筑　波　大　学

博 士（ 医 学 ）学 位 論 文

Bayesian method with horseshoe prior for incorporating multiple historical control data in randomized trials

(ランダム化比較試験で
既存対照データを組み込むための
馬蹄事前分布を用いたベイズ推定法)

2022

筑波大学大学院博士課程人間総合科学研究科
大東 智洋

Contents

Abbreviation

CI	Credible interval
CP	Calibrated power
DMPP	Dependent modified power prior
EBPP	Empirical Bayesian power prior
EHSS	Effective historical sample size
EMA	European Medicines Agency
ESS	Effective sample size
EX	Full exchangeability meta-analytic combined method
EXNEX	Robustified meta-analytic combined method
FDA	Food and Drug Administration
HS	Proposed method using horseshoe prior
ICH	International Conference on Harmonization
MAC	Meta-analytic combined
MAP	Meta-analytic predictive
MCMC	Markov chain Monte Carlo
MPP	Modified power prior
MPSD	Mean posterior standard deviation
PFS	Progression-free survival
RCT	Randomized controlled trial
RDMPP	Robust dependent modified power prior
RMSD	Root mean square deviation
SD	Standard deviation

List of Tables

List of Figures

Chapter 1

Background

1.1 Historical data in clinical trials

A randomized controlled trial (RCT) with placebo or standard care is the gold standard for determining the efficacy of a test treatment. RCTs contribute to the generation of unbiased treatment effect estimates and control for type I error in hypotheses testing. For medical product development in major disease areas, it is relatively easy to conduct RCTs with appropriate sample sizes, which regulatory agencies require. Appropriate RCTs play an important role in evidence-based medicine. However, it is often difficult to conduct appropriate RCTs involving participants with rare diseases or children.

The prevalence of rare diseases is defined as less than five per 10,000 persons (Orphan Medicinal Product Regulation, 2000), meaning the number of patients is small. However, due to the presence of several rare diseases in the world, many patients suffer from rare diseases (Unkel et al., 2016). There are several obstacles to the development of drugs and medical devices for the treatment of rare diseases. First, it is difficult to design and conduct appropriate clinical trials to investigate the efficacy of a test treatment because of

the small number of patients. Hence, in the development of rare diseases, it is difficult to construct sufficient evidence from two independent RCTs, as is the case with major diseases. Second, the development of rare diseases may be hampered by insufficient knowledge of their clinical course. Third, the small number of patients with rare diseases may diminish the commercial interest toward further development. To solve these problems, the research on clinical trial methodologies and statistical issues to evaluate novel therapies for rare diseases has increased in recent years. Additionally, the European Medicines Agency (EMA) published a guideline on clinical trials for small populations (European Medicines Agency, 2006). In this guideline, the options of internal controls or external controls, which may be historical, are presented as control groups for clinical trials. In many cases, it is preferable to use internal control groups; however, under certain circumstances, it is acceptable to use historical controls. Historical control data include participants assigned to placebo or the (current) standard of care in previous clinical trials. The U.S. Food and Drug Administration (FDA) guidance suggests that the use of historical controls may be acceptable, in particular, in cases of serious rare diseases with unmet medical needs (Food and Drug Administration, 2015). Therefore, regulatory agencies have recognized that historical controls can be used to solve the problem of insufficient sample sizes for rare diseases.

Conducting clinical trials in children presents several challenges. The ethical issues that place children at risk in clinical trials and small sample sizes have limited the development of pediatric medical products. To address these issues, design and analysis methodologies unique to pediatric clinical trials are being considered instead of applying the usual methodologies of adult clinical trials. One way to increase the feasibility of pediatric clinical trials is to extrapolate data from adult and other pediatric trials (Dunne et al., 2011; Gamalo-Siebers

et al., 2019). The International Conference on Harmonization (ICH) published the ICH E11A draft guideline which defines pediatric extrapolation as an approach to provide evidence in support of the effective and safe use of drugs in the pediatric population when it can be assumed that the course of the disease and the expected response to a medicinal product would be sufficiently similar between the pediatric and reference (adult or other pediatric) population (European Medicines Agency, 2017). Extrapolating external data (adult or other pediatric data) can be used to reduce necessary evidence obtained from the pediatric population in pediatric clinical trials, depending on the similarities between the target pediatric population and the source population of the external data. Regulatory agencies have taken notice of data extrapolation, and the FDA draft guidance (regarding medical devices) (Food and Drug Administration, 2016) and EMA reflection paper (European Medicines Agency, 2018) have been published. Therefore, in pediatric clinical trials, the use of external data is mainly considered in the context of extrapolating adult or other pediatric data.

Practically, there are various resources for historical/external data that can be used (Ghadessi et al., 2020). Examples include real-world data, patient registries, and completed clinical trials. Ghadessi et al. (2020) presented several examples of applications for approval that use these data as historical data. Many such cases where ethical issues prevent the establishment of concurrent control groups in RCTs, use historical data instead of concurrent control groups. Figure 1 (A) shows this approach. The advantage of using historical data instead of a concurrent control group is that it allows all participants in a current trial to receive a test treatment. Though very attractive, it has the disadvantage of not being able to evaluate the efficacy of a test treatment under comparison, as in RCTs. To overcome this disadvantage, an approach that

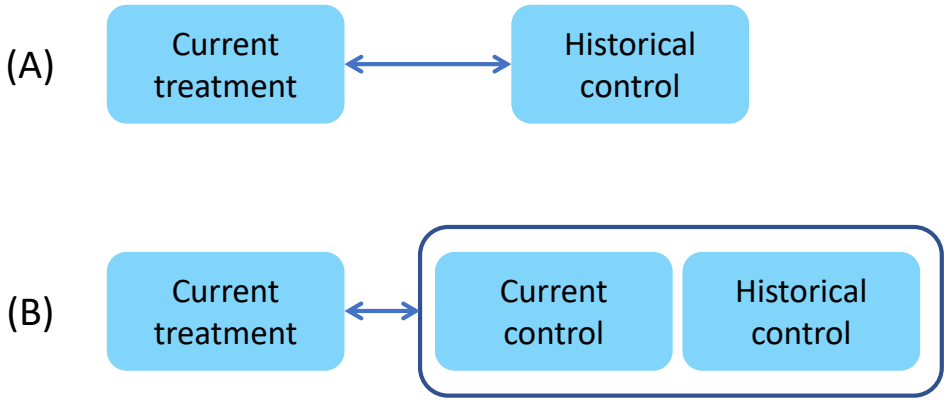


Figure 1: (A): Using historical data instead of a concurrent control group. (B) Incorporating historical data into current trials while assigning participants to the treatment and control groups.

incorporates historical data into current trials while assigning participants to the control group is attracting increasing attention (Lim et al., 2018). Figure 1 (B) shows this approach. Under this approach, since the current trial is an RCT, comparability can be ensured, and the degree to which historical data are used can be determined according to the similarity of historical data to the current control groups. Hereafter, when historical data are used for this purpose, they are referred to as historical control(s). Pocock (1976) proposed six conditions under which historical controls are acceptable. These conditions are intended to ensure the comparability of current treatment and control groups. In reality, the resources of historical controls do not always meet these conditions. Hence, when the comparability of current trials is violated, clinical trial designs and analysis methods that allow for a flexible adjustment of the degree of use of historical controls are needed.

Bayesian approaches have attracted attention as a flexible way of using historical controls (Viele et al., 2014). These approaches can incorporate beliefs

about the relationship between a parameter of control data in the current trial (current control) and parameters of historical controls into the prior distribution. The use of the Bayesian approach is also considered in the FDA draft guidance for medical devices and is increasingly used in practice (Food and Drug Administration, 2010). For example, a placebo-controlled phase II trial of secukinumab for ankylosing spondylitis incorporated historical controls using a Bayesian meta-analytic approach (Baeten et al., 2013). The cited study successfully demonstrated the efficacy of secukinumab in reducing allocation to the current control arm by incorporating historical controls. The use of historical controls using Bayesian approaches is gradually being realized and is expected to rise in the future.

1.2 Historical control data by Bayesian approaches

Several Bayesian approaches for incorporating historical controls into the analysis of the current trial have been hitherto proposed. These approaches make various assumptions regarding the relationships between the parameters of the current and historical controls. Spiegelhalter et al. (2004) classified the assumptions regarding this relationship into six categories.

The simple assumptions are "Irrelevance" and "Equal." "Irrelevance" assumes that historical controls provide no relevant information. If this is the case, the analysis is performed using the current trial without the use of historical controls. "Equal" assumes that parameter of interest θ is common in the current and historical controls. If this is the case, simple pooled current and historical controls are analyzed as the control group in the current trial.

Figure 2 illustrates "Exchangeable" assumption. "Exchangeable" assumes

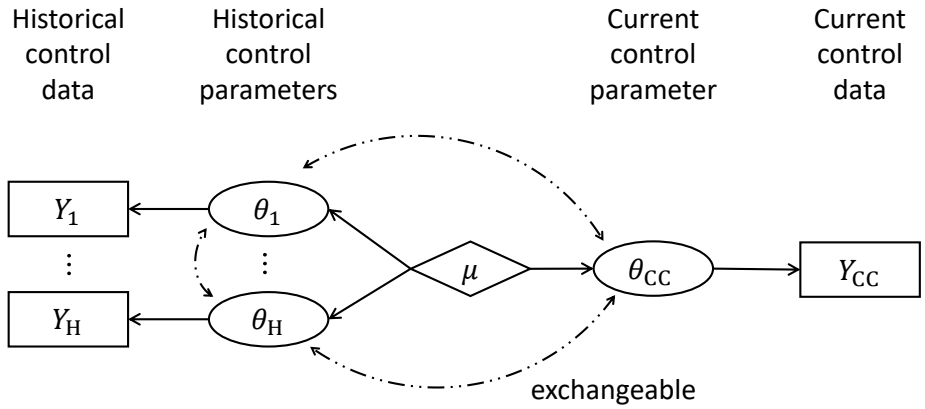


Figure 2: "Exchangeable" assumption relating to parameters of current and historical controls: single arrows represent a distribution, and long dashed double-dotted arrows represent exchangeable cases. μ denotes the overall mean parameter.

that the current and historical controls are similar, meaning that their parameters can be considered exchangeable. Several methods using this assumption have been proposed and are called meta-analytic approaches. The first meta-analytic approach, proposed by Neuenschwander et al. (2010), assumes that the parameters of interest of the current and historical controls follow an identical distribution and shrink toward the overall mean according to between-trial heterogeneity. Schmidli et al. (2014) proposed a robustified meta-analytic approach that considers the conflict between the current and historical controls. Banbeta et al. (2019) demonstrated that, when between-trial heterogeneity is large, the meta-analytic approaches can control for type I error rate to a significant level. However, when no between-trial heterogeneity exists, the statistical power of the meta-analytic approaches could not approach that of an analysis pooling the current and historical controls.

Figure 3 illustrates "Equal but discounted" assumption. "Equal but dis-

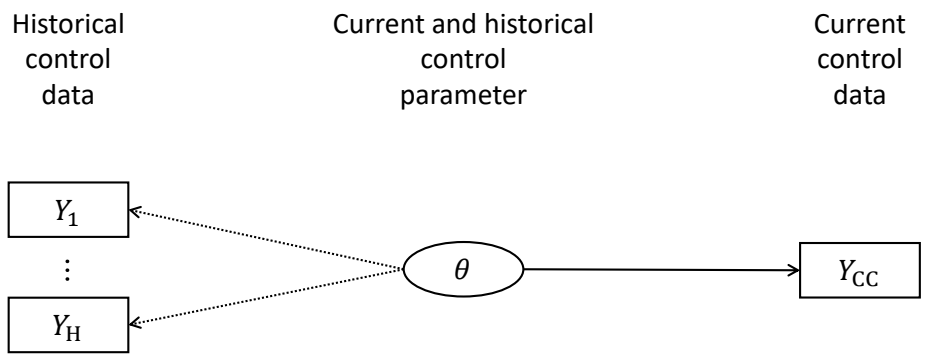


Figure 3: "Equal but discounted" assumption relating to parameters of current and historical controls: single arrows represent a distribution and dotted arrows represent discounting likelihood.

counted" assumes that the parameter of interest θ is common in the current and historical controls, but the precision decreases to discount the historical controls. Power priors use this assumption and down-weight the likelihood of the historical controls using power parameters (Chen et al., 2000). Banbeta et al. (2019) and Gravestock and Held (2019) extended the power prior for the case of multiple historical controls with binary endpoints. However, the power priors have yet to be developed for multiple historical controls with time-to-event endpoints.

Figure 4 illustrates "Potential biases" assumption. "Potential biases" assumes that differences in the parameters of interest between the current control and each historical control exist. These differences are called "potential" biases because their cause cannot be identified since they are due to a lack of quality or unmeasured factors in the historical controls. By assuming a Bayesian shrinkage prior distribution for the potential bias parameters, methods that assume the potential biases are implemented as if the current and

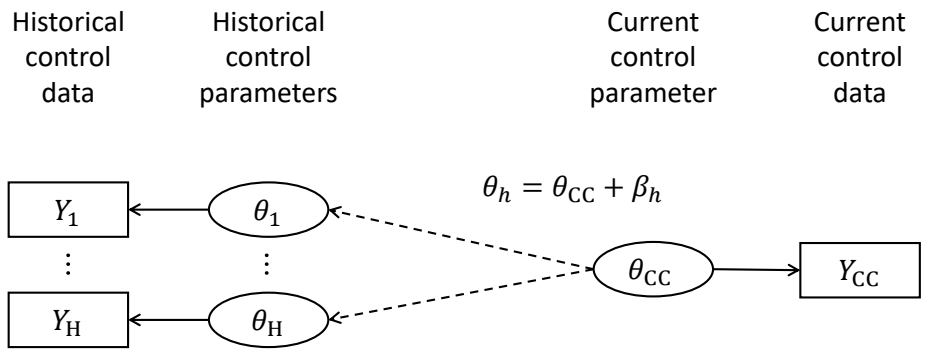


Figure 4: "Potential biases" assumption relating to parameters of current and historical controls: single arrows represent a distribution and long dashed arrows represent existing potential biases. β_h denotes the potential bias parameter which is the difference between the parameter of interest of the current control θ_{CC} and that of the hth historical control θ_h.

historical controls are simply pooled when there are no potential biases in the current and historical controls. Hence, the choice of prior distribution should be made carefully. Pocock (1976) proposed a method that assumes a normal distribution for the potential bias parameters. This method is necessary for determining the variance parameter of the normal distribution to adjust the degree of shrinkage of the potential bias parameters. However, as no procedure has been proposed to determine this adjustment parameter, Pocock's method is not commonly used. A commenturate prior (Hobbs et al., 2011) focuses on a scenario in which only one historical control exists. Hobbs et al. (2012) expanded the commenturate prior so that multiple historical controls can be used, but the relationship between multiple historical controls is not considered.

"Functional dependence" assumes that the parameter of the current control

is a logical function of the parameters of the historical controls but is not widely used.

As the meta-analytic approaches evaluate the overall mean including the current and historical controls, the parameter of interest of the current control is shrunk toward the overall mean. Hence, the meta-analytic approaches do not directly evaluate the differences in the parameters of interest between the current control and each historical control. The power priors cannot deal with multiple historical controls when time-to-event endpoints are assumed. For scenarios with multiple historical controls, the methods of assuming "Potential biases" are not well developed.

Chapter 2

Purpose

In this study, I propose a novel method for incorporating multiple historical controls based on a horseshoe prior, which is a type of global–local Bayesian shrinkage prior. The proposed method assumes that "Potential biases" between the parameters of interest of the current and historical controls exist. The horseshoe prior is applied to concentrate the potential bias on 0. The proposed method solves the problems of the existing methods. First, when no between-trial heterogeneity exists, unlike the meta-analytic approaches, the power of the proposed method is expected to approach that of an analysis pooling the current and historical controls. Second, the power priors have yet to be developed for multiple historical controls with time-to-event endpoints. However, the proposed method deals with them easily. Furthermore, when there are few heterogeneous historical controls, I expect that the proposed method can avoid the influence of the heterogeneous historical controls due to the properties of the horseshoe prior. This dissertation is based on Ohigashi et al. (2022). In the following chapters in this dissertation, through two analyses of clinical trial examples and simulation studies, I compare the proposed and existing methods.

Chapter 3

Methods

3.1 Existing methods

In this section, I review the meta-analytic approaches and the power priors to incorporate the multiple historical controls into a current control in a parallel randomized controlled trial.

3.1.1 Meta-analytic approach

I introduce the meta-analytic approach and its robustified version. Let D_{CC} denote the current control data. Let $D_h (h = 1, \ldots, \mathrm{H})$ denote the historical control data. Let θ_{CC} and θ_h denote interested model parameters for the current and historical controls, respectively.

Under the framework of meta-analytic approaches (Neuenschwander et al., 2010; Schmidli et al., 2014), the following two methods are available for evaluating θ_{CC}: the meta-analytic combined (MAC) and meta-analytic predictive (MAP) approaches. The MAC approach consists of a single analysis based on a hierarchical model for the current and historical controls and has a lower computational cost than the MAP approach. The MAP approach can con-

struct a prior distribution of the current control parameter (MAP prior) using a hierarchical model for the historical controls at the design stage of a current trial. Schmidli et al. (2014) showed that the MAC and MAP approaches are equivalent. However, the MAP approach requires an approximation of the Markov chain Monte Carlo (MCMC) sample when the MAP prior is constructed. Therefore, even a good approximation of the MAP prior does not ensure that the posterior distribution of the MAP approach (using the approximation) approximates that of the MAC approach (Holzhauer, 2020; Zhang et al., 2021). Hence, the MAC approach is applied as a meta-analytic approach in this study because it is simple.

