

Use of common control arms across substudies in a master protocol via ridge estimation

Akinori Nimura^{a*, b} and Tosiya Sato^{c, d}

^aDepartment of Biostatistics, Kyoto University School of Public Health, Kyoto, Japan;

^bBiostatistics, GlaxoSmithKline K.K., Tokyo, Japan; ^cInstitute of Statistical

Mathematics, Tokyo, Japan; ^dData Science and AI Innovation Research Promotion

Center, Shiga University, Shiga, Japan

Address for the corresponding author: Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

Email for the corresponding author: akinori.x.nimura@gsk.com

ORCID ID

Akinori Nimura <https://orcid.org/0009-0004-4533-5928>

Tosiya Sato <https://orcid.org/0000-0001-9830-013X>

Use of common control arms across substudies in a master protocol via ridge estimation

In a master protocol design, multiple substudies are employed to address various research questions, with each substudy potentially having common control arms. When analyzing a substudy of interest, borrowing a control arm from another substudy may enhance the precision of the estimates. Several Bayesian methodologies have been proposed for borrowing data from other studies, particularly in the context of historical control borrowing. These have operating characteristics according to which the closer the results are between current and historical studies, the larger the amount of data borrowed from the historical study. However, such approaches often require a diagnostic assessment of convergence in Markov chain Monte Carlo methods, and pre-specification of hyperparameters, which would not be feasible for not only the primary endpoint but also the secondary and safety endpoints. The shrinkage of the difference parameters between substudies is equivalent to partial borrowing from other substudies. In this study, we propose using ridge estimator for shrinkage, in which the tuning parameter is determined based on the observed data. The ridge estimator has a closed form given the tuning parameters, and the tuning parameters can be easily computed. A simulation study demonstrated that the proposed estimator is safe and exhibits conservative performance.

Keywords: master protocol design; substudies; control arms; ridge; borrowing

1. Introduction

A master protocol is defined as one designed with multiple substudies to address various research questions (Woodcock and LaVange 2017; FDA 2022). These substudies may have common control arms such as a placebo. Common control arms can accelerate drug development efficiently.

While analyzing data from a substudy of interest, borrowing a control arm from another substudy may improve the precision of the treatment effect and increase study power without requiring a large sample size in the substudy of interest. As a motivating example, a master protocol design was used in a phase II study (Chronic Pain Master

Protocol) on three types of chronic pain diseases (osteoarthritis, chronic low back pain, and diabetic peripheral neuropathic pain) and four study treatments (NCT number: NCT05986292). All substudies were placebo-controlled, with the common primary endpoint being changes from baseline in pain intensity measured using a numerical rating scale. In this case, the data from the placebo arms in each substudy were expected to be comparable under the same pain subtype, and the placebo arm may be borrowed for each substudy analysis. However, one must consider the potential differences in the common control arms between substudies owing to population differences and time trends. To handle these potential differences, a hierarchical Bayesian model has been discussed with the FDA as an approach to borrow placebo arm data from other substudies on the same pain disease (Price and Scott 2021; Chatterjee and Kemmer 2022).

In addition to the hierarchical models, several other Bayesian methodologies may be adopted. For example, power priors, commensurate priors, and robust meta-analytic predictive priors have been proposed in the context of historical control borrowing (Chen and Ibrahim 2000; Hobbs et al. 2011; Schmidli et al. 2014). In these methods, a prior distribution is formed such that the larger the difference between concurrent and historical studies, the smaller the amount of control arm data borrowed. This mitigates the risk that a large difference would have on the operating characteristics (i.e., robustness to the potential difference between the studies). However, these methods have two challenges. First, the hyperparameters must be determined before analyzing the data. It may be possible to determine the hyperparameters of the primary analysis for the primary efficacy endpoint using a simulation to assess the operating characteristics for each hyperparameter. It may be advantageous to select the hyperparameters based on prior ignorance and/or prior

knowledge. However, it would require considerable effort to determine the hyperparameters for other analyses, such as the secondary efficacy endpoints and occurrence of adverse events, in which borrowing these data would also be preferable. Second, Bayesian methods often require Markov chain Monte Carlo (MCMC) computation, which may encounter convergence problem while sampling from a posterior distribution and would require tuning MCMC settings such as the number of iterations, the number of burn-ins, and the thin parameter via assessment of proper convergence. Furthermore, it is known that convergence worsens in various cases: 1) when each study is assumed to follow a common normal distribution in hierarchical modeling, the variance between studies is close to zero (Lambert et al. 2005); and 2) the posterior is multimodal, such as a mixture distribution. These models are used as methodologies for historical borrowing and may require more sampling.

In this study, we propose a ridge-based method that offers a stable estimate of the treatment effect in a substudy and determines the tuning parameters based on the observed data without predefining them. This method is particularly applicable under a randomized umbrella study designs with a common control arm across substudies, as demonstrated in the motivating example. Ridge estimation was originally developed for non-orthogonal problems in linear regression (Hoerl and Kennard 1970). This estimation generates a shrinkage estimate by applying a penalty, enabling stable estimation even in the presence of multicollinearity. An important assumption for the borrowing control data is that no difference exists in the control data between the substudies of interest and the substudies for borrowing. Under this assumption, pooling the control data from all the substudies would be reasonable. One should not borrow data if the difference is large; partially borrowing data may be appropriate if the difference is not large. In a linear model, the process of pooling control data

corresponds to imposing zero difference parameters in the control arms between the substudies, and no borrowing is equivalent to an unrestricted estimation of the difference parameters. We propose using ridge-type shrinkage for the difference parameter to implement the partial borrowing of the control data. Moreover, tuning parameter, which is the penalty parameter, can be obtained via a simple procedure. The proposed method provides a closed form given the tuning parameters, leading to less effort when analyzing the data for reporting than existing methods requiring MCMC.

The remainder of this paper is organized as follows. In Section 2, we describe the details of the proposed method that uses ridge estimation. Simulation studies are conducted in Section 3 to assess the operating characteristics of the proposed method compared to existing approaches, including a hierarchical Bayesian model, demonstrating both the safe and conservative features of the proposed method. In Section 4, we apply the proposed method to the aggregated data in the Chronic Pain Master Protocol. Finally, a discussion is presented in Section 5.

2. Methods

In this section, we first describe the study design within the scope of the proposed method, assumed linear model, and least squares estimation. Next, the tuning parameter in the ridge estimation is introduced into the least squares estimation, and the determination of the tuning parameter is provided. Finally, interval estimation is described.

2.1. Study Design in Scope

Our proposed method can be applied to a master protocol in which multiple substudies have common control arms (see Figure 1). The Chronic Pain Master Protocol study comprised several independent substudies with placebo arms in common, making it

within the scope of our proposed method. Another study design in which the proposed method can be applied is discussed in Section 5.

2.2. *Least Squares Estimation*

As a hypothetical example, we consider one substudy of interest and the other two substudies for borrowing and consider estimating the treatment effect in a substudy of interest by borrowing data of the common control arm in a substudy for borrowing. Let x_i be an indicator variable for the i th participant ($i=1, 2, \dots, n$) coded as 1 for participants assigned to the study treatment and 0 for those assigned to the control; and z_{bi} is another indicator coded as 1 for participants from the b th substudy for borrowing ($b=1, 2$) or 0 for those from the substudy of interest. The linear model is expressed as follows:

$$y_i = \beta_0 + \beta_1 x_i + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \varepsilon_i,$$

where y_i is a continuous outcome variable, ε_i is a normal error term with mean 0 and variance σ_i^2 , $\boldsymbol{\beta} = (\beta_0, \beta_1)$ are mean parameters in the control and treatment difference in the substudy of interest, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ is the parameter of difference in the control arms between the substudy of interest and the substudy for borrowing. The design matrices involving $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are given as

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \text{ and } \mathbf{Z} = \begin{pmatrix} z_{11} & z_{21} \\ \vdots & \vdots \\ z_{1n} & z_{2n} \end{pmatrix}, \quad (1)$$

respectively. If one covariate is added, such as the baseline value of the outcome variable a_i , the design matrices can be rewritten as

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & a_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & a_n \end{pmatrix} \text{ and } \mathbf{Z} = \begin{pmatrix} z_{11} & z_{11}a_1 & z_{21} & z_{21}a_1 \\ \vdots & \vdots & \vdots & \vdots \\ z_{1n} & z_{1n}a_n & z_{2n} & z_{2n}a_n \end{pmatrix}.$$

In the general case, consider a linear model for the data from all substudies.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (2)$$

where $\mathbf{y}(n \times 1)$ is an outcome vector, $\mathbf{X}(n \times l)$ and $\mathbf{Z}(n \times m)$ are design matrices, and $\boldsymbol{\varepsilon}(n \times 1)$ is a vector of the normal error terms of which mean is equal to 0 (i.e., $E[\boldsymbol{\varepsilon}] = \mathbf{0}^{n \times 1}$). In addition, $\boldsymbol{\beta}(l \times 1) = (\beta_1, \dots, \beta_l)^t$ involves parameters of interest such as the treatment effect, and $\boldsymbol{\gamma}(m \times 1) = (\gamma_1, \dots, \gamma_m)^t$ is a parameter vector related to the difference between substudies of interest and the other substudy(studies) for borrowing. Note that n is the total sample size in the substudy of interest and substudies for borrowing, l and m are the numbers of dimensions of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Given this linear model, the least squares estimators are given by

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} - (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \mathbf{y} \\ (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \mathbf{y} \end{pmatrix},$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t$. From the above formula, $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t (\mathbf{y} - \mathbf{Z} \hat{\boldsymbol{\gamma}})$. The variance of $\hat{\boldsymbol{\gamma}}$ is given by

$$V[\hat{\boldsymbol{\gamma}}] = (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \times \text{cov}(\mathbf{y}) \times \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1},$$

where $\text{cov}(\mathbf{y})$ is a diagonal matrix of $\{\sigma_1^2, \dots, \sigma_n^2\}$ (i.e., heteroskedasticity is assumed).

Because the variance of the outcome vector is unknown, $V[\hat{\boldsymbol{\gamma}}]$ is obtained using the sandwich estimator given by

$$\hat{V}[\hat{\boldsymbol{\gamma}}] = (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \times \begin{pmatrix} \hat{\varepsilon}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\varepsilon}_n^2 \end{pmatrix} \times \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1},$$

where $\hat{\varepsilon}_i = y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\boldsymbol{\gamma}}$. Let $\mathbf{P}$ and $\hat{\mathbf{P}}$ be orthogonal transformation such that $V^{-1}[\hat{\boldsymbol{\gamma}}] = \mathbf{P} \mathbf{D} \mathbf{P}^t$ and $\hat{V}^{-1}[\hat{\boldsymbol{\gamma}}] = \hat{\mathbf{P}} \hat{\mathbf{D}} \hat{\mathbf{P}}^t$, where $\mathbf{D}$ and $\hat{\mathbf{D}}$ is the diagonal matrix of eigenvalues $\{d_1, \dots, d_m\}$ of $V^{-1}[\hat{\boldsymbol{\gamma}}]$ and $\{\hat{d}_1, \dots, \hat{d}_m\}$ of $\hat{V}^{-1}[\hat{\boldsymbol{\gamma}}]$. With the regularity conditions, $(\hat{\boldsymbol{\beta}}^t, \hat{\boldsymbol{\gamma}}^t)^t \xrightarrow{a.s.} (\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)^t$ and $n \times \hat{V}[\hat{\boldsymbol{\gamma}}] \xrightarrow{a.s.} n \times V[\boldsymbol{\gamma}]$ (White 1980). Therefore, $\hat{\mathbf{P}}^t \hat{\boldsymbol{\gamma}} \xrightarrow{a.s.} \mathbf{P}^t \boldsymbol{\gamma} = \boldsymbol{\alpha}$ by the continuous mapping theorem, and $\hat{\mathbf{P}}^t \hat{\boldsymbol{\gamma}} \xrightarrow{d} \mathbf{P}^t \boldsymbol{\gamma} \sim N(\boldsymbol{\alpha}, \mathbf{D}^{-1})$ by Slutsky's lemma. We consider shrinkage estimation of $\boldsymbol{\gamma}$ via $\boldsymbol{\alpha}$ based on $\hat{\boldsymbol{\alpha}} = \hat{\mathbf{P}}^t \hat{\boldsymbol{\gamma}}$.

