Statistical Science
2000, Vol. 15, No. 1, 46–60

Power Prior Distributions for Regression Models

Joseph G. Ibrahim and Ming-Hui Chen

Abstract. We propose a general class of prior distributions for arbitrary regression models. We discuss parametric and semiparametric models. The prior specification for the regression coefficients focuses on observable quantities in that the elicitation is based on the availability of historical data D_0 and a scalar quantity a_0 quantifying the uncertainty in D_0. Then D_0 and a_0 are used to specify a prior for the regression coefficients in a semiautomatic fashion. The most natural specification of D_0 arises when the raw data from a similar previous study are available. The availability of historical data is quite common in clinical trials, carcinogenicity studies, and environmental studies, where large data bases are available from similar previous studies. Although the methodology we present here is quite general, we will focus only on using historical data from similar previous studies to construct the prior distributions. The prior distributions are based on the idea of raising the likelihood function of the historical data to the power a_0, where $0 \leq a_0 \leq 1$. We call such prior distributions *power prior* distributions. We examine the power prior for four commonly used classes of regression models. These include generalized linear models, generalized linear mixed models, semiparametric proportional hazards models, and cure rate models for survival data. For these classes of models, we discuss the construction of the power prior, prior elicitation issues, propriety conditions, model selection, and several other properties. For each class of models, we present real data sets to demonstrate the proposed methodology.

Key words and phrases: Cure rate model, generalized linear model, Gibbs sampling, historical data, prior elicitation, model selection, proportional hazards model, random effects model.

1. INTRODUCTION

Prior elicitation perhaps plays the most crucial role in Bayesian inference. Although noninformative and improper priors may be useful and easier to specify for certain problems, they cannot be used in all applications, such as model selection or model comparison, as it is well known that proper priors are required to compute Bayes factors and posterior model probabilities. In addition, it is well known that Bayes factors are generally quite sensitive to the choices of hyperparameters of vague proper priors, and thus one cannot simply specify vague proper priors in model selection contexts to avoid informative prior elicitation. In addition, noninformative priors can cause instability in the posterior estimates and convergence problems for the Gibbs sampler. This can occur if the posterior surface is flat when using noninformative or improper priors. Moreover, noninformative priors do not make use of real prior information that one may have for a specific application. Thus, informative priors are essential in these situations, and, in general, they are useful in applied research settings where the investigator has access to previous studies measuring the same response and covariates as the cur-

Joseph G. Ibrahim is Associate Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney St., Boston, Massachusetts 02115. Ming-Hui Chen is Associate Professor, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Rd., Worcester, Massachusetts 01609-2280.

rent study. For example, in many cancer and AIDS clinical trials, current studies often use treatments that are very similar to or are slight modifications of treatments used in previous studies. We refer to data arising from previous similar studies as *historical data* throughout. In carcinogenicity studies, for example, large historical databases exist for the control animals from previous experiments. In all of these situations, it is natural to incorporate the historical data into the current study by quantifying it with a suitable prior distribution on the model parameters. The methodology discussed here can be applied to each of these situations as well as in other applications that involve historical data.

From a Bayesian perspective, historical data from past similar studies can be very helpful in interpreting the results of the current study. For example, historical control data can be very helpful in interpreting the results of a carcinogenicity study. According to Haseman, Huff and Boorman (1984), historical data can be useful when control tumor rates are low and when marginal significance levels are obtained in a test for dose effects. Suppose, for example, that 4 of 50 animals in an exposed group develop a specific tumor, compared with 0 of 50 in a control group. This difference is not statistically significant ($p = 0.12$, based on Fisher's exact test). However, the difference may be biologically significant if the observed tumor type is known to be extremely rare in the particular animal strain being studied. By specifying a suitable prior distribution on the control response rates that reflect the observed rates of a particular defect over a large series of past studies, one can derive a modified test statistic that incorporates historical information. If the defect is rare enough in the historical series, then even the difference of 4/50 versus 0/50 will be statistically significant based on a method that appropriately incorporates historical information.

To fix ideas, suppose we have historical data from a similar previous study, denoted by $D_0 = (n_0, y_0, X_0)$, where n_0 is the sample size of the historical data, y_0 is the $n_0 \times 1$ response vector, and X_0 is the $n_0 \times p$ matrix of covariates based on the historical data. The power prior is defined to be the likelihood function based on the historical data D_0, raised to a power a_0, where $0 \le a_0 \le 1$ is a scalar parameter that controls the influence of the historical data on the current data. One of the most useful applications of the power prior is for model selection problems, since these priors inherently automate the informative prior specification for all possible models in the model space. They are quite attractive in this context, since specifying meaningful informative prior distributions for the parameters in each model is a difficult task requiring contextual interpretations of a large number of parameters. In variable subset selection, for example, the prior distributions for all possible subset models are automatically determined once the historical data D_0 and the parameter a_0 are specified. Berger and Mallows (1988) refer to such priors as "semiautomatic" in their discussion of Mitchell and Beauchamp (1988). Chen, Manatunga and Williams (1998) use the power prior for heritability estimates from human twin data. Chen, Ibrahim and Yiannoutsos (1999) demonstrate the use of the power prior in variable selection contexts for logistic regression. Ibrahim, Chen and Ryan (2000) and Chen, Ibrahim, Shao and Weiss (1999) develop the power prior for the class of generalized linear mixed models. Ibrahim and Chen (1998), Ibrahim, Chen and MacEachern (2000), Chen, Ibrahim and Sinha (1999) and Chen, Dey and Sinha (1999) develop the power prior for various types of models for survival data.

The rest of this paper is organized as follows. In Section 2, we give the general development of the power prior for arbitrary regression models and discuss its interpretation and various advantages. In Section 3, we present the power prior for the class of generalized linear models and discuss two detailed applications. In Section 4, we present the power prior for the class of generalized linear mixed models and give an example illustrating variable subset selection. In Section 5, we examine the power prior for a specific class of semiparametric proportional hazards models. In Section 6, we study the power prior for a novel class of cure rate models for survival data. In Section 7, we discuss generalizations of the power prior and other elicitation techniques, and we compare our development to other methods. We close the article with a brief discussion.

2. THE POWER PRIOR

We consider the power prior for an arbitrary regression model. Let the data from the *current* study be denoted by $D = (n, y, X)$, where n denotes the sample size, 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 $L(\theta \mid D)$, where θ is a vector of indexing parameters. Thus, $L(\theta \mid D)$ is a general likelihood function for an arbitrary regression model, such as a generalized linear model, random effects model, nonlinear model or a survival model with right censored data. Now suppose we have historical data from a similar previous study, denoted by $D_0 = (n_0, y_0, X_0)$. Further, let $\pi_0(\theta \mid \cdot)$ denote the prior distribution for θ

before the historical data D_0 is observed. We shall call $\pi_0(\theta \mid \cdot)$ the *initial prior* distribution for θ. Given a_0, we define the *power prior* distribution of θ for the current study as

$$(2.1) \qquad \pi(\theta \mid D_0, a_0) \propto L(\theta \mid D_0)^{a_0} \pi_0(\theta \mid c_0),$$

where c_0 is a specified hyperparameter for the initial prior, and a_0 is a scalar prior parameter that weights the historical data relative to the likelihood of the current study. The prior parameter c_0 controls the impact of $\pi_0(\theta \mid c_0)$ on the entire prior, and the parameter a_0 controls the influence of the historical data on $\pi(\theta \mid D_0, a_0)$. The parameter a_0 can be interpreted as a relative precision parameter for the historical data. It is reasonable 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 θ. As a_0 becomes smaller, the tails of (2.1) become heavier. Setting $a_0 = 1$, (2.1) corresponds to the update of $\pi_0(\theta \mid c_0)$ using Bayes' theorem. That is, with $a_0 = 1$, (2.1) corresponds to the posterior distribution of θ from the previous study. When $a_0 = 0$, the prior does not depend on the historical data, and in this case $\pi(\theta \mid D_0, a_0 = 0) \equiv \pi_0(\theta \mid c_0)$. Thus, $a_0 = 0$ is equivalent to prior specification with no incorporation of historical data. Therefore, (2.1) can be viewed as a generalization of the usual Bayesian update of $\pi_0(\theta \mid c_0)$. 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 study, or when the sample sizes of the two studies are quite different.

The hierarchical power prior specification is completed by specifying a (proper) prior distribution for a_0. Thus we propose a joint power prior distribution for (θ, a_0) of the form

$$(2.2) \quad \pi(\theta, a_0 \mid D_0) \propto L(\theta \mid D_0)^{a_0} \pi_0(\theta \mid c_0) \pi(a_0 \mid \gamma_0),$$

where γ_0 is a specified hyperparameter vector. A natural choice for $\pi(a_0 \mid \gamma_0)$ is a beta prior. However, other choices, including a truncated gamma prior or a truncated normal prior can be used. These three priors for a_0 have similar theoretical properties, and our experience shows that they have similar computational properties. In practice, they yield similar results when the hyperparameters are appropriately chosen. Thus, for a clear focus and exposition, we will use a *beta* distribution for $\pi(a_0 \mid \gamma_0)$ throughout this article. The beta prior for a_0 appears to be the most natural prior to use and leads to the most natural elicitation scheme. The prior in (2.2) does not have a closed form in general, but it has several attractive theoretical and computational properties for the classes of models considered here. One attractive feature of (2.2) is that it creates heavier tails for the marginal prior of θ than the prior in (2.1), which assumes that a_0 is a fixed value. This is a desirable feature since it gives the investigator more flexibility in weighting the historical data. In addition, our construction of (2.2) is quite general, with various possibilities for $\pi_0(\theta \mid c_0)$. If $\pi_0(\theta \mid c_0)$ is proper, then (2.2) is guaranteed to be proper. Further, (2.2) can be proper even if $\pi_0(\theta \mid c_0)$ is an improper uniform prior. Specifically, Ibrahim, Ryan and Chen (1998) and Chen, Ibrahim and Yiannoutsos (1999) characterize the propriety of (2.2) for generalized linear models and show that, for fixed a_0, the prior converges to a multivariate normal distribution as $n_0 \to \infty$. For the class of generalized linear mixed models, Ibrahim, Chen and Ryan (2000), Chen et al. (1999) and Chen, Dey and Sinha (2000) characterize the propriety of (2.2) and derive various other theoretical properties of the power prior. Ibrahim, Chen and MacEachern (2000) and Ibrahim and Chen (1998) characterize various properties of (2.2) for proportional hazards models, and Chen, Ibrahim and Sinha (1999) examine various theoretical properties of (2.2) for a proposed class of cure rate models. We will briefly summarize the conditions for propriety as well as other properties for the above-mentioned models here, but refer the reader to these articles for details and proofs.

