Elastic Priors to Dynamically Borrow Information from Historical Data in Clinical Trials

Liyun Jiang\textsuperscript{1,3}, Lei Nie\textsuperscript{2}, and Ying Yuan\textsuperscript{1}

\textsuperscript{1}Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX
\textsuperscript{2}Center for Drug Evaluation and Research, Food and Drug Administration (FDA), Silver Spring, MD
\textsuperscript{3}Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing, China.

Abstract: Use of historical data and real-world evidence holds great potential to improve the efficiency of clinical trials. One major challenge is how to effectively borrow information from historical data while maintaining a reasonable type I error. We propose the elastic prior approach to address this challenge and achieve dynamic information borrowing. Unlike existing approaches, this method proactively controls the behavior of dynamic information borrowing and type I errors by incorporating a well-known concept of clinically meaningful difference through an elastic function, defined as a monotonic function of a congruence measure between historical data and trial data. The elastic function is constructed to satisfy a set of information-borrowing constraints prespecified by researchers or regulatory agencies, such that the prior will borrow information when historical and trial data are congruent, but refrain from information borrowing when historical and trial data are incongruent. In doing so, the elastic prior improves power and reduces the risk of data dredging and bias. The elastic prior is information borrowing consistent, i.e. asymptotically controls type I and II
errors at the nominal values when historical data and trial data are not congruent, a unique characteristic of the elastic prior approach. Our simulation study that evaluates the finite sample characteristic confirms that, compared to existing methods, the elastic prior has better type I error control and yields competitive or higher power.

KEY WORDS: Real-word data; Historical data; Dynamic information borrowing; Elastic prior; Elastic MAP prior; Adaptive design

1 Introduction

Real-world data (RWD) or evidence plays an increasingly important role in health care decisions. The 21st Century Cures Act, signed into law in 2016, emphasizes modernization of clinical trial designs, including the use of real-world evidence to support approval of new indications for approved drugs or to satisfy post-approval study requirements. The FDA released related guidance in the “Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices” [1] in 2017, and a draft guidance on “Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics Guidance for Industry” [2] in 2019.

Use of RWD to facilitate medical decisions is an extremely broad topic. We here focus on the use of historical data to improve the efficiency and guide decision making of randomized controlled trials (RCTs). For ease of exposition, we assume two-arm RCTs and historical data are only available on the control. It is straightforward to extend the proposed methodology to multiple-arm RCTs and to cases where historical data are also available for the treatment arm. The question of interest is how to leverage information from historical
data to increase the power of comparing the treatment efficacy between the control and treatment arms. This problem is also known as augmenting the control arm with historical data or RWD.

Under the Bayesian paradigm, such information borrowing is straightforward if historical data $D_h$ are congruent (or exchangeable) to control data $D_c$. Let $\theta$ denote the parameter of interest (e.g., mean of the efficacy endpoint). We start with assigning $\theta$ a non-informative or vague prior $\pi_0(\theta)$, combined with $D_h$, to obtain its posterior $\pi(\theta \mid D_h)$, and then use that posterior as the prior for $D_c$ to make the comparison between control and treatment arms. Such full information borrowing, however, is not appropriate when $D_h$ are partially or not congruent to $D_c$, leading to bias. If the bias favors treatment, the type I error rate will be inflated. If the bias favors control, the power of the study will reduce.

Various approaches have been proposed for dynamic information borrowing, such that the amount of information borrowed from $D_h$ is automatically adjusted according to the congruence between $D_h$ and $D_c$. Chen and Ibrahim \cite{3,4} proposed a power prior, which controls the degree of information borrowing through a “power parameter.” Hobbs et al. (2011) \cite{5} proposed a commensurate prior that allows for the commensurability of the information in the historical data and current data to determine how much historical information to use. Thall et al. (2003) \cite{6} and Berry et al. (2013) \cite{7} proposed to use the Bayesian hierarchical model to borrow information from different data resources or subgroups. Schmidli et al. (2014) \cite{8} proposed a robust meta-analytic-predictive (MAP) prior to borrow information from historical data via a mixture prior. Pan, Yuan, and Xia (2017) \cite{9} proposed a calibrated power prior, assuming the availability of patient-level historical data. However, most of these methods have difficulty achieving dynamic information borrowing, leading to substantially inflated type I error and bias, as noted previously by Neuenschwander et al. \cite{10}, Freidlin and Korn \cite{11}, and Chu and Yuan \cite{12}, among others.
In this paper, we propose a general Bayesian method with elastic priors to address the aforementioned issue. Unlike many existing approaches, the proposed method proactively controls the behavior of dynamic information borrowing through an elastic function, defined as a monotonic function of a congruence measure between $D_h$ and $D_c$. The elastic function is constructed to satisfy a set of prespecified information borrowing constraints. For example, a borrowing constraint can be set based on a prespecified clinically meaningful difference such that the amount of borrowing decreases when the difference between $D_h$ and $D_c$ increases. This control leads to a substantially reduced risk of bias. Asymptotically, the elastic prior approach maintains type I and II errors at the nominal value when $D_h$ and $D_c$ are not congruent. In contrast, most existing dynamic information borrowing methods, including the power prior, commensurate prior, and robust MAP prior, do not have this characteristic. The elastic prior also demonstrates superior finite sample characteristics. Our simulation study confirms that, compared to existing methods, the elastic prior approach controls type I errors better, yielding a competitive or higher power. Other desirable characteristics of the elastic prior approach include that it is straightforward to determine the prior effective sample size (PESS) contained in the elastic prior, and the elastic prior is defined independent of trial data $D_c$ and thus can be fully pre-specified.

The remainder of this article is organized as follows. In Section 2, we introduce the elastic prior method. In Section 3, we evaluate the operating characteristics of the proposed method using simulation, and we conclude with a brief discussion in Section 4.
2 Methods

Consider a two-arm RCT, let $y$ denote the efficacy endpoint that is a binary variable following a Bernoulli distribution or a continuous variable following a normal distribution. Let $\theta_c$ and $\theta_t$ denote $E(y)$ for the control and treatment arms, respectively. The objective of the trial is to compare $\theta_t$ with $\theta_c$ to determine whether the treatment is superior, noninferior, or equivalent to the control. Under the Bayesian paradigm, the decision can be made based on the following criterion: the treatment is deemed superior, noninferior, or equivalent to the control if $\Pr(M_L < \theta_t - \theta_c < M_H \mid D_c, D_t, D_h) > C$, where $M_L$ and $M_H$ are prespecified margins, $C$ is a probability cutoff. For example, superiority trials typically set $M_L = 0$ and $M_H = \infty$; noninferiority trials set $M_H = \infty$ and $M_L = -M$, where $M$ is the noninferiority margin; and equivalence trials set $(M_L, M_H) = (-E, E)$, where $E$ is the equivalence margin.

We assume that historical data $D_h$ are only available to the control. Thus, we focus on the posterior inference of $\theta_c$ and suppress its subscript when no confusion is caused. In the analysis, the posterior inference for $\theta_t$ will be done using standard Bayesian methods (e.g., using a conventional noninformative or vague prior).

