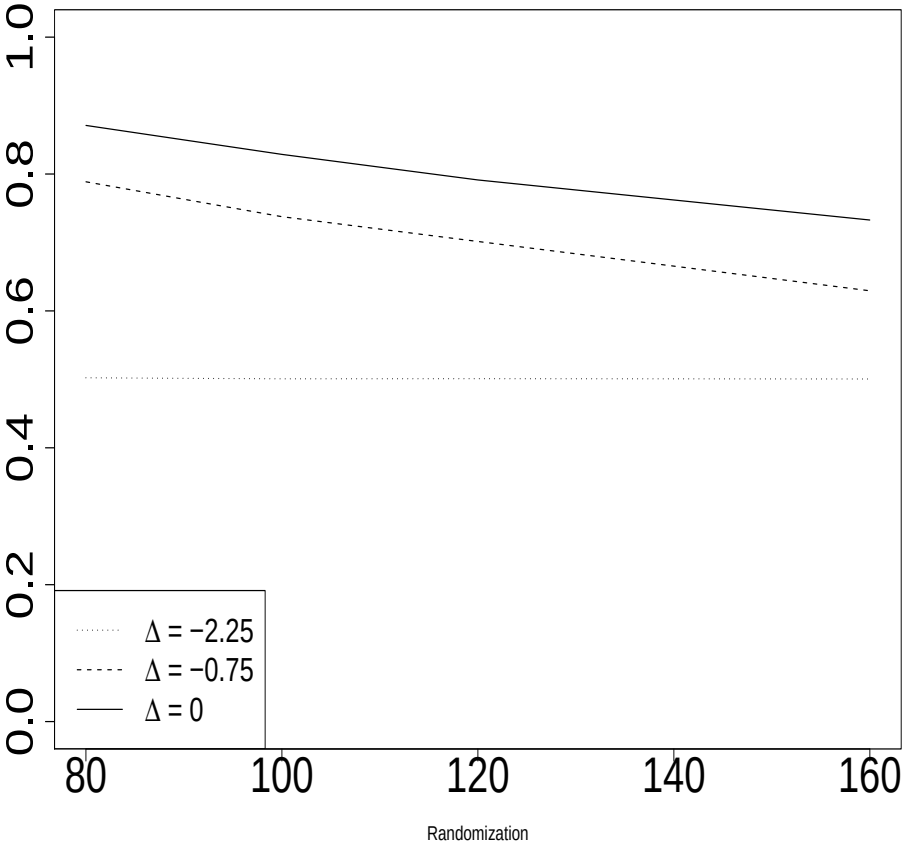

Figure 4.2: Plots of average value of balance function, δ, after randomization $80, 100, 120, 160$ for values of Δ, indicated by line type, corresponding to degrees of commensurability among the historical and current data.

Chapter 5

Discussion and Future work

In this dissertation we began by introducing classes of hierarchical models for Gaussian data using priors that facilitate adaptive borrowing from historical data when this is justified by its commensurability with the accumulating current data. Then we extended the commensurate prior method to facilitate adaptive borrowing from historical data for linear and linear mixed models as well as generalized linear and generalized linear mixed regression models in the context of two successive clinical trials. Furthermore, we proposed an adaptive randomization scheme rooted in our commensurate prior method that balances the allocation ratio with respect to the amount of incorporated historical information. 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).

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 in the current and historical studies differ; estimating nuisance parameters like σ^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 include time-dependent covariates, smoothed hazards, multiple events, and semi-parametric Cox models as e.g. Kalbfleisch (1978) or Ibrahim, Chen, and Sinha (2001, p.47). Our adaptive borrowing method might be extended to assess treatment compliance among subjects in historical and current trials in the context of a non-inferiority trial. Another promising avenue for evolution involves developing a general method for sequential analysis and monitoring of clinical trials using commensurate priors extended from decision theoretic or predictive probability approaches outlined in (Berry et al. 2010). Further expansions could incorporate "optimal balance" randomization schemes and multiple outcomes as in Zhao, Grambsch, and Neaton (2007).

