Bayesian Analysis (2006) **1**, Number 3, pp. 551–574

The Relationship Between the Power Prior and Hierarchical Models

Ming-Hui Chen*, and Joseph G. Ibrahim†

Abstract.

The power prior has emerged as a useful informative prior for the incorporation of historical data in a Bayesian analysis. Viewing hierarchical modeling as the "gold standard" for combining information across studies, we provide a formal justification of the power prior by examining formal analytical relationships between the power prior and hierarchical modeling in linear models. Asymptotic relationships between the power prior and hierarchical modeling are obtained for non-normal models, including generalized linear models, for example. These analytical relationships unify the theory of the power prior, demonstrate the generality of the power prior, shed new light on benchmark analyses, and provide insights into the elicitation of the power parameter in the power prior. Several theorems are presented establishing these formal connections, as well as a formal methodology for eliciting a guide value for the power parameter a_0 via hierarchical models.

Keywords: Generalized linear model, hierarchical model, historical data, power prior, prior elicitation, random effects model

1 Introduction

The power prior discussed in Ibrahim and Chen (2000) has emerged as a useful class of informative priors for a variety of situations in which historical data is available. Several applications to clinical trials and epidemiological studies using the power prior have appeared in the literature. Examples of the use of the power prior and its modifications in clinical trials and carcinogenicity studies include Berry (1991), Eddy et al. (1992), Berry and Hardwick (1993), Lin (1993), Spiegelhalter et al. (1994), Berry and Stangl (1996), Chen et al. (2000), Ibrahim and Chen (1998), Ibrahim et al. (1998), Ibrahim et al. (1999), Chen et al. (1999a), Chen et al. (1999b), Ibrahim and Chen (2000), and Ibrahim et al. (2003). Examples using the power prior in epidemiological studies include Greenhouse and Wasserman (1995), Brophy and Joseph (1995), Fryback et al. (2001), and Tan et al. (2002). A recent book illustrating the use of the power prior in epidemiological studies is Spiegelhalter et al. (2003).

The power prior for a general regression model can be constructed as follows. Suppose we have historical data from a similar previous study, denoted by $D_0 = (n_0, \boldsymbol{y}_0, X_0)$, where n_0 is the sample size of the historical data, y_0 is the $n_0 \times 1$ response vector, and

*Department of Statistics, University of Connecticut, Storrs, CT, http://www.stat.uconn.edu/~mhchen

†Department of Biostatistics, University of North Carolina, Chapel Hill, NC, http://www.sph.unc.edu/bios/facstaff/

© 2006 International Society for Bayesian Analysis ba0003

X_0 is the $n_0 \times p$ matrix of covariates based on the historical data. Further, let $\pi_0(\boldsymbol{\theta})$ denote the prior distribution for $\boldsymbol{\theta}$ before the historical data D_0 is observed. We shall call $\pi_0(\boldsymbol{\theta})$ the *initial prior* distribution for $\boldsymbol{\theta}$. Let the data from the *current* study be denoted by $D = (n, \boldsymbol{y}, X)$, where n denotes the sample size, $\boldsymbol{y}$ denotes the $n \times 1$ response vector, and X denotes the $n \times p$ matrix of covariates. Further, denote the likelihood for the current study by $L(\boldsymbol{\theta}|D)$, where $\boldsymbol{\theta}$ is a vector of indexing parameters. Thus, $L(\boldsymbol{\theta}|D)$ is a general likelihood function for an arbitrary regression model, such as linear models, generalized linear model, random effects model, nonlinear model, or a survival model with censored data. Given the discounting parameter a_0, we define the *power prior* distribution of $\boldsymbol{\theta}$ for the current study as

$$\pi(\boldsymbol{\theta}|D_0, a_0) \propto L(\boldsymbol{\theta}|D_0)^{a_0} \ \pi_0(\boldsymbol{\theta}) \ , \tag{1}$$

where a_0 weights the historical data relative to the likelihood of the current study, and thus the parameter a_0 controls the influence of the historical data on $L(\boldsymbol{\theta}|D)$. The parameter a_0 can be interpreted as a precision parameter for the historical data. Since D_0 is historical data, it is unnatural in most applications — including clinical trials and carcinogenicity studies — to weight the historical data more than the current data; thus it is scientifically more sound to restrict the range of a_0 to be between 0 and 1, and thus we take $0 \leq a_0 \leq 1$. One of the main roles of a_0 is that it controls the heaviness of the tails of the prior for $\boldsymbol{\theta}$. As a_0 becomes smaller, the tails of (1) become heavier. Setting $a_0 = 1$, (1) corresponds to the update of $\pi_0(\boldsymbol{\theta})$ using Bayes theorem. That is, with $a_0 = 1$, (1) corresponds to the posterior distribution of $\boldsymbol{\theta}$ based on the historical data. When $a_0 = 0$, then the prior does not depend on the historical data D_0; in this case, $\pi(\boldsymbol{\theta}|D_0, a_0 = 0) \equiv \pi_0(\boldsymbol{\theta})$. Thus, $a_0 = 0$ is equivalent to a prior specification with no incorporation of historical data. Therefore, (1) can be viewed as a generalization of the usual Bayesian update of $\pi_0(\boldsymbol{\theta})$. The parameter a_0 allows the investigator to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the previous and current studies, or when the sample sizes of the two studies are quite different. One of the most useful applications of the power prior is for model selection problems since it inherently automates the informative prior specification for all possible models in the model space (see (Chen et al. 1999b), (Ibrahim et al. 1999), and (Ibrahim and Chen 2000)).

One of the most common ways of combining several datasets or incorporating prior information is through hierarchical modeling. Hierarchical modeling is perhaps the most common and best known method for combining several sources of information. In this paper, we establish a formal analytic connection between the power prior and hierarchical models, both for the normal linear model and non-normal models including the class of generalized linear models. This relationship is accomplished by establishing an analytic relationship between the power parameter a_0 and the variance components of the hierarchical model. Establishing this relationship is critical since it formally justifies the power prior based on hierarchical modeling as well as guides the user into the choice of a_0 from which sensitivity analyses can be based. Such a "benchmark" power prior would be the prior that has an a_0 value which corresponds to equivalence between the hierarchical model and the power prior. Thus in this situation, the benchmark power prior analysis would be the one that corresponds to hierarchical modeling. Thus, the analytical connection we make is important and useful since it characterizes the relationship between the power parameter a_0 and the variance components in the hierarchical model, thereby motivating and justifying the use of the power prior as an informative prior for incorporating historical data, as well as providing a semi-automatic elicitation scheme for a_0 based on hierarchical modeling. Indeed, one of the most difficult and elusive issues in the use of the power prior is the choice of a_0. Since there will typically be little information in the historical data about a_0, it is not at all clear how to use the historical data in eliciting a_0. The analytic relationship between a_0 and the hyperparameters of the hierarchical model provides the data analyst with elicitation strategies for a_0 based on the historical data.

The rest of this paper is organized as follows. We consider the normal hierarchical linear model in Section 2 and develop formal analytical relationships between the power prior and the hierarchical regression model. In Section 3 we develop formal analytical relationships between the power prior and the hierarchical generalized linear model. In Section 4, we extend our results to multiple historical datasets. In Section 5, we use the analytic relationship between a_0 and the hyperparameters of the hierarchical model to devise a formal elicitation method for a_0 based on the historical data. In Section 6, we present a real dataset to demonstrate the main results.

2 Hierarchical Normal Models

2.1 I.I.D. Case

We first consider the i.i.d. case with a single historical dataset. The model for this case can be written as

$$y_i = \theta + \epsilon_i,\ i = 1, 2, \ldots, n, \quad \text{and} \quad y_{0i} = \theta_0 + \epsilon_{0i}, i = 1, 2, \ldots, n_0, \tag{2}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ denotes the current data with sample size of n and $\boldsymbol{y}_0 = (y_{01}, y_{02}, \ldots, y_{0n_0})$ denotes the historical data with sample size of n_0. In (2), we further assume that the ϵ_i are i.i.d. $N(0, \sigma^2)$ and the ϵ_{0i} are i.i.d. $N(0, \sigma_0^2)$ and independent of the ϵ_i's, where σ^2 and σ_0^2 are fixed parameters. Now the hierarchical model is completed by independently taking

$$\theta_0 \mid \mu, \tau^2 \sim N(\mu, \tau^2), \qquad \theta \mid \mu, \tau^2 \sim N(\mu, \tau^2), \tag{3}$$

and then taking

$$\mu \sim N(\alpha, \nu^2), \tag{4}$$

where α, ν^2, and τ^2 are all fixed hyperparameters. Within the development of this hierarchical model, our goal is to make inferences about the current study through the marginal posterior distribution $[\theta|\boldsymbol{y}, \boldsymbol{y}_0]$ ($[a|b]$ denotes the marginal distribution of a given b throughout). Here, θ is the parameter of interest for the current study, such as a treatment effect, and θ_0 denotes the corresponding parameter based on the historical study. The following theorem gives the form of the marginal posterior distribution $[\theta|\boldsymbol{y}, \boldsymbol{y}_0]$ obtained from (2), (3), and (4).

