Received: 28 March 2021 | Revised: 5 July 2021 | Accepted: 7 July 2021

DOI: 10.1002/sim.9146

RESEARCH ARTICLE

Statistics in Medicine WILEY

Unit information prior for adaptive information borrowing from multiple historical datasets

Huaqing Jin | **Guosheng Yin**

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong SAR, China

Correspondence
Guosheng Yin, Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong SAR, China.
Email: gyin@hku.hk

Funding information
Research Grants Council, University Grants Committee, Grant/Award Number: 17308420

Abstract

In clinical trials, there often exist multiple historical studies for the same or related treatment investigated in the current trial. Incorporating historical data in the analysis of the current study is of great importance, as it can help to gain more information, improve efficiency, and provide a more comprehensive evaluation of treatment. Enlightened by the unit information prior (UIP) concept in the reference Bayesian test, we propose a new informative prior called UIP from an information perspective that can adaptively borrow information from multiple historical datasets. We consider both binary and continuous data and also extend the new UIP to linear regression settings. Extensive simulation studies demonstrate that our method is comparable to other commonly used informative priors, while the interpretation of UIP is intuitive and its implementation is relatively easy. One distinctive feature of UIP is that its construction only requires summary statistics commonly reported in the literature rather than the patient-level data. By applying our UIP to phase III clinical trials for investigating the efficacy of memantine in Alzheimer's disease, we illustrate its ability to adaptively borrow information from multiple historical datasets. The Python codes for simulation studies and the real data application are available at https://github.com/JINhuaqing/UIP.

KEYWORDS

Bayesian design, clinical trial, Fisher's information, historical data, informative prior, multiple studies

1 | INTRODUCTION

Prior distributions are crucial in Bayesian data analysis and inference. By incorporating many sources of knowledge, such as expert opinions and historical data, a properly elicited prior distribution can help to analyze the current data more efficiently and thus reduce the cost and ethical hazard in clinical trials. With a large number of subjects followed for a long period of time, large-scale clinical trials are typically expensive. Clinical trials are rarely conducted in isolation, and thus it is critical to combine multiple samples systematically to improve the analysis of the current study. For example, several cancer clinical trials on the same or similar types of treatment are often conducted with patients of different ethnicity groups or disease sub-types and sometimes in different countries.[1-3] These trials typically have comparable settings with similar follow-ups and eligibility criteria, and sometimes use the same endpoints, hence information can be borrowed from them to improve efficiency of the current trial. It may also happen that several trials investigate different treatments with the same control arm used,[2,4] and thus the data in the control arm are valuable to the current trial as a common benchmark for comparison.

A major challenge associated with multiple historical datasets is to determine the amount of information borrowed from each dataset for the current study. Differences in patient populations or other trial-specific settings lead to heterogeneity among the current trial and historical trials, and thus the full-borrowing strategy is typically imprudent, as it would inflate the type I error substantially. Moreover, when the total sample size of the historical datasets is large, the information from historical datasets might overwhelm the analysis of the current study which is not desirable, as the data from the current study should typically dominate the analysis.

Several methods have been proposed for adaptively borrowing information from historical data. Pocock[5] considers the difference in the parameter of interest between the current and historical datasets and models this difference with a zero-mean random variable. Ibrahim et al[6] propose the joint power prior (JPP) to discount the historical dataset by a power parameter in the range of $[0, 1]$. Duan et al[7] and Neuenschwander et al[8] modify the JPP by adding a normalization term, which is referred to as the modified power prior (MPP). Banbeta et al[9] and Gravestock and Held[10] further extend the power prior to multiple historical datasets with binary endpoints. Using a hierarchical model for the between-trial heterogeneity, Neuenschwander et al[11] develop the meta-analytic-predictive (MAP) prior via deriving the predictive distribution of the model parameter resulting from the analysis of historical datasets. To further account for the prior-data conflict, Schmidli et al[12] make the MAP prior more robust by incorporating a non-informative component in the prior distribution[13] which is called robust MAP (rMAP). Hobbs et al[14] propose the commensurate prior by using a commensurate parameter directly to parameterize the commensurability or exchangeability between the historical and current data. All the aforementioned methods can borrow information according to the consistency or exchangeability between the historical and current datasets. Nonetheless, Pocock's method and the commensurate prior are typically applicable to the case with a single historical dataset. When multiple historical datasets exist, the naive extensions for both methods do not take the underlying interaction among historical datasets into consideration. For the commensurate prior, when multiple historical datasets are involved, the formula under the non-Gaussian case would be complicated which causes difficulty in practice and the interpretations of multiple commensurate parameters may not be intuitive as the commensurability concept defined by Hobbs et al[14] is typically for the case with a single historical dataset. The MAP prior relies on the exchangeability assumption and adopts a single parameter to parameterize the heterogeneity between the current and historical datasets. With a single parameter, the relative contributions of multiple historical datasets are not accounted for, that is, heterogeneity of the coherence between the current and each historical dataset cannot be incorporated to the model.

Kass and Wasserman[15] use the Fisher information to define the amount of information and set the amount of information in the prior equal to that of a single observation to conduct Bayesian hypothesis tests using Bayes factors. Motivated by the information in a single observation, we propose the unit information prior (UIP) as a new class of informative prior distributions to dynamically borrow information from multiple historical datasets. Unlike other priors which are constructed from the likelihood function of historical datasets, the UIP, originated from the information perspective, directly parameterizes the amount of Fisher's information in the prior distribution. The amount of information in the UIP is closely related to the effective sample size (ESS) defined by Morita et al,[16] and thus the UIP framework can straightforwardly control the ESS in the prior distribution. Moreover, our method considers the heterogeneity between the current and historical datasets, which guarantees the efficiency of information borrowing. As it is elicited based on the Fisher information, often the UIP only requires summary statistics of the historical data that are commonly reported in publications (eg, point estimates and 95% confidence intervals) rather than the patient-level data. The UIP is directly applicable to the case with multiple historical datasets, whose parameters have intuitive interpretations.

The rest of this article is organized as follows. In Section 2, we introduce the general framework of the UIP method. In Section 3, we illustrate the UIP framework in single-arm trials with binary and continuous data, respectively, and discuss its connection with the power prior, commensurate prior and MAP prior in terms of the conditional prior distribution as well as making an extension of the UIP to linear models. We also discuss the ESS in connection to the UIP.[16,17] Extensive simulation studies are presented in Section 4 where we demonstrate the dynamic borrowing property of the UIP, and compare different priors elicited from multiple historical datasets. Section 5 provides an example from six phase III clinical trials for Alzheimer's disease to illustrate the behavior of our UIP approach in the real data application. We conclude the article with a brief discussion in Section 6.

2 | UNIT INFORMATION PRIOR

Let $D = \{Y_1, \dots, Y_n\}$ denote the data of the current trial of sample size n. Suppose that there are K historical datasets $\{D_1, \dots, D_K\}$ related to the current study, where $D_k = \{Y_{k,1}, \dots, Y_{k,n_k}\}$ denotes the kth dataset of size n_k, for $k = 1, \dots, K$.

The parameter of interest is often the treatment effect, denoted by θ, for the current study, while the counterpart of θ for D_k is denoted as θ_k for $k = 1, \dots, K$. The likelihood function of θ_k based on D_k is denoted by $L^{(k)}(\theta_k|D_k)$.

The UIP is constructed directly from an information perspective under the normal approximation. When eliciting an informative prior for the parameter of interest θ, we are mainly interested in the accuracy and precision, that is, the correctness and the amount of information contained in the prior distribution. Under the normal approximation, the accuracy of the prior distribution is determined by the mean of the prior and the amount of the Fisher information in the prior is the inverse of the variance. Thus, the UIP framework focuses on the mean and variance of the prior distribution, and the amount of information in the prior distribution can be explicitly controlled. Moreover, because the UIP only requires the first and second moments, often the summary statistics of the historical data (eg, point estimates and standard errors or 95% confidence intervals) would be sufficient to derive the prior distribution, which is the typical case in practice as the patient-level historical datasets are typically inaccessible.

When considering information borrowing from historical datasets, the parameter of interest θ is assumed to be close to the counterpart of historical datasets θ_k. Due to heterogeneity among historical datasets, we introduce a weight parameter w_k for the historical dataset D_k to measure the relative closeness between the current dataset D and the historical one D_k. The mean of the prior is defined as $\sum_{k=1}^{K} w_k\theta_k$, with the weight parameter $w_k \in (0, 1)$ and $\sum_{k=1}^{K} w_k = 1$, where w_k can be also viewed as the measurement of the relative contribution from dataset D_k to the analysis of the current study. The larger value of w_k, the more contribution from D_k.

