

# UCSF

## UC San Francisco Previously Published Works

### Title

Bayesian adaptive design for covariate-adaptive historical control information borrowing.

### Permalink

<https://escholarship.org/uc/item/99w3q0wm>

### Journal

Statistics in Medicine, 42(29)

### Authors

Scheffler, Aaron

Kim, Mi-ok

Jiang, Fei

et al.

### Publication Date

2023-12-20

### DOI

10.1002/sim.9913

Peer reviewed



Published in final edited form as:

*Stat Med.* 2023 December 20; 42(29): 5338–5352. doi:10.1002/sim.9913.

## Bayesian adaptive design for covariate-adaptive historical control information borrowing

Huaqing Jin<sup>1</sup>, Mi-Ok Kim<sup>2</sup>, Aaron Scheffler<sup>2</sup>, Fei Jiang<sup>2</sup>

<sup>1</sup>Department of Radiology and Biomedical Imaging, University of California, San Francisco, California

<sup>2</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, California

### Abstract

Interest in incorporating historical data in the clinical trial has increased with the rising cost of conducting clinical trials. The intervention arm for the current trial often requires prospective data to assess a novel treatment, and thus borrowing historical control data commensurate in distribution to current control data is motivated in order to increase the allocation ratio to the current intervention arm. Existing historical control borrowing adaptive designs adjust allocation ratios based on the commensurability assessed through study-level summary statistics of the response agnostic of the distributions of the trial subject characteristics in the current and historical trials. This can lead to distributional imbalance of the current trial subject characteristics across the treatment arms as well as between current control data and borrowed historical control data. Such covariate imbalance may threaten the internal validity of the current trial by introducing confounding factors that affect study endpoints. In this article, we propose a Bayesian design which borrows and updates the treatment allocation ratios both covariate-adaptively and commensurate to covariate dependently assessed similarity between the current and historical control data. We employ covariate-dependent discrepancy parameters which are allowed to grow with the sample size and propose a regularized local regression procedure for the estimation of the parameters. The proposed design also permits the current and the historical controls to be similar to varying degree, depending on the subject level characteristics. We evaluate the proposed design extensively under the settings derived from two placebo-controlled randomized trials on vertebral fracture risk in post-menopausal women.

### Keywords

Bayesian; covariate-adaptive; high dimensional; historical sample; kernel

---

**Correspondence:** Fei Jiang, Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA. Fei.Jiang@ucsf.edu.

#### CONFLICT OF INTEREST STATEMENT

The authors have declared no conflicts of interest.

#### SUPPORTING INFORMATION

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

## 1 | INTRODUCTION

Historical data for the control arm of a prospective clinical trial is often available from previous randomized controlled clinical trials. Such data are conventionally used to establish design parameters in the planning stage, such as to perform sample size calculations based on variance estimates for targeted outcomes. Since the current trial may be different from the historical ones in many ways, for example, the characteristics of study participants and outcome definitions, it would be ideal to evaluate how comparable the historical and concurrent control data are and adaptively decide how much information can be borrowed to infer the parameters in the current trial control arm. Such evaluation can facilitate properly incorporating the historical information in order to improve the statistical power or reduce the sample size on the control arm. On the other hand, it is well known that balancing study participant profiles across different treatment arms in a trial with limited sample size is hard, whereas failure to do so could lead to biased estimates or lower power.<sup>1</sup> Thus, controlling the covariate imbalance across the treatment arms of the current trial is also important in borrowing historical control data. All of the existing adaptive designs control the amount of borrowing by assessing the discrepancy between the historical data and the current trial data at the study level and balancing of the covariates in the process is not a consideration (see, e.g., Jin and Yin,<sup>2</sup> Kim et al.<sup>3</sup>). Here, we assume the historical data are available at the individual participant level and propose covariate-adaptive evaluation and borrowing at the individual participant level.

Our motivation arises from two placebo-controlled randomized trials studying the effects of bisphosphonate treatment on vertebral fracture risk in post-menopausal women, the health outcomes and reduced incidence with zoledronic acid once yearly (HORIZON) pivotal fracture trial<sup>4</sup> and the fracture intervention trial (FIT).<sup>5,6</sup> The trials evaluated zoledronic acid and alendronate as interventions to reduce vertebral fracture risk. We consider a particular clinical population of interest, elderly women aged approximately 80 years or above who are at elevated vertebral fracture risk compared to younger women and thus suffer disproportionately with respect to disability and medical costs resulting from fracture events.<sup>4,7</sup> Both studies enrolled women using the similar enrolment criteria including baseline hip bone mineral density (BMD) and a lack of prior bisphosphonate use. We use the hip BMD measured at 24 months by dual-energy x-ray absorptiometry as our common study outcome for the two trials, an established objective surrogate endpoint for vertebral fracture to determine treatment efficacy.<sup>8</sup> These are large trials each enrolling 6459 and 7736, but the number enrolled for the target population of interest is 388 and 620 for the FIT and the HORIZON studies, respectively.

The placebo control arm data in the FIT study, which was completed in 1996, can be used as historical data for the latter initiated HORIZON trial, permitting borrowing of the control arm information corresponding to 69.5% (the ratio between historical control and current control sample sizes) of the current trial enrollment. The risk of vertebral fracture varies by factors such as the history of a vertebral fracture or a fall and the years since menopause. It would be ideal to match historical control participants with the current trial ones on those important prognostic factors to assess the similarity between the historical and the current trial data and to update allocation ratios to control balance of the important prognostic

factors by accounting for the borrowed historical information. No existing designs meet this need.

Little is found on developing a design that adaptively control both historical information borrowing and balance of the covariate distribution. The idea of using historical control data has been around in the clinical trial world since the seminal work by Pocock,<sup>9</sup> Dempster, Selwyn and Weeks<sup>10</sup> and Ryan<sup>11</sup> which combine the current and historical data to estimate the treatment effect by discounting historical data to account for the between-trial heterogeneity. Interest in the use of historical controls has increased in recent years. Hong, Fu and Carlin<sup>12</sup> and Wang et al.<sup>13</sup> synthesize historical information for better parameter estimation and inference in network meta-analysis. Several Bayesian designs address the between-trial heterogeneity by determining the degree of borrowing on the control arm using power prior,<sup>14,15</sup> commensurate prior,<sup>16</sup> meta-analytic-predictive prior<sup>17,18</sup> and the unit information prior.<sup>2</sup> Kim et al.<sup>3</sup> develop a sequential design that uses borrowed historical information to adjust allocation ratios in order to improve the current trial participants outcomes and raises the probability of early trial stopping.

In parallel, many covariate-adaptive randomization designs have been developed to achieve covariate balancing across treatment arms in clinical trials.<sup>19</sup> For discrete covariates, these methods include the biased coin covariate-adaptive randomization design,<sup>20,21</sup> which is an extension of the biased coin design<sup>22</sup> for balancing the sample size, and the Pocock–Simon design which is based on a minimization method for the sequential treatment assignment.<sup>23,24</sup> Recently, Hu and Hu<sup>25</sup> and Jiang et al.<sup>26</sup> extend the covariate-adaptive designs to handle both discrete and continuous covariates.

We propose a covariate-adaptive historical control borrowing Bayesian design (CAHB) integrating the historical trial information in a covariate-adaptive randomization design to achieve the goals of balancing the covariate specific information on the treatment arms when additional historical data are available. Specifically with the application of the design to the vertebral fracture risk study, the prior of the mean 24-month BMD outcome for the current control arm depends on the historical estimates, while the degree of the information borrowing is covariate-adaptively determined by a precision parameter that measures the difference between the mean 24-month BMD from historical and current control arms locally using study participants sharing similar characteristics. Furthermore, the mean 24-month BMD, its prior, and the precision parameter depend on the covariates so that the amount of information borrowing can vary across the subgroups of subjects with different covariate values. Thus, as study participants in the current trial are recruited, the probability of treatment assignment varies depending on the assessment of the agreement of the current and historical trials.

Control borrowing Bayesian design is related to but clearly different from existing historical borrowing and covariate-adaptive designs. First, CAHB essentially quantifies the information in the historical control arm using the effective sample size, which is similar in spirit as many existing adaptive designs.<sup>2,3,14–18,27</sup> Building on these approaches, we define the covariate-dependent effective control sample size (CECSS) as an estimate of the sample size required to achieve the same level of precision in the control arm if that

sample is a simple random sample with given covariate values. We utilize CECSS in the design to quantify the imbalance of the information in treatment arms such that the amount of historical information borrowed is locally defined and therefore is commensurate to the similarity of the historical data with the current trial data in the neighborhood defined by the covariates. Hence the amount of borrowed information can vary across the covariate domain. Second, we utilize a kernel-biased coin design to adaptively allocate subjects based on the covariate imbalance information, which is similar to the covariate-adaptive design proposed by Jiang et al. (2019).<sup>1</sup> But CAHB considers the additional information from the historical control arm when measuring the covariate imbalance, and hence CAHB tends to allocate more trial participants to experimental treatment arms when the information from the historical trial is sufficient to infer the current mean 24-month BMD outcome in the control arm. We illustrate the CAHB with a two-arm clinical trial, but the design is equally applicable to clinical trials with more than two arms. Also, the method can be easily generalized to consider discrete outcomes by adopting proper likelihoods.