If one assumes that the variances may be different between substudies but equal within the substudy, the restricted maximum likelihood (REML) method can be another way of estimating them (Searle, Casella, and McCulloch 1992). The variance of the outcome vectors is given by

$$\text{cov}(\mathbf{y}) = \begin{pmatrix} \sigma_1^2 \mathbf{I}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_B^2 \mathbf{I}_B \end{pmatrix},$$

where σ_b^2 ($b = 1, \dots, B$) is the variance in b^{th} substudy and $\mathbf{I}_b$ is an identity matrix of which rank is the same as the number of participants in b^{th} substudy. Therefore, $V[\hat{\boldsymbol{\gamma}}]$ is obtained using REML estimate as given by,

$$\hat{V}[\hat{\boldsymbol{\gamma}}] = (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \times \begin{pmatrix} \hat{\sigma}_1^2 \mathbf{I}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\sigma}_B^2 \mathbf{I}_B \end{pmatrix} \times \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1}.$$

2.3. Borrowing via Ridge Estimation

In the linear model (2), imposing $\boldsymbol{\gamma} = 0$ is equivalent to full borrowing of data from

substudies because this means that no difference is assumed between the substudies. In contrast, no restriction of $\boldsymbol{\gamma}$ is equivalent to no borrowing, which is discussed in Supplementary Section 1. Therefore, we propose using ridge estimation for a shrinkage of $\boldsymbol{\gamma}$ to 0 as partial borrowing. The estimating equation of $\hat{\boldsymbol{\gamma}}$ is given by $(\mathbf{Z}^t \mathbf{M} \mathbf{Z}) \boldsymbol{\gamma} = \mathbf{Z}^t \mathbf{M} \mathbf{y}$. Considering the above section, the estimating equation of $\hat{\boldsymbol{\alpha}}$ is given by,

$$\begin{aligned} \mathbf{Z}^t \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \times \text{cov}(\mathbf{y}) \times \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \mathbf{Z} \boldsymbol{\gamma} &= \mathbf{Z}^t \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \times \text{cov}(\mathbf{y}) \times \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \mathbf{y} \\ \Leftrightarrow \mathbf{P} \mathbf{D} \boldsymbol{\alpha} &= \mathbf{P} \mathbf{D} \mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \mathbf{y} \\ \Leftrightarrow \mathbf{D} \boldsymbol{\alpha} &= \mathbf{D} \mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \mathbf{y}. \end{aligned}$$

As well as the general ridge estimation (Hoerl and Kennard 1970), the ridge estimator of $\boldsymbol{\alpha}$ is defined by,

$$\hat{\boldsymbol{\alpha}}^* = (\mathbf{D} + \boldsymbol{\Lambda})^{-1} \mathbf{D} \mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \mathbf{y} = (\mathbf{D} + \boldsymbol{\Lambda})^{-1} \mathbf{D} \hat{\boldsymbol{\alpha}},$$

where $\boldsymbol{\Lambda}$ is the ridge tuning parameters formed by a diagonal matrix of $\{\lambda_1, \dots, \lambda_m\}$, $\forall \lambda_j \geq 0$. As $\lambda_j \rightarrow 0$, the ridge estimator reaches to the least squares estimator (i.e., no borrowing) whereas the ridge estimator of the difference in the control arm between the substudies ($\hat{\boldsymbol{\alpha}}^*$) goes to 0 (i.e. full borrowing) as $\lambda_j \rightarrow \infty$. Finally, the ridge estimators are given by:

$$\hat{\boldsymbol{\alpha}}^* = (\hat{\mathbf{D}} + \boldsymbol{\Lambda})^{-1} \hat{\mathbf{D}} \hat{\boldsymbol{\alpha}} = \mathbf{Q} \hat{\boldsymbol{\alpha}},$$

$$\hat{\boldsymbol{\gamma}}^* = \hat{\mathbf{P}} \hat{\boldsymbol{\alpha}}^*, \text{ and}$$

$$\hat{\boldsymbol{\beta}}^* = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t (\mathbf{y} - \mathbf{Z} \hat{\boldsymbol{\gamma}}^*).$$

Although $\hat{\boldsymbol{\beta}}^*$ is not strictly a ridge estimator, we refer to this as such for convenience in this article. Suppose that the parameter of interest is defined as $\mathbf{c}^t \boldsymbol{\beta}$. In the case of

design matrices (1), β_1 is interpreted as the treatment effect and one may consider $\mathbf{c} = (0, 1)^t$. We call $\mathbf{c}^t \hat{\boldsymbol{\beta}}^*$ the proposed ridge estimator.

2.4. Determination of Λ

We aimed to obtain an estimate of interest by borrowing the common control arm in terms of the mean square error (MSE) of the proposed ridge estimator, $\mathbf{c}^t \hat{\boldsymbol{\beta}}^*$, which is given by,

$$\begin{aligned} \text{MSE}[\mathbf{c}^t \hat{\boldsymbol{\beta}}^* | \Lambda] &= \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{Z} \mathbf{P} \mathbf{W} \boldsymbol{\alpha} \boldsymbol{\alpha}^t \mathbf{W} \mathbf{P}^t \mathbf{Z}^t \\ &\quad + (\mathbf{I} - \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (\mathbf{I} - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t)^t \times \text{cov}(\mathbf{y}) \\ &\quad \times (\mathbf{I} - \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (\mathbf{I} - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t)] \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}. \end{aligned}$$

Although $\boldsymbol{\alpha} \boldsymbol{\alpha}^t$ needs to be replaced with its least squares estimates, there is a risk of excessively borrowing the data through underestimation of the MSE due to the uncertainty in the estimate, especially in case of the smaller sample size. To consider the uncertainty in the least squares estimate in the MSE, we average the squared bias through the parametric bootstrap approach (Efron and Tibshirani 1993) to obtain the precision measure below.

$$\begin{aligned} G(\Lambda) &= \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{Z} \mathbf{P} \mathbf{W} (\hat{\boldsymbol{\alpha}} \hat{\boldsymbol{\alpha}}^t + \mathbf{D}^{-1}) \mathbf{W} \mathbf{P}^t \mathbf{Z}^t \\ &\quad + (\mathbf{I} - \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (\mathbf{I} - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t)^t \times \text{cov}(\mathbf{y}) \\ &\quad \times (\mathbf{I} - \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (\mathbf{I} - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t)] \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}, \end{aligned}$$

where $\mathbf{W} = \text{diag}(\mathbf{w}) = \mathbf{I} - \mathbf{Q}$. The variance of outcomes, $\text{cov}(\mathbf{y})$, and the corresponding $\mathbf{P} \mathbf{D} \mathbf{P}^t$ are replaced with their estimates, which are provided in the above section, to determine the ridge tuning parameters (Λ) by minimizing $G(\Lambda)$. The detailed

derivation of $G(\Lambda)$ is provided in Appendix A. In practice, determining $\mathbf{W}$ rather than Λ itself is more useful. Here, let

$$\begin{aligned}\mathbf{w} &= (\lambda_1/(\hat{d}_1 + \lambda_1), \dots, \lambda_m/(\hat{d}_m + \lambda_m))^t, \\ \mathbf{s}_1 &= \hat{\mathbf{P}}^t \mathbf{Z}^t \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}, \\ \mathbf{S}_1 &= \text{diag}(\mathbf{s}_1), \text{ and} \\ \mathbf{s}_2 &= \hat{\mathbf{P}}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \times \widehat{cov}(\mathbf{y}) \times \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}.\end{aligned}$$

The precision measure and its derivative can be written as follows:

$$G(\Lambda) = \text{const.} + \mathbf{w}^t \mathbf{S}_1 (2\hat{\mathbf{D}}^{-1} + \hat{\alpha}\hat{\alpha}^t) \mathbf{S}_1 \mathbf{w} - 2\mathbf{w}^t \mathbf{S}_1 (\hat{\mathbf{D}}^{-1} \mathbf{s}_1 - \mathbf{s}_2), \text{ and}$$

$$\frac{\partial}{\partial \mathbf{w}} G(\Lambda) = 2\mathbf{S}_1 (2\hat{\mathbf{D}}^{-1} + \hat{\alpha}\hat{\alpha}^t) \mathbf{S}_1 \mathbf{w} - 2\mathbf{S}_1 (\hat{\mathbf{D}}^{-1} \mathbf{s}_1 - \mathbf{s}_2).$$

Finally, $\mathbf{W} (= \text{diag}(w_1, \dots, w_m))$ is determined as follows.

$$(\hat{w}_1, \dots, \hat{w}_m)^t = \hat{\mathbf{w}} = \mathbf{S}_1^{-1} (2\hat{\mathbf{D}}^{-1} + \hat{\alpha}\hat{\alpha}^t)^{-1} (\hat{\mathbf{D}}^{-1} \mathbf{s}_1 - \mathbf{s}_2) \quad (3)$$

However, some $\hat{w}_j$ values may be lower than zero or greater than one. Therefore, we propose the following procedure for determining $\mathbf{w}$:

1. Obtain $\hat{\mathbf{w}} = (\hat{w}_1, \dots, \hat{w}_m)^t$ from (3).
2. If $\hat{w}_j$ was lower than zero, it was set to zero. Otherwise, if $\hat{w}_j$ is greater than 1, it is set as 1. By imposing this, we obtain $\hat{\mathbf{w}} = (\hat{w}_1, \dots, \hat{w}_m)^t$ using the reduced system. Repeat this step until all $(\hat{w}_1, \dots, \hat{w}_m)^t \in [0, 1]^m$.

$\hat{\mathbf{w}}$ obtained from this procedure satisfies the Karush–Kuhn–Tucker condition. The proof of this is provided in Appendix C. This procedure is similar to adaptive likelihood weights (Plante 2008). The estimator generated using this procedure converges to the

true value in probability (i.e., a consistent estimator). The proof of this is provided in Appendix D.