Following the work of Kass and Wasserman,[15] we adopt the Fisher information as the measurement of the amount of information in the data. As a result, we define the unit information (UI) for θ_k in the dataset D_k as

$$\mathrm{I_U}(\theta_k) = -\frac{1}{n_k}\frac{\partial^2 \log L^{(k)}(\theta_k|D_k)}{\partial\theta_k^2},$$

that is, $\mathrm{I_U}(\theta_k)$ is the observed Fisher information of D_k averaged at a unit observation level. By taking the heterogeneity of the historical datasets into consideration, the contribution of each historical dataset to the current study would be distinct. Therefore, the unit information from all the K historical datasets is defined as $\sum_{k=1}^{K} w_k \mathrm{I_U}(\theta_k)$. We introduce M as the sample size of the total amount of information borrowed from the K datasets, and then the amount of information contained in the prior is $M\sum_{k=1}^{K} w_k \mathrm{I_U}(\theta_k)$. Under the normal approximation, the variance of the prior distribution is $\left\{M\sum_{k=1}^{K} w_k \mathrm{I_U}(\theta_k)\right\}^{-1}$.

Therefore, to adaptively borrow information from different datasets, we formulate the UIP framework as follows,

$$\begin{aligned} &\theta|(M, w_1, \dots, w_K, D_1, \dots, D_K) \sim \pi(\theta|M, w_1, \dots, w_K, D_1, \dots, D_K),\\ &\text{with} \quad \mathbb{E}_\pi(\theta) = \sum_{k=1}^{K} w_k\theta_k,\\ &\mathrm{Var}_\pi(\theta) = \left\{M\sum_{k=1}^{K} w_k \mathrm{I_U}(\theta_k)\right\}^{-1}. \end{aligned} \tag{1}$$

As θ_k is typically unknown, we adopt the maximum likelihood estimator (MLE) $\hat{\theta}_k$ to replace θ_k while keeping $M, w_1, \dots, w_K$ as unknown parameters that are determined adaptively by the data. We use the MLE to replace θ_k mainly for two reasons: (i) Such a replacement helps to reduce the complexity of our method; and (ii) The MLE has desirable asymptotic properties,[18] for example, consistency and efficiency. Generally speaking, the MLE has a convergence rate of $O_p(n^{-1/2})$.[19] As phase III trials typically have relatively large sample size, the MLE is an accurate estimator of the parameter θ_k. In many cases, the MLE is simply the sample mean that can be easily obtained. In our experiments, the replacement of the true mean by the MLE does not affect the performance of the UIP method for the historical data size as low as 50.

As we adopt the normal approximation to link the amount of information and the variance of the prior, the specific form of π is typically chosen as a normal distribution. When the range of θ is not $(-\infty, \infty)$, we can use other distributions with approximately bell-shaped density curves in the range of θ (eg, a Beta distribution with a moderate variance) or a truncated normal distribution. Therefore, if θ is the mean parameter of a normal distribution, the UIP of θ exactly has a normal distribution; and if θ is the rate parameter of a Bernoulli distribution, the UIP of θ may conform to a Beta distribution.

The parameter M determines the total number of units corresponding to the amount of information borrowed from all historical datasets as a whole. It is shown that the value of M is closely related to the ESS defined by Morita et al.[16] In practice, M can be either fixed when we aim to control the amount of information borrowed from historical data or unknown by setting a hyper-prior on M. When there is a lack of prior information on M, a non-informative uniform prior is recommended, as it is a standard practice in Bayesian methods; otherwise, a Poisson distribution can be used to incorporate the prior information on M in the hyper-prior. When using a uniform prior, a common choice is $M \sim \text{Uniform}\left\{0, \min(n, \sum_{k=1}^{K} n_k)\right\}$, which can prevent the historical information from overwhelming the current study. When borrowing information from historical datasets, the current study should dominate the analysis, while historical data may play a supplemental role. For the real data application, we set $\text{Uniform}\left\{0, \min(n, \sum_{k=1}^{K} n_k)\right\}$ as the hyper-prior for M, while for simulation studies, unless otherwise specified, we choose $M \sim \text{Uniform}\left\{0, \sum_{k=1}^{K} n_k\right\}$ to make a fair comparison with other methods.

It is essential to determine the values of weight parameters which characterize the amount of information borrowed from each historical dataset. The values of w_k's reveal the relative importance of historical datasets. If one historical dataset is more consistent with the current study compared with the others, the corresponding weight parameter should be larger. Furthermore, Mw_k can be interpreted as the number of units of information borrowed from historical dataset D_k. Therefore, the amount of information borrowed from each dataset mainly relies upon the corresponding weight parameter, while the total amount of information borrowed from all historical datasets is controlled by M.

We propose two approaches to determining the values of the weight parameters. One is a fully Bayesian approach by imposing a hyper-prior on $(w_1, \dots, w_K)$. As all the values of w_k's are between 0 and 1 satisfying the constraint $\sum_{k=1}^{K} w_k = 1$, it is natural to assign a Dirichlet prior to $(w_1, \dots, w_K)$. We take the sample sizes of historical datasets into consideration by selecting suitable parameters for the Dirichlet distribution. Intuitively, it is preferable to assign a higher weight to the historical dataset with larger sample size, while we should also prevent a historical dataset with extremely large sample size from dominating the information borrowing. To strike a balance, we recommend setting the hyper-prior $(w_1, \dots, w_K) \sim \text{Dirichlet}(\gamma_1, \dots, \gamma_K)$ where $\gamma_k = \min(1, n_k/n)$. We refer to the UIP with a Dirichlet prior distribution as UIP-Dirichlet.

The other approach is to first measure the distances between the current dataset and the historical ones. To determine the values of weight parameters, a proper measure of the "distance" between two datasets is needed for measuring their similarity. The Jensen-Shannon (JS) divergence is a commonly used metric for measuring the dissimilarity between two probability distributions.[20,21] An alternative is the Kullback-Leibler (KL) divergence, while we adopt the JS divergence due to its symmetrical property.

Similar to the discussion in the UIP-Dirichlet method, the weight parameter w_k should tend to be small when the sample size of D_k is small, while w_k should not be too large even if the sample size of D_k is extremely large. The JS divergence automatically penalizes the relatively small sample size (compared with the current dataset) of the historical dataset. When the sample size of D_k is larger than that of the current dataset, we randomly select n samples from D_k to calculate the JS divergence and repeat this procedure for a large number of times to obtain the average. More specifically, the weight parameters are determined as follows.

1. Specify an initial non-informative prior (eg, Jeffreys' prior) for the parameter θ under the current dataset D and θ_k under each D_k. Based on the initial prior, we obtain the initial posteriors $f_{\text{ini}}(\theta|D)$.
2. For $k = 1, \dots, K$, when $n_k \le n$, we obtain the initial posterior $f_{\text{ini}}(\theta|D_k)$ and calculate the JS divergence,

$$d_k = \text{JS}(D|D_k) = \frac{\text{KL}(D|D_k) + \text{KL}(D_k|D)}{2},$$

   where $\text{KL}(D|D_k)$ represents the KL divergence between two density functions $f_{\text{ini}}(\theta|D)$ and $f_{\text{ini}}(\theta|D_k)$,

$$\text{KL}(D|D_k) = \int f_{\text{ini}}(\theta|D) \log\left\{\frac{f_{\text{ini}}(\theta|D)}{f_{\text{ini}}(\theta|D_k)}\right\} \, d\theta.$$

   When $n_k > n$, we randomly draw n samples from D_k for N times to obtain $\{D_k^{(l)}\}_{l=1}^{N}$ and compute the initial posteriors $\{f_{\text{ini}}(\theta|D_k^{(l)})\}_{l=1}^{N}$, and then calculate the distance d_k as the average $\sum_{l=1}^{N} \text{JS}(D|D_k^{(l)})/N$.
3. The weight parameters are defined as

$$w_k = \frac{1/d_k}{\sum_{s=1}^{K}(1/d_s)}$$

for $k = 1, \ldots, K$. We can also use $1/\sqrt{d_k}$ or other power values for the distance in the above formula.

In an extremely rare case for binary data, D and D_k may be exactly the same which results in the zero JS divergence. As a remedy, we add a small number, say 10^{-6}, to d_k to avoid the division-by-zero problem. We refer to the UIP in conjunction with the JS divergence as UIP-JS, where the weights are prespecified and the only unknown parameter is M in the UIP-JS.

It is also possible to use other methods (eg, the empirical Bayes method) to predetermine the weight parameters before sampling. However, in terms of the computation, the JS divergence is easier compared with the empirical Bayes method. Moreover, the JS divergence measures the dissimilarity between two datasets from an information perspective, which is consistent with the UIP framework for the prior elicitation.

3 | UIP WITH BINARY AND CONTINUOUS DATA

We illustrate the UIP methods in a single-arm trial with continuous and binary data, respectively. We also discuss the connections among the power prior, commensurate prior, MAP prior, and UIP for continuous data in terms of the conditional prior distribution, as well as extending our UIP to linear models. Moreover, the relationship between the amount parameter M and the ESS of the informative prior distribution is investigated.