The MAC approach assumes that $\theta_{\mathrm{CC}}, \theta_1, \ldots, \theta_{\mathrm{H}}$ independently follow a normal distribution,

$$\theta_{\mathrm{CC}}, \theta_1, \ldots, \theta_{\mathrm{H}} \sim \mathrm{N}(\mu, \tau^2), \tag{1}$$

where the hyperparameters μ and τ^2 denote the overall mean and the between-trial variance and are usually assigned as a non-informative prior and a weakly informative prior (e.g., half-normal distribution $\mathrm{N}^+(\cdot)$), respectively (Neuenschwander et al., 2010). However, a conflict between the current and historical controls often exists. Considering this case, a robustified meta-analytic approach is proposed (Schmidli et al., 2014). The robustified meta-analytic approach consists of a mixture prior of exchangeable and non-exchangeable components. More specifically, the robustified meta-analytic approach assumes that

$$\theta_{\mathrm{CC}} \sim (1 - w_R) \times \mathrm{N}(\mu, \tau^2) + w_R \times p_R, \tag{2}$$

where p_R is a robust component (vague prior); and w_R is a mixture proportion

of the robust component. In many cases, a constant value (e.g., 0.1) is usually used for w_R (Schmidli et al., 2014; Hupf et al., 2021).

For binary endpoints, the number of responses of the current and historical controls y_{CC}, y_h follow a binomial distribution with the sample size of n_{CC}, n_h at the response rate of π_{CC}, π_h. The logit transform ($\theta_{\mathrm{CC}} = \mathrm{logit}(\pi_{\mathrm{CC}}), \theta_h = \mathrm{logit}(\pi_h)$) is often used with the meta-analytic approaches (Neuenschwander et al., 2010; Schmidli et al., 2014). When time-to-event endpoints are used, I consider a scenario where individual participant data are available. The meta-analytic approach for summary data (e.g., Kaplan–Meier curves) was proposed by Roychoudhury and Neuenschwander (2020). For a participant i, I observed that time-to-event data (t_i, ν_i). t_i denotes an observed time, ν_i denotes whether t_i is a time to an event ($\nu_i = 1$) or right censoring ($\nu_i = 0$). I use indicator variables I_i^{CC}, I_i^h to distinguish among the hth historical control group ($I_i^{\mathrm{CC}} = 0$ and $I_i^h = 1$), the current control group ($I_i^{\mathrm{CC}} = 1$ and $I_i^h = 0$), and the current treatment group ($I_i^{\mathrm{CC}} = 0$ and $I_i^h = 0$). Using these indicator variables, I assume the proportional hazards model (Han et al., 2017; Smith et al., 2020):

$$\lambda_i = \lambda_{\mathrm{CT}} \exp\left(\theta_{\mathrm{CC}} I_i^{\mathrm{CC}} + \sum_{h=1}^{\mathrm{H}} \theta_h I_i^h\right), \tag{3}$$

where λ_i is the hazard for each participant; λ_{CT} is the hazard for the participant belonging to the current treatment group; $-\theta_{\mathrm{CC}}$ is the log hazard ratio of the current treatment group to the current control group; and $-\theta_h$ is the log hazard ratio of the current treatment group to the hth historical control group.

3.1.2 Power prior

3.1.2.1 Single historical control

Initially, I introduce a case of incorporating a single historical control. Let D_{H} denote the historical control data. Let $L(\cdot)$ denote the likelihood function. In the power prior, the interested model parameter θ is common between the current and historical controls. A power parameter δ_{H} is used for down-weighting the historical control.

The power prior for incorporating a single historical control is defined as (Ibrahim and Chen, 2000):

$$\pi(\theta|D_{\mathrm{H}}, \delta_{\mathrm{H}}) \propto L(\theta|D_{\mathrm{H}})^{\delta_{\mathrm{H}}} p(\theta).$$

The power parameter δ_{H} is a fixed value between 0 (ignoring the historical control) and 1 (pooling the historical control for the current control). To avoid arbitrary selection of the power parameter, a joint prior distribution for θ and δ_{H} is also proposed (Ibrahim and Chen, 2000):

$$\pi(\theta, \delta_{\mathrm{H}}|D_{\mathrm{H}}) \propto L(\theta|D_{\mathrm{H}})^{\delta_{\mathrm{H}}} p(\theta) p(\delta_{\mathrm{H}}). \tag{4}$$

The power prior in (4) does not satisfy the likelihood principle. In addition, the posterior distribution of the power parameter δ_{H} tends to be estimated at 0 regardless of the degree of conflict between the current and historical controls (Duan et al., 2006; Neuenschwander et al., 2009). To deal with this problem, a modified power prior (MPP) is proposed as follows (Duan et al., 2006):

$$\pi(\theta, \delta_{\mathrm{H}}|D_{\mathrm{H}}) \propto \frac{L(\theta|D_{\mathrm{H}})^{\delta_{\mathrm{H}}} p(\theta) p(\delta_{\mathrm{H}})}{C(\delta_{\mathrm{H}})},$$

where $C(\delta_{\mathrm{H}}) = \int_{\theta} L(\theta|D_{\mathrm{H}})^{\delta_{\mathrm{H}}} p(\theta) d\theta$ is the scaling constant used to satisfy the likelihood principle. In many cases, a computing of the $C(\delta_{\mathrm{H}})$ is complex; therefore, an algorithm-based approach or an approximation is used (van Rosmalen et al., 2018). However, in the scenario using binomial likelihood, $C(\delta_{\mathrm{H}})$ could be computed analytically (Banbeta et al., 2019). In this scenario, Gravestock et al. (2017) proposed an empirical Bayesian method for estimating the power parameter with one historical control for binary endpoints.

Based on these discussions, the power prior for binomial likelihood is extended to the case of multiple historical controls.

3.1.2.2 Multiple historical controls

It is necessary to consider a conflict between the current control and each historical control (Chen et al., 2000). For binary endpoints, several methods estimating δ_h are proposed. For time-to-event endpoints, a method estimating the power parameter for pooled historical controls is proposed (van Rosmalen et al., 2018; Brard et al., 2019). However, no method is proposed for estimating δ_h based on the degree of conflict between the current control and each historical control for time-to-event endpoints. Therefore, the power priors for binary endpoints are considered in this dissertation.

Banbeta et al. (2019) suggested three MPP approaches for multiple historical controls with binary endpoints. I introduce two methods considering the relationship between the historical controls. In many cases, the historical controls are checked for (their) achievement of Pocock's criteria (Pocock, 1976). Therefore, the incorporated historical control data are interpreted as homogeneous with each other. The dependent MPP (DMPP) assumes that

the power parameters follow a beta distribution, that is,

$$\delta_1, \ldots, \delta_{\mathrm{H}} \sim \mathrm{Beta}(\alpha_\delta, \beta_\delta).$$

The hyperparameters α_δ and β_δ adjust the degree for incorporating the historical controls. These hyperparameters are reparameterized to the mean μ_δ and the variance σ_δ^2 of the beta distribution as the mean $\mu_\delta = \frac{\alpha_\delta}{\alpha_\delta+\beta_\delta}$ and the variance $\sigma_\delta^2 = \frac{\mu_\delta(1-\mu_\delta)}{\alpha_\delta+\beta_\delta+1}$. For binary endpoints, the DMPP can be given by

$$\pi_{DMPP}(\theta,\boldsymbol{\delta}, \mu_\delta, \sigma_\delta^2|\boldsymbol{D}) \propto \frac{\theta^{\Sigma\delta_h y_h+\alpha_\theta-1}(1-\theta)^{\Sigma\delta_h(n_h-y_h)+\beta_\theta-1}\Pi p(\delta_h|\mu_\delta,\sigma_\delta^2)p(\mu_\delta)p(\sigma_\delta^2)}{B(\Sigma\delta_h y_h+\alpha_\theta, \Sigma\delta_h(n_h-y_h)+\beta_\theta)},$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{\mathrm{H}}), \boldsymbol{D} = (D_1, \ldots, D_{\mathrm{H}})$, $B(\cdot,\cdot)$ represents a beta function. α_θ and β_θ are the parameters for the initial prior distribution, $p(\theta) \sim \mathrm{Beta}(\alpha_\theta, \beta_\theta)$. The robust DMPP (RDMPP) accounts for the conflict between the current and historical controls by using a robust component. The prior distribution of $\boldsymbol{\delta}$ consists of a mixture distribution of the $p(\delta_h|\mu_\delta, \sigma_\delta^2)$ and a robust component as with the robustified meta-analytic approach. The prior distribution of $\boldsymbol{\delta}$ is given by

$$\delta_h \sim (1-w_R) \times p(\delta_h|\mu_\delta, \sigma_\delta^2) + w_R \times p_R(\delta_h).$$

If the conflict exists, the historical controls should not be incorporated. Thus, a half-normal distribution with a small variance concentrated close to zero is used for the robust component $p_R(\delta_h)$ (Banbeta et al., 2019).

Gravestock and Held (2019) proposed an empirical Bayesian method for estimating the power parameters for the multiple historical controls with binary endpoints. If I assume a flat initial prior on θ, the prior distribution for

θ with the current control could be given by

$$\begin{aligned}\pi(\theta, \boldsymbol{\delta}|\boldsymbol{D}) &= \frac{1}{C(\boldsymbol{\delta})} \prod \left(\binom{n_h}{y_h}^{\delta_h} \theta^{\delta_h y_h} (1-\theta)^{\delta_h (n_h - y_h)} \right) p(\boldsymbol{\delta}) \\ &= \frac{\prod \left(\binom{n_h}{y_h}^{\delta_h} \theta^{\delta_h y_h} (1-\theta)^{\delta_h (n_h - y_h)} \right) p(\boldsymbol{\delta})}{\prod \left(\binom{n_h}{y_h}^{\delta_h} \right) B(\Sigma \delta_h y_h + 1, \Sigma \delta_h (n_h - y_h) + 1)} \\ &= p(\theta|\boldsymbol{\delta}, \boldsymbol{D}) p(\boldsymbol{\delta}),\end{aligned}$$

where $p(\theta|\boldsymbol{\delta}, \boldsymbol{D}) \sim \text{Beta}(\theta|\Sigma\delta_h y_h + 1, \Sigma\delta_h(n_h - y_h) + 1)$ and scaling constant $C(\boldsymbol{\delta})$ is given by

$$\begin{aligned}C(\boldsymbol{\delta}) &= \int_0^1 \prod \left(\binom{n_h}{y_h}^{\delta_h} \theta^{\delta_h y_h} (1-\theta)^{\delta_h (n_h - y_h)} \right) d\theta \\ &= \prod \left(\binom{n_h}{y_h}^{\delta_h} \right) B(\Sigma \delta_h y_h + 1, \Sigma \delta_h (n_h - y_h) + 1).\end{aligned}$$

The marginal likelihood for the empirical Bayes is

$$\begin{aligned}p(\boldsymbol{\delta}|D_{\text{CC}}, \boldsymbol{D}) &\propto \int_0^1 L(\theta|y_{\text{CC}}, n_{\text{CC}}) p(\theta, \boldsymbol{\delta}|\boldsymbol{D}) d\theta \\ &= \int_0^1 \binom{n_{\text{CC}}}{y_{\text{CC}}} \theta^{y_{\text{CC}}} (1-\theta)^{n_{\text{CC}} - y_{\text{CC}}} \\ &\qquad \times \text{Beta}(\theta|\Sigma\delta_h y_h + 1, \Sigma\delta_h(n_h - y_h) + 1) d\theta \\ &= \binom{n_{\text{CC}}}{y_{\text{CC}}} \frac{B(y_{\text{CC}} + \Sigma\delta_h y_h + 1, n_{\text{CC}} - y_{\text{CC}} + \Sigma\delta_h(n_h - y_h) + 1)}{B(\Sigma\delta_h y_h + 1, \Sigma\delta_h(n_h - y_h) + 1)},\end{aligned}$$

where the integral is arranged using relationships among beta density, binomial density, and beta-binomial density. The empirical Bayes estimator is

$$\hat{\boldsymbol{\delta}} = \arg\max_{\boldsymbol{\delta}}(\boldsymbol{\delta}|D_{\mathrm{CC}}, \boldsymbol{D}).$$

Because the beta-binomial density is known that there is no analytical form for the maximum likelihood estimator, I use a numerical optimization method by the R-package `StudyPrior` (Gravestock and Held, 2019; Gravestock, 2018).

3.2 Proposed method

In this section, I review the horseshoe prior and propose a method for incorporating the multiple historical controls based on the horseshoe prior.

3.2.1 Horseshoe prior

For each of the observations $i = 1, \ldots, n$, consider the single continuous outcome y_i with several explanatory variables given by

$$y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i, \epsilon_i \sim \mathrm{N}(0, \sigma^2), i = 1, \ldots, n,$$

where $\boldsymbol{x}_i$ is the D-dimensional vector of the explanatory variable, $\boldsymbol{\beta}$ contains the corresponding weights, and σ^2 is the noise variance. The horseshoe prior is set for the regression coefficients $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_D)$ except the intercept β_0:

$$\beta_j|\lambda_j, \tau \sim \mathrm{N}(0, \lambda_j^2\tau^2), \lambda_j \sim \mathrm{C}^+(0, 1), \tau \sim \mathrm{C}^+(0, \sigma_\tau), j = 1, \ldots, D.$$

$\mathrm{C}^+(0, \sigma_\tau)$ is a half-Cauchy prior on the positive reals with a scale parameter σ_τ. The global shrinkage parameter τ adjusts the degree of shrinking of all

the coefficients toward zero. The local shrinkage parameter λ_j allows some of the coefficients to avoid shrinkage. Carvalho et al. (2010) proposed that a prior distribution for the global shrinkage parameter is $\tau \sim \mathrm{C}^+(0, 1)$. For ease of notation with comparing the density of the horseshoe with the other shrinkage prior, I assume fixed values of $\sigma = \tau = 1$ and suppress conditioning on these terms. Figure 5 shows the densities for the horseshoe, Cauchy, and Laplace priors. Although the density function of the horseshoe prior lacks a closed-form representation, it behaves essentially like $\log(1 + 2/\beta_j^2)$ and can be well approximated by elementary functions, as detailed in Theorem 1 of Carvalho et al. (2010). The horseshoe density is more concentrated on 0 than the Cauchy and Laplace densities. The tail of the horseshoe density is similar to that of the Cauchy density.

The horseshoe prior has the two good properties of tail-robustness and efficiency when a large signal is observed, and the true mean is 0. For more

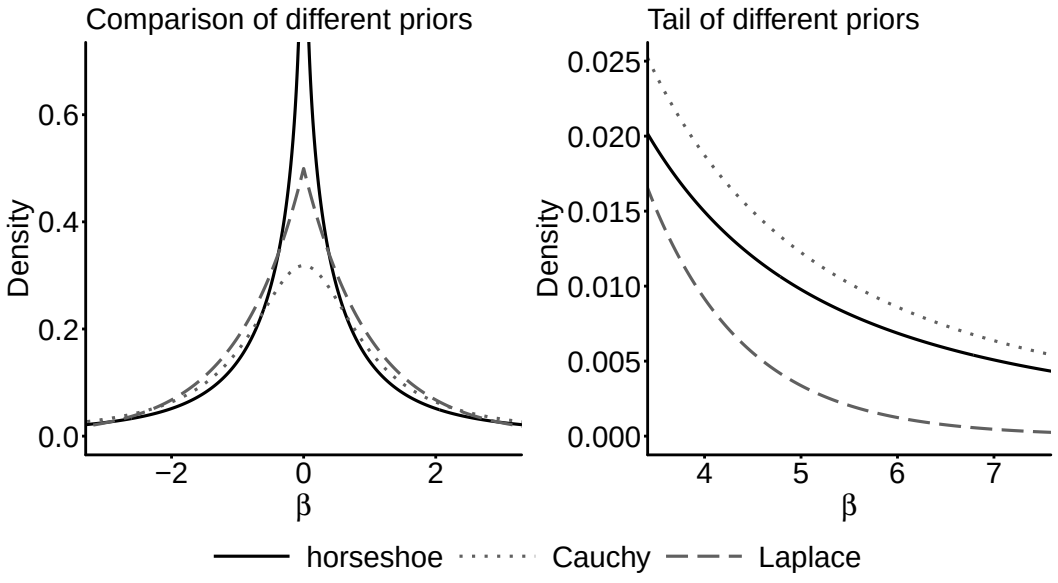


Figure 5: Comparison of the horseshoe, Cauchy, and Laplace densities. The horseshoe density is more concentrated on 0 than the Cauchy and Laplace densities. The tail of the horseshoe density is similar to that of the Cauchy density.

details, see Carvalho et al. (2010).

3.2.2 Proposed method

In this study, I propose a method in which the parameter estimates of the current and historical controls do not shrink toward an overall mean; rather, the historical controls shrink toward the current control. I define the potential bias β_h, namely, the difference in the parameter of interest between the current control θ_{CC} and the hth historical control θ_h, as $\theta_h = \theta_{\mathrm{CC}} + \beta_h$. However, since I evaluate whether or not the circumstances in which the historical controls have been performed are similar to those of the current control using Pocock's criteria (Pocock, 1976) in advance, I expect the historical controls to follow the same distribution as the current control. Therefore, the horseshoe prior is applied to concentrate the potential bias β_h on 0 as

$$\beta_h \sim \mathrm{N}(0, \lambda_h^2 \tau^2), \lambda_h \sim \mathrm{C}^+(0,1), \tau \sim \mathrm{C}^+(0,1).$$

When there is no conflict between the current control and hth historical control, the posterior distribution of the potential bias β_h is estimated to concentrate on 0 using the shrinkage characteristics of the horseshoe prior. When the current control and each of the few historical controls conflict, the potential biases between them avoid shrinkage by estimating the posterior density around 0 of λ_h to be low. If the many historical controls conflict with the current control, the posterior density around 0 of τ is also estimated to be low.

