The Scale Transformed Power Prior

with Applications to Studies with Different Endpoints

Joseph G. Ibrahim

Joint work with Brady Nifong and Matthew A. Psioda

UNC
GILLINGS SCHOOL OF
GLOBAL PUBLIC HEALTH

Introduction

- No two studies will ever be identical (Ibrahim et al. 2001)
- However, even if studies are not identical, historical data can provide useful information to a current study
- 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 (Ibrahim and Chen 2000, Ibrahim et al. 2015)

Practical Motivation

- Motivation: settings in which historical data and the current data involve outcomes with different distributions
  - Ex 1: comprehensive post-acute stroke services study (COMPASS)
  - Ex 2: Eastern Cooperative Oncology Group (ECOG) phase III clinical trials for melanoma
  - Ex 3: ECOG phase II clinical trials for liver cancer

Ex 1: COMPASS Study

- The COMPASS study (Duncan et al. 2017) was conducted at 41 hospitals (40 randomized units) in North Carolina.
- Interested in the effect of the COMPASS care model on the primary endpoint as well as several secondary endpoints
- The study includes part I and part II data with several correlated secondary endpoints.
    - Binary fall - measured whether the participant had fallen since hospital discharge or 90 days post-stroke
    - Continuous patient-recorded outcomes measurement information system (PROMIS) score for physical function
    - The PROMIS physical function score is only measured in part II

Ex 2: ECOG Melanoma Data

- ECOG phase III clinical trials, E2696 and E1694, focused on a post-operative chemotherapy, Interferon Alpha-2b (IFN) in comparison to a vaccine (GMK).
- The historical study, E2696, contains the primary, continuous, endpoint for immunoglobulin M (IgM) antibody levels at day 28, with the presence of such antibodies being shown to improve disease-free survival (Kirkwood et al. 2001).
- The current study, E1694, contains the a primary endpoint of survival time as described in Brown and Ibrahim (2003).
- The historical and current data sets contain the same covariates, such as age, treatment, and cancer stage.

Ex 3: ECOG Liver Data

- E2282 and E1286 phase II ECOG studies to to evaluate new treatments in patients with liver cancer (Lipsitz and Ibrahim 1996)
- The primary interest here is how the outcome, survival time from entry on the study until death, differs with respect to five dichotomous baseline covariates (Ibrahim et al. 1999).
- The historical data set E2282 contains the response variable nodes, which is a count variable for the number of nodes an individual patient has
- The current data set E1286, contains the response variable survival time

Use of Historical Data

- Two common ways to incorporate historical data are through the use of the power prior or the commensurate prior
- Let $\boldsymbol{\theta}$ be the parameter of interest
- Let $D_0 = \{\boldsymbol{y_0}, \boldsymbol{X}_0, n_0\}$ denote the historical data

Power Prior

The Power Prior (PP) developed by Ibrahim and Chen (2000) is as follows:

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

Where:

- $L(\boldsymbol{\theta}|D_0)$ is the historical data likelihood
- $0 \leq a_0 \leq 1$ weights the historical data relative to the likelihood of the current study
- $\pi_0(\boldsymbol{\theta})$ is the initial prior

Power Prior Variants - Fixed a_0

Partial Borrowing Power Prior: (Chen et al. 2011)

$$\pi(\boldsymbol{\theta}|D_0, a_0) \propto \left[\int L(\boldsymbol{\theta}|D_0, \boldsymbol{\xi})^{a_0} g(\boldsymbol{\xi}) d\boldsymbol{\xi}\right] \pi_0(\boldsymbol{\theta}),$$

where $\boldsymbol{\xi}$ is a vector of latent variables or nuisance parameters in the model, and $g(\boldsymbol{\xi})$ is the distribution of the latent (nuisance) variables (parameters)

Asymptotic Power Prior:

$$\pi(\boldsymbol{\theta}|D_0) \propto \exp\left\{\frac{-a_0}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0)^T I_0(\hat{\boldsymbol{\theta}}_0)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0)\right\},$$

where $\hat{\boldsymbol{\theta}}_0$ is the historical data maximum likelihood estimate and $I_0(\cdot)$ is the historical data likelihood information matrix

Commensurate Prior

The commensurate prior developed by Hobbs et al. (2011) utilizes historical information when there is strong evidence of agreement between the historical and current data.

- Unlike the power prior and it's variants, the commensurate prior assumes that the historical and current data parameters differ, denoted $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ respectively.
- This allows the use of a parameter that measures the agreement between $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, denoted τ
- Construct a prior for $\boldsymbol{\theta}$ that is normally distributed around $\boldsymbol{\eta}$ with a precision of τ, the commensurate parameter
    - Offers a clear interpretation of commensurablility between $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$

Commensurate Prior

The Commensurate Prior is then written as

$$\pi_c(\boldsymbol{\theta}|D_0, \boldsymbol{\eta}, \tau) \propto L(\boldsymbol{\eta}|D_0)\pi(\boldsymbol{\theta}|\boldsymbol{\eta}, \tau)\pi_0(\boldsymbol{\theta}), \tag{2}$$

where $\pi_0(\boldsymbol{\theta})$ is the initial prior for $\boldsymbol{\theta}$, $L(\boldsymbol{\theta} \mid D_0)$ is the historical data likelihood and $\pi(\boldsymbol{\theta}|\boldsymbol{\eta}, \tau)$ denotes the p dimensional normal density with mean $\boldsymbol{\eta}$ and covariance matrix $\tau^{-1}\boldsymbol{I}_p$, in which $\boldsymbol{I}_p$ is the $p \times p$ identity matrix.

- as $\tau \longrightarrow 0$, the historical data is effectively ignored
- as $\tau \longrightarrow \infty$, $\boldsymbol{\theta} \longrightarrow \boldsymbol{\eta}$, and the data is essentially pooled

The Scale Transformed Power Prior for Generalized Linear Models

Formulation

- We desire that the scale of $\boldsymbol{\theta}$ based on the historical data = scale of $\boldsymbol{\theta}$ based on the current data.
- We know that $I^{-1}(\boldsymbol{\theta})$ is the asymptotic covariance matrix of the MLE of $\boldsymbol{\theta}$.
- Let $\boldsymbol{\eta}$ and $I_0(\boldsymbol{\eta})$ denote the historical data parameter and Fisher information matrix
- Let $\boldsymbol{\theta}$ and $I_1(\boldsymbol{\theta})$ denote the current data parameter and Fisher information matrix.
- The corresponding standardized or "scaled" parameters are

$$I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta} \quad \text{and} \quad I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta}$$

Methodological Motivation

- Thus, we desire
$$I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta} = I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta}. \tag{3}$$
- In order to solve Equation (3) algebraically, one would have to solve for $\boldsymbol{\theta}$ (or $\boldsymbol{\eta}$), which is difficult unless the historical (current) information matrix does not depend on $\boldsymbol{\theta}$, such as when the historical data model is a normal linear model.
- We denote the transformation in general as $\boldsymbol{\eta} = g(\boldsymbol{\theta})$, and thus the scale transformed power prior is a function of $g(\boldsymbol{\theta})$.
- As a final step, in the presence of covariates, we compute I_1 based on the historical data covariates to avoid double use of the current data covariates.

Methodological Motivation - Special Case

- Under the case in which the historical data model is a linear model, we can solve algebraically for $\boldsymbol{\eta}$ as

$$\begin{aligned} \boldsymbol{\eta} &= I_0^{-1/2} I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta} \\ &\equiv A(\boldsymbol{\theta})\boldsymbol{\theta} \\ &\equiv g(\boldsymbol{\theta}), \end{aligned}$$

  where $A(\boldsymbol{\theta}) = I_0^{-1/2} I_1^{1/2}(\boldsymbol{\theta})$.
- Note, in the case where the current data follow a linear model, one can perform the analysis using the straPP by sampling from the posterior distribution for $\boldsymbol{\eta}$ and use the transformation $\boldsymbol{\theta} = A^{-1}(\boldsymbol{\eta})\boldsymbol{\eta} = g^{-1}(\boldsymbol{\eta})$ to obtain samples from the posterior distribution for $\boldsymbol{\theta}$.