The basic idea of an elastic prior is straightforward. Let $\pi_0(\theta)$ denote a vague initial prior that reflects prior knowledge about $\theta$ before $D_h$ is observed. Applying the prior $\pi_0(\theta)$ to $D_h$, we obtain a posterior distribution $\pi(\theta \mid D_h)$. The elastic prior is constructed by inflating the variance of $\pi(\theta \mid D_h)$ by a factor of $g(T)^{-1}$, where $T$ is a congruence measure between $D_h$ and $D_c$, and $g(T)$ is a monotonically decreasing function with values between 0 and 1. When $T \to 0$, reflecting a prefect congruence measure between $D_h$ and $D_c$, $g(T) \to 1$ and the $\pi(\theta \mid D_h)$ will be fully used as a prior. When $T \to \infty$, reflecting substantial incongruence measure between $D_h$ and $D_c$, $g(T) \to 0$ and the elastic prior will become a noninformative prior. In the next two subsections, we elaborate this approach using binary and Gaussian
2.1 Binary endpoint

Let $n_h$ and $n_c$ respectively denote the sample size of $D_h$ and $D_c$, $D_c = (y_{c,1}, \cdots, y_{c,n_c})$ and $D_h = (y_{h,1}, \cdots, y_{h,n_h})$, where $y_{h,i} \overset{i.i.d}{\sim} Bernoulli(\theta_h)$ and $y_{c,i} \overset{i.i.d}{\sim} Bernoulli(\theta)$. Let $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{h,i}$ and $\bar{y}_c = \frac{1}{n_c} \sum_{i=1}^{n_c} y_{c,i}$. Assuming a vague prior $\pi_0(\theta_h) \sim Beta(\alpha_0, \beta_0)$, with small values of $\alpha_0$ and $\beta_0$ (e.g., $\alpha_0 = \beta_0 = 0.1$), we multiply the historical likelihood function with $\pi_0(\theta_h)$, which results in a posterior of $\theta$ of the form.

$$
\pi(\theta_h | D_h) \propto Beta(\alpha_0 + n_h \bar{y}_h, \beta_0 + n_h - n_h \bar{y}_h).
$$

The elastic prior is given by

$$
\pi^*(\theta | D_h) \propto Beta((\alpha_0 + n_h \bar{y}_h)g(T), (\beta_0 + n_h - n_h \bar{y}_h)g(T)).
$$

The elastic prior $\pi^*(\theta | D_h)$ has the same mean as $\pi(\theta | D_h)$, but inflates the latter’s variance by a factor of $g(T)^{-1}$. Given $\pi^*(\theta | D_h)$, the posterior of $\theta$ after accounting for $D_c$ is

$$
\pi(\theta | D_h, D_c) = Beta((\alpha_0 + n_h \bar{y}_h)g(T) + n_c \bar{y}_c, (\beta_0 + n_h - n_h \bar{y}_h)g(T) + n_c - n_c \bar{y}_c).
$$

We now discuss how to choose congruence measure $T$ and elastic function $g(\cdot)$. For a binary endpoint, there are many different choices for congruence measure $T$. For example,
we may consider

\[ T = \frac{|\bar{y}_c - \bar{y}_h|}{\sqrt{\bar{y}(1 - \bar{y})\left(\frac{1}{n_c} + \frac{1}{n_h}\right)}}, \]

where \( \bar{y} = (\bar{y}_c n_c + \bar{y}_h n_h) / (n_c + n_h) \) is a pooled sample mean. While different choices of \( T \) have different advantages; in this paper, we choose the chi-square test statistic:

\[ T = \sum_{j=c,h} \frac{(O_{0j} - E_{0j})^2}{E_{0j}} + \sum_{j=c,h} \frac{(O_{1j} - E_{1j})^2}{E_{1j}}, \]

where \( O_{0j} \) and \( O_{1j} \) are the observed number of responders and non-responders for \( D_c \) and \( D_h \); \( E_{0j} \) and \( E_{1j} \) are the expected number of responders and non-responders, which are given by

\[ E_{0j} = n_j \sum_{j=c,h} \frac{n_j - \sum_{j=c,h} \sum_{i=1}^{n_j} y_{j,i}}{\sum_{j=c,h} n_j}, \quad E_{1j} = n_j \sum_{j=c,h} \frac{\sum_{i=1}^{n_j} y_{j,i}}{\sum_{j=c,h} n_j}. \]

A large value of \( T \in (0, \infty) \) indicates low congruence between \( D_c \) and \( D_h \).

Elastic function \( g(T) \) serves as a link function that maps congruence measure \( T \) to an information discount factor. Any monotonic function could be used as an elastic function, as long as \( g(T) \to 1 \) when the value of \( T \) corresponds to congruence and \( g(T) \to 0 \) when the value of \( T \) corresponds to incongruence. In this paper, we choose

\[ g(T) = \frac{1}{1 + \exp\{a + b \times \log(T)\}}, \tag{2} \]

where \( a \) and \( b > 0 \) are prespecified tuning parameters. We will discuss how to choose the values of \( a \) and \( b \) later. When appropriate, a more flexible elastic function \( g(T) = \frac{1}{1 + \exp\{a + b \times \log(T)\}} \) can be used to further control the rate of change from borrowing to no borrowing using the additional parameter \( c \) (see Figure 1 (a)). It can be shown that the resulting elastic prior has the following consistence property:
Theorem 1  The elastic prior defined in (1) is information-borrowing consistent. That is, when \( n_h \to \infty \) and \( n_c \to \infty \), it achieves full information borrowing if \( D_h \) and \( D_c \) are congruent (i.e., \( \theta_h = \theta \)), and discards \( D_h \) if \( D_h \) and \( D_c \) are incongruent (i.e., \( \theta_h \neq \theta \)).

The biggest concern and barrier for adopting information-borrowing methods in clinical trials is the potential risk of type I or II error inflation caused by the information borrowing when \( D_h \) and \( D_c \) are actually incongruent. Theorem 1 shows that, asymptotically, the elastic prior maintains a type I error at the nominal value when \( D_h \) and \( D_c \) are not congruent. In contrast, most existing dynamic information borrowing methods, including the power prior, commensurate prior, and robust MAP prior, do not have this property. To achieve the information-borrowing consistency, they typically require the number of historical datasets (not the number of observations within each historical dataset) goes to infinity, which is not the case in practice.

In finite samples, however, strictly controlling a type I error at its nominal value is impossible for any information-borrowing methods, including the elastic prior approach. The reason is simple: when \( \theta_h \neq \theta \), the type I error inflates whenever information-borrowing is triggered. With finite sample, even when \( \theta_h \neq \theta \), there is non-zero probability that the observed \( D_h \) and \( D_c \) are comparable and trigger (inappropriate) information borrowing, thus inflating the type I or II error.

Theorem 2  For any method that borrows information from historical or other external data, dynamically or non-dynamically, the inflation of type I or II error is inevitable under finite samples, depending on whether historical or other external data under- or over-estimate the treatment effect of the control arm when compared to the current data.

Theorem 2 is important, because it sets a realistic expectation for information-borrowing methods and avoids vain efforts to pursue a dynamic information borrowing method that
can strictly control type I errors in finite samples.

Since the inflation of type I or II errors is inevitable with information borrowing, one reasonable strategy is to control type I and II error inflation according to certain pre-specified criteria. This motivates the following procedure to choose the elastic function (2), as illustrated in Figure 2. Without loss of generality, we assume a large value of $T$ indicates larger incongruence between $D_h$ and $D_c$.

1. Elicit from subject matter experts a clinically meaningful difference (CMD), denoted as $\delta$, for $E(y)$. The CMD is routinely used in clinical trial planning, including for sample size determination and power calculation, and its determination often requires communication between sponsors and regulatory bodies.

2. (Congruent case) Simulate $R$ replicates of $D_c = (y_{c,1}, \cdots, y_{c,n_c})$ from $Bernoulli(\hat{\theta}_h)$, with $\hat{\theta}_h = \bar{y}_h$, and calculate congruence measure $T$ between $D_h$ and each simulated $D_c$, resulting in $T_0 = (T_1, \cdots, T_R)$, where $T_r$ denotes the value of $T$ based on the $r$th simulated $D_c$.

3. (Incongruent cases) Simulate $R$ replicates of $D_c$ from $Bernoulli(\hat{\theta}_h + 2\delta)$, and calculate congruence measure $T$ between $D_h$ and each simulated $D_c$, resulting in $T_{+1} = (T_{+1}^+, \cdots, T_{+1}^R)$, where $T_{+r}^+$ denotes the value of $T$ based on the $r$th simulated $D_c$. Repeat this with $D_c$ simulated from $Bernoulli(\hat{\theta}_h - 2\delta)$, resulting in $T_{-1} = (T_{-1}^-, \cdots, T_{-1}^R)$, where $T_{-r}^-$ denotes the value of $T$ between $D_h$ and the $r$th simulated $D_c$.

