Biometrics, 2024, 80(3), ujae091
https://doi.org/10.1093/biomtc/ujae091
Biometric Methodology

OXFORD

Unit information Dirichlet process prior

Jiaqi Gu [1,*] and Guosheng Yin [2]

[1]Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA 94304, United States, [2]Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong, 999077, China

*Corresponding author: Jiaqi Gu, Department of Neurology and Neurological Sciences, Stanford University, 453 Quarry Road, Stanford, CA 94304, USA (jiaqigu@stanford.edu).

ABSTRACT

Prior distributions, which represent one's belief in the distributions of unknown parameters before observing the data, impact Bayesian inference in a critical and fundamental way. With the ability to incorporate external information from expert opinions or historical datasets, the priors, if specified appropriately, can improve the statistical efficiency of Bayesian inference. In survival analysis, based on the concept of unit information (UI) under parametric models, we propose the unit information Dirichlet process (UIDP) as a new class of nonparametric priors for the underlying distribution of time-to-event data. By deriving the Fisher information in terms of the differential of the cumulative hazard function, the UIDP prior is formulated to match its prior UI with the weighted average of UI in historical datasets and thus can utilize both parametric and nonparametric information provided by historical datasets. With a Markov chain Monte Carlo algorithm, simulations and real data analysis demonstrate that the UIDP prior can adaptively borrow historical information and improve statistical efficiency in survival analysis.

KEYWORDS: Bayesian nonparametric; Fisher information; hazard function; Markov chain Monte Carlo; time-to-event data.

1 INTRODUCTION

Prior distributions play a crucial role in the paradigm of Bayesian inference. By incorporating external information from expert opinions or historical datasets, an elaborated informative prior can improve efficiency for statistical inference. In clinical trials, the effect of the same treatment may be investigated on patients of different ethnicity groups or different disease subtypes (Borghaei et al., 2015; Brahmer et al., 2015; Rittmeyer et al., 2017; Wu et al., 2019). Under similar experimental settings, a suitable prior that borrows information from historical datasets can alleviate the sample size requirement for achieving adequate power in the current trial. However, practitioners often face challenges in eliciting such an elaborated prior due to a lack of unified rules for determining the amount of information to be borrowed from multiple historical datasets. Typically, more information should be borrowed from historical datasets whose sample sizes are larger and data patterns are more similar to or commensurate with the current one. Another issue is how to control the total amount of information borrowed from historical datasets, which, in principle, should not be too large to overwhelm the current study, even when the total sample size of historical datasets is exceedingly large. In addition, as individual observations in historical datasets may not be accessible due to privacy protection or confidentiality, some historical information exists only in the form of summary statistics (eg, point estimates and confidence intervals of parameters) or curve estimates (eg, the Kaplan–Meier estimator for a survival curve or a cumulative incidence curve). It is thus desirable that the developed prior can incorporate both parametric and nonparametric historical information.

Several methods have been developed to use historical information for elaborating a prior. Pocock (1976) proposed a weighted estimator for the parameter of interest using the current and historical datasets, where weights are computed by modeling the differences of the parameter estimates across the current and historical datasets as a zero-mean random variable. The power prior proposed by Ibrahim and Chen (2000) discounts the likelihood functions of historical datasets with a power parameter in $[0, 1]$. The modified power prior is further developed to incorporate information from multiple historical datasets (Banbeta et al., 2019; Gravestock and Held, 2019). To account for the consistency between historical datasets and the current study, the meta-analytic-predictive prior (Neuenschwander et al., 2010; Schmidli et al., 2014) weighs likelihood functions of historical datasets according to the predictive distribution. The unit information (UI) prior (Jin and Yin, 2021) directly weighs the UI of historical datasets with respect to the current trial data. However, all the aforementioned methods require that the current dataset and historical datasets are inferred under the same parametric model. As a result, historical information provided by parameter estimation under different parametric models or nonparametric models cannot be incorporated, and the inference may suffer from model misspecification.

Following the definition of the UI for a parameter and the UI prior distribution (Kass and Wasserman, 1995; Jin and Yin, 2021) constructed from a historical dataset, we propose

Received: January 12, 2024; Revised: July 5, 2024; Accepted: August 21, 2024

the unit information Dirichlet process (UIDP) as a new class of priors for Bayesian nonparametric inference of time-to-event data. Based on the definition of the Dirichlet process (Ferguson, 1973; 1974), we derive the Fisher information under the Dirichlet process prior and historical datasets in terms of the cumulative hazard function (CHF), for which the independent increment property circumvents the major difficulty of constructing the nonparametric UI. The UIDP prior is formulated by matching its unit prior information with the weighted average of UI in historical datasets. Using the UIDP prior, both parametric and nonparametric information of historical datasets can be utilized to carry out Bayesian inference for the current time-to-event data. Unlike the work of Reimherr et al. (2021), where the effective sample size (ESS) of a prior is evaluated via frequentist assessment of a Bayesian estimator of the target parameter (Efron, 2015), ESS of the UIDP prior is selected via a cross-validation procedure to maximize the concordance between the prior and the current dataset. With a Markov chain Monte Carlo (MCMC) algorithm, extensive simulations and real data analysis show that the proposed UIDP prior can improve statistical efficiency in Bayesian nonparametric estimation of the cumulative distribution function (CDF) and adaptively borrow historical information according to the consistency or commensurability between the current dataset and historical datasets.

The remainder of this article is organized as follows. In Section 2, we propose the UIDP prior by deriving the information under the Dirichlet process prior of time-to-event data. We further develop an MCMC algorithm for making Bayesian nonparametric inference by borrowing historical information through the UIDP prior. The gain in statistical efficiency and adaptivity of information borrowing using the UIDP prior is investigated via extensive simulations in Section 3 and real data analysis in Section 4. Section 5 concludes with some discussion.

2 METHODOLOGY

2.1 Prior information under Dirichlet process

Consider a CDF, $F(x) = \mathbb{P}(X \le x)$, and let $dF(x)$ denote the differential of $F(x)$. In survival analysis, Càdlàg functions are commonly used, which are right-continuous with left limits. Let $F(x)$ be a Càdlàg function, and thus we define

$$dF(x) = F(x) - F(x-) \quad \text{and} \quad F(x-) = \mathbb{P}(X < x) = \lim_{u \nearrow x} F(u) \tag{1}$$

as the differential and the left limit of $F(\cdot)$ at x, respectively. When $F(x)$ is continuous, $dF(x)$ measures the probability in the infinitesimal interval $(x - dx, x)$; when $F(x)$ is discrete, $dF(x)$ measures the point mass probability at x. Let $\Lambda(x) = \int_{-\infty}^{x}\{1 - F(u-)\}^{-1} dF(u)$ be the CHF corresponding to $F(\cdot)$. To make Bayesian nonparametric inference on $F(\cdot)$, 1 popular way is to impose a Dirichlet process (DP; Ferguson, 1973, 1974) prior,

$$F \sim \mathrm{DP}(\alpha, G), \tag{2}$$

where $\alpha > 0$ is the concentration parameter measuring how close $F(\cdot)$ is to the base distribution $G(\cdot)$. The DP prior in Equation 2 renders that for any $-\infty < x_1 < x_2 < \cdots < x_{m-1} < x_m < \infty$, it holds that

$$\begin{aligned}&\Big(F(x_1), F(x_2) - F(x_1), \ldots, F(x_m) \\ &\quad -F(x_{m-1}), 1 - F(x_m)\Big) \\ &\sim \mathrm{Dir}\Big(\alpha G(x_1), \alpha\{G(x_2) - G(x_1)\}, \ldots, \alpha\{G(x_m) \\ &\quad -G(x_{m-1})\}, \alpha\{1 - G(x_m)\}\Big),\end{aligned} \tag{3}$$

which is the Dirichlet distribution with parameters $\alpha G(x_1), \alpha\{G(x_2) - G(x_1)\}, \ldots, \alpha\{G(x_m) - G(x_{m-1})\}$, $\alpha\{1 - G(x_m)\}$.

Remark 1 *Under the prior $DP(\alpha, G)$, F is a discrete distribution almost surely (Ferguson, 1973). If the base distribution G has probability point masses on $x_1, \ldots, x_m$, Equation 3 is well defined; if G has no probability point mass but non-zero densities on $x_1, \ldots, x_m$, $dF(x_1), \ldots, dF(x_m)$ are non-zero with infinitesimal probability.*

Analogously to Equation 1, we define

$$dG(x) = G(x) - G(x-), \quad \text{with } G(x-) = \lim_{u \nearrow x} G(u),$$

$$d\Lambda(x) = \Lambda(x) - \Lambda(x-) = \{1 - F(x-)\}^{-1} dF(x),$$

and present the properties of the CDF $F(x)$ as follows.