2.5. Variance and Confidence Interval

The variance of the proposed estimator is derived using the mean square error of $\hat{\beta}^*$ unconditional on $\mathbf{\Lambda}$ below. As $\hat{\mathbf{Q}}\hat{\mathbf{a}}$ is the function of $\mathbf{M}\mathbf{y}$ and, $cov(\mathbf{y})$ is assumed to be known, the unbiased estimate of MSE is given from Stein's lemma (Stein 1981) by

$$\begin{aligned} MSE[\mathbf{c}^t \hat{\beta}^*] &= E \left[\mathbf{c}^t (\hat{\beta}^* - \beta) (\hat{\beta}^* - \beta)^t \mathbf{c} \right] \\ &= \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \\ &\quad \times E \left[\mathbf{Z} \mathbf{P} \widehat{\mathbf{W}} \hat{\mathbf{a}} \hat{\mathbf{a}}^t \widehat{\mathbf{W}}^t \mathbf{P}^t \mathbf{Z}^t + \mathbf{H}_1 cov(\mathbf{y}) \mathbf{H}_1^t + \mathbf{H}_2 cov(\mathbf{y}) \mathbf{H}_1^t + \mathbf{H}_1 cov(\mathbf{y}) \mathbf{H}_2^t \right] \\ &\quad \times \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}, \end{aligned}$$

where:

$$\begin{aligned} \mathbf{H}_1 &= \mathbf{I} - \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M}, \text{ and} \\ \mathbf{H}_2 &= \mathbf{Z} \mathbf{P} \left\{ diag(\hat{\mathbf{a}}) \frac{\partial \widehat{\mathbf{W}}}{\partial \hat{\mathbf{a}}^t} + \widehat{\mathbf{W}} \right\} \mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M}. \end{aligned}$$

An approximately unbiased estimate of the MSE is obtained by replacing $cov(\mathbf{y})$ and related quantities, such as $\mathbf{P} \mathbf{D} \mathbf{P}^t$ with their estimates. A detailed derivation of this MSE is provided in Appendix B. Given that an unbiased estimate of $\mathbf{a} \mathbf{a}^t$ is $\hat{\mathbf{a}} \hat{\mathbf{a}}^t - \mathbf{D}^{-1}$, the variance estimator of $\mathbf{c}^t \hat{\beta}^*$ is defined below.

$$\begin{aligned} \hat{V}[\mathbf{c}^t \hat{\beta}^*] &= \widehat{MSE}[\mathbf{c}^t \hat{\beta}^*] - \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Z} \widehat{\mathbf{P}} \widehat{\mathbf{W}} (\hat{\mathbf{a}} \hat{\mathbf{a}}^t - \widehat{\mathbf{D}}^{-1}) \widehat{\mathbf{W}}^t \mathbf{P}^t \mathbf{Z}^t \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c} \\ &= \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \\ &\quad \times \left[\mathbf{X}^t \mathbf{Z} \widehat{\mathbf{P}} \widehat{\mathbf{W}} \widehat{\mathbf{D}}^{-1} \widehat{\mathbf{W}}^t \mathbf{P}^t \mathbf{Z}^t + \mathbf{H}_1 cov(\mathbf{y}) \mathbf{H}_1^t + \widehat{\mathbf{H}}_2 cov(\mathbf{y}) \mathbf{H}_1^t \right. \\ &\quad \left. + \mathbf{H}_1 cov(\mathbf{y}) \widehat{\mathbf{H}}_2^t \right] \times \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}. \end{aligned}$$

The minimum of the variance is given by $V_{\min}[\mathbf{c}^t \hat{\boldsymbol{\beta}}^*] = \min_{\mathbf{W}} V[\mathbf{c}^t \hat{\boldsymbol{\beta}}^* | \mathbf{W}]$, where

$$\begin{aligned} V[\mathbf{c}^t \hat{\boldsymbol{\beta}}^* | \mathbf{W}] &= \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \\ &\times [(I - \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (I - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t)^t \times \text{cov}(\mathbf{y}) \\ &\times (I - \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (I - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t)] \times \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}. \end{aligned}$$

Under homoskedasticity assumption (i.e., $\text{cov}(\mathbf{y}) = \sigma^2 \mathbf{I}$), this variance conditional on $\mathbf{W}$ is given below and this indicates that full borrowing (i.e., $\mathbf{W} = \mathbf{I}$) always achieves the minimum variance which is shown as $\sigma^2 \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}$.

$$\begin{aligned} V[\mathbf{c}^t \hat{\boldsymbol{\beta}}^* | \mathbf{W}] &= \sigma^2 \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \times [\mathbf{I} + \mathbf{Z} \mathbf{P} (I - \mathbf{W}) \mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (I - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t] \\ &\times \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}. \end{aligned}$$

However, under the heteroskedasticity assumption, there are no explicit values for the tuning parameters for the minimum variance, and full borrowing does not always minimize the variance. The minimum variance can be obtained using the same procedure mentioned in Section 2.4 because the variance is given by

$$\text{const.} + \mathbf{w}^t \mathbf{S}_1 \hat{\mathbf{D}}^{-1} \mathbf{S}_1 \mathbf{w} - 2 \mathbf{w}^t \mathbf{S}_1 (\hat{\mathbf{D}}^{-1} \mathbf{s}_1 - \mathbf{s}_2).$$

After all, if $\hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}^*] < V_{\min}[\mathbf{c}^t \hat{\boldsymbol{\beta}}^*]$, $\hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}^*]$ is defined as $V_{\min}[\mathbf{c}^t \hat{\boldsymbol{\beta}}^*]$. The confidence interval is computed based on the t-distribution, with $n - l - m$ as the degrees of freedom.

2.6. Analysis Implementation

Below is the algorithm to compute the proposed estimate and confidence interval.

Algorithm to conduct the proposed method

Input: Outcome vector $\mathbf{y}$, Design matrices $\mathbf{X}$ and $\mathbf{Z}$

Output: Treatment effect $\mathbf{c}^t \hat{\boldsymbol{\beta}}^*$ and its variance estimate $\hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}^*]$

1: Compute the least squares estimate from the input dataset

- 2: Determine the tuning parameter $\hat{\Lambda}$ (or $\hat{w}$) given by $\hat{\Lambda} = \arg \min_{\Lambda} G(\Lambda)$ subject to $\forall j, \lambda_j \geq 0$ as illustrated in Section 2.4
 - 3: Compute the treatment effect estimate using the least squares estimates and the selected tuning parameter
 - 4: Compute the variance estimate and construct the confidence interval based on t-distribution as mentioned in Section 2.5
-

An analysis of a substudy of interest using the proposed method may be conducted as follows, although another procedure may be considered.

1. Prespecify substudies that may be borrowed to analyze the substudy of interest in any relevant document (e.g., master study protocol).
2. Identify substudies that are included in the prespecified substudies for borrowing and have been completed at the database lock of the substudy of interest.
3. Apply the proposed method to data from a substudy of interest and identified substudies for borrowing.

It is important to pre-specify the substudies that may be potentially used for data borrowing in a master protocol prior to starting all the substudies from the perspective of study/analysis integrity if one wants to apply the proposed method as the primary analysis. For example, consider osteoarthritis substudies in the Chronic Pain Master Protocol. It may be easier to justify borrowing information from osteoarthritis substudies than from substudies on other diseases. Therefore, one may be able to pre-specify osteoarthritis substudies as those that may be utilized for data borrowing when analyzing the substudy in the same disease in a master study protocol.

3. Simulations

We conducted a simulation study to evaluate the performance of the proposed ridge

estimator for the treatment effect compared with existing methods. In all scenarios, we set up one substudy of interest and another substudy for borrowing. The outcome distributions for control and study treatment arms in the substudy of interest are $y_{ic} \sim N(0, 1.5^2)$ for $i = 1, \dots, n_c (= 50)$ and $y_{ia} \sim N(\mu_a, 1.5^2)$ for $i = 1, \dots, n_a (= 100)$, respectively. The scenarios with $\mu_a = 0$ and $\mu_a = 0.5$ were examined. In addition, the outcome distribution for the control arm in the substudy for borrowing was $y_{ib} \sim N(\mu_b, s_{borrow}^2)$ for $i = 1, \dots, n_b$. We explored the scenarios with combinations of $\mu_b = -1.5$ to 1.5 by 0.25 , $s_{borrow} = 0.5, 1.5, 3.0$ and $n_b = 50, 100$ (totally, 78 scenarios) given the results from the Chronic Pain Master Protocol (CPMP) (NCT number: NCT05986292).

We generated 10,000 (=M) simulation datasets for each scenario and computed the proposed ridge estimator for each dataset, along with the full borrowing estimator using data from both substudies and the no-borrowing estimator using data from the substudy of interest only. The design matrices in the proposed ridge estimator are given by

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \text{ and } \mathbf{Z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix},$$

where $n = n_a + n_c + n_b$, x_i is the treatment arm indicator coded as 0 (control) or 1 (study treatment), and z_i is the substudy indicator coded as 0 (substudy of interest) or 1 (substudy for borrowing). Given this design matrices, the true values of β and γ are μ_a and μ_b , respectively. In addition, the hierarchical Bayesian model, as in the method discussed for the CPMP, and power prior (the hyperparameter is determined based on the marginal likelihood) were fitted to the same datasets (Ibrahim et al. 2015). Details of the methodologies are provided in Supplementary Section 2.

The following measures were used to assess the performance:

$$Bias = \frac{1}{M} \sum_{j=1}^M (\hat{\beta}_j - \beta),$$

$$Mean Square Error = \frac{1}{M} \sum_{j=1}^M (\hat{\beta}_j - \beta)^2,$$

$$Power \text{ and Type 1 error rate} = \frac{1}{M} \sum_{j=1}^M I(\hat{\beta}_j^L > 0), \text{ and}$$

$$Coverage probability = \frac{1}{M} \sum_{j=1}^M I(\hat{\beta}_j^L < \beta < \hat{\beta}_j^U)$$

where β is the true value of the treatment effect, $\hat{\beta}_j$ is the point estimate in a frequentist method and the mean of the posterior distribution in a Bayesian method, and $\hat{\beta}_j^L, \hat{\beta}_j^U$ are the lower and upper limits of 95% confidence interval or credible interval. The code is available on GitHub (https://github.com/crist/ridge_master_protocol).

Figure 2 shows the simulation results with 100 participants and 1.5 standard deviation in the substudy for borrowing excluding the coverage probability (for the coverage probability, see Table S10). This figure does not include full borrowing for ease of presentation and does not include the proposed ridge estimator with sandwich variance because its operating characteristics are almost identical to those of the proposed ridge estimator with REML. The proposed ridge estimator, along with the other methods, had no bias when there was no difference between the substudies. The proposed estimator has less bias than the other methods, except for no borrowing, when there is a difference. Additionally, all borrowing methods had smaller MSE in the scenario of no/small difference (within ± 0.25) between substudies than no borrowing, which implied that the amount of improvement in the variance of the estimate outweighed the bias derived from the partial borrowing. In contrast, borrowing methods

had a larger MSE than no borrowing when there was a large difference, and the greater the difference between the substudies, the closer the MSE was to no borrowing. The proposed ridge estimator had a closer MSE to no borrowing than the other borrowing methods when there was a large difference between the substudies. However, it had a slightly greater MSE than the other methods when there was little or no difference.