The power prior defined in (2.2) can easily be generalized to multiple historical data sets. If there are L_0 historical studies, we define $D_{0k} = (n_{0k}, X_{0k}, y_{0k})$ to be the historical data based on the kth study, $k = 1, \ldots, L_0$ and $D_0 = (D_{01}, \ldots, D_{0L_0})$. In this case, it may be desirable to define a weight parameter a_{0k} for each historical study, and take the a_{0k}'s to be i.i.d. beta random variables with hyperparameters $\gamma_0 \equiv (\delta_0, \lambda_0)$, $k = 1, \ldots, L_0$. Letting $a_0 = (a_{01}, \ldots, a_{0L_0})$, the prior in (2.2) can be generalized as

$$(2.3) \quad \pi(\beta, a_0 \mid D_0) \propto \left(\prod_{k=1}^{L_0} \left[L(\beta \mid D_{0k}) \right]^{a_{0k}} \pi(a_{0k} \mid \gamma_0) \right) \cdot \pi_0(\beta \mid c_0).$$

3. POWER PRIOR FOR GENERALIZED LINEAR MODELS

Let y_{0i} denote the ith component of y_0, let $x'_{0i} = (x_{0i1}, x_{0i2}, \ldots, x_{0ip})$ denote the ith row of X_0 with $x_{0i1} = 1$ corresponding to an intercept, let $\eta_{0i} = x'_{0i}\beta$ denote the linear predictor based on the historical data, where β is a $p \times 1$ vector, and let $D_0 = (n_0, y_0, X_0)$ denote the historical data. Then, the likelihood function of β based on the historical data D_0 is given by

$$L(\beta \mid D_0) = \prod_{i=1}^{n_0} \exp\{\tau_0(y_{0i}\theta_{0i} - g(\theta_{0i})) + c(y_{0i}, \tau_0)\}, \tag{3.1}$$

where $\theta_{0i} = \theta(\eta_{0i})$, $\theta(\cdot)$ is a monotonic differentiable function often referred to as the θ-link, $g(\cdot)$ and $c(\cdot)$ are known functions and τ_0 is a known parameter. The power prior for the class of generalized linear models (GLM's) takes the form

$$\pi(\beta, a_0 \mid D_0) \propto [L(\beta \mid D_0)]^{a_0} \pi_0(\beta \mid c_0) \pi(a_0 \mid \gamma_0), \tag{3.2}$$

where $\pi(a_0 \mid \gamma_0) \propto a_0^{\delta_0 - 1}(1-a_0)^{\lambda_0 - 1}$ and $\gamma_0 = (\delta_0, \lambda_0)$. The prior in (3.2) will not have a closed form in general, but has several attractive properties. First, as shown in Ibrahim, Chen and Ryan (2000) and Chen, Ibrahim and Yiannoutsos (1999), if $\pi_0(\beta \mid c_0) \propto 1$, $L(\beta \mid D_0)$ satisfies mild regularity conditions and $\delta_0 > p$, then (3.2) is proper. In addition, if $\pi_0(\beta \mid c_0) \propto 1$ and a_0 is taken to be fixed, then Ibrahim, Ryan and Chen (1998) show that, as $n_0 \to \infty$, $\pi(\beta \mid D_0)$ converges to a normal distribution with mean $\hat{\beta}$ and covariance matrix $a_0^{-1}(X_0' V_0 X_0)^{-1}$, where $\hat{\beta}$ is the maximizer of $L(\beta \mid D_0)$ and V_0 is an $n_0 \times n_0$ diagonal matrix of variance functions of the GLM. When $\lambda_0 \to \infty$, $\pi(\beta, a_0 \mid D_0)$ becomes an improper uniform prior for β, resulting in no incorporation of the historical data. Also, when $\delta_0 \to \infty$, the historical data and the current data become equally weighted. For elicitation purposes, it is easier to work with the prior mean and standard deviation of a_0, that is, $\mu_{a_0} = \delta_0/(\delta_0 + \lambda_0)$ and $\sigma_{a_0} = (\mu_{a_0}(1-\mu_{a_0}))^{1/2}(\delta_0 + \lambda_0 + 1)^{-1/2}$. It is typically easier to specify $(\mu_{a_0}, \sigma_{a_0})$ and then solve for (δ_0, λ_0) from the implied equations. The investigator may choose μ_{a_0} to be small if he or she assigns low prior weight to the historical data. If a large prior weight is desired, then $\mu_{a_0} \geq 0.5$ may be suitable. In practice, we recommend that several choices of $(\mu_{a_0}, \sigma_{a_0})$ be used, including ones that give small and large weight to the historical data, and several sensitivity analyses be conducted. We do not recommend doing an analysis based on one set of prior parameters. The choices recommended here can be used as starting points from which sensitivity analyses can be based.

To illustrate the roles of the prior parameters in the power priors, we consider the following logistic regression model. We simulated a data set consisting of $n_0 = 200$ independent Bernoulli observations with success probability

$$p_{0i} = \frac{\exp\{-0.5 + 0.5x_{0i}\}}{1 + \exp\{-0.5 + 0.5x_{0i}\}}, \quad i = 1, \ldots, n_0,$$

where the x_{0i} are i.i.d. normal random variables with mean 0 and standard deviation 0.5. Using the Gibbs sampler, for each given set of (δ_0, λ_0), we generated 50,000 iterates from the joint prior distribution $\pi(\beta, a_0 \mid D_0)$ given by (3.2) taking $\pi_0(\beta \mid c_0) \propto 1$. The detailed implementation scheme of the Gibbs sampler can be found in Chen, Ibrahim and Yiannoutsos (1999). Figure 1 shows the marginal prior densities of β_1 (intercept) and β_2 (slope) for three choices of $(\mu_{a_0}, \sigma_{a_0})$. From Figure 1, we see that, as μ_{a_0} gets smaller, both marginal prior density curves get flatter, but the prior modes of β_1 and β_2 for all three choices of $(\mu_{a_0}, \sigma_{a_0})$ are almost the same. Although not shown in Figure 1, we also obtained the marginal prior densities for β_1 and β_2 for $(\delta_0, \lambda_0) = (3, 3)$, which are nearly uniform over the real line.

3.1 Illustrative Examples

Example 1. *AIDS Data.* For illustration we consider an analysis of the AIDS study ACTG036 using the data from study ACTG019 as historical data. 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 results of this study were published in Volberding et al. (1990). 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 development of AIDS-related complex (ARC), and a 0 indicates otherwise. Several covariates were also measured. The ACTG036 study was also a placebo-controlled clinical trial comparing AZT to placebo in patients with hereditary coagulation disorders. The results of this study have been published by Merigan et al. (1991). 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 development of AIDS-related complex (ARC), and a 0 indicates otherwise. We let D_0 denote the data from the ACTG019 study and D denote the data from the ACTG036 study.

Chen, Ibrahim and Yiannoutsos (1999) use the priors given by (3.2) and the logistic regression model to carry out Bayesian variable subset selection, which yields the model containing an intercept, CD4 count (cell count per cubic millimeter of serum), age and treatment as the model with the largest posterior probability. For that