Formulation

The Scale Transformed Power Prior (straPP) can be derived from the power prior in Equation (1) using the transformation $\boldsymbol{\eta} = g(\boldsymbol{\theta})$, as

$$\pi_s(\boldsymbol{\theta}|D_0) \propto L(\boldsymbol{g(\theta)}|D_0)^{a_0}\pi_0(\boldsymbol{g(\theta)}) \left|\frac{\boldsymbol{dg(\theta)}}{\boldsymbol{d\theta}}\right|, \tag{4}$$

where $|dg(\boldsymbol{\theta})/d\boldsymbol{\theta}|$ is the determinant of the Jacobian of the transformation.

The straPP reduces to the power prior in the case $g(\boldsymbol{\theta}) = \boldsymbol{\theta}$, that is, $A(\boldsymbol{\theta})$ is the identity matrix.

When $a_0 = 0$, we set the straPP equal to the initial prior

Expression for the Jacobian

In general, the transformation implied by (3) cannot be calculated algebraically. However, the Jacobian can be calculated via the chain rule and implicit differentiation as shown below.

$$
\begin{aligned}
\left\{\left[\frac{d}{d\boldsymbol{\eta}} I_0^{1/2}(\boldsymbol{\eta})\right]\boldsymbol{\eta} + I_0^{1/2}(\boldsymbol{\eta})\right\}\frac{d\boldsymbol{\eta}}{d\boldsymbol{\theta}} &= \left[\frac{d}{d\boldsymbol{\theta}} I_1^{1/2}(\boldsymbol{\theta})\right]\boldsymbol{\theta} + I_1^{1/2}(\boldsymbol{\theta}) \\
\Rightarrow \frac{d\boldsymbol{\eta}}{d\boldsymbol{\theta}} &= \left\{\left[\frac{d}{d\boldsymbol{\eta}} I_0^{1/2}(\boldsymbol{\eta})\right]\boldsymbol{\eta} + I_0^{1/2}(\boldsymbol{\eta})\right\}^{-1} \\
&\quad \times \left\{\left[\frac{d}{d\boldsymbol{\theta}} I_1^{1/2}(\boldsymbol{\theta})\right]\boldsymbol{\theta} + I_1^{1/2}(\boldsymbol{\theta})\right\}
\end{aligned}
$$

The derivative of $I_1^{1/2}(\boldsymbol{\theta})$ can be written as

$$
\frac{dI_1^{1/2}(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \left(\frac{dI_1^{1/2}(\boldsymbol{\theta})}{d\theta_0}, \ldots, \frac{dI_1^{1/2}(\boldsymbol{\theta})}{d\theta_{p-1}}\right).
$$

Expression for the Jacobian

For $j = 0, \dots, p-1$, the derivative can be decomposed using a direct application of the product rule as

$$dI_1(\boldsymbol{\theta})/d\theta_j \quad = \quad I_1^{1/2}(\boldsymbol{\theta})[dI_1^{1/2}(\boldsymbol{\theta})/d\theta_j] + [dI_1^{1/2}(\boldsymbol{\theta})/d\theta_j]I_1^{1/2}(\boldsymbol{\theta}). \quad (5)$$

Equation (5) can be expressed in the form of the Sylvester equation (Sylvester (1884)) . Let $\boldsymbol{I}_p$ denote the $p \times p$ identity matrix. Then, following Laub (2005), the required derivatives may be represented as follows:

$$\text{vec}\left(\frac{dI_1^{1/2}(\boldsymbol{\theta})}{d\theta_j}\right) \quad = \quad \left(I_1^{1/2}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_p + \boldsymbol{I}_p \otimes I_1^{1/2}(\boldsymbol{\theta})\right)^{-1} \text{vec}\left(\frac{dI_1(\boldsymbol{\theta})}{d\theta_j}\right),$$

where vec$(\cdot)$ denotes the vectorization of a matrix, in which columns are stacked to convert a $n \times p$ matrix into a $np \times 1$ vector. The derivative of $I_0^{1/2}(\boldsymbol{\eta})$ is calculated analogously.

Simple Normal IID Example

- Suppose the historical data is given by $y_{0i} \sim N(\eta, \sigma_0^2)$, are iid for $i = 1, \ldots, n_0$ and the current data is $y_i \sim N(\theta, \sigma_1^2)$ are iid for $i = 1, \ldots, n_1$.
- The Fisher information based on the historical data is n_0/σ_0^2 and the Fisher information from the current data likelihood, evaluated at the historical covariate vector is n_0/σ_1^2.

Simple Normal IID Example

- Then, applying the transformation for this example, we obtain

$$A(\theta) = \frac{\sigma_0}{\sigma_1},$$

so that the scaled transformed parameter in going from historical to current data is

$$\eta = \left(\frac{\sigma_0}{\sigma_1}\right)\theta.$$

- We see that η (the scaled θ) is the ratio of the standard deviations based on the historical and current data multiplied by θ.
- Moreover, if $\sigma_0 = \sigma_1$ in this example, then $A(\theta) = 1$, and the transformation is the identity transformation as expected.

Partial-Borrowing straPP

- Let $\boldsymbol{\theta}_{p\times 1}$ be partitioned into two independent vectors such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.
  - $\boldsymbol{\theta}_1$ is $r \times 1$, and $\boldsymbol{\theta}_2$ is $(p-r) \times 1$
- Similarly, let $\boldsymbol{\eta}_{s\times 1}$ be partitioned into two independent vectors such that $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$.
  - $\boldsymbol{\eta}_1$ is $r \times 1$, and $\boldsymbol{\eta}_2$ is $(s-r) \times 1$
- The partial-borrowing straPP is then derived according to the transformation

$$I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta}_1 = I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta}_1, \tag{6}$$

where here $I_0(\boldsymbol{\eta})$ and $I_1(\boldsymbol{\theta})$ denote the $r \times r$ submatrices of the Fisher information matrices for the current and historical data models, respectively.

Partial-Borrowing straPP

- In general, $\boldsymbol{\eta}_1$ can not be solved for in Equation (6), so we will refer to the transformation as $g_1(\boldsymbol{\theta}, \boldsymbol{\eta}_2)$
- Suppose we would like to use straPP on $\boldsymbol{\theta}_1$ and some other arbitrary prior, call this $\pi_0(\boldsymbol{\theta}_2)$, on $\boldsymbol{\theta}_2$. Then we can write the partial borrowing straPP as

