Robust Bayesian Inference with Power Prior under Covariate Missingness

By

Zinan Zhao

B.S., Xi'an Jiaotong University, 2024

Thesis

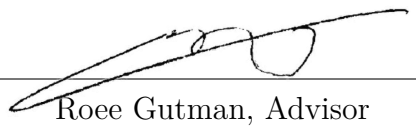
Submitted in fulfillment of the requirements for ScM in Biostatistics in the Brown University School of Public Health.

PROVIDENCE, RHODE ISLAND

MAY 2026

This thesis by Zinan Zhao is accepted in its present form
by the Department of Biostatistics as satisfying the
thesis requirements for the degree of Sc.M. in Biostatistics.

Date 4/30/2026



Roee Gutman, Advisor

Date 5/1/2026

Arman Oganisian, Reader

Date _____

_____

Stavroula Chrysanthopoulou, Director

Acknowledgements

I would like to express my deepest gratitude to my advisor, Professor Roee Gutman, for his invaluable guidance throughout this research. His profound knowledge, sharp insights, and patient support have been essential to both the development of the methodology and the completion of this thesis.

I am also sincerely grateful to my reader, Professor Arman Oganisian, for his thoughtful feedback and willingness to engage with my questions. His course on Bayesian methods has greatly enriched my understanding and played an important role in shaping this work. I would also like to thank the faculty members in the Department of Biostatistics at Brown University for their rigorous training and for introducing me to this field over the past two years.

I am thankful to my friends for their companionship and encouragement, which have made this journey both meaningful and enjoyable. I also appreciate the moments of inspiration and energy drawn from watching FC Barcelona, whose style of play has added brightness to my daily life.

Finally, I owe my deepest gratitude to my parents. Although far away, their unwavering support and encouragement have given me the strength and confidence to persevere.

Abstract of Robust Bayesian Inference with Power Prior under Covariate Missingness,
by Zinan Zhao, Sc.M., Brown University, May 2026.

Incorporating historical data can substantially improve statistical efficiency, but standard power prior methods typically assume that all covariates are fully observed across data sources. In many practical settings, however, key covariates are systematically missing in historical data, which may induce biased inference if external information is naively borrowed. In this thesis, we study Bayesian inference under covariate missingness within the power prior framework. We first consider settings in which a covariate is unobserved in the historical dataset but no additional inclusion mechanism is present, and derive a marginal likelihood formulation that enables valid borrowing based on partially observed records. We then extend the framework to settings with covariate-dependent historical inclusion. When selection depends on the missing covariate, we propose an importance-sampling weighted power prior to correct for the induced distributional shift. Simulation studies and a real-data application demonstrate that the proposed methods improve efficiency while maintaining robustness under a range of sensitivity scenarios.

Contents

List of Tables

1. Introduction

Randomized controlled trials (RCTs) are widely regarded as the gold standard for estimating the effects of interventions, but they are often constrained by limited sample sizes, high costs, and ethical considerations [Friedman et al., 2015]. Integrating historical data offers a principled way to augment trial evidence, improving efficiency and precision while potentially reducing the need for large concurrent control groups [Viele et al., 2013]. A substantial statistical literature has developed methods for incorporating historical data into current analyses, particularly within Bayesian frameworks that allow partial borrowing while guarding against bias from non-exchangeable sources. Approaches such as the power prior [Chen et al., 2000], commensurate prior [Hobbs et al., 2011], and meta-analytic-predictive prior [Schmidli et al., 2017] explicitly model the degree of similarity between historical and current data to enable adaptive borrowing. More recent work emphasizes robustification strategies—such as mixture priors and discounting—to mitigate prior-data conflict and improve frequentist operating characteristics [Schmidli et al., 2014, Viele et al., 2013]. These methods are now widely used in clinical trials and regulatory settings, especially for rare diseases and small-sample contexts where external information can substantially improve efficiency. However, these methods assume that the outcome variables and covariates that are used when analyzing the RCT are also available in the historical data. However, because of the prospective protocol-driven collection the data collected in randomized controlled trials is typically richer and more standardized than historical data, with more complete covariate information and consistent outcome definitions [Piantadosi, 2017, Pocock, 1976].

We will assume that the outcome variable and some covariates are recorded in both the RCT and the historical data, but some covariates are missing in the historical data. We will use the power prior method to integrate the RCT data and the historical data. The power prior is a Bayesian approach for incorporating historical data by raising the likelihood of the historical dataset to a power parameter that controls the degree of borrowing [Ibrahim and Chen, 2000]. This formulation allows analysts to discount historical information when there is concern about non-exchangeability, with extensions that treat the power parameter as random to enable data-adaptive borrowing [Müller and Quintana, 2004]. Subsequent work has further developed normalized and robust variants of the power prior to improve theoretical properties and operating characteristics in clinical trial settings [Ji et al., 2010, Schmidli et al., 2014]. We will discuss the assumptions and methods that are required

to integrate historical data in an RCT when covariates are missing from the historical data. We also develop sensitivity analysis to violations of these assumptions. Lastly, apply the proposed method to simulated and real data.

The remainder of the paper is organized as follows. Section 2 formalizes the data structure and missingness assumptions. Section 3 presents the modeling framework and our proposed extensions to the power prior. Section 4 shows the simulation results. Section 5 shows that real data application. Section 6 provides discussion and possible extensions.

1.1 Power Prior Introduction

The power prior [Ibrahim et al., 2015] is a Bayesian method for incorporating external data while accounting for potential incompatibility with current data. Let $\mathcal{D}_0$ denote the historical dataset and $\mathcal{D}$ the current data. Assuming a shared likelihood model $L(\theta; \mathcal{D})$, the power prior is defined as:

$$\pi(\theta \mid \mathcal{D}_0, \alpha) \propto L(\theta; \mathcal{D}_0)^{\alpha} \pi_0(\theta) \tag{1}$$

where $\pi_0(\theta)$ is a baseline prior and $\alpha \in [0, 1]$ controls the degree of borrowing. When $\alpha = 0$, the historical data is ignored; when $\alpha = 1$, it is fully incorporated as if fully compatible. Intermediate values allow partial borrowing.