4. Let $C_1$ and $C_2$ be constants close to 1 and 0, respectively, e.g., $C_1 = 0.99$ and $C_2 = 0.01$, and let $T_{q_0}$ denote the $q_0$th percentile of $T_0$, $T_{+1}^+$ and $T_{-1}^-$ denote the $q_1$th percentile of $T_{+1}^+$ and $T_{-1}^-$, respectively, and define $T_{q_1} = min(T_{+1}^+, T_{-1}^-)$. Determine the elastic function
(2) by solving the following two equations:

\[ C_1 = g(T_{q_0}), \]  
\[ C_2 = g(T_{q_1}), \]

where the first equation enforces (approximately) full information borrowing, and the second essentially enforces no information borrowing. This leads to the solution

\[ g(T) = \frac{1}{1 + \exp\{a + b \times \log(T)\}}, \]

where

\[ a = \log\left(\frac{1 - C_1}{C_1}\right) - \frac{\log\left(\frac{(1-C_1)C_2}{(1-C_2)C_1}\right) \log(T_{q_0})}{\log(T_{q_0}) - \log(T_{q_1})}, \]
\[ b = \frac{\log\left(\frac{(1-C_1)C_2}{(1-C_2)C_1}\right)}{\log(T_{q_0}) - \log(T_{q_1})}. \]

Several remarks are warranted. In step 3, we generate incongruent cases by simulating \( D_c \) from \( \text{Bernoulli}(\hat{\theta}_h \pm 2\delta) \), rather than \( \text{Bernoulli}(\hat{\theta}_h \pm \delta) \) (i.e., right at the CMD), because the objective of step 3 is to simulate highly incongruent cases to prevent information borrowing by equation (4) in step 4. As it is often regarded as reasonable to borrow some information when the difference between \( D_h \) and \( D_c \) is smaller than CMD, it is thus not appropriate to set the no-borrowing constraint right at the boundary. In step 4, as incongruence can occur in either direction (i.e., \( \theta_c \) is larger or smaller than \( \theta_h \)), we take \( T_{q_1} = \min(T_{q_1}^+, T_{q_1}^-) \) to ensure no information borrowing under the more conservative direction.

In step 4, \( q_0 \) and \( q_1 \) define the borrowing and no borrowing regions (see Figure 2). We may simply choose \( q_0 = q_1 = 0.5 \), i.e., median of \( T_0, T_1^+, \) and \( T_1^- \). A better and more
flexible approach is to choose \( q_0 \) and \( q_1 \) to maximize the trade-off between the power (in the congruent case) and type I error (in the incongruent case). Toward this goal, let \( \rho \) denote the power under the congruent case, \( \psi \) denote the type I error under the incongruent case described in Step 3, and \( \eta \) is a type I error threshold. We define the utility:

\[
U(q_0, q_1) = \rho - w_1 \psi - w_2 (\psi - \eta) I(\psi > \eta),
\]

where \( w_1 \) and \( w_2 \) are penalty weights. This utility imposes a penalty of \( w_1 \) for each unit increase of a type I error before it reaches \( \eta \), and then a penalty of \( w_1 + w_2 \). In our simulation, we set \( w_1 = 1 \), \( w_2 = 2 \), and \( \eta = 0.1 \), which means that before the type I error reaches 0.1, the penalty for a 1% increase of type I errors is to deduct the power by 1%; and once the type I error exceeds 0.1, the penalty for a 1% increase of type I errors increases to deduct the power by 3%. Through a grid search (see Appendix for the procedure), we can identify the \((q_0, q_1)\) that maximize \( U(q_0, q_1) \). Although this approach is more complicated than directly setting \( q_0 = q_1 = 0.5 \), it results in better performance, thus we generally recommend this approach.

A special form of the elastic function, with \( T_{q_0} \equiv T_{q_1} \) (see Figure 1 (b)), is the following step function

\[
g(T) = \begin{cases} 
1 & T \leq T_{q_0} \\
0 & T > T_{q_0}, 
\end{cases}
\]

where full information borrowing occurs if \( T \leq T_{q_0} \), and no information borrowing occurs if \( T > T_{q_0} \). Compared to smooth elastic function (2), one advantage of the step elastic function is that its calibration is simpler, needing only two steps:

1. (Congruent case) Simulate \( R \) replicates of \( D_c = (y_{c,1}, \ldots, y_{c,n_c}) \) from \( \text{Bernoulli}(\hat{\theta}_h) \),
with $\hat{\theta}_h = \bar{y}_h$, and calculate congruence measure $T$ between $D_h$ and each of the simulated $D_c$’s, resulting in $T_0 = (T_1, \ldots, T_R)$, where $T_r$ denote the value of $T$ based on the $r$th simulated $D_c$.

2. Use a grid search to identify the $T_{q_0}$ that maximizes utility $U(q_0)$.

Numerical study shows that the step elastic function can achieve similar operating characteristics as a smooth function, but with greater simplicity, making it a good choice for practical use.

The elastic prior approach has several desirable design characteristics, making it an appealing choice for prespecified analysis. One desirable characteristic is that the elastic function can be fully pre-specified and defined independent of trial data $D_c$. With the pre-specified elastic function, the amount of information borrowing is determined by a pre-specified congruence measure $T$ between historical and current trial data. We expect pre-specification would be a desired characteristics whenever possible. The elastic prior approach satisfies or goes beyond the requirement of pre-specification that “In general, Bayesian CID proposals should include a robust discussion of the prior distribution...a Bayesian proposal should also include a discussion explaining the steps the sponsor took to ensure information was not selectively obtained or used.In cases where downweighting or other non-data-driven features are incorporated in a prior distribution, the proposal should include a rationale for the use and magnitude of these features.” as briefly discussed in the draft Guidance for Industry on Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products.

Another desirable characteristic is the straightforward determination of the prior effective sample size (PESS) contained in the elastic prior, which is simply $g(T)n_h$ as $g(T)$ is a variance inflation factor. In contrast, determining PESS for existing methods (e.g., com-
mensurate prior and robust MAP prior) is more involved, and we found that different PESS calculations used by these methods \cite{13,15} often led to substantially different, sometimes improper results (e.g., PESS > \( n_h \)) \cite{13,16}.

### 2.2 Normal endpoint

Consider a normal endpoint \( y_{c,i} \overset{iid}{\sim} N(\theta, \sigma^2) \) and \( y_{h,i} \overset{iid}{\sim} N(\theta_h, \sigma_h^2) \), with interest in estimating \( \theta \). With a noninformative prior \( \pi_0(\theta_h) \propto 1 \) and historical data \( D_h \), the posterior of \( \theta_h \) is

\[
\pi(\theta_h|D_h, \sigma_h^2) \propto \pi_0(\theta_h) f(D_h|\theta_h, \sigma_h^2) = N(\bar{y}_h, \frac{\sigma_h^2}{n_h}).
\]

An unknown \( \sigma_h^2 \) is often replaced by its maximum likelihood estimate \( \hat{\sigma}_h^2 = \sum_{i=1}^{n_h} (y_{h,i} - \bar{y}_h)^2 / n_h \).

The elastic prior of \( \theta \) is obtained by inflating the variance of \( \pi(\theta_h|D_h, \sigma_h^2) \) with the elastic function \( g(T) \) as follows:

\[
\pi^*(\theta|D_h, \sigma_h^2) = N(\bar{y}_h, \frac{\sigma_h^2}{n_h g(T)}).
\]

An analogue to Section 2.1, the prior effective sample size for \( \pi^*(\theta|D_h, \sigma_h^2) \) is simply \( g(T)n_h \). Full information borrowing is achieved when \( g(T) = 1 \), and no information borrowing occurs when \( g(T) = 0 \). In this scenario, the power prior may obtain similar prior in (2.2). The key difference is that \( g(T) \) is pre-specified to proactively control type I and II error rates and its expected value is known prior to the trial conduct. In addition, as the power prior works by discounting the whole likelihood, it does not allow parameter-specific adaptive information borrowing, for example, when we are interested in estimating and information borrowing on both \( \theta \) and \( \sigma^2 \) as describe later.