Theorem 1 *If $F \sim DP(\alpha, G)$, then for all $-\infty < x_1 < x_2 < \infty$, (i) the mean and covariance matrix of the vector $\{F(x_1-), F(x_2-), dF(x_1), dF(x_2)\}^\top$ are, respectively, given by*

$$\mathbb{E}\{F(x_1-), F(x_2-), dF(x_1), dF(x_2)\}^\top = \{G(x_1-), G(x_2-), dG(x_1), dG(x_2)\}^\top,$$

$$\mathrm{Var}\{F(x_1-), F(x_2-), dF(x_1), dF(x_2)\}^\top = \frac{1}{\alpha+1}$$

$$\times \begin{pmatrix} G(x_1-)\{1-G(x_1-)\} & G(x_1-)\{1-G(x_2-)\} & -G(x_1-)dG(x_1) & -G(x_1-)dG(x_2) \\ G(x_1-)\{1-G(x_2-)\} & G(x_2-)\{1-G(x_2-)\} & \{1-G(x_2-)\}dG(x_1) & -G(x_2-)dG(x_2) \\ -G(x_1-)dG(x_1) & \{1-G(x_2-)\}dG(x_1) & dG(x_1)\{1-dG(x_1)\} & -dG(x_1)dG(x_2) \\ -G(x_1-)dG(x_2) & -G(x_2-)dG(x_2) & -dG(x_1)dG(x_2) & dG(x_2)\{1-dG(x_2)\} \end{pmatrix}.$$

(ii) For the cumulative hazard function, we have

$$\mathbb{E}\{d\Lambda(x_1)\} = \{1 - G(x_1-)\}^{-1} dG(x_1), \quad (4)$$

$$\text{Var}\{d\Lambda(x_1)\} = \frac{dG(x_1)\{1 - G(x_1)\}}{\{1 - G(x_1-)\}^2(\alpha + 1)}, \quad (5)$$

$$\text{Cov}\{d\Lambda(x_1), d\Lambda(x_2)\} = 0. \quad (6)$$

Theorem 1 provides the joint distribution of probabilities in intervals $(-\infty, x_1)$ and $(-\infty, x_2)$ and point mass probabilities at x_1 and x_2, for which the proof is provided in Section A1 of the Appendix. By the stick-breaking construction of DP (Ferguson, 1973), the distribution $F \sim \text{DP}(\alpha, G)$ is discrete almost surely. If the base distribution G is continuous at point x, $dF(x)$ is infinitesimal almost surely and so is $d\Lambda(x)$. If the base distribution G has non-zero point mass probability at point x, $dF(x)$ is a non-zero point mass probability almost surely and so is $d\Lambda(x)$. More importantly, by Equation 6, the increments of CHF $\Lambda(x)$ at different points x_1 and x_2 are independent, no matter whether the base distribution G is continuous or not at x_1 and x_2. Together with the prior variance in Equation 5, we can derive the prior information of $d\Lambda(x)$ under the DP prior in Equation 2 as follows:

$$\mathcal{I}\{d\Lambda(x)\} = [\text{Var}\{d\Lambda(x)\}]^{-1} = \frac{\{1 - G(x-)\}^2(\alpha + 1)}{dG(x)\{1 - G(x)\}}, \quad (7)$$

which can serve as the medium for information borrowing. Examples of $\mathbb{E}\{d\Lambda(x)\}$ and $\mathcal{I}\{d\Lambda(x)\}$ under different base functions $G(x)$ are given in Section A2 of the Appendix.

2.2 Unit information Dirichlet process

Suppose that there are K historical datasets $\mathcal{D}_1, \ldots, \mathcal{D}_K$ with corresponding sample sizes of $n_1, \ldots, n_K$, which are potentially related to the dataset of the current study $\mathcal{D}$. Let $\widehat{F}_k(x)$ and $\widehat{\Lambda}_k(x)$ denote the estimators of CDF $F_k(x)$ and CHF $\Lambda_k(x)$ under historical dataset $\mathcal{D}_k$, respectively. We define the estimated UI for $\Lambda_k(x)$ under $\mathcal{D}_k$ as

$$\widehat{\mathcal{I}}_U\{d\Lambda_k(x)\} = I(d\widehat{\Lambda}_k(x) > 0) \cdot [Y_k(x)\widehat{\text{Var}}\{d\widehat{\Lambda}_k(x)\}]^{-1},$$

where $Y_k(x)$ is the number of observations in $\mathcal{D}_k$ that contribute to $d\widehat{\Lambda}_k(x)$, and $I(\cdot)$ is the indicator function. In particular, we do not discriminate whether $\widehat{F}_k(x)$ and $\widehat{\Lambda}_k(x)$ are parametric or nonparametric estimators. For parametric historical information, if we take the exponential distribution as an example, we have $d\widehat{\Lambda}_k(x) = \widehat{\theta}_k dx$ as shown below, and thus all n_k observations in $\mathcal{D}_k$ contribute to the estimation of θ_k and $Y_k(x) = n_k$. For nonparametric historical information, as the calculation $d\widehat{\Lambda}_k(x)$ only requires the information of all observations in $\mathcal{D}_k$ with $Z_{ik} \geq x$, we have $Y_k(x) = \sum_{i=1}^{n_k} I(Z_{ik} \geq x)$.

For illustration, we consider historical datasets $\mathcal{D}_k = \{(Z_{ik}, \Delta_{ik}); i = 1, \ldots, n_k\}$ $(k = 1, \ldots, K)$, where $Z_{ik} = \min(X_{ik}, C_{ik})$ and $\Delta_{ik} = I(X_{ik} \leq C_{ik})$ are, respectively, the observed time and the censoring indicator obtained by the failure time X_{ik} and the censoring time C_{ik} of subject i in study k. Under the independent censoring assumption that X_{ik} and C_{ik} are mutually independent in each study k, we can derive $d\widehat{\Lambda}_k(x)$ and $\widehat{\mathcal{I}}_U\{d\Lambda_k(x)\}$ under a parametric model and a nonparametric model, respectively.

- **Parametric model:** If an exponential distribution $F_k(x) = 1 - \exp(-\theta_k x)$ $(0 < x < \infty)$ is fitted to the historical dataset $\mathcal{D}_k$, the maximum likelihood estimator and the estimated information of θ_k are, respectively, given by

$$\widehat{\theta}_k = \frac{\sum_{i=1}^{n_k} \Delta_{ik}}{\sum_{i=1}^{n_k} Z_{ik}} \quad \text{and} \quad \widehat{\mathcal{I}}(\widehat{\theta}_k) = \sum_{i=1}^{n_k} \left(\frac{\Delta_{ik}}{\widehat{\theta}_k} - Z_{ik}\right)^2.$$

Consequently, we can obtain $d\widehat{\Lambda}_k(x)$ and $\widehat{\mathcal{I}}_U\{d\widehat{\Lambda}_k(x)\}$ as follows:

$$d\widehat{\Lambda}_k(x) = \widehat{\theta}_k dx \quad \text{and}$$

$$\widehat{\mathcal{I}}_U\{d\widehat{\Lambda}_k(x)\} = \frac{1}{Y_k(x)dx} \sum_{i=1}^{n_k} \left(\frac{\Delta_{ik}}{\widehat{\theta}_k} - Z_{ik}\right)^2,$$

where $Y_k(x) = n_k$ for all $0 < x < \infty$.

- **Nonparametric model:** Without loss of generality, we assume that $Z_{(1)k} < \cdots < Z_{(M_k)k}$ are the M_k distinct values of event times in $\mathcal{D}_k$. Let $\widehat{F}_k(x)$ be the Kaplan–Meier estimator for $\mathcal{D}_k$,

$$\widehat{F}_k(x) = 1 - \prod_{m: Z_{(m)k} \leq x} \left\{1 - \frac{\sum_{i=1}^{n_k} I(Z_{ik} = Z_{(m)k}, \Delta_{ik} = 1)}{\sum_{i=1}^{n_k} I(Z_{ik} \geq Z_{(m)k})}\right\}.$$

Thus, we have $d\widehat{\Lambda}_k(x) = \left\{\sum_{i=1}^{n_k} I(Z_{ik} = x, \Delta_{ik} = 1)\right\} / Y_k(x)$, where $Y_k(x) = \sum_{i=1}^{n_k} I(Z_{ik} \geq x)$, and

$$\widehat{\mathcal{I}}_U\{d\widehat{\Lambda}_k(x)\} = \begin{cases} \dfrac{1}{Y_k(x)d\widehat{\Lambda}_k(x)\{1 - d\widehat{\Lambda}_k(x)\}}, & \text{if } x \in \{Z_{(1)k}, \ldots, Z_{(M_k)k}\}, \\ 0, & \text{otherwise.} \end{cases}$$