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\eta}_2 \mid D_0) &= \pi_s(\boldsymbol{\theta}_1, \boldsymbol{\eta}_2 \mid \boldsymbol{\theta}_2, D_0,)\pi_0(\boldsymbol{\theta}_2) \\ &\propto \mathcal{L}(g_1(\boldsymbol{\theta}, \boldsymbol{\eta}_2), \boldsymbol{\eta}_2 \mid \boldsymbol{\theta}_2, D_0)^{a_0} \pi_0(g_1(\boldsymbol{\theta}, \boldsymbol{\eta}_2), \boldsymbol{\eta}_2 \mid \boldsymbol{\theta}_2) \\ &\qquad \times \left| \frac{d g_1(\boldsymbol{\theta}, \boldsymbol{\eta}_2)}{d \boldsymbol{\theta}_1} \right| \pi_0(\boldsymbol{\theta}_2). \end{aligned} \tag{7}$$

straPP for Generalized Linear Models

- Assume the historical and current data sets are generalized linear models (GLMs)
- Let $\boldsymbol{y}_{0i}$ denote the i^{th} response for the historical data set and $\boldsymbol{y}_{1j}$ denotes the j^{th} response for the current data set
- Denote the i^{th} row of $\boldsymbol{X}_0$ as $\boldsymbol{x}_{0i}^T$, and let $\boldsymbol{x}_{1j}^T$ denote the j^{th} row of X_1

straPP for Generalized Linear Models

- Let $k = 0, 1$ be the index for the historical and current data, respectively.
- Let $\boldsymbol{\xi}_k = (\boldsymbol{\beta}, \phi_k)$, where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})'$ denotes the $p \times 1$ vector of regression parameters, and ϕ_k denotes the scalar dispersion parameter. For $i = 1, \ldots, n_k$, an individual response is distributed as follows:

$$f(y_{ki}|\boldsymbol{\xi}_k) = \exp\{\phi_k[y_{ki}h_k(\boldsymbol{x}_{ki}^T\boldsymbol{\beta}) - b_k(h_k(\boldsymbol{x}_{ki}^T\boldsymbol{\beta})) - c_k(y_{ki})] - \frac{1}{2}s_k(y_{ki}, \phi_k)\},$$

  where $h_k(\cdot)$ is the link function, and $b_k(\cdot)$, $c_k(\cdot)$, and $s_k(\cdot)$ are some known functions for data D_k, $k = 0, 1$.
- ϕ_0, ϕ_1 are scalar dispersion parameters.

Comparison to the Power Prior

Suppose the historical and current data arise from linear regression model with known variances, referred to as the normal-normal case. Then,

- $y_{0i} \sim N(\boldsymbol{x}_{0i}^T \boldsymbol{\eta}, \sigma_0^2 I_{n_0})$, i = 1,...,$n_0 \Rightarrow I_0(\boldsymbol{\eta}) = \frac{1}{\sigma_0^2} \boldsymbol{X}_0' \boldsymbol{X}_0$, and
- $y_{1i} \sim N(\boldsymbol{x}_{1i}^T \boldsymbol{\beta}, \sigma_1^2 I_{n_1})$, i = 1,...,$n_1 \Rightarrow I_1(\boldsymbol{\beta} | \boldsymbol{X}_0) = \frac{1}{\sigma_1^2} \boldsymbol{X}_0' \boldsymbol{X}_0$.

Further suppose the initial prior is uniform improper.

Then, under this normal-normal case, our scale-transformed parameter is

$$\boldsymbol{\eta} = g(\boldsymbol{\beta}) = A(\boldsymbol{\beta})\boldsymbol{\beta} = \frac{\sigma_0}{\sigma_1} \boldsymbol{\beta}.$$

Comparison *cont...*

Power Prior (PP):

$$\boldsymbol{\beta} \sim N\left((\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{Y}_0, \;\; \left(\frac{\sigma_0^2}{a_0}\right)(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1} \right)$$

straPP:

$$\begin{aligned} \boldsymbol{\beta} &\sim N\left(\frac{\sigma_1}{\sigma_0}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{Y}_0, \;\; \left(\frac{\sigma_0^2}{a_0}\right)\left(\frac{\sigma_1^2}{\sigma_0^2}\right)(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1} \right) \\ &\sim N\left(\frac{\sigma_1}{\sigma_0}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{Y}_0, \;\; \left(\frac{\sigma_1^2}{a_0}\right)(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1} \right) \end{aligned}$$

Properties of straPP

Theorem

Let β_j denote the (j+1)st element of $\boldsymbol{\beta}$, j $=$ 0,. . . , p-1.
Further let $\hat{\beta}_{sj}$ denote the posterior mean of β_j under the straPP and $\hat{\beta}_{ppj}$ denote the posterior mean of β_j under the PP.
Then, for the normal-normal case, where $g(\boldsymbol{\beta}) = \frac{\sigma_0}{\sigma_1}\boldsymbol{\beta}$, the straPP has a lower MSE than the power prior under the following condition:

$$\frac{\mathit{Var}(\hat{\beta}_{sj}) - \mathit{Var}(\hat{\beta}_{ppj})}{\left[\mathit{Percent\ Bias}(\hat{\beta}_{ppj})\right]^2} < \beta_j^2$$

In general, the Percent Bias of $\hat{\beta}_{ppj}$ depends on β_j.

Simulation Setup

Using this normal-normal set up, we assume our data has an intercept and a treatment indicator (p=2).
The following simulations use the parameter values:

- $a_0 = 0.5$
- $n_0 = 50;\ n_1 = 100$
- $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$
- $\beta_0 = 1;\ \beta_1 \in \{0, 1.8, \text{by} = 0.1\}$
- $\boldsymbol{\eta} = g(\boldsymbol{\beta}) = \frac{\sigma_0}{\sigma_1}\boldsymbol{\beta}$

Along with the straPP and PP, we also used a uniform improper prior (UIP). We generated 5,000 data sets and used Metropolis-Hastings MCMC for each model with a Monte Carlo sample size of 25,000 and a burn-in of 2,000.

In this simulation, the threshold value from the theorem can be calculated exactly given $(\sigma_0, \sigma_1, a_0, n_0, n_1)$:

$$\beta_1^* = \sqrt{\frac{\sigma_{s1}^2 \left[\frac{n_1 + a_0^2 n_0}{n_1 + a_0 n_0}\right] - \sigma_{pp1}^2 \left[\frac{n_1/\sigma_1^2 + a_0^2 n_0/\sigma_0^2}{n_1/\sigma_1^2 + a_0 n_0/\sigma_0^2}\right]}{\left(\frac{n_1}{\sigma_1^2} - \frac{a_0 n_0}{\sigma_0^2}\right)^{-2} \left(\frac{a_0 n_0}{\sigma_0}\right)^2 \left(\frac{1}{\sigma_1} - \frac{1}{\sigma_0}\right)^2}}$$

Where $\sigma_{s1}^2 = \frac{4\sigma_1^2}{n_1 + a_0 n_0}$ is the posterior variance of β_1 from the straPP and $\sigma_{pp1}^2 = 4\left(\frac{n_1}{\sigma_1^2} + \frac{a_0 n_0}{\sigma_0^2}\right)^{-1}$ is the posterior variance of β_1 from the power prior.

Figure 1: $\sigma_0 = 3 > \sigma_1 = 1$

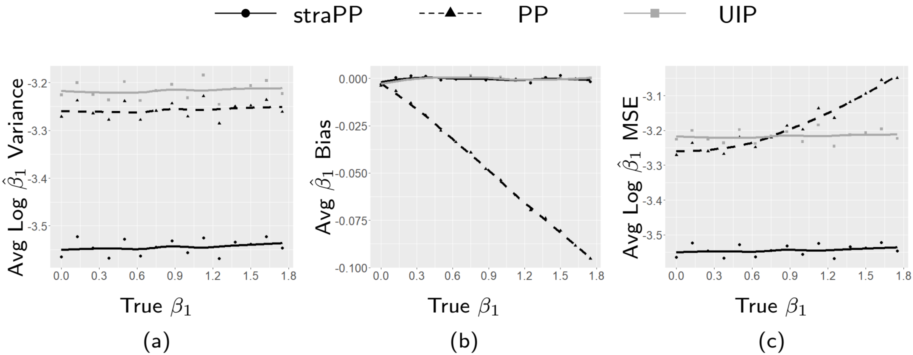
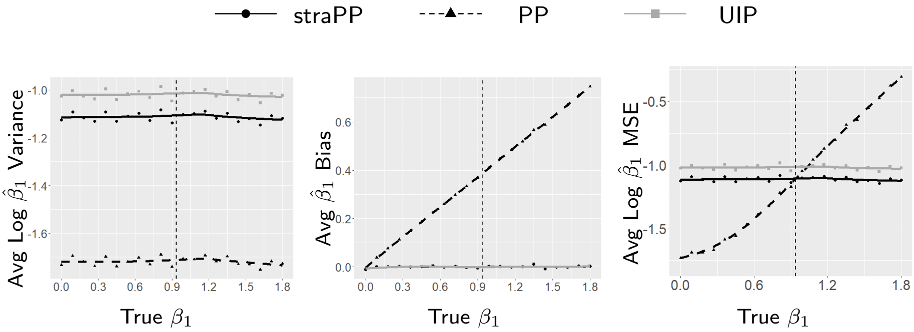


Figure 2: $\sigma_0 = 1 < \sigma_1 = 3$



(a) (b) (c)

Simulation Summary

- The posterior mean from the straPP is always unbiased in the normal-normal case.
- When $\sigma_0 > \sigma_1$, the MSE for straPP is always lower than PP.
- When $\sigma_0 < \sigma_1$, there is a trade off between the bias and variance that determines when the MSE is lower for the straPP.
- Changing the input parameter values shifts the magnitude of the graphs, but the pattern stays the same.