The elastic function \cite{2} or step elastic function \cite{7} can be used to dynamically
control information borrowing based on the congruence measure $T$. When subject-level data are available for $D_h$, the Kolmogorov-Smirnov (KS) statistic can be used as the congruence measure between $D_c$ and $D_h$.

$$T = \max_{i=1,\ldots,N} \{|F(Z(i)) - G(Z(i))|\}, \quad (9)$$

where $N = n_c + n_h$; $F(\cdot)$ and $G(\cdot)$ are the empirical distribution functions for $D_h$ and $D_c$, respectively; $Z(1) \leq \cdots \leq Z(N)$ are the $N = m + n$ ordered values for the combined sample of $D_h$ and $D_c$. When $D_h$ only contains summary statistics (e.g., mean and standard error), $t$ statistic is a reasonable choice for $T$,

$$T = \frac{|\bar{y}_c - \bar{y}_h|}{s \sqrt{\frac{1}{n_h} + \frac{1}{n_c}}}, \quad (10)$$

where $s = \sqrt{(n_c-1)s_c^2 + (n_h-1)s_h^2} \over n_c + n_h - 2$ with $s_c^2$ and $s_h^2$ denoting the sample variance of $D_c$ and $D_h$, respectively. For both congruence measures, a larger value of $T$ indicates less congruence between $D_h$ and $D_c$. Again, it can be shown that the resulting elastic prior is information-borrowing consistent, as described in Theorem 1. Also, the choice of $T$ is not unique (e.g., $t$ statistic can also be used when $D_h$ consists of individual-level data) and can be tailored to quantify inferential interest. For example, if the objective of the trial is to compare the variance between the treatment and control arms, the $F$ statistic of testing equal variance is an appropriate measure for the congruence of $D_h$ and $D_c$ in variance. The calibration procedure of elastic function $g(T; \phi)$ is similar to that for the binary endpoint and provided in the Appendix.

If estimation of $\theta$ and $\sigma^2$ is of interest, we can also construct the joint elastic prior for $(\theta, \sigma^2)$. We first apply the noninformative prior $\pi_0(\theta_h, \sigma_h^2) \propto (1/\sigma_h^2)^m$ to $D_h$, where $m$ is
a constant, resulting in the following posterior,

\[ \pi(\theta_h, \sigma^2_h|D_h) \propto \pi_0(\theta_h, \sigma^2_h)f(D_h|\theta_h, \sigma^2_h) \]

\[ \propto N_\theta(\bar{y}_h, \frac{\sigma^2_h}{n_h})IG_{\sigma^2}(\mu_h, \epsilon^2_h), \]

where \( IG(\cdot) \) is an inverse gamma distribution with mean \( \mu_h = \frac{n_h\hat{\sigma}^2_h}{n_h-5+2m} \) and variance \( \epsilon^2_h = \frac{(n_h\hat{\sigma}^2_h)^2}{(n_h-5+2m)^2(\frac{n_h}{2} + m)} \). The joint elastic prior for \( (\theta, \sigma^2) \) is obtained by inflating the variance of \( \pi(\theta_h, \sigma^2_h|D_h) \) with two elastic functions \( g_1(T_1) \) and \( g_2(T_2) \),

\[ \pi^*(\theta, \sigma^2|D_h) \propto N_\theta(\bar{y}_h, \frac{\sigma^2}{n_h g_1(T_1)})IG_{\sigma^2}(\mu_h, \frac{\epsilon^2_h}{g_2(T_2)}), \]

where \( T_1 \) is \( (9) \) or \( (10) \), and \( T_2 \) is the \( F \) statistic of testing equal variance. Allowing parameterspecific information borrowing renders the elastic prior more flexibility than the power prior.

Given the elastic prior and trial data \( D_c \), the posterior distribution for \( (\theta, \sigma^2) \) is

\[ \pi(\theta, \sigma^2|D_c, D_h) \propto N_\theta(\frac{n_h g_1(T_1)\bar{y}_h + n_c\bar{y}_c}{n_h g_1(T_1) + n_c}, \frac{\sigma^2}{n_h g_1(T_1) + n_c})IG_{\sigma^2}(\alpha^*, \beta^*), \]

where \( \alpha^* = \frac{n_c+4}{2} + (\frac{n_h-7}{2} + m)g_2(T_2) \), and \( \beta^* = \sum_{i=1}^{n_c} y^2_{c,i} + n_h g_1(T_1)\bar{y}_h^2 - \frac{\sigma^2}{2n_h g_1(T_1) + n_c} + \frac{n_h\hat{\sigma}^2_h}{n_h-5+2m}[1 + (\frac{n_h-7}{2} + m)g_2(T_2)]. \)

### 2.3 Extension to multiple historical datasets

The proposed method can be extended to borrow information from \( K \) independent historical datasets \( D_{h,1}, \cdots, D_{h,K} \). For notational brevity, we here suppress subscript “\( h \)”, and denote the \( k \)th historical dataset \( D_k = (y_{k,1}, \cdots, y_{k,n_k}) \) with sample mean \( \bar{y}_k = \sum_{i=1}^{n_k} y_{k,i}/n_k \).
The elastic prior can be obtained by sequentially applying the method described above to \( D_1, \ldots, D_K \). Using a binary endpoint as an example, the steps to obtain the elastic prior are

1. Starting with noninformative prior \( \pi_0(\theta) \sim \text{Beta}(\alpha_0, \beta_0) \), obtain the elastic prior \( \pi^*(\theta|D_1) \) for \( D_1 \), where \( \pi^*(\theta|D_1) = \text{Beta}(\alpha_1, \beta_2) \) with \( \alpha_1 = (\alpha_0 + n_1\bar{y}_1)g(T_1) \) and \( \beta_1 = (\beta_0 + n_1 - n_1\bar{y}_1)g(T_1) \).

2. Using \( \pi^*(\theta|D_1) \) as the prior, combining with \( D_2 \), obtain the elastic prior \( \pi^*(\theta|D_1, D_2) \) for \( D_1 \) and \( D_2 \), where \( \pi^*(\theta|D_1, D_2) = \text{Beta}(\alpha_2, \beta_2) \), where \( \alpha_2 = (\alpha_1 + n_2\bar{y}_2)g(T_2) \) and \( \beta_2 = (\beta_1 + n_2 - n_2\bar{y}_2)g(T_2) \).

3. Repeat step 2 sequentially to \( D_3, \ldots, D_K \) to obtain the elastic prior \( \pi^*(\theta|D_1, \ldots, D_K) \).

Elastic functions \( g(T_1), \ldots, g(T_K) \) are calibrated independently using the procedure described previously based on \( D_1, \ldots, D_K \), respectively. One advantage of this sequential elastic prior is that its allow study-specific dynamic information borrowing with minimal interference among \( D_1, \ldots, D_K \). For example, if \( D_1 \) is congruent to \( D_c \) and \( D_2 \) is not congruent to \( D_c \), the elastic prior will borrow more information from \( D_1 \) and less information from \( D_2 \).

Another approach is to aggregate historical information through meta-analysis of \( D_{h,1}, \ldots, D_{h,K} \), and then construct the elastic prior. This can be done using two steps: (1) perform meta-analysis on \( D_{h,1}, \ldots, D_{h,K} \) using the Bayesian hierarchical (or random-effects) model to obtain the posterior predictive distribution of \( \theta \) (i.e., MAP prior), \( \pi(\theta|D_{h,1}, \ldots, D_{h,K}) \); (2) inflate the variance of the MAP prior using the elastic function \( g(T) \) to obtain the elastic prior \( \pi^*(\theta|D_{h,1}, \ldots, D_{h,K}) \). One challenge is how to choose an appropriate statistic \( T \) to measure the congruence between \( D_c \) and \( K \) datasets. The congruence measure \( T \) discussed
previously is applicable to each of \( D_{h,1}, \cdots, D_{h,K} \), but it is not clear how to combine them into a single global congruence measure. To address this issue, we borrow the concept of the posterior predictive model assessment method \[17][18\]. The basic idea is that if \( D_c \) is congruent to \( D_{h,1}, \cdots, D_{h,K} \), we could expect that the actual observed \( D_c \) will be generally consistent with the data generated from \( \pi(D_c|D_{h,1}, \cdots, D_{h,K}) \). Therefore, if the observed \( D_c \) is located on the far tail of the predicted distribution of \( \pi(D_c|D_{h,1}, \cdots, D_{h,K}) \), then \( D_c \) is likely to be incongruent to the historical data. This motivates us to use the posterior predictive p value as the congruence measure \( T \). This approach is general and also can be used for a single historical dataset with various endpoints. Using a normal endpoint as example, \( T \) is calculated as follows:

1. Draw \( R \) samples of \( \theta \) from \( \pi(\theta|D_{h,1}, \cdots, D_{h,K}) \), denoted as \( \theta^{(1)}, \cdots, \theta^{(R)} \). Given \( \theta^{(r)} \), simulate trial data \( D_c = (y_{c,1}, \cdots, y_{c,n_c}) \), and denote its sample mean as \( \bar{y}^{(r)}_c \), \( r = 1, \cdots, R \). In our simulation, we use \( R = 10,000 \).