3.1 | Continuous data

Suppose that $\{Y_1, \ldots, Y_n\}$ are independent and identically distributed (i.i.d.) from $N(\theta, \sigma^2)$ and $\{Y_{k,1}, \ldots, Y_{k,n_k}\}$ are i.i.d. from $N(\theta_k, \sigma_k^2)$ for $k = 1, \ldots, K$. The parameter of interest is the mean θ and the unit information for D_k evaluated at the corresponding MLE $\hat{\theta}_k$ (which is the sample mean) is

$$\mathrm{I_U}(\hat{\theta}_k) = \frac{1}{\sigma_k^2},$$

where σ_k^2 can be simply replaced by its MLE $\hat{\sigma}_k^2$. We impose an inverse-Gamma prior for σ^2, for example, $\sigma^2 \sim \mathrm{InvGa}(0.01, 0.01)$. As a result, we obtain the UIP as

$$\theta|(M, w_1, \ldots, w_K, D_1, \ldots, D_K) \sim N(\mu, \eta^2),$$

where

$$\mu = \sum_{k=1}^{K} w_k \hat{\theta}_k,$$
$$\eta^2 = \left(M \sum_{k=1}^{K} \frac{w_k}{\hat{\sigma}_k^2} \right)^{-1}.$$

Under the normal distribution setting, the MPP, local commensurate prior (LCP), MAP prior, and UIP are closely related. We extend the MPP to the case with multiple historical datasets as follows,

$$\begin{aligned}
\pi^{\mathrm{MPP}}(\theta|\alpha_1, \ldots, \alpha_K, D_1, \ldots, D_K) &\propto \pi^{\mathrm{ini}}(\theta) \prod_{k=1}^{K} L^{(k)}(\theta|D_k)^{\alpha_k}, \\
\alpha_k &\sim \mathrm{Beta}(a, b),\ k = 1, \ldots, K, \\
\sigma^2 &\sim \mathrm{InvGa}(\zeta, \zeta), \\
\pi^{\mathrm{ini}}(\theta) &\propto 1,
\end{aligned}$$

where α_k is the power parameter for D_k and π^{ini} is the initial prior.

We also extend the LCP method to the case with K historical datasets. Denoting the commensurate parameter for dataset D_k by τ_k, the LCP is given by

$$\begin{aligned}
\pi^{\text{LCP}}(\theta|\tau_1, \dots, \tau_K, D_1, \dots, D_K) &\propto \pi^{\text{ini}}(\theta)\prod_{k=1}^{K}\int L^{(k)}(\theta_k|D_k)N\left(\theta;\theta_k,\frac{1}{\tau_k}\right)\,\mathrm{d}\theta_k,\\
\log(\tau_k) &\sim \text{Uniform}(\xi_1,\xi_2),\ k=1,\dots,K,\\
\sigma^2 &\sim \text{InvGa}(\zeta,\zeta),\\
\pi^{\text{ini}}(\theta) &\propto 1.
\end{aligned}$$

Denoting the between-trial standard deviation as τ, the MAP prior with multiple historical datasets can be written as

$$\begin{aligned}
\theta_1, \dots, \theta_K, \theta|\mu^{\text{MAP}}, \tau &\sim N(\mu^{\text{MAP}}, \tau^2),\\
\pi(\mu^{\text{MAP}}) &\propto 1,\\
\tau &\sim \text{HN}(0,1),
\end{aligned}$$

where HN denotes the half-normal distribution on the positive line.

The MPP and LCP, respectively, use the power parameter α_k and the commensurate parameter τ_k as the measurement of consistency or exchangeability between the current dataset and each historical dataset D_k. The MAP prior adopts the exchangeability assumption and utilizes a single between-trial dispersion parameter τ to measure the heterogeneity among the current and historical datasets. In the UIP method, the weight parameters measure the relative consistency of the historical datasets with respect to the current dataset, while the amount parameter M measures the total number of units of information borrowed from historical data.

Given the corresponding hyper-parameters (ie, weight and amount parameters for UIP, power parameters for MPP, the commensurate parameters for LCP, the dispersion parameter τ for MAP), the specific forms of the MPP, LCP, MAP, and UIP methods for continuous data are given as follows:

$$\begin{aligned}
\pi^{\text{MPP}}(\theta|\alpha_1, \dots, \alpha_K, D_1, \dots, D_K) &\sim N\left(\sum_{k=1}^{K}\frac{\alpha_k n_k/\hat{\sigma}_k^2}{\sum_{s=1}^{K}\alpha_s n_s/\hat{\sigma}_s^2}\hat{\theta}_k, \left(\sum_{k=1}^{K}\alpha_k n_k\frac{1}{\hat{\sigma}_k^2}\right)^{-1}\right),\\
\pi^{\text{LCP}}(\theta|\tau_1, \dots, \tau_K, D_1, \dots, D_K) &\sim N\left(\sum_{k=1}^{K}\frac{n_k\tau_k/(\tau_k\hat{\sigma}_k^2+n_k)}{\sum_{s=1}^{K}n_s\tau_s/(\tau_s\hat{\sigma}_s^2+n_s)}\hat{\theta}_k, \left(\sum_{k=1}^{K}\frac{n_k\tau_k\hat{\sigma}_k^2}{\tau_k\hat{\sigma}_k^2+n_k}\frac{1}{\hat{\sigma}_k^2}\right)^{-1}\right),\\
\pi^{\text{MAP}}(\theta|D_1, \dots, D_K, \tau) &\sim N\left(\sum_{k=1}^{K}\frac{n_k/(\hat{\sigma}_k^2+n_k\tau)}{\sum_{s=1}^{K}n_s/(\hat{\sigma}_s^2+n_s\tau)}\hat{\theta}_k, \frac{1}{\sum_{k=1}^{K}n_k/(\hat{\sigma}_k^2+n_k\tau)}+\tau^2\right),\\
\pi^{\text{UIP}}(\theta|M, w_1, \dots, w_K, D_1, \dots, D_K) &\sim N\left(\sum_{k=1}^{K}w_k\hat{\theta}_k, \left(\sum_{k=1}^{K}w_k M\frac{1}{\hat{\sigma}_k^2}\right)^{-1}\right).
\end{aligned}$$

Conditioning on the hyper-parameters, the MPP, LCP, MAP, and UIP approaches all lead to normal distributions with different means and variances. Interestingly, the means of all four priors can be written in the form of a weighted sum of individual sample means. The UIP adopts weight parameters in a much more direct way, while the MPP parameterizes weights as an increasing function of the power parameters. The LCP method utilizes commensurate parameters τ_k to measure the commensurability between the current and historical datasets, where a larger value of τ_k indicates a higher level of commensurability, and the weight naturally increases with τ_k. The MAP prior utilizes a single dispersion parameter τ to control the information borrowing, and thus the weights mainly rely on the historical variances rather than the dispersion parameter τ.

The precision (ie, inverse of the variance) of the four priors can be used to quantify the amount of information borrowed from historical datasets. In particular, the precision of the MPP, LCP, and UIP methods can be written as a weighted sum of the observed Fisher information from each dataset. For the MPP method, the number of units of information for D_k is determined by the product of the corresponding power parameter α_k and sample size n_k. The amount of information borrowed from D_k under the LCP approach adopts an increasing function of τ_k and n_k while that under the UIP framework corresponds to the product of the weight parameter w_k and the total amount parameter M, which has the most transparent form and intuitive interpretation. Under the exchangeability assumption, the MAP prior borrows information from multiple historical datasets as a whole, where the amount of information borrowed is a decreasing function of the dispersion parameter τ.

3.2 | Binary data

Suppose that $\{Y_1, \dots, Y_n\}$ are i.i.d. samples from Bernoulli(θ), and the historical data $\{Y_{k,1}, \dots, Y_{k,n_k}\}$ are from Bernoulli(θ_k) for $k = 1, \dots, K$. The UI under the kth historical dataset D_k evaluated at the MLE $\hat{\theta}_k$ is

$$\mathrm{I_U}(\hat{\theta}_k) = \frac{1}{\hat{\theta}_k(1-\hat{\theta}_k)}.$$

As the support of the rate parameter θ is $[0, 1]$, we assign a Beta prior on θ. Denote the prior mean and prior variance of UIP as $\mu = \sum_{k=1}^K w_k\hat{\theta}_k$ and $\eta^2 = \left\{M\sum_{k=1}^K w_k/\{\hat{\theta}_k(1-\hat{\theta}_k)\}\right\}^{-1}$. Thus, the UIP of θ can be written as

$$\theta|(M, w_1, \dots, w_K, D_1, \dots, D_K) \sim \mathrm{Beta}(\alpha, \beta),$$

where the two Beta distribution parameters can be easily derived by solving the mean and variance equations in (1),

$$\alpha = \mu\left\{\frac{\mu(1-\mu)}{\eta^2} - 1\right\},$$
$$\beta = (1-\mu)\left\{\frac{\mu(1-\mu)}{\eta^2} - 1\right\}.$$

3.3 | Linear regression

Under a linear regression model, $Y_i \sim N(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2)$, where $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^\top$ is the regression coefficients and $\mathbf{x}_i = (1, x_{1,i}, \dots, x_{p,i})^\top$ is the covariate vector associated with the outcome Y_i. For the kth historical dataset, $Y_i^{(k)} \sim N(\mathbf{x}_i^{(k)\top}\boldsymbol{\beta}^{(k)}, \sigma_k^2)$, where $\boldsymbol{\beta}^{(k)} = (\beta_0^{(k)}, \dots, \beta_p^{(k)})^\top$ and $\mathbf{x}_i^{(k)} = (1, x_{1,i}^{(k)}, \dots, x_{p,i}^{(k)})^\top$ for $k = 1, \dots, K$ and $i = 1, \dots, n_k$.