Binary-Normal Setup

Suppose both the historical and current data have and intercept and a treatment indicator with a 1:1 allocation, and are distributed

- $y_{0i} \sim Bern\left(p_{0i} = \frac{1}{1+\exp\{-\boldsymbol{x}'_{0i}\boldsymbol{\eta}\}}\right)$, i = 1,. . . , n_0
- $y_{1i} \sim N(\boldsymbol{x}'_{1i}\boldsymbol{\beta}, \sigma_1^2)$, i = 1,. . . ,$n_1$
    - σ_1 is known

Under this case, the historical data information matrix depends on the regression parameter, so we sample the complementary posterior to obtain samples for $\boldsymbol{\eta}$ and then transform them to obtain samples for $\boldsymbol{\beta}$.

$$\boldsymbol{\beta} = g^{-1}(\boldsymbol{\eta}) = A^{-1}(\boldsymbol{\eta})\boldsymbol{\eta} = [I_1^{-1/2} I_0^{1/2}(\boldsymbol{\eta})]\boldsymbol{\eta}$$

Priors

For the binary-normal simulation, we also consider the *asymptotic power prior*.

1. Uniform Improper: $\pi(\boldsymbol{\beta}|D_0) \propto 1$
2. Power Prior: $\pi(\boldsymbol{\beta}|D_0) \propto L(\boldsymbol{\beta}|D_0)^{a_0}$
3. complementary straPP: $\pi(\boldsymbol{\eta}|D_0) \propto L(\boldsymbol{\eta}|D_0)^{a_0} \left| \frac{dg^{-1}(\boldsymbol{\eta})}{d\boldsymbol{\eta}} \right|$
4. Asymptotic Power Prior: $\pi(\boldsymbol{\beta}|D_0) \propto \exp\{\frac{-a_0}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T I_0(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\}$

Developing $g^{-1}(\boldsymbol{\eta})$

For the binary-normal simulation, $g^{-1}(\boldsymbol{\eta}) = A^{-1}(\boldsymbol{\eta})\boldsymbol{\eta}$ is not simple:

$$g^{-1}(\boldsymbol{\eta}) = \frac{1}{\sigma_1}\begin{pmatrix} n_0 & n_0/2 \\ n_0/2 & n_0/2 \end{pmatrix}^{-1/2} \times \sqrt{\frac{2}{n_0}}\begin{pmatrix} p_{t0}(1-p_{t0}) + p_{t1}(1-p_{t1}) & p_{t1}(1-p_{t1}) \\ p_{t1}(1-p_{t1}) & p_{t1}(1-p_{t1}) \end{pmatrix}^{1/2} \boldsymbol{\eta}$$

in which:

- $p_{t0} = P(y_0 = 1|trt\ 0) = \frac{\exp(\eta_0)}{1+\exp(\eta_0)}$
- $p_{t1} = P(y_0 = 1|trt\ 1) = \frac{\exp(\eta_0+\eta_1)}{1+\exp(\eta_0+\eta_1)}$

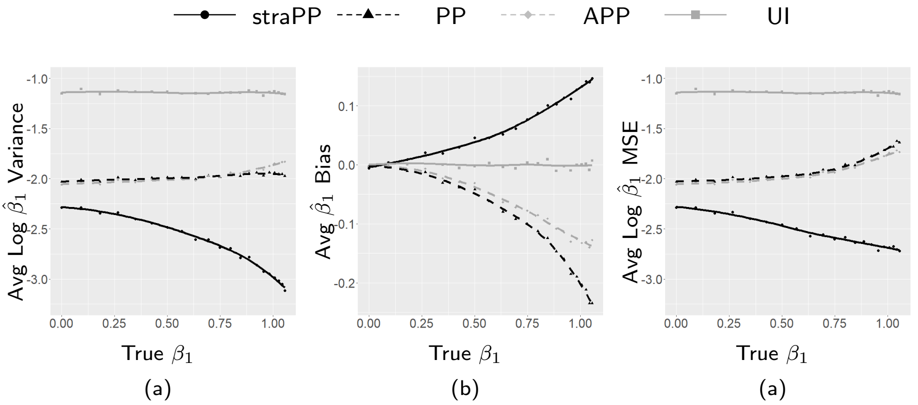
Simulation Setup

The following simulation uses the parameter values:

- $a_0 = 0.5$
- $n_0 = 100;\ n_1 = 50$
- $\sigma_1 = 2$
- $\eta_0 = 0.5$; $\eta_1 \in \{0, 2.0, \text{by} = 0.05\}$
- $\boldsymbol{\beta} = g^{-1}(\boldsymbol{\eta})$

Once again, we generated 5,000 data sets and used Metropolis-Hastings MCMC for each model with a Monte Carlo sample size of 25,000 and a burn-in of 2,000.

Figure 3


straPP
PP
APP
UI
Avg Log $\hat{\beta}_1$ Variance
-1.0
-1.5
-2.0
-2.5
-3.0
0.00
0.25
0.50
0.75
1.00
True β_1
(a)
Avg $\hat{\beta}_1$ Bias
0.1
0.0
-0.1
-0.2
0.00
0.25
0.50
0.75
1.00
True β_1
(b)
Avg Log $\hat{\beta}_1$ MSE
-1.0
-1.5
-2.0
-2.5
-3.0
0.00
0.25
0.50
0.75
1.00
True β_1
(a)

Simulation Summary

- The posterior mean based on the straPP has the smallest variance on average
- The posterior mean of the straPP no longer retains the unbiasedness property of the normal-normal case, but still has bias less than or the power prior in absolute value and less than or equal to the asymptotic power prior in absolute value
- The posterior mean based on the straPP has the lowest MSE compared to the other analyzed priors
- Changing the input parameter values shifts the magnitude of the graphs, but the pattern stays the same.

Generalized Scale Transformed Power Prior

- The straPP is derived under the assumption that the standardized parameter values for the historical and current data models are equal.
- This may be a reasonable transformation for rescaling the parameter in a power prior in settings where the historical data is not of the same data type as that of the current data.
- Nonetheless, this core assumption of the straPP may be violated in some cases, in which both a location change and a scale change are needed in the parameter.
- Thus, it is useful to develop a generalization of the straPP that provides a degree of robustness when the assumption of equal standardized parameter values does not hold. We call this generalization the *generalized scale transformed power prior (Gen-straPP)*

Gen-straPP Formulation

Here we specify

$$I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta} = I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta} + \boldsymbol{c}_{\mathbf{0}}, \tag{8}$$

where $\boldsymbol{c}_0$ is a $p \times 1$ vector that allows component-specific deviations from the assumption of equal standardized parameter values for $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$.

- We denote the transformation as $\boldsymbol{\eta} \equiv g_{\boldsymbol{c}_0}(\boldsymbol{\theta})$.
- $\boldsymbol{c}_0 = \mathbf{0}$ corresponds to the straPP.