When assuming binary or time-to-event endpoints or incorporating participant-level covariates, the proposed method can be used in the same way as the meta-analytic approaches described in Section 3.1.1.

3.2.3 Property of the proposed method

I examine the property of detecting the heterogeneous historical control by the proposed method in scenarios in which it conflicts with the current control. The effective number of non-zero coefficients m_{eff} proposed by Piironen and Vehtari (2017a) is an indicator of the number of regression coefficients that are not shrunk by the horseshoe prior and that could be used in situations applying generalized linear models. In this study, such an indicator could be interpreted as the number of heterogeneous historical controls. As a simple example, I consider the posterior distributions of m_{eff} in scenarios with one heterogeneous historical control. I assume scenarios in which four and eight historical controls exist; the numbers of participants are 30 and 90 per group; the response rate of both the current and the homogeneous historical controls is 0.5; and the response rate of the heterogeneous historical control is 0.5, 0.4, 0.3, and 0.2. Figures 6 and 7 show the posterior distributions of m_{eff} with 30 and 90 participants, respectively. In the scenario in which the response rate of the heterogeneous historical control is 0.2 with either 30 or 90 participants, the posterior distribution of m_{eff} is estimated to concentrate on 1. When the response rate of the heterogeneous historical control is 0.3 with 90 participants, the posterior distribution of m_{eff} is estimated to be around 1; however, with 30 participants, the posterior distribution of m_{eff} around 0 remains high. Figures 8 and 9 show the posterior distributions of the λ_h of the heterogeneous historical control and those of τ. The posterior distributions of τ also change according to the response rate of the heterogeneous historical control. According to the above examples, evaluating the magnitude of the conflict between the current and historical controls using the proposed method is influenced by between-trial heterogeneity and the sample size of each historical control.

Pocock's method (Pocock, 1976) and Holzhauer's method (Holzhauer, 2020)

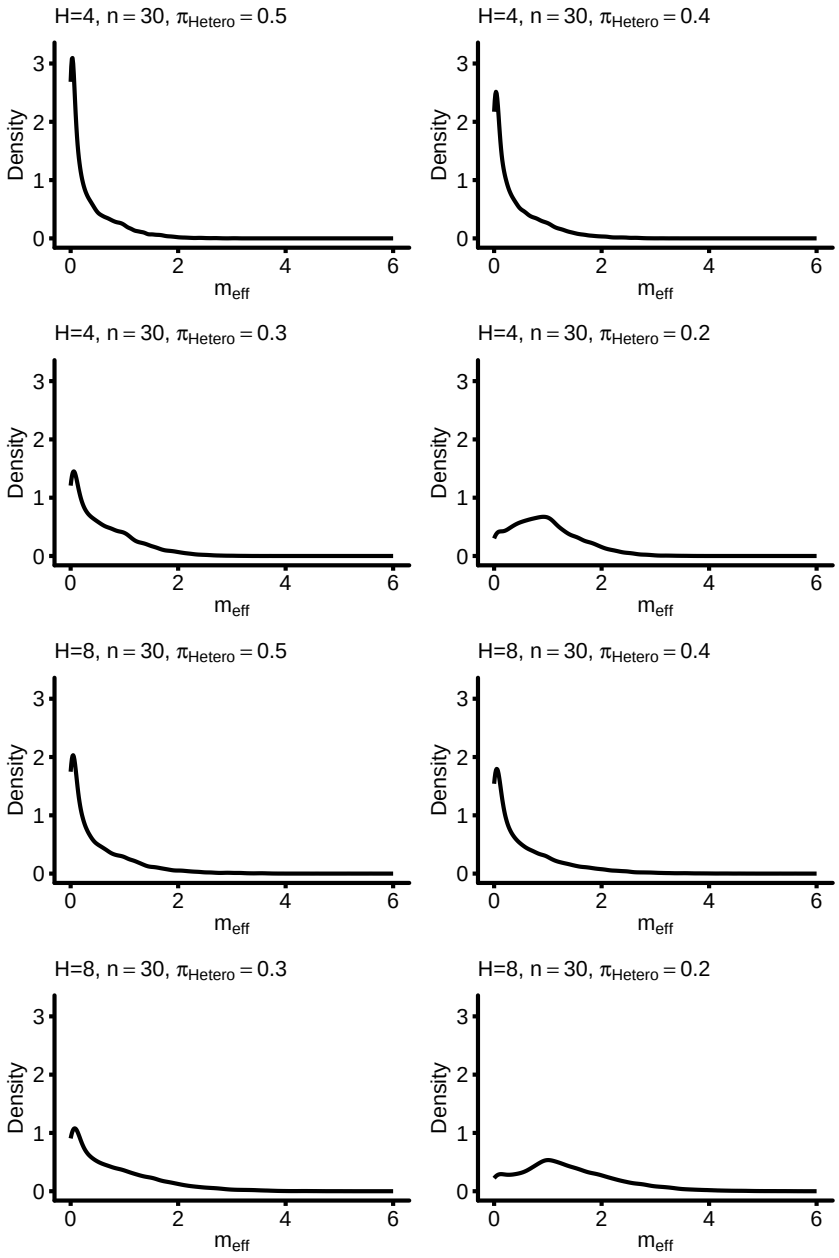


Figure 6: Posterior distributions of m_{eff} in the scenarios with 30 participants. As the response rate of the heterogeneous historical control increases, the density of $m_{\text{eff}} = 0$ decreases. When the response rate is 0.2, the posterior distribution of m_{eff} is estimated to concentrate on 1. m_{eff} denotes the effective number of non-zero coefficients. H denotes the number of historical controls. n denotes the number of participants per group. π_{Hetero} denotes the response rate of the heterogeneous historical control.

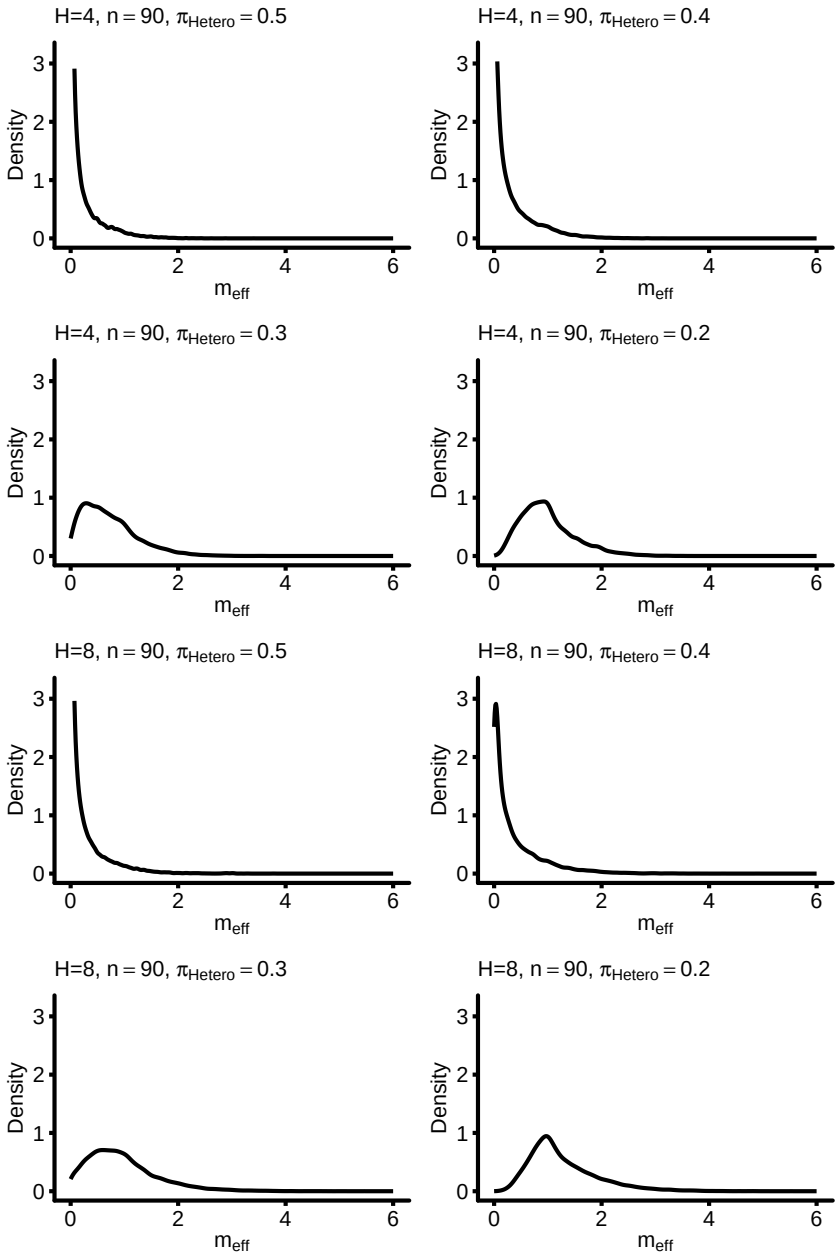


Figure 7: Posterior distributions of m_{eff} in the scenarios with 90 participants. As the response rate of the heterogeneous historical control increases, the density of $m_{\text{eff}} = 0$ decreases. Unlike the case of 30 participants, when the response rate is 0.3, the posterior distribution of m_{eff} is estimated to concentrate on 1. m_{eff} denotes the effective number of non-zero coefficients. H denotes the number of historical controls. n denotes the number of participants per group. π_{Hetero} denotes the response rate of the heterogeneous historical control.

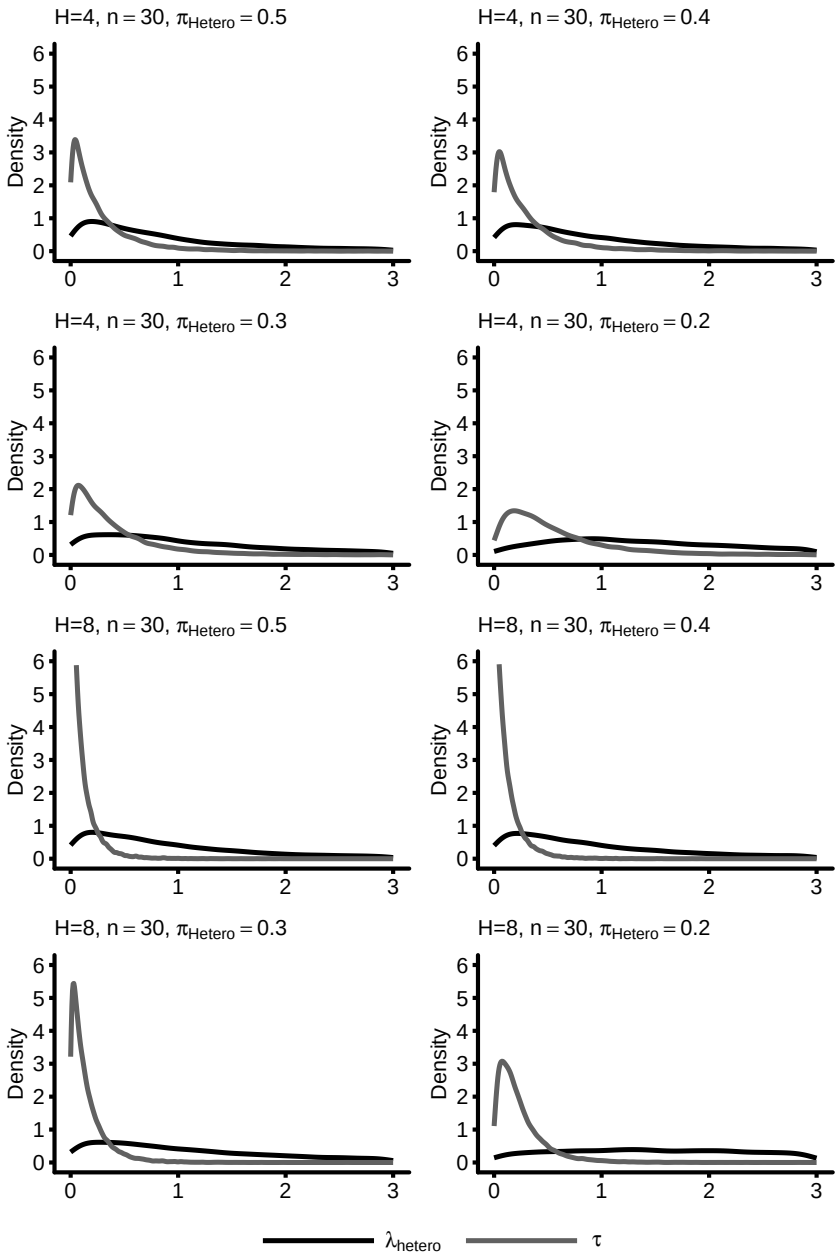


Figure 8: Posterior distributions of λ_h and τ in the scenarios with 30 participants. As the response rate of the heterogeneous historical control increases, the densities of $\tau = 0$ and $\lambda_{\text{hetero}} = 0$ decrease. The density of $\tau = 0$ with eight historical controls is higher than that with four historical controls. H denotes the number of historical controls. n denotes the number of participants per group. π_{Hetero} denotes the response rate of the heterogeneous historical control.

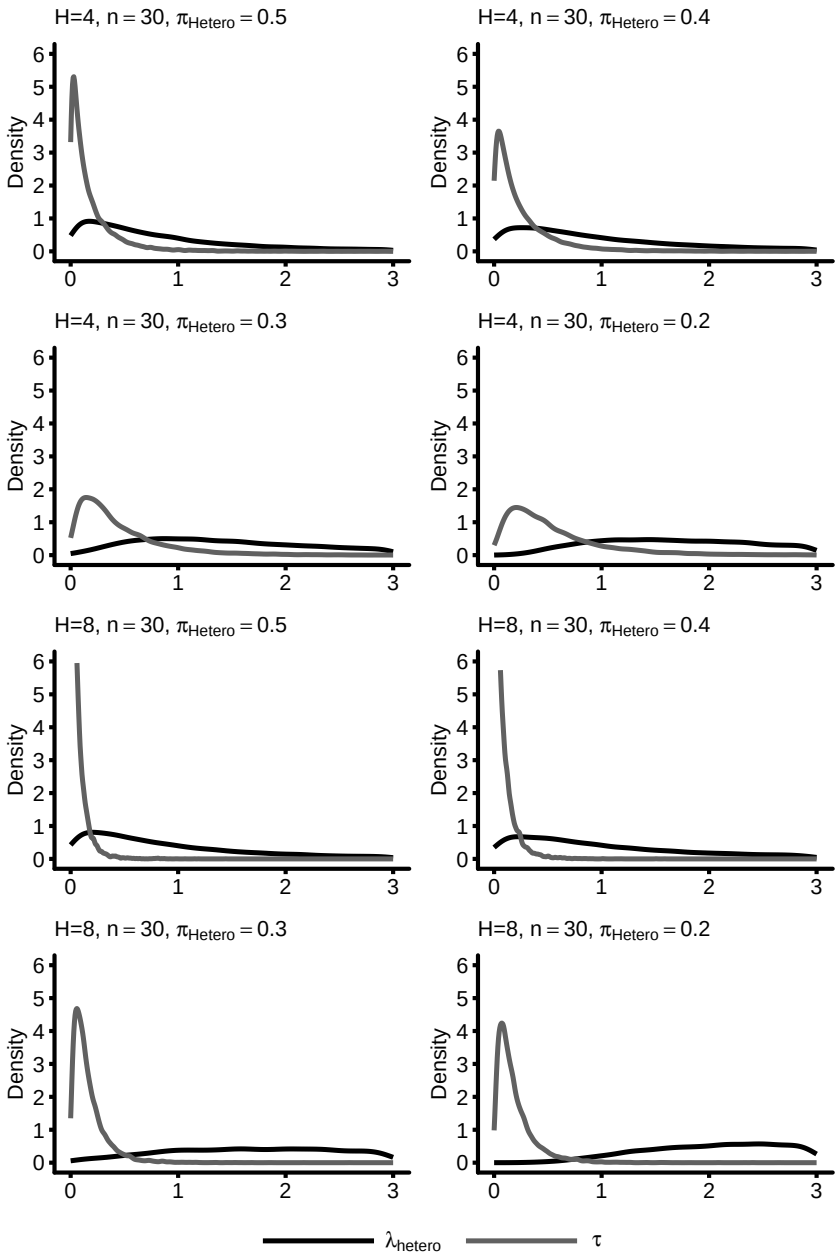


Figure 9: Posterior distributions of λ_h and τ in the scenarios with 90 participants. As the response rate of the heterogeneous historical control increases, the densities of $\tau = 0$ and $\lambda_{\text{hetero}} = 0$ decrease. The density of $\tau = 0$ with eight historical controls is higher than that with four historical controls. The density around lambda = 0 in the scenario with 90 participants is higher than that in the scenario with 30 participants. This suggests that heterogeneous historical controls can be clearly distinguished when the sample size is large. H denotes the number of historical controls. n denotes the number of participants per group. π_{Hetero} denotes the response rate of the heterogeneous historical control.

are similar to the proposed method. Pocock's method assumes that the historical controls are a potentially biased representation of the distribution of the current control, which is similar to the assumption of the proposed method. However, it is necessary to determine a parameter to adjust the degree of shrinkage of the potential biases. Therefore, I do not employ Pocock's method in the clinical trial examples and simulation studies because no rule for deciding the parameter has been proposed under that method. Holzhauer's method averages both a model in which the parameters of interest of the current and historical controls follow an identical distribution and a model in which the parameters of interest of the historical controls follow a distribution with a different mean from that of the current control. Holzhauer's method thus differs from the proposed method, as it does not evaluate the potential biases between the current control and each historical control.