The rest of the article is organized as follows. Section 2 presents the prior, likelihood, and the posterior distributions of the parameters in CAHB model. In Section 3, we introduce the concept of CECSS and develop the CAHB design to achieve the effective sample size balance between two treatment arms. Furthermore, Section 4 illustrates a nonparametric kernel-based procedure to estimate the parameters of interest based on their posterior distributions. Section 5 provides numerical studies and revisits our motivating examples. We conclude and discuss the design in Section 6.

## 2 | PROBABILITY MODEL

We let  $Y_i \in \mathbf{R}$  be the outcome of interest,  $Z_i$  be the treatment indicator with  $Z_i = 1$  and  $Z_i = 0$  indicating assignment to the experimental treatment and the control groups, respectively,  $\mathbf{X}_i \in \mathbb{R}^p$  be a  $p$ -dimensional vector containing the baseline covariates. We let  $\mu_0(\mathbf{X}) = E(Y | Z = 0, \mathbf{X})$  and  $\mu_1(\mathbf{X}) = E(Y | Z = 1, \mathbf{X})$ , and define  $\delta = \mu_1(\mathbf{X}) - \mu_0(\mathbf{X})$  to be the overall treatment effect, which does not depend on the covariates. Let  $\theta_0(\mathbf{X})$  be the conditional mean outcomes of the historical control arm. We assume

$$\mu_0(\mathbf{X}) \sim N\{\theta_0(\mathbf{X}), 1/\tau(\mathbf{X})\}$$

where  $\tau(\mathbf{X})$  is a covariate-dependent precision parameter. A larger value of  $\tau(\mathbf{X})$  implies that the historical and current conditional means of the outcome on the control arm are similar, and hence more information can be borrowed from the historical samples. We assume the prior of  $\tau(\mathbf{X})^{1/2}$  is half normal with scale parameter  $\gamma$ , and the prior density of  $\tau(\mathbf{X})$  is

$$\pi_{\tau}(\tau(\mathbf{X}); \gamma) = \frac{1}{\tau(\mathbf{X})^{1/2} \sqrt{2\pi}\gamma} \exp\left(-\frac{\tau(\mathbf{X})}{2\gamma^2}\right), \tau(\mathbf{X}) > 0.$$

Such prior specification has been widely used to model variance parameters, because it restricts the support of  $\tau(\mathbf{X})$  to be positive and the parameter  $\gamma$  prevents the resulting estimator from diverging to the infinite when  $\mu_0(\mathbf{X})$  and  $\theta_0(\mathbf{X})$  are close.<sup>28</sup> An improper

flat prior that is proportional to 1 is adopted for  $\mu_1(\mathbf{X})$  because we assume there is no prior information about the experimental treatment group.

Furthermore, let  $\mathcal{L}\{\mu_0(\mathbf{X}), \phi_0, Y\}$  and  $\mathcal{L}\{\mu_1(\mathbf{X}), \phi_1, Y\}$  be the likelihood functions for data from the control and experimental arms, respectively, where  $(\phi_0, \phi_1)$  are the auxiliary parameters in the model, such as the scale parameters in the normal likelihood. We assume the densities of  $\phi_0^2$  and  $\phi_1^2$  are  $\pi_\phi(\phi_0^2; a_0, b_0)$  and  $\pi_\phi(\phi_1^2; a_1, b_1)$ , respectively, where  $\pi_\phi(\cdot; a, b)$  is an inverse gamma density with shape and scale parameters  $a, b$ . We choose  $a_0 = b_0 = a_1 = b_1 = 0.01$  which results in noninformative priors that have little influence on the posterior distributions.<sup>2</sup> Denote  $\mathcal{D}_n = \{(Y_i, \mathbf{X}_i, Z_i), i = 1, \dots, n\}$  and we assume  $(Y_i, \mathbf{X}_i, Z_i), i = 1, \dots, n$  are independent. When we define  $\Delta_i(\mathbf{X}) = I(\mathbf{X}_i = \mathbf{X})$  as the indicator function, the conditional posterior distributions of  $\mu_0(\cdot), \tau(\cdot)$  and  $\phi_0$  given the observed  $n$  subjects are

$$\begin{aligned} &\mathcal{P}_{\mu_0}\{\mu_0(\mathbf{X}) \mid \phi_0, \tau(\cdot), \mathcal{D}_n\} \\ &\propto \prod_{i=1}^n \left[ \mathcal{L}\{\mu_0(\mathbf{X}_i), \phi_0, Y_i\} \exp\left(-\frac{\{\mu_0(\mathbf{X}_i) - \theta_0(\mathbf{X}_i)\}^2 \tau(\mathbf{X}_i)}{2}\right) \right]^{(1-Z_i)\Delta_i(\mathbf{X})}, \end{aligned} \tag{1}$$

$$\begin{aligned} &\mathcal{P}_\tau\{\tau(\mathbf{X}) \mid \mu_0(\cdot), \phi_0, \mathcal{D}_n\} \\ &\propto \pi_\tau\{\tau(\mathbf{X}); \gamma\} \prod_{i=1}^n \left[ \left(\frac{\tau(\mathbf{X}_i)}{2\pi}\right)^{1/2} \exp\left(-\frac{\{\mu_0(\mathbf{X}_i) - \theta_0(\mathbf{X}_i)\}^2 \tau(\mathbf{X}_i)}{2}\right) \right]^{(1-Z_i)\Delta_i(\mathbf{X})}, \end{aligned} \tag{2}$$

$$\begin{aligned} &\mathcal{P}_{\phi_0}\{\phi_0^2 \mid \mu_0(\cdot), \tau(\cdot), \mathcal{D}_n\} \\ &\propto \pi_\phi(\phi_0^2; a_0, b_0) \prod_{i=1}^n [\mathcal{L}\{\mu_0(\mathbf{X}_i), \phi_0, Y_i\}]^{(1-Z_i)}. \end{aligned} \tag{3}$$

And the conditional posterior distributions of  $\mu_1(\mathbf{X})$  and  $\phi_1$  are

$$\mathcal{P}_{\mu_1}\{\mu_1(\mathbf{X}) \mid \phi_1, \mathcal{D}_n\} \propto \prod_{i=1}^n \mathcal{L}\{\mu_1(\mathbf{X}_i), \phi_1, Y_i\}^{Z_i \Delta_i(\mathbf{X})}, \tag{4}$$

$$\mathcal{P}_{\phi_1}\{\phi_1^2 \mid \mu_1(\cdot), \mathcal{D}_n\} \propto \pi_\phi(\phi_1^2; a_1, b_1) \prod_{i=1}^n \{\mathcal{L}\{\mu_1(\mathbf{X}_i), \phi_1, Y_i\}\}^{Z_i}. \tag{5}$$

From (1), it is easy to see that the current samples and the historical samples provide the information through the likelihood function and the historical control mean  $\theta_0(\mathbf{X})$ , respectively, to estimate  $\mu_0(\mathbf{X})$ . From (1), (2) and (4), we can see that only the samples with covariate value  $\mathbf{X}$  contribute to the posteriors of  $\mu_0(\mathbf{X})$ ,  $\tau(\mathbf{X})$ ,  $\mu_1(\mathbf{X})$ . On the other hand, (3) and (5) indicate that all samples on the control and experimental treatment arms contribute to the posteriors of  $\phi_0^2$  and  $\phi_1^2$  respectively regardless their covariate values. It is worth mentioning that when  $\mathcal{L}$  is a normal likelihood, the maximizer of (1) is

$$\mu_0(\mathbf{X}) = \frac{\sum_{i=1}^n (1 - Z_i)\Delta_i(\mathbf{X})\{Y_i\phi_0^{-2} + \tau(\mathbf{X}_i)\theta_0(\mathbf{X}_i)\}}{\sum_{i=1}^n (1 - Z_i)\Delta_i(\mathbf{X})\{\phi_0^{-2} + \tau(\mathbf{X}_i)\}}. \tag{6}$$

It is easy to see that when  $\tau(\cdot)$ ,  $\phi_0$ ,  $\theta_0(\cdot)$  are given, the optimization procedure is similar to adding pseudo outcomes  $\theta_0(\mathbf{X}_i)$  into the current trial data weighted by  $(\mathbf{X}_i)$ ,  $i = 1, \dots, n$ . A higher  $\tau(\mathbf{X}_i)$  will give higher weight to these pseudo samples, where the amount of borrowed information from the historical control arm is represented by  $\tau(\mathbf{X}_i)\phi_0^2$ . In addition, the maximizer for (2) is

$$\tau(\mathbf{X}) = \frac{\sum_{i=1}^n (1 - Z_i)\Delta_i(\mathbf{X})}{\sum_{i=1}^n (1 - Z_i)\Delta_i(\mathbf{X})\{\mu_0(\mathbf{X}_i) - \theta_0(\mathbf{X}_i)\}^2 + 1/\gamma^2}.$$

We can see that  $\tau(\mathbf{X})$  is larger when  $\mu_0(\mathbf{X})$  and  $\theta_0(\mathbf{X})$  are closer, and therefore when combining with (6), the historical control samples will provide more information to estimate  $\mu_0(\cdot)$ .