To borrow information from historical datasets $\mathcal{D}_1, \ldots, \mathcal{D}_K$, we design weights $w_1, \ldots, w_K \in (0, 1)$ satisfying $\sum_{k=1}^{K} w_k = 1$, to characterize the contributions from $\mathcal{D}_1, \ldots, \mathcal{D}_K$. We let the prior of $\Lambda(x)$ concentrate at the weighted average of its empirical counterparts $\widehat{\Lambda}_1(x), \ldots, \widehat{\Lambda}_K(x)$ with the prior information approximated by the weighted average of $\widehat{\mathcal{I}}_U\{d\widehat{\Lambda}_1(x)\}, \ldots, \widehat{\mathcal{I}}_U\{d\widehat{\Lambda}_K(x)\}$. The UIDP prior of $F(x)$ is then formulated as follows:

$$F | M, w_1, \ldots, w_K, \mathcal{D}_1, \ldots, \mathcal{D}_K \sim \text{DP}(\alpha, G),$$

$$\text{with} \quad \mathbb{E}\{d\Lambda(x)\} = \sum_{k=1}^{K} w_k d\widehat{\Lambda}_k(x),$$

$$\mathcal{I}\{d\Lambda(x)\} = M \sum_{k=1}^{K} w_k \widehat{\mathcal{I}}_U\{d\widehat{\Lambda}_k(x)\}, \quad (8)$$

where M is the ESS corresponding to the total number of units borrowed from historical datasets. From Equations 4, 7, and 8, we can obtain the matching relations,

$$\frac{dG(x)}{1-G(x-)} = \sum_{k=1}^{K} w_k d\widehat{\Lambda}_k(x), \tag{9}$$

$$\frac{\{1-G(x-)\}^2(\alpha+1)}{dG(x)\{1-G(x)\}} = M\sum_{k=1}^{K} w_k \widehat{\mathcal{I}}_U\{d\widehat{\Lambda}_k(x)\}. \tag{10}$$

Although we can directly obtain $G(x)$ from Equation 9 that

$$G(x) = \begin{cases} 1-\exp[-\int_{-\infty}^{x}\sum_{k=1}^{K} w_k d\widehat{\Lambda}_k(u)], & \text{if } \widehat{\Lambda}_1(x),\ldots,\widehat{\Lambda}_K(x) \text{ are continuous,} \\ 1-\exp\left\{\int_{-\infty}^{x}\log[1-\sum_{k=1}^{K} w_k d\widehat{\Lambda}_k(u)]\right\}, & \text{if } \widehat{\Lambda}_1(x),\ldots,\widehat{\Lambda}_K(x) \text{ are discrete.} \end{cases}$$

The concentration parameter α is over-identified if doing so and there exists no solution satisfying Equation 10. In addition, it is difficult to make inference when some of $\widehat{\Lambda}_1(x),\ldots,\widehat{\Lambda}_K(x)$ are continuous and others are discrete.

To tackle such difficulties, we suggest imposing a continuous parametric model $G_{\boldsymbol{\theta}}(x)$ as the base distribution in the DP prior (2) to perform approximate matching. Here, $G_{\boldsymbol{\theta}}(x)$ represents one's prior belief on the family of parametric models that $F(x)$ belongs to. For example, we may use Weibull distribution for continuous CDFs on $(0,\infty)$ or negative-binomial distribution for discrete CDFs. Substituting $G_{\boldsymbol{\theta}}(x)$ in the matching relations (9) and (10), we have

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}^*} \sum_{l=1}^{m}\left[\int_{x_{l-1}}^{x_l}\{1-G_{\boldsymbol{\theta}^*}(u-)\}^{-1}dG_{\boldsymbol{\theta}^*}(u) - \sum_{k=1}^{K} w_k \int_{x_{l-1}}^{x_l} d\widehat{\Lambda}_k(u)\right]^2 \tag{11}$$

and

$$\alpha = M\cdot\frac{\sum_{l=1}^{m}\int_{x_{l-1}}^{x_l}\{1-G_{\boldsymbol{\theta}}(u)\}\{1-G_{\boldsymbol{\theta}}(u-)\}^{-2}dG_{\boldsymbol{\theta}}(u)}{\sum_{l=1}^{m}[\sum_{k=1}^{K} w_k\widehat{\mathcal{I}}_U\{\int_{x_{l-1}}^{x_l} d\widehat{\Lambda}_k(u)\}]^{-1}} - 1, \tag{12}$$

where $-\infty \le x_0 < x_1 < \cdots < x_m < \infty$, and the interval points $\{x_0, x_1, x_2, \ldots, x_{m-1}, x_m\}$ control both the quality and robustness of matching in Equations 11 and 12. The value of x_0 should be as small as possible to incorporate the most historical information, while an extremely large x_m may result in too large $\{1-G_{\boldsymbol{\theta}}(u-)\}^{-1}$ in Equation 11 and cause breakdown of computation. In practice, we recommend using a set of interval points $\{x_0, x_1, \ldots, x_m\}$ such that

(i) $x_0 = \inf(\cup_{k=1}^{K}\{x : \widehat{F}_k(x) > 0\})$;
(ii) $m = \sqrt{\min_k Y_k(x_0)}$;
(iii) $x_m = \min_k(\sup\{x : \widehat{F}_k(x) \le 0.9\})$; and
(iv) the total number of historical observations that contribute to $d\widehat{\Lambda}_1(x_0),\ldots,d\widehat{\Lambda}_K(x_0)$ is the same for all intervals $(x_0, x_1],\ldots,(x_{m-1}, x_m]$.

2.3 Bayesian inference

Let X_i denote the failure time and C_i the censoring time of subject i in the current study. The time-to-event data $\mathcal{D} = \{(Z_i,\Delta_i); i = 1,\ldots,n\}$ include $Z_i = \min(X_i, C_i)$ and $\Delta_i = I(X_i \le C_i)$. For the K historical datasets, we pre-determine the ESS M using a cross-validation procedure, and then set the hyper-prior $(w_1,\ldots,w_K) \sim \text{Dir}(\gamma_1,\ldots,\gamma_K)$ with $\gamma_k = \min(1, n_k/n)$ $(k = 1,\ldots,K)$ and pre-specify $-\infty \le x_0 < x_1 < x_2 < \cdots < x_{m-1} < x_m < \infty$. We propose a MCMC algorithm to draw posterior samples of the CDF $F(x)$ as detailed in Algorithm A1 in the Appendix.

It remains a task for selecting an appropriate ESS M whose impact on the posterior inference is enormous. Ideally, when historical datasets $\mathcal{D}_1,\ldots,\mathcal{D}_K$ are informative to the current dataset $\mathcal{D}$, we should choose a larger M to improve the statistical efficiency, while M should be smaller when $\mathcal{D}$ deviates much from $\mathcal{D}_1,\ldots,\mathcal{D}_K$. To select an appropriate value for M, we propose a cross-validation procedure as follows:

- Set the candidate values $1 \le M_1 < \cdots < M_H \le \min\{n, \sum_{k=1}^{K} n_k\}$ for M.
- Partition the current data $\mathcal{D}$ into V folds.
- For $v = 1,\ldots,V$:
  - Obtain $\widehat{F}_{(v)}(x)$ as the nonparametric maximum likelihood estimator (NPMLE) of $F(x)$ under the vth fold $\mathcal{D}_{(v)}$.
  - For $h = 1,\ldots,H$, apply Algorithm A1 with M_h to $\mathcal{D}_{(-v)}$, the current dataset excluding the vth fold, and obtain B posterior samples $F_{h(-v)}^{(1)}(x),\ldots,F_{h(-v)}^{(B)}(x)$.
  - Compute the Wilcoxon-type statistics as

    $$W_{h(v)} = \frac{1}{B}\sum_{b=1}^{B} W_{h(v)}^{(b)},$$

    where $W_{h(v)}^{(b)} = \Pr(U_1 < U_2) - \Pr(U_1 > U_2)$ with $U_1 \sim \widehat{F}_{(v)}$, $U_2 \sim F_{h(-v)}^{(b)}$. The detailed computation procedure is presented in Section A4 of the Appendix.
- Select $M = M_{h^*}$ where

  $$h^* = \arg\min_h \frac{1}{V}\sum_{v=1}^{V}|W_{h(v)}|.$$

This cross-validation procedure allows us to select the optimal M that minimizes the deviation,

$$|\Pr(U_1 < U_2) - \Pr(U_1 > U_2)|, \text{ with } U_1 \sim F,\ U_2 \sim G,$$

between $F(x)$ and the prior mean $G(x)$. Specifically, throughout the experiments in Sections 3 and 4, we use the candidate values $M_1 = 1$ and $(M_2, \ldots, M_{11}) = (0.1, 0.2, \ldots, 1) \times \min\{n, \sum_{k=1}^{K} n_k\}$. We prefer cross validation over a fully Bayesian approaches in selecting M because a fully Bayesian approach presumes the number of distinct values in the target dataset follows Chinese restaurant table distribution (which depicts the distribution of the number of tables under the Chinese restaurant process) with the parameter α under the DP prior (2). However, in practice, time-to-event observations $Z_1, \ldots, Z_n$ are usually distinct, often leading to too much historical information being borrowed. To mitigate the situation where the historical information is not helpful, we propose using the cross-validation method to choose M that maximizes the consistency between the target data and the posterior samples of $F(x)$. With the selected M, we then apply Algorithm A1 to the complete current data $\mathcal{D}$ to draw posterior samples $F^{(1)}(x), \ldots, F^{(B)}(x)$.