Under the linear model, we obtain the UI for D_k evaluated at $\hat{\beta}_l^{(k)}$ as

$$\mathrm{I_U}(\hat{\beta}_l^{(k)}) = \frac{1}{n_k \mathrm{Var}(\hat{\beta}_l^{(k)})}, \tag{2}$$

where $\hat{\beta}_l^{(k)}$ is the MLE of $\beta_l^{(k)}$ and $\mathrm{Var}(\hat{\beta}_l^{(k)})$ is the corresponding variance for $l = 0, \dots, p$. Thus, the UIPs of the regression coefficients are given by

$$\beta_l|(M, w_1, \dots, w_K, D_1, \dots, D_K) \sim N\left(\sum_{k=1}^K w_k\hat{\beta}_l^{(k)}, \left\{M\sum_{k=1}^K w_k \mathrm{I_U}(\hat{\beta}_l^{(k)})\right\}^{-1}\right),$$

for $l = 0, \dots, p$.

Certainly, we can assign the weight parameters and total amount parameter for each coefficient individually, while this strategy involves too many unknown parameters, leading to difficulties in the implementation of Markov chain Monte Carlo (MCMC). Hence, our parsimonious strategy is more desirable, that is, we use the same weights and M for all coefficients. If not all regression coefficients in the linear model are shared between the current data and historical data due to different sets of covariates, we can impose UIPs on those shared coefficients only and leave the unshared ones with non-informative priors.

3.4 Prior ESS

For any prior distribution, it is critical to quantify how much information is contained in the distribution in terms of the ESS.[11] In the sequel, we discuss the ESS and its connection with the amount parameter M in the UIP.

Given the weight parameters w_k and the amount parameter M, the ESS of the UIP can be easily obtained via the method of Morital et al.[16] For continuous data, the ESS is $\sigma^2 M \sum_{k=1}^{K}(w_k/\sigma_k^2)$. In practice, when incorporating the historical information, it is reasonable to assume $\sigma^2 \approx \sigma_1^2 \approx \dots \approx \sigma_K^2$. In such case, the ESS is approximately equal to the amount parameter M. For binary data, the ESS is

$$\alpha+\beta = \frac{\mu(1-\mu)}{\eta^2} - 1 = M\left\{\sum_{s=1}^{K} w_s\hat{\theta}_s\right\}\left\{\sum_{l=1}^{K} w_l(1-\hat{\theta}_l)\right\}\left\{\sum_{k=1}^{K}\frac{w_k}{\hat{\theta}_k(1-\hat{\theta}_k)}\right\} - 1,$$

which is approximately equal to $M-1$ when $\hat{\theta}_1 \approx \dots \approx \hat{\theta}_K$. Therefore, M in the UIP represents the total number of units in the informative prior.

It is also possible to obtain the ESS under the full Bayesian manner following Morita et al.[17] By integrating out the parameters $(w_1, \dots, w_K, M)$, the marginal informative prior is

$$\pi(\theta|D_1, \dots, D_K) = \int \pi(\theta|w_1, \dots, w_K, D_1, \dots, D_K, M)\pi(w_1, \dots, w_K, M)\, \mathrm{d}w_1 \dots \mathrm{d}w_K\, \mathrm{d}M.$$

We then define the ϵ-information conditional prior $\pi_\epsilon(\theta|w_1, \dots, w_K, M)$ such that it has the same mean but very large variance compared with the conditional informative prior. Suppose the dataset $\overline{D}^{(m)}$ contains m samples and all samples are $\overline{\theta}$ where $\overline{\theta}$ is the mean of the distribution $\pi(\theta|D_1, \dots, D_K)$. This leads to the expected posterior as

$$\pi_\epsilon(\theta|\overline{D}^{(m)}) \propto L(\theta|\overline{D}^{(m)}) \int \pi_\epsilon(\theta|w_1, \dots, w_K, M)\pi(w_1, \dots, w_K, M)\, \mathrm{d}w_1 \dots \mathrm{d}w_K\, \mathrm{d}M.$$

The ESS is defined as the value of m by minimizing $|\sigma_\pi^{-2} - \sigma_\epsilon^{-2}(\overline{D}^{(m)})|$, where σ_π^2 and $\sigma_\epsilon^2(\overline{D}^{(m)})$ are the variances under $\pi(\theta|D_1, \dots, D_K)$ and $\pi_\epsilon(\theta|\overline{D}^{(m)})$, respectively.

4 SIMULATION STUDIES

We conduct extensive simulations to assess the characteristics of the UIP with continuous data, and the results for binary data are presented in the Supplementary Material. First, we introduce the notation θ_0 as the true parameter value for the current data, while θ is the generic notation of the parameter of interest. We evaluate the ESS and the adaptive borrowing property of the UIP. We also compare the UIP with Jeffreys' prior, full-borrowing strategy, MPP, LCP, and rMAP priors under the single-arm trial in terms of the mean squared error (MSE) as well as hypothesis testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. The full-borrowing strategy refers to the analysis by directly pooling the current and historical datasets together and applying Jeffreys' prior for the pooled dataset. For the MPP and LCP methods, we adopt flat initial priors, that is, $\pi^{\mathrm{ini}}(\theta) \propto 1$. We assign a non-informative prior, Beta(1, 1), to the power parameter α_k of the MPP method, while a vague prior, Uniform(−30, 30), is imposed on the logarithm of the commensurate parameter $\log(\tau_k)$ for the LCP method. Following Neuenschwander et al,[11] the rMAP prior adopts the half-normal distribution with scale parameter 1 for the dispersion parameter τ and assigns weight 0.1 to the non-informative component.

4.1 Effective sample size

We justify the relationship between the amount parameter M and ESS in the prior distribution. For continuous data, we adopt three historical datasets of sample sizes 80,100, 120 with the sample size of the current dataset 100. The amount parameter M varies from 50 to 150. The true values of mean θ_0 and standard deviation σ of the current dataset are fixed at 0 and 1, while the means $(\theta_1, \theta_2, \theta_2)$ and standard deviations $(\sigma_1, \sigma_2, \sigma_3)$ of the historical datasets are randomly

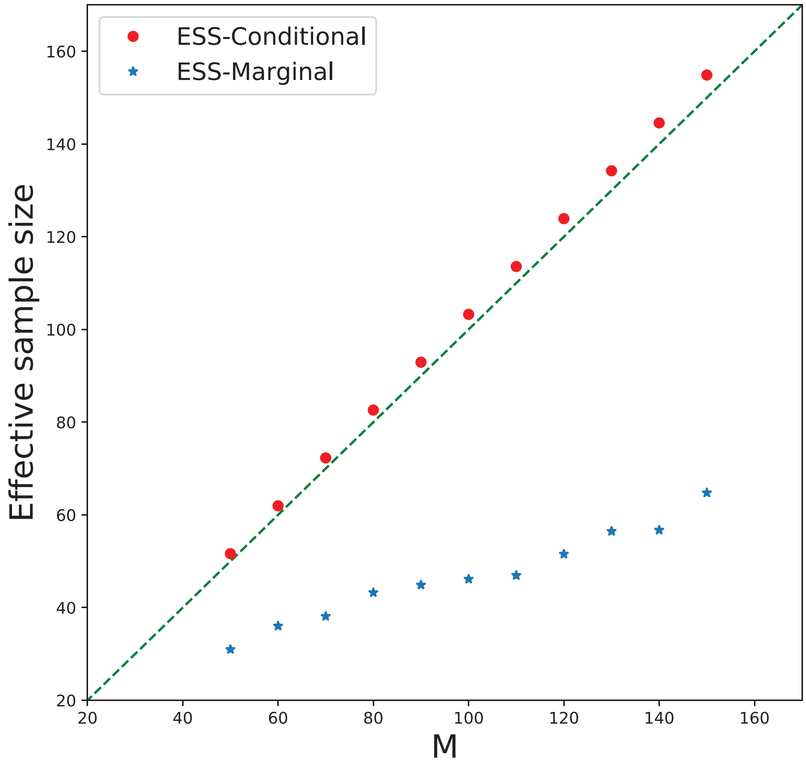

FIGURE 1 Effective sample size (ESS) of the conditional UIP and that of the marginal UIP for continuous data when varying M from 50 to 150 [Colour figure can be viewed at wileyonlinelibrary.com]

generated from ranges $[-0.5, 0.5]$ and $[0.9, 1.1]$, respectively. To obtain robust results, we repeat the experiment for 100 times.

The ESS of the conditional prior distribution[16] and that of the marginal prior distribution[17] under 100 repetitions are presented in Figure 1, where the weights for the conditional UIP are calculated via the JS method. For continuous data, the ESS contained in the conditional UIP is close to the amount parameter M. As the hyper-prior on weight parameters introduces more uncertainty, the ESS in the marginal UIP decreases compared with that in the conditional UIP given the same M as desired. However, the ESS of the marginal UIP still shows the same tendency with the amount parameter M, that is, it increases as M increases. The results for binary data show similar trends as shown in Web Figure 1 of the Supplementary Material.