Theorem 2.1 *Letting* $\bar{y}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0} y_{0i}$ *and* $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, *we have*

$$(\theta \mid \boldsymbol{y}, \boldsymbol{y}_0) \sim N(\mu_h, \sigma_h^2), \tag{5}$$

where

$$\mu_h = \sigma_h^2 \left[\frac{n\bar{y}}{\sigma^2} + \frac{\alpha}{\tau^2\nu^2\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)} + \frac{\frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}\left(\frac{n_0\bar{y}_0}{\sigma_0^2} + \frac{\alpha}{\tau^2\nu^2\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}\right)}{\frac{n_0}{\sigma_0^2} + \frac{1}{\tau^2} - \frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}} \right] \tag{6}$$

and

$$\sigma_h^2 = \left[\frac{n}{\sigma^2} + \frac{1}{\tau^2} - \frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)} - \frac{1}{\left(\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)\right)^2\left(\frac{n_0}{\sigma_0^2} + \frac{1}{\tau^2} - \frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}\right)} \right]^{-1}. \tag{7}$$

The proof of Theorem 2.1 is given in Appendix B. Now we consider a power prior formulation of the model in (2). To do this, we set $\theta_0 = \theta$, and the resulting model becomes

$$y_i = \theta + \epsilon_i, \quad \text{and} \quad y_{0i} = \theta + \epsilon_{0i}. \tag{8}$$

Thus in the power prior formulation, the ϵ_i are i.i.d. $N(0, \sigma^2)$, and the ϵ_{0i} are i.i.d. $N(0, \sigma_0^2)$ and independent of the ϵ_i's, where σ^2 and σ_0^2 are fixed. Under this model, the power prior based on the historical data y_0 using the initial prior $\pi_0(\theta) \propto 1$, is given by

$$\pi(\theta|\boldsymbol{y}_0) \propto \exp\left\{ -\frac{a_0}{2\sigma_0^2}\sum_{i=1}^{n_0}(y_{0i}-\theta)^2 \right\}. \tag{9}$$

Straightforward calculations show that

$$(\theta|\boldsymbol{y}, \boldsymbol{y}_0) \sim N(\mu_p, \sigma_p^2), \tag{10}$$

where

$$\mu_p = \frac{\frac{n\bar{y}}{\sigma^2} + a_0\frac{n_0\bar{y}_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + a_0\frac{n_0}{\sigma_0^2}}, \tag{11}$$

and

$$\sigma_p^2 = \frac{1}{\frac{n}{\sigma^2} + a_0\frac{n_0}{\sigma_0^2}}. \tag{12}$$

We now examine the relationship between (5) and (10). To do this, we need to find an explicit relationship between μ_h and μ_p as well as a relationship between σ_h^2 and σ_p^2. We are led to the following theorem which characterizes this relationship.

Theorem 2.2 *The posteriors in (5) and (10) match, i.e., $\mu_h = \mu_p$ and $\sigma_h^2 = \sigma_p^2$ if and only if $\alpha = 0$ and $\nu^2 \to \infty$ in (4), and*

$$a_0 = \frac{1}{\frac{2\tau^2 n_0}{\sigma_0^2} + 1}. \tag{13}$$

The proof of Theorem 2.2 is given in Appendix B.

Corollary 2.1 *The choice of a_0 given in (13) satisfies $0 < a_0 < 1$.*

Corollary 2.2 *The result in Theorem 2.2 can be alternatively obtained by taking a uniform improper prior for μ at the outset, i.e., $\pi(\mu) \propto 1$.*

The proofs of Corollaries 2.1 and 2.2 are straightforward. Theorem 2.2 gives us a useful characterization of the explicit relationship between the power prior and the hierarchical model, and we see from this theorem that the two models are equivalent if a_0 is chosen as (13). We see from (13) that a_0 is a monotonic function of τ, and if $\tau^2 \to 0$, then $a_0 \to 1$. This implies that if $\theta = \theta_0$ with probability 1, the historical and current data should be weighted equally. Also, the larger the sample size for the historical data, the less the weight given to the historical data. This is a desirable property since, in general, we would never want the historical data to dominate the posterior distribution of θ by simply increasing n_0.

2.2 Regression Model

We now extend the normal hierarchical model in (2) to the normal hierarchical regression model by setting $\boldsymbol{\theta} = \boldsymbol{x}_i'\boldsymbol{\beta}$ and $\theta_0 = \boldsymbol{x}_{0i}'\boldsymbol{\beta}_0$. This leads to the model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \ \ i = 1, 2, \ldots, n, \quad \text{and} \quad y_{0i} = \boldsymbol{x}_{0i}'\boldsymbol{\beta}_0 + \epsilon_{0i}, \ \ i = 1, 2, \ldots, n_0, \tag{14}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients for the current data, $\boldsymbol{\beta}_0$ is the $p \times 1$ vector of regression coefficients for the historical data, $\boldsymbol{x}_i$ is a $p \times 1$ vector of covariates for the i^{th} subject in the current dataset, and $\boldsymbol{x}_{0i}$ is a $p \times 1$ vector of covariates for the i^{th} subject in the historical dataset. Similar to (2), we further assume that the ϵ_i are i.i.d. $N(0, \sigma^2)$ and the ϵ_{0i} are i.i.d. $N(0, \sigma_0^2)$ and independent of the ϵ_i's, where σ^2 and σ_0^2 are fixed. In addition, we take

$$\boldsymbol{\beta} \mid \boldsymbol{\mu}, \Omega \sim N_p(\boldsymbol{\mu}, \Omega), \qquad \boldsymbol{\beta}_0 \mid \boldsymbol{\mu}, \Omega \sim N_p(\boldsymbol{\mu}, \Omega), \ \text{ and } \ \pi(\boldsymbol{\mu}) \propto 1, \tag{15}$$

where Ω is fixed. Here, $\boldsymbol{\beta}$ is the parameter vector of interest for the current study, and $\boldsymbol{\beta}_0$ denotes the corresponding parameter based on the historical study. The following theorem gives the form of the marginal posterior distribution $[\boldsymbol{\beta} | \boldsymbol{y}, \boldsymbol{y}_0]$ obtained from (14) and (15).

Theorem 2.3 *The marginal posterior distribution of* $\boldsymbol{\beta}$ *is given by*

$$(\boldsymbol{\beta} \mid \boldsymbol{y}, X, \boldsymbol{y}_0, X_0) \sim N_p(A_h^{-1}B_h, A_h^{-1}), \tag{16}$$

where

$$A_h = \sigma^{-2}X'X + \Omega^{-1} - \Omega^{-1}\left(2\Omega^{-1} - \Omega^{-1}\left(\Omega^{-1} + \sigma_0^{-2}X_0'X_0\right)^{-1}\Omega^{-1}\right)^{-1}\Omega^{-1}, \tag{17}$$

$$\begin{aligned} B_h = \sigma^{-2}X'\boldsymbol{y} + & \left[\Omega^{-1}\left(2\Omega^{-1} - \Omega^{-1}\left(\Omega^{-1} + \sigma_0^{-2}X_0'X_0\right)^{-1}\Omega^{-1}\right)^{-1}\right. \\ & \left.\times\Omega^{-1}\left(\Omega^{-1} + \sigma_0^{-2}X_0'X_0\right)^{-1}\sigma_0^{-2}X_0'\boldsymbol{y}_0\right], \end{aligned} \tag{18}$$

$X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)'$, *and* $X_0 = (\boldsymbol{x}_{01}, \boldsymbol{x}_{02}, \ldots, \boldsymbol{x}_{0n_0})'$.

The proof of Theorem 2.3 is given in Appendix B. Now we consider a power prior formulation of the model in (14). To do this, we set $\boldsymbol{\beta}_0 = \boldsymbol{\beta}$, and the resulting model becomes

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \text{ and } y_{0i} = \boldsymbol{x}_{0i}'\boldsymbol{\beta} + \epsilon_{0i}. \tag{19}$$

Thus in the power prior formulation, the ϵ_i are i.i.d. $N(0, \sigma^2)$, and the ϵ_{0i} are i.i.d. $N(0, \sigma_0^2)$ and independent of the ϵ_i's, where σ^2 and σ_0^2 are fixed. Under this model, the power prior based on the historical data $(n_0, \boldsymbol{y}_0, X_0)$ using the initial prior $\pi_0(\boldsymbol{\beta}) \propto 1$ is given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}_0, X_0, a_0) \propto \exp\left\{-\frac{a_0}{2\sigma_0^2}(y_0 - X_0\boldsymbol{\beta})'(\boldsymbol{y}_0 - X_0\boldsymbol{\beta})\right\}, \tag{20}$$

and the posterior distribution of $\boldsymbol{\beta}$ is given by

$$(\boldsymbol{\beta}|\boldsymbol{y}, X, \boldsymbol{y}_0, X_0, a_0) \sim N_p(A_p^{-1}B_p, A_p^{-1}), \tag{21}$$

where

$$A_p = \sigma^{-2}X'X + a_0\sigma_0^{-2}X_0'X_0 \tag{22}$$

and

$$B_p = \sigma^{-2}X'\boldsymbol{y} + a_0\sigma_0^{-2}X_0'\boldsymbol{y}_0. \tag{23}$$

To match (16) and (21), we need to find an explicit relationship between A_h and A_p as well as a relationship between B_h and B_p. Clearly for the hierarchical model and the power prior to have identical distributions, we need $A_h = A_p$ and $B_h = B_p$. We are led to the following theorem which characterizes this relationship.

Theorem 2.4 *Assume that* X_0 *is of full rank. Then, the posteriors in (16) and (21) match, i.e.,* $A_h = A_p$ *and* $B_h = B_p$ *if and only if*

$$a_0(I + 2\sigma_0^{-2}\Omega X_0'X_0) = I. \tag{24}$$

The proof of Theorem 2.4 is given in Appendix B. We note that when $p = 1$, (24) reduces to (13). Specifically, setting setting $\Omega = \tau^2$ and $X_0 = (1, 1, \ldots, 1)'$, (24) reduces to

$$a_0(1 + 2\sigma_0^{-2}\tau^2 n_0) = 1,$$

and thus, the condition for a_0 given by (13) is a special case of (24). We also see that Theorem 2.4 implies that for the two models to match, $\Omega(X_0'X_0)$ must be proportional to the identity matrix; that is, Ω must be of the form $\Omega = c(X_0'X_0)^{-1}$, where $c \geq 0$ is a scalar. This form of Ω is quite attractive especially in model selection contexts, and has been discussed by several authors as prior covariance matrix, including Zellner (1986), Ibrahim and Laud (1994), and Laud and Ibrahim (1995). Precisely, the form $\Omega = c(X_0'X_0)^{-1}$ corresponds to Zellner's g-prior, and has several nice properties as discussed in Zellner (1986), Ibrahim and Laud (1994), and Laud and Ibrahim (1995). Taking $\Omega = c(X_0'X_0)^{-1}$ is actually quite attractive in practice since it has the interpretation that, if there is historical data $D_0 = (n_0, y_0, X_0)$ for which a linear model is fit, $Y_0 = X_0\boldsymbol{\beta} + \epsilon$, $\epsilon \sim N(0, \sigma_0^2 I)$, then the information contained in $\boldsymbol{\beta}$ in the data D_0 is the Fisher information, given by $\sigma_0^{-2}(X_0'X_0)$. Thus, based on Fisher information arguments, it makes scientific sense to assign $c(X_0'X_0)^{-1}$ as the prior covariance matrix for $\boldsymbol{\beta}$. Our prior *does* indeed imply that the variation in $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ are linked via X_0. This is reasonable since $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ do have the same prior covariance matrix Ω, and $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ represent the same unobservable phenomenon. Therefore, allowing a multiple of the Fisher information in $\boldsymbol{\beta}_0$ to be the prior precision matrix makes scientific and intuitive sense. We are now led to the following theorem.