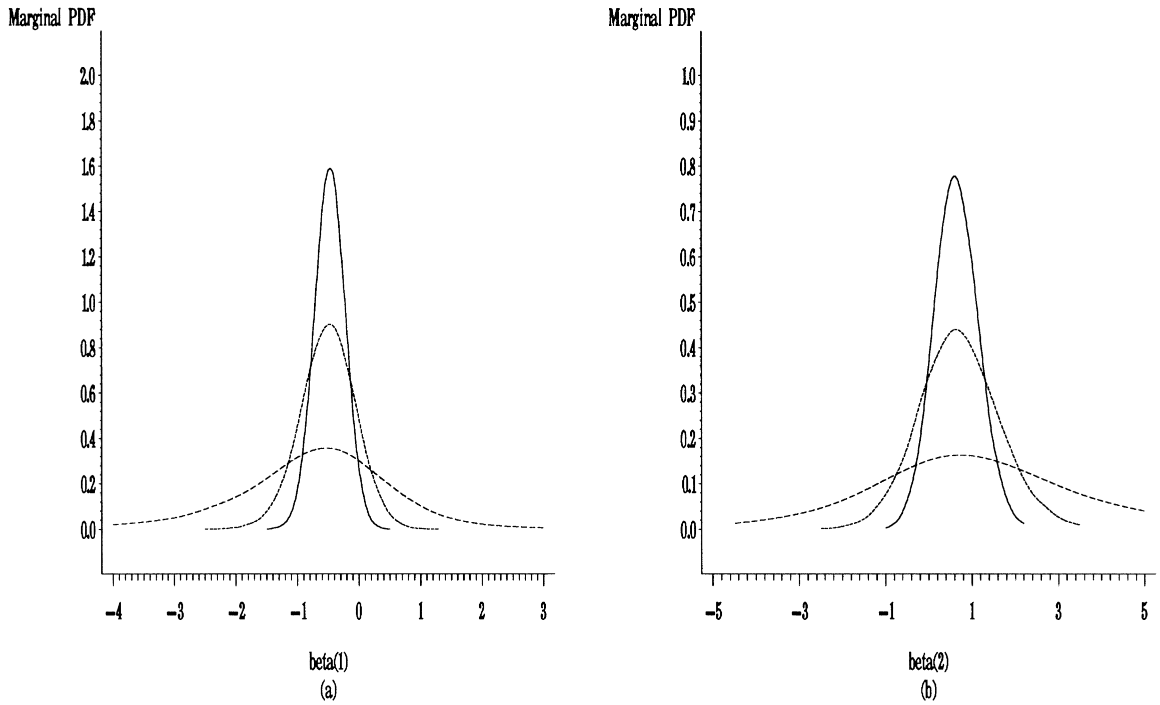

FIG. 1. *Plots of marginal posterior densities for* β_1 *and* β_2; (*solid curve*) $(\mu_{a_0}, \sigma_{a_0}) = (0.94, 0.031)$; (*dotted curve*) $(\mu_{a_0}, \sigma_{a_0}) = (0.5, 0.078)$; (*dashed curve*) $(\mu_{a_0}, \sigma_{a_0}) = (0.5, 0.151)$.

model, we here use the power prior (3.2), taking $\pi_0(\beta \mid c_0) \propto 1$, to obtain posterior estimates of the regression coefficients for various choices of $(\mu_{a_0}, \sigma_{a_0})$. From Table 1, we see that, as the weight for ACTG019 study increases, the posterior mean of a_0 [denoted $E(a_0 \mid D, D_0)$] increases, the posterior standard deviations (std. dev.) for all parameters decrease and the 95% highest probability density (HPD) intervals get narrower. Most noticeably, when $(\delta_0, \lambda_0) = (100, 1)$, none of the HPD intervals for the regression coefficients contains 0. Table 1 also indicates that the HPD intervals are not too sensitive to moderate changes in $(\mu_{a_0}, \sigma_{a_0})$. This is a comforting feature, because it implies that the HPD intervals are fairly robust with respect to the hyperparameters of a_0. This same robustness feature is also exhibited in posterior model probability calculations (see Chen, Ibrahim and Yiannoutsos, 1999). We mention that the Monte Carlo method of Chen and Shao (1999) was to calculate 95% highest probability density intervals for the parameters of interest.

EXAMPLE 2. *Carcinogenicity study.* Consider a study involving $r+1$ groups of test animals, one of which serves as a control and the remaining r receive a test compound at increasing dose levels. Denote the dose levels by $d_1 < d_2 < \cdots < d_{r+1}$, where $d_1 \equiv 0$ denotes the dose level for the control group. Let n_i denote the number of animals receiving the ith dose and define

$$y_{ij} = \begin{cases} 1, & \text{if the } j\text{th animal in the } i\text{th dose group has a tumor,} \\ 0, & \text{otherwise.} \end{cases}$$

Let $x_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ be a $p \times 1$ vector of covariates for the jth animal in the ith dose group for $j = 1, \ldots n_i$, $i = 1, \ldots, r+1$. Denote by θ_{ij}, the probability that animal j in the ith dose level develops a tumor. We assume that y_{ij} has a Bernoulli distribution with parameter θ_{ij}, which depends on the covariates through a logistic model,

$$\theta_{ij} = \frac{\exp\{\beta_0 + bd_i + \beta_1' x_{ij}\}}{1 + \exp\{\beta_0 + bd_i + \beta_1' x_{ij}\}}, \tag{3.3}$$

where x_{ij} is the covariate vector for animal ij, β_0 is the intercept, b is the dose coefficient and β_1 is a $p \times 1$ vector of regression coefficients corresponding to x_{ij}, $i = 1, \ldots, r+1$, $j = 1, \ldots, n_i$. We write the $(p+1) \times 1$ vector of regression coefficients as

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

and expand x_{ij} to include an intercept. The main goal for this problem is to derive a test statistic for

TABLE 1
Posterior estimates for AIDS data

(δ_0, λ_0)	$(\mu_{a_0}, \sigma_{a_0})$	$E(a_0 \mid D, D_0)$	Posterior variable	Posterior mean	95% HPD std. dev.	Interval
(5, 5)	(0.50, 0.151)	0.02	Intercept	−4.389	0.725	(−5.836, −3.055)
			CD4 count	−1.437	0.394	(−2.238, −0.711)
			Age	0.135	0.221	(−0.314, 0.556)
			Treatment	−0.120	0.354	(−0.817, 0.570)
(20, 20)	(0.50, 0.078)	0.09	Intercept	−3.803	0.511	(−4.834, −2.868)
			CD4 count	−1.129	0.300	(−1.723, −0.559)
			Age	0.176	0.195	(−0.214, 0.552)
			Treatment	−0.223	0.300	(−0.821, 0.364)
(30, 30)	(0.50, 0.064)	0.13	Intercept	−3.621	0.436	(−4.489, −2.809)
			CD4 count	−1.028	0.265	(−1.551, −0.515)
			Age	0.194	0.185	(−0.170, 0.557)
			Treatment	−0.259	0.278	(−0.805, 0.288)
(50, 1)	(0.98, 0.019)	0.26	Intercept	−3.337	0.323	(−3.978, −2.715)
			CD4 count	−0.865	0.211	(−1.276, −0.448)
			Age	0.233	0.160	(−0.081, 0.548)
			Treatment	−0.314	0.230	(−0.766, 0.138)
(100, 1)	(0.99, 0.010)	0.53	Intercept	−3.144	0.231	(−3.601, −2.705)
			CD4 count	−0.746	0.161	(−1.058, −0.429)
			Age	0.271	0.135	(0.001, 0.529)
			Treatment	−0.356	0.181	(−0.717, −0.011)

TABLE 2
Summary of historical data from 70 *studies of female* B6C3F1 *mice*

	Mean	Std. dev.	Minimum	Maximum
Animals	49.83	4.50	43	79.00
Tumors	12.44	9.41	1	54.00

the null hypothesis H_0: $b = 0$ using the likelihood based on (3.3) along with the power prior (3.2). From a frequentist perspective, two natural test statistics include the score and likelihood ratio tests. These test statistics are based on the marginal likelihood of b, denoted $L(b)$, which is found by integrating the joint posterior density of (β, a_0) given b, with respect to (β, a_0). Details of the computations of the score test and the likelihood ratio test can be found in Ibrahim, Ryan and Chen (1998).

To illustrate this methodology, we consider an experiment conducted on female mice at the National Toxicology Program (NTP) with a commercial disinfectant, called *o*-benzyl-*p*-chlorophenol (see Alden, 1994). A detailed and comprehensive analysis of these data using the power priors can be found in Ibrahim, Ryan and Chen (1998).

Consider an analysis of liver tumors in female mice. Our data set involves $n = 144$ female mice exposed at 0, 120 and 240 ppm, labeled d_1, d_2 and d_3, respectively. The numbers of animals at the three dose levels are 50, 44 and 44 respectively, with 13, 15 and 17 tumors, respectively, resulting. There were a total of $L_0 = 70$ studies in the historical database, with a total of 3,488 animals, 871 of which had liver tumors. We take $D_{0k} = (n_{0k}, y_{0k}, X_{0k})$ to be the historical data from the kth study, $k = 1, \ldots, 70$. Table 2 gives a summary of the animal and tumor counts for all $L_0 = 70$ studies. There are two covariates available in the historical and current data, time to death (T) and weight at 12 months (W). We consider a logistic regression model including the covariates time to death and weight at 12 months. The dose covariate is always included in the model for the current data. To determine the evidence of a dose trend, we first conducted a logistic regression on the current data. The p-value for the dose effect was 0.15, suggesting a nonsignificant dose trend. The logistic regression, however, gave significant p-values for each of the covariates. Thus, the logistic regression analysis for the current data appears to suggest a nonsignificant dose effect, but significance in the regression coefficients for all of the covariates.

We use a uniform improper initial prior for β in the analysis, that is, $\pi_0(\beta \mid c_0) \propto 1$. To determine the effect of the historical data, we first conducted the score test using the point mass prior, $a_0 = (0, \ldots, 0)$ with probability 1. This corresponds to the usual score test with no incorporation of historical data, which yielded a nonsignificant p-value $(p = 0.281)$. On the other hand, using a point mass prior at $a_0 = (1, \ldots, 1)$ with the model including only dose (i.e., no covariates) yields a significant result $(p = 0.041)$. This demonstrates that the in-

TABLE 3
Model with (T, W)

μ_{a_0}	σ_{a_0}	Score test	p-value	LR	p-value
0.01	0.003	2.79	0.095	2.70	0.100
0.05	0.032	3.80	0.051	3.73	0.053
0.1	0.1	4.22	0.040	3.95	0.047
0.5	0.288	5.37	0.020	5.11	0.024
0.5	0.152	6.27	0.012	5.84	0.015
0.8	0.381	7.10	0.008	6.45	0.011

corporation of the historical data yields a significant score test for the dose trend. Table 3 shows the results of the score test and likelihood ratio test (denoted by LR) with covariates T and W for several values of the prior parameters $(\mu_{a_0}, \sigma_{a_0})$. With $\mu_{a_0} = 0.01$, the p-value for the score and likelihood ratio tests is not significant, whereas for values of $\mu_{a_0} \geq 0.05$, we see that the p-values become more and more significant as μ_{a_0} increases. When no covariates are included in the model, the p-value for the score test is not significant for values of $\mu_{a_0} \leq 0.1$ and becomes significant when $\mu_{a_0} \geq 0.5$. This shows that, when the covariates are included, significant results are obtained with much smaller values of μ_{a_0}, thus demonstrating the importance of the covariates in the analysis. The score test and likelihood ratio test become much more significant as μ_{a_0} increases.