Among all possible ways to compute the Bayesian estimator of $F(x)$, the most straightforward estimator is the posterior mean $B^{-1}\sum_{b=1}^{B} F^{(b)}(x)$, which, however, is not the most efficient. Instead, we utilize the Rao–Blackwell theorem and compute the Bayesian estimator based on $(\alpha^{(1)}, \boldsymbol{\theta}^{(1)}), \ldots, (\alpha^{(B)}, \boldsymbol{\theta}^{(B)})$ as follows:

$$\widehat{F}_{\text{Bayes}}(x) = B^{-1}\sum_{b=1}^{B} \mathbb{E}\{F^{(b)}(x)|\mathcal{D}, \alpha^{(b)}, \boldsymbol{\theta}^{(b)}\}. \tag{13}$$

It can be shown that the 2 estimators give the same mean,

$$\mathbb{E}\left\{B^{-1}\sum_{b=1}^{B} F^{(b)}(x)\middle|\mathcal{D}\right\} = B^{-1}\sum_{b=1}^{B}\mathbb{E}\{F^{(b)}(x)|\mathcal{D}\}$$
$$= B^{-1}\sum_{b=1}^{B}\mathbb{E}[\mathbb{E}\{F^{(b)}(x)|\mathcal{D}, \alpha^{(b)}, \boldsymbol{\theta}^{(b)}\}|\mathcal{D}]$$
$$= \mathbb{E}\{\widehat{F}_{\text{Bayes}}(x)|\mathcal{D}\},$$

but the Rao–Blackwell Bayesian estimator yields a smaller variance,

$$\text{Var}\left\{B^{-1}\sum_{b=1}^{B} F^{(b)}(x)\middle|\mathcal{D}\right\}$$
$$= \text{E}\left[\text{Var}\left\{B^{-1}\sum_{b=1}^{B} F^{(b)}(x)\middle|\mathcal{D}, \alpha^{(b)}, \boldsymbol{\theta}^{(b)}\right\}\middle|\mathcal{D}\right]$$
$$+\text{Var}\left[\text{E}\left\{B^{-1}\sum_{b=1}^{B} F^{(b)}(x)\middle|\mathcal{D}, \alpha^{(b)}, \boldsymbol{\theta}^{(b)}\right\}\middle|\mathcal{D}\right]$$
$$\geq \text{Var}\left[\text{E}\left\{B^{-1}\sum_{b=1}^{B} F^{(b)}(x)\middle|\mathcal{D}, \alpha^{(b)}, \boldsymbol{\theta}^{(b)}\right\}\middle|\mathcal{D}\right]$$
$$= \text{Var}\{\widehat{F}_{\text{Bayes}}(x)|\mathcal{D}\}.$$

Thus, the Rao–Blackwell Bayesian estimator $\widehat{F}_{\text{Bayes}}(x)$ is used throughout the numerical studies. Based on B posterior samples $F^{(1)}(x), \ldots, F^{(B)}(x)$ obtained by Algorithm A1, the $100(1 - q)\%$ pointwise credible interval of the CDF $F(x)$ can be obtained as $[\widehat{L}_q(x), \widehat{U}_q(x)]$, where

$$\widehat{L}_q(x) = \max\left\{y : \frac{1}{B}\sum_{b=1}^{B} I(F^{(b)}(x) \leq y) < q/2\right\},$$
$$\widehat{U}_q(x) = \min\left\{y : \frac{1}{B}\sum_{b=1}^{B} I(F^{(b)}(x) \leq y) \geq 1 - q/2\right\}. \tag{14}$$

3 SIMULATIONS

To illustrate the performance of the proposed UIDP prior and the MCMC algorithm, we conduct experiments to compare the Bayesian estimator of the CDF $F(x)$ with the classic NPMLE.

3.1 Exponential distributions

We simulate 1000 target datasets with the study period $[0, 2]$ and sample size $n = 200$, each denoted as $\mathcal{D} = \{(Z_i, \Delta_i); i = 1, \ldots, n\}$, where

$$Z_i = \min(X_i, C_i, 2), \quad \Delta_i = I(X_i \leq C_i, X_i \leq 2),$$
$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F(x) = 1 - \exp(-\theta x),$$
$$C_1, \ldots, C_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(0.5),$$

for each value of the parameter $\theta \in \{0.7, 0.75, \ldots, 1.3\}$. Corresponding to each target dataset, we generate $K = 2$ historical datasets $\mathcal{D}_1, \mathcal{D}_2$, where $\mathcal{D}_k = \{(Z_{ik}, \Delta_{ik}); i = 1, \ldots, n_k\}$ for $k = 1, 2$, and

$$Z_{ik} = \min(X_{ik}, C_{ik}, 2), \quad \Delta_{ik} = I(X_{ik} \leq C_{ik}, X_{ik} \leq 2),$$
$$X_{1k}, \ldots, X_{n_k k} \overset{\text{i.i.d.}}{\sim} F_k(x) = 1 - \exp(-\theta_k x),$$
$$C_{1k}, \ldots, C_{n_k k} \overset{\text{i.i.d.}}{\sim} \text{Exp}(0.5),$$

with parameters $(\theta_1, \theta_2) = (0.9, 1.1)$ and sample sizes $n_1 = n_2 = 100$.

Although historical datasets may be accessible in the public databases or through extensive collaborations, there remain practical situations where only summary statistics are available due to data privacy or confidentiality. To mimic the latter scenarios, we compute the Nelson–Aalen (NA) estimators (Nelson, 1969, 1972; Aalen, 1978) $\widehat{\Lambda}_1(x)$ and $\widehat{\Lambda}_2(x)$ based on the 2 historical datasets, respectively, for historical information borrowing. To perform the matching in Equations 11 and 12, we use data-dependent interval points $x_0, \ldots, x_5$ with $x_0 = 0$ and $x_5 = \min_k(\sup\{x \in [0, 2] : \widehat{F}_k(x) \leq 0.9\})$ and keep the numbers of observed events in the 2 historical datasets the same for all intervals $(x_0, x_1], \ldots, (x_4, x_5]$. We adopt the cross-validation procedure in Section 2.3 to select the ESS M, and then apply Algorithm A1 to draw posterior samples, from which we compute the Bayesian estimator in comparison with the KM estimator. The 2 estimators are compared using the mean squared error (MSE) for estimation of the CDF $F(x) = 1 - \exp(-\theta x)$.

Figure 1A presents the ratio of MSEs between the Bayesian estimator in Equation 13 and the KM estimator over the study period $[0, 2]$ under different values of θ for the target dataset $\mathcal{D}$. The UIDP prior improves the estimation efficiency when historical datasets are similar to the target dataset. The Bayesian

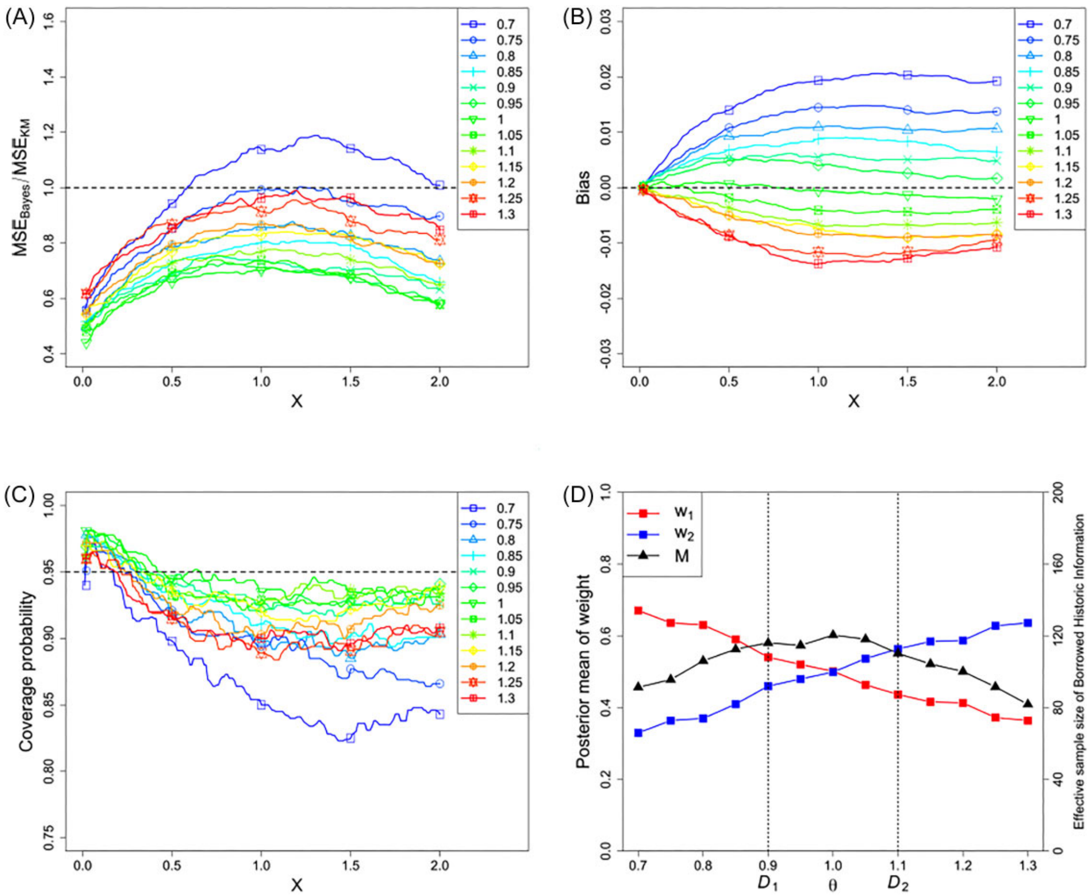

FIGURE 1 Performance of the Bayesian estimator of $F(x) = 1 - \exp(-\theta x)$ over $[0, 2]$ using the unit information Dirichlet process (UIDP) prior under the exponential distribution datasets. (A) The mean squared error (MSE) ratio between the Bayesian estimator in Equation 13 and the KM estimator under different values of θ. (B) Bias of the Bayesian estimator under different values of θ. (C) Pointwise coverage rates of 95% credible intervals by Equation 14 under different values of θ. (D) Average values of the selected M and posterior means of (w_1, w_2) under different values of θ.