Theorem 2.5 *If the g-prior form is taken for Ω, that is, $\Omega = c(X_0'X_0)^{-1}$, then the power prior and the hierarchical model are identical when a_0 is taken to satisfy*

$$a_0(I + 2\sigma_0^{-2}cI) = I \quad \text{or} \quad a_0 = \frac{1}{1 + \frac{2c}{\sigma_0^2}}.$$

This theorem yields a very appealing result in that it shows the power prior and the hierarchical model are equivalent in the regression setting when a g-prior is chosen for Ω and a_0 is chosen as in Theorem 2.4. This result thus gives more appeal to the g-prior for use in Bayesian inference.

3 Hierarchical Generalized Linear Models

The analytic relationships given in the previous section can be extended to any non-normal model that has an asymptotically normal likelihood. To be specific, we demonstrate the nature of such approximations for the class of generalized linear models. Consider the hierarchical generalized linear model given by

$$p(y_i|\theta_i, \tau) = \exp\left\{a_i^{-1}(\tau)(y_i\theta_i - b(\theta_i)) + c(y_i, \tau)\right\},\ i = 1, \ldots, n,$$

and

$$p(y_{0i}|\theta_{0i}, \tau) = \exp\left\{a_i^{-1}(\tau)(y_{0i}\theta_{0i} - b(\theta_{0i})) + c(y_{0i}, \tau)\right\},\ i = 1, \ldots, n_0,$$

indexed by the canonical parameter θ_i (θ_{0i}) and the scale parameter τ. The functions b and c determine a particular family in the class, such as the binomial, normal, Poisson, etc. The functions $a_i(\tau)$ are commonly of the form $a_i(\tau) = \tau^{-1}w_i^{-1}$, where the w_i's are known weights. For ease of exposition, we assume that $w_i = 1$ and τ is known throughout. We further assume the θ_i's and θ_{0i}'s satisfy the equations

$$\theta_i = \theta(\eta_i) \ , \ i = 1, 2, \ldots, n, \ \ \theta_{0i} = \theta(\eta_{0i}) \ , \ i = 1, 2, \ldots, n_0,$$

and

$$\boldsymbol{\eta} = X\boldsymbol{\beta} \ \text{ and } \ \boldsymbol{\eta}_0 = X_0\boldsymbol{\beta}_0,$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$, and $\boldsymbol{\eta}_0 = (\eta_{01}, \ldots, \eta_{0n_0})$. The function $\theta(\eta_i)$ is called the θ-link, and for a canonical link, $\theta_i = \eta_i$. Given $(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$, $(\boldsymbol{y}, \boldsymbol{y}_0)$ are independent, and thus we can write the joint likelihood of $(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$ in vector notation for the generalized linear model (GLM) as

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{y}_0 | \boldsymbol{\beta}, \boldsymbol{\beta}_0) &= \exp\{\tau\Big(\boldsymbol{y}'\theta(X\beta) - b[\theta(X\boldsymbol{\beta})]\Big) + c(\boldsymbol{y}, \tau)\} \\ &\quad \times \exp\{\tau\Big(\boldsymbol{y}_0'\theta(X_0\beta_0) - b[\theta(X_0\boldsymbol{\beta}_0)]\Big) + c(\boldsymbol{y}_0, \tau)\}, \end{aligned} \tag{25}$$

where $\boldsymbol{\theta}(X\boldsymbol{\beta})$ is a componentwise function of $X\boldsymbol{\beta}$ that depends on the link. When a canonical link is used, $\boldsymbol{\theta}(X\boldsymbol{\beta}) = X\boldsymbol{\beta}$. To complete the hierarchical GLM specification, we specify priors for $(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$ as in the normal linear regression model. That is, we independently take

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \Omega) \ \text{ and } \ \boldsymbol{\beta}_0 \sim N(\boldsymbol{\mu}, \Omega), \tag{26}$$

where Ω is fixed. We further take $\pi(\boldsymbol{\mu}) \propto 1$. The marginal posterior distribution of $\boldsymbol{\beta}$ is required to make inferences about $\boldsymbol{\beta}$ in this framework. Due to the complexity of the hierarchical GLM specification, it is not possible to obtain a closed form expression for the marginal posterior distribution of $\boldsymbol{\beta}$. To overcome this difficulty, we consider an asymptotic approximation to the posterior similar to that of Chen (1985). In the context considered here, "asymptotic" means $n_0 \to \infty$ and $n \to \infty$. Toward this goal, let

$$p(\boldsymbol{\beta} | \boldsymbol{y}, X) = \exp\{\tau\Big(\boldsymbol{y}'\boldsymbol{\theta}(X\boldsymbol{\beta}) - \tau b[\boldsymbol{\theta}(X\boldsymbol{\beta})]\Big) + c(\boldsymbol{y}, \tau)\}$$

and

$$p(\boldsymbol{\beta}_0 | \boldsymbol{y}_0, X_0) = \exp\{\tau\Big(\boldsymbol{y}_0'\boldsymbol{\theta}(X_0\boldsymbol{\beta}_0) - b[\boldsymbol{\theta}(X_0\boldsymbol{\beta}_0)]\Big) + c(\boldsymbol{y}_0, \tau)\}.$$

Following Chen (1985) and ignoring constants that are free of the parameters, we have

$$p(\boldsymbol{\beta} | \boldsymbol{y}, X) \approx \exp\Big\{ -\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\widehat{\Sigma}^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\Big\} \tag{27}$$

and

$$p(\boldsymbol{\beta}_0 | \boldsymbol{y}_0, X_0) \approx \exp\Big\{ -\frac{1}{2}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_0)'\widehat{\Sigma}_0^{-1}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_0)\Big\}, \tag{28}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_0$ are the respective maximum likelihood estimators (MLEs) of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ based on the data $(n, \boldsymbol{y}, X)$ and $(n_0, \boldsymbol{y}_0, X_0)$ under the GLM,

$$\widehat{\Sigma} = \left\{ -\frac{\partial^2 \ln p(\boldsymbol{\beta}|\boldsymbol{y}, X)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right\}^{-1}, \text{ and } \widehat{\Sigma}_0 = \left\{ -\frac{\partial^2 \ln p(\boldsymbol{\beta}_0|\boldsymbol{y}_0, X_0)}{\partial \boldsymbol{\beta}_0 \partial \boldsymbol{\beta}_0'} \bigg|_{\boldsymbol{\beta}_0=\hat{\boldsymbol{\beta}}_0} \right\}^{-1}.$$

It is straightforward to show that under the model in (25),

$$\widehat{\Sigma}^{-1} = \tau X' \hat{\Delta}^2 \hat{V} X \text{ and } \widehat{\Sigma}_0^{-1} = \tau X_0' \hat{\Delta}_0^2 \hat{V}_0 X_0, \tag{29}$$

where $\hat{\Delta}$ and $\hat{V}$ are $n \times n$ diagonal matrices with i^{th} diagonal elements $\delta_i = \delta_i(\boldsymbol{x}_i'\boldsymbol{\beta}) = d\theta_i/d\eta_i$ and $v_i = v_i(\boldsymbol{x}_i'\boldsymbol{\beta}) = d^2 b(\theta_i)/d\theta_i^2$ evaluated at $\hat{\boldsymbol{\beta}}$, and $\hat{\Delta}_0$ and $\hat{V}_0$ are $n_0 \times n_0$ diagonal matrices with i^{th} diagonal elements $\delta_{0i} = \delta_{0i}(\boldsymbol{x}_{0i}'\boldsymbol{\beta}_0) = d\theta_i/d\eta_i$ and $v_{0i} = v_{0i}(\boldsymbol{x}_{0i}'\boldsymbol{\beta}_0) = d^2 b(\theta_{0i})/d\theta_{0i}^2$ evaluated at $\hat{\boldsymbol{\beta}}_0$. The approximations in (27) and (28) are valid for large n and large n_0, respectively.

Using (27) and (28) and the proof of Theorem 2.3, the marginal posterior distribution of $\boldsymbol{\beta}$ is then approximated by

$$(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{y}_0, X, X_0) \overset{approx.}{\sim} N_p(\hat{A}_h^{-1}\hat{B}_h, \hat{A}_h^{-1}), \tag{30}$$

where

$$\hat{A}_h = \hat{\Sigma}^{-1} + \Omega^{-1} - \Omega^{-1}\left(2\Omega^{-1} - \Omega^{-1}\left(\Omega^{-1} + \hat{\Sigma}_0^{-1}\right)^{-1}\Omega^{-1}\right)^{-1}\Omega^{-1} \tag{31}$$

and

$$\hat{B}_h = \hat{\Sigma}^{-1}\hat{\boldsymbol{\beta}} + \Omega^{-1}\left(2\Omega^{-1} - \Omega^{-1}\left(\Omega^{-1} + \hat{\Sigma}_0^{-1}\right)^{-1}\Omega^{-1}\right)^{-1}\Omega^{-1}\left(\Omega^{-1} + \hat{\Sigma}_0^{-1}\right)^{-1}\hat{\Sigma}_0^{-1}\hat{\boldsymbol{\beta}}_0. \tag{32}$$