2. Let \( \bar{y}_c \) denote the actual observed sample mean of \( D_c \); the congruence measure is defined as

\[
T = -\log(PP),
\]

where \( PP = 2 \times \min(\sum_{r=1}^{R} I(\bar{y}^{(r)}_c > \bar{y}_c)/R, \sum_{r=1}^{R} I(\bar{y}^{(r)}_c < \bar{y}_c)/R) \) is the two-sided posterior-predictive p value.

Of note, Theorem 1 and 2, as well as desirable design characteristics, which were discussed in Section 2.1, also apply to Section 2.2 and 2.3.
3 Simulation studies

In this section, we evaluate the finite-sample properties of the elastic prior approach and compare them to some existing methods.

3.1 Simulation setting

We considered scenarios that involve a two-arm superiority trial with one historical data $D_c$, where the endpoint is either a continuous variable with Gaussian distribution or a binary variable with a Bernoulli distribution. For a continuous endpoint, the sample sizes for historical data, control arm, and treatment arm were $n_h = 50$, $n_c = 25$, and $n_t = 50$, respectively. We generated control data $D_c$ from $N(\theta_c, 1^2)$ with $\theta_c = 1$, and treatment data $D_t$ from $N(\theta_t, 1^2)$ with $\theta_t = 1$ and $1.5$. The CMD is $\delta = 0.5$. We generated the historical data $D_h$ from $N(\theta_h, 1^2)$ and varied its mean $\theta_h$ to simulate the scenarios where $D_h$ is congruent or incongruent. For a binary endpoint, the sample sizes for the historical data, control arm, and treatment arm were $n_h = 100$, $n_c = 40$, and $n_t = 80$, respectively. We generated $D_c$ from $Bernoulli(\theta_c)$ with $\theta_c = 0.4$, and $D_t$ from $Bernoulli(\theta_t)$ with $\theta_t = 0.4$, $0.55$, and $0.6$. The CMD is $\delta = 0.12$. We generated $D_h$ from $Bernoulli(\theta_h)$ and varied its mean $\theta_h$ to simulate the scenarios where $D_h$ is congruent or incongruent to $D_c$. We considered the smooth elastic function (2) and step function (7) and denoted them as elastic prior 1 (EP1) and elastic prior 2 (EP2), respectively.

We compared the proposed elastic prior with the commensurate prior (CP), (normalized) power prior (PP), and conventional non-informative prior (NP) that ignores historical data. For CP, we considered two priors for its shrinkage parameter $\tau$ used in publications: $\log(\tau) \sim Unif(-30, 30)$ [5] (denoted as CP1), and spike-and-slab prior with a slab of $(1, 2)$,
spike of 20 and \( \Pr(\text{slab}) = 0.98 \) (denoted as CP2) \cite{16}. For PP, uniform prior \( \text{Unif}(0, 1) \) is used for the power parameter. For EP1 and EP2, we set \( w_1 = 1, w_2 = 2 \), and \( \eta = 0.1 \) in utility to determine the elastic function. For fair comparison, the same criterion is used across the methods to evaluate the efficacy of the treatment, i.e., the treatment is deemed superior to the control if \( \Pr(\theta_t - \theta_c > 0 \mid D_c, D_t, D_h) > C \). The probability cutoff \( C \) is calibrated for each method with 10,000 simulated trials such that under the null (i.e., \( \theta_c = \theta_t = \theta_h \), corresponding to scenario 1 in Tables 1 and 2), the type I error is 5%. The treatment arm does not involve information borrowing and the posterior of \( \theta_t \) is obtained based on the conventional noninformative prior. Under other simulation configurations, we conducted 1000 simulations.

### 3.2 Simulation results

Table 1 shows the results for a normal endpoint. In scenarios 1 and 2, \( D_h \) and \( D_c \) are congruent. When the treatment is not effective (i.e., scenario 1), all methods control the type I error rate at its nominal value of 5%. When the treatment is effective (i.e., scenario 2), EP1, EP2, CP1, and PP offer substantial power gain over NP. For example, the power of EP1 is 27.1% higher than NP, and also slightly higher than CP1 and PP. EP2 has comparable performance to EP1. In contrast, CP2 provides little power improvement, indicating that the spike-and-slab prior is too conservative to borrow information. Similar results are observed in scenarios 3 to 4, where \( D_h \) and \( D_c \) are approximately congruent. Scenarios 5-8 consider the case that \( D_h \) and \( D_c \) are incongruent. Specifically, in scenarios 5 and 6, the treatment is not effective, and the results are type I errors. Compared to CP1 and PP, EP1 and EP2 offer better type I error control. For example, in scenario 5, the type I errors of EP1 and EP2 are 7.7% and 7.3%, whereas the type I errors of CP1 and PP are 14.6% and 30%, respectively.
CP2 has little type I inflation because it barely borrows information, demonstrated by its low power when the $D_h$ and $D_c$ are congruent (i.e., scenarios 3 and 4). In scenarios 7 and 8, the treatment is effective, and the results are power. EP1 and EP2 yield higher power to detect the treatment effect than CP1 and PP. For example, in scenario 7, the power of EP2 is 15.0% and 34.8% higher than CP1 and PP, respectively.

Table 2 shows the results for a binary endpoint, which are generally consistent with these for normal endpoint. Scenarios 1 to 4 consider the case that $D_h$ and $D_c$ are congruent or approximately congruent. In scenario 1, the treatment is not effective; all methods control the type I error rate at its nominal value of 5%. In scenario 2, the treatment is effective; EP1, EP2, CP1, and PP offer substantial power gain over NP. For example, the power of EP1 is 15.9% higher than NP, and comparable to CP1 and PP. Akin to the normal endpoint, CP2 is similar to NP with little information borrowing. Similar results are observed in scenarios 3 and 4, where $D_h$ and $D_c$ are approximately congruent. Scenarios 5-8 consider the case that $D_h$ and $D_c$ are incongruent. Specifically, in scenarios 5 and 6 the treatment is not effective, and the results are type I errors. Compared to CP1, CP2 and PP, EP1 and EP2 offer better type I error control. For example, in scenario 5, the type I error of EP1 and EP2 is approximately 1/2 and 1/4 of that of CP1, 1/3 and 1/5 of PP, and 3.6% (7.4%) lower than CP2. In scenarios 7 and 8, the treatment is effective, and the results are power. EP1 and EP2 yield higher power to detect the treatment effect than found with CP1 and PP. For example, in scenario 7, the power of EP1 and EP2 are more than double that of CP1 and PP.
3.3 Multiple historical datasets

Taking a similar setting as the simulation with one historical dataset, we generated control arm data $D_c$ from $N(\theta_c, 1^2)$ with $\theta_c = 1$ and sample size $n_c = 25$, and treatment arm data $D_t$ from $N(\theta_t, 1^2)$ with $\theta_t = 1, 1.5$ and sample size $n_t = 50$. The CMD is $\delta = 0.5$. We considered four historical datasets with sample size 40, 50, 45, and 55, respectively, generated from the following hierarchical model:

$$y_k \sim N(\theta_k, 1^2), k = 1, \cdots, 4,$$

$$\theta_1, \theta_2, \theta_3, \theta_4 \sim N(\theta_h, 0.1^2).$$

We varied $\theta_h$ to simulate scenarios where $D_h$ is congruent or incongruent to $D_c$. Similarly, we considered both the smooth elastic and step functions, and denoted them as elastic MAP 1 (EMAP1) and elastic MAP 2 (EMAP2), respectively.

We compared the elastic MAP priors with the robust MAP prior. Following Schmidli et al (2014) [8], we considered two versions of the robust MAP prior: Mix50 with a weight of 0.5 and the Mix90 design with a weight of 0.1 assigned to MAP. As the benchmark, we also considered the conventional NP that ignores historical data. The treatment is deemed superior to the control if $\Pr(\theta_t - \theta_c > 0 \mid D_c, D_t, D_h) > C$. The probability cutoff $C$ is calibrated for each method with 10,000 simulated trials such that under the null (i.e., $\theta_c = \theta_t = \theta_h$, corresponding to scenario 1 in Table 3), the type I error is 5%. Under other simulation configurations, we conducted 1000 simulations.