estimator outperforms the KM estimator and the former's advantage amplifies as x decreases (or increases) when $x \in [0, 1]$ (or $x \in [1, 2]$). Such a trend of the MSE ratio mainly originates from the larger estimation uncertainty of $F(x)$ when $F(x)$ is close to 0.5. The Bayesian estimator performs better for $\theta \in [0.9, 1.1]$ than $\theta < 0.9$. However, when $\theta > 1.1$, MSE reduction of the Bayesian estimator over the KM estimator becomes smaller as θ increases. The main reason is that as θ increases, historical datasets become more dissimilar to the target dataset and thus borrowing historical information becomes more detrimental. Similar trends can also be observed from the bias of the Bayesian estimator and the coverage rate of the 95% credible interval $[\widehat{L}_{0.05}(x), \widehat{U}_{0.05}(x)]$ in Figure 1B and C. When the parameter θ of the target dataset is between θ_1 and θ_2, the Bayesian estimator yields smaller bias, and the coverage rate of credible intervals is close to the nominal level 95%. When θ is not between θ_1 and θ_2, the bias of the Bayesian estimator becomes larger compared with the KM estimator (see Section A5.1 of the Appendix) and the coverage rate is no longer close to the nominal level 95%. However, the coverage rate remains above 85% in most of the cases, suggesting that our method can reduce the negative impact of poor historical information on credible intervals. From Figure 1D, it is observed that as θ grows, the average value of the selected M increases until θ exceeds θ_1 and then decreases after θ exceeds θ_2, while the posterior mean of w_2 keeps increasing. This implies that the proposed UIDP prior can adaptively adjust the amount of information borrowed from historical datasets according to the similarity between the target dataset and historical datasets.

We further investigate how our method performs as sample size grows. Specifically, in Section A5.2 of the Appendix, we perform experiments where sample sizes of the target dataset and historical datasets are $(n, n_1, n_2) = (400, 200, 200)$ and $(n, n_1, n_2) = (800, 400, 400)$. Compared with Figure 1, it is clear that increasing sample size would decrease both the bias of the Bayesian estimator and its advantage in MSE reduction over the KM estimator when $\theta \notin [\theta_1, \theta_2]$. The main reason is that when the sample size is larger, the relative amount of borrowed historical information is lower if $\theta \notin [\theta_1, \theta_2]$, as shown in Figure A4 of the Appendix.

3.2 Weibull distribution

We simulate 1000 target datasets with the study period $[0, 2]$ and sample size $n = 200$, each denoted as $\mathcal{D} = \{(Z_i, \Delta_i); i =$

TABLE 1 Average values of the selected M and posterior means of $w_1, \ldots, w_5$ under different combinations of Weibull distribution parameters (ν, η).

Scenario	ν	η	M	$w_1(\%)$	$w_2(\%)$	$w_3(\%)$	$w_4(\%)$	$w_5(\%)$
1	0.90	0.90	110.32	24.48	30.62	17.98	12.71	14.21
2	0.95	0.95	134.76	22.11	28.69	18.63	14.20	16.38
3	1.00	1.00	144.80	19.53	26.57	19.35	15.38	19.18
4	1.05	1.05	143.03	17.51	24.74	19.53	16.21	22.01
5	1.10	1.10	128.23	15.58	23.11	19.29	16.97	25.05
6	0.90	1.10	126.75	24.58	25.16	19.20	15.47	15.58
7	0.95	1.05	137.30	21.96	25.54	19.34	15.64	17.52
8	1.05	0.95	146.63	17.77	27.49	19.26	14.57	20.91
9	1.10	0.90	144.54	15.81	28.65	19.15	14.00	22.39
	Historical datasets		ν_k	0.90	1.00	1.00	1.00	1.10
			η_k	1.00	0.90	1.00	1.10	1.00

$1, \ldots, n\}$, where

$$Z_i = \min(X_i, C_i, 2), \quad \Delta_i = I(X_i \le C_i, X_i \le 2),$$

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F(x) = 1 - \exp\{-(x/\eta)^\nu\},$$

$$C_1, \ldots, C_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(0.5),$$

for each combination of parameters (ν, η) listed in Table 1.

For each target dataset, we generate $K = 5$ historical datasets $\mathcal{D}_1, \ldots, \mathcal{D}_5$ and each $\mathcal{D}_k = \{(Z_{ik}, \Delta_{ik}); i = 1, \ldots, n_k\}$, where

$$Z_{ik} = \min(X_{ik}, C_{ik}, 2), \quad \Delta_{ik} = I(X_{ik} \le C_{ik}, X_{ik} \le 2),$$

$$X_{1k}, \ldots, X_{n_k k} \overset{\text{i.i.d.}}{\sim} F_k(x) = 1 - \exp\{-(x/\eta_k)^{\nu_k}\},$$

$$C_{1k}, \ldots, C_{n_k k} \overset{\text{i.i.d.}}{\sim} \text{Exp}(0.5),$$

with Weibull distribution parameters (ν_k, η_k) listed in Table 1 and sample size $n_k = 300$ for $k = 1, \ldots, 5$.

From Table 1, it is observed that Scenarios 1, 5, 6, and 9 correspond to the case that $\mathcal{D}$ deviates from $\mathcal{D}_1, \ldots, \mathcal{D}_K$. We also compute the NA estimators $\widehat{\Lambda}_1(x), \ldots, \widehat{\Lambda}_5(x)$ for historical information borrowing and use data-dependent interval points $x_0, \ldots, x_{10}$ for the matching in Equations 11 and 12, with $x_0 = 0$, $x_{10} = \min_k(\sup\{x \in [0, 2] : \widehat{F}_k(x) \le 0.9\})$ and keep the numbers of observed events in 5 historical datasets the same for all intervals $(x_0, x_1], \ldots, (x_9, x_{10}]$. We evaluate the Bayesian estimator in comparison with the KM estimator in terms of MSE for estimation of the CDF $F(x) = 1 - \exp\{-(x/\eta)^\nu\}$.

Figure 2A presents the ratio of MSEs between the Bayesian estimator and the KM estimator in estimating $F(x)$ over the study period [0,2] under different values of parameters ν and η of the target datasets. Similar to Figure 1A, the UIDP prior dramatically improves the efficiency in estimating $F(x)$ with about 50% reduction in the MSE when $\mathcal{D}$ does not deviate from $\mathcal{D}_1, \ldots, \mathcal{D}_K$ (Scenarios 2, 3, 4, 7, and 8). Even when $\mathcal{D}$ deviates from $\mathcal{D}_1, \ldots, \mathcal{D}_K$ (Scenarios 1, 5, 6, and 9), the MSE of the Bayesian estimator is still smaller than that of the KM estimator. In Figure 2B, it is observed that for all scenarios, the bias of the Bayesian estimator is negligible. As shown in Figure 2C, the coverage rates of 95% credible intervals $[\widehat{L}_{0.05}(x), \widehat{U}_{0.05}(x)]$ by Equation 14 are close to the nominal level 95% when $\mathcal{D}$ does not deviate from $\mathcal{D}_1, \ldots, \mathcal{D}_K$ (Scenarios 2, 3, 4, 7, and 8). We also present the average value of the selected M and posterior means of $w_1, \ldots, w_K$ under all scenarios in Table 1. The selected M is larger when $\mathcal{D}$ is more similar to $\mathcal{D}_1, \ldots, \mathcal{D}_K$ and the average value of the posterior means of w_k's reflects the similarities between $\mathcal{D}$ and $\mathcal{D}_k$'s.