As in the normal linear regression model, we consider a power prior formulation of the hierarchical GLM given by (25) by setting $\boldsymbol{\beta}_0 = \boldsymbol{\beta}$. The power prior based on the historical data $(n_0, \boldsymbol{y}_0, X_0)$ using the initial prior $\pi_0(\boldsymbol{\beta}) \propto 1$ is then given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}_0, X_0, a_0) \propto \exp\{a_0\tau\Big(\boldsymbol{y}_0'\boldsymbol{\theta}(X_0\boldsymbol{\beta}_0) - b[\boldsymbol{\theta}(X_0\boldsymbol{\beta}_0)]\Big) + c(\boldsymbol{y}_0, \tau)\}, \tag{33}$$

and the posterior distribution of $\boldsymbol{\beta}$ is given by

$$\begin{aligned} \pi(\boldsymbol{\beta}|\boldsymbol{y}, X, \boldsymbol{y}_0, X_0, a_0) \quad \propto \quad & \exp\{\tau\Big(\boldsymbol{y}'\boldsymbol{\theta}(X\boldsymbol{\beta}) - \tau b[\boldsymbol{\theta}(X\boldsymbol{\beta})]\Big) + c(\boldsymbol{y}, \tau)\} \\ & \times \exp\{a_0\tau\Big(\boldsymbol{y}_0'\boldsymbol{\theta}(X_0\boldsymbol{\beta}) - b[\boldsymbol{\theta}(X_0\boldsymbol{\beta})]\Big) + c(\boldsymbol{y}_0, \tau)\}. \end{aligned} \tag{34}$$

Using (27) and (28), and after some algebra, the posterior distribution of $\boldsymbol{\beta}$ given by (34) is approximated by

$$\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{y}_0, X, X_0, a_0 \overset{approx.}{\sim} N_p(\hat{A}_p^{-1}\hat{B}_p, \hat{A}_p^{-1}), \tag{35}$$

where

$$\hat{A}_p = \hat{\Sigma}^{-1} + a_0\hat{\Sigma}_0^{-1}, \quad \hat{B}_p = \hat{\Sigma}^{-1}\hat{\boldsymbol{\beta}} + a_0\hat{\Sigma}_0^{-1}\hat{\boldsymbol{\beta}}_0, \tag{36}$$

$\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_0$ are the MLEs, and $\hat{\Sigma}$ and $\hat{\Sigma}_0$ are defined by (29).

Similar to the normal linear regression model, we now match the *approximate* posterior distributions of $\boldsymbol{\beta}$ under the GLM. We are now led to the following theorem.

Theorem 3.1 *The approximate posteriors in (30) and (35) match, i.e., $\hat{A}_h = \hat{A}_p$ and $\hat{B}_h = \hat{B}_p$ if and only if*

$$a_0(I + 2\Omega\hat{\Sigma}_0^{-1}) = I. \tag{37}$$

The proof of Theorem 3.1 is similar to that of Theorem 2.4, and therefore the details are omitted for brevity. If a generalized g-prior form is taken for Ω, i.e., $\Omega = c\hat{\Sigma}_0$, then the approximate power prior and the hierarchical model are identical when $a_0 = 1/(1+2c)$. Thus, again, we see that for the class of GLM's, the connection between the hierarchical model and the power prior is facilitated by taking a generalized g-prior for Ω.

4 Extensions to Multiple Historical Datasets

In this section, we generalize the results given in Theorems 2.2, 2.4, and 3.1 to multiple historical datasets. First, we generalize the hierarchical model in (14) and the power prior in (20) to multiple historical datasets as follows. Suppose we have L_0 historical datasets. For the current study, the hierarchical model can be written as

$$y_i = \theta + \epsilon_i, \tag{38}$$

where $i = 1, 2, \ldots, n$, and the ϵ_i's are i.i.d. $N(0, \sigma^2)$ random variables, where σ^2 is fixed. For the historical data, the hierarchical model can be written as

$$y_{0ki} = \theta_{0k} + \epsilon_{0ki}, \tag{39}$$

where $i = 1, 2, \ldots, n_{0k}$, the ϵ_{0ki}'s are i.i.d. $N(0, \sigma_{0k}^2)$ and independent of the ϵ_i's, and the σ_{0k}^2's are fixed for $k = 1, 2, \ldots, L_0$. Further, we assume that

$$\theta \mid \mu, \tau^2 \sim N(\mu, \tau^2) \tag{40}$$

and the θ_{0k} are independently distributed as

$$\theta_{0k} \mid \mu, \tau^2 \sim N(\mu, \tau^2), \tag{41}$$

where $k = 1, 2, \ldots, L_0$, and τ^2 is a fixed hyperparameter. Based on the results established in Theorem 2.2, we take a uniform improper prior for μ, i.e., $\pi(\mu) \propto 1$, and let $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0L_0})'$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$, and $\boldsymbol{y}_0 = (\boldsymbol{y}_{01}', \boldsymbol{y}_{02}', \ldots, \boldsymbol{y}_{0L_0}')'$, where $\boldsymbol{y}_{0k} = (y_{0k1}, y_{0k2}, \ldots, y_{0kn_{0k}})'$ for $k = 1, 2, \ldots, L_0$.

The following theorem gives the form of the marginal posterior distribution $[\theta|\boldsymbol{y}, \boldsymbol{y}_0]$ under the model given by (38), (39), (40), and (41).

Theorem 4.1 *Using (38), (39), (40), and (41), the marginal posterior distribution of θ is given by*

$$\theta \mid \boldsymbol{y}, \boldsymbol{y}_0 \sim N(\mu_{mh}, \sigma^2_{mh}), \tag{42}$$

where

$$\mu_{mh} = \sigma^2_{mh} \left\{ \frac{n\bar{y}}{\sigma^2} + \left[\sum_{k=1}^{L_0} \frac{\frac{n_{0k}\bar{y}_{0k}}{\sigma^2_{0k}}}{\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2}} \right] \left[\tau^4 \left(\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0} \frac{1}{\tau^4 (\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2})} \right) \right]^{-1} \right\}, \tag{43}$$

$$\sigma^2_{mh} = \left[\frac{n}{\sigma^2} + \frac{1}{\tau^2} - \frac{1}{\tau^4 \left(\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0} \frac{1}{\tau^4 (\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2})} \right)} \right]^{-1}, \tag{44}$$

$\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$, *and* $\bar{y}_{0k} = \frac{1}{n_{0k}}\sum_{i=1}^{n_{0k}} y_{0ki}$ *for* $k = 1, 2, \ldots, L_0$.

The proof of Theorem 4.1 is given in Appendix B. The power prior formulation of the model with multiple historical datasets can be obtained by letting $\theta_{0k} = \theta$ in (40) and (41). The model for the current data can be written as

$$y_i = \theta + \epsilon_i, \tag{45}$$

and the model for the historical data can be written as

$$y_{0ki} = \theta + \epsilon_{0ki}, \tag{46}$$

where the ϵ_{0ki} are i.i.d. $N(0, \sigma^2_{0k})$ for $k = 1, 2, \ldots, L_0$. Now the power prior for θ under multiple historical datasets is given by

$$\pi(\theta | \boldsymbol{y}_{0k}, k = 1, 2, \ldots, L_0, \boldsymbol{a}_0) \propto \exp\left\{ -\frac{1}{2} \sum_{k=1}^{L_0} \frac{a_{0k}}{\sigma^2_{0k}} \sum_{i=1}^{n_{0k}} (y_{0ki} - \theta)^2 \right\}, \tag{47}$$

where $\boldsymbol{a}_0 = (a_{01}, a_{02}, \ldots, a_{0L_0})'$ with $0 < a_{0k} < 1$ for $k = 1, 2, \ldots, L_0$. Using (45) and (46) along with a uniform improper initial prior for θ, (i.e., $\pi_0(\theta) \propto 1$), we are led to

$$(\theta \mid \boldsymbol{y}, \boldsymbol{y}_0, \boldsymbol{a}_0) \sim N(\mu_{mp}, \sigma^2_{mp}), \tag{48}$$

where

$$\mu_{mp} = \frac{\frac{n\bar{y}}{\sigma^2} + \sum_{k=1}^{L_0} a_{0k} \frac{n_{0k}\bar{y}_{0k}}{\sigma^2_{0k}}}{\frac{n}{\sigma^2} + \sum_{k=1}^{L_0} a_{0k} \frac{n_{0k}}{\sigma^2_{0k}}} \tag{49}$$

and

$$\sigma^2_{mp} = \frac{1}{\frac{n}{\sigma^2} + \sum_{k=1}^{L_0} a_{0k} \frac{n_{0k}}{\sigma^2_{0k}}}. \tag{50}$$

We are now led to the following theorem characterizing the relationship between the power prior and the hierarchical model with multiple historical datasets.

Theorem 4.2 *Suppose we take $\pi_0(\theta) \propto 1$ for the power prior and $\pi(\mu) \propto 1$ for the hierarchical model. Then $\mu_{mh} = \mu_{mp}$ in (43) and (49), and $\sigma^2_{mh} = \sigma^2_{mp}$ in (44) and (50) if and only if*

$$a_{0k} = \tau^{-4}\left(\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2}\right)^{-1}\left(\frac{L_0+1}{\tau^2} - \sum_{i=1}^{L_0}\tau^{-4}\left(\frac{n_{0i}}{\sigma^2_{0i}} + \frac{1}{\tau^2}\right)^{-1}\right)^{-1}. \tag{51}$$

The proof of Theorem 4.2 is given in Appendix B.

Corollary 4.1 *The choice of a_{0k} given in (51) satisfies $0 < a_{0k} < 1$. Furthermore, if $\tau^2 \geq \frac{\sigma^2_{0k}}{n_{0k}}$ for $k = 1, 2, \ldots, L_0$, then the choice of a_{0k} given in (51) also satisfies $\sum_{k=1}^{L_0} a_{0k} < 1$.*

The proof of Corollary 4.1 is straightforward. We also note that when $L_0 = 1$, then with some obvious adjustments in notation, μ_{mh} in (43), σ^2_{mh} in (44), and a_{01} in (51) all reduce to μ_h in (6), σ^2_h in (7), and a_0 in (13), respectively.