All methods generally controlled the type 1 error rate to less than 2.5% when there was no difference between the substudies. However, the error rate was inflated when the mean of the control arm was smaller in the substudy for borrowing since the treatment difference was overestimated and the null hypothesis of no treatment effect was excessively rejected in these scenarios only. The proposed method had 2–3% inflation compared with the no-borrowing (2.8%), of which type 1 error rate was generally smaller than the other borrowing methods and greater by <1% than the hierarchical Bayesian model with $\tau \sim U(0, 1)$ when the difference between the substudies was 0 to -0.5 (see Table S6 in Supplementary Section 3.3). When the difference between the substudies was less than -1.0, the proposed ridge estimator demonstrated a smaller type 1 error rate than the other methods. Furthermore, the proposed ridge estimator had a higher power than no borrowing when there was no difference between the substudies. Finally, the proposed estimator had no worse coverage probability than the other methods (Table S10 in Supplementary Section 3.5).

The simulation results for the other scenarios are presented in Supplementary Section 3. A similar trend was observed, except for cases in which the power prior was applied and the standard deviation in the substudy for borrowing was not 1.5.

In scenarios with 1.5 standard deviation in the substudy for borrowing, the power prior had a smaller MSE and higher power when there was no difference

between the substudies compared with the proposed ridge estimator. However, a larger bias and an inflated type 1 error rate were observed when the difference was large.

In the substudy for borrowing, when the standard deviation was 3.0, the power prior exhibited a larger bias and MSE, and a lower power compared with the other methods. In addition, contrasting results were obtained when the standard deviation was 0.5. Other borrowing methods that can account for unequal variance between substudies showed similar trends between scenarios of equal and unequal variances.

The performance of the hierarchical Bayesian model depends on hyperprior specifications. The vaguer hyperprior led to less borrowing data from the other substudy across the scenarios and $\tau \sim U(0, 1)$ had similar performance to the proposed estimator in this simulation when the difference between the substudies was within 0.25.

4. Application to the Chronic Pain Master Protocol

As of 16th June 2024, nine substudies (three substudies for each disease subtype) were published at ClinicalTrials.gov, and one substudy was terminated early (NCT number: NCT04707157). As mentioned in Section 1, all substudies were placebo controlled, and the placebo arm may be borrowed to analyze the substudy of interest. In Table 1, we show that the substudies in bold are of interest and the other substudies are borrowed. The primary efficacy analysis was a Bayesian mixed-model repeated-measures analysis for the average pain intensity of weekly time intervals, as measured by the NRS. The posterior mean of the outcome at week 8 in each treatment arm and the 95% credible interval from the above analyses are presented in Table 1. Table 1 shows the heterogeneity among almost all substudies, although the mean was closer between NCT04476108 and NCT05177094 in diabetic peripheral neuropathic pain (DPNP) and between NCT04627038 and NCT05080660 in osteoarthritis (OA). In cases where there were small differences in control between substudies, borrowing data from the common

control arms was expected to yield a more precise estimate. The detailed procedure for computing all methods is provided in the codes available in GitHub (https://github.com/crist/ridge_master_protocol).

The proposed ridge estimator and comparator methods were applied to the data. For each disease subtype, there were two substudies for borrowing. Therefore, three estimates were computed: two by borrowing each substudy and one by borrowing two substudies. The results for DPNP and OA are presented in Tables 2 (DPNP) and 3 (OA). When borrowing NCT04476108 in DPNP and NCT04627038 in OA, all borrowing methods produced estimates with narrower confidence/credible intervals compared with no borrowing. Additionally, all borrowing methods resulted in more precise estimates compared to no borrowing when both NCT04476108 and NCT04707157 were borrowed for DPNP and when both NCT04627038 and NCT04456686 were borrowed for OA. The hierarchical Bayesian model and power prior yielded point estimates closer to those in full borrowing (i.e., the difference in control from the substudy for borrowing was closer to 0) than the proposed ridge estimator in most cases, which was consistent with the trend in the bias of the simulation results.

The results in Chronic low back pain (CLBP) substudies are provided in Supplementary Section 4.

5. Discussion

This study proposes a new estimation method that borrows data from the common control arm of another substudy using a single master protocol study. The proposed method provides a closed form given the tuning parameters, and a simple method for computing the tuning parameters. This allows for stable estimation without any convergence assessment, such as MCMC, and requires less effort when analyzing data

for reporting purposes. As a result, it is more feasible to apply the proposed method to both primary and secondary/exploratory continuous endpoints.

In addition, the proposed method does not require prespecification of the tuning parameters at the design stage. As shown in the simulation results, the hyperparameters in the hierarchical Bayesian model affect the operating characteristics. Typically, the hyperparameters are determined based on the operating characteristics assessed through simulations. The scenarios in these simulations may include assumptions about the difference in the mean parameters in the control arms between the substudies, the sample size, the number of substudies, etc., resulting in the need for extensive simulations will be required. The proposed method does not require such simulations and requires less effort during the design stage.

In the simulation, the proposed estimator was compared with the power prior, which does not require the pre-specification of hyperparameters. Simulation results demonstrated the more conservative features of the proposed estimator, because less information was borrowed from the other substudy compared with the power prior in scenarios with equal variance between the substudy of interest and the substudy for borrowing. This is because the power prior borrows full data when there is no or a small observed difference between the substudies (Han, Ye, and Wang 2023), whereas the proposed estimator cannot borrow full data when one substudy is borrowed and equal variance is assumed between the substudies. This feature indicates that the proposed method leads to a safer conclusion (e.g., the efficacy of a study treatment is unlikely to be falsely demonstrated under the null hypothesis of no treatment effect) than the power prior. In the early phase, such as in a Phase II study, less information may be available regarding the difference in the mean parameters between the substudies at the design stage, and there might be a weaker prior belief to borrow the information from the other

substudy. In such a case, the proposed estimator is safer and more conservative than the power prior and may be recommended. In the scenario with a larger variance in the substudy for borrowing than in the substudy of interest, the power prior showed poor operating characteristics because it assumed equal variance of the outcome variable between substudies, leading to an underestimation of the variance in the substudy for borrowing and excessive borrowing of data from it.

In some platform trials, multiple study treatments were started sequentially alongside a shared control arm (Figure 3). For example, the I-SPY2 platform study had multiple treatment arms with one shared control arm (Barker et al. 2009). In this setting, the data from the control arm are separated into concurrent and non-concurrent control arms for analysis. Three approaches have been proposed for this study design: 1) direct modeling of the time trend in the outcome over time (Bofill Roig et al. 2022; Saville et al. 2022), 2) modeling the time trend by accounting for the difference in covariates between concurrent and non-concurrent controls using propensity score weighting (Guo, Wang, and Yuan 2024) and 3) a network meta-analysis approach that considers this study as a network of direct and indirect comparisons (Marschner and Schou 2022). The proposed estimator can be used in this study design by considering the concurrent and non-concurrent control arms as a substudy of interest and a substudy for borrowing, respectively.

The proposed estimator is currently only applicable to continuous outcomes. However, in future work, it may be extended to other types of variables, such as categorical and time-to-event outcomes. The ridge tuning parameters are determined based on the precision measure. However, the measures for these types of variables may be nonlinear in terms of the tuning parameters, and the computation may be more complicated to determine them. Although a normal approximation of parameter

estimates in generalized linear models may be one solution, this approximation may not work well with a small sample size, such as in cases where data borrowing is required. Furthermore, incorporating the effect of the time trend into the design matrix to improve the precision of the estimate is an area of future work for the I-SPY2-like platform study design discussed in the previous paragraph. Determining the appropriate sample size is another challenge. Traditionally, the sample size is determined in each substudy, but optimizing measures, such as the utility function formed by data from each substudy, can be a solution. Finally, no scenarios in which a covariate was included in the linear model were used in the simulation. As shown in the hypothetical example in Section 2.2, the parameter of the difference in covariate effects can be included in the model, and the covariate effect can be borrowed by shrinking this parameter to zero. It is still unclear whether the covariate effect should be borrowed using the proposed method or separately estimated for each substudy when the control mean is borrowed.

In conclusion, the proposed method provides a stable estimate and determines tuning parameters based on observed data without pre-definition. This method is both safe and conservative and may be widely recommended in master protocol studies.

6. Supplementary Materials

Supplementary material is available online.

7. Acknowledgements

We are grateful to Kentaro Kudo for invaluable comments on this study. We would also like to thank Editage (www.editage.jp) for the English language editing.

8. Disclosure Statement

Akinori Nimura is employed by GlaxoSmithKline.

9. Funding

This work was supported by AMED under Grant Number: JP211k0201702.

10. Appendices

10.1. Appendix A: Derivation of precision measure to determine the tuning parameters

The MSE of $\mathbf{c}^t \hat{\boldsymbol{\beta}}^*$ is given by,

$$\begin{aligned} \text{MSE}[\mathbf{c}^t \hat{\boldsymbol{\beta}}^* | \Lambda] &= \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{Z} \mathbf{P} \mathbf{W} \boldsymbol{\alpha} \boldsymbol{\alpha}^t \mathbf{W} \mathbf{P}^t \mathbf{Z}^t \\ &\quad + (\mathbf{I} - \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (\mathbf{I} - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t)^t \times \text{cov}(\mathbf{y}) \\ &\quad \times (\mathbf{I} - \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} (\mathbf{I} - \mathbf{W}) \mathbf{P}^t \mathbf{Z}^t)] \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}. \end{aligned}$$

This MSE may be estimated by replacing $\boldsymbol{\alpha}$ with its least squares estimates. It is important to note that there is a risk of excessive borrowing of data because of uncertainty in the estimate, especially in the case of a smaller sample size. To address this, we considered averaging this estimated MSE in terms of $\hat{\boldsymbol{\alpha}}$. The precision measure is calculated as the average obtained using the parametric bootstrap principle.

Based on the principle of the parametric bootstrap approach, the distribution for the outcome vector can be estimated using the least squares estimates given by $\tilde{\mathbf{y}} \sim N(\mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{Z} \tilde{\boldsymbol{\gamma}}, \text{cov}(\mathbf{y}))$. From this estimated distribution, the least squares estimator of $\boldsymbol{\gamma}$ is estimated by,

$$\tilde{\boldsymbol{\gamma}} = (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \tilde{\mathbf{y}}.$$

With the same approach mentioned in the main paper, the least squares estimate of $\boldsymbol{\alpha}$ and its sample distribution can be obtained as follows.

$$\tilde{\boldsymbol{\alpha}} = \mathbf{P} \tilde{\boldsymbol{\gamma}} \sim N(\hat{\boldsymbol{\alpha}}, \mathbf{D}^{-1}).$$

Note that $\mathbf{P}, \mathbf{D}$ are determined independently of bootstrapped sample $\tilde{\mathbf{y}}$. The precision measure is defined as

$$\begin{aligned}
G(\Lambda) &= E_{\tilde{\alpha}|\hat{\alpha}}[\mathbf{c}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t[\mathbf{ZPW}\tilde{\alpha}\tilde{\alpha}^t\mathbf{W}^t\mathbf{P}^t\mathbf{Z}^t \\
&\quad + (\mathbf{I} - \mathbf{MZ}(\mathbf{Z}^t\mathbf{MZ})^{-1}\mathbf{P}(\mathbf{I} - \mathbf{W})\mathbf{P}^t\mathbf{Z}^t)^t \times \text{cov}(\mathbf{y}) \\
&\quad \times (\mathbf{I} - \mathbf{MZ}(\mathbf{Z}^t\mathbf{MZ})^{-1}\mathbf{P}(\mathbf{I} - \mathbf{W})\mathbf{P}^t\mathbf{Z}^t)]\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{c}] \\
&= \mathbf{c}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\{\mathbf{ZPW}(\hat{\alpha}\hat{\alpha}^t + \mathbf{D}^{-1})\mathbf{W}^t\mathbf{P}^t\mathbf{Z}^t \\
&\quad + (\mathbf{I} - \mathbf{MZ}(\mathbf{Z}^t\mathbf{MZ})^{-1}\mathbf{P}(\mathbf{I} - \mathbf{W})\mathbf{P}^t\mathbf{Z}^t)^t \times \text{cov}(\mathbf{y}) \\
&\quad \times (\mathbf{I} - \mathbf{MZ}(\mathbf{Z}^t\mathbf{MZ})^{-1}\mathbf{P}(\mathbf{I} - \mathbf{W})\mathbf{P}^t\mathbf{Z}^t)\}\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{c}.
\end{aligned}$$