Details on $\boldsymbol{c}_0$

We can take $\boldsymbol{c}_0$ to be fixed or random.

- Fixed: $\boldsymbol{c}_0 = d_0\boldsymbol{J}$, where $\boldsymbol{J}$ is a vector of 1's and d_0 is some scalar
- Random: take $\boldsymbol{c}_0 \sim N(0, \omega_0 I)$, where ω_0 is a specified positive scalar

There are innumerable choices for fixed c0 and this setting is best equipped to identify violations in the assumptions of the straPP when the violations occur across all covariates, but poorly equipped to identify violations which present in only one covariate.

Therefore, to accommodate various combinations of violations, we suggest taking c0 to be a random vector.

Then, the Gen-straPP can be derived from the power prior in (1) using the transformation in (8). The Gen-straPP is

$$
\begin{aligned}
\pi_g(\boldsymbol{\theta}, \boldsymbol{c}_0 \mid D_0) &= \pi_s(\boldsymbol{\theta} \mid \boldsymbol{c}_0, D_0)\pi_0(\boldsymbol{c}_0) \\
&\propto \left[\mathcal{L}(g_{\boldsymbol{c}_0}(\boldsymbol{\theta}) \mid \boldsymbol{c}_0, D_0)^{a_0} \pi_0(g_{\boldsymbol{c}_0}(\boldsymbol{\theta}) \mid \boldsymbol{c}_0) \left| \frac{dg(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right| \right] \pi_0(\boldsymbol{c}_0),
\end{aligned} \tag{9}
$$

where $\pi_0(\boldsymbol{c}_0)$ denotes the prior for $\boldsymbol{c}_0$.

Similarly to the partial-borrowing straPP, we can develop the partial borrowing Gen-straPP from the transformation

$$I_0^{-1/2}(\boldsymbol{\eta})\boldsymbol{\eta}_1 = I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta}_1 + \boldsymbol{c}_0, \tag{10}$$

where here $\boldsymbol{c}_0$ is an $r \times 1$ vector. We denote the transformation implied by (10) as $g_{1,\boldsymbol{c}_0}(\boldsymbol{\theta}, \boldsymbol{\eta}_2)$. The partial-borrowing Gen-straPP is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\eta}_2, \boldsymbol{c}_0 \mid D_0) = \pi_s(\boldsymbol{\theta}_1, \boldsymbol{\eta}_2 \mid \boldsymbol{c}_0, \boldsymbol{\theta}_2, D_0,)\pi_0(\boldsymbol{c}_0)\pi_0(\boldsymbol{\theta}_2), \tag{11}$$

where $\pi_s(\boldsymbol{\theta}_1, \boldsymbol{\eta}_2 \mid \boldsymbol{c}_0, \boldsymbol{\theta}_2, D_0,)$ is proportional to the expression

$$\mathcal{L}(g_{1,\boldsymbol{c}_0}(\boldsymbol{\theta}, \boldsymbol{\eta}_2), \boldsymbol{\eta}_2 \mid \boldsymbol{c}_0, \boldsymbol{\theta}_2, D_0)^{a_0}\pi_0(g_{1,\boldsymbol{c}_0}(\boldsymbol{\theta}, \boldsymbol{\eta}_2), \boldsymbol{\eta}_2 \mid \boldsymbol{c}_0, \boldsymbol{\theta}_2)\left|\frac{dg_1(\boldsymbol{\theta}, \boldsymbol{\eta}_2)}{d\boldsymbol{\theta}_1}\right|.$$

To derive the relationship between the Gen-straPP and the commensurate prior, we derive the Gen-straPP in an alternate manner. The Gen-straPP transformation in (8), can be re-written as

$$I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta} = I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta} - \boldsymbol{c_0}. \tag{12}$$

When $\boldsymbol{c_0} \sim N_p(\mathbf{0}, \omega_0 \boldsymbol{I}_p)$, the standardized current parameter, $\boldsymbol{\theta}^* = I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta}$, is distributed normally about the standardized historical parameter as

$$\boldsymbol{\theta}^* \mid \boldsymbol{\eta} \sim N_p(I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta},\ \omega_0 \boldsymbol{I}_p).$$

Prior for $(\boldsymbol{\theta}^*, \boldsymbol{\eta})$

To complete the specification of the joint prior for the historical and standardized current parameters, we specify a power prior on $\boldsymbol{\eta}$. The joint prior is

$$\pi(\boldsymbol{\theta}^*, \boldsymbol{\eta}) \propto N_p(\theta^* | I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta},\ \omega_0 \boldsymbol{I}_p) \mathcal{L}(\boldsymbol{\eta}|D_0)^{a_0} \pi_0(\boldsymbol{\eta}).$$

To derive the Gen-straPP, we must calculate the joint distribution of the untransformed current and historical parameters. Let

$$\xi = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} h(\boldsymbol{\theta}^*) \\ \boldsymbol{\eta} \end{pmatrix},$$

then, when $a_0 = 1$ we can write the Gen-straPP as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\eta}) \propto \mathcal{L}(\boldsymbol{\eta}|D_0) N_p(I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta} \mid I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta},\ \omega_0 \boldsymbol{I}_p) \pi_0(\boldsymbol{\eta}) \left| \frac{dh(\boldsymbol{\theta}^*)}{d\boldsymbol{\theta}^*} \right|. \quad (13)$$

This can be thought of as the commensurate prior from (2) in which the standardized current parameter is normally distributed about the standardized historical parameter when the commensurate parameter is equal to the inverse of ω_0.

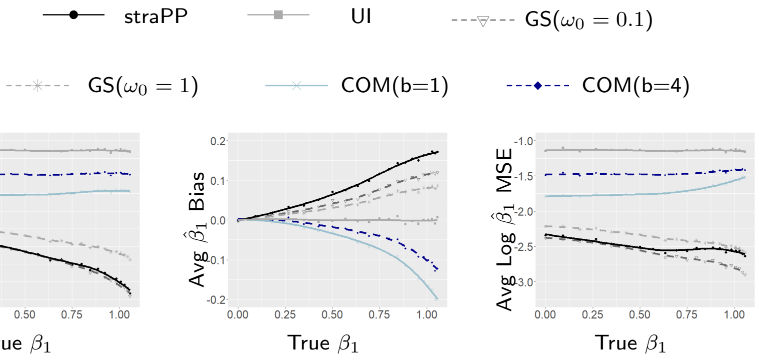
Additional Binary-Normal Simulations

Using the same parameter values from the earlier binary-normal simulations, except with $a_0 = 1.0$, we compare the straPP to the Gen-straPP, and the Gen-straPP to the commensurate prior.

- Using $\omega_0 = 0.1, 0.5, 1.0$
- Using $\tau \sim Gam(2, b)$ for the commensurate parameter of the commensurate prior, with $b = 1, 4$

Once again, we generated 5,000 data sets and used Metropolis-Hastings MCMC for each model with a Monte Carlo sample size of 25,000 and a burn-in of 2,000.

Figure 4


straPP
UI
GS($\omega_0 = 0.1$)
GS($\omega_0 = 1$)
COM(b=1)
COM(b=4)
Avg Log $\hat{\beta}_1$ Variance
-1.0
-1.5
-2.0
-2.5
-3.0
0.00 0.25 0.50 0.75 1.00
True β_1
Avg $\hat{\beta}_1$ Bias
0.2
0.1
0.0
-0.1
-0.2
0.00 0.25 0.50 0.75 1.00
True β_1
Avg Log $\hat{\beta}_1$ MSE
-1.0
-1.5
-2.0
-2.5
-3.0
0.00 0.25 0.50 0.75 1.00
True β_1

(a) (b) (a)

Simulation Summary

- The posterior mean based on the straPP has the smallest variance on average
- The posterior mean based on the Gen-straPP has smaller bias than the straPP and similar bias to the commensurate prior in absolute value
- The posterior mean based on the straPP has the lowest MSE compared to the other analyzed priors, until the true $\beta_1 > 0.6$, in which case the Gen-straPP with $\omega_0 = 0.1$ has the lowest MSE

Real Data Example: COMPASS Study

- We had access to the data from hospital site 30, which consisted of part I (historical) and part II (current) data from the trial.
- We removed observations with missing values of the covariates of interest for complete case analysis ($n_0 = 244$, $n_1 = 385$).