We propose an iterative procedure Algorithm 1 in Section 4 to obtain the posterior estimators for  $\mu_0(\cdot)$ ,  $\tau(\cdot)$ ,  $\phi_0$ ,  $\mu_1(\cdot)$ ,  $\phi_1$ , denoted by  $\hat{\mu}_0(\cdot)$ ,  $\hat{\tau}(\cdot)$ ,  $\hat{\phi}_0$ ,  $\hat{\mu}_1(\cdot)$ ,  $\hat{\phi}_1$ . Here, instead of using a parametric method, we use a *nonparametric* kernel technique to approximate the unknown functions  $\mu_0(\cdot)$ ,  $\tau(\cdot)$ ,  $\mu_1(\cdot)$ , which is more robust to the mis-specifications of their parametric forms. Next, given the above estimators, we illustrate the CAHB design in detail.

### 3 | COVARIATE-ADAPTIVE HISTORICAL BORROWING BAYESIAN DESIGN

The goal of the covariate-adaptive historical borrowing Bayesian design is to assign more subjects to the treatments in order to balance the total information between the control and the experimental arms after accounting for the covariate-adaptively borrowed information on the control arm. To achieve this goal, it is necessary to quantify the CECSS from the historical and current control arms<sup>3</sup> for the given covariates. We define the CECSS by extending the effect sample size defined in Hobbs, Carlin and Sargent (2013).<sup>29</sup> We define

$$R_n(\mathbf{X}) = \widehat{\text{var}}_{\text{ref}}\{\mu_0(\mathbf{X}) \mid \hat{\phi}_0, \mathcal{D}_n\} / \widehat{\text{var}}\{\mu_0(\mathbf{X}) \mid \hat{\phi}_0, \hat{\tau}(\mathcal{D}_n), \mathcal{D}_n\} \tag{7}$$

to be the ratio of the information on the control arm considering and not considering the historical data. *The CECSS given  $\mathbf{X}$  is then defined by  $R_n(\mathbf{X})$  times the size of*

the actual control samples with covariate  $\mathbf{X}$ . A detailed derivations of general forms of  $\widehat{\text{var}}_{\text{ref}}\{\mu_0(\mathbf{X}) \mid \widehat{\phi}_0, \mathcal{D}_n\}$  and  $\widehat{\text{var}}\{\mu_0(\mathbf{X}) \mid \widehat{\phi}_0, \widehat{\tau}(\mathcal{D}_n), \mathcal{D}_n\}$ , and their specific forms when  $\mathcal{L}$  is a normal likelihood are presented in the supplementary material. To measure the covariate imbalance, we define a similarity measure between each of  $n$  previously enrolled subjects in the trial and  $(n + 1)$ th new subject as follows:

$$w_i = \prod_{k=1}^p K_{h_k}(X_{ik} - X_{(n+1)k}), i = 1, \dots, n$$

where  $K_{h_k}(x) = K(x/h_k)/h_k$  is a kernel function satisfies  $K(\cdot) \geq 0$ ,  $K(0) = 1$  and  $h_k > 0$  is a bandwidth. The CAHB design utilizing the CECSS and the similarity measure is outline below. Specifically in Step 5 the similarity measure is used to evaluate the imbalance between the two arms below, and the new subject is assigned to the treatment arm based on the allocation probability that is designed to reduce the imbalance in Step 6.

**Step 1:** Enroll  $n_{\text{initial}}$  number of subjects and equally assign to the control and experimental treatment arms.

**Step 2:** When the  $(n + 1)$  th subject is enrolled and up for treatment allocation, we use Algorithm 1 to obtain the estimators  $\widehat{\mu}_0(\mathbf{X})$ ,  $\widehat{\tau}(\mathcal{D}_n)$ ,  $\widehat{\phi}_0$ ,  $\widehat{\mu}_1(\mathbf{X})$  and  $\widehat{\phi}_1$ .

**Step 3:** Obtain  $R_n(\mathbf{X}_{n+1})$  according to (7) based on the current trial data in  $n$  enrolled subjects.

**Step 4:** Calculate the similarity measure of the new subject with each of the existing  $n$  subjects in trial to obtain  $w_i$ ,  $i = 1, \dots, n$  based on (8).

**Step 5:** For the control arm (arm 0) and the experimental treatment arm (arm 1), define the respective CECSSs as

$$n_0 = R_n(\mathbf{X}_{n+1}) \sum_{i=1}^n w_i(1 - Z_i), n_1 = \sum_{i=1}^n w_i Z_i.$$

Obtain the imbalance measure  $g_u = n_u/(n_0 + n_1)$ ,  $u = 0, 1$ .

**Step 6:** Define the allocation probability  $\pi_u(g_0, g_1) = (g_u^{-1} - 1)/(g_0^{-1} + g_1^{-1} - 2)$ ,<sup>26,30</sup> which is a decreasing function of  $g_u$ . Assign the new subject to arm 0 and arm 1 with probability  $\pi_0(g_0, g_1)$  and  $\pi_1(g_0, g_1)$ , respectively.

**Step 7:** Continue **Step 2 – Step 6** until it reaches the maximum sample size  $N$ .

If  $\Pr[N^{-1} \sum_{i=1}^N \{\mu_1(\mathbf{X}_i) - \mu_0(\mathbf{X}_i)\} > \delta_0 \mid \widehat{\phi}_0, \widehat{\phi}_1, \widehat{\tau}(\mathcal{D}_N), \mathcal{D}_N] > c$ , we reject the null hypothesis  $H_0: \delta = 0$ , where  $\delta_0$  is a cut-off for the indifference region and  $c$  is a pre-specific value to control the type I error rate. The point estimate of the treatment effect is  $\widehat{\delta} = N^{-1} \sum_{i=1}^N \{\widehat{\mu}_1(\mathbf{X}_i) - \widehat{\mu}_0(\mathbf{X}_i)\}$ .

Here  $R_n(\mathbf{X}_{n+1})$  represents the ratio of the information with and without historical borrowing. When no information is borrowed,  $R_n(\mathbf{X}_{n+1}) = 1$  and  $n_0 = \sum_{i=1}^n w_i(1 - Z_i)$  in **Step 5**, which reduces to the information on the control arm at stage  $n$  in Jiang, Ma and Yin (2018).<sup>26</sup> Furthermore, to achieve a balanced allocation ratio, the allocation probability must satisfy (1)  $\pi_u(g_0, g_1)$  is a decreasing function of  $g_u$  and (2)  $\pi_u(g_0, g_1)$  is twice continuously differentiable function of vector  $(g_0, g_1)^\top$  with a uniformly bounded Hessian matrix. Smith (1984)<sup>31</sup> shows that the proposed allocation probability satisfies these two conditions, and hence could yield balanced allocation between the treatment and control arms. In our design, a balanced allocation means the effective sample sizes  $n_1$  and  $n_0$  defined in Step 5 are the same for both arms. Here  $n_1$  and  $n_0$  measure the information in the two arms when considering the borrowed data from historical control arm. It is worth noting that to implement the design, while the patient-level data are required for the current trial, they are not necessary for the historical data. All the historical information needed is an estimation of the conditional mean outcome  $\theta_0(\mathbf{X})$  from the historical control arm. This property simplifies practical implementation of this design.

#### 4 | PARAMETER ESTIMATION

Directly optimizing (1)–(5) to obtain the parameter estimators can be difficult when there are continuous covariates in the data. We assume  $\mu_0(\cdot), \mu_1(\cdot), \tau(\cdot)$  are smooth functions with finite second derivatives, and then adopt a kernel device to estimate them without imposing parametric assumptions. Specifically, we update  $\mu_0(\mathbf{X}), \tau(\mathbf{X})$  and  $\phi_0$  iteratively as follows: when  $l = (1, \dots, L)$  is the iteration index and the superscript  $(l)$  denotes the  $l$ th iteration, we first update  $\mu_0(\mathbf{X})$  at the  $l$ th iteration given  $\tau(\mathbf{X}), \phi_0$  at the  $l - 1$ th iteration as

$$\begin{aligned} \mu_0^{(l)}(\mathbf{X}) = & \operatorname{argmax}_{\mu} \sum_{i=1}^n (1 - Z_i) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{X}) \\ & \times \left[ \log \mathcal{L}(\mu, \phi_0^{(l-1)}, Y_i) - \frac{\{\mu - \theta_0(\mathbf{X}_i)\}^2 \tau^{(l-1)}(\mathbf{X}_i)}{2} \right], \end{aligned} \tag{8}$$

where  $K_{\mathbf{H}}(\mathbf{X}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}\mathbf{x})$  is a multivariate kernel function with bandwidth matrix  $\mathbf{H}$  satisfies  $K(0) = 0$  and  $K(\cdot) > 0$ . When  $\mathbf{X}$  only contains discrete random variables, the kernel function can be reduced to an indicator function. Therefore, the kernel function handles both the continuous and discrete covariates. After obtaining  $\mu_0^{(l)}$ , we update  $\tau(\mathbf{X})$  as