10.2. Appendix B: Derivation of $\widehat{MSE}(\hat{\beta}^*)$

MSE of $\hat{\beta}^*$ is given by,

$$\begin{aligned}
MSE[\hat{\beta}^*] &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \\
&\quad \times E\left[\mathbf{ZP}(\alpha - \hat{Q}\hat{\alpha})(\alpha - \hat{Q}\hat{\alpha})^t\mathbf{P}^t\mathbf{Z}^t + \varepsilon\varepsilon^t + \mathbf{ZP}(\alpha - \hat{Q}\hat{\alpha})\varepsilon^t \right. \\
&\quad \left. + \varepsilon(\alpha - \hat{Q}\hat{\alpha})^t\mathbf{P}^t\mathbf{Z}^t\right] \times \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1} \\
&= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \times [\mathbf{ZP} \\
&\quad \times E\{(\alpha - \hat{\alpha})(\alpha - \hat{\alpha})^t + \widehat{W}\hat{\alpha}\hat{\alpha}^t\widehat{W} - \widehat{W}\hat{\alpha}(\hat{\alpha} - \alpha)^t - (\hat{\alpha} - \alpha)\hat{\alpha}^t\widehat{W}\} \\
&\quad \times \mathbf{P}^t\mathbf{Z}^t + \text{cov}(\mathbf{y}) - \mathbf{ZP} \times E\{\hat{Q}\hat{\alpha}\varepsilon^t\} - E\{\varepsilon\hat{\alpha}^t\hat{Q}\} \times \mathbf{P}^t\mathbf{Z}^t] \times \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1} \\
&= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \times \left[\mathbf{ZP} \times \left\{\mathbf{D}^{-1} + E(\widehat{W}\hat{\alpha}\hat{\alpha}^t\widehat{W}) - E\{\widehat{W}\hat{\alpha}(\hat{\alpha} - \alpha)^t\} - E\{(\hat{\alpha} - \alpha)\hat{\alpha}^t\widehat{W}\}\right\} \right. \\
&\quad \left. \times \mathbf{P}^t\mathbf{Z}^t + \text{cov}(\mathbf{y}) - \mathbf{ZP} \times E(\hat{Q}\hat{\alpha}\varepsilon^t) - E(\varepsilon\hat{\alpha}^t\hat{Q}) \times \mathbf{P}^t\mathbf{Z}^t\right] \times \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}.
\end{aligned}$$

From Stein's Lemma, two equalities are obtained as follows.

$$\begin{aligned}
E[\widehat{W}\hat{\alpha}(\hat{\alpha} - \alpha)^t] &= E\begin{bmatrix} \hat{w}_1\hat{\alpha}_1(\hat{\alpha}_1 - \alpha_1) & \cdots & \hat{w}_1\hat{\alpha}_1(\hat{\alpha}_m - \alpha_m) \\ \vdots & \ddots & \vdots \\ \hat{w}_m\hat{\alpha}_m(\hat{\alpha}_1 - \alpha_1) & \vdots & \hat{w}_m\hat{\alpha}_m(\hat{\alpha}_m - \alpha_m) \end{bmatrix} \\
&= E\begin{bmatrix} d_1^{-1}\left(\frac{\partial \hat{w}_1}{\partial \hat{\alpha}_1}\hat{\alpha}_1 + \hat{w}_1\right) & \cdots & d_m^{-1}\frac{\partial \hat{w}_1}{\partial \hat{\alpha}_m}\hat{\alpha}_1 \\ \vdots & \ddots & \vdots \\ d_1^{-1}\frac{\partial \hat{w}_m}{\partial \hat{\alpha}_1}\hat{\alpha}_m & \cdots & d_m^{-1}\left(\frac{\partial \hat{w}_m}{\partial \hat{\alpha}_m}\hat{\alpha}_m + \hat{w}_m\right) \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
&= E \left[\text{diag}(\hat{\alpha}) \frac{\partial \hat{\mathbf{W}}}{\partial \hat{\alpha}^t} + \hat{\mathbf{W}} \right] \times \mathbf{D}^{-1}. \\
E(\hat{\mathbf{Q}} \hat{\alpha} \boldsymbol{\varepsilon}^t) &= E \begin{bmatrix} \hat{q}_1 \hat{\alpha}_1 \varepsilon_1 & \cdots & \hat{q}_1 \hat{\alpha}_1 \varepsilon_n \\ \vdots & \ddots & \vdots \\ \hat{q}_m \hat{\alpha}_m \varepsilon_1 & \cdots & \hat{q}_m \hat{\alpha}_m \varepsilon_n \end{bmatrix} \\
&= E \begin{bmatrix} \sigma_1^2 \left(\frac{\partial \hat{q}_1}{\partial \hat{\alpha}^t} \frac{\partial \hat{\alpha}}{\partial \varepsilon_1} \hat{\alpha}_1 + \hat{q}_1 \frac{\partial \hat{\alpha}_1}{\partial \varepsilon_1} \right) & \cdots & \sigma_n^2 \left(\frac{\partial \hat{q}_1}{\partial \hat{\alpha}^t} \frac{\partial \hat{\alpha}}{\partial \varepsilon_n} \hat{\alpha}_1 + \hat{q}_1 \frac{\partial \hat{\alpha}_1}{\partial \varepsilon_n} \right) \\ \vdots & \ddots & \vdots \\ \sigma_1^2 \left(\frac{\partial \hat{q}_m}{\partial \hat{\alpha}^t} \frac{\partial \hat{\alpha}}{\partial \varepsilon_1} \hat{\alpha}_m + \hat{q}_m \frac{\partial \hat{\alpha}_m}{\partial \varepsilon_1} \right) & \cdots & \sigma_n^2 \left(\frac{\partial \hat{q}_m}{\partial \hat{\alpha}^t} \frac{\partial \hat{\alpha}}{\partial \varepsilon_n} \hat{\alpha}_m + \hat{q}_m \frac{\partial \hat{\alpha}_m}{\partial \varepsilon_n} \right) \end{bmatrix} \\
&= E \left[\text{diag}(\hat{\alpha}) \frac{\partial \hat{\mathbf{Q}}}{\partial \hat{\alpha}^t} \frac{\partial \hat{\alpha}}{\partial \boldsymbol{\varepsilon}^t} \text{cov}(\mathbf{y}) + \hat{\mathbf{Q}} \frac{\partial \hat{\alpha}}{\partial \boldsymbol{\varepsilon}^t} \text{cov}(\mathbf{y}) \right] \\
&= E \left[\mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \times \text{cov}(\mathbf{y}) \right. \\
&\quad \left. - \left(\text{diag}(\hat{\alpha}) \frac{\partial \hat{\mathbf{W}}}{\partial \hat{\alpha}^t} + \hat{\mathbf{W}} \right) \mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \times \text{cov}(\mathbf{y}) \right].
\end{aligned}$$

Then,

$$\begin{aligned}
MSE[\hat{\boldsymbol{\beta}}^*] &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \\
&\quad \times \left[\mathbf{Z} \mathbf{P} \left\{ \mathbf{D}^{-1} + E(\hat{\mathbf{W}} \hat{\alpha} \hat{\alpha}^t \hat{\mathbf{W}}) - E \left\{ \text{diag}(\hat{\alpha}) \frac{\partial \hat{\mathbf{W}}}{\partial \hat{\alpha}^t} + \hat{\mathbf{W}} \right\} \times \mathbf{D}^{-1} \right. \right. \\
&\quad \left. \left. - \mathbf{D}^{-1} \times E \left\{ \left(\frac{\partial \hat{\mathbf{W}}}{\partial \hat{\alpha}^t} \right)^t \text{diag}(\hat{\alpha}) + \hat{\mathbf{W}} \right\} \right\} \mathbf{P}^t \mathbf{Z}^t + \text{cov}(\mathbf{y}) \right. \\
&\quad \left. - \mathbf{Z} \mathbf{P} E \left\{ \mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \times \text{cov}(\mathbf{y}) \right. \right. \\
&\quad \left. \left. - \left(\text{diag}(\hat{\alpha}) \frac{\partial \hat{\mathbf{W}}}{\partial \hat{\alpha}^t} + \hat{\mathbf{W}} \right) \mathbf{P}^t (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{M} \times \text{cov}(\mathbf{y}) \right\} \right. \\
&\quad \left. - E \left\{ \text{cov}(\mathbf{y}) \times \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} \right. \right. \\
&\quad \left. \left. - \text{cov}(\mathbf{y}) \times \mathbf{M} \mathbf{Z} (\mathbf{Z}^t \mathbf{M} \mathbf{Z})^{-1} \mathbf{P} \left(\left(\frac{\partial \hat{\mathbf{W}}}{\partial \hat{\alpha}^t} \right)^t \text{diag}(\hat{\alpha}) + \hat{\mathbf{W}} \right) \right\} \mathbf{P}^t \mathbf{Z}^t \right] \\
&\quad \times \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1}
\end{aligned}$$

$$\begin{aligned}
&= (X^t X)^{-1} X^t \times E \left[Z P \widehat{W} \widehat{\alpha} \widehat{\alpha}^t \widehat{W} P^t Z^t \right. \\
&\quad + (I - MZ(Z^t MZ)^{-1} Z^t)^t \times cov(y) \times (I - MZ(Z^t MZ)^{-1} Z^t) \\
&\quad + ZP \left(diag(\widehat{\alpha}) \frac{\partial \widehat{W}}{\partial \widehat{\alpha}^t} + \widehat{W} \right) P^t (Z^t MZ)^{-1} Z^t M \times cov(y) \\
&\quad \times (I - MZ(Z^t MZ)^{-1} Z^t) \\
&\quad + (I - MZ(Z^t MZ)^{-1} Z^t)^t \times cov(y) \\
&\quad \left. \times MZ(Z^t MZ)^{-1} P \left(\left(\frac{\partial \widehat{W}}{\partial \widehat{\alpha}^t} \right)^t diag(\widehat{\alpha}) + \widehat{W} \right) P^t Z^t \right] \times X(X^t X)^{-1}.
\end{aligned}$$