In summary, the proposed method has the following two novelties: (1) it directly evaluates the differences in the parameters of interest between the current control and each historical control; and (2) unlike power priors, it easily deals with the multiple historical controls when time-to-event endpoints are assumed. As the meta-analytic approaches evaluate the overall mean including the current and historical controls, the parameter of interest of the current control θ_{CC} is shrunk toward the overall mean. Thus, when a few heterogeneous historical controls exist, I expect the proposed method to improve the estimated accuracy of θ_{CC} more than the meta-analytic approaches. This is the primary difference between the proposed and meta-analytic approaches.

Chapter 4

Results

4.1 Case study

In this section, I apply the proposed and existing methods discussed in Sections 3.1 and 3.2 to two clinical trial examples. Section 4.1.1 provides an example with a binary endpoint. Section 4.1.2 provides an example with a time-to-event endpoint.

4.1.1 Binary endpoint

This example contains five trials included in a Cochrane review (Prefontaine et al., 2009; Isogawa et al., 2020). These trials target Crohn's disease and compare the efficacy of azathioprine at a dose of 2.0 mg/(kg·day) with a placebo. The outcome is a binary endpoint (i.e., the achievement of maintenance of remission). Table 1 lists the observed data. The most recent D'Haens trial is regarded as the current trial, while the other trials are regarded as the historical trials. Therefore, only the information on the placebo group is extracted from the historical trials. Since the response rate of the Lemann trial is much higher than those of the other historical controls and the current control, it is

Table 1: Observed response rate of the azathioprine and placebo groups from five trials.

		Remission / Total (%)	
Trial	Source	Azathioprine	Placebo
Current	D'Haens et al. (2008)	18 / 32 (56.3)	9 / 29 (31.0)
Historical 1	Lemann et al. (2005)		36 /43 (83.7)
Historical 2	O'Donoghe et al. (1978)		8 / 27 (29.6)
Historical 3	Rosenberg et al. (1975)		4 / 10 (40.0)
Historical 4	Willoughby et al. (1971)		2 / 5 (40.0)

desirable to avoid the large impact of incorporating the Lemann trial.

A Beta$(1, 1)$ prior is assumed for the response rate of the current control in all the fully Bayesian methods. A N$(0, 100^2)$ prior is assumed for the treatment effect on the response rate scale. In the "Pooled" data analysis that includes the current and historical controls without accounting for between-trial heterogeneity, a Beta$(1, 1)$ prior is assumed for the common response rate of the current and historical controls. In the "EX" method, the full exchangeability MAC approach is applied with the between-trial standard deviation (SD) following a N$^+(0, 1)$ prior and the overall mean following a N$(0, 100^2)$ prior. In the "EXNEX" method, the robustified MAC approach is applied with the robust component constructed as $w_R = 0.1$ and using a N$(\mu, 100^2)$ prior. In the "DMPP" method, the hyperparameters μ_δ and σ^2_δ are assumed to have a U$(0, 1)$ prior and a InverseGamma$(0.01, 0.01)$ prior, respectively. In the "RDMPP" method, δ_h is assumed to have a N$^+(0, \delta^2_\delta/6.25)$ prior for the robust component $p_R(\delta_h)$ (Banbeta et al., 2019). In the "EBPP" method, the empirical Bayesian power prior described in Section 3.1.2.2 is applied using the R package `StudyPrior`. In the "HS" method, the proposed method presented in Section 3.2.2 is applied. The computations involve MCMC computations in fully Bayesian methods. These are conducted using Stan via the `cmdstanr`

package (version 0.4.0) within R version 4.1.1 for Windows. For all the fully Bayesian methods, a quadruple chain of sufficient length is run. Convergence is assessed using $\hat{R}$ (Gelman et al., 2013), which provides information on the convergence of the algorithm. The length of the chain is fixed to reach a criterion of $\hat{R} < 1.03$.

Table 2 shows the posterior distributions of the treatment effect (difference between the response rates of the azathioprine and placebo groups) estimated using the above methods. The posterior mean of the treatment effect in the "Current" data analysis is 23.7%. In the "Pooled" data analysis, the posterior mean is 4.4%. The posterior means of the other methods lie between these two percentages. The posterior means of EX, EXNEX, EBPP, and HS are close to that of the "Current" data analysis. This means that EX, EXNEX, EBPP, and HS are less influenced by the heterogeneous historical control (i.e., Lemann trial).

The posterior SD of the treatment effect using RDMPP is larger than

Table 2: Summary statistics of posterior distribution of the treatment effect in terms of the response rate (%).

Method	Mean	SD	95% CI
Current	23.7	11.7	0.0, 45.6
Pooled	4.4	9.6	−14.3, 22.9
EX	22.5	11.7	−0.9, 44.7
EXNEX	22.6	11.8	−1.1, 45.0
DMPP	11.2	11.2	−10.6, 33.7
RDMPP	17.1	12.2	−7.0, 40.3
EBPP	23.0	10.2	2.7, 42.4
HS	22.1	10.6	0.8, 42.3

SD: standard deviation; CI: credible interval; Current, current data analysis; Pooled, pooled data analysis; EX, full exchangeability meta-analytic combined method; EXNEX, robustified meta-analytic combined method; DMPP, dependent modified power prior; RDMPP, robust dependent modified power prior; EBPP, empirical Bayesian power prior; HS, proposed method using horseshoe prior.

that of the "Current" data analysis. This means that the RDMPP is strongly influenced by the heterogeneous historical control. The posterior SDs using EX and EXNEX are equivalent to that of the "Current" data analysis, whereas those using EBPP and HS are smaller. This means that the EBPP and HS could incorporate positively the homogeneous historical controls and avoid the impact of incorporating the heterogeneous historical control.

The 95% credible intervals (CIs) of the treatment effect using EBPP and HS do not include zero, whereas those of the other methods include zero. Figure 10 shows the posterior distributions of the potential bias β_h between the log-odds of the current control and each historical control using HS. Only

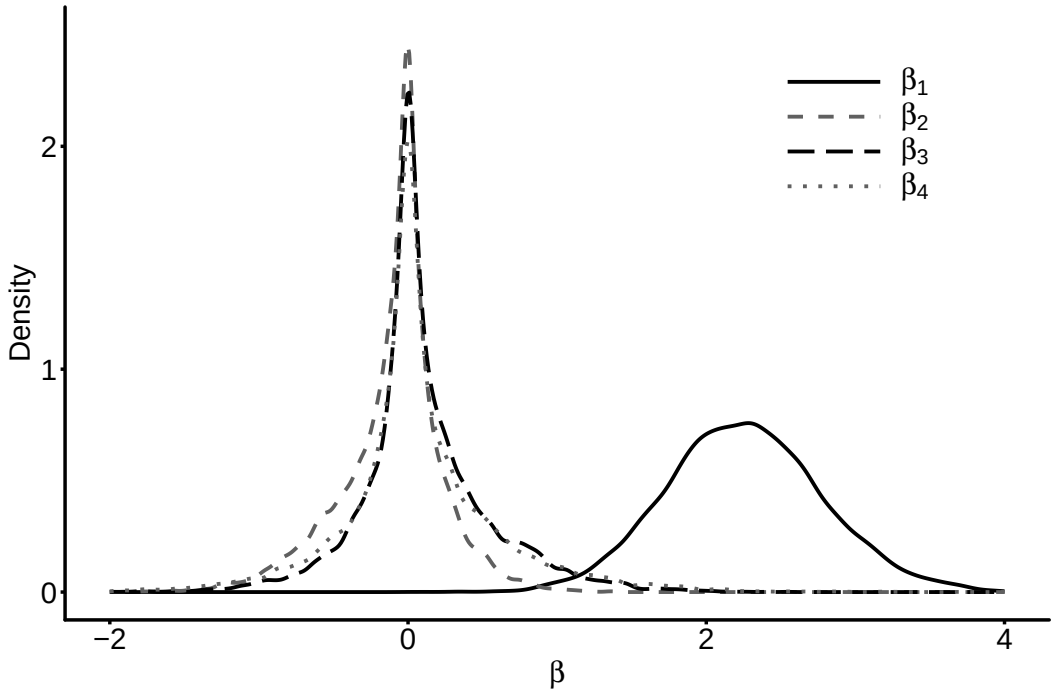


Figure 10: Posterior distributions of the potential bias between the log-odds of the current control and each historical control using HS. Only the posterior distribution of the potential bias of the heterogeneous historical control (Lemann trial) is estimated to be far from zero. This suggests that HS could distinguish the heterogeneous historical controls. β_h denotes the potential bias parameter which is the difference between the log-odds of the current control and that of the hth historical control. HS, proposed method using horseshoe prior.

the posterior distribution of the potential bias of the heterogeneous historical control (Lemann trial) is estimated to be far from zero.

4.1.2 Time-to-event endpoint

This example contains four trials from Project Data Sphere (`https://www.projectdatasphere.org/`), an open-source repository of individual-level patient data from phase IIB/III oncology trials. Table 3 shows some basic information on the trials used in this example, which target extensive-stage small cell lung cancer. The interventions of the current and historical controls are platinating agents (carboplatin or cisplatin) and etoposide. The interventions of the current treatment are LY2510924, carboplatin, and etoposide. The outcome is progression-free survival (PFS). LY2510924 does not show a significant effect in the current trial (Salgia et al., 2017). However, when the treatment effect limited to the intervention period (4 months) is focused on, the hazard ratio suggests a slight treatment effect. Hence, in this example, the time-to-event data are only considered up to 4 months, and all data thereafter are censored for 4 months. Figure 11 illustrates the Kaplan–Meier plots for PFS in each group, showing no clear difference between the current control and three historical controls. I also conduct an analysis that considers participant-level

Table 3: Available information in Project Data Sphere on extensive stage small cell lung cancer trials.

	ClinicalTrials.gov ID	Intervention	Useable N
Current	NCT01439568	LY2510924 Carboplatin+Etoposide	47
		Carboplatin+Etoposide	42
Historical 1	NCT00003299	Cisplatin+Etoposide	283
Historical 2	NCT00363415	Carboplatin+Etoposide	203
Historical 3	NCT00143455	Cisplatin+Etoposide	450

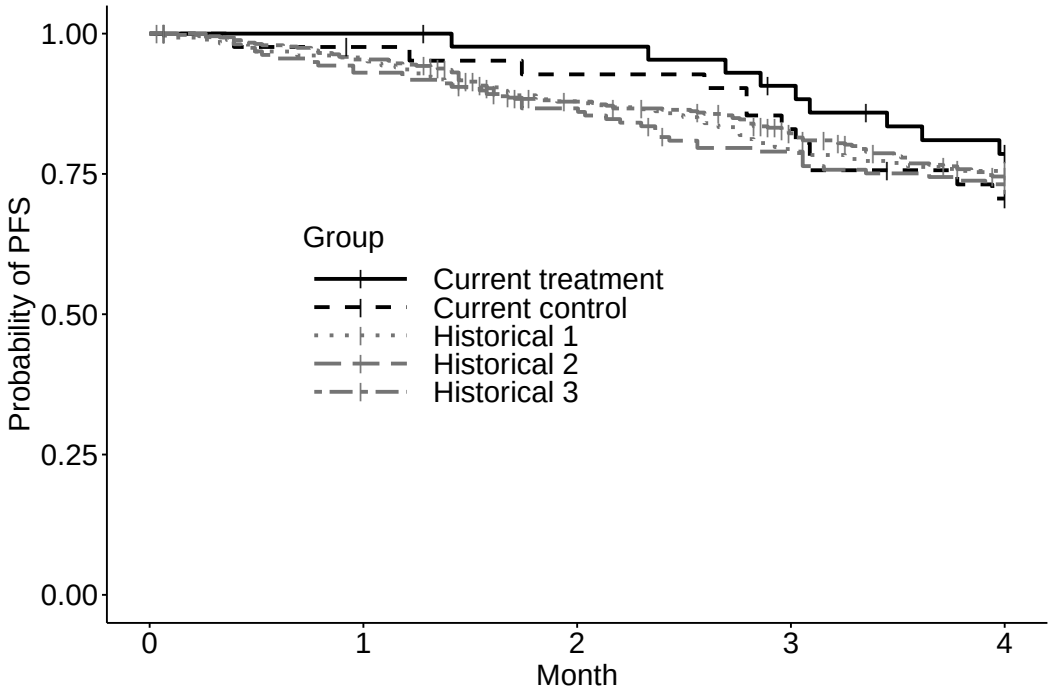


Figure 11: Kaplan–Meier plots for each group. There is no clear difference between the Kaplan–Meier plot of the current control and those of the three historical controls, and the survival rates of PFS are also similar. PFS, progression-free survival.

covariates. Table 4 shows the baseline characteristics recorded in common for the current and historical trials.

In all the methods, a Gamma$(0.01, 0.01)$ prior is assumed for the baseline hazard that describes the risk for individuals in the current treatment group. In the "Current" data analysis, a $\mathrm{N}(0, 10^2)$ prior is assumed for the log-hazard ratio between the current control and current treatment. In the "Pooled" data analysis, a $\mathrm{N}(0, 10^2)$ prior is assumed for the log-hazard ratio between all the control groups and current treatment. In the "EX" method, the full exchangeability MAC approach is applied for the proportional hazards model with the overall mean and between-trial SD following a $\mathrm{N}(0, 10^2)$ and a $\mathrm{N}^+(0, 0.5)$ prior, respectively. In the "EXNEX" method, the robustified MAC approach is applied with the robust component constructed as $w_R = 0.1$ and

Table 4: Baseline characteristics recorded in common for the current and historical trials.

Study	N	Sex (female) (%)	Age (≥ 65) (%)	Number of metastases (≥ 3) (%)
Current Treatment	47	25 (53.2)	24 (51.1)	41 (87.2)
Current Control	42	25 (59.5)	26 (61.9)	39 (92.9)
Historical 1	283	130 (45.9)	96 (33.9)	61 (21.6)
Historical 2	203	48 (27.1)	67 (33.0)	125 (61.6)
Historical 3	450	122 (23.6)	177 (39.3)	246 (54.7)

using a $N(\mu, 1)$ prior. This robust component consists of a weakly informative prior with a mean centered at the mean of the MAC and a variance equal to one (an approximate unit-information prior for exponential data on the log scale) (Roychoudhury and Neuenschwander, 2020). In the "HS" method, the HS prior described in Section 3.2.2 is applied for the proportional hazards model. When we consider the participant-level covariates, a $N(0, 10^2)$ prior is assumed for the regression coefficients for the covariate vector. The computations involve MCMC computations in fully Bayesian methods. These are conducted using Stan via the `cmdstanr` package (version 0.4.0) within R version 4.1.1 for Windows.

Table 5 shows the posterior distributions of the unadjusted and adjusted hazard ratios of the current treatment group to the current control group estimated using the above methods. The posterior means of the hazard ratio using EX, EXNEX, and HS lie between those of the "Current" and "Pooled" data analyses, with or without the covariate adjustment. In particular, the posterior distribution using HS with the covariate adjustment is close to that using the "Pooled" data analysis with the covariate adjustment. Figure 12 and 13 show the posterior distributions of the potential bias β_h between the

Table 5: Summary statistics of posterior distributions of the unadjusted and adjusted hazard ratio of the current treatment group to the current control group.

	Unadjusted			Adjusted		
Method	Mean	SD	95% CI	Mean	SD	95% CI
Current	0.746	0.348	0.271, 1.611	0.783	0.368	0.286, 1.688
Pooled	0.776	0.266	0.346, 1.375	0.843	0.275	0.391, 1.468
EX	0.760	0.280	0.328, 1.399	0.835	0.298	0.369, 1.523
EXNEX	0.760	0.291	0.317, 1.455	0.836	0.298	0.372, 1.524
HS	0.768	0.269	0.344, 1.384	0.842	0.284	0.390, 1.489

SD: standard deviation; CI: credible interval; Current, current data analysis; Pooled, pooled data analysis; EX, full exchangeability meta-analytic combined method; EXNEX, robustified meta-analytic combined method; DMPP, dependent modified power prior; RDMPP, robust dependent modified power prior; EBPP, empirical Bayesian power prior; HS, proposed method using horseshoe prior.

log-hazard ratios of the current treatment group to the current control group and those of the current treatment group to each historical control group using HS. The posterior density of $\beta_h = 0$ with the covariate adjustment is lower than that without.

4.2 Simulation study

I evaluate the performance of the proposed and existing methods using simulation studies based on the examples with binary and time-to-event endpoints in Section 4.1. To analyze the simulation data, I use the same methods and settings as for the case studies.