$$\begin{aligned} \tilde{\tau}(\mathbf{X}) = & \operatorname{argmax}_{\tau} \sum_{i=1}^n (1 - Z_i) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{X}) \\ & \times \left[ \frac{1}{2} \log(\tau) - \frac{\{\mu_0^{(l)}(\mathbf{X}_i) - \theta_0(\mathbf{X}_i)\}^2 \tau}{2} \right] + \log\{\pi_{\tau}(\tau; \gamma)\}, \end{aligned} \tag{9}$$

$$\boldsymbol{\tau}^{(l)}(\mathcal{D}_n) = \text{Proj}\{\tilde{\boldsymbol{\tau}}(\mathcal{D}_n), \lambda_1, \lambda_2\}, \quad (10)$$

where  $\tilde{\boldsymbol{\tau}}(\mathcal{D}_n) = \{\tilde{\boldsymbol{\tau}}(\mathbf{X}_i), i = 1, \dots, n\}^\top$ ,  $\boldsymbol{\tau}^{(l)}(\mathcal{D}_n) = \{\tau^{(l)}(\mathbf{X}_i), i = 1, \dots, n\}^\top$ . The function  $\text{Proj}(\tilde{\boldsymbol{\tau}}^{(l)}(\mathcal{D}_n), \lambda_1, \lambda_2)$  projects  $[\tilde{\boldsymbol{\tau}}(\mathbf{X}_i)I\{\tilde{\boldsymbol{\tau}}(\mathbf{X}_i) > \lambda_1\}, i = 1, \dots, n]^\top$  onto the space  $\mathbb{B}_1(\lambda_2)$ , a simplex in  $\mathbb{R}^n$ , which can be implemented by the linear programming method proposed by Duchi et al. (2008).<sup>32</sup> When  $\mathcal{L}$  is a normal likelihood, the solution for (9) is

$$\tilde{\boldsymbol{\tau}}(\mathbf{X}) = \frac{\sum_{i=1}^n (1 - Z_i)K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{X})}{\sum_{i=1}^n (1 - Z_i)K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{X})\{\mu_0^{(l)}(\mathbf{X}_i) - \theta_0(\mathbf{X}_i)\}^2 + 1/\gamma^2}.$$

Here, the parameter  $\gamma$  prevents  $\tilde{\boldsymbol{\tau}}(\mathbf{X})$  from exploding. In addition, the  $\text{Proj}(\cdot)$  function induces sparseness in  $\boldsymbol{\tau}^{(l)}(\mathcal{D}_n)$ , a parameter vector with growing dimension, and forces  $\tau^{(l)}(\mathbf{X})$  to zero if the distance between  $\mu_0^{(l)}(\mathbf{X})$  and  $\theta_0(\mathbf{X})$  is large. This is a desirable property as it prevents the information borrowing if the historical and current control samples are not from the same population. Finally, we update  $\phi_0$  as

$$\phi_0^{(l)} = \text{argmax}_{\phi_0} \sum_{i=1}^n (1 - Z_i) [\log \mathcal{L}\{\mu_0^{(l)}(\cdot), \phi_0, Y_i, \mathbf{X}_i\}] + \log\{\pi_{\phi}(\phi_0^2; a_0, b_0)\}. \quad (11)$$

We solve for  $\mu_0(\mathbf{X})$ ,  $\boldsymbol{\tau}(\mathbf{X})$ ,  $\phi_0$  iteratively until the algorithm converges.

Similarly, following (4) and (5), we obtain the estimators for  $\mu_1(\mathbf{X})$  and  $\phi_1$  as follows:

$$\hat{\mu}_1(\mathbf{X}) = \text{argmax}_{\mu} \sum_{i=1}^n [Z_i K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{X}) \log \mathcal{L}(\mu, \phi_1, Y_i)] \quad (12)$$

$$\hat{\phi}_1 = \text{argmax}_{\phi_1} \sum_{i=1}^n Z_i [\log \mathcal{L}\{\hat{\mu}_1(\cdot), \phi_1, Y_i, \mathbf{X}_i\}] + \log\{\pi_{\phi}(\phi_1^2; a_1, b_1)\}. \quad (13)$$

We summarize the estimation procedure in Algorithm 1.

The selections of  $\lambda_1$ ,  $\lambda_2$  and  $\gamma$  are crucial to determine the amount of information borrowing from the historical control. We select  $\lambda_2 = 300 \log(n)$  and  $\gamma = \sqrt{3}$ , which are chosen to minimize the average estimation errors of the treatment effects in the simulation. We select  $\lambda_1$  to be the 10% quantile of the estimated precision parameters when assuming the historical

and current control data are consistent. We illustrate the selection of  $\lambda_1$  in Algorithm 2 and summarize how to select the tuning parameters in Table 1.

The tuning parameter  $\lambda_1$  is used to eliminate the spurious estimators of  $(\mathbf{X}_i)$ ,  $i = 1, \dots, n$ . Because the posterior distribution of  $\tau(\mathbf{X}_i)$  contains the term  $\log\{\tau(\mathbf{X}_i)\}$ , directly maximizing the posterior distribution is impossible to obtain a correct estimator of  $\tau(\mathbf{X}_i)$  at  $\tau(\mathbf{X}_i) = 0$ . Instead, the algorithm will provide very small estimators of  $\tau(\mathbf{X}_i)$  at these points. Therefore, to eliminate the bias, we remove the  $\tau(\mathbf{X}_i)$  estimates whose values are less than  $\lambda_1$ . This selection rule in Algorithm 2 ensures that when the historical and control arm have the same distribution, the majority of the information (over 90%) is retained. It's worth noting that the 10% quantile can be adjusted according to specific requirements; other values such as 5% or 20% could also be used. This selection of  $\lambda_1$  leads to satisfactory type I and II errors, and yields the smallest estimation error of the treatment effect in our simulation studies.

### ALGORITHM 1.

Estimation of parameters

---

**Input:** The maximal number of iterations  $L$ , the observed dataset  $\{(Y_i, \mathbf{X}_i, Z_i)\}_{i=1}^n$ , the historical mean model for the control arm  $\theta_0(\mathbf{X})$ .

- 1 **for**  $l = 1, \dots, L$  **do**
- 2     Update  $\{\mu_0^{(l)}(\mathbf{X}), \tau^{(l)}(\mathcal{D}_n), \phi_0^{(l)}\}$  using (8)–(11), respectively.
- 3     Define  

$$\max \left[ \sum_{i=1}^n (\mu_0^{(l)}(\mathbf{X}_i) - \mu_0^{(l-1)}(\mathbf{X}_i))^2 / n, \sum_{i=1}^n (\tau^{(l)}(\mathbf{X}_i) - \tau^{(l-1)}(\mathbf{X}_i))^2 / n, \{(\phi_0^{(l)})^2 - (\phi_0^{(l-1)})^2\}^2 \right]$$
 as iteration error.
- 4     **if** iteration error  $< 10^{-5}$  **then**
- 5         **break**
- 6     **end if**
- 7   **end for**
- 8
- 9    $\hat{\mu}_0(\mathbf{X}) = \mu_0^{(l)}(\mathbf{X}), \hat{\tau}(\mathcal{D}_n) = \tau^{(l)}(\mathcal{D}_n), \hat{\phi}_0 = \phi_0^{(l)}$
- 10   Estimate  $\{\hat{\mu}_1(\mathbf{X}), \hat{\phi}_1\}$  with (12) and (13), respectively.

**Output:** The estimated parameters  $\{\hat{\mu}_0(\mathbf{X}), \hat{\tau}(\mathcal{D}_n), \hat{\phi}_0, \hat{\mu}_1(\mathbf{X}), \hat{\phi}_1\}$ .

---

### ALGORITHM 2.

Tuning the parameter  $\lambda_1$

---

**Input:** The observed data in the current study  $\{(Y_i, \mathbf{X}_i, Z_i)\}_{i=1}^n$ .

- 1: Estimate the conditional mean outcome of the current control samples without considering the historical data.
- 2: Treat the estimated conditional mean as the historical mean model  $\theta_0(\mathbf{X})$  and obtain  $\hat{\tau}(\mathcal{D}_n) = \{\hat{\tau}(\mathbf{X}_i), i = 1, \dots, n\}^\top$  using Algorithm 1 by setting  $\lambda_1 = 0$ .
- 3: Select  $\lambda_1$  as the 10% quantile of the  $\hat{\tau}(\mathbf{X}_i)$ 's.

Output:  $\lambda_i$ .