Table 3 shows the results. When historical data and control data are congruent (i.e., scenarios 1 to 4), EMAP1 and EMAP2 have comparable performance to Mix50 and Mix90. All methods control type I errors at the nominal value of 5% (scenario 1), and they
yield substantially higher power than the NB due to borrowing information from historical datasets. Scenarios 5-8 consider the case that historical data and control data are incongruent. In scenarios 5-6, the treatment is ineffective and the results are type I errors. EMAP1 and EMAP2 offer better type I error control than the robust MAP. For example, in scenario 5, the type I error of EMAP1 and EMAP2 are 8.5% and 7.8%, whereas that of Mix50 and Mix90 are 14.1% and 26.4%, respectively. In addition, EMAP1 and EMAP2 provide substantial power gain over Mix50 and Mix90. For example, in scenario 7, the power of EMAP1 is 17.5% and 25.9% higher than Mix50 and Mix90, respectively, and EMAP2 has 23.3% and 31.7% higher power than Mix50 and Mix90, respectively.

4 Conclusion

We have proposed the elastic prior to dynamically borrow information from historical data. Through the use of elastic function, the elastic prior approach adaptively borrows information based on the congruence between trial data and historical data. The elastic function is constructed based on a set of information-borrowing constraints prespecified such that the prior will borrow information when historical and trial data are congruent, and refrain from information borrowing when historical and trial data are incongruent. The elastic prior is information-borrowing consistent, and is easy to quantify using a prior effective sample size. Simulation study shows that, compared to existing methods, the elastic prior has better type I error control, and yields competitive or higher power. In addition, we provide insights on what can and cannot be achieved using the information-borrowing method, which is useful for guiding future methodology development.

The good performance of the elastic prior stems from the use of elastic function to
regulate the behavior of information borrowing within the range of the parameter space of practical interest. That is, the elastic prior does not completely rely on the data to determine information borrowing. It also incorporates the subject matter knowledge (e.g., when it should borrow or not) to enhance and govern the performance. In contrast, many existing methods intend to achieve dynamic information borrowing by estimating the information-borrowing parameter (e.g., power parameter in power prior or shrinkage parameter in commensurate prior), jointly with model parameters, based on data. However, the data contain extremely limited information for estimating the information-borrowing parameter because the observation unit contributing to the estimation is the dataset, not subject-level observations. For example, one historical data and one trial data actually provide only two observations to estimate the information-borrowing parameter. This is a well-known issue in meta-analysis for estimating the between-study variation. As a result, these dynamic information borrowing methods cannot reliably sense the congruence/incongruence between historical data and trial data to perform appropriate information borrowing.

The idea of an elastic prior is general, and it also can be applied to both commensurate and power priors to improve their operating characteristics. We outline this approach in the Appendix. In addition, we have focused on two-arm randomized superiority trials with binary or normal endpoints. The methodology can be applied to single-arm and multiple-arm trials, as well as other types of trials, for example, noninferiority trials. Extension of the elastic prior to the time-to-event endpoint is of practical interest and warrants further research.

The type I error considered in this paper was referred as the usual view, where type I error rate is based on current trial(s) alone. Pennello and Thompson (2008) also discussed a view 2, considering type I error rate based on current trial(s) and prior data considered together. Type I error rate with view 2 might be considered when we
extrapolating adult results to a pediatric setting, when know before the adult trials, that the analysis in the pediatric setting will borrow from the adult results because of similarity in the course of disease, response to treatment, pharmacokinetic, and pharmacodynamic.

**Disclaimer**

This article reflects the views of the author, and it should not be construed to represent FDA views or policies.
Figure 1: (a) A class of smooth elastic functions defined by $g(T) = \frac{1}{1+\exp[a+b \times \{\log(T)\}]^c}$, and (b) a step elastic function, where $g(T) = 1$ leads to full information borrowing and $g(T) = 0$ leads to no information borrowing.

Figure 2: Dynamic information borrowing through the elastic function.
Table 1: Simulation results for a normal endpoint using a noninformative prior (NP), elastic prior with the smooth elastic function (EP1) and step elastic function (EP2), commensurate prior with uniform prior (CP1) and spike-and-slab prior (CP2), and power prior (PP).

| Scenario | $\theta_h$ | $\theta_c$ | $\theta_t$ | NP       | EP1       | EP2       | CP1       | CP2       | PP       |
|----------|------------|------------|------------|----------|----------|----------|----------|----------|----------|
| Congruent|            |            |            |          |          |          |          |          |          |
| 1*       | 1          | 1          | 1          | 5.0      | 5.06(48.87) | 5.17(49.20) | 5.23(42.37) | 4.92(4.41) | 5.04(28.70) |
| 2        | 1          | 1          | 1.5        | 66.3     | 93.4(48.87) | 93.6(49.20) | 91.6(42.37) | 68.4(4.41) | 88.3(28.70) |
| 3        | 0.9        | 1          | 1.5        | 66.3     | 94.6(48.28) | 95.0(48.75) | 93.0(40.74) | 69.7(4.39) | 92.0(28.48) |
| 4        | 1.1        | 1          | 1.5        | 66.3     | 86.0(48.14) | 86.3(48.75) | 85.5(40.70) | 67.7(4.39) | 83.1(28.47) |
| Incongruent|          |            |            |          |          |          |          |          |          |
| 5*       | 0          | 1          | 1          | 5.0      | 7.7(0.31)  | 7.3(0.35)  | 14.6(0.41) | 7.0(3.60) | 30.0(9.55)  |
| 6*       | -0.5       | 1          | 1          | 5.0      | 7.1(0.00)  | 7.0(0.00)  | 10.0(1.34) | 8.8(3.24) | 18.3(3.19)  |
| 7        | 2          | 1          | 1.5        | 66.3     | 72.3(0.24) | 72.6(0.35) | 57.6(0.36) | 60.7(3.58) | 37.8(9.47)  |
| 8        | 2.5        | 1          | 1.5        | 66.3     | 72.3(0.00) | 72.6(0.00) | 69.2(1.37) | 57.6(3.25) | 49.1(3.14)  |

*Type I error
Table 2: Simulation results for a binary endpoint using a noninformative prior (NP), elastic prior with the smooth elastic function (EP1) and step elastic function (EP2), commensurate prior with uniform prior (CP1) and spike-and-slab prior (CP2), and power prior (PP).

| Scenario | $\theta_h$ | $\theta_c$ | $\theta_t$ | Percentage of claiming efficacy (PESS) |
|----------|------------|------------|------------|--------------------------------------|
|          |            |            |            | NP | EP1 | EP2 | CP1 | CP2 | PP |
| Congruent|            |            |            |    |    |    |     |     |    |
| 1*       | 0.4        | 0.4        | 0.4        | 5  | 5.09 | 5.04 | 4.97 | 5.10 | 5.27 | 5.16 |
| 2         | 0.4        | 0.4        | 0.6        | 5  | 5.09 | 91.3 | 90.8 | 91.2 | 76.0 | 90.1 |
| 3         | 0.35       | 0.4        | 0.55       | 5  | 5.09 | 80.6 | 80.9 | 83.1 | 58.4 | 79.9 |
| 4         | 0.42       | 0.4        | 0.6        | 5  | 5.09 | 88.0 | 87.3 | 90.0 | 75.0 | 87.3 |
| Incongruent|           |            |            |    |    |    |     |     |    |
| 5*       | 0.16       | 0.4        | 0.4        | 5  | 5.09 | 10.3 | 6.5  | 23.7 | 13.9 | 30.1 |
| 6*       | 0.10       | 0.4        | 0.4        | 5  | 5.09 | 7.0  | 6.0  | 14.8 | 18.6 | 24.6 |
| 7         | 0.6        | 0.4        | 0.55       | 5  | 5.09 | 48.4 | 46.0 | 18.6 | 41.0 | 22.2 |
| 8         | 0.6        | 0.4        | 0.6        | 5  | 5.09 | 67.6 | 65.9 | 38.2 | 64.8 | 43.8 |

*Type I error
Table 3: Simulation results for multiple historical studies using a noninformative prior (NP), elastic MAP prior with the smooth elastic function (EMAP1) and step elastic function (EMAP2), robust MAP priors with 50% mixture (Mix50) and 90% mixture (Mix90).