Next, we extend the normal hierarchical regression model (14) to multiple historical datasets as follows. Suppose we have L_0 historical datasets. In the vector notation, we assume

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{and} \quad \boldsymbol{y}_{0k} = X_{0k}\boldsymbol{\beta}_{0k} + \boldsymbol{\epsilon}_{0k}, \quad k = 1, 2, \ldots, L_0, \tag{52}$$

where X is an $n \times p$ matrix, X_{0k} is an $n_{0k} \times p$ matrix for $k = 1, 2, \ldots, L_0$, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{0k}$, $k = 1, 2, \ldots, L_0$, are the $p \times 1$ vectors of regression coefficients, $\boldsymbol{\epsilon} \sim N_n(0, \sigma^2 I)$ with σ^2 fixed, $\boldsymbol{\epsilon}_{0k} \sim N_{n_{0k}}(0, \sigma^2_{0k} I)$ with σ^2_{0k} fixed for $k = 1, 2, \ldots, L_0$, and $\boldsymbol{\epsilon}$ is independent of $\boldsymbol{\epsilon}_{0k}$. The following theorem gives the marginal posterior distribution of $\boldsymbol{\beta}$ under this setting.

Theorem 4.3 *Assume that $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{0k}$, $k = 1, 2, \ldots, L_0$, are i.i.d. $N_p(\boldsymbol{\mu}, \Omega)$ random vectors, where Ω is fixed, and $\pi(\boldsymbol{\mu}) \propto 1$. Using (52), the marginal posterior distribution of $\boldsymbol{\beta}$ is given by*

$$\boldsymbol{\beta} \sim N_p(A^{-1}_{mh}B_{mh}, A^{-1}_{mh}), \tag{53}$$

where

$$A_{mh} = \frac{1}{\sigma^2}X'X + \Omega^{-1} - \Omega^{-1}\left[(L_0+1)\Omega^{-1} - \sum_{k=1}^{L_0}\Omega^{-1}\Big(\Omega^{-1} + \frac{1}{\sigma^2_{0k}}X'_{0k}X_{0k}\Big)^{-1}\Omega^{-1}\right]^{-1}\Omega^{-1} \tag{54}$$

and

$$B_{mh} = \frac{1}{\sigma^2}X'\boldsymbol{y} + \Bigg\{\Omega^{-1}\left[(L_0+1)\Omega^{-1} - \sum_{k=1}^{L_0}\Omega^{-1}\Big(\Omega^{-1} + \frac{1}{\sigma^2_{0k}}X'_{0k}X_{0k}\Big)^{-1}\Omega^{-1}\right]^{-1} \times\Omega^{-1}\Big(\sum_{k=1}^{L_0}\Big(\Omega^{-1} + \frac{1}{\sigma^2_{0k}}X'_{0k}X_{0k}\Big)^{-1}\frac{1}{\sigma^2_{0k}}X'_{0k}\boldsymbol{y}_{0k}\Bigg\}. \tag{55}$$

The proof of Theorem 4.3 is similar to that of Theorem 2.3 for the normal hierarchical model with a single historical dataset. Thus, the proof is omitted for brevity. By letting $\boldsymbol{\beta} = \boldsymbol{\beta}_{01} = \cdots = \boldsymbol{\beta}_{0k}$ in (52), the power prior for the normal linear model with multiple historical datasets can be obtained as

$$\pi(\boldsymbol{\beta}|(\boldsymbol{y}_{0k}, X_{0k}), k = 1, 2, \ldots, L_0, \boldsymbol{a}_0) \propto \exp\left\{-\frac{1}{2}\sum_{k=1}^{L_0}\frac{a_{0k}}{\sigma_{0k}^2}(\boldsymbol{y}_{0k} - X_{0k}\boldsymbol{\beta})'(\boldsymbol{y}_{0k} - X_{0k}\boldsymbol{\beta})\right\}, \tag{56}$$

where $0 < a_{0k} < 1$ for $k = 1, 2, \ldots, L_0$. Assuming $\pi_0(\boldsymbol{\beta}) \propto 1$, the posterior distribution of $\boldsymbol{\beta}$ is given by

$$(\boldsymbol{\beta} \mid \boldsymbol{y}, X, (\boldsymbol{y}_{0k}, X_{0k}), k = 1, 2, \ldots, L_0) \sim N_p(A_{mp}^{-1}B_{mp}, A_{mp}^{-1}), \tag{57}$$

where

$$A_{mp} = \frac{1}{\sigma^2}X'X + \sum_{k=1}^{L_0}\frac{a_{0k}}{\sigma_{0k}^2}X_{0k}'X_{0k} \text{ and } B_{mp} = \frac{1}{\sigma^2}X'\boldsymbol{y} + \sum_{k=1}^{L_0}\frac{a_{0k}}{\sigma_{0k}^2}X_{0k}'\boldsymbol{y}_{0k}.$$

The following theorem characterizes the relationship between the power prior and the normal hierarchical regression model with multiple historical datasets.

Theorem 4.4 *The posterior distribution of $\boldsymbol{\beta}$ in (57) under the power prior (56) matches (53) under the normal hierarchical regression model if and only if*

$$a_{0k}\left[(L_0 + 1)I - \sum_{k=1}^{L_0}\left(I + \frac{1}{\sigma_{0k}^2}\Omega X_{0k}'X_{0k}\right)^{-1}\right] = \left(I + \frac{1}{\sigma_{0k}^2}\Omega X_{0k}'X_{0k}\right)^{-1} \tag{58}$$

for $k = 1, 2, \ldots, L_0$.

The proof of Theorem 4.4 directly follows from that of Theorem 2.4. We note that when $L_0 = 1$, (58) reduces to (24), and if $p = 1$, $\Omega = \tau^2$, and $X = X_{01} = \cdots = X_{0L_0} = (1, 1, \ldots, 1)'$, (58) reduces to (51). Finally, the result given in Theorem 3.1 for GLM's can be extended to multiple historical datasets, and details of this extension can be found in Appendix A.

5 Elicitation of a_0

In this section, we discuss two methods for specifying Ω in a Bayesian analysis: (i) using the analytic relationship between the power prior and hierarchical models, and taking a fixed value for Ω — namely $\Omega = c(X_0'X_0)^{-1}$ — where c is a fixed hyperparameter, as well as a fixed value for σ_0^2, and (ii) taking Ω (and σ_0^2) to be random, in which a prior is specified for Ω (and σ_0^2).

We first consider situation (i). The results of the previous section, yielding equivalence between the power prior and hierarchical models, shed light on how to elicit a guide value for a_0 using the historical data. The basic idea is to use the relationship between the power prior and the hierarchical model to specify a guide value for a_0 from the hyperparameters of the hierarchical model. Specifically, we can use the relationship given by (24) to elicit a_0. One natural way to specify a_0 is take the trace of both sides of (24) and take a_0 to be the solution to that equation. Upon taking traces of both sides of equation (24), we are led to

$$a_0 \mathrm{tr}(I + 2\sigma_0^{-2}\Omega X_0' X_0) = \mathrm{tr}(I). \tag{59}$$

Since I is a $p \times p$ matrix, $tr(I) = p$, and thus (59) implies that a guide value of a_0 is taken to be

$$\hat{a}_0 = \frac{p}{p + 2\sigma_0^{-2} tr(\Omega X_0' X_0)}. \tag{60}$$

If $\Omega = c(X_0' X_0)^{-1}$, (60) reduces to

$$\hat{a}_0 = \frac{1}{1 + 2\sigma_0^{-2} c}. \tag{61}$$

We see that (61) corresponds to the relationship between the power prior and the hierarchical model in the case that θ is a univariate parameter, i.e., the i.i.d. case. We can clearly see that (59) satisfies $0 < a_0 < 1$. The guide value for a_0 given in (60) provides a very useful benchmark value for a_0 from which further sensitivity analyses can be conducted. This guide value has a solid theoretical justification and motivation: it is the value of a_0 that results in equivalence between the power prior and the hierarchical model. In the example of Section 6, we show that the guide value given by (60) works quite well in practice.

We now consider situation (ii). If Ω (and σ_0^2) are random, then (60) can be easily modified. In this case, our guide value for a_0 would be the posterior expectation of (60) with respect to the joint posterior distribution of (σ_0^2, Ω), thus leading to

$$\hat{a}_0 = E\left(\frac{p}{p + 2\sigma_0^{-2} tr(\Omega X_0' X_0)}\right), \tag{62}$$

where the expectation is taken with respect to the joint posterior distribution of (σ_0^2, Ω).

Similar formulas for the guide value for a_0 are available for GLM's using (37). For the class of GLM's under situation (i), the guide value for a_0 is given by

$$\hat{a}_0 = \frac{p}{p + 2tr(\Omega \hat{\Sigma}_0^{-1})}. \tag{63}$$

When Ω is random under situation (ii) for GLM's, the guide value is obtained by taking the expectation of (63) with respect to the posterior distribution of Ω, leading to

$$\hat{a}_0 = E\left(\frac{p}{p + 2tr(\Omega \hat{\Sigma}_0^{-1})}\right). \tag{64}$$

Finally, we briefly discuss how to obtain a guide value for a_{0k} in the presence of multiple historical datasets under situation (ii). For ease of exposition, we consider the normal linear model with multiple historical datasets. Using (58), we take the guide value for a_{0k} as

$$\hat{a}_{0k} = E\left\{ \frac{tr\left[\left(I + \frac{1}{\sigma_{0k}^2}\Omega X_{0k}' X_{0k}\right)^{-1}\right]}{(L_0+1)p - \sum_{k=1}^{L_0} tr\left[\left(I + \frac{1}{\sigma_{0k}^2}\Omega X_{0k}' X_{0k}\right)^{-1}\right]} \right\},$$

where the expectation is taken with respect to the joint posterior distribution of $(\Omega, \sigma_{01}^2, \ldots, \sigma_{0L_0}^2)$. After some matrix algebra, it can be shown that $\hat{a}_{0k} < 1$ for $k = 1, 2, \ldots, L_0$. When multiple historical datasets are available, we have more information to estimate Ω, and thus Ω is estimated with more precision in this case. In this sense, inference with multiple historical datasets can be quite advantageous in eliciting the power parameter in the power prior.

6 AIDS Data

We consider the two AIDS studies ACTG019 and ACTG036, where ACTG036 represents the current data and ACTG019 represents the historical data. The purpose of this example is to demonstrate the proposed methodology for obtaining the guide value for a_0 discussed in Section 5 in the context of logistic regression.