4. POWER PRIORS FOR GENERALIZED LINEAR MIXED MODELS

Consider the generalized linear mixed model (GLMM),

$$\begin{aligned} &p(y_{it} \mid \beta, b_i, \tau) \\ &\quad = \exp\{\tau[y_{it}\theta(\eta_{it}) - g(\theta(\eta_{it}))] + c(y_{it}, \tau)\}, \end{aligned} \tag{4.1}$$

where $\eta_{it} = x'_{it}\beta + z'_{it}b_i$, b_i is a $q \times 1$ vector of random effects and z'_{it} and x'_{it} are vectors of covariates. Let X_i denote the $n_i \times p$ matrix with ith row x'_{it}, and let Z_i denote the $n_i \times q$ matrix with ith row z'_{it}. Letting $b = (b'_1, \ldots, b'_N)'$, $y = (y_{11}, \ldots, y_{Nn_N})'$ and $X = (X'_1, \ldots, X'_N)'$, $Z = \text{diag}(Z_1, \ldots, Z_N)$, the joint density of (y, b) based on N subjects for the GLMM is

$$p(y, b \mid \beta, T) = \prod_{i=1}^{N} \prod_{t=1}^{n_i} p(y_{it} \mid \beta, b_i)\pi(b_i \mid T), \tag{4.2}$$

where $\pi(b_i \mid T)$ is the normal distribution with mean 0 and covariance matrix $V = T^{-1}$. For ease of exposition, we will assume one previous study, because the generalization of the prior to multiple previous studies proceeds as in (2.3). Suppose there exist historical data with N_0 subjects that yielded the $n_{0i} \times 1$ response vector y_{0i} for subject i.

Let X_{0i} be an $n_{0i} \times p$ matrix of fixed covariates, and let Z_{0i} be an $n_{0i} \times q$ matrix of covariates for the $q \times 1$ vector of random effects b_{0i} for subject i, $i = 1, 2, \ldots, N_0$ for the historical data. Also let b_0, y_0, X_0 and Z_0 be defined similar to b, y, X and Z. Finally let $D_0 = (N_0, X_0, y_0, Z_0)$ denote the historical data. Given a_0, we propose to take the power prior distribution for β to be of the form

$$\begin{aligned} &\pi(\beta \mid D_0, T, a_0) \\ &\quad \propto \prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [p(y_{0it} \mid \beta, b_{0i})]^{a_0} \right. \\ &\qquad\qquad \left. \cdot\, \pi(b_{0i} \mid T)\, db_{0i} \right) \pi_0(\beta \mid c_0), \end{aligned} \tag{4.3}$$

where $p(y_{0it} \mid \beta, b_{0i}, \tau)$ is (4.1) with $(y_{0it}, b_{0i}, \tau_0)$ in place of (y_{it}, b_i, τ). That is, $p(y_{0it} \mid \beta, b_{0i})$ is the GLMM based on the historical data y_{0it}. We note that the construction of the power prior in (4.3) is based on exponentiating the historical data likelihood *given* the random effects, as opposed to exponentiating the marginal historical data likelihood after the random effects have been integrated out. The prior in (4.3) turns out to have several advantages and several attractive computational properties compared to a power prior based on the marginal historical likelihood. For example, a power prior based on the marginal historical data likelihood is computationally intractable, and it is not at all clear how to implement Markov chain Monte Carlo (MCMC) methods with such a prior, whereas MCMC methods for (4.3) are relatively straightforward to implement.

The power prior specification is completed by specifying priors for (a_0, σ_b^2, ρ). We take these parameters independent a priori. We specify a beta prior for a_0, an inverse gamma prior for σ_b^2, denoted $\text{IG}(\alpha_0, \omega_0)$, and a scaled beta prior for ρ, denoted $\text{scbeta}(\phi_0, \psi_0)$. Thus, we propose a joint power prior distribution of the form

$$\begin{aligned} &\pi(\beta, a_0, \sigma_b^2, \rho \mid D_0) \\ &\quad \propto \prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [p(y_{0it} \mid \beta, b_{0i})]^{a_0} \right. \\ &\qquad\qquad \left. \cdot\, \pi(b_{0i} \mid T)\, db_{0i} \right) \pi_0(\beta \mid c_0) \\ &\qquad \cdot\, a_0^{\delta_0 - 1}(1 - a_0)^{\lambda_0 - 1} \\ &\qquad \cdot\, (\sigma_b^2)^{-(\alpha_0+1)} \exp(-\sigma_b^{-2}\omega_0) \\ &\qquad \cdot\, (1 + \rho)^{\phi_0 - 1}(1 - \rho)^{\psi_0 - 1}, \end{aligned} \tag{4.4}$$

where $\theta_0 \equiv (\delta_0, \lambda_0, \alpha_0, \omega_0, \phi_0, \psi_0)$ are known prior parameters. In our analyses, we take $\pi_0(\beta|c_0) \propto 1$ and take choices of the prior parameters θ_0 leading to vague but proper prior distributions. If we assume $\pi_0(\beta|c_0) \propto 1$, mild regularity conditions on $p(y_0|\beta, b_0)$, $\delta_0 > p$ and $\alpha_0 > p/2$, then (4.4) is proper. We refer the reader to Chen et al. (1999) for detailed theorems and proofs characterizing the propriety of (4.4)

4.1 Example: School Nurse Visits

We illustrate our methodology with a data set involving repeated measures of school nurse visits. We also illustrate variable subset selection for the GLMM with these data. The implementation of the variable subset selection procedure is identical to that described in Section 5 and thus is omitted here for brevity. The response for each of 51 grade school children with complete data is a two-dimensional vector of the yearly nurse visits for each of two years. The covariance structure of the b_i is an AR-1 model for all models with $q \equiv 2$ and each Z a 2×2 identity matrix. Children participated in a laboratory cold pressor pain paradigm experiment with four trials of arm immersion in very cold water. The goal of this analysis was to see if children's health care usage as measured by nurse visits could be predicted from the results of the experiment. The full model contains seven covariates and an intercept term, implying $2^7 = 128$ possible subset models. The seven covariates are age (x_1), two treatment indicator variables (x_2 and x_3), coping style (x_4), tolerance (x_5), rating (x_6) and a coping style by rating interaction (x_7). The response variable y is the total number of nurse visits, which we model as a Poisson GLMM. For these data, we have $N = 33$, $N_0 = 18$, and all of the n_i's and n_{0i}'s are equal to 2. Since the n_i's, n_{0i}'s and q are all equal, we can directly apply the complete hierarchical centering reparameterization of Gelfand, Sahu and Carlin (1996).

Table 4 gives results for the top three models with $\delta_0 = 10$, $\lambda_0 = 10$, that is, $\mu_{a_0} = 0.5$ and $\sigma_{a_0} = 0.11$. In addition, we take $\pi_0(\beta|c_0) \propto 1$ and take vague priors for σ_b^2 and ρ. Specifically, for σ_b^2, we take $(\alpha_0, \omega_0) = (0.005, 0.005)$ and, for ρ, we take $\phi_0 = \psi_0 = 1$. Table 4 indicates that treatment, rating, coping style and rating by coping style interaction are important covariates for explaining the number of nurse visits. To examine the sensitivity of model selection to the choices of $(\mu_{a_0}, \sigma_{a_0})$, we computed posterior model probabilities for several choices of $(\mu_{a_0}, \sigma_{a_0})$. From Table 5, we see that, for several choices of $(\mu_{a_0}, \sigma_{a_0})$, the (x_2, x_4, x_6, x_7) model obtains the largest posterior probability. The pattern of the posterior probability structure for the other models for these choices of prior parameters is similar to that of Table 4. However, model selection does become sensitive to the choice of $(\mu_{a_0}, \sigma_{a_0})$ when we give large weight to the historical data, as demonstrated in the last line of Table 5. Here, we see that the top model is $(x_2, x_3, x_4, x_5, x_6, x_7)$. Thus, it appears for this data set that there is no clearcut top model, but perhaps two or three adequate models, which all contain the covariates treatment, rating, coping style and rating by coping style interaction.