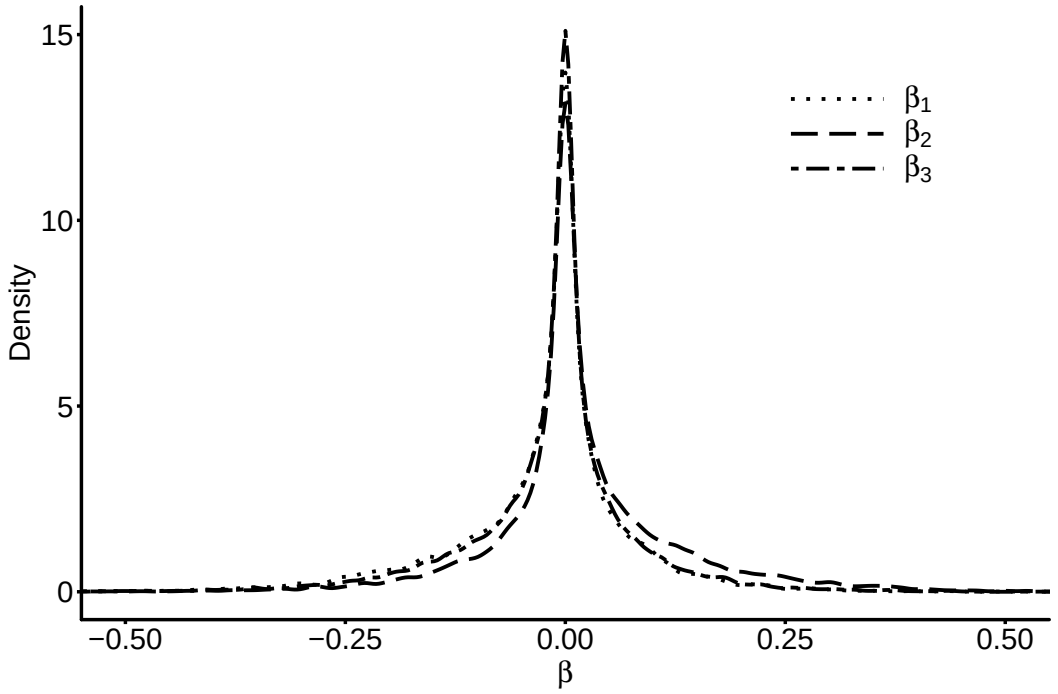


Figure 12: Posterior distributions of the potential bias between the log-hazard ratios of the current treatment group to the current control group and those of the current treatment group to each historical control group using HS with unadjusted analysis. Because the Kaplan–Meier plot of the current control and those of the three historical controls are similar, the posterior distribution of β_h is estimated to concentrate on 0. β_h denotes the potential bias parameter which is the difference between the log-hazard ratios of the current treatment group to the current control group and those of the current treatment group to hth historical control group. HS, proposed method using horseshoe prior.

4.2.1 Binary endpoint

4.2.1.1 Setting

First, I compare the performance of each method in a simple setting. I set the numbers of participants to 30 and 90 and the numbers of historical controls to four and eight. Following the clinical trial examples and also considering scenarios with many historical controls, I set the numbers of historical controls. Hupf et al. (2021) suggests that even when multiple historical controls are available, the number is usually less than 10. I also consider scenarios in the current trial with a 1:2 allocation ratio (i.e., n_{CC} and n_{CT} are 20 and 40,

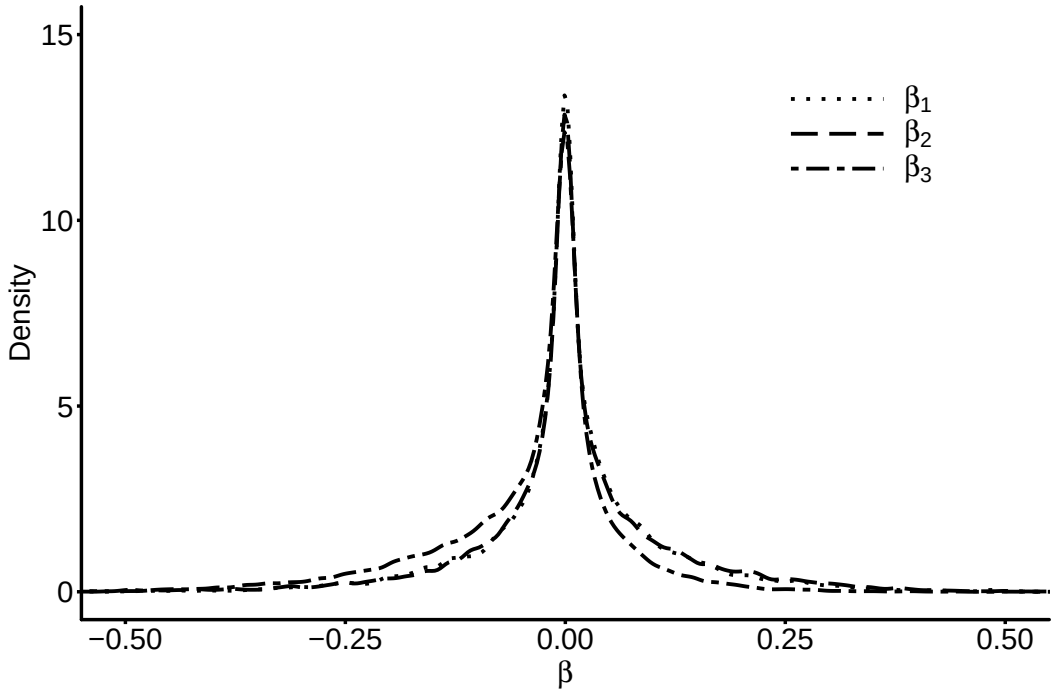


Figure 13: Posterior distributions of the potential bias between the log-hazard ratios of the current treatment group to the current control group and those of the current treatment group to each historical control group using HS with adjusted analysis. Because the Kaplan–Meier plot of the current control and those of the three historical controls are similar, the posterior distribution of β_h is estimated to concentrate on 0. The posterior density of $\beta_h = 0$ with the covariate adjustment is lower than that without. β_h denotes the potential bias parameter which is the difference between the log-hazard ratios of the current treatment group to the current control group and those of the current treatment group to hth historical control group. HS, proposed method using horseshoe prior.

and 60 and 120, respectively). To assess the influence of a small number of historical controls on each method, I also consider scenarios with two and three historical controls. The numbers of responses are generated from a binomial distribution for the historical controls and both the treatment and the control groups of the current trial. The response rate for the jth group π_j, where $j = 1, \ldots, \mathrm{H}, \mathrm{CC}, \mathrm{CT}$ with $1, \ldots, \mathrm{H}$ for the historical controls, CC for the

current control, and CT for the current treatment, is calculated as follows:

$$\pi_j = 1/(1+\exp(-Z_j)), Z_j = \alpha_0 + \alpha_1 G_j + \epsilon_j, \epsilon_j \sim \mathrm{N}(0, \tau_Z^2),$$

where G_j is a treatment indicator equal to 0 for the control group and 1 for the treatment group, and $\epsilon_{\mathrm{CC}} = \epsilon_{\mathrm{CT}}$ because the current trial is a RCT. τ_Z^2 is the variance of the trial-specific effect on the log-odds scale that varies by scenario. Spiegelhalter et al. (2004) and Neuenschwander et al. (2010) suggested that between-trial heterogeneity often lies between 0.01 and 0.25 on the log-odds scale in practice. I consider $\tau_Z^2 = 0, 0.01, 0.09$, and 0.25 to be no, low, moderate, and high between-trial heterogeneity, respectively. Approximating these heterogeneities as the SD of the response rate yields 0, 2.5, 7.5, and 12.5 (%), respectively. In addition, I consider the scenario in which one of the historical controls follows a different distribution from the current control. I set the expected difference between the response rates of the current and heterogeneous historical controls to 30% and between-trial heterogeneity τ_Z^2 to 0. I set the expected response rate to 50% for the current control. Accordingly, the baseline log-odds is set to $\alpha_0 = \log(0.5/(1-0.5)) = 0$. In the scenarios with a treatment effect, I set it to 24.52% (14.45%) when the number of participants in each group is 30 (90) so that the power can reach 50% according to the chi-square test for the current trial when its response rate is 50%. I show a list of settings in Table 6, with the main settings in bold.

I simulate 10,000 datasets for each scenario and setting. The methods are then compared using the frequentist type I error rate and power based on the equal tail 95% CI of the posterior distribution of the treatment effect. I also calculate the calibrated power (CP). For the CP, the rejection region is based on the equal tail CI of the posterior distribution, which yields an observed type

Table 6: Settings for the number of historical controls, the number of participants, the allocation ratio, and the between trial heterogeeity for each scenario in the simulation study with binary endpoint. The mainly discussed settings are shown in **bold font**.

Variable	Settings
Number of historical controls	$2, 3, \mathbf{4}, \mathbf{8}$
Number of participants for each group	$\mathbf{30}, \mathbf{90}$
Allocation ratio for the control group n_{CC} vs. treatment group n_{CT} in the current trial	$\mathbf{1:1}$ (30 vs. 30, 90 vs. 90), 1:2 (20 vs. 40, 60 vs. 120)
Between-trial heterogeneity (τ_Z^2)	**No (0), Low (0.01)**, **Moderate (0.09)**, **High (0.25)**, **OneHetero**

I error rate of approximately 5% in the simulations (Banbeta et al., 2019). This calibration uses different rejection regions for each scenario and method. I also calculate the bias of the posterior mean of the treatment effect as well as the root mean square deviation (RMSD) of the posterior mean of the treatment effect and the mean posterior SD (MPSD) of the posterior distribution of the treatment effect. I generate more than 10,000 data sets and show the results of the 10,000 times that reach the convergence criteria shown above. The computations involve MCMC computations in fully Bayesian methods. These were conducted using Stan via the `rstan` version 2.19.3 package within R versions 3.6.3 for Linux.

4.2.1.2 Results

Figures 14 and 15 (Table A.1 and A.2) show the type I error rates and power with four and eight historical controls, respectively.

In the scenarios in which between-trial heterogeneity is moderate or high, the type I error rates using DMPP, RDMPP, and HS are slightly inflated. With an increasing number of participants, the inflated amount of type I error rates using these methods increases. However, the inflated amount using HS is less

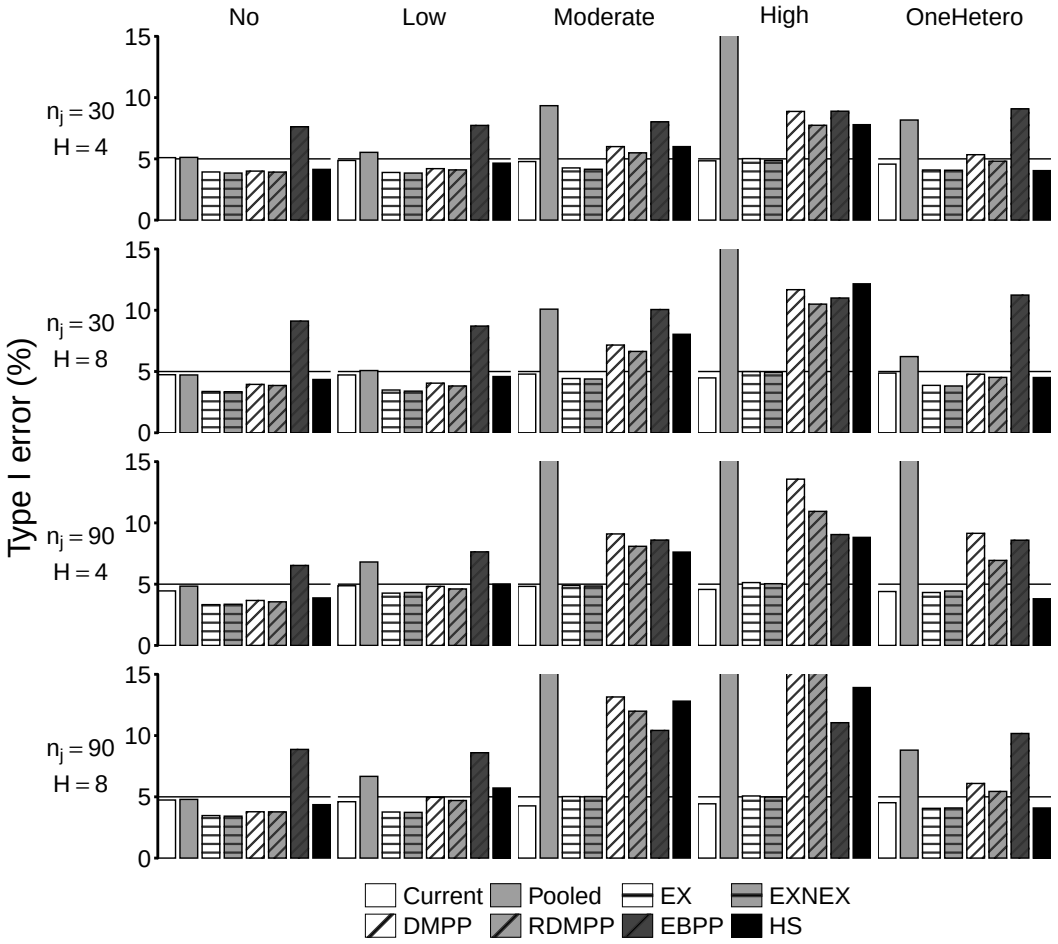


Figure 14: Type I error rate (%) of the treatment effect in the simulation study with a binary endpoint. In the scenarios in which between-trial heterogeneity is moderate or high, the type I error rates using DMPP, RDMPP, and HS are slightly inflated. In the scenario in which one heterogeneous historical control exists, the type I error rates using HS are adequately controlled under 5%, while those using DMPP and RDMPP are sometimes not controlled under 5%. In all the scenarios and settings, the type I error rates using EX and EXNEX (EBPP) are controlled (not controlled) under 5%. n_j denotes the number of participants per group. H denotes the number of historical controls. OneHetero, one heterogeneity historical control; Current, current data analysis; Pooled, pooled data analysis; EX, full exchangeability meta-analytic combined method; EXNEX, robustified meta-analytic combined method; DMPP, dependent modified power prior; RDMPP, robust dependent modified power prior; EBPP, empirical Bayesian power prior; HS, proposed method using horseshoe prior.

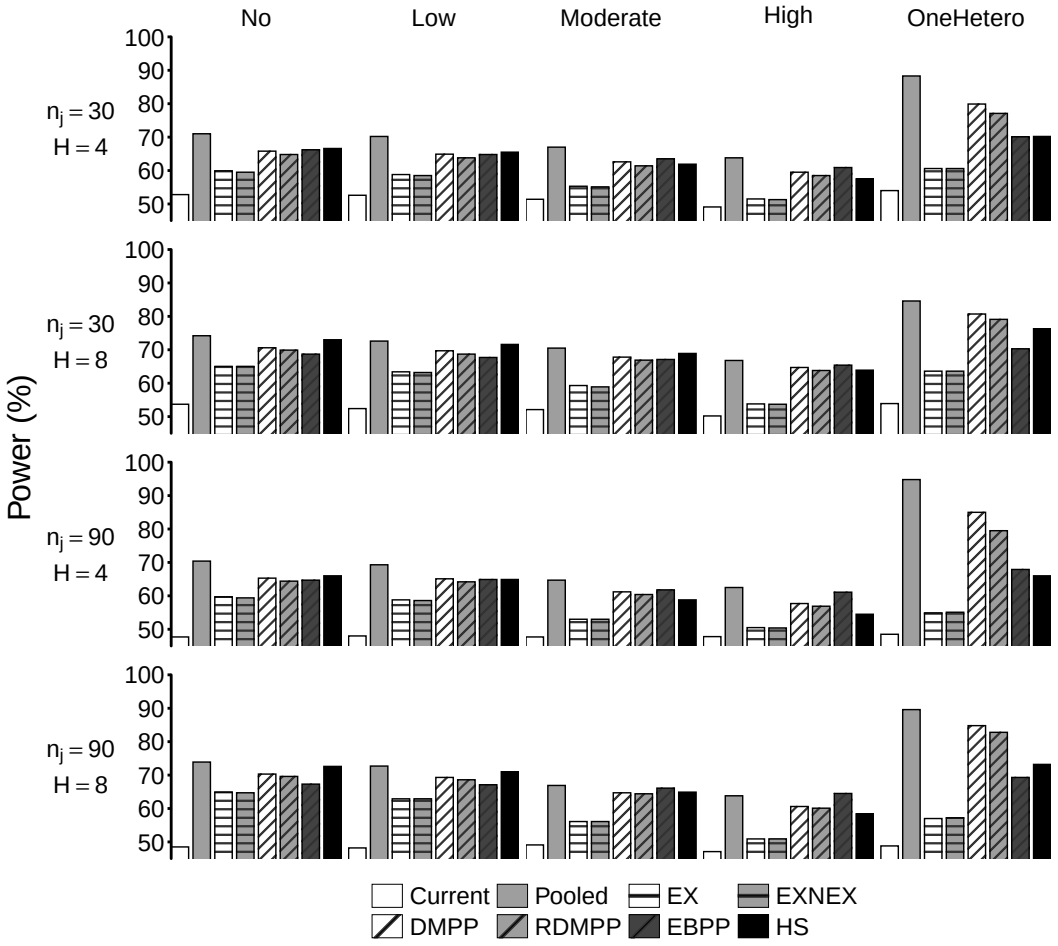


Figure 15: Power (%) of the treatment effect in the simulation study with a binary endpoint. In the scenarios in which there is no between-trial heterogeneity, the power using HS is the highest among the methods incorporating historical controls. When between-trial heterogeneity is moderate or high, or one heterogeneous historical control exists, the power using DMPP, RDMPP, and HS increases due to the influence of inflating type I error rates. n_j denotes the number of participants per group. H denotes the number of historical controls. OneHetero, one heterogeneity historical control; Current, current data analysis; Pooled, pooled data analysis; EX, full exchangeability meta-analytic combined method; EXNEX, robustified meta-analytic combined method; DMPP, dependent modified power prior; RDMPP, robust dependent modified power prior; EBPP, empirical Bayesian power prior; HS, proposed method using horseshoe prior.

than those using DMPP and RDMPP. In the scenario in which one heterogeneous historical control exists, the type I error rates using HS are adequately controlled under 5%, while those using DMPP and RDMPP are sometimes not controlled under 5%. In all the scenarios and settings, the type I error rates using EX and EXNEX (EBPP) are controlled (not controlled) for under 5%. In the scenarios in which there is no between-trial heterogeneity, the power using HS is highest among the methods incorporating the historical controls. When between-trial heterogeneity is moderate or high, or one heterogeneous historical control exists, the power using DMPP, RDMPP, and HS increases due to the influence of inflating type I error rates. However, the increased amount of power using HS is less than that using DMPP and RDMPP.