Proposition 1. *Assume that the weight function $w(x)$ is such that*

$$sup_x |w(x) - c| \xrightarrow{p} 0 \tag{2}$$

where c is a positive constant. Then

$$\frac{1}{n} \sum_{i=1}^{n} w(x_i) s(x_i \mid \theta) \xrightarrow{p} cE(s(X \mid \theta)) = 0. \tag{3}$$

The above proposition can guarantee the unbiaseness of the posterior estimates obtained through power prior method. One of the advantages of using power prior is that it can reduce the posterior variance. In other words, it allow us to use trial data with smaller size to obtain good posterior estimates. However, the power prior can be sensitive to misspecification in $\mathcal{D}_0$. In the presence of covariate shifts or structural missingness, naive application of the power prior may yield biased posteriors. Addressing such complications motivates the methods developed in this work.

2. Data Structure and Notation

We consider a data integration setting where the target inference model is based on the joint relationship between an outcome Y and two covariate vectors, $X_1 \in \mathbb{R}^{p_1}$ and $X_2 \in \mathbb{R}^{p_2}$. The available data consists of two sources:

- A *current trial dataset* $\mathcal{D} = \{(Y_i, X_{1i}, X_{2i})\}_{i=1}^{n}$, where all variables are fully observed.
- An *historical dataset* $\mathcal{D}_h = \{(Y_{hi}, X_{1hi})\}_{i=1}^{n_h}$, in which the covariate vector $\mathbf{X}_2$ is systematically missing and not observed for any unit.

We assume a generalized linear model:

$$Y \mid X_1, X_2 \sim \text{Exponential Family}, \tag{4}$$

$$\eta = \beta_0 + \boldsymbol{\beta}_1^T X_1 + \boldsymbol{\beta}_2^T X_2, \tag{5}$$

$$g(\mu) = \eta. \tag{6}$$

where g is the link function.

Suppose θ is the set of parameters for $f(Y \mid X_1, X_2)$, we will focus on the estimates of $(\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \subset \theta$. These are what we are interested in when dealing with causal inference problems.

The challenge arises due to the absence of X_2 in the historical dataset $\mathcal{D}_h$, which precludes direct likelihood contributions from these observations without additional assumptions or modeling strategies. We have a baseline strategy that entirely discards the external dataset and performs Bayesian inference using the current trial data alone. While this approach ensures validity under no structural assumptions, it fails to exploit valuable auxiliary information and leads to higher posterior variance, especially in moderate-sample settings.

In the next section, we develop methods to use the partial information in $\mathcal{D}_h$ for inference on $(\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ under this missingness structure, including marginalization and latent modeling approaches.

3. Methodology

3.1 Compatibility Assumption

Our framework relies on the assumption that the trial dataset provides a valid reference distribution for the historical population after conditioning on observed co-

variates. Specifically, we assume that the conditional outcome model and the latent covariate structure are shared across the two data sources.

Formally, we assume that

$$f_T(Y \mid X_1, X_2; \theta) = f_H(Y \mid X_1, X_2; \theta) \tag{7}$$

and

$$f_T(X_2 \mid X_1, \delta) = f_H(X_2 \mid X_1, \delta). \tag{8}$$

The first condition ensures that the regression parameters $(\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ have the same interpretation in both datasets, while the second condition allows us to estimate the conditional distribution of the missing covariates X_2 using the trial data and integrate over its uncertainty in the historical likelihood.

3.2 No Inclusion Mechanism

In this case, we only need to construct the marginal historical likelihood by integrating X_2. The observed-data likelihood contribution for subject i in historical data is

$$L_{hi}(\theta) \;=\; \int f(Y_{hi} \mid X_{1hi}, X_{2hi}, \theta)\; f(X_{2hi} \mid X_{1hi}, \delta)\; dX_{2hi}. \tag{9}$$

A common method to approximate (9) is Monte Carlo sampling. Specifically, for given δ, draw $X_{2hi}^{(s)} \sim f(X_{2hi} \mid X_{1hi}, \delta)$ for $s = 1, \ldots, S$, and use

$$\widehat{L}_i(\theta) \;=\; \frac{1}{S} \sum_{s=1}^{S} f\Big(Y_{hi} \mid X_{1hi}, X_{2hi}^{(s)}, \theta\Big). \tag{10}$$

Notice that to do this, we need to fit $X_2 \sim X_1$ based on the trial data.

If the missing covariate is one-dimension and we have normality assumption for $X_2 \sim X_1$, the above integral can be more efficiently approximated by Gauss-Hermite method. After constructing the historical likelihood, we can directly apply the power prior method to incorporate historical data.

3.3 Extension to MAR and MNAR: Modeling the Inclusion Mechanism

We now extend our model to settings where the external dataset is not a random sample from the population, but is subject to a missing-not-at-random (MNAR) inclusion mechanism. [Neuhaus et al., 2013] Specifically, we assume that inclusion in the external data depends on the covariates via a logistic model:

$$\text{logit}P(M = 1 \mid X_1, X_2) = \xi_0 + \boldsymbol{\xi}_1 X_1 + \boldsymbol{\xi} X_2 \tag{11}$$

3.3.1 $\boldsymbol{\xi}_2 = 0$

Now we have:

$$\text{logit}P(M = 1 \mid X_1) = \xi_0 + \boldsymbol{\xi}_1 X_1 \tag{12}$$

In other words, we have $M \perp X_2 | X_1$ which is MAR.