Therefore, the MSE of the proposed estimator is estimated by replacing $cov(y)$ and related PDP^t using its estimates as follows:

$$\begin{aligned}
\widehat{MSE}[\widehat{\beta}^*] &= (X^t X)^{-1} X^t \\
&\quad \times \left[Z \widehat{P} \widehat{W} \widehat{\alpha} \widehat{\alpha}^t \widehat{W} \widehat{P}^t Z^t \right. \\
&\quad + (I - MZ(Z^t MZ)^{-1} Z^t)^t \times \widehat{cov}(y) \times (I - MZ(Z^t MZ)^{-1} Z^t) \\
&\quad + Z \widehat{P} \left(diag(\widehat{\alpha}) \frac{\partial \widehat{W}}{\partial \widehat{\alpha}^t} + \widehat{W} \right) \widehat{P}^t (Z^t MZ)^{-1} Z^t M \times \widehat{cov}(y) \\
&\quad \times (I - MZ(Z^t MZ)^{-1} Z^t) \\
&\quad + (I - MZ(Z^t MZ)^{-1} Z^t)^t \times \widehat{cov}(y) \\
&\quad \left. \times MZ(Z^t MZ)^{-1} \widehat{P} \left(\left(\frac{\partial \widehat{W}}{\partial \widehat{\alpha}^t} \right)^t diag(\widehat{\alpha}) + \widehat{W} \right) \widehat{P}^t Z^t \right] \times X(X^t X)^{-1}
\end{aligned}$$

10.3. Appendix C: Karush–Kuhn–Tucker (KKT) condition in determining the tuning parameters

The precision measure can be written as

$$G(\Lambda) = w^t S_1 (2\widehat{D}^{-1} + \widehat{\alpha} \widehat{\alpha}^t) S_1 w - 2w^t (S_1 \widehat{D}^{-1} s_1 - S_1 s_2) + C,$$

where $\mathbf{w} = (w_1, \dots, w_m)^t$ and C is a constant independent of $\mathbf{w}$. Thus, our optimization problem is

$$\text{minimize } G(\mathbf{\Lambda}) \text{ subject to } \forall j \in \{1, \dots, m\}, w_j(w_j - 1) \leq 0,$$

and the KKT condition is given as follows:

$$\left. \frac{\partial G(\mathbf{\Lambda})}{\partial \mathbf{w}} \right|_{\mathbf{w}=\hat{\mathbf{w}}} + \sum_{j=1}^m \mu_j \left. \frac{\partial \{w_j(w_j - 1)\}}{\partial \mathbf{w}} \right|_{\mathbf{w}=\hat{\mathbf{w}}} = 0 \quad (\text{A1.1})$$

$$\mu_j \geq 0 \quad (\text{A1.2})$$

$$\hat{w}_j(\hat{w}_j - 1) \leq 0, \text{ and} \quad (\text{A1.3})$$

$$\mu_j \hat{w}_j(\hat{w}_j - 1) = 0. \quad (\text{A1.4})$$

Lemma 1. The hessian of $G(\mathbf{\Lambda})$ is positive definite.

Proof of Lemma 1. The hessian of $G(\mathbf{\Lambda})$ is given as

$$\mathbf{S}_1(2\hat{\mathbf{D}}^{-1} + \hat{\mathbf{\alpha}}\hat{\mathbf{\alpha}}^t)\mathbf{S}_1.$$

$\hat{\mathbf{D}}^{-1}$ is the diagonal matrix of positive values, and $\hat{\mathbf{\alpha}}\hat{\mathbf{\alpha}}^t$ is semi-positive definite. The sum of the semi-positive and positive-definite matrices is positive definite. Thus,

$2\hat{\mathbf{D}}^{-1} + \hat{\mathbf{\alpha}}\hat{\mathbf{\alpha}}^t$ is positive definite. Furthermore, for arbitrary $\mathbf{x}$,

$$\mathbf{x}^t \mathbf{S}_1(2\hat{\mathbf{D}}^{-1} + \hat{\mathbf{\alpha}}\hat{\mathbf{\alpha}}^t)\mathbf{S}_1 \mathbf{x} = (\mathbf{S}_1 \mathbf{x})^t (2\hat{\mathbf{D}}^{-1} + \hat{\mathbf{\alpha}}\hat{\mathbf{\alpha}}^t) \mathbf{S}_1 \mathbf{x} > 0.$$

Theorem 1. Kuhn–Tucker conditions are satisfied with the procedure.

Proof of Theorem 1.

Let

$$J = \{1, \dots, m\},$$

$$J_0 = \{j \in J; \hat{w}_j = 0\},$$

$$J_1 = \{j \in J; \hat{w}_j = 1\}, \text{ and}$$

$$J_c = J \setminus (J_0 \cup J_1)$$

through the procedure. Note that $\hat{\mathbf{w}}$ is satisfied with $\hat{w}_j(\hat{w}_j - 1) \leq 0$. Suppose that the order of the elements of $\mathbf{w}$ is such that $w_j = 0$ or 1 is the last. That is,

$$\hat{\mathbf{w}} = (\hat{\mathbf{w}}_{J_c}^t, \mathbf{0}_{J_0}^t, \mathbf{1}_{J_1}^t)^t.$$

Thus, the reduced system of the algorithm is given below.

$$G(\Lambda)$$

$$= C$$

$$\begin{aligned} & + (\mathbf{w}_{J_c}^t, \mathbf{0}_{J_0}^t, \mathbf{1}_{J_1}^t)^t \begin{pmatrix} \mathbf{S}_{1J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{1J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{1J_1} \end{pmatrix} \begin{pmatrix} 2\hat{\mathbf{D}}_{J_c}^{-1} + \hat{\alpha}_{J_c}\hat{\alpha}_{J_c}^t & \hat{\alpha}_{J_c}\hat{\alpha}_{J_0}^t & \hat{\alpha}_{J_c}\hat{\alpha}_{J_1}^t \\ \hat{\alpha}_{J_0}\hat{\alpha}_{J_c}^t & 2\hat{\mathbf{D}}_{J_0}^{-1} + \hat{\alpha}_{J_0}\hat{\alpha}_{J_0}^t & \hat{\alpha}_{J_0}\hat{\alpha}_{J_1}^t \\ \hat{\alpha}_{J_1}\hat{\alpha}_{J_c}^t & \hat{\alpha}_{J_1}\hat{\alpha}_{J_0}^t & 2\hat{\mathbf{D}}_{J_1}^{-1} + \hat{\alpha}_{J_1}\hat{\alpha}_{J_1}^t \end{pmatrix} \\ & \times \begin{pmatrix} \mathbf{S}_{1J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{1J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{1J_1} \end{pmatrix} \begin{pmatrix} \mathbf{w}_{J_c} \\ \mathbf{0}_{J_0} \\ \mathbf{1}_{J_1} \end{pmatrix} \\ & - 2(\mathbf{w}_{J_c}^t, \mathbf{0}_{J_0}^t, \mathbf{1}_{J_1}^t) \begin{pmatrix} \mathbf{S}_{1J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{1J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{1J_1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{D}}_{J_c}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{D}}_{J_0}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \hat{\mathbf{D}}_{J_1}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{1J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{1J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{1J_1} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{J_c} \\ \mathbf{1}_{J_0} \\ \mathbf{1}_{J_1} \end{pmatrix} \\ & + 2(\mathbf{w}_{J_c}^t, \mathbf{0}_{J_0}^t, \mathbf{1}_{J_1}^t) \begin{pmatrix} \mathbf{S}_{1J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{1J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{1J_1} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{2J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{2J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{2J_1} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{J_c} \\ \mathbf{1}_{J_0} \\ \mathbf{1}_{J_1} \end{pmatrix} \\ & = C + \mathbf{w}_{J_c}^t \mathbf{S}_{1J_c} (2\hat{\mathbf{D}}_{J_c}^{-1} + \hat{\alpha}_{J_c}\hat{\alpha}_{J_c}^t) \mathbf{S}_{1J_c} \mathbf{w}_{J_c} + \mathbf{1}_{J_1}^t \mathbf{S}_{1J_1} \hat{\alpha}_{J_1} \hat{\alpha}_{J_c}^t \mathbf{S}_{1J_c} \mathbf{w}_{J_c} \\ & \quad + \mathbf{w}_{J_c}^t \mathbf{S}_{1J_c} \hat{\alpha}_{J_c} \hat{\alpha}_{J_1}^t \mathbf{S}_{1J_1} \mathbf{1}_{J_1} + \mathbf{1}_{J_1}^t \mathbf{S}_{1J_1} (2\hat{\mathbf{D}}_{J_1}^{-1} + \hat{\alpha}_{J_1}\hat{\alpha}_{J_1}^t) \mathbf{S}_{1J_1} \mathbf{1}_{J_1} \\ & \quad - 2(\mathbf{w}_{J_c}^t \mathbf{S}_{1J_c} \hat{\mathbf{D}}_{J_c}^{-1} \mathbf{S}_{1J_c} \mathbf{1}_{J_c} + \mathbf{1}_{J_1}^t \mathbf{S}_{1J_1} \hat{\mathbf{D}}_{J_1}^{-1} \mathbf{S}_{1J_1} \mathbf{1}_{J_1}) \\ & \quad + 2(\mathbf{w}_{J_c}^t \mathbf{S}_{1J_c} \mathbf{S}_{2J_c} \mathbf{1}_{J_c} + \mathbf{1}_{J_1}^t \mathbf{S}_{1J_1} \mathbf{S}_{2J_1} \mathbf{1}_{J_1}) \\ & = C^* + \mathbf{w}_{J_c}^t \mathbf{S}_{1J_c} (2\hat{\mathbf{D}}_{J_c}^{-1} + \hat{\alpha}_{J_c}\hat{\alpha}_{J_c}^t) \mathbf{S}_{1J_c} \mathbf{w}_{J_c} \\ & \quad - 2\mathbf{w}_{J_c}^t \mathbf{S}_{1J_c} (\hat{\mathbf{D}}_{J_c}^{-1} \mathbf{S}_{1J_c} \mathbf{1}_{J_c} - \mathbf{S}_{2J_c} \mathbf{1}_{J_c} - \hat{\alpha}_{J_c} \hat{\alpha}_{J_1}^t \mathbf{S}_{1J_1} \mathbf{1}_{J_1}), \end{aligned}$$

where:

$$\mathbf{S}_1 = \begin{pmatrix} \mathbf{S}_{1J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{1J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{1J_1} \end{pmatrix},$$

$$\mathbf{S}_2 = \begin{pmatrix} \mathbf{S}_{2J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{2J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{2J_1} \end{pmatrix}$$

$$\hat{\alpha} = (\hat{\alpha}_{J_c}^t, \hat{\alpha}_{J_0}^t, \hat{\alpha}_{J_1}^t)^t$$

$$\hat{\mathbf{D}} = \begin{pmatrix} \hat{\mathbf{D}}_{J_c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{D}}_{J_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \hat{\mathbf{D}}_{J_1} \end{pmatrix}, \text{ and}$$

C^* is constant independent of $\mathbf{w}$.