TABLE 4
Posterior model probabilities for $(\mu_{a_0}, \sigma_{a_0}) = (0.5, 0.11)$

$\boldsymbol{m}$	$\boldsymbol{p(m\|D)}$
(x_2, x_4, x_6, x_7)	0.119
$(x_2, x_3, x_4, x_6, x_7)$	0.111
$(x_2, x_4, x_5, x_6, x_7)$	0.059

TABLE 5
Posterior model probabilities for several values of $(\mu_{a_0}, \sigma_{a_0})$

$(\boldsymbol{\mu_{a_0}, \sigma_{a_0}})$	$\boldsymbol{m}$	$\boldsymbol{p(m\|D)}$
(0.5, 0.078)	(x_2, x_4, x_6, x_7)	0.100
(0.5, 0.064)	(x_2, x_4, x_6, x_7)	0.085
(0.5, 0.050)	(x_2, x_4, x_6, x_7)	0.073
(0.91, 0.027)	$(x_2, x_3, x_4, x_5, x_6, x_7)$	0.046

5. PROPORTIONAL HAZARDS MODELS

A proportional hazards model is defined by a hazard function of the form

$$h(t, x) = h_b(t) \exp(x'\beta), \tag{5.1}$$

where $h_b(t)$ denotes the baseline hazard function at time t, x denotes the $p \times 1$ covariate vector for an arbitrary individual in the population and β denotes a $p \times 1$ vector of regression coefficients. We first construct a finite partition of the time axis as in Ibrahim and Chen (1998). Let $0 \le s_0 < s_1 < \cdots < s_M$ denote this partition with $s_M > \max_i(t_i)$. Further, let

$$\delta_i = h_b(s_i) - h_b(s_{i-1})$$

denote the increment in the baseline hazard in the interval $(s_{i-1}, s_i]$, $i = 1, \ldots, M$, and let $\Delta = (\delta_1, \ldots, \delta_M)$. We follow Ibrahim and Chen (1998) for constructing the likelihood function of (β, Δ). To construct the likelihood function, we use a piecewise-constant baseline hazard model and use only information about which interval the failure times fall into. For an arbitrary individual in the population, the cumulative distribution function for the proportional hazards model at time s is given by

$$F(s) = 1 - \exp\left\{-\exp\{\eta\}\int_0^s h_b(t)\,dt\right\}$$

$$\simeq 1 - \exp\left\{-\exp\{\eta\}\left((s-s_0)h_b(s_0) + \sum_{i=1}^{M}\delta_i(s-s_{i-1})^+\right)\right\}, \tag{5.2}$$

where $(t)^+ = t$ if $t > 0$, $(t)^+ = 0$ otherwise and $\eta = x'\beta$. Let p_i denote the probability of a failure in the interval $(s_{i-1}, s_i]$, d_i denote the number of failures and let c_i be the number of right censored observations in the ith interval, respectively, $i = 1, \ldots, M$. For ease of exposition, we order the observations so that in the ith interval the first d_i are failures and the remaining c_i are right censored, $i = 1, \ldots, M$. Let x_{ik} denote the vector of covariates for the kth individual in the ith interval and define

$$u_{ik}(\beta) = \exp\{x'_{ik}\beta\},$$

$$a_i = \sum_{j=i+1}^{M}\sum_{k=1}^{d_j} u_{jk}(\beta)(s_{j-1} - s_{i-1}),$$

$$b_i = \sum_{j=i}^{M}\sum_{k=d_j+1}^{d_j+c_j} u_{jk}(\beta)(s_j - s_{i-1}),$$

$$T_i(\Delta) = (s_i - s_{i-1})\sum_{j=1}^{i}\delta_j.$$

Let $D = (n, y, X, \nu)$ denote the data for the current study, where $\nu = (\nu_1, \ldots, \nu_n)'$ is the $n \times 1$ vector of censoring indicators. The likelihood function for the current study over all M intervals is given by

$$L(\beta, \Delta \mid D) = \left\{\prod_{i=1}^{M}\exp\{-\delta_i(a_i + b_i)\}\right\} \cdot \left\{\prod_{i=1}^{M}\prod_{k=1}^{d_i}(1 - \exp\{-u_{ik}(\beta)T_i(\Delta)\})\right\}.$$

For ease of exposition, we assume that we have one previous study. Let $D_0 = (n_0, y_0, X_0, \nu_0)$ denote the historical data and let $\pi_0(\beta, \Delta)$ denote the initial prior distribution for (β, Δ). The joint power prior distribution for (β, Δ, a_0) takes the form

$$\pi(\beta, \Delta, a_0 \mid D_0) \propto L(\beta, \Delta \mid D_0)^{a_0}\pi_0(\beta, \Delta)a_0^{\delta_0 - 1}(1 - a_0)^{\lambda_0 - 1}, \tag{5.3}$$

where $L(\beta, \Delta \mid D_0)$ is the likelihood function of (β, Δ) based on the historical data. We note that, in (5.3), (β, Δ) are not independent, and also the components of Δ are not independent a priori. To simplify the prior specification, we take $\pi_0(\beta, \Delta) = \pi_0(\beta \mid c_0)\,\pi_0(\Delta \mid \theta_0)$, where c_0 and θ_0 are fixed hyperparameters. Specifically, we take a p-dimensional multivariate normal density for $\pi_0(\beta \mid c_0)$ with mean 0 and covariance matrix $c_0 W_0$, where c_0 is a specified scalar and W_0 is a specified $p \times p$ diagonal matrix. We take $\pi_0(\Delta \mid \theta_0)$ to have a gamma density of the form $\pi_0(\Delta \mid \theta_0) \propto \prod_{i=1}^{M}\delta_i^{f_{0i}-1}\exp\{-\delta_i g_{0i}\}$, where $\theta_0 = (f_{01}, g_{01}, \ldots, f_{0M}, g_{0M})$. If $\pi_0(\beta \mid c_0) \propto 1$, then (5.3) is proper if $\pi_0(\Delta)$ is proper and $\delta_0 > p$. Details of these results and the Gibbs sampling techniques for this model can be found in Ibrahim and Chen (1998) and Ibrahim, Chen and MacEachern (2000).

5.1 Applications to Variable Selection

The power priors lead to a novel formulation for eliciting prior model probabilities in the variable subset selection problem. Let $\mathscr{M}$ denote the model space, and let m be a specific model in $\mathscr{M}$. Further, under model m, let $\beta^{(m)}$ denote the vector of regression coefficients, let $X_0^{(m)}$ denote the covariate matrix and let $D_0^{(m)} = (n_0, y_0, X_0^{(m)}, \nu_0)$ denote the historical data. Let

$$p_0^*(\beta^{(m)}, \Delta \mid D_0^{(m)}) = L(\beta^{(m)}, \Delta \mid D_0^{(m)})\pi_0(\beta^{(m)} \mid d_0)\pi_0(\Delta \mid \kappa_0) \tag{5.4}$$

denote the unnormalized posterior density of $(\beta^{(m)}, \Delta)$ based only on the historical data $D_0^{(m)}$, and (d_0, κ_0) are specified hyperparameters. We propose to take the prior probability of model m as

$$p(m) \equiv p(m \mid D_0^{(m)}) = \frac{\iint p_0^*(\beta^{(m)}, \Delta \mid D_0^{(m)})\,d\beta^{(m)}\,d\Delta}{\sum_{m\in\mathscr{M}}\iint p_0^*(\beta^{(m)}, \Delta \mid D_0^{(m)})\,d\beta^{(m)}\,d\Delta}. \tag{5.5}$$

Because Δ is viewed as a nuisance parameter, we recommend taking $\kappa_0 = \theta_0$ to simplify the elicitation scheme. The prior parameter d_0 controls the impact of $\pi_0(\beta^{(m)} \mid d_0)$ on the prior model probability $p(m)$. This choice for $p(m)$ has several nice interpretations. First, $p(m)$ in (5.5) corresponds to the posterior probability of model m based on the data $D_0^{(m)}$ using a uniform prior for the previous study. That is, $p_0(m) = 2^{-p}$ for $m \in \mathscr{M}$, where $p_0(m)$ is the prior probability of model m before observing the historical data $D_0^{(m)}$. Therefore, $p(m) \propto p(m \mid D_0^{(m)})$, and thus $p(m)$ corresponds to the usual Bayesian update of $p_0(m)$ using $D_0^{(m)}$ as the data. Second, as $d_0 \to 0$, $p(m)$ reduces to a uniform prior on the model space. Therefore, as $d_0 \to 0$, the historical data $D_0^{(m)}$ have a minimal impact in determining

$p(m)$. On the other hand, with a large value of d_0, $\pi_0(\beta^{(m)}|d_0)$ plays a minimal role in determining $p(m)$, and in this case the historical data play a larger role in determining $p(m)$. Thus as $d_0 \to \infty$, $p(m)$ will be regulated by the historical data. The parameter d_0 plays the same role as c_0 and thus serves as a tuning parameter to control the impact of $D_0^{(m)}$ on the prior model probability $p(m)$. We refer the reader to Ibrahim and Chen (1998) for more details on the variable selection problem for proportional hazards models.

5.2 Example: Myeloma Data

We consider two studies in multiple myeloma. Krall, Uthoff and Harley (1975) analyzed data from a study (historical data) on multiple myeloma in which $n_0 = 65$ patients were treated with alkylating agents. A few years later, another multiple myeloma study (current study) using similar alkylating agents was undertaken by the Eastern Cooperative Oncology Group (ECOG). This study, labeled E2479, had $n = 479$ patients with the same set of covariates being measured as the historical data. Here, y_0 consists of the 65 survival times from the historical study and $X_0^{(m)}$ is an $n_0 \times p_m$ matrix of covariates, where p_m denotes the number of covariates under model m.

Our main goal in this example is to illustrate the proposed power priors for variable selection. A detailed data analysis can be found in Ibrahim and Chen (1998). We also examine the sensitivity of the posterior probabilities to the choices of $(\mu_{a_0}, \sigma_{a_0})$, c_0 and d_0. Our analysis is based on $p = 8$ covariates for the full model. These are blood urea nitrogen (x_1), hemoglobin (x_2), platelet count (x_3) (1 if normal, 0 if abnormal), age (x_4), white blood cell count (x_5), fractures (x_6), percentage of the plasma cells in bone marrow (x_7) and serum calcium (x_8). We conduct sensitivity analyses with respect to (i) c_0, (ii) d_0 and (iii) $(\mu_{a_0}, \sigma_{a_0})$. To compute the prior and posterior model probabilities, 50,000 Gibbs iterations were used to get convergence.