4.2 | Adaptive borrowing property

We demonstrate that under the UIP, the values of the total amount parameter M, the weight parameters w_k's, and the Mw_k's can adapt to the level of consistency between the historical datasets and current dataset. If the historical datasets are close to the current, the UIP borrows more information from historical datasets, that is, the value of M would be large. When a certain dataset D_k is more consistent with D relative to other datasets, more weight would be assigned to D_k.

To examine the trend of the total amount parameter M, we consider two historical datasets D_1 and D_2 generated from $N(-0.3, 1)$ and $N(0.3, 1)$, respectively, with the same sample size $n_1 = n_2 = 40$. We vary θ_0, the true value of the mean for the current dataset D also with sample size $n = 40$, from 0.3 to 1.6 and fix the standard deviation at $\sigma = 1$. The hyper-prior for M is set as $M \sim \text{Uniform}(0, 40)$. We draw the posterior samples of the total amount parameter M under both UIP-Dirichlet and UIP-JS and take the posterior mean of M as the estimate. We replicate the experiment for 100 times, and the averages of posterior means of M under both priors are shown in the left column of Figure 2. When the level of consistency (or exchangeability) between the population mean of D and those of historical datasets decreases, the value of M decreases, indicating that less information is borrowed from historical datasets. It is worth noting that when inconsistency becomes extremely severe, the amount parameter M may drop close to 0, which indicates that there

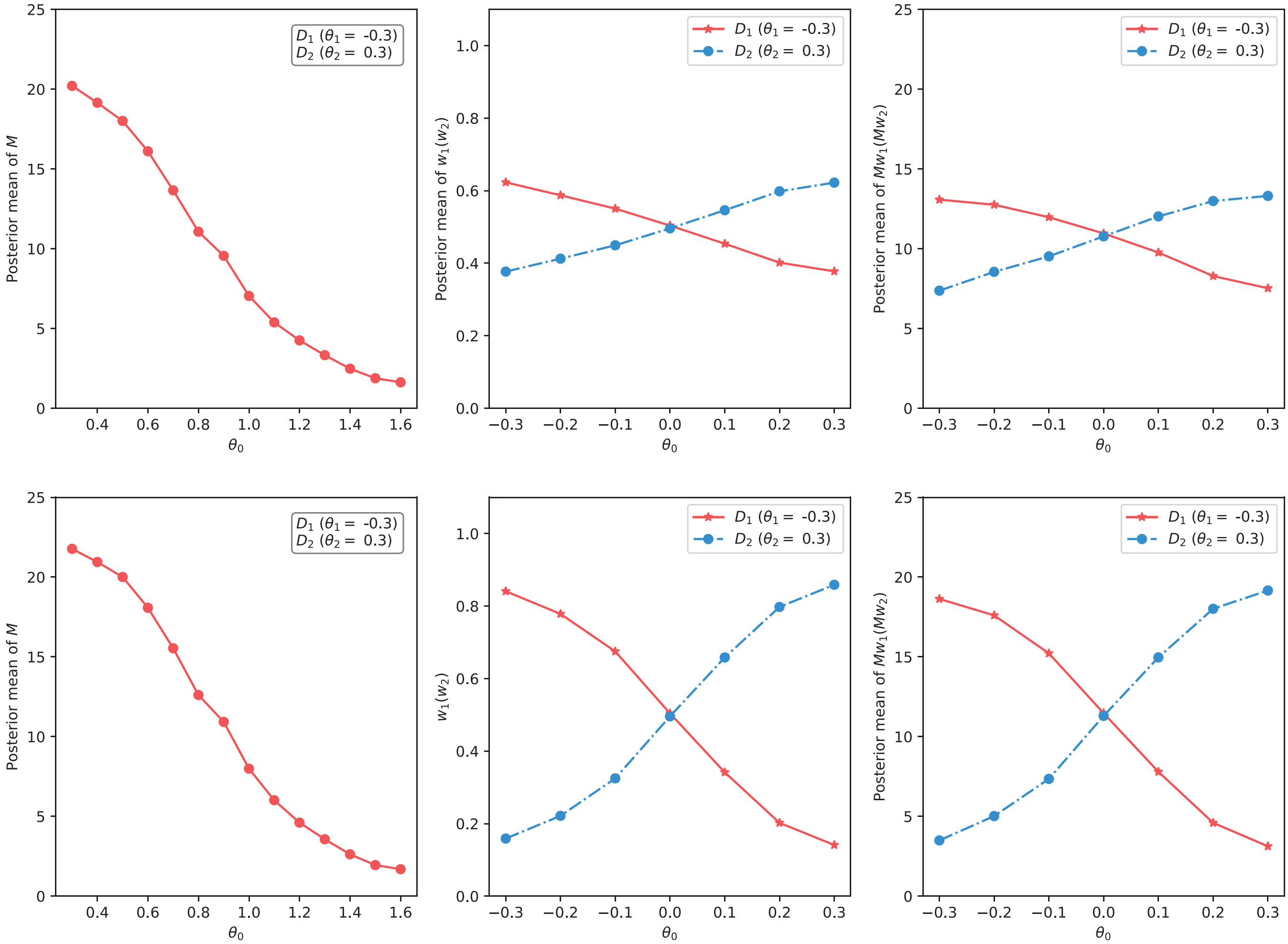

FIGURE 2 The trend of M (left panels) when the population mean of the current dataset θ_0 varies from 0.3 to 1.6, and the trends of w_1 and w_2 (middle panels) and those of Mw_1 and Mw_2 (right panels) when θ_0 varies from −0.3 to 0.3 under the UIP-Dirichlet (top) and UIP-JS (bottom) methods for continuous data with 100 repetitions [Colour figure can be viewed at wileyonlinelibrary.com]

is almost no information borrowed from the historical data. The experiments for the amount parameter M show that the UIP incorporates the historical information adaptively according to the overall consistency between the current and historical datasets.

We further utilize two historical datasets to investigate the trends of weight parameter w_k and Mw_k for the UIP-Dirichlet and UIP-JS methods. The historical data D_1 and D_2 are drawn from $N(-0.3, 1)$ and $N(0.3, 1)$, respectively, with sample sizes $n_1 = n_2 = 40$ while the population mean of D with $n = 40$ varies from −0.3 to 0.3 with a fixed standard deviation $\sigma = 1$. The hyper-prior for M is set as $M \sim \text{Uniform}(0, 40)$. We replicate the experiment for 100 times to draw the plots of the averages of estimates of weight parameters w_k's and Mw_k's. The right column of Figure 2 demonstrates that when the population mean of the current dataset θ_0 is closer to that of D_1, Mw_1 is larger, indicating more information is borrowed from D_1 compared with D_2; and a similar trend is observed for Mw_2. The tendency of the weight parameters w_k is similar to that of Mw_k. It is reasonable because the overall level of consistency between the current and historical datasets remains approximately the same when varying θ_0 between θ_1 and θ_2. It is also worth noting that the distinction of weight parameters under the UIP-JS method is larger than that under the UIP-Dirichlet method. For example, when $\theta_0 = 0.3$, the UIP-JS method assigns a weight of almost 0.8 to the historical dataset D_1 (whose population mean is also 0.3), while the weight parameter of D_1 under the UIP-Dirichlet method is around 0.6, which is less extreme than UIP-JS.

The results for binary data are presented in Web Figure 2 of the Supplementary Material, where similar phenomena can be observed.