Under this assumption, $f(\theta \mid D, D^h)$ is given by

$$\begin{aligned}
&f(\beta, \sigma \mid D, D^h) \\
&\propto \int\int\int \pi(\theta) \left[\prod_{i=1}^{n_h} f(M_i, Y_{hi}, X_{1hi})\right]^{\alpha} \prod_{i=1}^{n} f(Y_i \mid X_{1i}, X_{2i}, \theta) f(X_{1i}, X_{2i} \mid \theta_2) d\xi d\theta_1 d\theta_2 \\
&\propto C_1 \int\int \pi(\theta) \left[\prod_{i=1}^{n_h} f(M_i \mid X_{1hi}, \xi) f(Y_{hi} \mid X_{1hi}, \omega) f(X_{1hi} \mid \theta_1)\right]^{\alpha} \prod_{i=1}^{n} f(Y_i \mid X_{1i}, X_{2i}, \theta) d\xi d\theta_1 \\
&\propto C_2 \pi(\theta) \left[\prod_{i=1}^{n_h} f(Y_{hi} \mid X_{1hi}, \omega)\right]^{\alpha} \prod_{i=1}^{n} f(Y_i \mid X_{1i}, X_{2i}, \theta),
\end{aligned} \tag{13}$$

where C_2 is the term intergreted out, which will not change the posterior of our interest parameters. We have shown that $\omega = \theta$ in section 3.2. In this way, we can do the same posterior draw in `Stan`. In other words, inclusion mechanism depending on the observed covariates has no effect on the results.

3.3.2 $\boldsymbol{\xi}_2 \neq 0$

In this case, inclusion depends on the missing covariates, which is MNAR. To focus on the effect of the unobserved covariate in the inclusion mechanism, we consider the simplified model

$$\text{logit}\mathbb{P}(M = 1 \mid X_2) = \xi_0 + \boldsymbol{\xi}_2 X_2, \qquad \xi_2 \neq 0. \tag{14}$$

Denote the marginal trial density by $g_T(y, x_1) \equiv f_{Y,X_1}(y, x_1)$, and the historical observed marginal density by $g_H(y, x_1) \equiv f_{Y,X_1}(y, x_1 \mid M = 1)$. By Bayes' rule and the selection model, we have

$$\begin{aligned}
g_H(y, x_1) &= \int f_{Y,X_1,X_2}(y, x_1, x_2 \mid M = 1)\, dx_2 \\
&= \int \frac{\mathbb{P}(M = 1 \mid x_2)\, f_{Y,X_1,X_2}(y, x_1, x_2)}{\mathbb{P}(M = 1)}\, dx_2 \\
&= \frac{f_{Y,X_1}(y, x_1)}{\mathbb{P}(M = 1)} \int \pi(x_2)\, f_{X_2|Y,X_1}(x_2 \mid y, x_1)\, dx_2 \\
&= \frac{f_{Y,X_1}(y, x_1)\, \mathbb{E}[\pi(X_2) \mid Y = y, X_1 = x_1]}{\mathbb{P}(M = 1)}.
\end{aligned} \tag{15}$$

Therefore, the density-ratio (importance-sampling) weight that transports historical draws of (Y, X_1) back to the trial marginal satisfies

$$w(y, x_1) \propto \frac{g_T(y, x_1)}{g_H(y, x_1)} = \frac{f_{Y,X_1}(y, x_1)}{f_{Y,X_1}(y, x_1)\,\mathbb{E}[\pi(X_2) \mid y, x_1]/\mathbb{P}(M = 1)} \propto \frac{1}{\mathbb{E}[\pi(X_2) \mid Y = y, X_1 = x_1]}. \tag{16}$$

The proportionality constant $\mathbb{P}(M = 1)$ is common to all observations and thus does not affect weighted likelihood-based estimation (or weighted regression), so we use the above result as the operative weight. In addition,

$$\mathbb{E}[\pi(X_2) \mid Y, X_1] = \int \pi(X_2) f(X_2 \mid Y, X_1) dX_2 = \frac{\int \pi(X_2) f(Y \mid X_1, X_2) f(X_2 \mid X_1) dX_2}{\int f(Y \mid X_1, X_2) f(X_2 \mid X_1) dX_2} \tag{17}$$

Here, we need to numerically integrate both numerator and denominator.

Given these weights, we can construct the full likelihood. Let $\ell_T(\theta) = \sum_{i=1}^{N} \log f_T(Y_i \mid X_{1i}, X_{2i}; \theta)$ denote the trial log-likelihood. The posterior distribution of θ is

$$\pi(\theta \mid D_T, D_H) \propto \exp\left\{ \ell_T(\theta) + \alpha \sum_{h=1}^{N_h} w_h \log f(y_h \mid x_{1h}; \theta) \right\} \pi_0(\theta), \tag{18}$$

where $\alpha \in [0, 1]$ is the global borrowing parameter and $\pi_0(\theta)$ is the baseline prior. Relative to the standard power prior, the only change is the replacement $\alpha \sum_h \log p_T(\cdot)$ by $\alpha \sum_h w_h \log p_T(\cdot)$, so that the historical contribution is transported back to the trial (Y, X_1) distribution under covariate-dependent inclusion.

4. Simulation

We conduct a series of simulation studies to evaluate the performance of our proposed power prior models under a variety of missingness scenarios. To evaluate the effectiveness and robustness of the methods, we assess the posterior inference results for the parameters of the outcome model, including posterior estimates, 95% credible intervals, and standard deviation.