As is shown in Table A.3, in the scenarios in which between-trial heterogeneity is between none and high, the CPs using HS and DMPP are similar. When one heterogeneous historical control exists, the CP using HS is lower than that using DMPP. Table A.4 shows the average biases. In the scenario in which one heterogeneous historical control exists, the average biases using HS are smaller than those using DMPP. Tables A.5 and A.6 show the RMSDs and MPSDs, respectively. The RMSDs show the degree to which the estimation of the treatment effect improves by incorporating the historical controls. In the scenarios in which no between-trial heterogeneity exists, or the heterogeneity is low, the RMSDs using DMPP, RDMPP, and HS are close to those in the "Pooled" data analysis. However, when between-trial heterogeneity is moderate or high, only the RMSDs using EX, EXNEX, and EBPP are lower than those in the "Current" data analysis. The MPSDs show the degree of incorporating the historical controls. For all the methods, the MPSDs lie between those in the "Current" and "Pooled" data analyses. When no between-trial heterogeneity exists or the heterogeneity is low with eight historical controls,

the MPSDs using HS are lower than those using the other methods. In all the scenarios and settings, the MPSDs using EX and EXNEX are higher than those using the other methods and are close to those in the "Current" data analysis.

Tables A.7 to A.12 show the results of the simulation study in the settings with a 1:2 allocation ratio. The power using DMPP, RDMPP, and HS is 7–10% higher than those in the settings with a 1:1 allocation ratio, suggesting that the loss of power due to setting a 1:2 allocation ratio is covered by incorporating historical controls. Since the power using EX and EXNEX increases by 4–8% in the settings with a 1:2 allocation ratio, the DMPP, RDMPP, and HS could incorporate historical controls more positively than EX and EXNEX.

Tables A.13 to A.24 show the results of the simulation study in the settings with two and three historical controls, highlighting that the type I error rates using HS are roughly similar to those in the setting with four historical controls. When no between-trial heterogeneity exists and $n_j = 30$, the power using HS is lower than that using DMPP, RDMPP, and EBPP, but higher than that using EX and EXNEX. When $n_j = 90$ and between-trial heterogeneity is no and low, the power using HS and DMPP is the same.

4.2.2 Time-to-event endpoint

4.2.2.1 Settings

As with the settings of the simulation study with binary endpoints, I consider a current trial and two, three, four, and eight historical controls with 30 and 90 participants per group in each trial as well as scenarios with a 1:2 allocation ratio (i.e., n_{CC} and n_{CT} are 20 and 40, and 60 and 120, respectively). I mainly discuss the results with four and eight historical controls, as in the

binary endpoint case. Time-to-event data are generated from an exponential distribution of the historical controls and both the treatment and the control groups of the current trial. The hazard for the ith participant of the jth group λ_{ij}, where $i = 1, \ldots, n_j$ and $j = 1, \ldots, \mathrm{H}, \mathrm{CC}, \mathrm{CT}$ with $1, \ldots, \mathrm{H}$ for the historical controls, CC for the current control, and CT for the current treatment, is calculated as follows:

$$\lambda_{ij} = \lambda_0 \exp(\gamma G_{ij} + \epsilon_j), \epsilon_j \sim \mathrm{N}(0, \tau_\lambda^2),$$

where G_{ij} is a treatment indicator equal to 0 for the control group and 1 for the treatment group, and $\epsilon_{\mathrm{CC}} = \epsilon_{\mathrm{CT}}$ because the current trial is a RCT. τ_λ^2 is the variance in the trial-specific effect on the log-hazard scale that varies with the scenario. As with the settings of the simulation with the binary endpoint, I consider $\tau_\lambda^2 = 0, 0.01, 0.09$, and 0.25 to be no, low, moderate, and high between-trial heterogeneity, respectively. Approximating these heterogeneities as the SD of a 4 months survival rate yields 0, 3.5, 10.4, and 17.3 (%), respectively. In addition, I consider the scenario in which there is only one heterogeneous historical control. I set the hazard ratio of the current control to the heterogeneous historical control to 0.7. In this scenario, I set between-trial heterogeneity τ_λ^2 to 0. The baseline hazard λ_0 is determined so that the 4 months survival rate is 0.5. Since the time to censoring is generated from an exponential distribution, the censoring proportion of each trial and group is 0.2. In the scenarios with a treatment effect γ, I set it to 0.586 (0.331) when the number of participants in each group is 30 (90) so that the power can reach 50% according to the log-rank test for the current trial. I show a list of settings in Table 7, with the main settings in bold.

I simulate 10,000 datasets for each scenario and setting. The methods are

Table 7: Settings for the number of historical controls, the number of participants, the allocation ratio, and the between trial heterogeeity for each scenario in the simulation study with time-to-event endpoint. The mainly discussed settings are shown in **bold font**.

Variable	Settings
Number of historical controls	$2, 3, \mathbf{4}, \mathbf{8}$
Number of participants for each group	$\mathbf{30}, \mathbf{90}$
Allocation ratio for the control group n_{CC} vs. treatment group n_{CT} in the current trial	$\mathbf{1:1}$ (30 vs. 30, 90 vs. 90), 1:2 (20 vs. 40, 60 vs. 120)
Between-trial heterogeneity (τ_λ^2)	**No (0), Low (0.01)**, **Moderate (0.09)**, **High (0.25)**, **OneHetero**

compared using the frequentist type I error rate and power based on the equal tail 95% CI of the posterior distribution of the treatment effect. I also calculate the CP, the bias of the posterior mean of the treatment effect, the RMSD of the posterior mean of the treatment effect, and the MPSD of the posterior distribution of the treatment effect. I generate more than 10,000 data sets and show the results of the 10,000 times that reach the convergence criteria shown above. The software used for the MCMC computation and settings related to convergence are the same as in Section 4.2.1.1.

4.2.2.2 Results

Figures 16 and 17 (Table A.25 and A.26) show the type I error rates and power, respectively. In the scenarios in which between-trial heterogeneity is moderate or high, the type I error rates using HS are slightly inflated. When between-trial heterogeneity is low other than when $n_h = 30$ and four historical controls exist, the type I error rates using HS are slightly inflated. Although the type I error rates using EX and EXNEX are slightly inflated in several scenarios, the primary causes are speculated to be the small sample size and censoring. When one heterogeneous historical control exists, the type I error

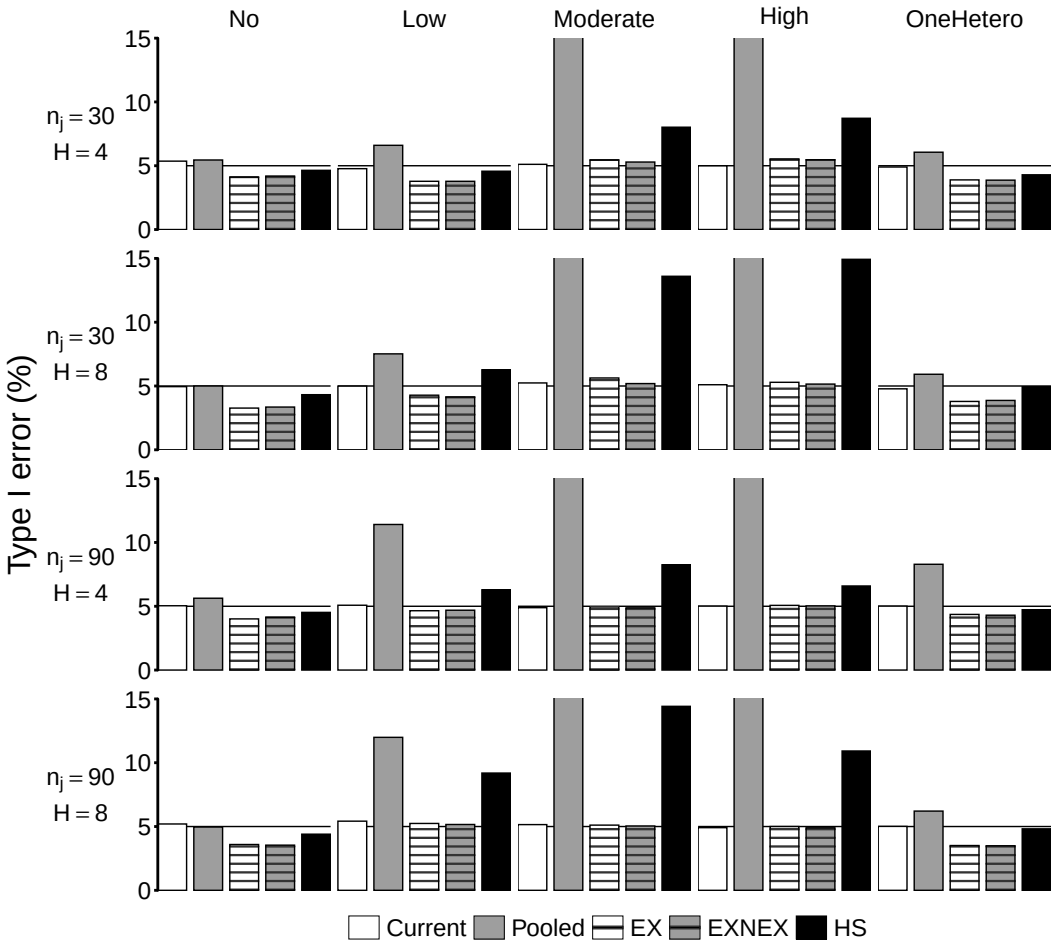


Figure 16: Type I error rate (%) of the treatment effect in the simulation study with a time-to-event endpoint. In the scenarios in which between-trial heterogeneity is moderate or high, the type I error rates using HS are slightly inflated. When between-trial heterogeneity is low other than when $n_h = 30$ and four historical controls exist, the type I error rates using HS are slightly inflated. When one heterogeneous historical control exists, the type I error rates using HS are adequately controlled under 5%. n_j denotes the number of participants per group. H denotes the number of historical controls. OneHetero, one heterogeneity historical control; Current, current data analysis; Pooled, pooled data analysis; EX, full exchangeability meta-analytic combined method; EXNEX, robustified meta-analytic combined method; HS, proposed method using horseshoe prior.

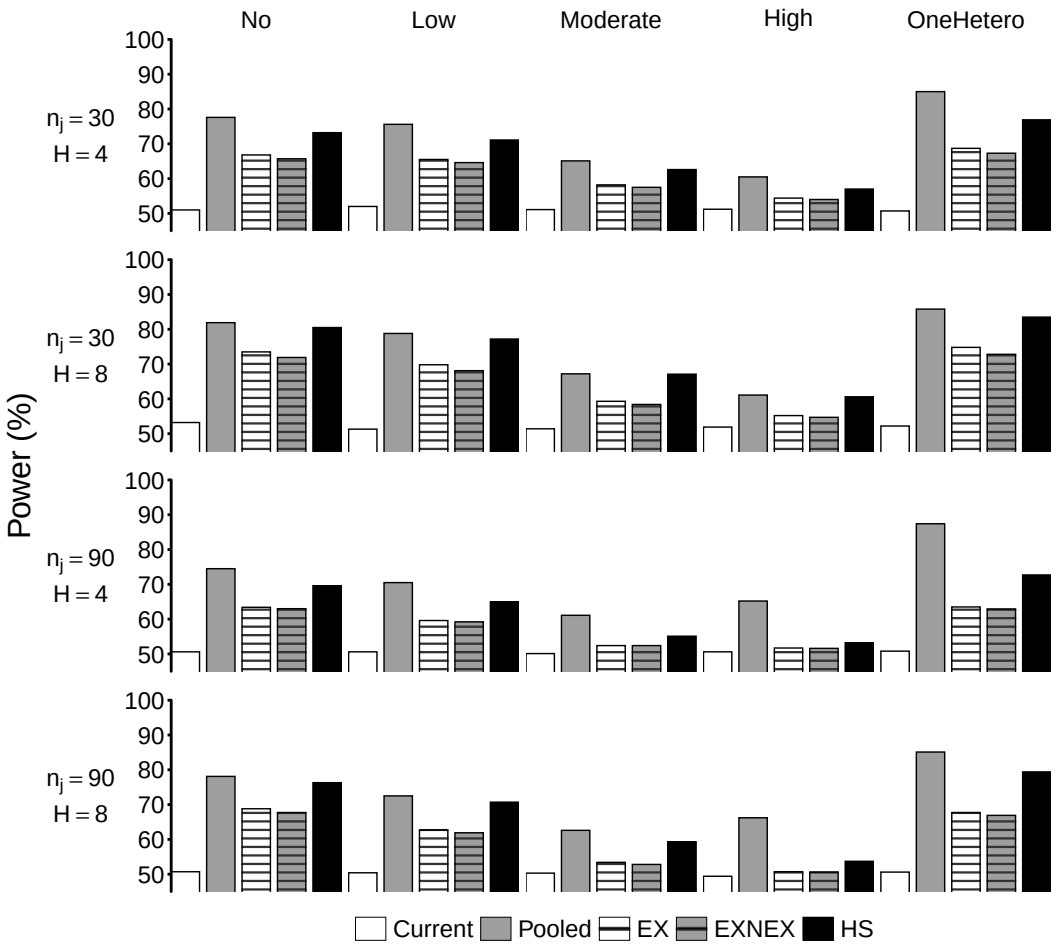


Figure 17: Power (%) of the treatment effect in the simulation study with a time-to-event endpoint. In the scenario in which no between-trial heterogeneity exists, the power using HS is approximately 25% higher than that in the "Current" data analysis. Moreover, the difference in power between the HS and "Pooled" data analysis is 2–4%, and the power using HS is approximately 6–8% higher than that using EX and EXNEX. n_j denotes the number of participants per group. H denotes the number of historical controls. OneHetero, one heterogeneity historical control; Current, current data analysis; Pooled, pooled data analysis; EX, full exchangeability meta-analytic combined method; EXNEX, robustified meta-analytic combined method; HS, proposed method using horseshoe prior.

rates using HS are adequately controlled under 5%.

In the scenario in which no between-trial heterogeneity exists, the power using HS is approximately 25% higher than that in the "Current" data analysis. Moreover, the difference in power between the HS and "Pooled" data analysis is 2–4%, and the power using HS is approximately 6–8% higher than that using EX and EXNEX. When between-trial heterogeneity is moderate or high, the difference in power between HS and "Current" data analysis decreases, although the differences caused by the inflating type I error rates still exist.

Table A.27 shows the CP. In the scenario in which no between-trial heterogeneity exists, the CP using HS is higher than that using EX and EXNEX, and close to the CP in the "Pooled" data analysis. Table A.28 shows the average biases. When one heterogeneous historical control exists, the average biases using HS are larger than those using EX and EXNEX, and smaller than those in the "Pooled" data analysis. Tables A.29 and A.30 show the RMSDs and MPSDs, respectively. In the scenario in which no between-trial heterogeneity exists, the RMSDs using HS are close to those in the "Pooled" data analysis. When one heterogeneous historical control exists, the RMSDs using HS are close to the RMSDs in the "Pooled" data analysis, while the type I error rates using HS are not inflated. In all the scenarios and settings, the MPSDs using EX and EXNEX are higher than those using HS.

Tables A.31 to A.36 show the results of the simulation study in the settings with a 1:2 allocation ratio. The power using HS is 6–9% higher than that in the settings with a 1:1 allocation ratio. As the power using EX and EXNEX is higher by 2–6% with a 1:2 allocation ratio, the HS could incorporate historical controls more positively than EX and EXNEX.

Tables A.37 to A.48 show the results of the simulation study in the settings with two and three historical controls, highlighting that the type I error rates

using HS are roughly similar to those in the setting with four historical controls. When no between-trial heterogeneity exists, the difference in power between the "Pooled" data analysis and HS with two and three historical controls is larger than that with four and eight historical controls, although the power using HS is higher than that using EX and EXNEX.

Chapter 5

Discussion

In this study, I proposed a novel method for incorporating the multiple historical controls based on the horseshoe prior, a type of global–local shrinkage prior. When no conflicts between the current control and each historical control exist (i.e., no heterogeneous historical controls), the proposed method works in a similar way to the pooled data analysis shown in this dissertation. When there are few heterogeneous historical controls, it works in a similar way to an analysis excluding these controls, assuming that they follow a potentially biased distribution from the distribution of the current control. However, when many heterogeneous historical controls exist, the proposed method cannot adequately consider the potential biases between the current control and each heterogeneous historical control. Thus, this method improves the posterior SD of the treatment effect under a relatively low number of heterogeneous historical controls. Moreover, it easily deals with the multiple historical controls even if time-to-event endpoints are assumed, unlike power priors.

I verified these features by analyzing two clinical trial examples and conducting the simulation studies with binary and time-to-event endpoints in order to compare the performances of the proposed and existing methods.

The analysis of the clinical trial example with a binary endpoint showed that the proposed method reduced the posterior SD of the treatment effect by adequately considering the potential bias between the current and heterogeneous historical controls. The posterior SDs using the meta-analytic approaches (EX and EXNEX) were almost equivalent to that of the current data analysis. These results suggest that the meta-analytic approaches were strongly influenced by the heterogeneous historical control. If there are historical controls with heterogeneous and relatively large sample sizes, the use of the meta-analytic approaches is not recommended. The posterior SD using DMPP was smaller than that of the current data analysis, but the posterior mean using DMPP was strongly influenced by the heterogeneous historical control. The posterior mean using RDMPP was not influenced by the heterogeneous historical control as much as the posterior means using DMPP, but the posterior SD using RDMPP was higher than that of the current data analysis. These results suggest that DMPP strongly incorporates even if there are heterogeneous historical controls, and that RDMPP mitigates DMPP's property but does not effectively incorporate historical controls.

The analysis of the clinical trial example with a time-to-event endpoint showed that the proposed method reduced the posterior SD of the treatment effect because the current and historical controls were homogeneous. The posterior SDs using EX and EXNEX were higher than that of proposed method. These results confirm the findings by Banbeta et al. (2019) that the statistical power of meta-analytic approaches could not approach that of an analysis pooling current and historical controls when no between-trial heterogeneity exists. These results were similar in the analysis with the participant-level covariates adjustment, indicating the superiority of the proposed method.