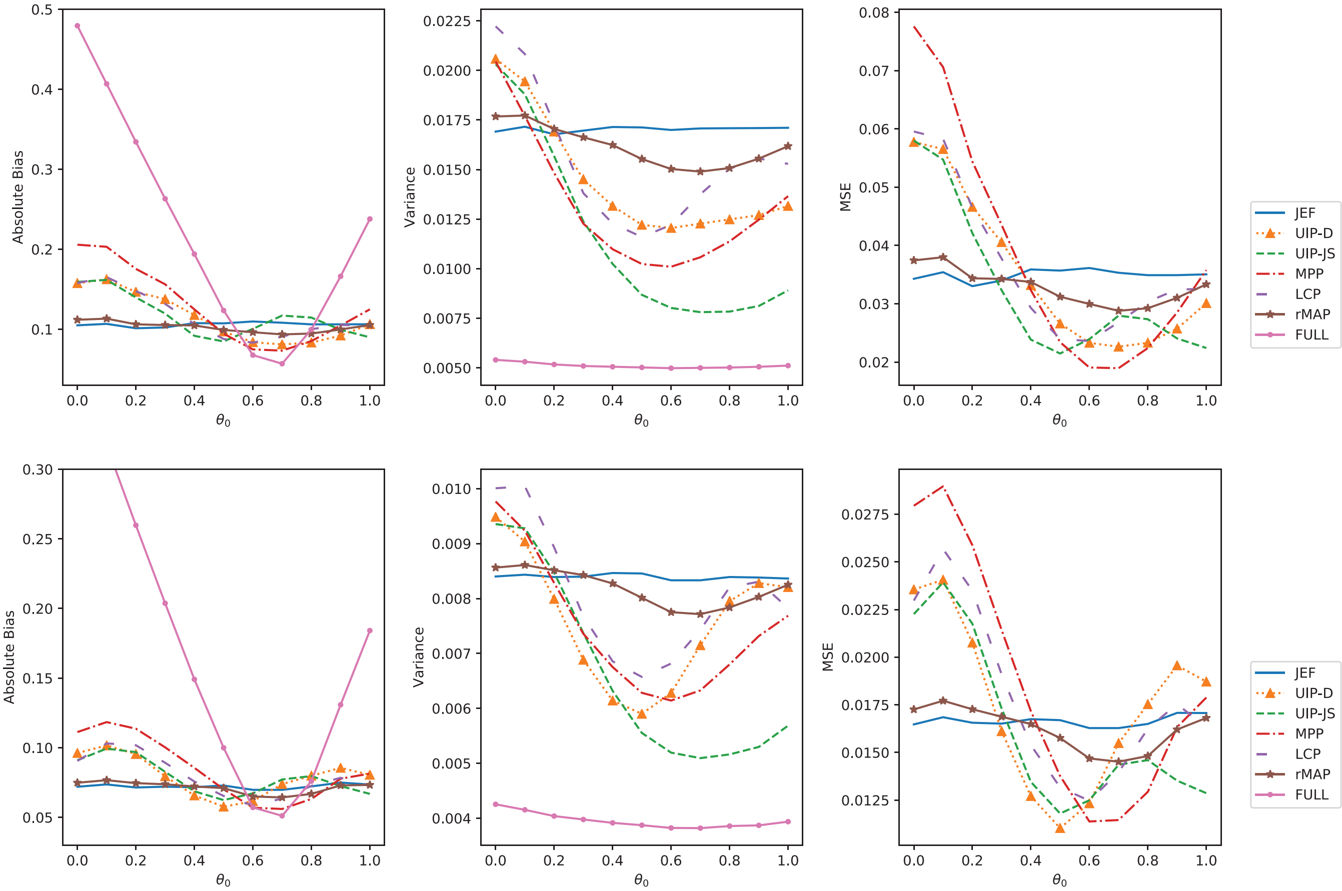

FIGURE 3 The absolute bias, variance, and mean squared error (MSE) when varying θ_0 from 0 to 1 under Jeffreys' prior (JEF), UIP-Dirichlet (UIP-D), UIP-JS, MPP, LCP, rMAP, and full-borrowing (FULL) methods with sample size $n = 60$ (top) and $n = 120$ (bottom) for continuous data [Colour figure can be viewed at wileyonlinelibrary.com]

4.3 | Single-arm trial scenario

We further compare our UIPs with Jeffreys' prior, the full-borrowing method, MPP, LCP, and rMAP priors for continuous data. We consider two historical datasets with sample sizes $n_1 = 100$ and $n_2 = 50$: D_1 from $N(0.5, 1)$ and D_2 from $N(1, 1)$. The variance for the current dataset D is also fixed as 1.

To assess the performance of different methods, we show the absolute biases, variances, and MSEs in Figure 3 when varying the mean parameter of the current dataset θ_0 from 0 to 1.0 for $n = 60$ and 120, respectively. The bias of θ, defined as $\mathrm{E}\{(\theta - \theta_0)|D\}$, measures the accuracy for the posterior mean of θ. The variance, denoted by $\mathrm{Var}(\theta|D)$, measures the precision of the posterior distribution. The MSE compromises both the accuracy and precision of the posterior distribution, which is defined as $\mathrm{E}\left\{(\theta - \theta_0)^2|D\right\}$. We omit the MSE curve for the full-borrowing method in order to display other MSE curves better. We replicate 1000 experiments and take the average for each metric.

All five informative priors show better performances when the historical datasets are more consistent with the current dataset. Among them, the rMAP prior yields the most robust results in terms of all three metrics. However, when the mean parameter of the current dataset is close to the counterparts of the historical datasets, the variance of rMAP is only slightly smaller than that under Jeffreys' prior. It illustrates that the rMAP prior tends to be too conservative to borrow enough information in some cases. The other four informative priors have a similar trend for $\theta_0 \leq 0.5$, that is, they all borrow information more aggressively when the historical datasets are coherent with the current one, yet sacrifice the robustness. When $\theta_0 \in [0.5, 1]$, the UIP-JS method has the relatively better performance as it yields comparable absolute bias and relatively lower variance among the UIP-Dirichlet, UIP-JS, MPP, and LCP methods. In terms of MSE, when sample size is small ($n = 60$), all five informative priors perform better than Jeffreys' prior for $\theta_0 > 0.4$. However, if we increase the sample size to $n = 120$, as more information is available for Jeffreys' prior, it shows lower MSE than the

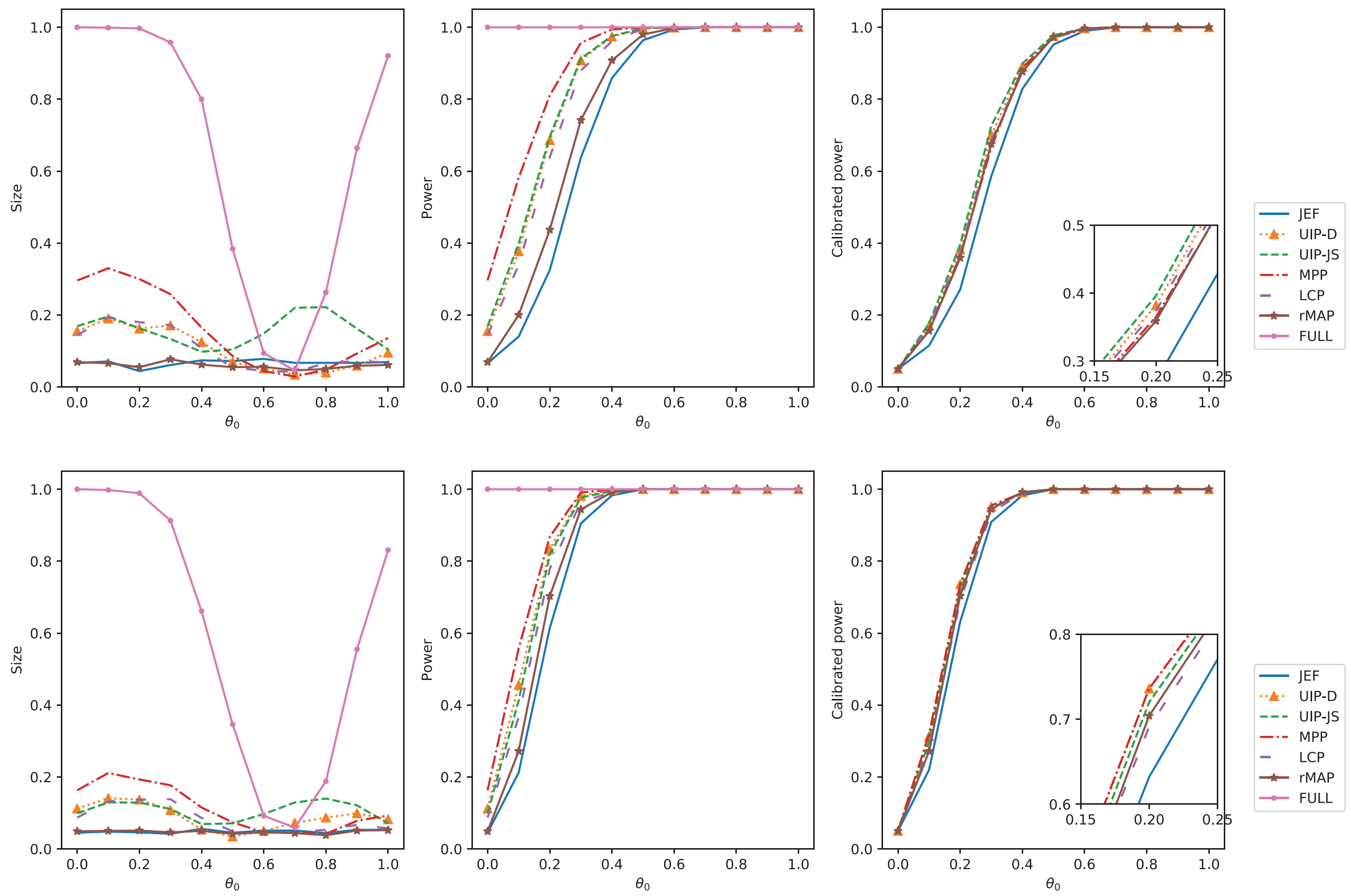

FIGURE 4 The size, power, and calibrated power of hypothesis testing when varying the true mean θ_0 from 0 to 1 for continuous data with sample size $n = 60$ (top) and $n = 120$ (bottom) under Jeffreys' prior (JEF), UIP-Dirichlet (UIP-D), UIP-JS, MPP, LCP, rMAP, and full-borrowing (FULL) methods; left panels: the test size under $H_0 : \theta = \theta_0$; middle panels: power under $H_0 : \theta = 0$.; right panels: the calibrated power under $H_0 : \theta = 0$ when controlling the test size at 0.05 [Colour figure can be viewed at wileyonlinelibrary.com]

UIP-Dirichlet, MPP, and LCP methods around $\theta_0 = 1$. It is also worth noting that the UIP-JS method enjoys consistently lower MSE compared with Jeffreys' prior and rMAP under both sample sizes for $\theta_0 > 0.4$.