3.3 Model misspecification

To evaluate whether the UIDP prior is vulnerable to model misspecification, we perform Bayesian sensitivity analysis on Weibull distributed data using the UIDP prior with the exponential base distribution $G_{\boldsymbol{\theta}}$. We simulate 1000 target datasets with the study period $[0, 2]$ and sample size $n = 200$, each dataset denoted as $\mathcal{D} = \{(Z_i, \Delta_i); i = 1, \ldots, n\}$, where

$$Z_i = \min(X_i, C_i, 2), \quad \Delta_i = I(X_i \le C_i, X_i \le 2),$$

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F(x) = 1 - \exp\{-(x/\eta)^\nu\},$$

$$C_1, \ldots, C_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(0.5).$$

for each combination of $\eta = 0.7, 0.75, \ldots, 1.3$ and $\nu = 0.8, 1.2$. Under the Weibull distribution, the value of ν determines the shape of the hazard function: $\nu = 1$ leads to a constant hazard (corresponding to an exponential model), $\nu < 1$ yields a decreasing hazard over time, and $\nu > 1$ results in an increasing hazard over time. For each target dataset, we generate $K = 2$ historical datasets $\mathcal{D}_1$ and $\mathcal{D}_2$ with $\mathcal{D}_k = \{(Z_{ik}, \Delta_{ik}); i = 1, \ldots, n_k\}$ for $k = 1, 2$, where

$$Z_{ik} = \min(X_{ik}, C_{ik}, 2), \quad \Delta_{ik} = I(X_{ik} \le C_{ik}, X_{ik} \le 2),$$

$$X_{1k}, \ldots, X_{n_k k} \overset{\text{i.i.d.}}{\sim} F_k(x) = 1 - \exp\{-(x/\eta_k)^{\nu_k}\},$$

$$C_{1k}, \ldots, C_{n_k k} \overset{\text{i.i.d.}}{\sim} \text{Exp}(0.5),$$

and we set parameters $(\eta_1, \eta_2) = (0.9, 1.1)$ and $\nu_1 = \nu_2$ equal to the ν used in generating the target dataset. Using an exponential CDF as the base distribution $G_{\boldsymbol{\theta}}$, we evaluate the Bayesian estimator in comparison with the KM estimator in terms of the MSE for estimating the CDF $F(x) = 1 - \exp\{-(x/\eta)^\nu\}$.

As shown in Figure 3, the Bayesian estimator using the UIDP prior with an exponential base distribution has comparable accuracy with the KM estimator. Specifically, when $\nu = 0.8$ and η are large or when $\nu = 1.2$ and η are small, our Bayesian estimator outperforms the KM estimator throughout the entire interval $[0, 2]$. The Bayesian estimator under the UIDP prior can

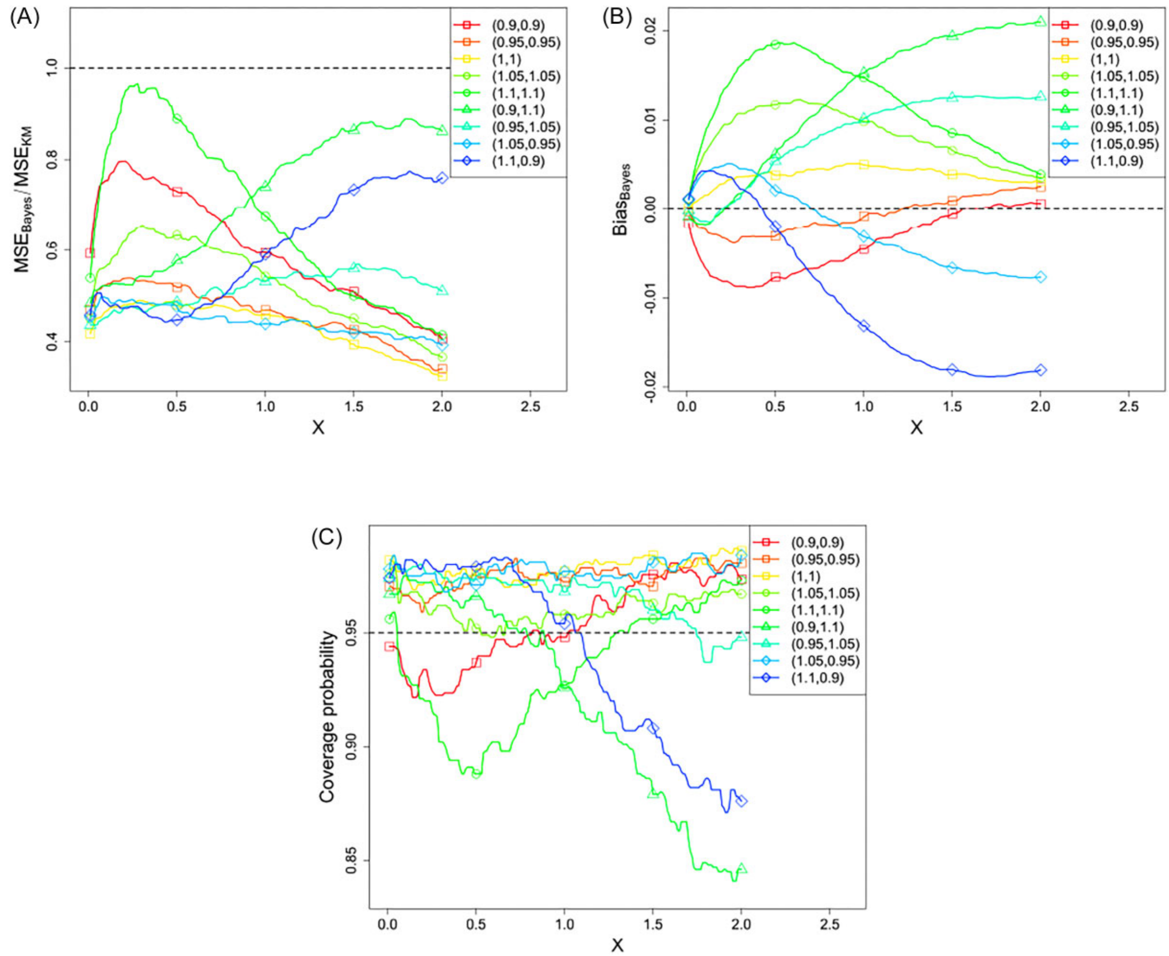

FIGURE 2 Performance of the Bayesian estimator of $F(x) = 1 - \exp\{-(x/\eta)^{\nu}\}$ over $[0, 2]$ using the unit information Dirichlet process (UIDP) prior under the Weibull distribution datasets. (A) The mean squared error (MSE) ratio between the Bayesian estimator and the KM estimator under different values of (ν, η). (B) Bias of the Bayesian estimator under different values of (ν, η). (C) Pointwise coverage rates of 95% credible intervals under different values of (ν, η).

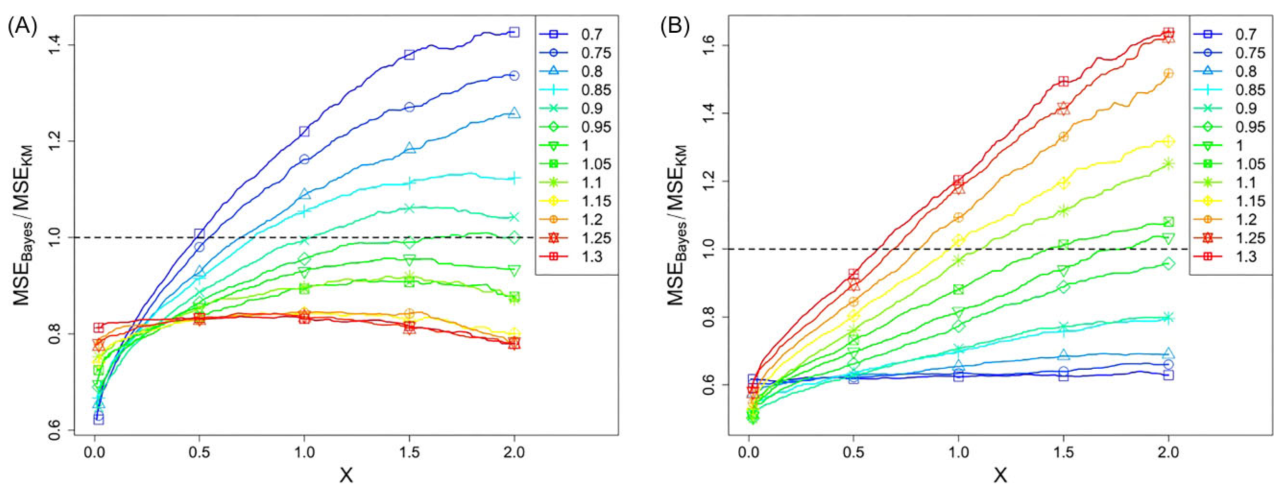

FIGURE 3 Sensitivity analysis of the Bayesian estimator in Equation 13 for estimating the Weibull cumulative distribution function $F(x)$ over $[0, 2]$ using the unit information Dirichlet process (UIDP) prior with the exponential base distribution under the Weibull distribution datasets. (A) The mean squared error (MSE) ratio between the Bayesian estimator and the KM estimator under different values of η and $\nu = 0.8$. (B) The MSE ratio between the Bayesian estimator and the KM estimator under different values of η and $\nu = 1.2$.