In the simulation studies, I found that the proposed method has the follow-

ing two advantages over the existing methods: (1) when there is no between-trial heterogeneity, and four and eight historical controls, its statistical power is higher than that using the meta-analytic approaches and the power priors; and (2) when one heterogeneous historical control exists, its average biases are smaller than those using the power priors. However, when between-trial heterogeneity is moderate or high, the type I error rates using the proposed method are somewhat inflated. This result reveals the following limitation of the proposed method: it may not distinguish between the homogeneous and heterogeneous historical controls when the current and historical controls are dispersed. Indeed, as the numbers of the heterogeneous historical controls rise, the performance of the proposed method deteriorates, which is synonymous with the fact that as between-trial heterogeneity increases, it performs worse. Nonetheless, if the comparability of the current and historical controls—often evaluated using Pocock's criteria—is lacking, the influence of the heterogeneous historical controls may be avoided by using the proposed method.

If the observed response rate is equal to 0.5 in the logistic regression analysis, Piironen and Vehtari (2017a) recommended that the prior distribution for the global shrinkage parameter be set to $\mathrm{C}^{+}(0, 2)$. However, because their recommendation was derived (empirically) by a simulation study restricted to certain settings, I did not adopt it in this study. Instead, I set the prior distribution for the global shrinkage parameter to $\mathrm{C}^{+}(0, 1)$.

In terms of practical implications, if I can characterize the numbers of the heterogeneous historical controls in advance, I can suitably determine the scale parameter of the prior distribution for the global shrinkage parameter based on the effective number of non-zero coefficients (Piironen and Vehtari, 2017b). However, because it is generally difficult to characterize the extent to which the multiple historical controls differ from the current control when planning a

current trial, determining the scale parameter of the prior distribution for the global shrinkage parameter using the effective number of non-zero coefficients is unrealistic.

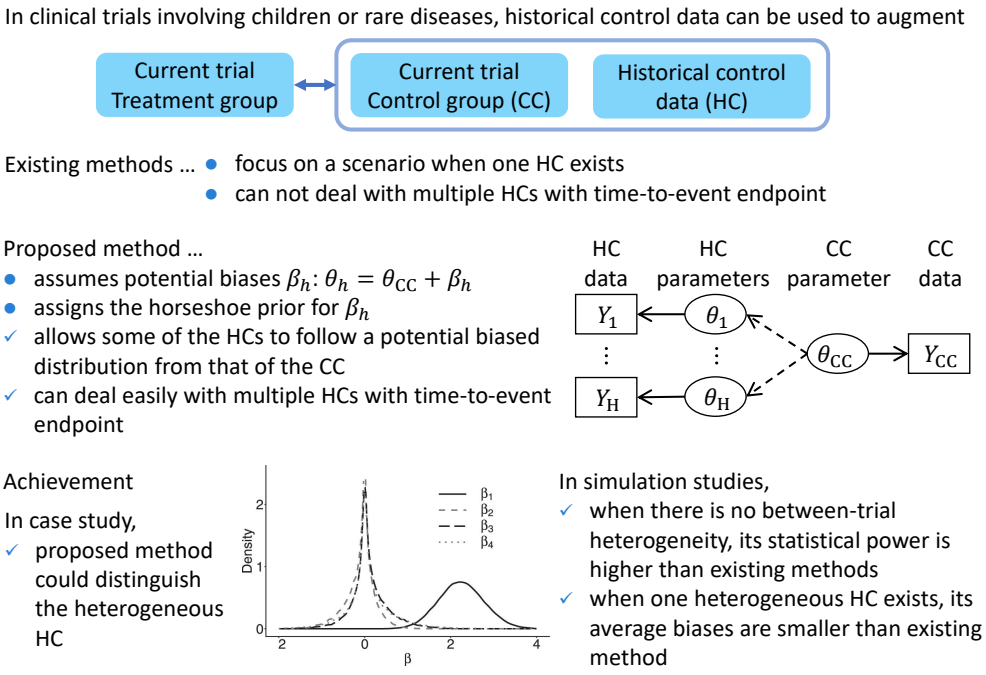
To express the amount of information that the prior/posterior distribution has as the number of subjects, an effective sample size (ESS) is often used. When using the historical controls, it is worth considering evaluating the amount of information corresponding to the number of participants incorporated from historical controls. Several ESS calculation methods have been proposed (Morita et al., 2008, 2012; Hobbs et al., 2013; Neuenschwander et al., 2020). An effective historical sample size (EHSS), an extension of ESS, is also proposed (Hobbs et al., 2013; Wiesenfarth and Calderazzo, 2020; Bennett et al., 2021). Based on ESS and EHSS, a method to control the amount of information corresponding to the number of participants incorporated from historical controls may be developed in the future.

Chapter 6

Conclusion

In this dissertation, I proposed a novel method for incorporating the multiple historical controls based on a horseshoe prior, a type of global–local Bayesian shrinkage prior. The proposed method assumed that "Potential biases" between the parameters of interest of the current and historical controls exist. For making the potential bias close to 0, the horseshoe prior was applied to the potential bias parameters. To clarify the usefulness of the proposed method with respect to the existing methods, I analyzed the two clinical trial examples and conducted the simulation studies with binary and time-to-event endpoints. In conclusion, I recommend using the proposed method when a sufficient number of historical controls exist (e.g., $H \geq 4$), and no or only a few heterogeneous historical controls are expected.

Summary figure



Historical control data are used in clinical trials involving children or rare diseases. Several methods have been proposed for using historical control data. For this situation, I proposed a novel method to apply the horseshoe prior to potential biases. The proposed method (1) could distinguish the heterogeneous historical control in case study and (2) improved performance in simulation studies. β_h denotes the potential bias parameter which is the difference between the parameter of interest of the current control θ_{CC} and that of the hth historical control θ_h.

References

Baeten, D., Baraliakos, X., Braun, J., Sieper, J., Emery, P., van der Heijde, D., McInnes, I., van Laar, J. M., Landewé, R., Wordsworth, P., Wollenhaupt, J., Kellner, H., Paramarta, J., Wei, J., Brachat, A., Bek, S., Laurent, D., Li, Y., Wang, Y. A., Bertolino, A. P., Gsteiger, S., Wright, A. M., and Hueber, W. (2013). Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: A randomised, double-blind, placebo-controlled trial. *The Lancet* **382,** 1705–1713.

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

Bennett, M., White, S., Best, N., and Mander, A. (2021). A novel equivalence probability weighted power prior for using historical control data in an adaptive clinical trial design: A comparison to standard methods. *Pharmaceutical Statistics* **20,** 462–484.

Brard, C., Hampson, L. V., Gaspar, N., Le Deley, M.-C., and Le Teuff, G. (2019). Incorporating individual historical controls and aggregate treatment effect estimates into a Bayesian survival trial: A simulation study. *BMC Medical Research Methodology* **19,** 85.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97,** 465–480.

Chen, M.-H., Ibrahim, J. G., and Shao, Q.-M. (2000). Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference* **84,** 121–137.

Duan, Y., Ye, K., and Smith, E. P. (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics* **17,** 95–106.

Dunne, J., Rodriguez, W. J., Murphy, M. D., Beasley, B. N., Burckart, G. J., Filie, J. D., Lewis, L. L., Sachs, H. C., Sheridan, P. H., Starke, P., and Yao, L. P. (2011). Extrapolation of Adult Data and Other Data in Pediatric Drug-Development Programs. *Pediatrics* **128,** e1242–e1249.

European Medicines Agency (2006). Guideline on clinical trials in small populations.

European Medicines Agency (2017). ICH E11(R1) guideline on clinical investigation of medicinal products in the pediatric population: Step 5.

European Medicines Agency (2018). Reflection paper on the use of extrapolation in the development of medicines for paediatrics - Final.

Food and Drug Administration (2010). Guidance for the use of Bayesian statistics in medical device clinical trials.

Food and Drug Administration (2015). Rare diseases: Common issues in drug development. Draft guidance for industry.

Food and Drug Administration (2016). Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices: Draft Guidance for Industry and Food and Drug Administration Staff.

Gamalo-Siebers, M., Hampson, L., Kordy, K., Weber, S., Nelson, R. M., and Portman, R. (2019). Incorporating Innovative Techniques Toward Extrapolation and Efficient Pediatric Drug Development. *Therapeutic Innovation & Regulatory Science* **53,** 567–578.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis, Third Edition*. CRC Press.

Ghadessi, M., Tang, R., Zhou, J., Liu, R., Wang, C., Toyoizumi, K., Mei, C., Zhang, L., Deng, C. Q., and Beckman, R. A. (2020). A roadmap to using historical controls in clinical trials – by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG). *Orphanet Journal of Rare Diseases* **15,** 69.

Gravestock, I. (2018). Studyprior package. https://r-forge.r-project.org/projects/studyprior/.

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

Gravestock, I., Held, L., and On behalf of the COMBACTE-Net consortium (2017). Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical Statistics* **16,** 349–360.

Han, B., Zhan, J., John Zhong, Z., Liu, D., and Lindborg, S. (2017). Covariate-adjusted borrowing of historical control data in randomized clinical trials. *Pharmaceutical Statistics* **16,** 296–308.

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,** 1047–1056.

Hobbs, B. P., Carlin, B. P., and Sargent, D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials* **10,** 430–440.

Hobbs, B. P., Sargent, D. J., and Carlin, B. P. (2012). Commensurate Priors for Incorporating Historical Information in Clinical Trials Using General and Generalized Linear Models. *Bayesian Analysis* **7,** 639–674.

Holzhauer, B. (2020). Methods for Using Aggregate Historical Control Data in Meta-Analyses of Clinical Trials With Time-to-Event Endpoints. *Statistics in Biopharmaceutical Research* **12,** 107–116.

Hupf, B., Bunn, V., Lin, J., and Dong, C. (2021). Bayesian semiparametric meta-analytic-predictive prior for historical control borrowing in clinical trials. *Statistics in Medicine* **40,** 3385–3399.

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

Isogawa, N., Takeda, K., Maruo, K., and Daimon, T. (2020). A Comparison Between a Meta-analytic Approach and Power Prior Approach to Using Historical Control Information in Clinical Trials With Binary Endpoints. *Therapeutic Innovation & Regulatory Science* **54,** 559–570.

Lim, J., Walley, R., Yuan, J., Liu, J., Dabral, A., Best, N., Grieve, A., Hampson, L., Wolfram, J., Woodward, P., Yong, F., Zhang, X., and Bowen, E. (2018). Minimizing Patient Burden Through the Use of Historical Subject-Level Data in Innovative Confirmatory Clinical Trials: Review of Methods

and Opportunities. *Therapeutic Innovation & Regulatory Science* **52,** 546–559.

Morita, S., Thall, P. F., and Müller, P. (2008). Determining the Effective Sample Size of a Parametric Prior. *Biometrics* **64,** 595–602.

Morita, S., Thall, P. F., and Müller, P. (2012). Prior Effective Sample Size in Conditionally Independent Hierarchical Models. *Bayesian Analysis* **7,** 591–614.

Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). A note on the power prior. *Statistics in Medicine* **28,** 3562–3566.

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

Neuenschwander, B., Weber, S., Schmidli, H., and O'Hagan, A. (2020). Predictively consistent prior effective sample sizes. *Biometrics* **76,** 578–587.

Ohigashi, T., Maruo, K., Sozu, T., and Gosho, M. (2022). Using horseshoe prior for incorporating multiple historical control data in randomized controlled trials. *Statistical Methods in Medical Research* **31,** 1392–1404.

Orphan Medicinal Product Regulation (2000). Regulation (EC) No. 141/2000 of the European Parliament and of the Council of 16 December 1999 on Orphan Medicinal Products. *Official Journal* **L 018,** 0001–0005.

Piironen, J. and Vehtari, A. (2017a). On the Hyperprior Choice for the Global Shrinkage Parameter in the Horseshoe Prior. In Singh, A. and Zhu, J., editors, *Proceedings of the 20th International Conference on Artificial In-*

telligence and Statistics, volume 54 of *Proceedings of Machine Learning Research*, pages 905–913, Fort Lauderdale, FL, USA. PMLR.

Piironen, J. and Vehtari, A. (2017b). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics* **11,** 5018–5051.

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

Prefontaine, E., Sutherland, L. R., MacDonald, J. K., and Cepoiu, M. (2009). Azathioprine or 6-mercaptopurine for maintenance of remission in Crohn's disease. *Cochrane Database of Systematic Reviews* .

Roychoudhury, S. and Neuenschwander, B. (2020). Bayesian leveraging of historical control data for a clinical trial with time-to-event endpoint. *Statistics in Medicine* **39,** 984–995.

Salgia, R., Stille, J. R., Weaver, R. W., McCleod, M., Hamid, O., Polzer, J., Roberson, S., Flynt, A., and Spigel, D. R. (2017). A randomized phase II study of LY2510924 and carboplatin/etoposide versus carboplatin/etoposide in extensive-disease small cell lung cancer. *Lung Cancer* **105,** 7–13.

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

Smith, C. L., Thomas, Z., Enas, N., Thorn, K., Lahn, M., Benhadji, K., and Cleverly, A. (2020). Leveraging historical data into oncology development programs: Two case studies of phase 2 Bayesian augmented control trial designs. *Pharmaceutical Statistics* **19,** 276–290.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons.

Unkel, S., Röver, C., Stallard, N., Benda, N., Posch, M., Zohar, S., and Friede, T. (2016). Systematic reviews in paediatric multiple sclerosis and Creutzfeldt-Jakob disease exemplify shortcomings in methods used to evaluate therapies in rare conditions. *Orphanet Journal of Rare Diseases* **11,** 16.

van Rosmalen, J., Dejardin, D., van Norden, Y., Löwenberg, B., and Lesaffre, E. (2018). Including historical data in the analysis of clinical trials: Is it worth the effort? *Statistical Methods in Medical Research* **27,** 3167–3182.

Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S., Micallef, S., Roychoudhury, S., and Thompson, L. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics* **13,** 41–54.

Wiesenfarth, M. and Calderazzo, S. (2020). Quantification of prior impact in terms of effective current sample size. *Biometrics* **76,** 326–336.

Zhang, H., Chiang, A. Y., and Branson, M. (2021). On the implementation of robust meta-analytical-predictive prior. *Statistics in Biopharmaceutical Research* **0,** 1–9.

Acknowledgments

I would like to express my sincere appreciation to my supervisor Professor Masahiko Gosho and Associate Professor Kazushi Maruo for the invaluable support of my research. They provided not only beneficial comments about my research but also important advice about research activity as a biostatistician. Without their encouragement and guidance, this dissertation would not have been possible.

I am deeply indebted to all of my referees, Professor Masato Homma, Professor Tomoko Sankai, Associate Professor Kazuyoshi Yata, and Assistant Professor Takeshi Machino, who provided instructive comments to improve this dissertation.

I greatly appreciate all the members of Central Coordinatory Unit, Clinical Research Service Center, Tsukuba Clinical Research & Development Organization, University of Tsukuba, especially Professor Koichi Hashimoto for their substantial support.

I would like to thank all the members of the Gosho laboratory, Dr. Masashi Shimura, Dr. Ryota Ishii, Dr. Keisuke Tada, Mr. Kenichi Takahashi, and Mr. Satoshi Yoshida, for useful discussions.

I would like to thank Professor Takashi Sozu, Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science. Thanks to his interesting lecture, I became interested in biostatistics.

Even after I finished my master's degree, he was always concerned about my research progress and gave me useful advice.

Finally, I am grateful to my parents for supporting and encouraging my research activity.

Source

This dissertation is based on the following original article published in

- Ohigashi T, Maruo K, Sozu T, Gosho M. Using horseshoe prior for incorporating multiple historical control data in randomized controlled trials. Statistical Methods in Medical Research (Volume 31, Issue 7) pp. 1392–1404. Copyright © 2022 (SAGE Publications). DOI: 10.1177/09622802221090752.

The contents are re-used in this dissertation following the guideline from SAGE Publications (SAGE ' s Author Archiving and Re-Use Guidelines, https://uk.sagepub.com/en-gb/asi/journal-author-archiving-policies-and-re-use).

Appendix

List of Tables in Appendix

Appendix A

Additional tables in the simulation study

A.1 Binary endpoint

Table A.1: Type I error rate (%) of the treatment effect in the simulation study with a binary endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	5.10	4.87	4.78	4.84	4.57
		Pooled	5.12	5.53	9.34	16.63	8.17
		EX	3.93	3.89	4.26	5.02	4.09
		EXNEX	3.84	3.84	4.15	4.87	4.07
		DMPP	4.00	4.20	6.00	8.87	5.34
		RDMPP	3.92	4.10	5.49	7.74	4.81
		EBPP	7.62	7.73	8.02	8.89	9.08
		HS	4.14	4.65	6.00	7.79	4.04
	H = 8	Current	4.74	4.72	4.79	4.48	4.87
		Pooled	4.72	5.08	10.09	18.17	6.22
		EX	3.37	3.49	4.43	4.99	3.87
		EXNEX	3.35	3.40	4.39	4.94	3.82
		DMPP	3.95	4.05	7.16	11.68	4.78
		RDMPP	3.86	3.82	6.64	10.50	4.52
		EBPP	9.12	8.71	10.06	11.00	11.24
		HS	4.35	4.59	8.04	12.16	4.51
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	4.45	4.87	4.82	4.57	4.40
		Pooled	4.84	6.81	18.52	32.32	18.04
		EX	3.33	4.27	4.89	5.13	4.32
		EXNEX	3.36	4.32	4.85	5.04	4.44
		DMPP	3.67	4.82	9.10	13.57	9.15
		RDMPP	3.56	4.61	8.09	10.94	6.94
		EBPP	6.53	7.64	8.60	9.05	8.59
		HS	3.88	5.00	7.62	8.82	3.81
	H = 8	Current	4.74	4.60	4.26	4.43	4.52
		Pooled	4.79	6.67	20.54	33.54	8.81
		EX	3.47	3.76	5.02	5.07	4.06
		EXNEX	3.42	3.73	5.02	4.98	4.09
		DMPP	3.78	4.96	13.15	18.35	6.09
		RDMPP	3.77	4.70	11.99	16.00	5.44
		EBPP	8.87	8.60	10.42	11.05	10.17
		HS	4.36	5.72	12.81	13.92	4.09

One Hetero, one heterogeneity historical control.