4.1 Example I: Normality Assumption

In this part, we assume that trial data and historical data share the same normal distribution

$$(Y, X_1, X_2) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma). \tag{19}$$

And the outcome model is

$$Y \mid X_1, X_2 \sim \mathcal{N}(\beta_0 + \beta_1 X_1 + \beta_2 X_2, \sigma^2) \tag{20}$$

Under this setting, we have

$$Y \mid X_1 \sim \mathcal{N}(\gamma_0 + \gamma_1 X_1, \sigma^2_{Y|X_1}), \tag{21}$$

where $\gamma_0 = \beta_0 + \beta_2\mu_{X_2} - \beta_2\rho_{X_1X_2}\frac{\sigma_{X_2}}{\sigma_{X_1}}\mu_{X_1}, \gamma_1 = \beta_1 + \beta_2\rho_{X_1X_2}\frac{\sigma_{X_2}}{\sigma_{X_1}}, \sigma^2_{Y|X_1} = \sigma^2 + \beta_2^2\sigma^2_{X_2}(1-\rho^2_{X_1X_2})$.

4.1.1 Stage I: Comparing Power Prior with Trial-only Analysis

We want to test the performance of two approaches: (1) using only the trial dataset and (2) applying the power prior method based on marginal likelihood of $Y \mid X_1$. Trial data are generated from a trivariate normal distribution with parameters:

$$\mu = (5, 2, 3), \quad \Sigma = \text{specified via correlations } \rho_{YX_1} = 0.5, \rho_{YX_2} = 0.3, \rho_{X_1X_2} = 0.4,$$
$$\sigma_Y = 1.5, \sigma_{X_1} = 2, \sigma_{X_2} = 2.5.$$

The historical data is generated with the same distribution and then remove X_2. Given this information, we can compute the real parameters, which is (4.11, 0.34, 0.07, 1.29).

The size of the trial data D is $N = 1000$, and the size of the historical data D_h is $N_h = 5000$. Each simulation is repeated for 50 iterations. For each iteration, we record posterior summaries (mean, standard deviation, 95% credible interval width) of the regression parameters $(\beta_0, \beta_1, \beta_2, \sigma^2)$ and then summarize the average performance. The MCMC sampling iterations is set to be 2000 and we fix $\alpha = 0.6$ as the power parameter. Below are our results:

Table 1: Posterior summaries comparing trial-only and power prior models

	β_0			β_1			β_2			σ		
Model	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
Only Trial	3.98	0.067	[3.85, 4.11]	0.35	0.021	[0.32, 0.37]	0.09	0.018	[0.05, 0.10]	1.26	0.028	[1.21, 1.32]
Power Prior	4.03	0.048	[3.94, 4.13]	0.34	0.013	[0.33, 0.36]	0.09	0.018	[0.06, 0.11]	1.28	0.015	[1.25, 1.31]

We observe that the model using only trial data underestimates β_0 and β_2, likely due to insufficient information from a limited sample.

In contrast, the power prior method achieves a balance: it yields accurate point estimates for all parameters, particularly improving estimation of β_2, and reduced the posterior SD compared with Bayesian using only trial. This demonstrates its effectiveness in integrating external data in a coherent Bayesian fashion, outperforming trial-only inference in terms of bias and uncertainty calibration.

4.1.2 Stage II: Study of Historical Inclusion Mechanism

In the second stage, we test our proposed power prior framework under covariate-dependent inclusion mechanism in the historical dataset.

Trial data are generated from a trivariate normal distribution with parameters:

$$\mu = (3, 1, 2), \quad \Sigma = \text{specified via correlations } \rho_{YX_1} = 0.6, \rho_{YX_2} = 0.45, \rho_{X_1X_2} = 0.2,$$
$$\sigma_Y = 2.2, \sigma_{X_1} = 1.5, \sigma_{X_2} = 1.8.$$

The sample sizes for trial data is $N = 1000$. The complete historical data with size $N_{hc} = 10000$ is generated with the same distribution. Then it is selected by the inclusion mechanism. Finally we remove X_2 to obtain the biased D_h and record the sample size N_h. Given this information, we can compute the real parameters, which is (1.38, 0.78, 0.42, 1.60).

Scenario A: Shift driven by X_1. We simulate missingness using a logit inclusion probability $P(M = 1 \mid X_1) = \text{logit}^{-1}(\xi_0 + \xi_1 X_1)$ and retain only samples with $M = 1$ in the external data. Since X_1 is observed, the marginal power prior remains valid. Here we set $(\xi_0, \xi_1) = (-1, 0.5)$.

Table 2: Posterior summaries under inclusion mechanism depending on X_1

	β_0			β_1			β_2			σ		
Model	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
Only Trial	1.34	0.075	[1.19, 1.48]	0.74	0.034	[0.67, 0.81]	0.44	0.027	[0.39, 0.49]	1.53	0.035	[1.46, 1.60]
Power Prior	1.33	0.055	[1.22, 1.44]	0.73	0.020	[0.69, 0.77]	0.45	0.022	[0.41, 0.50]	1.55	0.026	[1.51, 1.61]

Note. True value: (1.3806, 0.7792, 0.4201, 1.5964)

Results in the above table support our derivation. Standard deviation of the power prior method is smaller.

Scenario B: Shift driven by X_2. We now let inclusion depend on X_2 via $P(M = 1 \mid X_2) = \text{logit}^{-1}(\xi_0 + \xi_2 X_2)$, and treat X_2 as unobserved in the external data. We set $(\xi_0, \xi_2 = (-1, 0.5))$.

Firstly, we want to know whether the original method using marginal $Y_h \mid X_{1h}$ works, though we know the marginal distribution will shift due to the inclusion mechanism.

We can notice that the true β_0 is not in the 95% confidence interval, so the original method no longer works.