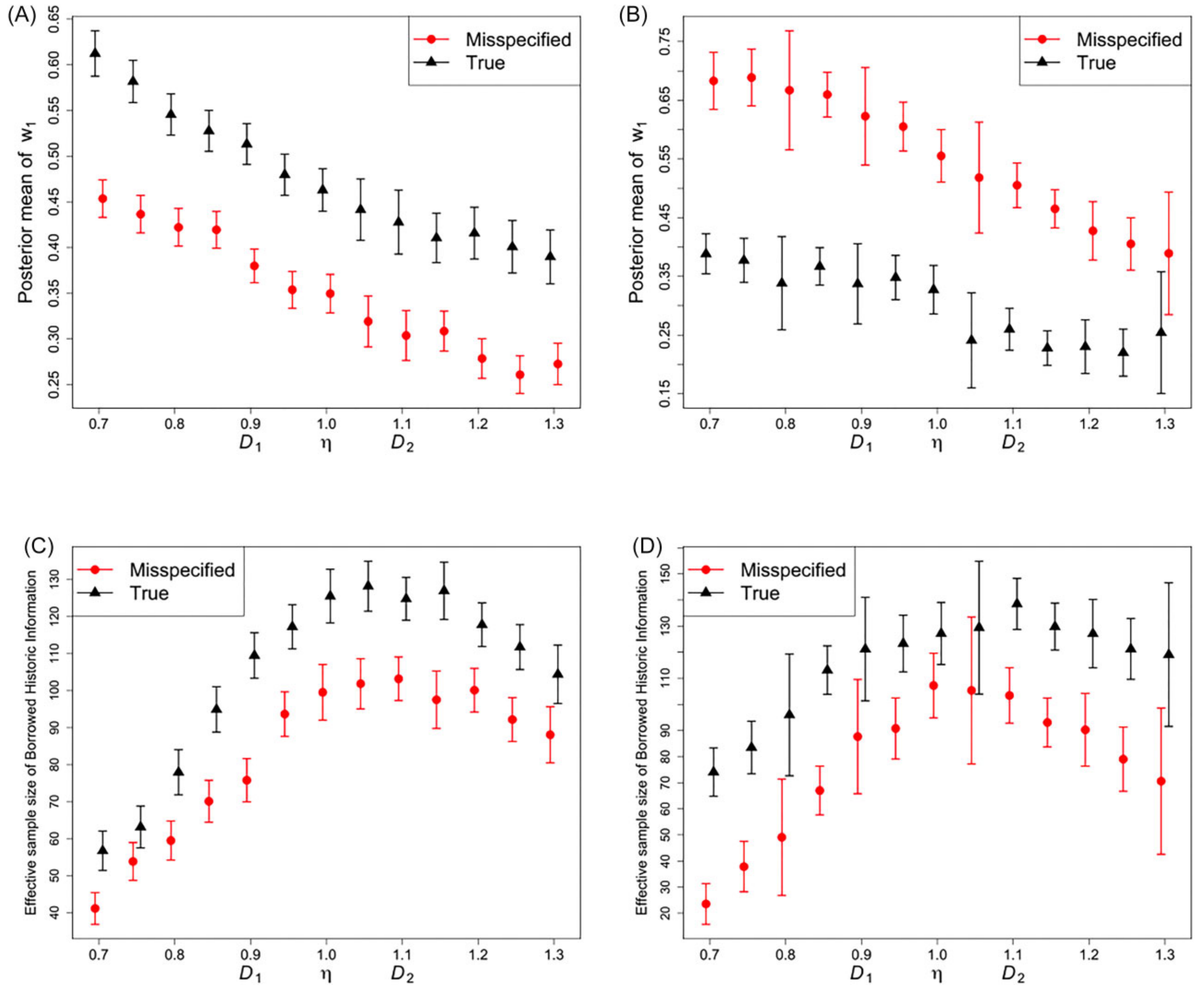

FIGURE 4 Comparison between Bayesian estimators in Equation 13 using the unit information Dirichlet process (UIDP) prior with a misspecified exponential base distribution and the true Weibull base distribution in estimating weights and the effective sample size of historical datasets. (A) Estimated weight w_1 under different values of η and $\nu = 0.8$. (B) Estimated weight w_1 under different values of η and $\nu = 1.2$. (C) The effective sample size M under different values of η and $\nu = 0.8$. (D) The effective sample size M under different values of η and $\nu = 1.2$.

generally maintain low MSE when the base distribution is specified incorrectly, because our UIDP prior can adaptively reduce the impact of a misspecified base distribution $G_{\boldsymbol{\theta}}$ on the Bayesian estimator, as shown by the posterior mean of weight w_1 and the average value of selected ESS M in Figure 4. As a benchmark for comparison, we also perform Bayesian inference using the UIDP prior with the correctly specified Weibull base distribution on the same datasets. As shown in Figure 4A and B, the posterior mean of w_1 using the UIDP prior with a misspecified base distribution is different from that with the correctly specified base distribution. Under the UIDP prior with a misspecified base distribution, the posterior mean of w_1 is smaller when $\nu = 0.8$ and is larger when $\nu = 1.2$ than that using a correctly specified base distribution. Specifically, when $\nu = 0.8$, less information is borrowed from $\mathcal{D}_1$ (w_1 is smaller than 0.5), while more information is borrowed from $\mathcal{D}_2$ ($w_2 = 1 - w_1$ is larger). Because the target dataset $\mathcal{D}$ is more similar to $\mathcal{D}_2$ than $\mathcal{D}_1$ when η is large, borrowing more information from $\mathcal{D}_2$ can provide a more accurate estimate of $F(x)$. The advantage of our Bayesian estimator over the KM estimator when $\nu = 1.2$ and η are small can also be explained analogously. Figure 4C and D show that when the parameter η does not lie between η_1 and η_2, the selected ESS M decreases as η moves further away from η_1 and η_2. This indicates that under the misspecified base distribution, the UIDP prior can adaptively select the ESS according to the similarity between the target dataset and historical datasets.

4 REAL DATA ANALYSIS

To illustrate the empirical performance of the UIDP prior in improving the efficiency of statistical inference, we analyze the Brazil cancer dataset, which contains patient data in Brazil cancer centers from 2000 to 2019. In our study, we extracted observations from 7 states, including Acre (AC), Alagoas (AL), Minas Gerais (MG), Pará (PA), Paraná (PR), São Paulo (SP), and Sergipe (SE). In the extracted 7 datasets, each patient record contains the dates of diagnosis, last contact, and the death. We treat records with the dates of death as uncensored observations, and

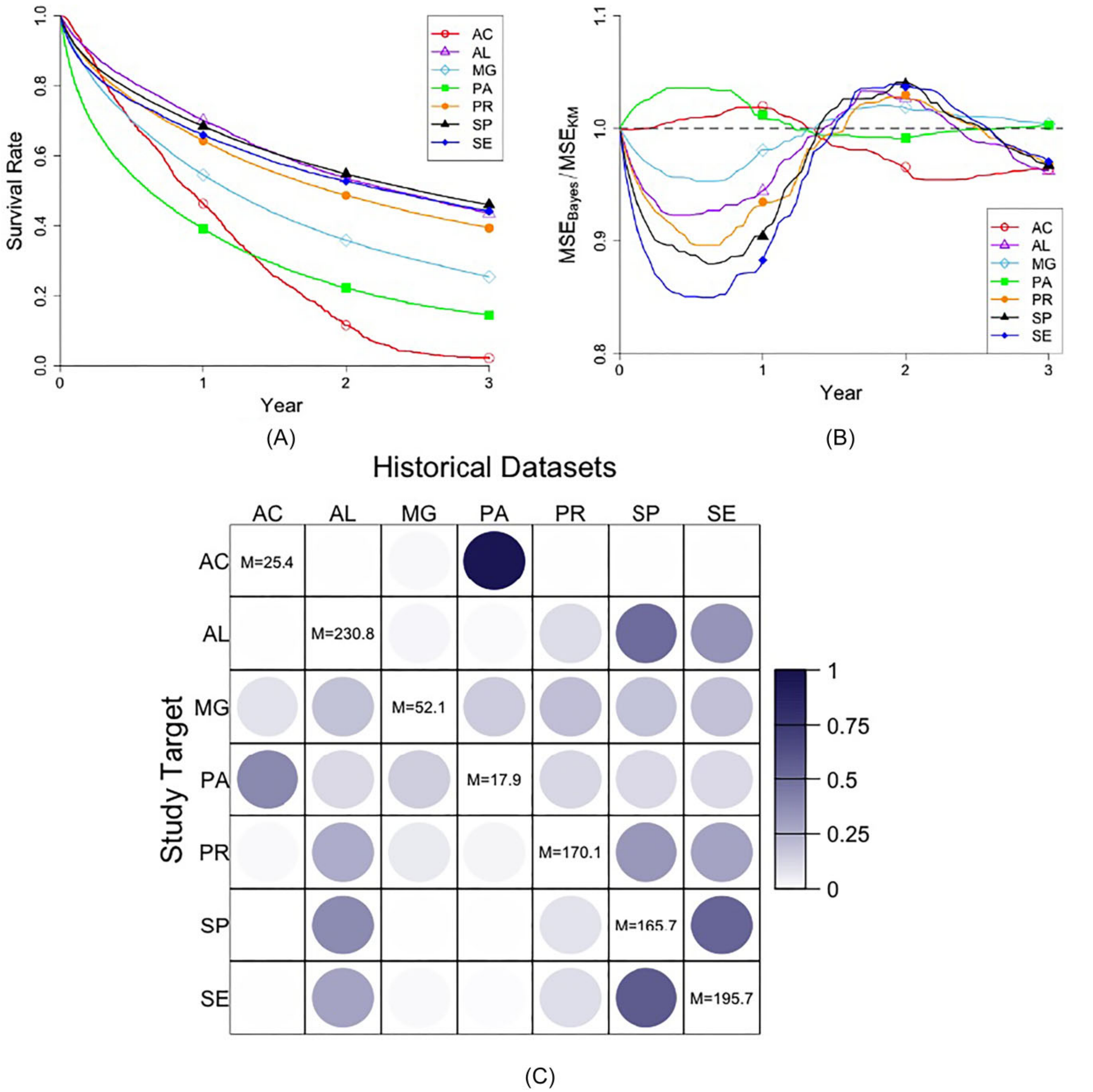

FIGURE 5 (A) Kaplan–Meier estimators of the survival functions of cancer patients in the 7 states of Brazil within the first 3 years of follow-up after diagnosis. (B) MSE ratios between the Bayesian estimator and the KM estimator in estimating survival functions of cancer patients in different states. (C) The amount and the proportion of information borrowed from historical datasets when we alternate each state as the study target and keep the remaining states as historical datasets.