A final, perhaps most interesting path for the future lies in adaptively borrowing from concurrent observational arms for the purpose of sequential assessment of futility, in order to stop a trial early or, perhaps drop treatment arms that are revealed to

be ineffective. Prentice et al. (2005), show that incorporating historical controls from observational studies in seemingly identical populations requires considerable care to control for confounding and minimize bias. However, these methods would be valuable in settings where interventions are costly to produce and maintain, sample sizes are smaller, randomization may not be ethical or feasible, or numerous promising investigations therapies await further study.

References

Allocco, D. J., Cannon, L. A., Britt, A., Heil, J. E., Nersesov, A., Wehrenberg, S., Dawkins, K. D., and Kereiakes, D. J. (2010). A prospective evaluation of the safety and efficacy of the TAXUS Element paclitaxel-eluting coronary stent system for the treatment of de novo coronary artery lesions: Design and statistical methods of the PERSEUS clinical program. *Trials* **11**, 1–15.

Berry, D. A. (1993). A case for Bayesianism in clinical trials (with discussion). *Statistics in Medicine* **12**, 1377–1404.

Berry, D. A. (2006). Bayesian clinical trials. *Nature Reviews Drug Discovery* **5**, 27–36.

Berry, S. M., Carlin, B. P., Lee, J. J., and Müller, P. (2010). *Bayesian Adaptive Methods for Clinical Trials*. Boca Raton, FL: Chapman and Hall/CRC Press.

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

Biswas, S., Liu, D. D., Lee, J. J., and Berry, D. A. (2009). Bayesian clinical trials at the University of Texas M. D. Anderson Cancer Center. *Clinical Trials* **6** 205–216.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Browne, W. J. and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* **1**, 473–514.

Berg-Wolf, M., Peng, G., Xiang, Y., Huppler-Hullsiek, K., MacArthur, R.D., Novak, R.M., Kozal, M., Schmetter, B., Henely, C., and Dehlinger, M. (2006). Long-term comparison of Nevirapine versus Efavirenz when combined with other antiretroviral drugs in HIV-1 positive antiretroviral-naive persons: the NNRTI substudy of the CPCRA 058 FIRST Study. Poster presented at the XVI International AIDS Conference, August 17, 2006, Toronto, Canada.

Carlin, B. P., Kadane, J. B., and Gelfand, A. E. (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics* **54** 964-975.

Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data Analysis*. 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Chaloner, K. (1987). A Bayesian approach to the estimation of variance components in the unbalanced one-way random-effects model. *Technometrics*, **29**, 323–337.

Chen, M-H. and Ibrahim, J. G. (2006). The relationship between the power prior and hierarchical models. *Bayesian Analysis* **1**, 554–571.

Cheng, Y. and Shen, Y. (2005). Bayesian adaptive designs for clinical trials. *Biometrika* **92**, 633–646.

Cui, Y., Hodges J. S., Kong, X., and Carlin, B. P. (2010). Partitioning degrees of freedom in hierarchical and other richly-parameterised models. *Technometrics* **52**, 124–136.

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

Duan, Y., Ye, K., and Smith, E. P. (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics* **17**, 95–106.

Freedman L. S, and Spiegelhalter, D. J. (1989). Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Controlled Clinical Trials* **10** 357-367.

Friedman L., Furberg C., DeMets, D. L. (1998). *Fundamentals of Clinical Trials.* 3rd ed. New York, NY: Springer-Verlag.

Fúquene, J. A., Cook, J. D., and Pericchi, L. R. (2009). A case for robust Bayesian priors with applications to clinical trials. *Bayesian Analysis* **4**, 817–846.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–534.

Goldberg, R. M., Sargent, D. J., Morton, R. F., Fuchs, C. S., Ramanathan, R. K., Williamson, S. K., Findlay, B. P., Pitot, H. C., and Alberts, S. R. (2004). A randomized controlled trial of Fluorouracil Plus Leucovorin, Irinotecan, and Oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *Journal of Clinical Oncology* **22**, 23–30.