The ACTG019 study was a double blind placebo-controlled clinical trial comparing zidovudine (AZT) to placebo in persons with CD4 counts less than 500. The sample size for this study, excluding cases with missing data, was $n_0 = 823$. The response variable (y_0) for these data is binary with a 1 indicating death, development of AIDS, or AIDS related complex (ARC), and a 0 indicates otherwise. Several covariates were also measured. The ones we use here are CD4 count (x_{01}) (cell count per mm^3 of serum), age (x_{02}), and treatment, (x_{03}). The covariates CD4 count and age are continuous, whereas the treatment covariate is binary. The ACTG036 study was also a placebo-controlled clinical trial comparing AZT to placebo in patients with hereditary coagulation disorders. The sample size in this study, excluding cases with missing data, was $n = 183$. The response variable (y) for these data is binary with a 1 indicating death, development of AIDS, or AIDS related complex (ARC), and a 0 indicates otherwise. Several covariates were measured for these data. The ones we use here are CD4 count (x_1), age (x_2), and treatment (x_3).

We consider the hierarchical logistic regression model to fit the ACTG019 and ACTG036 data. In (26), we take Ω to be a 4×4 diagonal matrix and we further assume that $\Omega_{jj} \overset{i.i.d.}{\sim} \text{IG}(1, 0.005)$, $j = 0, 1, \ldots, 3$. Using (64), we specify a_0 by

$$\hat{a}_0 = E\Big[\frac{4}{4 + 2tr(\Omega\hat{\Sigma}_0^{-1})}\Big], \tag{65}$$

where $\hat{\Sigma}_0^{-1}$ is computed via (29), and the expectation is taken with respect to the posterior distribution of Ω under the hierarchical logistic regression model. The Gibbs sampling algorithm is used to sample $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$, $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01}, \beta_{02}, \beta_{03})'$, $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2, \mu_3)'$, and Ω from their respective posterior distributions under the hierarchical model. Table 1 gives the posterior estimates of $\boldsymbol{\beta}$, $\boldsymbol{\beta}_0$, $\boldsymbol{\mu}$, and Ω based on 20,000 Gibbs iterations after a "burn-in" of 1,000 iterations. Using (65), we obtain $\hat{a}_0 = 0.415$. Table 1 shows the posterior estimates of $\boldsymbol{\beta}$ using the power prior in (33) with several values of a_0, including $a_0 = \hat{a}_0 = 0.415$. From Tables 1 and 2, we can see that the posterior estimates of $\boldsymbol{\beta}$ using the power prior with $a_0 = 0.415$ are fairly close to those obtained under the hierarchical model. From Table 2, we also see that the posterior estimates are slightly different for the different values of a_0. In particular, when more weight is given to the historical data, age and treatment become more "significant", that is, their 95% Highest Posterior Density (HPD) intervals do not include 0. Thus, the power prior gives the investigator great flexibility for incorporating the historical data to analyze the ACTG036 trial.

Parameter	Posterior Mean	Posterior Std Dev	Parameter	Posterior Mean	Posterior Std Dev
β_0	-3.128	0.238	β_{00}	-3.044	0.177
β_1	-0.728	0.161	β_{01}	-0.671	0.129
β_2	0.261	0.138	β_{02}	0.323	0.118
β_3	-0.336	0.184	β_{03}	-0.387	0.144
μ_0	-3.083	0.224	Ω_{00}	0.031	0.227
μ_1	-0.702	0.170	Ω_{11}	0.021	0.091
μ_2	0.293	0.149	Ω_{22}	0.020	0.134
μ_3	-0.361	0.179	Ω_{33}	0.021	0.104

Table 1: Posterior Estimates Under the Hierarchical Logistic Regression Model

a_0	Parameter	Posterior Mean	Posterior Std Dev	95% HPD Interval
0	β_0	-4.781	0.849	(-6.461, -3.223)
	β_1	-1.636	0.449	(-2.539, -0.791)
	β_2	0.122	0.234	(-0.334, 0.587)
	β_3	-0.057	0.380	(-0.802, 0.698)
0.415	β_0	-3.196	0.253	(-3.691, -2.708)
	β_1	-0.779	0.175	(-1.121, -0.434)
	β_2	0.259	0.142	(-0.026, 0.531)
	β_3	-0.344	0.196	(-0.724, 0.043)
1	β_0	-3.041	0.169	(-3.379, -2.722)
	β_1	-0.677	0.123	(-0.917, -0.437)
	β_2	0.302	0.110	(0.083, 0.513)
	β_3	-0.377	0.139	(-0.654, -0.109)

Table 2: Posterior Estimates Using the Power Prior for the Logistic Regression Model

Appendix A: Extension to Multiple Historical Datasets for GLMs

In this appendix, we generalize the result given in Theorem 3.1 to multiple historical datasets.

First, we consider the hierarchical GLM with multiple historical datasets. In vector notation, the likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$ given $(\boldsymbol{y}, \boldsymbol{y}_{01}, \ldots, \boldsymbol{y}_{0L_0})$ is given by

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{y}_{01}, \ldots, \boldsymbol{y}_{0L_0} | \boldsymbol{\beta}, \boldsymbol{\beta}_0) &= \exp\{\tau\Big(\boldsymbol{y}'\boldsymbol{\theta}(X\boldsymbol{\beta}) - b[\boldsymbol{\theta}(X\boldsymbol{\beta})]\Big) + c(\boldsymbol{y}, \tau)\} \\ &\quad \times \prod_{k=1}^{L_0} \exp\{\tau\Big(\boldsymbol{y}_{0k}'\boldsymbol{\theta}(X_{0k}\boldsymbol{\beta}_{0k}) - b[\boldsymbol{\theta}(X_{0k}\boldsymbol{\beta}_{0k})]\Big) + c(\boldsymbol{y}_{0k}, \tau)\}. \end{aligned}$$

We independently take $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}, \Omega)$, $\boldsymbol{\beta}_{0k} \sim N_p(\boldsymbol{\mu}, \Omega)$ for $k = 1, 2, \ldots, L_0$, and $\pi(\boldsymbol{\mu}) \propto 1$. Using (3.3), (3.4), and Theorems 2.3 and 4.3, the marginal posterior distribution of $\boldsymbol{\beta}$ is approximated by (i.e., for n and n_0 large)

$$(\boldsymbol{\beta} | \boldsymbol{y}, X, (\boldsymbol{y}_{0k}, X_{0k}), k = 1, 2, \ldots, L_0) \overset{approx.}{\sim} N_p(\hat{A}_{mh}^{-1}\hat{B}_{mh}, \hat{A}_{mh}^{-1}), \tag{A.1}$$

where

$$\hat{A}_{mh} = \hat{\Sigma}^{-1} + \Omega^{-1} - \Omega^{-1}\left[(L_0+1)\Omega^{-1} - \sum_{k=1}^{L_0}\Omega^{-1}\Big(\Omega^{-1} + \hat{\Sigma}^{-1}\Big)^{-1}\Omega^{-1}\right]^{-1}\Omega^{-1},$$

$$\begin{aligned} \hat{B}_{mh} = \hat{\Sigma}^{-1}\hat{\boldsymbol{\beta}} + \Bigg\{ &\quad \Omega^{-1}\left[(L_0+1)\Omega^{-1} - \sum_{k=1}^{L_0}\Omega^{-1}\Big(\Omega^{-1} + \hat{\Sigma}_{0k}^{-1}\Big)^{-1}\Omega^{-1}\right]^{-1} \\ &\times \Omega^{-1}\Big(\sum_{k=1}^{L_0}\Big(\Omega^{-1} + \hat{\Sigma}_{0k}^{-1}\Big)^{-1}\hat{\Sigma}_{0k}^{-1}\hat{\boldsymbol{\beta}}_{0k}\Bigg\}, \end{aligned}$$

$\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$ based on the data $D = (n, \boldsymbol{y}, X)$, and $\hat{\boldsymbol{\beta}}_{0k}$ is the MLE of β_{0k} based on the data $D_{0k} = (n_{0k}, \boldsymbol{y}_{0k}, X_{0k})$, $k = 1, \ldots, L_0$ under the GLM, $\hat{\Sigma}^{-1}$ and $\hat{\Sigma}_{0k}^{-1}$ are defined by (3.5) based on the current data D and the historical data D_{0k}, for $k = 1, 2 \ldots, L_0$.

The power prior based on the L_0 historical datasets $\{(\boldsymbol{y}_{0k}, X_{0k}),\ \ k = 1, 2, \ldots, L_0\}$ using the initial prior $\pi_0(\boldsymbol{\beta}) \propto 1$ is given by

$$\begin{aligned} &\pi(\boldsymbol{\beta} | (\boldsymbol{y}_{0k}, \boldsymbol{X}_{0k}),\ k = 1, 2, \ldots, L_0, \boldsymbol{a}_0) \\ \propto &\prod_{k=1}^{L_0} \exp\{a_{0k}\tau\Big(\boldsymbol{y}_{0k}'\boldsymbol{\theta}(X_{0k}\boldsymbol{\beta}) - b[\boldsymbol{\theta}(X_{0k}\boldsymbol{\beta})]\Big) + c(\boldsymbol{y}_0, \tau)\}, \end{aligned} \tag{A.2}$$

and the posterior distribution of $\boldsymbol{\beta}$ is given by

$$\begin{aligned} & \pi(\beta|\boldsymbol{y}, X, (\boldsymbol{y}_{0k}, \boldsymbol{X}_{0k}),\ k = 1, 2, \ldots, L_0, \boldsymbol{a}_0) \\ \propto\ & \exp\{\tau\Big(\boldsymbol{y}'\boldsymbol{\theta}(X\boldsymbol{\beta}) - \tau b[\boldsymbol{\theta}(X\boldsymbol{\beta})]\Big) + c(\boldsymbol{y}, \tau)\} \\ & \times \prod_{k=1}^{L_0} \exp\{a_{0k}\tau\Big(\boldsymbol{y}'_{0k}\boldsymbol{\theta}(X_{0k}\boldsymbol{\beta}_{0k}) - b[\boldsymbol{\theta}(X_{0k}\boldsymbol{\beta})]\Big) + c(\boldsymbol{y}_0, \tau)\}. \end{aligned} \tag{A.3}$$

Using (3.3) and (3.4) and after some algebra, the posterior distribution of $\boldsymbol{\beta}$ given by (A.3) is approximated by (i.e., for large n and n_{0k}, $k = 1, \ldots, L_0$)

$$(\boldsymbol{\beta} \mid \boldsymbol{y}, X, (\boldsymbol{y}_{0k}, X_{0k}) = 1, 2, \ldots, L_0, \boldsymbol{a}_0) \overset{approx.}{\sim} N_p(\hat{A}_p^{-1}\hat{B}_p, \hat{A}_p^{-1}), \tag{A.4}$$

where

$$\hat{A}_p = \hat{\Sigma}^{-1} + \sum_{k=1}^{L_0} a_{0k}\hat{\Sigma}_{0k}^{-1} \ \text{ and } \ \hat{B}_p = \hat{\Sigma}^{-1}\hat{\boldsymbol{\beta}} + \sum_{k=1}^{L_0} a_{0k}\hat{\Sigma}_{0k}^{-1}\hat{\boldsymbol{\beta}}_{0k}.$$

Then we are now led to the following theorem.