## 5 | NUMERICAL STUDY

### 5.1 | Comparison with a study level historical borrowing design.

We compare our method with a study level historical borrowing design (SLHBD) that ignores the covariate information. In the SLHBD, we assume  $\mu_0, \mu_1, \tau$  are independent of  $\mathbf{X}$  and have the same prior specifications as those used in our model. We obtain their posterior distributions and  $R_n$  as the ratio of the posterior variances of  $\mu_0$  with a flat prior and with our proposed prior. We also define the effective control sample size as  $R_n$  times the size of the actual control samples. Without considering the covariates, SLHBD utilizes the classical biased coin design<sup>21</sup> to balance the effective control sample size and the sample size on the treatment arm. We consider the scenarios where the relationship between historical and current samples varies across different subgroups. More specifically, we generate the outcome in the current study from the model

$$Y = \delta Z + \mu_0(\mathbf{X}) + \epsilon,$$

where  $\mathbf{X} = (1, X)^\top$ ,  $\mu_0(\mathbf{X}) \equiv \beta_0 + \beta_1 X$ ,  $X$  is a binary covariate with equal probability to be 0 or 1,  $\epsilon$  is the mean zero Gaussian noise with variance 0.25,  $\delta = 0.3$  and  $\beta_0 = \beta_1 = 1$ . We assume the mean of the outcome in the historical sample follows  $\theta_0(\mathbf{X}) = \alpha_0 + \alpha_1 X$ .

We consider two choices of  $(\alpha_0, \alpha_1)$ :

- Scenario 1:  $(\alpha_0, \alpha_1) = (1, 10)$ . In this setting,  $\mu_0(\mathbf{X}) = \theta_0(\mathbf{X})$  only when  $X = 0$ , but  $E\{\mu_0(\mathbf{X})\} \neq E\{\theta_0(\mathbf{X})\}$ .
- Scenario 2:  $(\alpha_0, \alpha_1) = (6, -9)$ . In this setting,  $\mu_0(\mathbf{X}) \neq \theta_0(\mathbf{X})$ , but  $E\{\mu_0(\mathbf{X})\} = E\{\theta_0(\mathbf{X})\} = 1.5$ .

We simulate historical and current samples 2000 times with sample size  $N = 100$ , and present the resulting allocation ratios to the treatment arm, absolute errors between the estimated and true treatment effects, and statistical powers under the alternative hypothesis that  $\delta = 0.3$  in Table 2. Note that a design that borrows more information from the historical samples on the control arm would have higher allocation ratio onto the experimental treatment arm, because less information on the control arm is needed from the current study to achieve the same information level as that in the experimental treatment arm. On the other hand, if a design does not borrow historical information, the subjects in the current study will approximately equally allocated in the two arms. In all settings, we adjust  $\delta_0$  and  $c$  in **Step 7** to control the type I errors at 0.05. All the other tuning parameters are selected based on Table 1. The results show that when ignoring the covariates, SLHBD does not borrow the historical information in Scenario 1 with the allocation ratio approximately 0.5 in both subgroups, because the outcomes in historical and current control samples have different marginal means. CAHB borrows the information from the historical observations with

$X = 0$ , which yields higher allocation ratio in the subgroup with  $X = 0$ . In Scenario 2, by only considering the marginal similarity between the historical and current control samples, SLHBD mistakenly borrows the historical information, which results in more than 80% subjects being allocated to the experimental treatment arm even if there are large differences between the historical and current control samples. On the other hand, CAHB correctly identifies the difference between historical and control samples in the subgroups and refuses to borrow any information from the historical data with allocation ratio approximately 0.5 across the subgroups. Furthermore, with the correct information borrowing mechanism, CAHB enjoys higher estimation accuracy and higher statistical power in both scenarios.

## 5.2 | Simulations based on a real data example

The motivating HORIZON and FIT trials are used to inform simulation studies to assess the benefit of the proposed CAHB method. The FIT and the HORIZON are similarly conducted placebo-controlled randomized trials that studied the effects of bisphosphonate treatments on vertebral fracture risk in post-menopausal women. The clinical population of interest is elderly women aged approximately 80 years or above who are at elevated vertebral fracture risk compared to younger women and thus suffer disproportionately with respect to disability and medical costs resulting from fracture events.<sup>4,7</sup> We restrict the evaluation to consider only women with white race/ethnicity since the FIT study lacks diverse recruitment. The HORIZON enrolls 620 of such women and borrows the placebo control arm information in the FIT study permitted utilizing the control arm information collected in its sample corresponding to 62.6% of the HORIZON sample.

Because the primary outcome, vertebral fracture by 36 months, is adjudicated slightly differently between the two trials,<sup>4-6</sup> we instead use the hip BMD measured at 24 months by dual-energy x-ray absorptiometry as our common study outcome for the two trials, an established objective surrogate endpoint for vertebral fracture to determine treatment efficacy.<sup>8</sup> Using a more readily observable surrogate has been well accepted in the adaptive trial design literature.<sup>34-36</sup> Four covariates associated with vertebral fracture risk<sup>37-40</sup> are used to match subjects from the historical FIT study and the current HORIZON study: fall history ( $X_1$ ,  $X_1 = 1$  if there is any fall history, otherwise 0), vertebral fracture history ( $X_2$ ,  $X_2 = 1$  if there is any fracture history, otherwise 0), the years since menopause ( $X_3$ ) and the baseline hip BMD ( $X_4$ ). The two binary covariates divide the data to four subgroups. The distributions of the BMD at 24 months in the FIT and the HORIZON control arm are similar in subgroups 2 and 3 and rather discrepant in subgroups 1 and 4 (Figure 1). The primary outcome distributions are also similar in subgroups 2 and 3, but rather different in subgroups 1 and 4. We utilize this real data setting and examine the performance of the CAHB design for adaptively assigning the subjects based on the covariate information so that more subjects can be assigned to the experimental treatment arm when sufficient control information is provided from both current and historical trials.

More specifically, we let  $\mu_0(\mathbf{X}) = \beta_0(X_1, X_2) + \beta_1(X_1, X_2)X_3 + \beta_2(X_1, X_2)X_4$ , and simulate the current trial samples from the model

$$Y = \delta Z + \mu_0(\mathbf{X}) + \epsilon,$$

where  $X_1, X_2$  are two binary covariates with probability 31.0%, 71.9% to be one, respectively. The values of  $X_1, X_2$  divide samples to four subgroups which comprise (19.8%, 8.2%, 49.2%, and 22.7%) of the total sample size, respectively. The covariates  $X_3$  and  $X_4$  are simulated from estimated distributions of the standardized years since menopause and the standardized baseline hip BMD in the HORIZON data, and  $\epsilon$  is a mean zero Gaussian random error with variance 0.11. Here  $\delta = 0.349$  is the overall treatment effect, and  $\beta_0(X_1, X_2), \beta_1(X_1, X_2), \beta_2(X_1, X_2)$  are regression coefficients, which vary across the subgroups. We summarize the true values of  $\beta_0(X_1, X_2)$ , and  $\beta_1(X_1, X_2), \beta_2(X_1, X_2)$  and the subgroup specifications in Table 3. The values of  $\phi_0^2, \delta$  and  $\beta_0, \beta_1, \beta_2$  are all estimated from the HORIZON dataset.

Furthermore, we let  $\theta_0 = \alpha_0(X_1, X_2) + \alpha_1(X_1, X_2)X_3 + \alpha_2(X_1, X_2)X_4$  and assume the historical samples are generated from the model,

$$Y = \delta_h Z + \theta_0(\mathbf{X}) + \epsilon_h, \quad (14)$$

where  $\alpha_0(X_1, X_2), \alpha_1(X_1, X_2), \alpha_2(X_1, X_2)$  are subgroup-specific parameters,  $\epsilon_h$  is a mean zero random error. We vary the values of  $\alpha_0, \alpha_1, \alpha_2$  to represent different degrees of consistency between historical and current control data.

To start the trial, we first equally randomize 20 subjects to the experimental treatment and control arms. When implementing CAHB, we specify  $\mathcal{L}$  to be a normal likelihood,  $K_{\mathbf{H}}$  to be a multivariable gaussian kernel with  $\mathbf{H} = \text{diag}(0.1, 0.1, w_1, w_2)$ , where  $w_1, w_2$  are the rule-of-thumb bandwidths discussed in Scott (2015).<sup>33</sup> The bandwidth 0.1 is selected for the discrete covariates so that the kernel function works like an indicator function in the first two dimensions. Following Jiang, Ma and Yin,<sup>26</sup>  $K_{h_n}$  is Epanechnikov kernel, with bandwidths  $h_1 = h_2 = 1.1$  for the binary covariate and,  $h_3 = h_4 = 1.3$  for the continuous covariates. In all simulations, we set  $\delta_0 = 0$  and adjust  $c$  in **Step 7** to achieve 0.05 type I error rate.

**5.2.1 | Simulation studies under varying discrepancy between the historical and current control.**—We first investigate the performance of CAHB under varying amounts of discrepancy between historical and current control data when the relationships between  $\beta_k(X_1, X_2)$  and  $\alpha_k(X_1, X_2)$ ,  $k = 0, 1, 2$  are the same across the subgroups. We specify  $\alpha_k(X_1, X_2) = \beta_k(X_1, X_2)d$ ,  $k = 0, 1, 2$ , where  $d$  is generated from a normal distribution with mean  $(1 + \xi)$  and variance  $0.3I(\xi \neq 0)$ . It can be seen that when  $\xi = 0$ ,  $d = 1$  so that  $\alpha_k(X_1, X_2) = \beta_k(X_1, X_2)$ , and the difference between  $\alpha_k(X_1, X_2)$  and  $\beta_k(X_1, X_2)$  increases with the magnitude of  $\xi$ . Here, we assume  $d$  is random, because  $\theta_0(\mathbf{X})$  is often estimated from the historical sample, which is a random variable in practice. We vary  $\xi$  from 0 to 1 and generate 2000 datasets with the maximal sample size  $N = 200$  under different specifications of  $\xi$ . We compare CAHB with the kernel-based biased coin design (KBCD)<sup>26</sup> for (a) the mean estimation error  $|\hat{\delta} - \delta|$  over the simulations (b) the power of a hypothesis test with the null that  $\delta = 0$  when the true effect is 0.349 at 5% significance level and (c) the mean percentages of subjects assigned to the experimental treatment arm.