Table 3: Posterior summaries under MNAR inclusion ($\xi_2 \neq 0$)

	β_0			β_1			β_2			σ		
Model	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
Only Trial	1.34	0.075	[1.19, 1.48]	0.74	0.034	[0.67, 0.81]	0.44	0.027	[0.39, 0.49]	1.53	0.035	[1.46, 1.60]
Power Prior	1.51	0.057	[1.41, 1.63]	0.76	0.019	[0.72, 0.80]	0.45	0.024	[0.39, 0.49]	1.55	0.026	[1.50, 1.60]

Note. True value: (1.3806, 0.7792, 0.4201, 1.5964)

Then we test the efficiency of our proposed weighted power prior method based on importance sampling. We also want to observe its consistency under large sample size. So We do another experiment with $(N, N_{hc}) = (10000, 100000)$.

Table 4: Posterior summaries under MNAR inclusion with importance sampling correction

	β_0			β_1			β_2			σ		
Model	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
Only Trial	1.34	0.075	[1.19, 1.48]	0.74	0.034	[0.67, 0.81]	0.44	0.027	[0.39, 0.49]	1.53	0.035	[1.46, 1.60]
Power Prior(IS)	1.43	0.053	[1.33, 1.54]	0.74	0.021	[0.70, 0.78]	0.41	0.026	[0.37, 0.46]	1.60	0.025	[1.55, 1.65]
Power Prior(IS, much larger sample size)	1.39	0.017	[1.37, 1.43]	0.77	0.006	[0.76, 0.78]	0.42	0.007	[0.40, 0.43]	1.60	0.008	[1.58, 1.61]

Note. True value: (1.3806, 0.7792, 0.4201, 1.5964); $(\xi_0, \xi_2) = (-1, 0.5)$

Under covariate-dependent inclusion ($\xi_2 \neq 0$), the importance-sampling–based weighted power prior exhibits both efficiency gains and desirable large-sample behavior. Relative to using the trial data alone, incorporating historical information through importance-sampling weights substantially reduces posterior uncertainty for all parameters, while avoiding the bias induced by naive borrowing under selection. More importantly, when both the trial and historical sample sizes increase to $(N, N_{hc}) = (10000, 100000)$, the posterior means under the proposed method concentrate tightly around the true parameter values, and posterior standard deviations shrink at the expected rate. This empirical evidence supports the consistency of the weighted power prior estimator and demonstrates that, even in the presence of covariate-dependent selection, the proposed reweighting strategy can effectively recover the target trial distribution and deliver reliable inference in large samples.

4.2 Example II: Logit Link GLM with inclusion mechanism depending on the missing X_2 in historical data

In this part, we assume that trial data and historical data share the same generalized linear model

$$Y \sim \text{Bernouli}(\eta), \quad \eta = \text{logit}^{-1}(\beta_0 + \beta_1 X_1 + \beta_2 X_2) \tag{22}$$

and two covariates have joint normal distribution

$$(X_1, X_2) \sim \mathcal{N}(\boldsymbol{\mu}_X, \Sigma_X) \tag{23}$$

In addtion, we have inclusion mechanism in historical data

$$P(M = 1 \mid X_{2h}) = \text{logit}^{-1}(\xi_0 + \xi_2 X_{2h}). \tag{24}$$

We set $(\beta_0, \beta_1, \beta_2) = (-0.5, 0.6, 0.5)$, $\boldsymbol{\mu}_X = (1, 2)$, $\sigma_{X_1} = 1.5$, $\sigma_{X_2} = 1.8$, $\rho = 0.2$, $(\xi_0, \xi_2) = (-1, 0.5)$.

Then we test the efficiency of our proposed weighted power prior method based on importance sampling. We also want to observe its consistency under large sample size. So We do another experiment with $(N, N_{hc}) = (10000, 100000)$.

Table 5: Posterior summary(Logit link, $\xi_2 \neq 0$)

	β_0			β_1			β_2		
Model	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
Only Trial	-0.27	0.114	[-0.49, -0.04]	0.52	0.061	[0.40, 0.64]	0.45	0.048	[0.36, 0.54]
Power Prior(IS)	-0.39	0.094	[-0.57, -0.20]	0.58	0.033	[0.51, 0.65]	0.45	0.051	[0.35, 0.55]
Power Prior(IS, much larger sample size)	-0.53	0.040	[-0.61, -0.45]	0.61	0.021	[0.56, 0.65]	0.50	0.016	[0.47, 0.54]

Note. True value: (-0.5, 0.6, 0.5); $(\xi_0, \xi_2) = (-1, 0.5)$

The results show that our proposed weighted power prior method has more efficient posterior estimates which also show consistency under large sample.

4.3 Sensitivity Analysis

In the simulation study, the inclusion mechanism parameters $\boldsymbol{\xi}$ and the borrowing strength α are treated as fixed. To assess the robustness of the proposed method to misspecification of the inclusion mechanism and to different levels of borrowing, we conduct a sensitivity analysis over a two-dimensional grid.

Specifically, we fix $\xi_0 = -1$ at its true value and vary the slope parameter ξ_2 over a symmetric grid

$$\xi_2 \in \{-\xi_{2,\max}, \ldots, \xi_{2,\max}\}, \tag{25}$$

where the upper bound $\xi_{2,\max}$ is calibrated using the empirical distribution of X_2 in the trial data. In particular, letting $q_{0.05}$ and $q_{0.95}$ denote the 5th and 95th percentiles of X_2, we choose $\xi_{2,\max}$ such that the induced difference on the logit scale,

$$\xi_{2,\max}(q_{0.95} - q_{0.05}), \tag{26}$$

corresponds approximately to a contrast between inclusion probabilities 0.05 and 0.95. This ensures that the grid covers a wide range of selection strengths, from no selection to strong MNAR mechanisms in both directions. The true value $\xi_2 = 0.5$ is included in the grid.