Covariates and Response Variables

- Covariates of interest were:
  - Indicator for intervention: electronic care plan (eCare plan)
  - Indicator for history of stroke or transient ischemic attack (TIA)
  - Categorized NIH stroke scale score (NIHSS): 0 = no stroke symptoms, 1-4 = minor stroke symptoms, $\geq$ 5 moderate-to-severe stroke symptoms.
  - Indicator for non white race
- Historical response: binary fall - whether the participant had fallen since hospital discharge and 90 days post-stroke
- Current response: continuous PROMIS score for physical function

Historical and Current Data Distributions

We assume the historical patient outcomes are independently distributed according to a logistic regression model and the current patient outcomes are independently distributed according to a linear regression model such that

- $y_{0i} \sim Bern(p_i)$, $i = 1, \ldots, 244$
  - $p_i = 1/(1 + \exp\{-\boldsymbol{x}_{0i}^T \boldsymbol{\eta}\})$
  - $\boldsymbol{\eta}^T = (\eta_0, \boldsymbol{\eta}_1^T)$
- $y_{1i} \sim N(\boldsymbol{x}_{1i}^T \boldsymbol{\beta}, \sigma_1^2)$, $i = 1, \ldots, 385$
  - $\boldsymbol{\beta}^T = (\beta_0, \boldsymbol{\beta}_1^T)$
  - $\phi = 1/\sigma_1^2$ is an unknown precision parameter.
- $\boldsymbol{\eta}_1$ and $\boldsymbol{\beta}_1$ denote the historical and current covariate regression parameters.

Gen-strapp and straPP Setup

- We chose not to borrow from the historical data intercept in each model tested
- We specify the partial-borrowing straPP and Gen-straPP for the covariates, and specify a uniform improper prior for the intercept.
- Similar to the binary-normal simulation, we sample the complementary posterior to obtain samples for $\boldsymbol{\eta}_1$ and then use the corresponding transformation to obtain samples for $\boldsymbol{\beta}_1$
  - straPP transformation:

$$\boldsymbol{\beta}_1 = g_1^{-1}(\boldsymbol{\eta}) = [I_1^{-1/2} I_0^{1/2}(\boldsymbol{\eta})]\boldsymbol{\eta}_1$$

  - Gen-straPP transformation:

$$\boldsymbol{\beta}_1 = g_{1,c_0}^{-1}(\boldsymbol{\eta}) = [I_1^{-1/2} I_0^{1/2}(\boldsymbol{\eta})]\boldsymbol{\eta}_1 + [I_1^{-1/2}]\boldsymbol{c}_0$$

Additional Model Setup

- Additionally we compute the partial-borrowing power prior, partial-borrowing asymptotic power prior and uniform improper prior.
- For the partial-borrowing power prior and asymptotic power prior, we specify a uniform improper prior for the intercept
- For the partial-borriwng commensurate prior, we specify a gamma prior for τ with shape and inverse scale parameters equal to 2
- For all models, we specify a gamma prior for ϕ with shape and inverse scale parameters equal to 0.01
- For all models, we used Metropolis-Hastings MCMC with 20,000 MC samples.

Choosing a_0, ω_0

- We ran the Gen-straPP model on the COMPASS data with varying values of (ω_0, a_0).
  - $a_0 \in \{0, 0.1, 0.25, 0.5, 0.75, 1.0\}$
  - $\omega_0 \in \{0, 0.1, 0.25\}$
- Values of (ω_0, a_0) will be chosen based on the lowest deviance information criterion (DIC), developed by Spiegelhalter et al. (2002). A similar approach was taken by Ibrahim et al. (2015).
  - The DIC is calculated as

$$\text{DIC}(a_0) = 2E(\text{Dev}(\boldsymbol{\theta}) \mid D, D_0, a_0) - \text{Dev}(\bar{\boldsymbol{\theta}}),$$

    where $\bar{\boldsymbol{\theta}} = E(\boldsymbol{\theta} \mid D, D_0, a_0)$ and $\text{Dev}(\boldsymbol{\theta}) = -2\sum_{i=1}^{n} \log f(y_i \mid \boldsymbol{x}_i, \boldsymbol{\theta})$.
  - Lower values of DIC indicate better performance of the associated prior.
- Then we will use the chosen a_0 for the original straPP, power prior, and asymptotic power prior.

Table 1: DIC for the Gen-straPP with Various (a_0, ω_0)

	a_0					
ω_0	0.0**	0.1	0.25	0.5	0.75	1.0
0.00	2816.56	2815.23	2815.72	2817.20	2818.88	2820.61
0.10	2816.56	2816.06	2816.17	2817.16	2818.51	2819.63
0.25	2816.56	2816.13	2816.13	2816.77	2817.87	2818.84

Note: *$\omega_0 = 0$ is equivalent to the straPP and **$a_0 = 0$ is equivalent to the uniform improper prior.

We note that we calculated the DIC for the partial-borrowing Gen-straPP for higher values of ω_0 than those shown (up to $\omega_0 = 1$) which resulted in similar patterns of DIC with varying a0 for a given value of ω_0.

Table 4: Posterior Estimates when $a_0 = 0.1$

Model	DIC	eCare Plan		History of Stroke		Minor NIHSS		Moderate-Severe NIHSS		Non-white	
		Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD
straPP	2815.23	0.69 (0.95)	(-1.16, 2.55)	-0.93 (1.16)	(-3.29, 1.26)	-1.42 (1.07)	(-3.53, 0.65)	-3.78 (1.08)	(-5.89, -1.66)	-1.55 (1.56)	(-4.23, 1.70)
APP	2815.85	0.28 (0.77)	(-1.21, 1.79)	-0.37 (0.96)	(-2.23, 1.56)	-0.73 (0.80)	(-2.31, 0.83)	-3.07 (1.17)	(-5.42, -0.82)	-2.17 (1.88)	(-5.92, 1.46)
PP	2815.97	0.29 (0.77)	(-1.16, 1.86)	-0.36 (0.96)	(-2.27, 1.48)	-0.72 (0.81)	(-2.30, 0.86)	-3.04 (1.20)	(-5.47, -0.78)	-2.19 (1.87)	(-5.86, 1.40)
Gen-straPP	2816.06	0.72 (0.98)	(-1.26, 2.57)	-0.89 (1.17)	(-3.17, 1.42)	-1.40 (1.08)	(-3.47, 0.77)	-3.77 (1.20)	(-6.14, -1.45)	-1.53 (1.64)	(-4.55, 1.96)
UIP	2816.56	0.85 (0.98)	(-1.08, 2.79)	-1.18 (1.24)	(-3.60, 1.27)	-1.69 (1.10)	(-3.81, 0.50)	-4.69 (1.32)	(-7.19, -2.00)	-2.23 (2.12)	(-6.40, 1.91)
COM	2817.45	0.45 (0.84)	(-1.16, 2.14)	-0.47 (0.99)	(-2.43, 1.44)	-0.48 (0.93)	(-2.39, 1.21)	-2.35 (1.25)	(-4.79, 0.03)	-1.53 (1.37)	(-4.31, 1.07)

straPP, scale transformed power prior; APP, asymptotic power prior; PP, power prior; Gen-straPP, generalized scale transformed power prior; UIP, uniform improper prior; COM, commensurate prior.

Gen-straPP was run with $\omega_0 = 0.1$.