Hobbs, B. P. and Carlin, B. P. (2008). Practical Bayesian design and analysis for drug

and device clinical trials. *Journal of Biopharmaceutical Statistics* **18**, 54–80.

Hodges, J. S. and Sargent, D. J. (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika* **88**, 367–379.

Ibrahim, J. G. and Chen, M-H. (2000). Power prior distributions for regression models. *Statistical Science* **15**, 46–60.

Ibrahim, J. G., Chen, M-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag.

Ibrahim, J. G., Chen, M-H., and Sinha, D. (2003). On optimality properties of the power prior. *Journal of the American Statistical Association* **98**, 204–213.

Irony, T. (2008). Personal communication.

Kalbfleisch, J. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society (Series B)* **40**, 214–221.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd ed. New York: Wiley.

Kass, R. E. and Natarajan, R. (2006). A Default conjugate prior for variance components in generalized linear mixed models. *Bayesian Analysis* **1**, 535–542.

Klein J. P. and Moeschberger M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. New York: Springer-Verlag.

Lan K. K. G., and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials.*Biometrika* **70**, 659-663.

MacArthur, R.D., Chen, L., Mayers, D.L., Besch, C.L., Novak, R., Berg-Wolf, M., Yurik, T., Peng, G., Schmetter, B., Brizz, B., and Abrams, D. for the CPCRA 058 FIRST Trial Study Team (2001). The Rationale and Design of the CPCRA (Terry Beirn Community Programs for Clinical Research on AIDS) 058 FIRST (Flexible Initial Retrovirus Suppressive Therapies) Trial. *Controlled Clinical Trials*, **22**, 176–190.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* New York: Wiley.

Morita, S., Thall, P. F., and Müller, P. (2008). Determining the effective sample size of a parametric prior. *Biometrics* **64**, 595–602.

Neelon B., O'Malley A. J., and Margolis P. A. (2008). Bayesian Analysis Using Historical Data with Application to Pediatric Quality of Care. In *Proceedings of the Joint Statistical Meeting*, pages 2960–2967, Denver. ASA Section on Bayesian Statistical Sciences Section.

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

Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). A note of the power prior. *Statistics in Medicine* **28**, 3562–3566.

Neuenschwander, B., Capkun-Niggli, G., Branson, M., and Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials* **7**, 5–18.

OBrien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35**, 549-556.

Pericchi, L. R. (2009). Personal communication.

Pocock S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64**, 191-199.

Prentice, R. L., Langer, R., Stefanick, M. L., Howard, B. V., Pettinger, M., Anderson, G., Barad, D., Curb, J. D., Kotchen, J., Kuller, L., Limacher, M., and Wactawski-Wende, J., for the Womens Health Initiative Investigators. (2005). Combined Postmenopausal Hormone Therapy and Cardiovascular Disease: Toward Resolving the Discrepancy between Observational Studies and the Womens Health Initiative Clinical Trial. *American Journal of Epidemiology* **162**, 404–414.

Reich, B. J. and Hodges, J. S. (2008). Identification of the variance components in the general two-variance linear model. *Journal of Statistical Planning and Inference* **138**, 1592–1604.

Saltz L. B., Cox J. V., Blanke C., et al. and the Irinoteca Study Group (2000). Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. *The New England Journal of Medicine* **343**, 905–914.

Severini, T. A. (1994). On the relationship between Bayesian and non-Bayesian interval estimates. *Journal of the Royal Statistical Society (Series B)*, **53**, 611–618.

Spiegelhalter, D.J., Freedman, L.S., and Parmar, M.K.B. (1994). Bayesian approaches to randomized trials (with discussion). *Journal of the Royal Statistical Society (Series A)*, **157**, 357–416.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* London: Wiley.

Wang, B. and Titterington, D. M. (2006). Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model. *Bayesian Analysis* **1**, 625–650.

Zhao, Y., Grambsch, P. M., Neaton, J. D. (2007). A decision rule for sequential monitoring of clinical trials with a primary and supportive outcome. *Clinical Trials* **4**, 140–153.