We further conduct hypothesis testing for $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ using the 95% equal-tailed credible interval (CI), that is, if the CI does not contain θ_0, we reject H_0. In the left column of Figure 4, we present the sizes when varying the mean parameter for the current dataset from 0 to 1.0 under $n = 60{,}120$ with 1000 repetitions. In terms of the type I error, all five informative priors are robust compared with the full-borrowing method. The MPP method has the largest type I error among the five informative priors for $\theta_0 \leq 0.5$. For $\theta_0 \in [0.5, 1]$, the UIP-JS shows higher type I error while the other informative priors perform similarly to Jeffreys' prior. The results of the rMAP prior are most close to those of Jeffreys' prior for all values of the θ_0's. We show the power in the middle column of Figure 4 under the hypothesis test $H_0 : \theta = 0$ when varying the true mean θ_0 from 0 to 1.0. The rMAP prior has the lowest power among the five informative priors, as it is conservative in borrowing information from historical datasets. While yielding the largest type I error, the MPP method also has largest power for $\theta_0 \in [0, 0.5]$. For $\theta_0 \in [0.5, 1]$, all the methods yield power close to 1.

To make a fair comparison in power, we recalibrate the test size of $H_0 : \theta = 0$ for all seven methods to be 0.05 (ie, adjust the coverage probability of CI for the hypothesis test) and present the power curves under $H_0 : \theta = 0$ in the right column of Figure 4. For continuous data, it is impossible to control the size at 0.05 for the full-borrowing method, which is thus omitted. After calibration, there are significant gaps between the informative priors and Jeffreys' prior, revealing that all the informative priors gain information from historical datasets. The rMAP prior has the lowest calibrated power, which is consistent with the observation that it is conservative in borrowing information. In terms of the calibrated power, the UIP methods are consistently better than rMAP. Under the small sample size ($n = 60$), the UIP-JS method has the largest calibrated power, while the MPP and UIP-Dirichlet methods lead to the best performance for $n = 120$.

The test sizes under the MPP, LCP, and UIP methods are significantly inflated compared with that under Jeffreys' prior. In fact, size inflation is common for informative priors and a similar phenomenon is observed in the simulations by other works.[9,10,14] When incorporating information from historical data to increase the power, it also tends to inflate the test size. To solve this issue, we can calibrate the size by enlarging the coverage probability of the credible interval. In real data application, the calibration can be implemented by resampling methods, for example, bootstrap or permutation, to reconstruct the null distribution. In the simulation studies, all the MPP, LCP, and UIP methods yield significantly higher power even after calibration of the size compared with Jeffreys' prior, which demonstrates that these informative priors can effectively borrow information from the historical data. Moreover, not only are informative priors used for frequenstist hypothesis testing, but they can also help to estimate the parameter of interest. Figure 3 shows that when the historical datasets do not deviate dramatically from the current dataset, informative priors can improve the parameter estimates by reducing MSE compared with Jeffreys' prior.

5 | APPLICATION

As an illustration, we apply the UIP-Dirichlet and UIP-JS methods to six phase III clinical trials to investigate the efficacy of memantine in Alzheimer's disease (AD).[22,23] All the six trials were double-blind and placebo-controlled. Among the six trials, only the trial MRZ-9605[24] had a treatment period of 28 weeks, while others had a duration of 24 weeks. Trials MEM-MD-02[25] and MEM-MD-12[26] took memantine as an add-on therapy in patients who already received acetylcholinesterase inhibitors (AChEIs) and other trials assessed memantine as a monotherapy. Trials MRZ-9605, MEM-MD-01,[27] and MEM-MD-02 recruited patients with moderately severe to severe AD, while trials LU-99679,[28] MEM-MD-10,[29] and MEM-MD-12 enrolled patients with mild to moderate AD. The severity of AD was defined by scores of the mini-mental state exam (MMSE).

In our analysis, we regard the trial MEM-MD-12 as the current study and the remaining trials as historical studies. The efficacy of the behavioral domain could be measured at the end of the trial by the change of the neuropsychiatric inventory (NPI) score from the baseline, and a decrease in the NPI score indicated clinical improvement.[30] Among the six historical trials, the results of LU-99679 and MEM-MD-12 appeared to be similar in comparison with other trials. For trials LU-99679 and MEM-MD-12, the changes in the NPI scores indicated that the efficacy of memantine was inferior to the placebo in the behavioral domain, while the rest of the trials demonstrated the opposite results.

To analyze the changes in the NPI scores in the control and treatment groups, it is natural to fit a linear model by regressing the change in the NPI scores Y_i on the group indicator X_i. However, based on previous studies,[31,32] the NPI scores do not conform to a Gaussian distribution, while the normality of the change in the NPI scores cannot be assessed as we only have summary statistics of the datasets. To be conservative, instead of modeling the patient-level data, we choose to model the sample mean of each group. As the sample size of each of the six studies is large, the normality of the sample mean is guaranteed by the central limit theorem.

Specifically, let Y_i be the change in the NPI score for patient i, X_i be an indicator variable taking a value of 1 if patient i received memantine, and 0 otherwise, $\overline{Y}_T$ and $\overline{Y}_C$ be the sample means of Y_i's for the treatment and control groups, respectively and (n_T, n_C) are the corresponding sample sizes. We assume

$$\mathbb{E}[Y_i] = \beta_0 + \beta_1 X_i, \quad \mathrm{Var}(Y_i) = \sigma^2,$$
$$\overline{Y}_T \sim N(\beta_0 + \beta_1, \sigma^2/n_T), \quad \overline{Y}_C \sim N(\beta_0, \sigma^2/n_C).$$

Our goal is to determine whether memantine is superior to placebo in the behavioral domain, that is, whether β_1 is significantly smaller than 0. The parameter σ^2 is the nuisance parameter, which is replaced by the sample variance in all the analyses. We first analyze the data of the six trials separately by classical Bayesian linear models with non-informative priors for β_0 and β_1, that is, $\beta_0 \sim N(0, 10^2)$ and $\beta_1 \sim N(0, 10^2)$.

The results in Table 1 show that among all the six studies, MEM-MD-02 is the only trial with a statistically significant result as its upper bound of the 95% equal-tailed CI for β_1 is below 0. The estimates of β_1 in trials LU-99679 and MEM-MD-12 are positive and close to each other, and thus we expect that more information would be borrowed from LU-99679 compared with other historical datasets.

TABLE 1 Posterior mean estimates and 95% credible intervals (CIs) under non-informative priors for each of the six trials with the sample size of each trial and the corresponding weight assigned to each dataset using UIP-Dirichlet and UIP-JS

		β_0		β_1		UIP-Dirichlet	UIP-JS
Trials	**Sample size**	**Estimate**	**95% CI**	**Estimate**	**95% CI**	**Weight**	**Weight**
MEM-MD-12	261	0.881	(−1.055, 2.798)	0.069	(−2.579, 2.663)	1	1
LU-99679	210	−2.146	(−4.593, 0.275)	1.772	(−1.149, 4.712)	0.237	0.357
MEM-MD-01	260	0.423	(−2.072, 2.923)	−2.473	(−5.966, 0.958)	0.217	0.159
MEM-MD-02	323	2.698	(0.720, 4.644)	−3.412	(−6.086, −0.748)	0.201	0.073
MEM-MD-10	225	2.739	(0.228, 5.288)	−1.924	(−5.485, 1.697)	0.196	0.256
MRZ-9605	181	2.741	(−0.568, 6.131)	−2.565	(−7.128, 1.915)	0.148	0.155

TABLE 2 Posterior mean estimates and 95% credible intervals (CIs) using the UIP-Dirichlet, UIP-JS, MPP, LCP, and rMAP methods for the current study MEM-MD-12 by borrowing information from five historical trials LU-99679, MEM-MD-01, MEM-MD-02, MEM-MD-10, and MRZ-9605

Parameters		**UIP-Dirichlet**	**UIP-JS**	**MPP**	**LCP**	**rMAP**
β_0	Estimate	1.237	1.129	0.857	1.187	1.163
	95% CI	(−0.559, 3.008)	(−0.587, 2.805)	(−1.059, 2.800)	(−0.490, 2.886)	(−0.690, 2.943)
β_1	Estimate	−0.623	−0.406	−0.828	−0.511	−0.482
	95% CI	(−2.774, 1.643)	(−2.311, 1.663)	(−3.084, 1.421)	(−2.691, 1.472)	(−2.631, 2.054)

We apply the UIP-Dirichlet and UIP-JS methods to analyze the data from MEM-MD-12 by incorporating the five historical datasets in the prior, in comparison with MPP, LCP, and rMAP. As our main interest focuses on the parameter β_1, we only impose an informative prior on β_1 while we adopt non-informative priors for the other parameters. To prevent the historical data from overwhelming the current one, we set $M \sim \text{Uniform}(0, n)$ as the hyper-prior for the total amount parameter M for the UIP methods, where n is the sample size of the current trial MEM-MD-12.