Given this reduced system, the solution of w_{J_c} is given as follows:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}_{J_c}} G(\Lambda) &= 2\mathbf{S}_{1J_c}(2\hat{\mathbf{D}}_{J_c}^{-1} + \hat{\alpha}_{J_c}\hat{\alpha}_{J_c}^t)\mathbf{S}_{1J_c}\mathbf{w}_{J_c} \\ &\quad - 2\mathbf{S}_{1J_c}(\hat{\mathbf{D}}_{J_c}^{-1}\mathbf{S}_{1J_c}\mathbf{1}_{J_c} - \mathbf{S}_{2J_c}\mathbf{1}_{J_c} - \hat{\alpha}_{J_c}\hat{\alpha}_{J_1}^t\mathbf{S}_{1J_1}\mathbf{1}_{J_1}) = 0 \end{aligned}$$

$$\hat{\mathbf{w}}_{J_c} = \mathbf{S}_{1J_c}^{-1}(2\hat{\mathbf{D}}_{J_c}^{-1} + \hat{\alpha}_{J_c}\hat{\alpha}_{J_c}^t)^{-1}(\hat{\mathbf{D}}_{J_c}^{-1}\mathbf{S}_{1J_c}\mathbf{1}_{J_c} - \mathbf{S}_{2J_c}\mathbf{1}_{J_c} - \hat{\alpha}_{J_c}\hat{\alpha}_{J_1}^t\mathbf{S}_{1J_1}\mathbf{1}_{J_1})$$

Note that $2\hat{\mathbf{D}}_{J_c}^{-1} + \hat{\alpha}_{J_c}\hat{\alpha}_{J_c}^t$ has a unique inverse matrix since this is positive definite.

For $j \in J \setminus (J_0 \cup J_1)$, $\{\partial/\partial w_j\}G(\hat{\Lambda}) = 0$. Therefore, $\mu_j (= 0)$ is satisfied with

(A1.1), (A1.2) and (A1.4). Secondly, for $j \in J_0$, the global unconstrained minimum of the convex function $G(\Lambda)$ is outside $w_j \geq 0$ and $\{\partial G(\Lambda)/\partial w_j\}|_{w_j=\hat{w}_j} > 0$. In addition,

$$\left\{\partial(w_j(w_j - 1))/\partial w_j\right\}|_{w_j=\hat{w}_j} = 2\hat{w}_j - 1 = -1 < 0. \text{ Therefore,}$$

$$\frac{\partial}{\partial w_j} G(\hat{\Lambda}) + \mu_j(2\hat{w}_j - 1) = 0 \Rightarrow \mu_j = \frac{\partial}{\partial w_j} G(\hat{\Lambda}) > 0,$$

and this μ_j is satisfied with (A1.1), (A1.2) and (A1.4). Likewise, for $j \in J_1$, the KKT conditions are satisfied.

10.4. Appendix D: Consistency in the proposed estimator

Theorem 2. If the least squares estimator $\mathbf{c}^t \hat{\boldsymbol{\beta}}$ converges to $\mathbf{c}^t \boldsymbol{\beta}$ in L^2 and $n \times \hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}] \xrightarrow{p} n \times V[\mathbf{c}^t \hat{\boldsymbol{\beta}}]$, then the proposed estimator $\mathbf{c}^t \hat{\boldsymbol{\beta}}^*$ converges to $\mathbf{c}^t \boldsymbol{\beta}$ in probability.

Proof of Theorem 2. Consider the distance between $\mathbf{c}^t \hat{\boldsymbol{\beta}}^*$ and $\mathbf{c}^t \hat{\boldsymbol{\beta}}$:

$$|\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}|^2 = \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Z} \mathbf{P} \widehat{\mathbf{W}} \hat{\boldsymbol{\alpha}} \hat{\boldsymbol{\alpha}}^t \widehat{\mathbf{W}} \mathbf{P}^t \mathbf{Z}^t \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c} \leq G(\hat{\boldsymbol{\Lambda}}),$$

where $\hat{\boldsymbol{\Lambda}}$ denotes the optimal values that minimize the precision measure subjects to $\forall j = 1, \dots, m, w_j \in [0,1]$. Let $\boldsymbol{\Lambda}_0$ be a $(m \times m)$ matrix that has zero in all the elements.

Then, the precision measure with $\boldsymbol{\Lambda}_0$ is equal to the variance estimate of $\mathbf{c}^t \hat{\boldsymbol{\beta}}$ and,

$$|\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}|^2 \leq G(\hat{\boldsymbol{\Lambda}}) \leq G(\boldsymbol{\Lambda}_0) = \hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}].$$

Therefore, for every $\varepsilon_1 > 0$, there exists $n_1 \in \mathbb{N}$ such that $n \geq n_1 \Rightarrow \varepsilon_1 > \delta_1 =$

$\sqrt{V[\mathbf{c}^t \hat{\boldsymbol{\beta}}]}$ and,

$$\begin{aligned} \Pr(|\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}| > \varepsilon_1) &\leq \Pr(\hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}] > \varepsilon_1^2) \leq \Pr(|\hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}] - V[\mathbf{c}^t \hat{\boldsymbol{\beta}}]| + V[\mathbf{c}^t \hat{\boldsymbol{\beta}}] > \varepsilon_1^2) \\ &\leq \Pr\left(n|\hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}] - V[\mathbf{c}^t \hat{\boldsymbol{\beta}}]| > n(\varepsilon_1^2 - \delta_1^2)\right) \\ &\rightarrow 0 \left(\because n \times \hat{V}[\mathbf{c}^t \hat{\boldsymbol{\beta}}] \xrightarrow{p} n \times V[\mathbf{c}^t \hat{\boldsymbol{\beta}}]\right). \end{aligned}$$

This indicates that the proposed estimator converges to the least squares estimator in probability. In addition, for every $\varepsilon > 0$, there exists $n_2 \in \mathbb{N}$ such that $n \geq n_2 \geq n_1 \Rightarrow \varepsilon > \varepsilon_1 > \delta_1$ and,

$$\begin{aligned} \Pr(|\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \boldsymbol{\beta}| > \varepsilon) &\leq \Pr(|\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}| + |\mathbf{c}^t \hat{\boldsymbol{\beta}} - \mathbf{c}^t \boldsymbol{\beta}| > \varepsilon) \\ &= \Pr(|\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}| + |\mathbf{c}^t \hat{\boldsymbol{\beta}} - \mathbf{c}^t \boldsymbol{\beta}| > \varepsilon, |\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}| \leq \varepsilon_1) \\ &\quad + \Pr(|\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}| + |\mathbf{c}^t \hat{\boldsymbol{\beta}} - \mathbf{c}^t \boldsymbol{\beta}| > \varepsilon, |\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}| > \varepsilon_1) \\ &\leq \Pr(|\mathbf{c}^t \hat{\boldsymbol{\beta}} - \mathbf{c}^t \boldsymbol{\beta}| > \varepsilon - \varepsilon_1) + \Pr(|\mathbf{c}^t \hat{\boldsymbol{\beta}}^* - \mathbf{c}^t \hat{\boldsymbol{\beta}}| > \varepsilon_1) \rightarrow 0. \end{aligned}$$

The last convergence is based on the assumption that the least squares estimator converges to the true value in probability.

11. ORCID

Akinori Nimura <https://orcid.org/0009-0004-4533-5928>

Tosiya Sato <https://orcid.org/0000-0001-9830-013X>

12. References

- Barker, A. D., Sigman, C. C., Kelloff, G. J., Hylton, N. M., Berry, D. A., and Esserman, L. J. (2009), "I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy," *Clinical Pharmacology and Therapeutics*, 86 (1), 97-100. DOI: 10.1038/clpt.2009.68.
- Bofill Roig, M., Krotka, P., Burman, C.-F., Glimm, E., Gold, S. M., Hees, K., et al.. (2022), "On model-based time trend adjustments in platform trials with non-concurrent controls," *BMC Medical Research Methodology*, 22 (1), 228. DOI: 10.1186/s12874-022-01683-w.
- Chatterjee, S., and Kemmer, P. (2022), "Learnings from chronic pain master protocol: FDA complex innovative designs pilot program," [online], 2024 (29 July). Available at: http://www.bayesianscientific.org/wp-content/uploads/2022/10/DIA-BSWG-KOL-CPMP-21OCT2022_final.pdf.
- Chen, M.-H., and Ibrahim, J. G. (2000), "Power prior distributions for regression models," *Statistical Science*, 15 (1), 46-60, 15. DOI: 10.1214/ss/1009212673.
- Efron, B., and Tibshirani, R. J. (1993), *An introduction to the bootstrap* (Vol. 57), New York: Chapman & Hall.
- FDA (2022), "Master protocols: efficient clinical trial design strategies to expedite development of oncology drugs and biologics. Guidance for industry," [online], 2024 (29 July). Available at: <https://www.fda.gov/media/120721/download>.
- Guo, B., Wang, L., and Yuan, Y. (2024), "Treatment Comparisons in Adaptive Platform Trials Adjusting for Temporal Drift," *Statistics in Biopharmaceutical Research*. DOI: 10.1080/19466315.2023.2292238.

- Han, Z., Ye, K., and Wang, M. (2023), "A study on the power parameter in power prior Bayesian analysis," *The American Statistician*, 77 (1), 12-19. DOI: 10.1080/00031305.2022.2046160.
- Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011), "Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials," *Biometrics*, 67 (3), 1047-1056. DOI: 10.1111/j.1541-0420.2011.01564.x.
- Hoerl, A. E., and Kennard, R. W. (1970), "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, 12 (1), 55-67. DOI: 10.1080/00401706.1970.10488634.
- Ibrahim, J. G., Chen, M. H., Gwon, Y., and Chen, F. (2015), "The power prior: theory and applications," *Statistics in Medicine*, 34 (28), 3724-3749. DOI: 10.1002/sim.6728.
- Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., and Jones, D. R. (2005), "How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS," *Statistics in Medicine*, 24 (15), 2401-2428. DOI: 10.1002/sim.2112.
- Marschner, I. C., and Schou, I. M. (2022), "Analysis of adaptive platform trials using a network approach," *Clinical Trials*, 19 (5), 479-489. DOI: 10.1177/17407745221112001.
- Plante, J.-F. (2008), "Nonparametric adaptive likelihood weights," *Canadian Journal of Statistics*, 36, 443-461. DOI: 10.1002/cjs.5550360308.
- Price, D., and Scott, J. (2021), "The U.S. Food and Drug Administration's complex innovative trial design pilot meeting program: progress to date," *Clinical Trials*, 18 (6), 706-710. DOI: 10.1177/17407745211050580.
- Saville, B. R., Berry, D. A., Berry, N. S., Viele, K., and Berry, S. M. (2022), "The Bayesian Time Machine: Accounting for temporal drift in multi-arm platform trials," *Clinical Trials*, 19 (5), 490-501. DOI: 10.1177/17407745221112013.
- Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., and Neuenschwander, B. (2014), "Robust meta-analytic-predictive priors in clinical trials with historical control information," *Biometrics*, 70 (4), 1023-1032. DOI: 10.1111/biom.12242.

- Searle, S. R., Casella, G., and McCulloch, C. E. (1992), "Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML)," in *Variance Components*, eds. S. R. Searle, G. Casella and C. E. McCulloch, New York: Wiley, pp. 232-257.
- Stein, C. M. (1981), "Estimation of the mean of a multivariate normal distribution," *Annals of Statistics*, 9 (6), 1135-1151. DOI: 10.1214/aos/1176345632.
- White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48 (4), 817-838. DOI: 10.2307/1912934.
- Woodcock, J., and LaVange, L. M. (2017), "Master protocols to study multiple therapies, multiple diseases, or both," *New England Journal of Medicine*, 377 (1), 62-70. DOI: 10.1056/NEJMra1510062.