Table 4: Posterior Estimates when $a_0 = 0.1$

		eCare Plan		History of Stroke		Minor NIHSS		Moderate-Severe NIHSS		Non-white	
Model	DIC	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD
straPP	2815.23	0.69 (0.95)	(-1.16, 2.55)	-0.93 (1.16)	(-3.29, 1.26)	-1.42 (1.07)	(-3.53, 0.65)	-3.78 (1.08)	(-5.89, -1.66)	-1.55 (1.56)	(-4.23, 1.70)
APP	2815.85	0.28 (0.77)	(-1.21, 1.79)	-0.37 (0.96)	(-2.23, 1.56)	-0.73 (0.80)	(-2.31, 0.83)	-3.07 (1.17)	(-5.42, -0.82)	-2.17 (1.88)	(-5.92, 1.46)
PP	2815.97	0.29 (0.77)	(-1.16, 1.86)	-0.36 (0.96)	(-2.27, 1.48)	-0.72 (0.81)	(-2.30, 0.86)	-3.04 (1.20)	(-5.47, -0.78)	-2.19 (1.87)	(-5.86, 1.40)
Gen-straPP	2816.06	0.72 (0.98)	(-1.26, 2.57)	-0.89 (1.17)	(-3.17, 1.42)	-1.40 (1.08)	(-3.47, 0.77)	-3.77 (1.20)	(-6.14, -1.45)	-1.53 (1.64)	(-4.55, 1.96)
UIP	2816.56	0.85 (0.98)	(-1.08, 2.79)	-1.18 (1.24)	(-3.60, 1.27)	-1.69 (1.10)	(-3.81, 0.50)	-4.69 (1.32)	(-7.19, -2.00)	-2.23 (2.12)	(-6.40, 1.91)
COM	2817.45	0.45 (0.84)	(-1.16, 2.14)	-0.47 (0.99)	(-2.43, 1.44)	-0.48 (0.93)	(-2.39, 1.21)	-2.35 (1.25)	(-4.79, 0.03)	-1.53 (1.37)	(-4.31, 1.07)

straPP, scale transformed power prior; APP, asymptotic power prior; PP, power prior; Gen-straPP, generalized scale transformed power prior; UIP, uniform improper prior; COM, commensurate prior.

Table 4: Posterior Estimates when $a_0 = 0.1$

		eCare Plan		History of Stroke		Minor NIHSS		Moderate-Severe NIHSS		Non-white	
Model	DIC	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD
straPP	2815.23	0.69 (0.95)	(-1.16, 2.55)	-0.93 (1.16)	(-3.29, 1.26)	-1.42 (1.07)	(-3.53, 0.65)	-3.78 (1.08)	(-5.89, -1.66)	-1.55 (1.56)	(-4.23, 1.70)
APP	2815.85	0.28 (0.77)	(-1.21, 1.79)	-0.37 (0.96)	(-2.23, 1.56)	-0.73 (0.80)	(-2.31, 0.83)	-3.07 (1.17)	(-5.42, -0.82)	-2.17 (1.88)	(-5.92, 1.46)
PP	2815.97	0.29 (0.77)	(-1.16, 1.86)	-0.36 (0.96)	(-2.27, 1.48)	-0.72 (0.81)	(-2.30, 0.86)	-3.04 (1.20)	(-5.47, -0.78)	-2.19 (1.87)	(-5.86, 1.40)
Gen-straPP	2816.06	0.72 (0.98)	(-1.26, 2.57)	-0.89 (1.17)	(-3.17, 1.42)	-1.40 (1.08)	(-3.47, 0.77)	-3.77 (1.20)	(-6.14, -1.45)	-1.53 (1.64)	(-4.55, 1.96)
UIP	2816.56	0.85 (0.98)	(-1.08, 2.79)	-1.18 (1.24)	(-3.60, 1.27)	-1.69 (1.10)	(-3.81, 0.50)	-4.69 (1.32)	(-7.19, -2.00)	-2.23 (2.12)	(-6.40, 1.91)
COM	2817.45	0.45 (0.84)	(-1.16, 2.14)	-0.47 (0.99)	(-2.43, 1.44)	-0.48 (0.93)	(-2.39, 1.21)	-2.35 (1.25)	(-4.79, 0.03)	-1.53 (1.37)	(-4.31, 1.07)

straPP, scale transformed power prior; APP, asymptotic power prior; PP, power prior; Gen-straPP, generalized scale transformed power prior; UIP, uniform improper prior; COM, commensurate prior.

Table 4: Posterior Estimates when $a_0 = 0.1$

Model	DIC	eCare Plan		History of Stroke		Minor NIHSS		Moderate-Severe NIHSS		Non-white	
		Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD	Mean (SD)	95%HPD
straPP	2815.23	0.69 (0.95)	(-1.16, 2.55)	-0.93 (1.16)	(-3.29, 1.26)	-1.42 (1.07)	(-3.53, 0.65)	-3.78 (1.08)	(-5.89, -1.66)	-1.55 (1.56)	(-4.23, 1.70)
APP	2815.85	0.28 (0.77)	(-1.21, 1.79)	-0.37 (0.96)	(-2.23, 1.56)	-0.73 (0.80)	(-2.31, 0.83)	-3.07 (1.17)	(-5.42, -0.82)	-2.17 (1.88)	(-5.92, 1.46)
PP	2815.97	0.29 (0.77)	(-1.16, 1.86)	-0.36 (0.96)	(-2.27, 1.48)	-0.72 (0.81)	(-2.30, 0.86)	-3.04 (1.20)	(-5.47, -0.78)	-2.19 (1.87)	(-5.86, 1.40)
Gen-straPP	2816.06	0.72 (0.98)	(-1.26, 2.57)	-0.89 (1.17)	(-3.17, 1.42)	-1.40 (1.08)	(-3.47, 0.77)	-3.77 (1.20)	(-6.14, -1.45)	-1.53 (1.64)	(-4.55, 1.96)
UIP	2816.56	0.85 (0.98)	(-1.08, 2.79)	-1.18 (1.24)	(-3.60, 1.27)	-1.69 (1.10)	(-3.81, 0.50)	-4.69 (1.32)	(-7.19, -2.00)	-2.23 (2.12)	(-6.40, 1.91)
COM	2817.45	0.45 (0.84)	(-1.16, 2.14)	-0.47 (0.99)	(-2.43, 1.44)	-0.48 (0.93)	(-2.39, 1.21)	-2.35 (1.25)	(-4.79, 0.03)	-1.53 (1.37)	(-4.31, 1.07)

straPP, scale transformed power prior; APP, asymptotic power prior; PP, power prior; Gen-straPP, generalized scale transformed power prior; UIP, uniform improper prior; COM, commensurate prior.

Posterior Distribution of the straPP

The straPP as developed on slides 12-15 can be written as,

$$\pi_s(\boldsymbol{\theta} \mid D_0, a_0) \propto \mathcal{L}(\boldsymbol{\eta} \mid D_0)^{a_0} \pi_0(\boldsymbol{\eta}) \left| \frac{d\boldsymbol{\eta}}{d\boldsymbol{\theta}} \right| \mathsf{I}\left[\boldsymbol{\eta} = g\left(\boldsymbol{\theta}\right)\right], \tag{14}$$

where $\mathsf{I}\left[A\right]$ is an indicator that A is true, and $\boldsymbol{\eta}$ is calculated by solving

$$I_0^{1/2}(\boldsymbol{\eta})\boldsymbol{\eta} = I_1^{1/2}(\boldsymbol{\theta})\boldsymbol{\theta}. \tag{15}$$

Posterior Distribution of the straPP

When neither information matrix is free of the parameter, it is still possible to analyze data using the straPP by considering a posterior representation involving both $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, and using a Metropolis-Hastings sampling algorithm where proposed values of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ satisfy the constraint in Equation (15).