Table 6 gives the model with the largest posterior probability using $(\mu_{a_0}, \sigma_{a_0}) = (0.5, 0.063)$ (i.e., $\delta_0 = \lambda_0 = 30$) for several values of c_0. For each value of c_0 in Table 6, the model $(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$ obtains the largest posterior probability, and thus model choice is not sensitive to these values. In addition, for $d_0 = 3$ and for any $c_0 \geq 3$, the $(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$ model obtains the largest posterior probability. Although not shown in Table 6, values of $c_0 < 3$ do not yield $(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$ as the top model. Thus, model choice may become sensitive to the choice of c_0 when $c_0 < 3$. When d_0 is changed, the $(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$ model obtains the largest prior probability when $d_0 \geq 3$. With values of $d_0 < 3$, however, model choice may be sensitive to the choice of d_0. For example, when $d_0 = 0.0001$ and $c_0 = 10$, the top model is $(x_1, x_2, x_4, x_5, x_7, x_8)$ with posterior probability of 0.42 and the second best model is $(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$ with posterior probability of 0.31. Finally, we mention that as both c_0 and d_0 become large, the $(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$ model obtains the largest posterior model probability. A monotonic decrease in the posterior probability of model $(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$ occurs as c_0 and d_0 are increased. This indicates that there is a moderate impact of the historical data on model choice. A sensitivity analysis was also conducted with respect to $(\mu_{a_0}, \sigma_{a_0})$, and model choice is not sensitive to the choice of $(\mu_{a_0}, \sigma_{a_0})$. For a wide variety of choices for $(\mu_{a_0}, \sigma_{a_0})$, $(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$ obtains the largest posterior probability. In addition, there is a monotonic increase in the posterior model probability as more weight is given to the historical data.

TABLE 6
Posterior model probabilities for $(\mu_{a_0}, \sigma_{a_0}) = (0.5, 0.063)$, $d_0 = 3$ and various choices of c_0

c_0	m	$p(m)$	$p(D\|m)$	$p(m\|D)$
3	$(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$	0.015	0.436	0.769
10	$(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$	0.015	0.310	0.679
30	$(x_1, x_2, x_3, x_4, x_5, x_7, x_8)$	0.015	0.275	0.657

6. CURE RATE MODELS

Cure rate models have become increasingly popular for analyzing survival data, because for many diseases a significant proportion of patients are "cured" after sufficient follow-up. Here we present a recently proposed model of Chen, Ibrahim and Sinha (1999) to demonstrate the power priors for this class of models. A similar frequentist formulation of the model is also discussed in Tsodikov (1998).

The model can be derived as follows. Suppose that, for an individual in the population, we let C denote the number of metastasis-competent tumor cells for that individual left active after the initial treatment, and assume that C has a Poisson distribution with mean ω. Also let Z_i denote the random time for the ith metastasis-competent cells to produce a metastatic tumor. That is, Z_i can be viewed as an incubation time for the ith metastastic tumor cell. The variables Z_i, $i = 1, 2, \ldots$, are assumed to be independent conditional on C, and identically distributed with a common distribution function $F(t) = 1 - S(t)$. The time to relapse of cancer can be defined by the random variable $T = \min\{Z_i, 0 \le i \le C\}$, where $P(Z_0 = \infty) = 1$. The survival function for T is given by

$$
\begin{aligned}
S_p(t) &= P(\text{no metastatic cancer by time } t)\\
&= P(C=0) + P(Z_1 > t, \ldots, Z_C > t, C \ge 1)\\
&= \exp(-\omega) + \sum_{k=1}^{\infty} S(t)^k \frac{\omega^k}{k!} \exp(-\omega)\\
&= \exp(-\omega + \omega S(t)) = \exp(-\omega F(t)).
\end{aligned} \tag{6.1}
$$

The cure fraction (i.e., cure rate) is given by $S_p(\infty) \equiv P(C=0) = \exp(-\omega)$.

Suppose we have n subjects, and let C_i denote the number of metastasis-competent tumor cells for the ith subject. Further, we assume that the C_i's are i.i.d. Poisson random variables with mean ω, $i = 1, \ldots, n$. Further, suppose $Z_{i1}, \ldots, Z_{i,C_i}$ are the i.i.d. incubation times for the C_i tumor cells for the ith subject, which are unobserved, and all have cumulative distribution function $F(\cdot)$, $i = 1, \ldots, n$. We specify a Weibull distribution for $F(\cdot)$. We denote the indexing parameter (possibly vector valued) by γ, and thus write $F(\cdot\,|\,\gamma)$ and $S(\cdot\,|\,\gamma)$. Let t_i denote the failure time for subject i, where t_i may be right censored. Let v_i denote the censoring time so that we observe $y_i = \min(t_i, v_i)$, where the censoring indicator $\nu_i = I(t_i \le v_i)$ equals 1 if t_i is a failure time and 0 if it is right censored. We can represent the observed data by the vector (n, y, ν), where $y = (y_1, \ldots, y_n)$ and $\nu = (\nu_1, \ldots, \nu_n)$. Also, let $C = (C_1, \ldots C_n)$. We incorporate covariates for the cure rate model (6.1) through the cure rate parameter ω. Let $x_i' = (x_{i1}, \ldots, x_{ip})$ denote the $p \times 1$ vector of covariates for the ith subject, and let $\beta = (\beta_1, \ldots, \beta_p)'$ denote the corresponding vector of regression coefficients. We relate ω to the covariates by $\omega_i = \exp(x_i'\beta)$, so that the cure rate for subject i is $\exp(-\omega_i) = \exp(-\exp(x_i'\beta))$, $i = 1, \ldots, n$. With this relation, we can write the complete data likelihood of (β, γ) as

$$
\begin{aligned}
&L(\beta, \gamma\,|\,D)\\
&= \left(\prod_{i=1}^{n} S(y_i\,|\,\gamma)^{C_i - \nu_i} \big(C_i f(y_i\,|\,\gamma)\big)^{\nu_i}\right)\\
&\quad \cdot \exp\left\{\sum_{i=1}^{n} \big[C_i x_i'\beta - \log(C_i!) - \exp(x_i'\beta)\big]\right\},
\end{aligned} \tag{6.2}
$$

where $D = (n, y, X, \nu, C)$, X is the $n \times p$ matrix of covariates, $f(y_i\,|\,\gamma) = \alpha y^{\alpha-1} \exp\{\lambda - y^{\alpha}\exp(\lambda)\}$ and $S(y_i\,|\,\gamma) = \exp(-y_i^{\alpha}\exp(\lambda))$.

Let C_0 denote the unobserved vector of latent counts, and let $D_0 = (n_0,\ y_0,\ X_0,\ \nu_0,\ C_0)$ denote the complete historical data. Further, let $\pi_0(\beta, \gamma)$ denote the initial prior distribution for (β, γ). The joint power prior distribution for (β, γ, a_0) takes the form

$$
\begin{aligned}
&\pi(\beta, \gamma, a_0\,|\,D_{0,\text{obs}})\\
&\quad \propto \left[\sum_{C_0} L(\beta, \gamma\,|\,D_0)\right]^{a_0}\\
&\qquad \cdot \pi_0(\beta, \gamma) a_0^{\delta_0 - 1}(1 - a_0)^{\lambda_0 - 1},
\end{aligned} \tag{6.3}
$$

where $L(\beta, \gamma\,|\,D_0)$ is the complete data likelihood given in (6.2) with D being replaced by the historical data D_0, and $D_{0,\text{obs}} = (n_0, y_0, X_0, \nu_0)$. We mention that the sum over C_0 in (6.3) has a closed form, making it computationally tractable. We take a noninformative prior for $\pi_0(\beta, \gamma)$. Specifically, we take β and γ to be independent at this stage and take an improper uniform prior for β. For $\gamma = (\alpha, \lambda)$, we take a gamma prior for α with small shape parameter α_0 and small scale parameter τ_0. Also, we take an independent normal prior for λ with mean 0 and variance c_0, where c_0 is large. The prior in (6.3) does not have a closed form but has several attractive theoretical properties. First, we note that if $\pi_0(\beta, \gamma)$ is proper, then (6.3) is guaranteed to be proper. Further, let X_0^* denote the $n_0 \times p$ matrix with rows $\nu_{0i} x_{0i}'$. Then, if (i) X_0^* is of full rank and (ii) $\delta_0 > p$ and $\pi(\lambda)$ is proper, then (6.3) is proper. We refer the reader to Chen, Ibrahim and Sinha (1999) for details of the theorems and proofs.

6.1 Example: Melanoma Data

We consider data from a phase III melanoma clinical trial conducted by the Eastern Cooperative Oncology Group. The current study, denoted E1684, was a two-arm clinical trial involving patients randomized to one of two treatment arms: high-dose interferon (IFN) or observation. Three covariates and an intercept are included in the analyses. The covariates are age (x_1), gender (x_2) (male, female) and performance status (x_3) (fully active, other). Performance status is abbreviated by PS in Tables 7. After deleting missing observations, a total of $n = 284$ observations are used in the analysis. A Kaplan–Meier plot for overall survival shows a "plateau" (see Chen, Ibrahim and Sinha, 1999) in the survival curve, and thus a cure rate model appears to be suitable for these data.