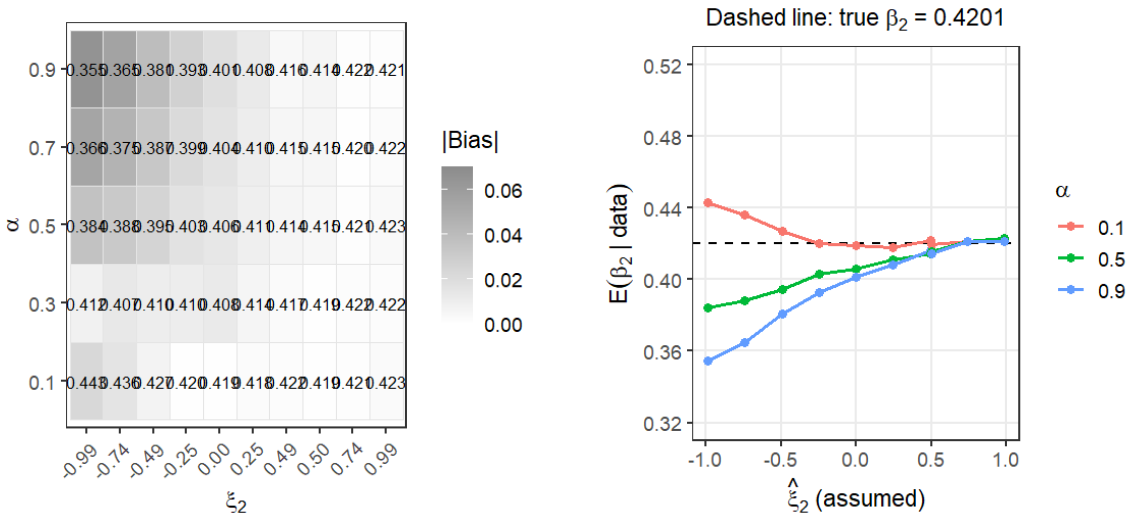
For the borrowing strength, we consider

$$\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}. \tag{27}$$

Under the normality assumption, for each combination (ξ_2, α), the importance-sampling weights for the historical data are recomputed based on the assumed inclusion mechanism, and the weighted power prior model is refitted to obtain posterior summaries of $(\beta_0, \beta_1, \beta_2, \sigma)$.

Figure 1(a) summarizes the sensitivity of the posterior mean of β_2. The heatmap displays the absolute bias relative to the true value $\beta_2 = 0.4201$, while the numerical values in each cell correspond to the posterior means. Overall, the proposed method demonstrates strong robustness: across a wide range of ξ_2 and α, the posterior mean of β_2 remains close to the truth, with only mild bias even under substantial misspecification of the inclusion mechanism.

Figure 1(b) further illustrates that, although the estimated β_2 varies systematically with the assumed ξ_2, the variation is moderate and diminishes as ξ_2 approaches the true value. Moreover, the effect of α on the posterior mean is relatively limited, primarily influencing the degree of shrinkage rather than inducing substantial bias. These results indicate that the proposed importance-sampling weighted power prior approach is stable with respect to both the selection mechanism and the borrowing strength.



(a) Heatmap of |Bias|

(b) Posterior mean under selected α

Figure 1: Sensitivity analysis of the posterior mean of β_2.

5. Application

5.1 Data Description

We illustrate the proposed framework using data from the National Health and Nutrition Examination Survey (NHANES). The outcome of interest is log-transformed glycohemoglobin (HbA1c), denoted as $\log(gh)$. We consider a set of covariates including treatment status (`tx`), age, sex, and body mass index (BMI).

The variable `tx` indicates whether an individual is currently using insulin or diabetes-related medication. In this analysis, `tx` is treated as a primary exposure variable of interest, and its regression coefficient is used as the target parameter for methodological comparison.

5.2 Construction of Trial and Historical Data

To mimic a data generation process with missing covariates, we partition the dataset into a trial subset and a historical subset with a ratio of approximately 1:5. The trial dataset contains complete information on all covariates, while in the historical dataset BMI is systematically missing.

Let Y denote the outcome, X_1 represent observed covariates (age and sex), and X_2 denote BMI. The outcome model is specified as:

$$Y = \beta_0 + \beta_{\text{tx}}\,\text{tx} + \beta_{\text{age}}\,\text{age} + \beta_{\text{sex}}\,\text{sex} + \beta_{\text{bmi}}\,\text{bmi} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2). \tag{28}$$

For the historical data, where BMI is unobserved, we integrate over its conditional distribution $p(\text{bmi} \mid \text{tx}, \text{age}, \text{sex})$, estimated from the trial data.

5.3 Modeling Framework

We compare the following approaches:

1. **Trial-only analysis**, using only the complete trial data;
2. **Power prior**, which incorporates historical data via a weighted likelihood with fixed power parameter α;
3. **Inclusion mechanisms**, where the availability of historical observations depends on covariates. We will consider both the inclusion depending on the observed covariates and incluision depending on the missing covariate BMI.

5.4 Results

We report posterior summaries for the coefficient β_{tx}, which serves as the primary parameter of interest.

Table 6: Posterior summary for β_{tx} under different models (NHANES data)

	β_{tx} (log scale)				
Model	Mean	SD	2.5%	50%	97.5%
Only Trial	0.263	0.0110	0.241	0.262	0.284
Power Prior	0.235	0.0060	0.223	0.235	0.247
Power Prior (IM1)	0.232	0.0073	0.217	0.233	0.246
Weighted PP (IM2)	0.218	0.0087	0.201	0.218	0.235

5.5 Summary

When no inclusion mechanism is present, the power prior substantially decreases the posterior standard deviation while moderately shifting the posterior mean. Under

inclusion depending on observed covariates (IM1), the standard power prior remains valid and achieves similar efficiency gains.