Theorem A.1 *The posterior distribution of $\boldsymbol{\beta}$ in (A.4), based on the power prior (A.2), matches (A.1) under the hierarchical GLM if and only if*

$$a_{0k}\left[(L_0+1)I - \sum_{k=1}^{L_0}\left(I + \Omega\hat{\Sigma}_{0k}^{-1}\right)^{-1}\right] = \left(I + \Omega\hat{\Sigma}_{0k}^{-1}\right)^{-1} \tag{A.5}$$

for $k = 1, 2, \ldots, L_0$.

Appendix B: Proofs of Theorems

Proof of Theorem 2.1:

Under the hierarchical model in (2.1), (2.2), and (2.3), we can write the joint posterior of (θ, θ_0) as

$$\pi(\theta, \theta_0, \mu|\boldsymbol{y}, \boldsymbol{y}_0) \propto \exp\left\{-\frac{n(\bar{y}-\theta)^2}{2\sigma^2} - \frac{n_0(\bar{y}_0-\theta_0)^2}{2\sigma_0^2} - \frac{(\theta-\mu)^2 + (\theta_0-\mu)^2}{2\tau^2} - \frac{(\mu-\alpha)^2}{2\nu^2}\right\}.$$

Integrating out μ gives

$$\pi(\theta, \theta_0|\boldsymbol{y}, \boldsymbol{y}_0) \propto \exp\left\{-\frac{n\theta^2 - 2n\bar{y}\theta}{2\sigma^2} - \frac{n_0\theta_0^2 - 2n_0\bar{y}_0\theta_0}{2\sigma_0^2} - \frac{\theta^2+\theta_0^2}{2\tau^2} - \frac{\left(\frac{\theta+\theta_0}{\tau^2} + \frac{\alpha}{\nu^2}\right)^2}{2(\frac{1}{\nu^2} + \frac{2}{\tau^2})}\right\}.$$

After integrating out θ_0, straightforward calculations lead to

$$\begin{aligned}
&\pi(\theta|\boldsymbol{y},\boldsymbol{y}_0)\\
\propto\ &\exp\Bigg\{-\frac{1}{2}\Bigg[\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})}\right)\theta^2-2\left(\frac{n\bar{y}}{\sigma^2}+\frac{\alpha}{\tau^2\nu^2(\frac{1}{\nu^2}+\frac{2}{\tau^2})}\right)\theta\\
&-\frac{\left(\frac{n_0\bar{y}_0}{\sigma_0^2}+\frac{\alpha}{\tau^2\nu^2(\frac{1}{\nu^2}+\frac{2}{\tau^2})}+\frac{\theta}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})}\right)^2}{\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})}}\Bigg]\Bigg\}\\
\propto\ &\exp\Bigg\{-\frac{1}{2}\Bigg[\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})}-\frac{1}{(\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2}))^2(\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})})}\right)\theta^2\\
&-2\left(\frac{n\bar{y}}{\sigma^2}+\frac{\alpha}{\tau^2\nu^2\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}+\frac{\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})}(\frac{n_0\bar{y}_0}{\sigma_0^2}+\frac{\alpha}{\tau^2\nu^2(\frac{1}{\nu^2}+\frac{2}{\tau^2})})}{\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})}}\right)\theta\Bigg]\Bigg\}. \qquad \text{(B.1)}
\end{aligned}$$

Thus, (2.4) directly follows from (B.1).

<u>Proof of Theorem 2.2</u>:

It is easy to show that μ_h in (2.5) equals μ_p in (2.10) if and only if $\alpha = 0$. When $\alpha = 0$, μ_h reduces to

$$\mu_h = \sigma_h^2\left[\frac{n\bar{y}}{\sigma^2}+\frac{\frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}\cdot\frac{n_0\bar{y}_0}{\sigma_0^2}}{\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}}\right]. \qquad \text{(B.2)}$$

To match (B.2) to μ_p in (2.10), we have to set

$$a_0 = \frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})(\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})})} \qquad \text{(B.3)}$$

and

$$\frac{n}{\sigma^2}+a_0\frac{n_0}{\sigma_0^2}=\sigma_h^{-2}. \qquad \text{(B.4)}$$

Note that (B.4) directly yields that σ_h^2 in (2.6) is equal to σ_p^2 in (2.11) Now, using (B.3), (B.4) reduces to

$$\begin{aligned}
&\frac{\frac{n_0}{\sigma_0^2}}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})(\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})})}\\
=\ &\frac{1}{\tau^2}-\frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}-\frac{1}{\left(\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)\right)^2\left(\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}\right)}. \qquad \text{(B.5)}
\end{aligned}$$

After some algebra, we obtain

$$\begin{aligned}
&\frac{\frac{n_0}{\sigma_0^2}}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})(\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4(\frac{1}{\nu^2}+\frac{2}{\tau^2})})} \\
= \quad & \frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)} - \frac{\frac{\tau^2}{\nu^2}+1}{\left(\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)\right)^2\left(\frac{n_0}{\sigma_0^2}+\frac{1}{\tau^2}-\frac{1}{\tau^4\left(\frac{1}{\nu^2}+\frac{2}{\tau^2}\right)}\right)}.
\end{aligned}$$

Thus, (B.5) holds if and only if $\nu^2 \to \infty$. When $\nu^2 \to \infty$, (B.3) reduces to (2.12), which completes the proof of Theorem 2.2.

Proof of Theorem 2.3:

Using (2.13) and (2.14), we have

$$\begin{aligned}
&\pi(\boldsymbol{\beta}, \boldsymbol{\beta}_0, \boldsymbol{\mu}|\boldsymbol{y}, X, \boldsymbol{y}_0, X_0) \\
\propto \quad & \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-X\boldsymbol{\beta})'(\boldsymbol{y}-X\boldsymbol{\beta})\}\exp\{-\frac{1}{2\sigma_0^2}(\boldsymbol{y}_0-X_0\boldsymbol{\beta}_0)'(\boldsymbol{y}_0-X_0\boldsymbol{\beta}_0)\} \\
& \times\exp\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\}\exp\{-\frac{1}{2}(\boldsymbol{\beta}_0-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{\beta}_0-\boldsymbol{\mu})\}.
\end{aligned}$$

Integrating out $\boldsymbol{\beta}_0$ leads to

$$\begin{aligned}
\pi(\boldsymbol{\beta}, \boldsymbol{\mu}|\boldsymbol{y}, X, \boldsymbol{y}_0, X_0) \quad \propto \quad & \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-X\boldsymbol{\beta})'(\boldsymbol{y}-X\boldsymbol{\beta})-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\} \\
& \times\exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\mu}'(\Omega^{-1}-\Omega^{-1}(\Omega^{-1}+\frac{1}{\sigma_0^2}X_0'X_0)^{-1}\Omega^{-1})\boldsymbol{\mu} \\
& \quad -2\boldsymbol{\mu}'\Omega^{-1}(\Omega^{-1}+\frac{1}{\sigma_0^2}X_0'X_0)^{-1}\frac{1}{\sigma_0^2}X_0'\boldsymbol{y}_0\Big]\Big\}.
\end{aligned}$$

To obtain the marginal distribution of $\boldsymbol{\beta}$, we need to integrate out $\boldsymbol{\mu}$. After some further algebra, we obtain

$$\begin{aligned}
&\pi(\boldsymbol{\beta}|\boldsymbol{y}, X, \boldsymbol{y}_0, X_0) \\
\propto \quad & \exp\Big\{-\frac{1}{2}\Big(\boldsymbol{\beta}'\Big[\frac{1}{\sigma^2}X'X+\Omega^{-1}-\Omega^{-1}(2\Omega^{-1}-\Omega^{-1}(\Omega^{-1}+\frac{1}{\sigma_0^2}X_0'X_0)^{-1}\Omega^{-1})^{-1}\Omega^{-1}\Big]\boldsymbol{\beta} \\
& -2\boldsymbol{\beta}'\Big[\frac{1}{\sigma^2}X'Y+\Big(\Omega^{-1}(2\Omega^{-1}-\Omega^{-1}(\Omega^{-1}+\frac{1}{\sigma_0^2}X_0'X_0)^{-1}\Omega^{-1})^{-1}\Omega^{-1} \\
& \quad \times(\Omega^{-1}+\frac{1}{\sigma_0^2}X_0'X_0)^{-1}\frac{1}{\sigma_0^2}X_0'\boldsymbol{y}_0\Big)\Big]\Big)\Big\}.
\end{aligned}$$

Thus, (2.15) directly follows from the above equation.