censored observations if otherwise. As a result, the number of observations and the censoring rate for these 7 states are, respectively, (1) AC: (2146, 48.70%); (2) AL: (6100, 39.62%); (3) MG: (42077, 22.87%); (4) PA: (18729, 4.76%); (5) PR: (42009, 37.25%); (6) SP: (208753, 28.41%); and (7) SE: (10160, 20.68%), corresponding to 7 time-to-event datasets $\mathcal{D}_{\text{state}} = \{(Z_{i,\text{state}}, \Delta_{i,\text{state}}); i = 1, \ldots, n_{\text{state}}\}$ (state =AC, AL, MG, PA, PR, SP, and SE). The observed event time of an uncensored observation is the number of days from diagnosis to death, and the observed follow-up time for a censored observation is the number of days from diagnosis to the last contact. Figure 5A shows the KM estimators of survival functions for cancer patients in different states within the first 3 years (1095 days) after diagnosis.

To compare the NPMLE and our Bayesian UIDP approach in estimating the CDF $F(x) = 1 - S(x)$, we set the state AC as the study target and synthesize 1000 datasets $\mathcal{D}_{\text{AC}}^{(b)}$ $(b = 1, \ldots, 1000)$ of size 600 by randomly drawing observations without replacement from the AC dataset $\mathcal{D}_{\text{AC}}$. For $b = 1, \ldots, 1000$, we also synthesize 6 datasets

$\mathcal{D}_{\rm AL}^{(b)}, \mathcal{D}_{\rm MG}^{(b)}, \ldots, \mathcal{D}_{\rm SE}^{(b)}$ of each size 400 to serve as historical datasets and incorporate them to compute the Bayesian estimator using an exponential base distribution $G_{\boldsymbol{\theta}}(x)$. We take the KM estimator of the original dataset $\mathcal{D}_{\rm AC}$ (presented as the red line in Figure 5A) as the true survival function, and compare the MSEs of the Bayesian estimator and the KM estimator over 1000 synthetic datasets. We also investigate how much information is borrowed from different historical datasets in analyzing $\mathcal{D}_{\rm AC}^{(b)}$'s by summarizing the selected M and posterior means of $w_{\rm AL}, \ldots, w_{\rm SE}$ over 1000 synthetic datasets. We perform the same experiment by, respectively, taking each individual state as the study target and the remaining states as historical studies to examine how similarities among different datasets affect the amount (reflected by the selected M) and the proportion (reflected by posterior means of w_k's) of information borrowed.

From Figure 5B, it is observed that our Bayesian estimator has comparable performance to the KM estimator in estimating the CDF. Specifically, when the study target is AC or PA, the KM estimator has lower MSE than our Bayesian estimator because both survival curves of AC and PA substantially differ from the rest. As a result, borrowing information from other states might bias the inference, and our UIDP prior adaptively limits the amount of information borrowed from other datasets with the average selected $M = 25.4$ for AC and 17.9 for PA, as shown in Figure 5C. Because our cross-validation procedure is used to select M from $\{1, 60, 120, \ldots, 600\}$, such a small average selected M implies that our procedure selects $M = 1$ for more than 60% replicates. However, when we take either AL, PR, SP, or SE as the study target, the MSE of our Bayesian estimator is generally lower than the KM estimator, because these 4 states have similar survival curves and mutually borrowing information can significantly improve the efficiency. As shown in Figure 5C, our UIDP prior adaptively borrows more historical information with a larger average value of M and, more importantly, the proportions of information borrowed among these 4 states (AL, PR, SP, and SE) are much higher than the proportions from the other 3 states (AC, MG, and PA), implying the ability of our UIDP prior in adaptively adjusting the proportions of information borrowed according to similarities between the datasets. When the study target is MG, the amount of information borrowed from other datasets is moderate (the average selected M equals 52.1) with quite balanced weights among other datasets. The main reason is that the survival curve of MG lies in the middle of all survival curves, and the inference of MG can borrow information from datasets with both higher survival rates (AL, PR, SP, and SE) and lower survival rates (AC and PA).

5 CONCLUSION

As an extension of UI in Bayesian nonparametrics, we propose the UIDP prior as a venue for borrowing information from historical data. Based on the Fisher information for the differential of CHF derived under the Dirichlet process, the UIDP prior is defined by matching the prior mean and prior UI of CHF with the weighted average of the estimated CHFs and weighted UI from historical datasets. Without any parametric assumption imposed on the target dataset and historical information, both parametric and nonparametric information of historical datasets can be incorporated in the UIDP prior to improve the statistical efficiency of Bayesian nonparametric inference. With an elaborated MCMC algorithm to draw posterior samples and a cross-validation procedure to determine the amount of historical information to be borrowed, simulations and real data analysis reveal the advantages of the UIDP prior over the classic nonparametric maximum likelihood approach in estimating the CDF of the target dataset. Moreover, the amount of information borrowed from different historical datasets is shown to be consistent with their similarities to the target dataset.

As the UI prior is defined only in terms of the prior mean and UI, it can be generalized to other nonparametric priors. However, because the independent increment property $\mathrm{Cov}\{d\Lambda(x_1), d\Lambda(x_2)\} = 0$ (for all $x_1 \neq x_2$) may not hold under other Bayesian nonparametric models, their corresponding UI priors warrant further investigation. Another possible future direction for research is to incorporate the UIDP prior into the framework of the empirical Bayes. As shown in Section 3.3, misspecification of the parametric base distribution inevitably has negative impact on the performance of our Bayesian inference. Thus, it is interesting to develop an empirical Bayes procedure that uses the nonparametrically estimated base distribution to address the misspecification issue.

ACKNOWLEDGMENTS

The authors would like to thank the associate editor, referee, and co-editor for their many constructive and insightful comments that have led to significant improvements in this paper.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices, Tables, Figures, and data and R code referenced in Sections 2-4 are available with this paper at the Biometrics website on Oxford Academic.

FUNDING

This research was partially supported by funding from the Research Grants Council of Hong Kong (17308321).

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The data that support the findings in this paper are available in https://www.kaggle.com/datasets/joaopedromedeiros/cancer-data-brazil. The data processed for analysis are also included in the online supplementary materials.

REFERENCES

Aalen, O. (1978). Nonparametric inference for a Family of counting processes. *The Annals of Statistics*, 6, 701–726.

Banbeta, A., van Rosmalen, J., Dejardin, D. and Lesaffre, E. (2019). Modified power prior with multiple historical trials for binary endpoints. *Statistics in Medicine*, 38, 1147–1169.

Borghaei, H., Paz-Ares, L., Horn, L., Spigel, D. R., Steins, M., Ready, N. E. et al. (2015). Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *New England Journal of Medicine*, 373, 1627–1639.

Brahmer, J., Reckamp, K. L., Baas, P., Crinò, L., Eberhardt, W. E., Poddubskaya, E. et al. (2015). Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *New England Journal of Medicine*, 373, 123–135.

Efron, B. (2015). Frequentist accuracy of Bayesian estimates. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77, 617–646.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1, 209–230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2, 615–629.

Gravestock, I. and Held, L. (2019). Power priors based on multiple historical studies for binary outcomes. *Biometrical Journal*, 61, 1201–1218.

Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science*, 15, 46–60.

Jin, H. and Yin, G. (2021). Unit information prior for adaptive information borrowing from multiple historical datasets. *Statistics in Medicine*, 40, 5657–5672.

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz Criterion. *Journal of the American Statistical Association*, 90, 928–934.

Nelson, W. (1969). Hazard Plotting for incomplete failure data. *Journal of Quality Technology*, 1, 27–52.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14, 945–966.

Neuenschwander, B., Capkun-Niggli, G., Branson, M. and Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials*, 7, 5–18.

Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, 29, 175–188.

Reimherr, M., Meng, X.-L. and Nicolae, D. L. (2021). Prior sample size extensions for assessing prior impact and prior-likelihood discordance. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83, 413–437.

Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., von Pawel, J. et al. (2017). Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *The Lancet*, 389, 255–265.

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D. and Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70, 1023–1032.

Wu, Y.-L., Lu, S., Cheng, Y., Zhou, C., Wang, J., Mok, T. et al. (2019). Nivolumab versus docetaxel in a predominantly chinese patient population with previously treated advanced NSCLC: CheckMate 078 randomized phase III clinical trial. *Journal of Thoracic Oncology*, 14, 867–875.

Received: January 12, 2024; Revised: July 5, 2024; Accepted: August 21, 2024