As shown in Table 2, all the five informative priors demonstrate the ability of adaptively borrowing information for β_1 from historical data, because the 95% CIs of β_1 are narrower than those without any prior information in Table 1. The UIP-Dirichlet, UIP-JS, and LCP methods yield similar results in terms of β_0 and β_1. Among the five informative priors, the 95% CI using the rMAP prior is the widest, as the rMAP prior is more conservative in borrowing information. Furthermore, even when we leverage the same non-informative prior for β_0, the 95% CIs of β_0 under UIP, LCP, and rMAP are narrower than those in Table 1, while for MPP, the estimate and 95% CI of β_0 are essentially unchanged compared with the original results for MEM-MD-12 in Table 1.

For the total amount parameter M, the UIP-Dirichlet and UIP-JS methods lead to comparable results, $M =$ 137 vs 146, indicating intermediate borrowing of the historical data compared with the sample size of the current data 261. Nonetheless, the weight parameters of the two methods are slightly different. The weight parameters under the UIP-Dirichlet method are $(0.239, 0.217, 0.201, 0.196, 0.148)$ for the five trials LU-99679, MEM-MD-01, MEM-MD-02, MEM-MD-10, and MRZ-9605, respectively, while those under the UIP-JS method are $(0.357, 0.159, 0.073, 0.256, 0.155)$. As expected, both methods assign notably larger weights to the trial LU-99679 compared with other historical datasets, while the distinction of the weight parameters under UIP-JS is larger than that under UIP-Dirichlet.

In summary, under all the five informative priors, although the CIs of β_1 become narrower, they still cover 0. Thus, in terms of the NPI score, the efficacy of memantine in the behavioral domain is not shown to be superior to that of placebo, which is consistent with the original conclusion in Porsteinsson et al.[26]

The statistics of the NPI scores for both the memantine and placebo groups of the six trials are presented in Web Table 1 of the Supplementary Material, which also contains more information on the numerical studies.

6 | DISCUSSION

We propose an adaptively informative prior using historical data, which is elicited from an information perspective. We demonstrate that the UIP framework has many similarities to other commonly used adaptive priors and yields comparable performances. The proposed UIP methods are easy to implement for multiple historical datasets, whose parameters have intuitive interpretations. The weight parameters w_k can be interpreted as the relative importance of the historical datasets through competition against each other. The amount parameter M reveals the total units of information contained in the prior. For both binary and continuous data, we show that the amount parameter M typically has a comparable value with the prior ESS defined by Morita et al.[16]

The UIP method would be useful in the clinical trial field, as it is not uncommon to find multiple related trials for any ongoing study, especially, for the control arm (eg, the standard of care). While we mainly illustrate the UIP under the single-arm trial case, it is also extended to the linear model settings. In practice, it is typically not easy to obtain the patient-level historical data. An important feature of the UIP framework is that it does not need the patient-level historical data while some informative priors (eg, MPP and LCP methods) need such data. For example, in a study involving a linear regression model with multiple covariates, to adopt the UIP-Dirichlet method for the parameter of interest we only need the estimate of that parameter and its corresponding confidence interval which are commonly reported in publications of the historical study. However, as the MPP and LCP are derived from the likelihood, the complete patient-level historical data are required.

ACKNOWLEDGEMENTS

We would like to thank two anonymous referees and associate editor for their insightful suggestions in greatly improving the quality of this article. The research was supported by a grant (17308420) for Guosheng Yin from the Research Grants Council of Hong Kong.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the Supplementary Material of this article.

ORCID

Huaqing Jin https://orcid.org/0000-0002-1932-1336
Guosheng Yin https://orcid.org/0000-0003-3276-1392

REFERENCES

1. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non–small-cell lung cancer. *N Engl J Med*. 2015;373(2):123-135.
2. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. *N Engl J Med*. 2015;373(17):1627-1639.
3. Wu YL, Lu S, Cheng Y, et al. Nivolumab versus docetaxel in a predominantly Chinese patient population with previously treated advanced NSCLC: CheckMate 078 randomized phase III clinical trial. *J Thorac Oncol*. 2019;14(5):867-875.
4. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet*. 2017;389(10066):255-265.
5. Pocock SJ. The combination of randomized and historical controls in clinical trials. *J Chronic Dis*. 1976;29(3):175-188.
6. Ibrahim JG, Chen MH. Power prior distributions for regression models. *Stat Sci*. 2000;15(1):46-60.
7. Duan Y, Ye K, Smith EP. Evaluating water quality using power priors to incorporate historical information. *Environmetr Official J Int Environmetr Soc*. 2006;17(1):95-106.
8. Neuenschwander B, Branson M, Spiegelhalter DJ. A note on the power prior. *Stat Med*. 2009;28(28):3562-3566.
9. Banbeta A, van Rosmalen J, Dejardin D, Lesaffre E. Modified power prior with multiple historical trials for binary endpoints. *Stat Med*. 2019;38(7):1147-1169.
10. Gravestock I, Held L. Power priors based on multiple historical studies for binary outcomes. *Biom J*. 2019;61(5):1201-1218.
11. Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clin Trials*. 2010;7(1):5-18.
12. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*. 2014;70(4):1023-1032.
13. Weber S, Li Y, Seaman J, Kakizume T, Schmidli H. Applying meta-analytic predictive priors with the R Bayesian evidence synthesis tools; 2019. arXiv preprint arXiv:1907.00603.

14. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*. 2011;67(3):1047-1056.
15. Kass RE, Wasserman L. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J Am Stat Assoc*. 1995;90(431):928-934.
16. Morita S, Thall PF, Müller P. Determining the effective sample size of a parametric prior. *Biometrics*. 2008;64(2):595-602.
17. Morita S, Thall PF, Müller P. Prior effective sample size in conditionally independent hierarchical models. *Bayesian Anal*. 2012;7(3):591-614.
18. Hunter DR. *Notes for a Graduate-Level Course in Asymptotics for Statisticians*. Pennsylvania, PA: Penn State University; 2014.
19. Halonen B. Rate of Convergence of Maximum Likelihood Estimators under Relaxed Smoothness Conditions on the Likelihood Function [master's report]. Houghton, MI: Michigan Technological University; 2015.
20. Itzkovitz S, Hodis E, Segal E. Overlapping codes within protein-coding sequences. *Genome Res*. 2010;20(11):1582-1589.
21. Sims GE, Jun SR, Wu GA, Kim SH. Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. *Proc Natl Acad Sci*. 2009;106(8):2677-2682.
22. Winblad B, Jones RW, Wirth Y, Stöffler A, Möbius HJ. Memantine in moderate to severe Alzheimer's disease: a meta-analysis of randomised clinical trials. *Dement Geriatr Cogn Disord*. 2007;24(1):20-27.
23. Rive B, Gauthier S, Costello S, Marre C, François C. Synthesis and comparison of the meta-analyses evaluating the efficacy of memantine in moderate to severe stages of Alzheimer's disease. *CNS Drugs*. 2013;27(7):573-582.
24. Reisberg B, Doody R, Stöffler A, Schmitt F, Ferris S, Möbius HJ. Memantine in moderate-to-severe Alzheimer's disease. *N Engl J Med*. 2003;348(14):1333-1341.
25. Tariot PN, Farlow MR, Grossberg GT, et al. Memantine treatment in patients with moderate to severe Alzheimer disease already receiving donepezil: a randomized controlled trial. *J Am Med Assoc*. 2004;291(3):317-324.
26. Porsteinsson AP, Grossberg GT, Mintzer J, Olin JT. Memantine treatment in patients with mild to moderate Alzheimer's disease already receiving a cholinesterase inhibitor: a randomized, double-blind, placebo-controlled trial. *Curr Alzheimer Res*. 2008;5(1):83-89.
27. van Dyck CH, Tariot PN, Meyers B, Resnick EM. A 24-week randomized, controlled trial of memantine in patients with moderate-to-severe Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2007;21(2):136-143.
28. Bakchine S, Loft H. Memantine treatment in patients with mild to moderate Alzheimer's disease: results of a randomised, double-blind, placebo-controlled 6-month study. *J Alzheimers Dis*. 2008;13(1):97-107.
29. Peskind ER, Potkin SG, Pomara N, et al. Memantine treatment in mild to moderate Alzheimer disease: a 24-week randomized, controlled trial. *Am J Geriatr Psychiatry*. 2006;14(8):704-715.
30. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The neuropsychiatric inventory: comprehensive assessment of psychopathology in dementia. *Neurology*. 1994;44(12):2308-2308.
31. Mega MS, Masterman DM, O'Connor SM, Barclay TR, Cummings JL. The spectrum of behavioral responses to cholinesterase inhibitor therapy in Alzheimer disease. *Arch Neurol*. 1999;56(11):1388-1393.
32. Lai CK. The merits and problems of neuropsychiatric inventory as an assessment tool in people with dementia and other neurological disorders. *Clin Interv Aging*. 2014;9:1051.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Jin H, Yin G. Unit information prior for adaptive information borrowing from multiple historical datasets. *Statistics in Medicine*. 2021;1–16. https://doi.org/10.1002/sim.9146