Proof of Theorem 2.4:

A_h in (2.16) matches (equals) A_p in (2.21) if and only if

$$\frac{a_0}{\sigma_0^2}X_0'X_0 = \Omega^{-1}-\Omega^{-1}\left(2\Omega^{-1}-\Omega^{-1}\left(\Omega^{-1}+\sigma_0^{-2}X_0'X_0\right)^{-1}\Omega^{-1}\right)^{-1}\Omega^{-1}. \tag{B.6}$$

After some algebra, (B.6) can be written as

$$\frac{a_0}{\sigma_0^2}\Omega X_0' X_0 = I - \left(2I - \left(I + \sigma_0^{-2}\Omega X_0' X_0\right)^{-1}\right)^{-1},$$

which is equivalent to

$$\frac{a_0}{\sigma_0^2}\Omega X_0' X_0 \left(2I - \left(I + \sigma_0^{-2}\Omega X_0' X_0\right)^{-1}\right) = I - \left(I + \sigma_0^{-2}\Omega X_0' X_0\right)^{-1}. \tag{B.7}$$

Multiplying both sides of (B.7) by $(I + \sigma_0^{-2}\Omega X_0' X_0)$ gives

$$\frac{a_0}{\sigma_0^2}\Omega X_0' X_0 \left[2(I + \sigma_0^{-2}\Omega X_0' X_0) - I\right] = \sigma_0^{-2}\Omega X_0' X_0. \tag{B.8}$$

Since $X_0' X_0$ is invertible, then (B.8) reduces to (2.23). We further note that if (2.23) holds, B_h in (2.17) automatically matches B_p in (2.22) which completes the proof.

Proof of Theorem 4.1:

Let $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \ldots, \theta_{0k})'$. Under the hierarchical model in (4.1), (4.2), (4.3), and (4.4), we can write the joint posterior of $(\theta, \boldsymbol{\theta}_0, \mu)$ as

$$\pi(\theta, \boldsymbol{\theta}_0, \mu | \boldsymbol{y}, \boldsymbol{y}_0) \propto \exp\left\{-\frac{n(\bar{y}-\theta)^2}{2\sigma^2} - \frac{1}{2}\sum_{k=1}^{L_0}\frac{n_{0k}(\bar{y}_{0k}-\theta_{0k})^2}{\sigma_{0k}^2} - \frac{(\theta-\mu)^2}{2\tau^2} - \frac{1}{2\tau^2}\sum_{k=1}^{L_0}(\theta_{0k}-\mu)^2\right\}.$$

After integrating out $\boldsymbol{\theta}_0$, we obtain

$$\pi(\theta, \mu | \boldsymbol{y}, \boldsymbol{y}_0) \propto \exp\left\{-\frac{1}{2}\left[\frac{n(\bar{y}-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2} + \frac{L_0\mu^2}{\tau^2} - \sum_{k=1}^{L_0}\frac{\left(\frac{n_{0k}\bar{y}_{0k}}{\sigma_{0k}^2} + \frac{\mu}{\tau^2}\right)^2}{\frac{n_{0k}}{\sigma_{0k}^2} + \frac{1}{\tau^2}}\right]\right\}. \tag{B.9}$$

Integrating out μ from (B.9) yields

$$\begin{aligned}
\pi(\theta | \boldsymbol{y}, \boldsymbol{y}_0) \quad &\propto \quad \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)\theta^2 - \frac{2n\bar{y}\theta}{\sigma^2} - \frac{\left(\frac{\theta}{\tau^2} + \sum_{k=1}^{L_0}\frac{\frac{n_{0k}\bar{y}_{0k}}{\sigma_{0k}^2}}{\tau^2(\frac{n_{0k}}{\sigma_{0k}^2} + \frac{1}{\tau^2})}\right)^2}{\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0}\frac{1}{\tau^4(\frac{n_{0k}}{\sigma_{0k}^2} + \frac{1}{\tau^2})}}\right]\right\} \\
&\propto \quad \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2} - \frac{\tau^{-4}}{\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0}\frac{1}{\tau^4(\frac{n_{0k}}{\sigma_{0k}^2} + \frac{1}{\tau^2})}}\right)\theta^2\right.\right. \\
&\qquad \left.\left. -2\left(\frac{n\bar{y}}{\sigma^2} + \frac{\sum_{k=1}^{L_0}\frac{n_{0k}\bar{y}_{0k}}{\sigma_{0k}^2}\left(\frac{n_{0k}}{\sigma_{0k}^2} + \frac{1}{\tau^2}\right)^{-1}}{\tau^4\left(\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0}\frac{1}{\tau^4(\frac{n_{0k}}{\sigma_{0k}^2} + \frac{1}{\tau^2})}\right)}\right)\theta\right]\right\}.
\end{aligned} \tag{B.10}$$

Thus, (4.5) directly follows from (B.10).

Proof of Theorem 4.2:

To show $\mu_{mh} = \mu_{mp}$ in (4.6) and (4.12) and $\sigma^2_{mh} = \sigma^2_{mp}$ in (4.7) and (4.13), we take a_{0k} in (4.14). It suffices to show

$$\frac{n}{\sigma^2} + \sum_{k=1}^{L_0} a_{0k} \frac{n_{0k}}{\sigma^2_{0k}} = \sigma^{-2}_{mh}. \tag{B.11}$$

Note that if (B.11) is true, then we have $\sigma^2_{mh} = \sigma^2_{mp}$. Using (4.13), (B.11) reduces to

$$\sum_{k=1}^{L_0} a_{0k} \frac{n_{0k}}{\sigma^2_{0k}} = \frac{1}{\tau^2} - \frac{1}{\tau^4\left(\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0} \frac{1}{\tau^4(\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2})}\right)}. \tag{B.12}$$

Using (4.14), we are led to

$$\begin{aligned}\sum_{k=1}^{L_0} a_{0k} \frac{n_{0k}}{\sigma^2_{0k}} &= \frac{\sum_{k=1}^{L_0} \frac{\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2} - \frac{1}{\tau^2}}{\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2}}}{\tau^4\left(\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0} \frac{1}{\tau^4(\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2})}\right)} = \frac{L_0 - \frac{1}{\tau^2}\sum_{k=1}^{L_0} \frac{1}{\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2}}}{\tau^4\left(\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0} \frac{1}{\tau^4(\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2})}\right)} \\ &= \frac{\tau^2\left[\frac{L_0+1}{\tau^2} - \frac{1}{\tau^4}\sum_{k=1}^{L_0} \frac{1}{\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2}} - \frac{1}{\tau^2}\right]}{\tau^4\left(\frac{L_0+1}{\tau^2} - \sum_{k=1}^{L_0} \frac{1}{\tau^4(\frac{n_{0k}}{\sigma^2_{0k}} + \frac{1}{\tau^2})}\right)},\end{aligned}$$

which is equal to the right-hand side of (B.12). This completes the proof.

References

Berry, D. A. (1991). "Bayesian Methodology in Phase III Trails." *Drug Information Association Journal*, 25: 345–368. 551

Berry, D. A. and Hardwick, J. (1993). "Using Historical Controls in Clinical Trials: Application to ECMO." In Berry, D. A. and Stangl, D. K. (eds.), *Statistical Decision Theory and Related Topics V*, 141–156. New York: Springer-Verlag. 551

Berry, D. A. and Stangl, D. K. (1996). "Bayesian Methods in Health-Related Research." In Berry, D. A. and Stangl, D. K. (eds.), *Bayesian Biostatistics*, 3–66. New York: Marcel Dekker. 551

Brophy, J. M. and Joseph, L. (1995). "Placing Trials in Context using Bayesian Analysis - GUSTO Revisited by Reverend Bayes." *Journal of the American Medical Association*, 273: 871–875. 551

Chen, M.-H., Ibrahim, J. G., and Shao, Q.-M. (2000). "Power Prior Distributions for Generalized Linear Models." *Journal of Statistical Planning and Inference*, 84: 121–137. 551

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999a). "A New Bayesian Model for Survival Data with a Surviving Fraction." *Journal of the American Statistical Association*, 94: 909–919. 551

Chen, M.-H., Ibrahim, J. G., and Yiannoutsos, C. (1999b). "Prior Elicitation, Variable Selection and Bayesian Computation for Logistic Regression Models." *Journal of the Royal Statistical Society, Series B*, 61: 223–242. 551, 552

Eddy, D. M., Hasselblad, V., and Schaachter, R. (1992). *Meta-analysis by the Confidence Profile Method: The Statistical Synthesis of Evidence*. New York: Academic Press. 551

Fryback, D. G., Chinnis, J. O., and Ulvila, J. W. (2001). "Bayesian Cost-effectiveness Analysis - An Example Using the GUSTO Trial." *International Journal of Treatment and Healh Care*, 17: 83–97. 551

Greenhouse, J. B. and Wasserman, L. (1995). "Robust Bayesian Methods for Monitoring Clinical Trials." *Statistics in Medicine*, 14: 1379–1391. 551

Ibrahim, J. G. and Chen, M.-H. (1998). "Prior Distributions and Bayesian Computation for Proportional Hazards Models." *Sankhya, Series B*, 60: 48–64. 551

— (2000). "Power Prior Distributions for Regression Models." *Statistical Science*, 15: 46–60. 551, 552

Ibrahim, J. G., Chen, M.-H., and MacEachern, S. N. (1999). "Bayesian Variable Selection for Proportional Hazards Models." *Canadian Journal of Statistics*, 27: 701–717. 551, 552

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. 551

Ibrahim, J. G. and Laud, P. W. (1994). "A Predictive Approach to the Analysis of Designed Experiments." *Journal of the American Statistical Association*, 89: 309–319. 557

Ibrahim, J. G., Ryan, L. M., and Chen, M.-H. (1998). "Use of Historical Controls to Adjust for Covariates in Trend Tests for Binary Data." *Journal of the American Statistical Association*, 93: 1282–1293. 551

Laud, P. W. and Ibrahim, J. G. (1995). "Predictive Model Selection." *Journal of the Royal Statistical Society, Series B*, 57: 247–262. 557

Lin, Z. (1993). "Statistical Methods for Combining Historical Controls With Clinical Trial Data." *Unpublished Ph.D. Dissertation*, Duke University, Durham, NC. 551

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2003). *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. Chichester: Wiley. 551

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. 551

Tan, S., Machin, D., Tai, B. C., and Foo, K. F. (2002). "A Bayesian Re-assessment of Two Phase II Trials of Gemcitabine in Metastatic Nasopharyngeal Cancer." *British Journal of Cancer*, 86: 843–850. 551

Zellner, A. (1986). "On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions." In Goel, P. and Zellner, A. (eds.), *Bayesian Inference and Decision Techniques*, 233–243. Amsterdam: Elsevier Science Publishers B.V. 557

Acknowledgments

The authors wish to thank the Editor and two referees for several suggestions which have greatly improved the paper. This research was partially supported by NIH grants #CA 70101 and #CA 74015.