When selection depends on the missing covariate (IM2), the proposed weighted power prior approach corrects for selection bias and yields stable estimates with reduced uncertainty.

6. Discussion and Conclusion

In this work, we studied Bayesian inference with historical data under systematic covariate missingness. Motivated by practical settings where key covariates are unavailable or structurally missing in external datasets, we developed a series of principled extensions that enable efficient and robust borrowing while maintaining valid posterior inference for the target regression parameters. Under a joint multivariate normal model, we showed that when historical missingness is ignorable or conditionally independent of the missing covariate, a closed-form marginal likelihood for the historical data can be derived and seamlessly embedded into the power prior, yielding substantial efficiency gains over trial-only analysis.

When historical inclusion depends on the unobserved covariate, leading to an MNAR mechanism, we demonstrated that naive application of marginal power priors can result in biased inference due to distributional shifts in the observed historical sample. To address this challenge, we proposed an importance-sampling-weighted power prior that explicitly corrects for selection-induced bias by reweighting historical likelihood contributions to recover the target trial distribution. The resulting estimator preserves the variance-reduction benefits of borrowing while restoring consistency, as supported by both finite-sample simulations and large-sample experiments.

Extensive simulation studies and sensitivity analyses further showed that the proposed weighted power prior is robust to a wide range of assumptions on the inclusion mechanism and borrowing strength. In particular, inference on the regression coefficient associated with the missing covariate remains stable across grids of sensitivity parameters, highlighting the practical reliability of the method even under partial misspecification. Taken together, these results demonstrate that power priors, when carefully adapted to account for covariate missingness and selection bias, can serve as a flexible and robust tool for integrating historical data in modern Bayesian analyses.

References

Ming-Hui Chen, Joseph G. Ibrahim, and Qi-Man Shao. Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference*, 84(1):121–137, 2000. ISSN 0378-3758. doi: https://doi.org/10.1016/S0378-3758(99)00140-8. URL https://www.sciencedirect.com/science/article/pii/S0378375899001408.

Alexander Friedman, Daigo Homma, Leif G. Gibb, Ken-ichi Amemori, Samuel J. Rubin, Adam S. Hood, Michael H. Riad, and Ann M. Graybiel. A corticostriatal path targeting striosomes controls decision-making under conflict. *Cell*, 161(6):1320–1333, 2015. ISSN 0092-8674. doi: 10.1016/j.cell.2015.04.049. URL http://dx.doi.org/10.1016/j.cell.2015.04.049.

Richard J. Hobbs, Lauren M. Hallett, Paul R. Ehrlich, and Harold A. Mooney. Intervention ecology: Applying ecological science in the twenty-first century. *BioScience*, 61(6):442–450, 06 2011. ISSN 0006-3568. doi: 10.1525/bio.2011.61.6.6. URL https://doi.org/10.1525/bio.2011.61.6.6.

Joseph G. Ibrahim and Ming-Hui Chen. Power prior distributions for regression models. *Statistical Science*, 15(1):46–60, 2000.

Joseph G Ibrahim, Ming-Hui Chen, Yeonhee Gwon, and Fang Chen. The power prior: theory and applications. *Statistics in Medicine*, 34(28):3724–3749, 2015. doi: 10.1002/sim.6728.

Yuan Ji, Ping Liu, Yisheng Li, and {B. Nebiyou} Bekele. A modified toxicity probability interval method for dose-finding trials. *Clinical Trials*, 7(6):653–663, December 2010. ISSN 1740-7745. doi: 10.1177/1740774510382799.

Peter Müller and Fernando A. Quintana. Nonparametric bayesian data analysis. *Statistical Science*, 19(1), February 2004. ISSN 0883-4237. doi: 10.1214/088342304000000017. URL http://dx.doi.org/10.1214/088342304000000017.

Jonathan M Neuhaus, James J Schlesselman, and Nicholas P Jewell. Missing at random and missing completely at random are not the same: results of imputation and analysis when a planned missing data design is used. *American Journal of Epidemiology*, 178(9):1405–1410, 2013. doi: 10.1093/aje/kwt084.

Steven Piantadosi. *Clinical Trials: A Methodologic Perspective*. John Wiley & Sons, Hoboken, NJ, 3 edition, 2017.

Stuart J. Pocock. The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, 29(3):175–188, 1976. ISSN 0021-9681. doi: https://doi.org/10.1016/0021-9681(76)90044-8. URL https://www.sciencedirect.com/science/article/pii/0021968176900448.

Heinz Schmidli, Sandro Gsteiger, Satrajit Roychoudhury, Anthony O'Hagan, David Spiegelhalter, and Beat Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, 2014. doi: https://doi.org/10.1111/biom.12242. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12242.

Heinz Schmidli, Beat Neuenschwander, and Tim Friede. Meta-analytic-predictive use of historical variance data for the design and analysis of clinical trials. *Computational Statistics Data Analysis*, 113:100–110, 2017. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2016.08.007. URL https://www.sciencedirect.com/science/article/pii/S0167947316301888.

Kert Viele, Scott Berry, Beat Neuenschwander, Billy Amzal, Fang Chen, Nathan Enas, Brian Hobbs, Joseph G. Ibrahim, Nelson Kinnersley, Stacy Lindborg, Sandrine Micallef, Satrajit Roychoudhury, and Laura Thompson. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, 13(1):41–54, August 2013. ISSN 1539-1612. doi: 10.1002/pst.1589. URL http://dx.doi.org/10.1002/pst.1589.

Keying Ye, Zifei Han, Yuyan Duan, and Tianyu Bai. Normalized power prior bayesian analysis. *Journal of Statistical Planning and Inference*, 216:29–50, 2022. ISSN 0378-3758. doi: https://doi.org/10.1016/j.jspi.2021.05.005. URL https://www.sciencedirect.com/science/article/pii/S0378375821000501.