Compared with KBCD method, the CAHB method yields smaller mean  $|\hat{\delta} - \delta|$  when the historical and current control are similar ( $\xi \leq 0.2$ ) and slightly larger mean  $|\hat{\delta} - \delta|$  when  $\xi > 0.25$  (Figure 2 a). Whereas both CAHB and KBCD achieve 85% of power, CAHB has higher power when the distributions of the historical and current control are consistent ( $\xi \leq 0.1$ ) (Figure 2b). Finally, the allocation ratio to the experimental arm is always higher for the CAHB. The allocation ratio decreases as the more discrepant the historical and the current trial control arm data become, that is, the distance between  $\alpha_k(X_1, X_2)$  and  $\beta_k(X_1, X_2)$  increases (Figure 2c). Across the different subgroups, the allocation ratio changes by the subgroup sample size: the allocation ratio to the experimental treatment is higher in the subgroups with larger sample sizes. In conclusion, these results suggest that CAHB yields comparable mean  $|\hat{\delta} - \delta|$  and power as the KBCD does, while more subjects have been allocated to the experimental arms by using CAHB than that by using KBCD. Furthermore, the allocation ratio is larger when the historical and current control samples are similar. Moreover, by using CAHB, more subjects are assigned to the experimental arm in the subgroups with larger samples on the population.

We further investigate the operating characteristics of CAHB when the relationship between  $\alpha_k(X_1, X_2)$  and  $\beta_k(X_1, X_2)$  varies across the subgroups. More specifically, we assume  $\alpha_k(X_1, X_2) = \beta_k(X_1, X_2)$  in Subgroups 1 and 4, and  $\alpha_k(X_1, X_2) = \beta_k(X_1, X_2)d$ ,  $k = 0, 1, 2$ , in Subgroups 2 and 3. Figure 3a and b show the similar overall properties as those observed in Figure 2 that CAHB yields better  $|\hat{\delta} - \delta|$  and power on average than the KBCD does, while more subjects have been allocated to the experimental arms by using CAHB than that by using KBCD. In addition, Figure 3 (c) shows that the allocation ratio in Subgroups 1 and 4 are consistent to be around 0.58, and the allocation ratios in Subgroup 2 and 3 decrease dramatically when the discrepancy between historical and current control samples increases ( $\xi$  from 0 to 1). Furthermore, when the discrepancy between historical and current control samples is considerably large ( $\xi = 1$ ), the allocation ratio reduces to 0.5.

### 5.2.2 | Simulation studies under varying current and historical sample sizes.

—We use the data from FIT trial to estimate the parameters  $\alpha_k(X_1, X_2)$ , and  $k = 0, 1, 2$ . The conditional mean outcomes from FIT and HORIZON are similar in certain subgroups as shown in Figure 1 and we expect that CAHB will have more subjects allocated to experimental treatment arm when considering the historical data.

We first assess the performance of CAHB when the sample size of the current data varies. We simulate 2000 datasets for the settings when current sample size  $N$  varies from 40 to 400. To capture the variability of the historical model, in each simulation, we sample  $N_0 = 350$  observations without replacement from the historical datasets to estimate  $\theta_0(\mathbf{X})$  based on (14).

We summarize the mean  $|\hat{\delta} - \delta|$ , the power, and the allocation ratio to the experimental arm for CAHB and KBCD in Figure 4. Figure 4 shows that CAHB yields consistently better  $|\hat{\delta} - \delta|$  and power on average than the KBCD does, while the allocation ratio from CAHB is always greater than 0.5. This suggests that CAHB allocates more subjects to the experimental treatment arm without sacrificing estimation accuracy and power. It is

worth mentioning that the allocation ratio starts to decrease when  $N$  reaches 350. This phenomenon indicates that when there are sufficient samples on both arms, CAHB tends to take less information from historical studies.

Lastly, we study the operating characteristics of CAHB when the historical sample size  $N_0$  varies. We vary the  $N_0$  from 30 to 350 and simulate 2000 datasets under each setting. In each simulation, we sample  $N_0$  observations without replacement to estimate  $\theta_0(\mathbf{X})$  based on (14), and fix the current maximal sample size at  $N = 200$ .

We summarize the mean  $|\hat{\delta} - \delta|$ , the power, and the allocation ratio to the experimental arm for CAHB and KBCD in Figure 5. Figure 5 shows that the mean  $|\hat{\delta} - \delta|$ , power and allocation ratio improve when the historical sample size  $N_0$  increases. This is because when the historical control sample size increases, the estimation variation of  $\theta_0(\mathbf{X})$  decreases. Hence the estimator becomes closer to the true values, and in turn is also closer to the  $\mu_0(\mathbf{X})$  in the current data. In summary, when the historical and current control samples are consistent, larger historical control samples yield smaller mean  $|\hat{\delta} - \delta|$ , higher power, and better allocation ratio to the experimental treatment arm.

## 6 | CONCLUSION

We proposed a CAHB design, which incorporates the historical information and adaptively allocates subjects to balance the conditional effect sample size. CAHB automatically adjusts the amount of information borrowed by covariate-adaptively evaluating the similarity between the distributions of the current and historical control samples. Compared with the study level historical borrowing design, when the agreement between the historical and current studies in the subgroups are different from those at the study level, CAHB is more likely to make the correct treatment assignment decision, which yields higher power and more accurate effect size estimation. When the historical and current control samples are not from the same population, CAHB yields the similar estimation accuracy and statistical power and achieves covariate balancing as the kernel-based biased coin design<sup>26</sup> does, which does not utilize the historical information. Importantly, compared with the kernel-based biased coin design, CAHB has better estimation accuracy, statistical power, and assigns more subjects to the experimental treatment arm when the distributions of the historical and current control samples are consistent.

This new design has several practical implications. First, it can help to reduce distributional imbalance of the current trial subject characteristics across the treatment arms. This is important because covariate imbalance can threaten the internal validity of the current trial by introducing confounding factors that affect study endpoints. Second, the proposed design can be used to borrow historical control data from trials with different distributions of subject characteristics. This can be useful in settings where there is limited historical control data available, or where the available historical control data are not well-matched to the current trial. Overall, the proposed design is a promising new approach to historical control borrowing in clinical trials. It has the potential to reduce distributional imbalance, increase the efficiency of trials, and improve the accuracy of treatment effect estimates.

Our primary outcome hip BMD at 24 months is continuous, however, CAHB can be extended to handle discrete outcomes, where  $\mu_b(\mathbf{X})$  can be  $\exp(\mathbf{b}^T \mathbf{X}) / \{1 + \exp(\mathbf{b}^T \mathbf{X})\}$  for the logistic model and can be  $\exp(\mathbf{b}^T \mathbf{X})$  in the Poisson model with  $\mathbf{b}$  to be the unknown parameter. Algorithm 1 and the covariate-adaptive design discussed in Section 3 are directly applicable for the parameter estimation and the adaptive allocation of the subjects.

Although our prior selection is based on the Gaussian distribution, our likelihood function is designed to be flexible to accommodate various data distributions. By using a Gaussian prior, we implicitly used weighted prior mean that contains historical information as pseudo samples as shown in (6) to improve the estimation efficiency of the parameter of interest in the current study. Additionally, when the likelihood is Gaussian, selecting a Gaussian prior leads to a closed-form solution of the posterior distribution of the parameters of interest. This facilitates us to analyze how each model parameter affects the information borrowing from the historical study.

Finally, an important limitation of our design is its inability to consider the variance of the estimated mean outcome from the historical control arm during information borrowing. Consequently, the amount of information borrowed from the historical study remains unchanged as long as the estimated mean remains constant, even if the historical sample size increases. To address this limitation, it is necessary to clearly define how the variability of historical estimators impacts the accurate borrowing of information. This aspect presents an avenue for the future research and is worth exploring in further studies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thankfully acknowledge the permission of the Fracture Intervention Trial (FIT) and HORIZON PFT Trial Steering Committees (Dennis Black, chair) to utilize individual patient data from these trials to provide correlations that were used in our simulations. We would like to express our gratitude to the Associate Editor and two reviewers for their insightful comments, which have greatly enhanced the quality of our article.