Several years earlier, a similar melanoma study with the same patient population was conducted by ECOG. This study, denoted by E1673, serves as the historical data for our Bayesian analysis of E1684. A total of $n_0 = 650$ patients are used in the historical data. Using the E1673 study as historical data,

TABLE 7
Melanoma data: posterior estimates of the model parameters with $\alpha \sim \text{Gamma}(1, 0.01)$ *and* $\lambda \sim N(0, 10,000)$

$E(a_0 \mid D_{obs}, D_{0,obs})$	Variable	Posterior mean	Posterior std. dev.	95% HPD interval
0 (with probability 1)	Intercept	0.094	0.106	(−0.115, 0.301)
	Age	0.091	0.073	(−0.054, 0.231)
	Gender	−0.125	0.159	(−0.435, 0.186)
	PS	−0.226	0.260	(−0.733, 0.281)
	α	1.312	0.087	(1.145, 1.484)
	λ	−1.356	0.123	(−1.596, −1.114)
0.064	Intercept	0.212	0.108	(0.005, 0.426)
	Age	0.108	0.068	(−0.025, 0.242)
	Gender	−0.159	0.148	(−0.447, 0.133)
	PS	−0.160	0.236	(−0.630, 0.292)
	α	1.117	0.066	(0.989, 1.245)
	λ	−1.525	0.127	(−1.779, −1.282)
0.142	Intercept	0.251	0.100	(0.051, 0.446)
	Age	0.119	0.063	(−0.004, 0.243)
	Gender	−0.196	0.137	(−0.470, 0.068)
	PS	−0.094	0.215	(−0.533, 0.309)
	α	1.062	0.057	(0.949, 1.174)
	λ	−1.619	0.118	(−1.849, −1.389)
0.288	Intercept	0.257	0.089	(0.081, 0.431)
	Age	0.132	0.057	(0.019, 0.242)
	Gender	−0.241	0.123	(−0.481, 0.001)
	PS	−0.006	0.187	(−0.382, 0.352)
	α	1.028	0.050	(0.932, 1.127)
	λ	−1.700	0.106	(−1.909, −1.495)
1 (with probability 1)	Intercept	0.224	0.062	(0.106, 0.349)
	Age	0.159	0.041	(0.077, 0.239)
	Gender	−0.319	0.087	(−0.495, −0.153)
	PS	0.142	0.127	(−0.111, 0.386)
	α	0.997	0.036	(0.927, 1.067)
	λ	−1.822	0.076	(−1.970, −1.673)

we consider an analysis with the proposed priors (6.3). We take $\pi_0(\beta) \propto 1$ and for $\pi_0(\alpha \mid \nu_0, \tau_0)$ we take $\alpha_0 = 1$ and $\tau_0 = 0.01$ to ensure a proper prior. The parameter λ is taken to have a normal distribution with mean 0 and variance 10,000.

Table 7 gives posterior estimates of β based on several values of (δ_0, λ_0) using the proposed model (6.1). In Table 7 we obtain, for example, $E(a_0 \mid D_{obs},\ D_{0,\text{obs}}) = 0.064$ and 0.142 by taking $(\delta_0, \lambda_0) = (100, 100)$ and $(200, 1)$, respectively. Table 7 indicates a fairly robust pattern of behavior. The estimates of the posterior mean, standard deviation or highest posterior density (HPD) intervals of β do not change a great deal if a low or moderate weight is given to the historical data. However, if a higher than moderate weight is given to the historical data, these posterior summaries can change a lot. For example, when the posterior mean of a_0 is less than 0.064, we see that all of the HPD intervals for β include 0, and when the posterior mean of a_0 is greater than or equal to 0.064, some HPD intervals for β do not include 0. Thus, when we give more weight to the historical data, this has the potential of affecting our inference about β. The HPD interval for age does not include 0 when the posterior mean of a_0 is 0.288, and it includes 0 when the posterior mean of a_0 is less than 0.288. This finding is interesting, since it indicates that age is a potentially important prognostic factor for predicting survival in melanoma. Such a conclusion is not possible based on a frequentist or Bayesian analysis of the current data alone.

In addition, when the historical data and the current data are equally weighted, that is, $a_0 = 1$ with probability 1, the HPD intervals for age and gender both do not include 0, thus demonstrating the importance of gender in predicting overall survival. Thus, we see the potential impact of the historical data on the posterior analysis of β, and hence the potential impact on the posterior estimates of the cure rates. Another feature of Table 7 is that the posterior standard deviations of the β's become smaller and the HPD intervals become narrower as the posterior mean of a_0 increases. This is a strong feature of our model since it demonstrates that incorporation of historical data can yield more precise posterior estimates of β.

Incorporation of historical data can also affect the posterior estimates of the cure rates. The posterior estimates in the cure rates are quite different in the model with $E(a_0 \mid D_0, D_{0,\mathrm{obs}}) = 0.288$ compared to the one with no incorporation of historical data. The mean and standard deviations are 0.361 and 0.048 ($a_0 = 0$) and 0.310 and 0.062 for the model with $E(a_0 \mid D_{\mathrm{obs}}, D_{0,\mathrm{obs}}) = 0.288$. Thus we see that the mean cure rate drops from 0.361 to 0.310 when the historical data is incorporated. A partial explanation of this result is due to the fact that the historical data are much more mature than the current data, with nearly 20 years of follow-up and a smaller fraction of censored cases. These results are not surprising, and in fact appealing, since they give us a better estimate of the cure rate compared to an estimate based on the current data alone. Such a conclusion is not possible by a frequentist or Bayesian analysis using the current data alone. We also conducted a detailed sensitivity analysis for the regression coefficients by varying the hyperparameters for a_0 (i.e., (δ_0, λ_0)) and varying the hyperparameters for $\gamma = (\alpha, \lambda)$. Table 7 shows that the posterior estimates of the parameters are fairly robust as the hyperparameters (δ_0, λ_0) are varied. When we vary the hyperparameters for γ, the posterior estimates of β are also robust for a wide range of hyperparameter values.

7. GENERALIZATIONS AND COMPARISONS WITH OTHER METHODS

If historical data are not available from which to construct $D_0 = (n_0, y_0, X_0)$, then y_0 can be obtained via a prior prediction, including specifications based on a theoretical prediction model, expert opinion or case-specific information. For example, a theoretical model of the form $y_0 = g(X_0)$ may be available for obtaining the prior predictions, where X_0 is the covariate matrix corresponding to some model m_0, and g is a known function. Such prediction models are often used, for example, in respiratory studies measuring forced vital capacity and forced expiratory volume. Also, when historical data are not available, a common approach is take X_0 to be the covariate matrix of the current study, that is, $X_0 = X$ and $n_0 = n$. This approach has been motivated and considered by many, including Zellner (1986), Ibrahim and Laud (1994) and Laud and Ibrahim (1995). Thus, the power prior is in fact quite general and can be constructed even if historical data from a previous study is *not* available. In any case, the existence of historical data from a similar previous study leads to the most natural specification of D_0 and serves as the primary motivation for (2.2). Taking D_0 to be the raw data from a similar previous study results in a more natural, interpretable and automated specification for (2.2).

It sometimes occurs that the set of covariates measured in the previous study is a subset of the covariates measured in the current study. This may occur because the investigators discover "new" and potentially useful covariates to measure in the current study that were not measured in previous studies. In this case, we can modify (2.2) as follows. Let X_1 denote the $n \times r$ matrix of covariates in the current study that are common to the covariates in the previous study, and let X_2 be the $n \times s$ matrix of new covariates in the current study which are not measured in the previous study. Write

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix},$$

and let X_{01} represent the $n_{01} \times r$ matrix of covariates from the previous study, let X_{02} be an $n_{02} \times s$ matrix of covariates representing the new covariates and let $p = r + s$. The most natural choice for X_{01} is the raw covariate matrix from the historical data, and to take $X_{02} = X_2$. In this specification, we assume that the new covariates have small or negligible correlation to the common covariates, that is, $\mathrm{Corr}(X_1, X_2) \approx 0$. This may be a sensible assumption if in fact the new set of covariates in the current study is being scientifically investigated for the first time. Also let $D_{0j} = (n_{0j}, y_{0j}, X_{0j})$, where y_{0j} is the historical data corresponding to X_{0j}, $j = 1, 2$. Finally, we assume a priori independence between θ_1 and θ_2, which leads to the power prior

$$\begin{aligned} &\pi(\theta \mid D_0, a_0) \\ &\quad = \pi_1(\theta_1 \mid D_{01}, a_{01})\pi_2(\theta_2 \mid D_{02}, a_{02}) \\ &\quad \propto L(\theta_1 \mid D_{01})^{a_{01}} L(\theta_2 \mid D_{02})^{a_{02}} \pi_0(\theta_1, \theta_2 \mid c_0). \end{aligned} \tag{7.1}$$

The prior specification is completed by specifying independent beta priors for (a_{01}, a_{02}). A natural choice for y_{01} is the raw response vector from the previous study. The elicitation of y_{02} is less automatic since no a priori information is available for it. One can use expert opinion, fitted values or predictions for specifying y_{02}. For example, in logistic regression, one possible choice is to pick $y_{02} = (1/2, \dots, 1/2)$. The prior parameters for a_{02} are chosen to reflect vague prior beliefs, and thus a uniform prior for a_{02} would be reasonable. We mention that if we take a

TABLE 8
Comparisons to other methods

Method	Intercept (SD)	β_1 (SD)	β_2 (SD)	β_3 (SD)
$a_0 = 0$ using power prior	−4.78 (0.85)	−1.64 (0.45)	0.12 (0.23)	−0.05 (0.38)
ML for ACTG036	−4.40 (0.77)	−1.51 (0.42)	0.12 (0.22)	−.004 (0.36)
$a_0 = 1$ using power prior	−3.04 (0.17)	−0.68 (0.12)	0.30 (0.11)	−0.38 (0.14)
ML with pooled	−3.01 (0.17)	−0.67 (0.12)	0.30 (0.11)	−0.37 (0.14)
$b_i \sim N(0, 0.1)$	−3.13 (0.30)	−0.69 (0.12)	0.30 (0.11)	−0.38 (0.14)
$b_i \sim N(0, 10)$	−.3.27 (2.38)	−0.70 (0.12)	0.30 (0.11)	−0.38 (0.14)
meta-analysis	−3.19 (0.20)	−0.75 (0.13)	0.32 (0.11)	−0.34 (0.14)

point mass prior for a_{02} at $a_{02} = 0$, then (7.1) is improper. The prior in (7.1) reduces to (2.1) if the sets of covariates from the previous and current studies are identical. If the set of covariates in the current study is a subset of the covariates in the previous study, then we can construct a submatrix by omitting those columns corresponding to covariates not in the current study and take X_0 to be that submatrix. For more on these issues, see Ibrahim and Chen (1997) and Chen, Ibrahim and Yiannoutsos (1999). Finally, we mention that if $\text{Corr}(X_1, X_2)$ is not neglible, then the prior in (7.1) may not be adequate, and in this case a more general development is needed. This is an open research problem under current investigation.