Algorithm for Sampling Posterior Distribution of the straPP

Let t denote the current sample. We propose the following algorithm for sampling the posterior distribution of the straPP.

1. Propose $\boldsymbol{\theta}^{(t)} \sim N\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t-1)}, \hat{\Sigma}\right)$, where $\hat{\Sigma}$ is the maximum likelihood estimate (MLE) for the new data analysis.
2. Compute $c(\boldsymbol{\theta}^{(t)}) = I_1^{1/2}(\theta^{(t)})\boldsymbol{\theta}^{(t)}$
3. Solve for $\boldsymbol{\eta}^{(t)}$ in $I_0^{1/2}(\boldsymbol{\eta}^{(t)})\boldsymbol{\eta}^{(t)} = c(\boldsymbol{\theta}^{(t)})$, via a nonlinear programming (NLP) solver in SAS or R.
4. Perform a Metropolis-Hastings step based on the proposal value on $\left(\boldsymbol{\eta}^{(t)}, \boldsymbol{\theta}^{(t)}\right)$.

The Scale Transformed Power Prior for Survival Data

- We allow the historical and current data set to be a GLM or a survival model

$$\pi_s(\boldsymbol{\theta}|D_0) \propto L(g(\boldsymbol{\theta})|D_0)^{a_0}\pi_0(g(\boldsymbol{\theta}))\left|\frac{dg(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right| \tag{16}$$

  where $\boldsymbol{\theta} = (\phi, \boldsymbol{\beta}, \boldsymbol{\lambda})$.
  - ϕ is the dispersion parameter
  - $\boldsymbol{\beta}$ is vector of regression parameters
  - $\boldsymbol{\lambda}$ is vector of baseline hazard parameters
- Here we propose that the current data is survival data, and that we do not wish to borrow from the historical dispersion parameter or current baseline hazard parameters

- We construct a finite partition of the time axis, $0 < s_1 < s_2 < \ldots < s_J$,with $s_J > y_i$ for all $i = 1, 2, \ldots, n$.
- In the j^{th} interval, assume a constant baseline hazard $h_0(y) = \lambda_j$ for $y \in I_j = (s_{j-1}, s_j]$.
- Let $D = (n, \boldsymbol{y}, \boldsymbol{X}, \nu)$ denote the observed data, where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$, $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_n)'$.
    - $\nu_i = 1$ if the i^{th} subject failed and 0 otherwise

We can write the likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ for the n subjects as

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\lambda}|D) = \prod_{i=1}^{n} \prod_{j=1}^{J} \left(\lambda_j \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right)^{\delta_{ij}\nu_i} \exp\Bigg\{ -\delta_{ij}\Bigg[\lambda_j(y_i - s_{j-1}) \\
+ \sum_{g=1}^{j-1} \lambda_g(s_g - s_{g-1})\Bigg] \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\Bigg\}, \qquad (17)
\end{aligned}
$$

Proposed straPP for Survival Data

- For the piecewise exponential model, the information matrix of the current data does not depend on the regression or baseline hazard parameters, therefore we can solve for $\boldsymbol{\eta} = g_1(\boldsymbol{\beta}, \phi) = A(\boldsymbol{\beta}, \phi)\boldsymbol{\beta}$.
- An attractive computational form of the partial-borrowing straPP is given by

$$\begin{aligned}\pi(\phi, \boldsymbol{\beta}, \boldsymbol{\lambda} | D_0) &= \pi_s(\phi, \boldsymbol{\beta} | D_0)\pi(\boldsymbol{\lambda}) \\ &\propto L(\phi, g_1(\boldsymbol{\beta}, \phi) | \boldsymbol{\lambda}, D_0)^{a_0} \pi_0(g_1(\boldsymbol{\beta}, \phi)) \\ &\quad \times \left| \frac{d g_1(\boldsymbol{\beta}, \phi)}{d\boldsymbol{\beta}} \right| \pi(\phi)\pi(\boldsymbol{\lambda}), \end{aligned} \tag{18}$$

where $L(\cdot | D_0)$ is the historical data likelihood.

Study Design based on the Scale Transformed Power Prior

Motivation

- The Prescription Drug User Fee Amendments of 2017 (PDUFA VI) and the 21st Century Cures Act contain provisions to make the use of complex innovative trial designs (CID) easier
- In 2018, the FDA launched a pilot program to incorporate the use of CID (CDER 2018)
- Goals:
  - maximize clinical trial efficiency
  - apply innovative approaches where traditional methods may not be feasible

- Chen et al. (2011) developed a Bayesian approach to sample size determination (SSD) in noninferiority clinical trials - partial-borrowing power prior and normalized power prior
- Found that in using historical data, the sample size needed for the current study was reduced from 1480 to 1080 subjects for a type I error rate of 0.05 and 80% power
- However, the study found that the approach works best when the historical and current data are compatible.

straPP for Clinical Trial Design

- We propose to use the straPP to develop a new Bayesian sample size determination (SSD) procedure.
- We believe that by using the straPP to develop a new Bayesian SSD procedure, we can reduce the sample size while maintaining type I error rate and power.
- The straPP is able to account for the incompatibility of the historical and current data in terms of different data types.

Brown, E. R., and Ibrahim, J.G., (2003), "A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data," *Biometrics*, 2, 221-228.

CDER (2018),"CDER Conversation: Complex Innovative Trial Designs, *News and Events for Human Drugs*.

Chen, M.-H., Dey, D. K., and Shao, Q.-M. (2001), "Bayesian analysis of binary data using skewed logitmodels," *Calcutta Statistical Association Bulletin*, 51, 11-30.

Chen, M.-H., Ibrahim, J. G., Lam, P., Yu, A., and Zhang, Y. (2011), "Bayesian Design of Noninferiority Trials for Medical Devices Using Historical Data," *Biometrics*, 67, 1163-1170.

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

Duncan, P. W., Bushnell, C. D., Rosamond, W. D., Berkeley, S. B. J., Gesell, S. B., Jr, R. B. et al. (2017), "The Comprehensive Post-Acute Stroke Services (COMPASS) study: design and methods for a cluster-randomized pragmatic trial," BMC *Neurology*, 17.

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J. and Sargent, D. J. (2011), "Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials." *Biometrics*, 67, 1047-1056.

Ibrahim, J. G. and Chen, M.-H. (2000), "Power Prior Distributions for Regression Models," *Statistical Science*, 15, 46-60.

Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015), "The Power Prior: Theory and Applications," *Statistics in Medicine*, 34, 3724-3749.

Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (1999), "Monte Carlo EM for Missing Covariates in Parametric Regression Models," *Biometrics*, 55, 591-596.

References III

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001), *Bayesian Survival Analysis*, New York: Springer, chap. 3, pp. 47-86.

Kirkwood, J. M., Ibrahim, J.G., Sosman, J.A., Sondak, V.K., Agarwala, S.S, Ernstoff, M. S., and Rao, U. (2001), "High-dose interferon alfa-2b significantly prolongs relapse-free and overall survival compared with the GM2-KLH/QS-21 vaccine in patients with resected stage IIB-III melanoma: results of intergroup trial E1694/S9512/C509801," *Journal of Clinical Oncology*, 19, 2370-2380.

Laub, A.J. (2005), "Matrix Analysis for Scientists and Engineers," *Society for Industrial and Applied Mathematics*.

Lipsitz, S. R. and Ibrahim, J. G. (1996), "Using the EM-Algorithm for Survival Data with Incomplete Categorical Covariates," *Lifetime Data Analysis*, 2, 5-14.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2003), "Bayesian Approaches to Clinical Trials and Health-Care Evaluation," *John Wiley and Sons.*

Sylvester, J. (1884), "Sur lequations en matrices $px = xq$," *C.R. Acad. Sci. Paris*, 99, 67-71, 115-116.