| Scenario | $\theta_h$ | $\theta_c$ | $\theta_t$ | Percentage of claiming efficacy (PESS) |
|----------|------------|------------|------------|---------------------------------------|
|          |            |            |            | NP  | EMAP1     | EMAP2     | Mix50    | Mix90    |
| **Congruent** |            |            |            |     |            |            |          |          |
| 1*       | 1          | 1          | 1          | 5.00| 5.18(29.32)| 5.37(29.09)| 4.99(14.99)| 4.91(29.18) |
| 2        | 1          | 1          | 1.5        | 66.3| 91.0(29.32)| 90.7(29.09)| 90.1(14.99)| 90.9(29.18) |
| 3        | 0.9        | 1          | 1.5        | 66.3| 94.2(29.12)| 93.3(28.85)| 93.1(14.99)| 94.1(29.18) |
| 4        | 1.1        | 1          | 1.5        | 66.3| 84.6(28.90)| 84.3(28.59)| 83.9(14.99)| 84.2(29.18) |
| **Incongruent** |            |            |            |     |            |            |          |          |
| 5*       | 0          | 1          | 1          | 5.0 | 8.5(0.65)  | 7.8(0.38) | 14.1(14.99)| 26.4(29.18) |
| 6*       | -0.2       | 1          | 1          | 5.0 | 7.7(0.03)  | 7.6(0.00) | 9.7(15.00)| 16.2(29.18) |
| 7        | 1.6        | 1          | 1.5        | 66.3| 61.3(11.45)| 67.1(8.41)| 43.8(14.99)| 35.4(29.18) |
| 8        | 2          | 1          | 1.5        | 66.3| 75.3(0.00) | 75.1(0.15)| 63.0(15.00)| 47.8(29.18) |

*Type I error
References

[1] US Food and Drug Administration. (2017). Use of real-world evidence to support regulatory decision-making for medical devices. Guidance for industry and food and drug administration staff.

[2] US Food and Drug Administration. (2019). Submitting documents using real-world data and real-world evidence to FDA for drugs and biologics: guidance for Industry: draft guidance. Rockville, MD: US Food and Drug Administration.

[3] Ibrahim, J. G., & Chen, M. H. (2000). Power prior distributions for regression models. Statistical Science, 15(1), 46-60.

[4] Ibrahim, J. G., Chen, M. H., & Sinha, D. (2003). On optimality properties of the power prior. Journal of the American Statistical Association, 98(461), 204-213.

[5] Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., & Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics, 67(3), 1047-1056.

[6] Thall, P. F., Wathen, J. K., Bekele, B. N., Champlin, R. E., Baker, L. H., & Benjamin, R. S. (2003). Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. Statistics in Medicine, 22(5), 763-780.

[7] Berry, S. M., Broglio, K. R., Groshen, S., & Berry, D. A. (2013). Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. Clinical Trials, 10(5), 720-734.

[8] Schmidli, H., Gsteiger, S., Roychoudhury, S., O’Hagan, A., Spiegelhalter, D., & Neuenschwander, B. (2014). Robust metaanalytic-predictive priors in clinical trials with historical control information. Biometrics, 70(4), 1023-1032.
[9] Pan, H., Yuan, Y., & Xia, J. (2017). A calibrated power prior approach to borrow information from historical data with application to biosimilar clinical trials. Journal of the Royal Statistical Society: Series C (Applied Statistics), 66(5), 979-996.

[10] Neuenschwander, B., Branson, M., & Spiegelhalter, D. J. (2000). A note on the power prior, Statistics in Medicine, 28(28), 3562-3566.

[11] Freidlin, B., & Korn, E. L. (2013). Borrowing information across subgroups in phase II trials: is it useful?. Clinical Cancer Research, 19(6), 1326-1334.

[12] Chu, Y., & Yuan, Y. (2018). BLAST: Bayesian latent subgroup design for basket trials accounting for patient heterogeneity. Journal of the Royal Statistical Society: Series C (Applied Statistics), 67(3), 723-740.

[13] Hobbs, B. P., Carlin, B. P., & Sargent, D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. Clinical Trials, 10(3), 430-440.

[14] Morita, S., Thall, P. F., & Müller, P. (2008). Determining the effective sample size of a parametric prior. Biometrics, 64(2), 595-602.

[15] Neuenschwander, B., Weber, S., Schmidli, H., & OHagan, A. (2020). Predictively consistent prior effective sample sizes. Biometrics, 1-10.

[16] Chen, N., Carlin, B. P., & Hobbs, B. P. (2018). Web-based statistical tools for the analysis and design of clinical trials that incorporate historical controls. Computational Statistics & Data Analysis, 127, 50-68.

[17] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis. CRC press.

[18] Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica, 6(4), 733-760.
[19] Pennello, G., & Thompson, L. (2008). Experience with Reviewing Bayesian Medical Device Trials. Journal of Biopharmaceutical Statistics, 18(1), 81-115.
Appendix

A. Proof of Theorem 1
Suppose the chi-square test statistic is used to measure the congruency $T$ between $D_h$ and $D_c$. Since the chi-square statistic of homogeneity is consistent, $T \to 0$ as $n_h \to \infty$ and $n_c \to \infty$, when $D_h$ and $D_c$ are congruent. Given $b > 0$, $g(T) = \frac{1}{1 + \exp\{a + b \times \log(T)\}} \to 1$. Consequently, the elastic prior fully borrows historical information. When $D_h$ and $D_c$ are incongruent, $T \to \infty$ as $n_h \to \infty$ and $n_c \to \infty$. Given $b > 0$, $g(T) = \frac{1}{1 + \exp\{a + b \times \log(T)\}} \to 0$, and thus no historical information will be borrowed.

B. Grid search for percentile combination $(q_0, q_1)$
Let $(q_0^{(1)}, \cdots, q_0^{(J)})$ and $(q_1^{(1)}, \cdots, q_1^{(K)})$ denote the prespecified searching grid for $q_0$ and $q_1$, respectively. We used $q_0^{(1)} = q_1^{(1)} = 0.3$ and $q_0^{(J)} = q_1^{(K)} = 0.9$, and set a grid step of 0.1. The following steps are used to find the $(q_0, q_1)$ that optimizes the utility $U(q_0, q_1)$.

1. Given a specific grid $(q_0^{(j)}, q_1^{(k)})$, determine the elastic function using equation (5).

2. Given the obtained elastic function, under the congruent case ($\theta_h = \theta_c$), calibrate the probability cutoff $C$ to control the type I error rate at a nominal value of 5% and compute the power ($\rho$) through simulation.

3. Given the cutoff $C$, compute the type I error ($\psi$) under the incongruent case (e.g., $\theta_c = \theta_h - 2\delta$).

4. Identify $(q_0^{(j)}, q_1^{(k)})$ that produces the largest value of $U(q_0^{(j)}, q_1^{(k)}) = \rho - w_1\psi - w_2(\psi - \eta)I(\psi > \eta)$.

For the step elastic function, the calibration of $q_0$ is similar to that shown above.
The main difference is that we only need to search over a one-dimensional grid \( (q_0^{(1)}, \cdots, q_0^{(J)}) \), which greatly reduces the optimization time.

C. Determination of elastic function for a normal endpoint

The steps to determine elastic function are similar to these for the binary endpoint, and described as follows:

1. Estimate the mean and variance of \( D_h \) by \( \hat{\theta}_h = \bar{y}_h \) and \( \hat{\sigma}_h^2 = \frac{\sum_{i=1}^{n_h} (y_{h,i} - \bar{y}_h)}{(n_h - 1)} \) with \( \bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{h,i}}{n_h} \).

2. Elicit from subject matter experts a clinically meaningful difference \( \delta \) for \( E(y) \).

3. (Congruent case) Simulate \( R \) replicates of \( D_c = (y_{c,1}, \cdots, y_{c,nc}) \) from \( N(\hat{\theta}_h, \hat{\sigma}_h^2) \), and calculate congruence measure \( T \) between \( D_h \) and each simulated \( D_c \), resulting in \( T_0 = (T_1, \cdots, T_R) \), where \( T_r \) denote the value of \( T \) based on the \( r \)th simulated \( D_c \).