The power prior in (2.1) gives results that are equivalent to other methods for special values of a_0. For example, when $a_0 = 0$ and $\pi_0(\theta | c_0) \propto 1$, then the power prior is the uniform improper prior and thus yields estimates similar to maximum likelihood. Table 8 shows results for $a_0 = 0$ for the AIDS data and results of a maximum likelihood analysis of ACTG036. We see that the estimates are nearly identical. When $a_0 = 1$, (2.1) corresponds to the posterior distribution of β based on the historical data. Therefore, taking $a_0 = 1$ essentially corresponds to pooling the historical and current data. Table 8 shows posterior estimates of β using $a_0 = 1$ for the AIDS data, and a maximum likelihood analysis based on pooling the ACTG019 and ACTG036 data sets. We see that the estimates are remarkably similar. The rows in Table 8 correspond to the estimates of β along with the corresponding standard deviation (SD) given in parentheses, for the various methods.

In addition, we mention that Bayesian inference using the power prior is related to maximum likelihood inference using a random effects model. For example, for the AIDS data, we can fit a random effects logistic regression model for the combined datasets ACTG019 and ACTG036, where the random effect accounts for the heterogeneity between studies. Denote the random effect by $b_i \sim N(0, \sigma_b^2)$. Table 8 shows results from a random effects model for the AIDS data using several values of σ_b^2. We see that, for small σ_b^2, we get results very similar to those of $a_0 = 1$. As σ_b^2 gets large, the estimates and standard errors of β are fairly robust, and the standard errors are slightly larger than those of $a_0 = 1$. We also note that meta-analysis type estimates are related to the power prior. For example, for the AIDS data, we can construct a meta-analysis type estimate of β as $\hat{\beta}_{\text{meta}} = w_0\hat{\beta}_1 + (1 - w_0)\hat{\beta}_2$, where $w_0 = n_0/(n_0 + n)$, $\hat{\beta}_1$ is the maximum likelihood estimate of β from the ACTG019 data alone and $\hat{\beta}_2$ is the maximum likelihood estimate of β based on the ACTG036 data alone. Table 8 shows estimates and standard errors for β for the AIDS data based on this meta analysis approach. We see that the estimates are quite comparable to the Bayesian analysis with $a_0 = 1$.

The relationship between the power prior and a maximum likelihood analysis using a random effects models has also been investigated for survival models. Chen, Harrington and Ibrahim (1999) examine relationships between the power prior and the frailty model and obtain similar conclusions as those reported here. We refer the reader to Chen, Harrington and Ibrahim (1999) for a detailed discussion.

8. DISCUSSION

We have presented a general class of prior distributions for arbitrary regression models, called the power priors. The power priors are constructed from historical data and were demonstrated in detail for several specific classes of models. These priors are quite useful in a wide variety of applications, including carcinogenicity studies and clinical trials. They are also quite useful in model selection contexts since they automate the prior elicitation procedure for the prior on the model space, as well as the model parameters arising from the different models in the model space. The priors are also quite robust under a variety of settings. Further research work is needed to study further computational properties of these priors, as well as other properties and modifications of the proposed priors.

ACKNOWLEDGMENTS

The authors thank the Executive Editor and an Editor for their helpful comments and suggestions, which have led to an improvement in this article. This work was supported in part by NSF Grant DMS-97-02172 and NIH Grants CA 70101 and CA 74015.

REFERENCES

ALDEN, C. J. (1994). Toxicology and carcinogenesis studies of *o*-benzyl-*p*-chlorophenol in F344/N rats and B6C3F_1 mice. Technical Report NTP 424, U.S. Dept. Health and Human Services.

BERGER, J. O. and MALLOWS, C. L. (1988). Discussion of Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83** 103-3-1034.

CHEN, M.-H., DEY, D. K. and SINHA, D. (2000). Bayesian analysis of multivariate mortality data with large families. *Appl. Statist.* **49** 129–144.

CHEN, M.-H., HARRINGTON, D. P. and IBRAHIM, J. G. (1999). Bayesian models for high-risk melanoma: a case study of ECOG trial E1690. Technical Report MS-06-99-22, Dept. Mathematical Sciences, Worcester Polytechnic Inst.

CHEN, M.-H., IBRAHIM, J. G., SHAO, Q.-M. and WEISS, R. E. (1999). Prior elicitation for model selection and estimation in generalized linear mixed models. Technical Report MS-01-99-17, Dept. Mathematical Sciences, Worcester Polytechnic Inst.

CHEN, M.-H., IBRAHIM, J. G. and SINHA, D. (1999). A new Bayesian model for survival data with a surviving fraction. *J. Amer. Statist. Assoc.* **94** 909–919.

CHEN, M.-H., IBRAHIM, J. G. and YIANNOUTSOS, C. (1999). Prior elicitation, variable selection and Bayesian computation for logistic regression models. *J. Roy. Statist. Soc. Ser. B* **61** 223–242.

CHEN, M.-H., MANATUNGA, A. K. and WILLIAMS, C. J. (1998). Heritability estimates from human twin data by incorporating historical prior information. *Biometrics* **54** 1348–1362.

CHEN, M.-H. and SHAO, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Comput. Graph. Statist.* **8** 69–92.

GELFAND, A. E., SAHU, S. K. and CARLIN, B. P. (1996). Efficient parametrisations for generalized linear mixed models (with discussion). In *Bayesian Statistics 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 165–180. Oxford Univ. Press.

HASEMAN, J. K., HUFF, J. and BOORMAN, G. A. (1984). Use of historical control data in carcinogenicity studies in rodents. *Toxocologic Pathology* **12** 126–135.

IBRAHIM J. G. and LAUD, P. W. (1994). A predictive approach to the analysis of designed experiments. *J. Amer. Statist. Assoc.* **89** 309–319.

IBRAHIM, J. G. and CHEN, M.-H. (1997). Predictive variable selection in the multivariate linear model. *Biometrics* **53** 465–478.

IBRAHIM, J. G. and CHEN, M.-H. (1998). Prior distributions and Bayesian computation for proportional hazards models. *Sankhyā Ser. B* **60** 48–64.

IBRAHIM, J. G., CHEN, M.-H. and MACEACHERN, S. N. (2000). Bayesian variable selection for proportional hazards models. *Canad. J. Statist.* To appear.

IBRAHIM, J. G., CHEN, M.-H. and RYAN, L. M. (2000). Bayesian variable selection for time series count data. *Statist. Sinica*. To appear.

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. *J. Amer. Statist. Assoc.* **93** 1282–1293.

KRALL, J. M., UTOFF, V. A. and HARLEY, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics* **31** 49–57.

LAUD , P. W. and IBRAHIM, J. G. (1995). Predictive model selection. *J. Roy. Statist. Soc. Ser. B* **57** 247–262.

MERIGAN, T. C., AMATO, D. A., BALSLEY, J., POWER, M., PRICE, W. A., BENIOT, S., PEREZ-MICHAEL, A., BROWNSTEIN, A., KRAMER, A. S., BRETTLER, D., ALEDORT, L., RAGNI, M. V., ANDES, A. W., GILL, J. C., GOLDSMITH, J., STABLER, S., SANDERS, N., GJERSET, G., USHER, J. and the NHF-ACTG 036 Study Group (1991). Placebo-controlled trial to evaluate zidovudine in treatment of human immunodeficiency virus infection in asymptomatic patients with hemophilia. *Blood* **78** 900–906.

MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression (with discussion). *J. Amer. Statist. Assoc.* **83** 1023–1036.

TSODIKOV, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics* **54** 1508–1516.

VOLBERDING, P. A., LAGAKOS, S. W., KOCH, M. A., PETTINELLI, C., MYERS, M. W., BOOTH, D. K., BALFOUR, H. H., REICHMAN, R. C., BARTLETT, J. A., HIRSCH, M. S., MURPHY, R. L., HARDY, D., SOEIRO, R., FISCHL, M. A., BARTLETT, J. G., MERIGAN, T. C., HYSLOP, N. E., RICHMAN, D. D., VALENTINE, F. T., COREY, L. and the AIDS CLINICAL TRIALS GROUP OF THE NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES (1990). Zidovudine in asymptomatic human immunodeficiency virus infection. *New England J. Medicine* **322** 941–949.

ZELLNER, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Studies in Bayesian Econometrics and Statistics* (P. K. Goel and A. Zellner, eds.) 233–243. North-Holland, Amsterdam.