Appendix A. Normality Assumption

We next propose a principled extension of the power prior framework to accommodate missing covariates under the assumption that (Y, X_1, X_2) follows a trivariate normal distribution: Ye et al. [2022]

$$(Y, X_1, X_2) \sim \mathcal{N}(\mu, \Sigma)$$

Letting $Y \mid X_1, X_2 \sim \mathcal{N}(\beta_0+\beta_1 X_1+\beta_2 X_2, \sigma^2)$, we derive the marginal distribution $Y \mid X_1$ by integrating out X_2.

Using the formulas of total expectation and total variance, we have:

- $$\begin{aligned}\mathbb{E}[Y \mid X_1] &= \mathbb{E}[\mathbb{E}[Y \mid X_1, X_2]] \\ &= \beta_0 + \beta_1 X_1 + \beta_2 \mathbb{E}[X_2 \mid X_1] \\ &= \beta_0 + \beta_1 X_1 + \beta_2 \left(\mu_{X_2} + \rho_{X_1X_2}\frac{\sigma_{X_2}}{\sigma_{X_1}}(X_1 - \mu_{X_1})\right) \\ &= \underbrace{\beta_0 + \beta_2\mu_{X_2} - \beta_2\rho_{X_1X_2}\frac{\sigma_{X_2}}{\sigma_{X_1}}\mu_{X_1}}_{\gamma_0} + \underbrace{\beta_1 + \beta_2\rho_{X_1X_2}\frac{\sigma_{X_2}}{\sigma_{X_1}}}_{\gamma_1} X_1\end{aligned}$$
- $$\begin{aligned}\mathrm{Var}(Y \mid X_1) &= \mathbb{E}[\mathrm{Var}(Y \mid X_1, X_2)] + \mathrm{Var}(\mathbb{E}[Y \mid X_1, X_2]) \\ &= \sigma^2 + \beta_2^2 \mathrm{Var}(X_2 \mid X_1) \\ &= \sigma^2 + \beta_2^2 \sigma_{X_2}^2 (1 - \rho_{X_1X_2}^2)\end{aligned}$$

The properties of multi-normal distribution ganruatee that $Y \mid X_1$ is also normally distributed. In conclusion, we get:

$$Y \mid X_1 \sim \mathcal{N}(\gamma_0 + \gamma_1 X_1, \sigma^2_{Y|X_1}),$$

where the parameters are shown above and we can observe that it is based on $(\beta_0, \beta_1, \beta_2, \sigma^2)$ and some statistics.

Using this marginal likelihood, we define a power prior that incorporates the external data by discounting its likelihood contribution:

$$f(\beta, \sigma^2 \mid \mathcal{D}, \mathcal{D}_h, \alpha) \propto \pi(\beta, \sigma^2) \cdot \prod_{i=1}^{n} \mathcal{N}(Y_i; \beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}, \sigma^2) \cdot \left[\prod_{i=1}^{n_h} \mathcal{N}(Y_{hi}; \gamma_0 + \gamma_1 X_{1hi}, \sigma^2_{Y|X_1})\right]^{\alpha}$$

Here $(\gamma_0, \gamma_1, \sigma^2_{Y|X_1})$ can be directly computed with (β, σ^2) and statistics

$$(\mu_{X_1}, \mu_{X_2}, \sigma_{X_1}, \sigma_{X_2}, \rho_{X_1X_2})$$

obtained from the trial data. In this way, we successfully introducing the power prior structure without adding uncertainty.

Appendix B. Gauss-Hermite Method

To compute $\mathbb{E}[\pi(X_2) \mid y, x_1]$, we leverage the conditional distribution of $X_2 \mid (Y, X_1)$ implied by the trial model. In our Gaussian joint model for (Y, X_1, X_2), the conditional distribution is Normal:

$$X_2 \mid (Y = y, X_1 = x_1) \ \sim \ \mathcal{N}(m(y, x_1),\, v)\,,$$

where $m(y, x_1)$ and v are the standard conditional mean and variance obtained from the joint Normal parameters (estimated from the trial sample). Then

$$\mathbb{E}[\pi(X_2) \mid y, x_1] = \int \pi(z)\, \varphi(z; m(y, x_1), v)\, dz.$$

Apply the standardization $z = m(y, x_1) + \sqrt{2v}\, t$ with $t \sim \mathcal{N}(0, 1/2)$, yielding

$$\mathbb{E}[\pi(X_2) \mid y, x_1] = \int \pi\Big(m(y, x_1) + \sqrt{2v}\, t\Big)\ \frac{1}{\sqrt{\pi}} e^{-t^2}\, dt.$$

Let $\{t_k, w_k\}_{k=1}^K$ be the Gauss–Hermite nodes and weights for approximating integrals of the form $\int h(t) e^{-t^2} dt$. Then we have the following approximation

$$\mathbb{E}[\pi(X_2) \mid y, x_1] \ \approx \ \frac{1}{\sqrt{\pi}} \sum_{k=1}^{K} w_k\, \pi\Big(m(y, x_1) + \sqrt{2v}\, t_k\Big)\,.$$

In conclusion, the implementable weight for a historical record (y_h, x_{1h}) is

$$w_h \ \propto \ \left[\frac{1}{\sqrt{\pi}} \sum_{k=1}^{K} w_k\, \pi\Big(m(y_h, x_{1h}) + \sqrt{2v}\, t_k\Big)\right]^{-1}, \qquad \pi(u) = \mathrm{expit}(\xi_0 + \xi_2 u).$$

In practice we normalize $\{w_h\}$ (e.g., divide by $\frac{1}{N_H} \sum_h w_h$) for numerical stability; this normalization does not change the implied weighted likelihood up to an overall constant.