### Funding information

NIH, Grant/Award Number: K25AG071840; UCSF Resource Allocation Program Funding (FJ); NSF, Grant/Award Number: DMS-2210206

## DATA AVAILABILITY STATEMENT

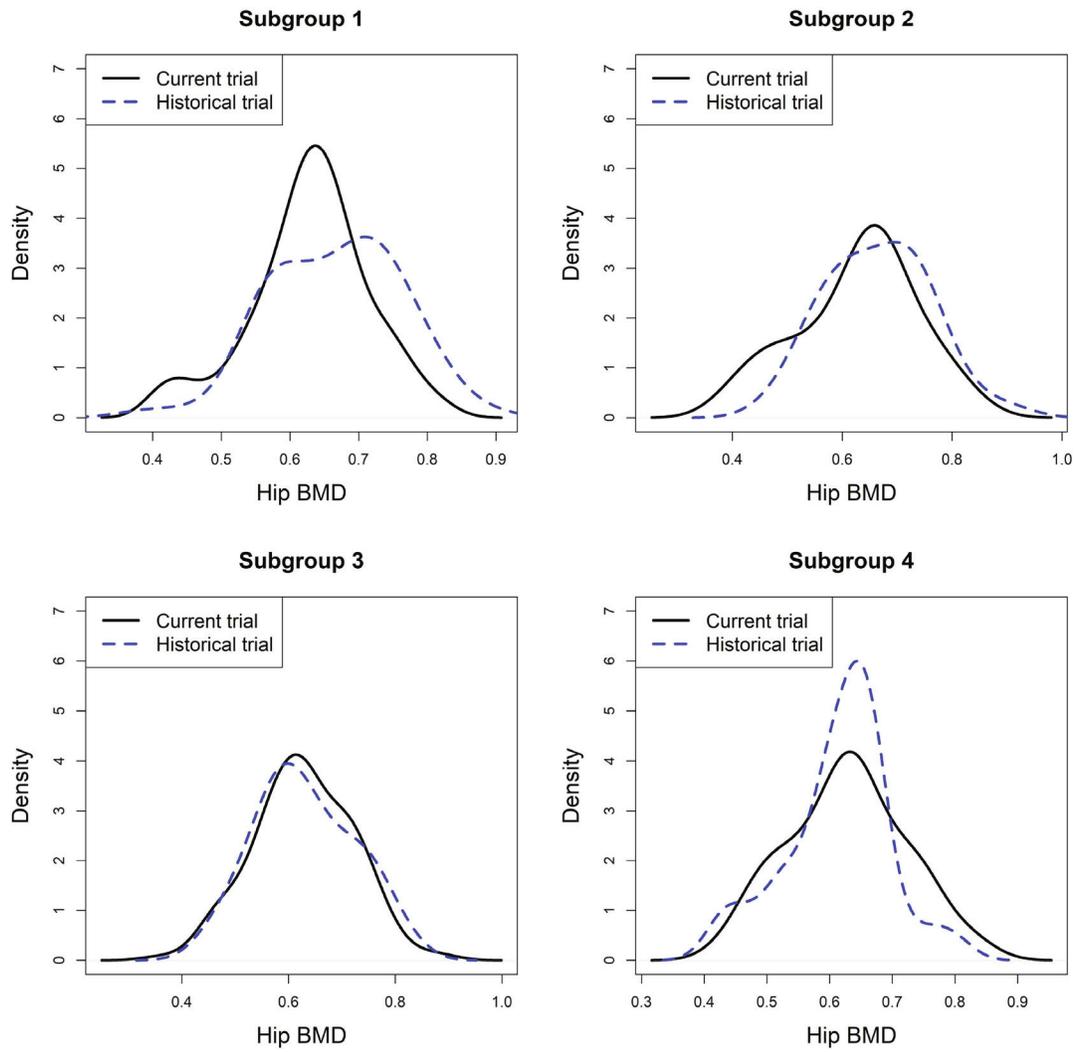
Research data are not shared.

## REFERENCES

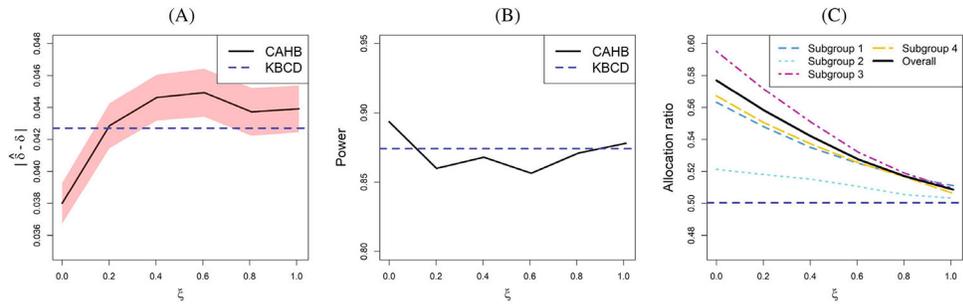
1. Jiang F, Tian L, Fu H, Hasegawa T, Wei L. Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study. *J Amer Statist Assoc.* 2019;114(528):1854–1864.
2. Jin H, Yin G. Unit information prior for adaptive information borrowing from multiple historical datasets. *Stat Med.* 2021;40(25):5657–5672. [PubMed: 34302378]

3. Kim MO, Harun N, Liu C, Khoury JC, Broderick JP. Bayesian selective response-adaptive design using the historical control. *Stat Med*. 2018;37(26):3709–3722. [PubMed: 29900577]
4. Black DM, Delmas PD, Eastell R, et al. Once-yearly zoledronic acid for treatment of postmenopausal osteoporosis. *New Engl J Med*. 2007;356(18):1809–1822. [PubMed: 17476007]
5. Black DM, Cummings SR, Karpf DB, et al. Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. *Lancet*. 1996;348(9041):1535–1541. [PubMed: 8950879]
6. Cummings SR, Black DM, Thompson DE, et al. Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the fracture intervention trial. *J Am Med Assoc*. 1998;280(24):2077–2082.
7. Chen P, Kregge JH, Adachi JD, et al. Vertebral fracture status and the World Health Organization risk factors for predicting osteoporotic fracture risk. *J Bone Mineral Res*. 2009;24(3):495–502.
8. Black DM, Bauer DC, Vittinghoff E, et al. Treatment-related changes in bone mineral density as a surrogate biomarker for fracture risk reduction: meta-regression analyses of individual patient data from multiple randomised controlled trials. *Lancet Diabetes Endocrinol*. 2020;8(8):672–682. [PubMed: 32707115]
9. Pocock SJ. The combination of randomized and historical controls in clinical trials. *J Chron Dis*. 1976;29(3):175–188. [PubMed: 770493]
10. Dempster AP, Selwyn MR, Weeks BJ. Combining historical and randomized controls for assessing trends in proportions. *J Amer Statist Assoc*. 1983;78(382):221–227.
11. Ryan L. Using historical controls in the analysis of developmental toxicity data. *Biometrics*. 1993;49(4):1126–1135. [PubMed: 8117906]
12. Hong H, Fu H, Carlin BP. Power and commensurate priors for synthesizing aggregate and individual patient level data in network meta-analysis. *J R Statist Soc: Ser C (Appl Stat)*. 2018;67(4):1047–1069.
13. Wang Z, Lin L, Murray T, Hodges JS, Chu H. Bridging randomized controlled trials and single-arm trials using commensurate priors in arm-based network meta-analysis. *Ann Appl Stat*. 2021;15(4):1767–1787. [PubMed: 36032933]
14. Ibrahim JG, Chen MH. Power prior distributions for regression models. *Statist Sci*. 2000;15(1):46–60.
15. Duan Y, Ye K, Smith EP. Evaluating water quality using power priors to incorporate historical information. *Environmetr: Off J Int Environmetr Soc*. 2006;17(1):95–106.
16. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*. 2011;67(3):1047–1056. [PubMed: 21361892]
17. Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clin Trials*. 2010;7(1):5–18. [PubMed: 20156954]
18. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*. 2014;70(4):1023–1032. [PubMed: 25355546]
19. Lin Y, Zhu M, Su Z. The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials. *Contemp Clin Trials*. 2015;45:21–25. [PubMed: 26244705]
20. Antognini AB, Giovagnoli A. A new 'biased coin design' for the sequential allocation of two treatments. *J R Statist Soc: Ser C (Appl Stat)*. 2004;53(4):651–664.
21. Wei LJ. The adaptive biased coin design for sequential experiments. *Ann Stat*. 1978;6(1):92–100.
22. Efron B. Forcing a sequential experiment to be balanced. *Biometrika*. 1971;58(3):403–417.
23. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. *Clin Pharmacol Therapeut*. 1974;15(5):443–453.
24. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975;31(1):103–115. [PubMed: 1100130]
25. Hu Y, Hu F. Balancing treatment allocation over continuous covariates: a new imbalance measure for minimization. *J Probab Stat*. 2012;40(3):2012.