Table 1. Study results in Chronic Pain Master Protocol as of June 2024

NCT number	Treatment arm	Week	n	Posterior mean	95% credible interval	
					Lower	Upper
Chronic low back pain (CLBP)						
NCT04529096	Placebo	8	41	-2.09	-2.60	-1.59
NCT04874636	Placebo	8	38	-1.77	-2.31	-1.22
NCT05086289	LY3526318	8	87	-1.47	-1.82	-1.12
NCT05086289	Placebo	8	47	-1.20	-1.67	-0.73
Diabetic peripheral neuropathic pain (DPNP)						
NCT04476108	Placebo	8	38	-1.56	-2.17	-0.96
NCT04707157	Placebo	8	14	-1.19	-2.14	-0.24
NCT05177094	LY3526318	8	85	-1.80	-2.17	-1.42
NCT05177094	Placebo	8	38	-1.78	-2.32	-1.24
Osteoarthritis (OA)						
NCT04456686	Placebo	8	31	-2.00	-2.61	-1.39
NCT04627038	Placebo	8	46	-1.76	-2.23	-1.29
NCT05080660	LY3526318	8	90	-1.26	-1.60	-0.93
NCT05080660	Placebo	8	51	-1.45	-1.92	-0.99

Table 2. Application results in DPNP substudies

Estimator	Treatment difference (95%CI or CrI)	Difference in control from substudies for borrowing
NCT04476108 & NCT04707157		
No borrow	-0.02 (-0.68, 0.64)	0.22 & 0.59
Ridge (Sandwich)	-0.09 (-0.62, 0.43)	0.13 & 0.35
Ridge (REML)	-0.10 (-0.63, 0.43)	0.13 & 0.35
Full borrow	-0.20 (-0.73, 0.32)	NA
Power prior	-0.13 (-0.68, 0.42)	NA
HBM (s_tau=0.5)	-0.14 (-0.73, 0.47)	0.06 & 0.14
HBM (s_tau=1)	-0.12 (-0.67, 0.44)	0.08 & 0.18
NCT04476108		
No borrow	-0.02 (-0.68, 0.64)	0.22
Ridge (Sandwich)	-0.06 (-0.61, 0.49)	0.13
Ridge (REML)	-0.06 (-0.62, 0.49)	0.13
Full borrow	-0.13 (-0.68, 0.42)	NA
Power prior	-0.13 (-0.68, 0.42)	NA
HBM (s_tau=0.5)	-0.06 (-0.63, 0.44)	0.07
HBM (s_tau=1)	-0.07 (-0.67, 0.55)	0.12
NCT04707157		
No borrow	-0.02 (-0.68, 0.64)	0.59

Estimator	Treatment difference (95%CI or CrI)	Difference in control from substudies for borrowing
Ridge (Sandwich)	-0.07 (-0.66, 0.53)	0.42
Ridge (REML)	-0.07 (-0.68, 0.54)	0.41
Full borrow	-0.18 (-0.78, 0.42)	NA
Power prior	-0.16 (-0.77, 0.46)	NA
HBM (s_tau=0.5)	-0.10 (-0.69, 0.52)	0.18
HBM (s_tau=1)	-0.10 (-0.70, 0.56)	0.30

HBM, Hierarchical Bayesian model; CI, confidence interval; CrI, credible interval; REML, restricted maximum likelihood.

Table 3. Application results in OA substudies

Estimator	Treatment difference (95%CI or CrI)	Difference in control from substudies for borrowing
NCT04456686 & NCT04627038		
No borrow	0.19 (-0.38, 0.76)	-0.55 & -0.31
Ridge (Sandwich)	0.25 (-0.20, 0.69)	-0.40 & -0.26
Ridge (REML)	0.24 (-0.20, 0.69)	-0.41 & -0.26
Full borrow	0.43 (-0.01, 0.88)	NA
Power prior	0.34 (-0.13, 0.81)	NA
HBM (s_tau=0.5)	0.35 (-0.14, 0.89)	-0.20 & -0.12
HBM (s_tau=1)	0.30 (-0.28, 0.83)	-0.28 & -0.18
NCT04456686		
No borrow	0.19 (-0.38, 0.76)	-0.55
Ridge (Sandwich)	0.24 (-0.34, 0.82)	-0.42
Ridge (REML)	0.24 (-0.34, 0.82)	-0.42
Full borrow	0.40 (-0.10, 0.90)	NA
Power prior	0.29 (-0.24, 0.82)	NA
HBM (s_tau=0.5)	0.32 (-0.24, 0.83)	-0.21
HBM (s_tau=1)	0.26 (-0.30, 0.84)	-0.39
NCT04627038		
No borrow	0.19 (-0.38, 0.76)	-0.31

Estimator	Treatment difference (95%CI or CrI)	Difference in control from substudies for borrowing
Ridge (Sandwich)	0.24 (-0.23, 0.71)	-0.20
Ridge (REML)	0.24 (-0.23, 0.72)	-0.20
Full borrow	0.34 (-0.13, 0.81)	NA
Power prior	0.34 (-0.13, 0.81)	NA
HBM (s_tau=0.5)	0.26 (-0.28, 0.80)	-0.15
HBM (s_tau=1)	0.29 (-0.28, 0.85)	-0.17

HBM, Hierarchical Bayesian model; CI, confidence interval; CrI, credible interval; REML, restricted maximum likelihood.

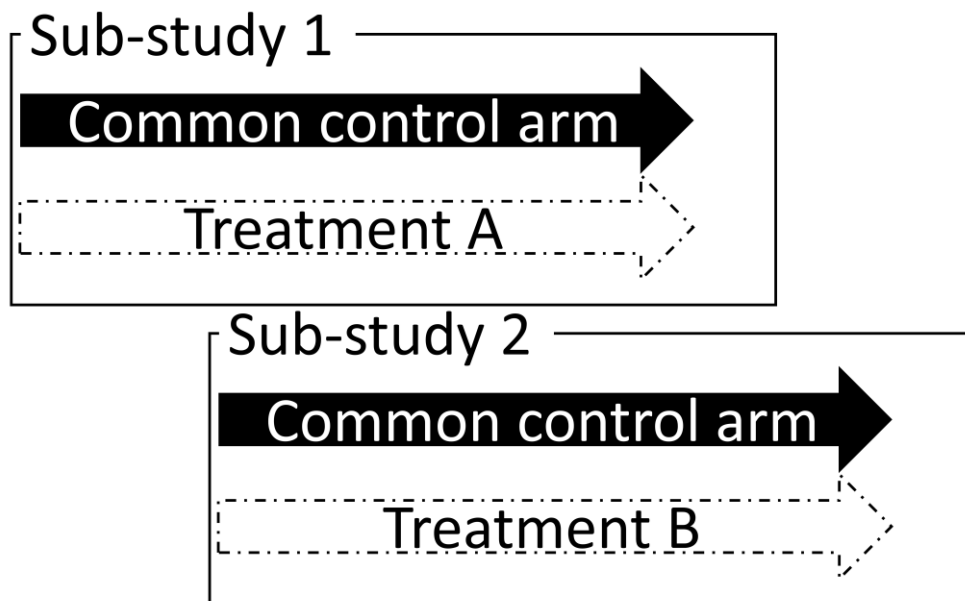


Figure 1. Study Schematic in a master protocol with independent substudies

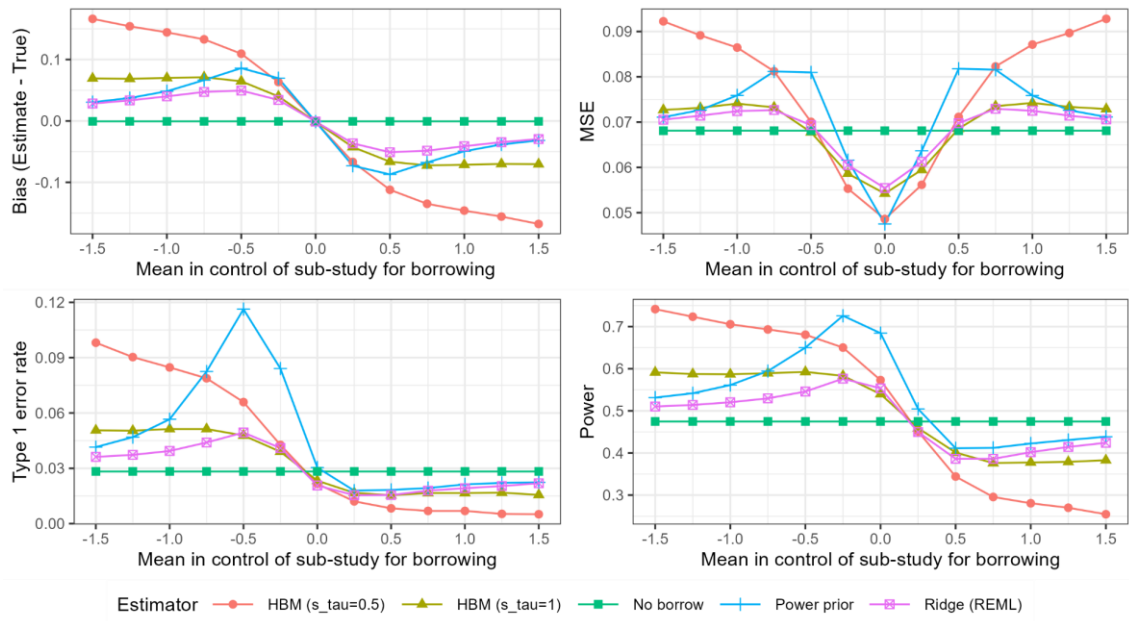


Figure 2. Simulation results where the sample size is 100 and the standard deviation is 1.5 in the substudy for borrowing. HBM: Hierarchical Bayesian model.

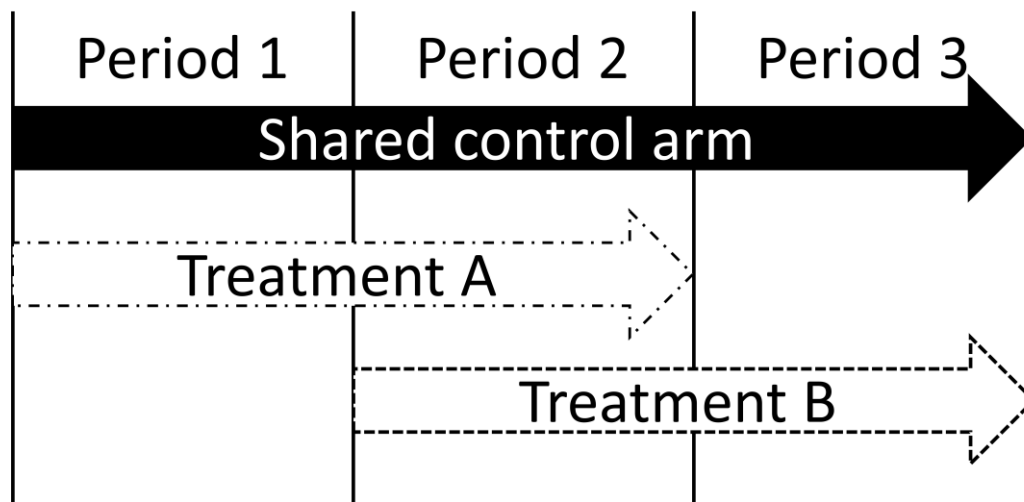


Figure 3. Study Schematic in a master protocol where study treatments are sequentially introduced with the shared control arm