4. (Incongruent cases) Simulate \( R \) replicates of \( D_c \) from \( N(\hat{\theta}_h + 2\delta, \hat{\sigma}_h^2) \), and calculate congruence measure \( T \) between \( D_h \) and each simulated \( D_c \), resulting in \( T_1^+ = (T_1^+, \cdots, T_R^+) \), where \( T_r^+ \) denote the value of \( T \) based on the \( r \)th simulated \( D_c \). Repeat this with \( D_c \) simulated from \( N(\hat{\theta}_h - 2\delta, \hat{\sigma}_h^2) \), resulting in \( T_1^- = (T_1^-, \cdots, T_R^-) \), where \( T_r^- \) denote the value of \( T \) between \( D_h \) and the \( r \)th simulated \( D_c \).

5. Let \( C_1 \) and \( C_2 \) be constants close to 1 and 0, respectively, e.g., \( C_1 = 0.99 \) and \( C_2 = 0.01 \), and let \( T_{q_0} \) denotes the \( q_0 \)th percentile of \( T_0 \), \( T_{q_1}^+ \) and \( T_{q_1}^- \) denote the \( q_1 \)th percentile of \( T_1^+ \) and \( T_1^- \), respectively, and define \( T_{q_1} = min(T_{q_1}^+, T_{q_1}^-) \).

6. Based on \( T_{q_0} \) and \( T_{q_1} \), determine the elastic function (2) by equation (5).

D. Elastic power prior and elastic commensurate prior

The idea of an elastic prior also can be applied to the power prior and commensurate prior,
and we refer to them as elastic power prior and elastic commensurate prior.

**D1. Elastic power prior**

With the power prior, the power parameter $\delta$ is treated as an unknown parameter, while with an elastic power prior, $\delta$ is linked with $T$ by an elastic function $g(\cdot)$, that is,

$$\delta = g(T; \phi). \quad (11)$$

Then the elastic power prior is given by

$$\pi^*(\theta | D_h) \propto \pi_0(\theta) f(D_h | \theta)^{g(T)}, \quad (12)$$

where the elastic function $g(T)$ is same as equation [2], which maps support of $T$ to $[0, 1]$. Actually, the elastic power prior is identical to the calibrated power prior proposed by Pan et al (2017) [9].

Following the notations and initial priors described in Section 2, we display the elastic power prior for normal and binary endpoints. For the normal endpoint, the joint elastic power prior of $(\theta, \sigma^2)$ is

$$\pi^*(\theta, \sigma^2 | D_h) \propto \left( \frac{1}{\sigma^2} \right)^{m + \frac{\sigma^2 n_h}{2}} \exp \left[ - \frac{g(T)n_h}{2\sigma^2} \left( \hat{\sigma}_h^2 + (\theta - \bar{y}_h)^2 \right) \right]$$

$$\propto N(\bar{y}_h, \frac{\sigma^2}{g(T)n_h})IG_{\sigma^2}(m + \frac{g(T)n_h - 3}{2}, \frac{g(T)n_h\hat{\sigma}_h^2}{2}). \quad (13)$$

Given current data, the posterior distribution for $(\theta, \sigma^2)$ is

$$\pi(\theta, \sigma^2 | D, D_h) \propto N(\frac{g(T)n_h\bar{y}_h + n_c\bar{y}_c}{g(T)n_h + n_c}, \frac{\sigma^2}{g(T)n_h + n_c})IG_{\sigma^2}(\alpha^\Delta, \beta^\Delta), \quad (14)$$

34
\[ \alpha^\Delta = m + \frac{g(T)n_h + n_c - 3}{2}, \quad \beta^\Delta = \frac{\Sigma_{i=1}^{n_i} y_{i,h}^2 + g(T)n_h y_h^2 + g(T)n_h \sigma_h^2}{2n_h g(T) + n_c} - \frac{(g(T)n_h y_h + n_c y_c)^2}{2n_h g(T) + n_c}. \]

For a binary endpoint, the elastic power prior of \( p \) is

\[ \pi^*(p|D_h) \propto p^{g(T)n_h y_h + \alpha_0 - 1}(1 - p)^{g(T)(n_h - n_h y_h) + \beta_0 - 1} \]

\[ \propto Beta(g(T)n_h y_h + \alpha_0, g(T)(n_h - n_h y_h) + \beta_0). \]  

(15)

Based on the current data \( D_c \), the posterior of \( p \) is given as

\[ \pi(p|D, D_h) \propto Beta(g(T)n_h y_h + \alpha_0 + n_c y_c, g(T)(n_h - n_h y_h) + \beta_0 + n_c - n_c y_c). \]  

(16)

### D2. Elastic commensurate prior

With a commensurate prior, shrinkage parameter \( \tau \) controls the degree that \( \theta \) shrinks to \( \theta_h \), and it is assumed unknown with a prior. However, with an elastic commensurate prior, \( \tau \) is determined by \( T \) through the elastic function \( g(T) \), i.e.,

\[ \tau = g(T; \phi). \]  

(17)

Then the elastic commensurate prior for \( \theta \) is

\[ \pi^*(\theta|D_h, g(T)) \propto \int_{\theta_h} f(D_h|\theta_h) \pi(\theta|\theta_h, g(T)) \pi_0(\theta_h) d\theta_h. \]  

(18)

Since \( \tau \) is located in \((0, +\infty)\), we adopt the following elastic function:

\[ g(T) = \exp(a + b \cdot \log(T)). \]  

(19)
If a larger value of $T$ indicates more incongruence between $D_c$ and $D_h$, we require $b < 0$ to ensure that a larger value of $T$ leads to a smaller value of $g(T)$ (i.e., a larger variance inflation). The calibration of $g(T)$ is similar to that described in Section 2.

Let us return to the Gaussian case. We first focus on the historical information borrowing for location parameter $\theta$, that is $\theta|\theta_h \sim N(\theta_h, \tau^{-1})$, where $\tau = g(T)$. Assuming $\pi_0(\theta_h) \propto 1$ and integrating out the nuisance parameter $\theta_h$, the elastic commensurate prior for $\theta$ is

$$
\pi^*(\theta|D_h, g(T)) \propto N(\bar{y}_h, \frac{1}{g(T)} + \frac{\hat{\sigma}^2_h}{n_h}).
$$  \hspace{1cm} (20)

Multiplying the above elastic commensurate prior with the current likelihood, we obtain the following posterior distribution for $\theta$:

$$
\pi(\theta|D, D_h, \sigma^2) \propto N(\frac{n_c\bar{y}_c\Delta + \sigma^2\bar{y}_h}{n_c\Delta + \sigma^2}, \frac{\sigma^2\Delta}{n_c\Delta + \sigma^2}),
$$  \hspace{1cm} (21)

where $\Delta = \frac{1}{g(T)} + \frac{\hat{\sigma}^2_h}{n_h}$.

If the information borrowing both for location parameter $\theta$ and scale parameter $\sigma^2$ are required, a new precision parameter $\zeta$ is introduced to measure the commensurate between $\sigma^2$ and $\sigma^2_h$. Specifically, we assume $\sigma^2$ a prior that is centered at $\sigma^2_h$ with precision $\zeta$, e.g., $\sigma^2|\sigma^2_h \sim IG(\sigma^2_h, \zeta^{-1})$, where $IG(\cdot)$ is an inverse gamma distribution with mean $\sigma^2_h$ and variance $\zeta^{-1}$. With an elastic commensurate prior, precision $\tau = g_1(T_1)$ and $\zeta = g_2(T_2)$. Given historical data $D_h$, assuming a prior $\pi_0(\sigma^2_h) \propto (\sigma^2_h)^{-m}$ for $\sigma^2_h$ and integrating out $\theta_h$,
the joint elastic commensurate prior for $(\theta, \sigma^2)$ is

\[
\pi^*(\theta, \sigma^2, \sigma_h^2|D_h, g_1(T_1), g_2(T_2)) \propto f(D_h|\theta_h, \sigma_h^2) N(\theta_h, g_1(T_1)^{-1}) IG_{\sigma^2}(\alpha', \beta') \times (\sigma_h^2)^{-m}
\]

\[
\times N(\bar{y}_h, \frac{1}{g_1(T_1)} + \frac{\sigma_h^2}{n_h}) IG_{\sigma^2}(\alpha', \beta')
\]

\[
\times IG_{\sigma_h^2}(\frac{n_h - 3}{2} + m, \frac{n_h \hat{\sigma}^2}{2}),
\]

where $\alpha' = g_2(T_2)\sigma^4_h + 2$ and $\beta' = \sigma^2_h(g_2(T_2)\sigma^4_h + 1)$. 

37