26. Jiang F, Ma Y, Yin G. Kernel-based adaptive randomization toward balance in continuous and discrete covariates. *Statistica Sinica*. 2018;28(4):2841–2856.
27. Murray TA, Thall PF, Schortgen F, Asfar P, Zohar S, Katsahian S. Robust Adaptive Incorporation of Historical Control Data in a Randomized Trial of External Cooling to Treat Septic Shock. *Bay Anal*. 2021;16(3):825–844.
28. Gelman A Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayes Anal*. 2006;1(3):515–534.
29. Hobbs BP, Carlin BP, Sargent DJ. Adaptive adjustment of the randomization ratio using historical control data. *Clin Trials*. 2013;10(3):430–440. [PubMed: 23690095]
30. Atkinson AC. Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*. 1982;69(1):61–67.
31. Smith RL. Properties of biased coin designs in sequential clinical trials. *Ann Stat*. 1984;12(3):1018–1034.
32. Duchi J, Shalev-Shwartz S, Singer Y, Chandra T. Efficient Projections onto the L1-Ball for Learning in High Dimensions. In *Proceedings of the 25th international conference on Machine learning*. 2008:272–279.
33. Scott DW. Multivariate density estimation and visualization. In: Gentle J, Härdle W, Mori Y, eds. *Handbook of Computational Statistics*. Springer; 2012.
34. Kim MO, Liu C, Hu F, Lee JJ. Outcome-adaptive randomization for a delayed outcome with a short-term predictor: imputation-based designs. *Stat Med*. 2014;33(23):4029–4042. [PubMed: 24889540]
35. Hu F, Zhang LX, Cheung SH, Chan WS. Doubly adaptive biased coin designs with delayed responses. *Can J Stat*. 2008;36(4):541–559.
36. Huang X, Ning J, Li Y, Estey E, Issa JP, Berry DA. Using short-term response information to facilitate adaptive randomization for survival clinical trials. *Stat Med*. 2009;28(12):1680–1689. [PubMed: 19326367]
37. Shieh A, Ruppert KM, Greendale GA, et al. Associations of Age at Menopause With Postmenopausal Bone Mineral Density and Fracture Risk in Women. *J Clin Endocrinol Metabol*. 2022;107(2):e561–e569.
38. Cummings SR, Nevitt MC, Browner WS, et al. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. *N Engl J Med*. 1995;332(12):767–773. [PubMed: 7862179]
39. Black DM, Cauley JA, Wagman R, et al. The Ability of a Single BMD and Fracture History Assessment to Predict Fracture Over 25 Years in Postmenopausal Women: The Study of Osteoporotic Fractures. *J Bone Miner Res*. 2018;33(3):389–395. [PubMed: 28719727]
40. Black DM, Steinbuch M, Palermo L, et al. An assessment tool for predicting fracture risk in postmenopausal women. *Osteoporos Int*. 2001;12(7):519–528. [PubMed: 11527048]

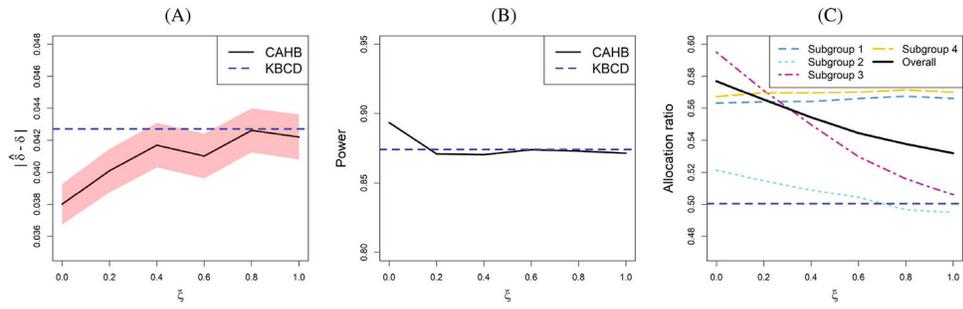


**FIGURE 1.** The distributions of the BMD at 24 months under different subgroups from the control arms of the current and historical trials.



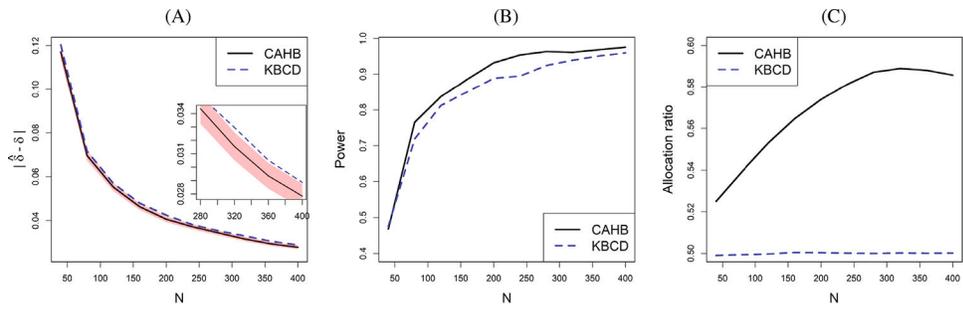
**FIGURE 2.**

The results from 2000 simulations. The absolute error (a), the power when the type I error is 0.05 (b), and the percentage of the subjects assigned to experimental treatment arm (c) when varying  $\xi$  from 0 to 1. The shadow indicates the 95% confidence intervals of the absolute error. A larger  $\xi$  yields a larger difference of the outcome-covariates association between the historical and current control samples.



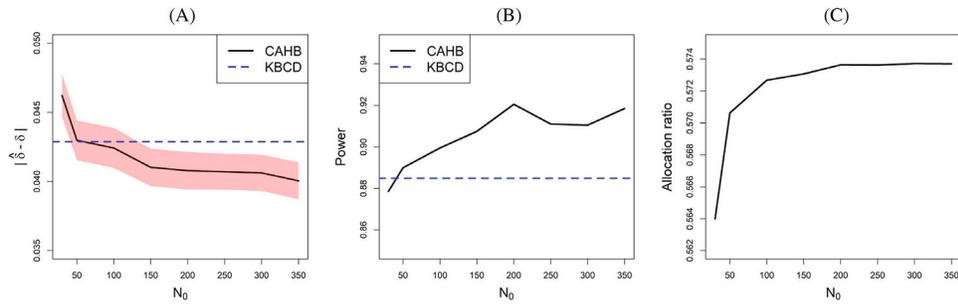
**FIGURE 3.**

The results from 2000 simulations. The absolute error (a), the power when the type I error is 0.05 (b), and the percentage of the subjects assigned to experimental treatment arm (c) when varying  $\xi$  from 0 to 1 in subgroups 2 and 3. The shadow indicates the 95% confidence intervals of the absolute error. A larger  $\xi$  yields a larger difference of the outcome-covariates association between the historical and current control samples.



**FIGURE 4.**

The results from 2000 simulations. The absolute error (a), the power when the type I error is 0.05 (b), and the percentage of the subjects assigned to experimental treatment arm (c) when varying the total sample size  $N$  from 40 to 400 when the historical sample size is  $N_0 = 350$ . The shadow indicates the 95% confidence intervals of the absolute error.



**FIGURE 5.** The results from 2000 simulations. The absolute error (a), the power when the type I error is 0.05 (b), and the percentage of the subjects assigned to experimental treatment arm (c) when varying the total historical sample size  $N_0$  from 30 to 350 when the current sample size is  $N = 200$ . The shadow indicates the 95% confidence intervals of the absolute error.

**TABLE 1**

Tuning parameter selections.

	<b>Algorithm 2</b>
$\lambda_1$	Minimize the average estimation error of the treatment effect in the simulation.
$\lambda_{2,\delta}$	Less than the smallest distance between two covariate values in the historical study
The kernel bandwidth for discrete covariate	Rule-of-thumb bandwidth discussed in Scott (2015) <sup>33</sup>
The kernel bandwidth for continuous covariate	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**TABLE 2**

The comparison between CAHB and SLHBD designs. The 95% confidence intervals (CIs) are reported. The power is reported when the type I error is controlled at 0.05.

Design	Allocation ratio			Absolute error	Power
	Overall	$X = 0$	$X = 1$	$ \delta - \hat{\delta} $	
Scenario 1: $(\alpha_0, \alpha_1) = (1, 10)$					
CAHB	0.643	0.817	0.468	0.070	0.959
95% CI	[0.642, 0.644]	[0.816, 0.819]	[0.466, 0.469]	[0.068, 0.072]	
SLHBD	0.504	0.502	0.506	0.146	0.681
95% CI	[0.503, 0.505]	[0.500, 0.504]	[0.503, 0.508]	[0.142, 0.151]	
Scenario 2: $(\alpha_0, \alpha_1) = (6, -9)$					
CAHB	0.501	0.500	0.501	0.080	0.912
95% CI	[0.500, 0.502]	[0.499, 0.502]	[0.499, 0.502]	[0.077, 0.082]	
SLHBD	0.814	0.812	0.815	0.340	0.166
95% CI	[0.812, 0.815]	[0.810, 0.815]	[0.813, 0.817]	[0.329, 0.352]	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**TABLE 3**

The subgroup-specific parameters used in the simulation study to generate current study samples, which are estimated from the HORIZON dataset.

	Subgroup 1 ( $X_1 = 0, X_2 = 0$ )	Subgroup 2 ( $X_1 = 1, X_2 = 0$ )	Subgroup 3 ( $X_1 = 0, X_2 = 1$ )	Subgroup 4 ( $X_1 = 1, X_2 = 1$ )
$\beta_0(X_1, X_2)$	-0.159	-0.170	-0.175	-0.189
$\beta_1(X_1, X_2)$	-0.003	-0.055	0.009	0.004
$\beta_2(X_1, X_2)$	0.922	0.954	0.939	0.910