Table A.2: Power (%) of the treatment effect in the simulation study with a binary endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	52.8	52.6	51.4	49.1	54.0
		Pooled	71.0	70.2	67.0	63.8	88.3
		EX	59.9	58.8	55.3	51.5	60.6
		EXNEX	59.5	58.5	55.1	51.3	60.6
		DMPP	65.8	64.9	62.6	59.5	79.9
		RDMPP	64.8	63.8	61.4	58.5	77.1
		EBPP	66.2	64.8	63.5	60.9	70.1
		HS	66.6	65.5	61.9	57.5	70.2
	H = 8	Current	53.7	52.4	52.1	50.2	53.9
		Pooled	74.2	72.6	70.5	66.8	84.6
		EX	65.0	63.4	59.3	53.8	63.6
		EXNEX	65.0	63.2	58.9	53.7	63.6
		DMPP	70.6	69.7	67.8	64.7	80.7
		RDMPP	69.9	68.7	66.9	63.8	79.1
		EBPP	68.7	67.7	67.1	65.4	70.3
		HS	73.0	71.6	68.9	63.9	76.3
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	47.7	48.0	47.7	47.8	48.5
		Pooled	70.4	69.3	64.7	62.5	94.8
		EX	59.7	58.8	53.0	50.5	54.9
		EXNEX	59.4	58.6	53.0	50.4	55.1
		DMPP	65.3	65.1	61.2	57.7	85.0
		RDMPP	64.4	64.2	60.4	56.9	79.5
		EBPP	64.7	64.9	61.8	61.1	67.9
		HS	66.0	64.9	58.8	54.5	66.0
	H = 8	Current	48.5	48.2	49.1	47.1	48.8
		Pooled	73.9	72.7	66.9	63.8	89.6
		EX	64.9	62.9	56.1	50.9	57.0
		EXNEX	64.7	62.9	56.1	50.9	57.2
		DMPP	70.3	69.3	64.7	60.6	84.8
		RDMPP	69.6	68.6	64.4	60.1	82.8
		EBPP	67.3	67.1	66.1	64.5	69.3
		HS	72.6	71.0	64.9	58.4	73.2

One Hetero, one heterogeneity historical control.

Table A.3: Calibrated power (%) of the treatment effect in the simulation study with a binary endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	46.8	47.6	49.8	47.8	53.6
		Pooled	71.6	68.7	55.8	39.8	82.9
		EX	64.4	63.4	59.0	51.9	63.5
		EXNEX	64.3	63.3	59.1	52.0	63.3
		DMPP	70.3	67.8	60.8	49.3	79.8
		RDMPP	69.8	67.6	60.8	50.4	78.1
		EBPP	58.5	58.5	55.7	50.7	59.7
		HS	70.0	67.2	59.2	49.8	73.7
	H = 8	Current	48.0	49.6	50.6	50.3	47.4
		Pooled	74.8	73.5	56.4	40.7	82.2
		EX	70.2	69.6	61.6	54.3	67.5
		EXNEX	70.2	69.3	61.4	54.4	67.3
		DMPP	73.7	73.6	60.1	48.4	80.3
		RDMPP	73.4	73.4	60.6	49.7	79.6
		EBPP	58.7	58.1	54.9	52.0	58.0
		HS	74.4	73.4	59.3	46.9	77.1
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	51.1	49.6	50.2	48.8	51.3
		Pooled	72.2	65.0	39.6	NA	81.3
		EX	66.3	61.5	53.9	49.0	57.5
		EXNEX	66.2	61.1	53.9	49.1	57.4
		DMPP	71.1	66.0	49.4	39.5	77.6
		RDMPP	70.7	65.3	50.9	41.8	74.6
		EBPP	62.4	57.0	53.1	48.9	58.9
		HS	71.0	65.3	50.7	42.9	69.4
	H = 8	Current	51.0	50.9	50.2	48.3	51.1
		Pooled	75.5	67.9	37.6	NA	83.7
		EX	71.1	66.9	54.2	49.4	60.2
		EXNEX	71.2	67.3	54.0	49.6	60.1
		DMPP	74.7	69.5	46.7	36.5	82.8
		RDMPP	74.5	69.1	48.2	37.7	81.4
		EBPP	59.0	58.4	52.7	48.1	57.7
		HS	74.8	68.7	45.3	38.5	76.2

One Hetero, one heterogeneity historical control.

Table A.4: Average bias of the treatment effect in the simulation study with a binary endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	−0.017	−0.018	−0.022	−0.026	−0.014
		Pooled	−0.016	−0.017	−0.021	−0.026	0.045
		EX	−0.017	−0.017	−0.022	−0.026	0.005
		EXNEX	−0.017	−0.017	−0.022	−0.026	0.004
		DMPP	−0.016	−0.017	−0.022	−0.026	0.030
		RDMPP	−0.016	−0.017	−0.021	−0.026	0.026
		EBPP	−0.017	−0.018	−0.021	−0.026	−0.003
		HS	−0.016	−0.017	−0.021	−0.026	0.009
	H = 8	Current	−0.015	−0.018	−0.019	−0.025	−0.014
		Pooled	−0.016	−0.018	−0.019	−0.024	0.018
		EX	−0.016	−0.018	−0.019	−0.025	0.001
		EXNEX	−0.016	−0.018	−0.019	−0.025	0.000
		DMPP	−0.016	−0.018	−0.019	−0.024	0.013
		RDMPP	−0.016	−0.018	−0.019	−0.024	0.012
		EBPP	−0.015	−0.018	−0.019	−0.025	−0.008
		HS	−0.016	−0.018	−0.019	−0.024	−0.001
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	−0.006	−0.005	−0.008	−0.009	−0.004
		Pooled	−0.005	−0.004	−0.006	−0.009	0.057
		EX	−0.005	−0.004	−0.007	−0.009	0.004
		EXNEX	−0.005	−0.004	−0.007	−0.009	0.004
		DMPP	−0.005	−0.004	−0.006	−0.008	0.041
		RDMPP	−0.005	−0.004	−0.006	−0.008	0.033
		EBPP	−0.005	−0.004	−0.007	−0.008	0.004
		HS	−0.005	−0.004	−0.006	−0.008	0.004
	H = 8	Current	−0.004	−0.005	−0.005	−0.009	−0.003
		Pooled	−0.003	−0.004	−0.004	−0.009	0.031
		EX	−0.003	−0.004	−0.005	−0.009	0.004
		EXNEX	−0.003	−0.004	−0.005	−0.009	0.004
		DMPP	−0.003	−0.004	−0.005	−0.009	0.025
		RDMPP	−0.003	−0.004	−0.005	−0.009	0.023
		EBPP	−0.003	−0.004	−0.005	−0.009	0.002
		HS	−0.003	−0.004	−0.004	−0.009	0.001

One Hetero, one heterogeneity historical control.

Table A.5: RMSD of the treatment effect in the simulation study with a binary endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	0.113	0.114	0.114	0.114	0.113
		Pooled	0.086	0.088	0.099	0.117	0.095
		EX	0.094	0.095	0.099	0.104	0.104
		EXNEX	0.094	0.095	0.099	0.104	0.104
		DMPP	0.087	0.089	0.094	0.103	0.091
		RDMPP	0.088	0.089	0.094	0.102	0.092
		EBPP	0.103	0.104	0.106	0.109	0.108
		HS	0.088	0.089	0.096	0.104	0.092
	H = 8	Current	0.115	0.114	0.115	0.114	0.112
		Pooled	0.081	0.084	0.095	0.116	0.082
		EX	0.087	0.089	0.095	0.102	0.094
		EXNEX	0.087	0.089	0.095	0.102	0.094
		DMPP	0.082	0.084	0.091	0.106	0.082
		RDMPP	0.083	0.085	0.091	0.104	0.082
		EBPP	0.106	0.106	0.109	0.111	0.107
		HS	0.082	0.085	0.093	0.107	0.082
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	0.071	0.071	0.071	0.070	0.072
		Pooled	0.054	0.058	0.079	0.104	0.079
		EX	0.059	0.060	0.066	0.069	0.069
		EXNEX	0.059	0.060	0.066	0.069	0.069
		DMPP	0.055	0.057	0.069	0.079	0.070
		RDMPP	0.055	0.057	0.068	0.075	0.068
		EBPP	0.064	0.064	0.067	0.069	0.068
		HS	0.055	0.058	0.068	0.073	0.060
	H = 8	Current	0.072	0.071	0.071	0.070	0.071
		Pooled	0.053	0.056	0.078	0.103	0.061
		EX	0.056	0.057	0.065	0.068	0.066
		EXNEX	0.056	0.058	0.065	0.068	0.066
		DMPP	0.054	0.056	0.071	0.084	0.059
		RDMPP	0.054	0.056	0.070	0.081	0.059
		EBPP	0.066	0.066	0.068	0.069	0.068
		HS	0.053	0.056	0.071	0.078	0.054

One Hetero, one heterogeneity historical control.

Table A.6: MPSD of the treatment effect in the simulation study with a binary endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	0.115	0.115	0.114	0.113	0.115
		Pooled	0.086	0.086	0.086	0.086	0.086
		EX	0.101	0.101	0.104	0.107	0.109
		EXNEX	0.101	0.102	0.104	0.107	0.109
		DMPP	0.094	0.094	0.094	0.095	0.094
		RDMPP	0.095	0.095	0.096	0.097	0.097
		EBPP	0.091	0.091	0.092	0.092	0.090
		HS	0.092	0.093	0.095	0.098	0.098
	H = 8	Current	0.115	0.115	0.114	0.113	0.115
		Pooled	0.082	0.082	0.082	0.081	0.082
		EX	0.095	0.096	0.100	0.105	0.104
		EXNEX	0.096	0.096	0.100	0.105	0.104
		DMPP	0.088	0.088	0.088	0.088	0.088
		RDMPP	0.089	0.089	0.089	0.090	0.089
		EBPP	0.086	0.086	0.087	0.086	0.085
		HS	0.084	0.085	0.086	0.089	0.087
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	0.072	0.072	0.071	0.070	0.072
		Pooled	0.055	0.055	0.054	0.054	0.055
		EX	0.063	0.064	0.067	0.068	0.071
		EXNEX	0.064	0.064	0.067	0.068	0.070
		DMPP	0.059	0.059	0.060	0.061	0.061
		RDMPP	0.060	0.060	0.061	0.063	0.063
		EBPP	0.058	0.058	0.059	0.058	0.058
		HS	0.059	0.059	0.062	0.065	0.063
	H = 8	Current	0.072	0.072	0.071	0.071	0.072
		Pooled	0.053	0.052	0.052	0.052	0.053
		EX	0.060	0.061	0.066	0.068	0.069
		EXNEX	0.060	0.061	0.066	0.068	0.069
		DMPP	0.056	0.056	0.056	0.058	0.056
		RDMPP	0.056	0.056	0.057	0.059	0.057
		EBPP	0.056	0.056	0.056	0.055	0.055
		HS	0.054	0.054	0.057	0.061	0.056

One Hetero, one heterogeneity historical control.

Table A.7: Type I error rate (%) of the treatment effect in the simulation study with a binary endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	4.25	4.72	4.86	4.32	4.62
		Pooled	5.03	5.90	11.90	21.23	10.26
		EX	2.88	3.36	4.17	4.72	3.98
		EXNEX	2.84	3.40	4.18	4.69	3.96
		DMPP	3.22	3.58	6.78	10.88	5.85
		RDMPP	3.04	3.46	6.07	9.26	5.00
		EBPP	8.36	9.18	10.68	11.75	12.32
		HS	3.60	4.24	6.81	9.56	4.15
	H = 8	Current	4.45	4.54	4.77	4.48	4.56
		Pooled	4.58	5.91	13.14	24.26	6.37
		EX	2.59	2.93	4.46	5.20	2.91
		EXNEX	2.52	2.99	4.37	5.08	2.79
		DMPP	3.27	3.97	8.75	16.26	4.42
		RDMPP	3.02	3.68	7.96	14.64	4.09
		EBPP	11.21	12.21	13.55	14.66	14.68
		HS	3.92	5.16	10.22	16.47	4.03
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	4.25	4.71	4.36	4.39	4.30
		Pooled	4.74	7.83	23.91	38.74	22.95
		EX	2.83	3.62	4.87	5.28	4.28
		EXNEX	2.85	3.59	4.84	5.34	4.26
		DMPP	3.22	4.76	12.69	18.94	11.04
		RDMPP	2.98	4.29	10.84	15.43	8.05
		EBPP	8.05	9.48	11.36	12.57	12.66
		HS	3.42	5.07	10.10	11.10	3.15
	H = 8	Current	4.88	4.37	4.32	4.00	4.10
		Pooled	5.46	7.88	26.22	41.35	10.92
		EX	2.99	3.57	5.25	5.15	3.44
		EXNEX	2.99	3.48	5.27	5.03	3.43
		DMPP	3.77	5.24	17.79	27.86	6.91
		RDMPP	3.59	4.86	15.95	24.50	5.94
		EBPP	11.34	11.53	14.34	15.91	14.47
		HS	4.61	6.29	17.01	20.58	3.62

One Hetero, one heterogeneity historical control.

Table A.8: Power (%) of the treatment effect in the simulation study with a binary endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	49.1	49.1	48.2	46.8	49.4
		Pooled	80.2	79.9	75.9	70.8	94.9
		EX	63.2	62.9	59.1	53.3	61.0
		EXNEX	62.9	62.2	58.8	53.2	61.1
		DMPP	73.6	73.4	70.9	66.6	87.9
		RDMPP	71.8	71.7	69.4	65.0	84.4
		EBPP	71.3	71.2	69.7	67.1	73.0
		HS	74.6	74.1	70.2	63.2	77.2
	H = 8	Current	49.5	49.8	48.0	46.7	49.2
		Pooled	84.3	83.6	78.2	73.8	92.6
		EX	71.7	70.5	62.8	56.2	66.5
		EXNEX	71.3	70.3	62.7	56.1	66.4
		DMPP	80.4	79.6	74.9	71.2	88.7
		RDMPP	79.5	78.7	74.1	70.4	87.3
		EBPP	73.9	73.7	71.5	70.4	74.5
		HS	83.1	81.8	76.1	70.5	84.9
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	46.2	45.2	44.7	43.6	45.8
		Pooled	80.2	77.5	69.7	66.3	98.2
		EX	63.2	61.0	53.5	48.4	53.8
		EXNEX	62.9	60.8	53.4	48.6	53.6
		DMPP	74.0	71.6	65.6	61.0	91.7
		RDMPP	72.3	69.8	64.4	59.7	86.4
		EBPP	71.0	69.9	67.4	65.0	73.6
		HS	75.0	71.5	62.4	55.8	71.3
	H = 8	Current	46.3	45.8	44.9	43.3	45.5
		Pooled	85.4	81.3	72.4	68.0	95.8
		EX	72.2	68.7	56.2	49.2	56.3
		EXNEX	71.9	68.4	56.1	49.3	56.2
		DMPP	81.3	78.0	70.5	65.0	92.5
		RDMPP	80.4	77.1	69.8	64.1	90.6
		EBPP	73.5	72.9	70.7	69.2	74.8
		HS	83.9	79.7	69.2	61.5	82.5

One Hetero, one heterogeneity historical control.

Table A.9: Calibrated power (%) of the treatment effect in the simulation study with a binary endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	47.4	47.3	46.7	45.5	47.2
		Pooled	80.2	77.7	60.0	40.1	89.5
		EX	72.5	68.6	61.5	53.5	64.8
		EXNEX	71.9	68.4	61.0	53.7	64.5
		DMPP	79.2	76.9	65.0	50.7	86.5
		RDMPP	78.7	76.3	65.0	52.1	84.6
		EBPP	64.2	63.0	56.1	51.2	59.3
		HS	78.6	75.9	63.0	49.8	79.9
	H = 8	Current	48.0	47.8	46.5	46.7	47.3
		Pooled	85.9	82.7	62.9	41.1	91.0
		EX	80.3	77.4	65.7	55.0	74.1
		EXNEX	80.0	77.4	66.0	55.2	74.0
		DMPP	85.4	82.7	67.2	49.0	89.9
		RDMPP	85.2	82.1	67.7	50.8	89.3
		EBPP	63.0	60.8	55.6	50.8	59.4
		HS	85.5	82.3	64.3	46.6	86.9
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	46.5	45.2	47.2	45.2	48.3
		Pooled	81.5	71.6	36.9	NA	88.8
		EX	72.1	66.5	53.9	46.2	56.4
		EXNEX	72.1	66.2	53.3	46.1	56.1
		DMPP	80.1	72.9	47.2	38.0	83.9
		RDMPP	79.6	72.4	49.5	40.2	81.3
		EBPP	65.0	60.3	52.1	46.0	57.6
		HS	79.6	71.1	48.2	39.2	78.6
	H = 8	Current	44.2	46.5	46.7	46.6	47.5
		Pooled	84.5	76.0	39.5	NA	90.7
		EX	78.9	73.5	54.1	47.0	62.6
		EXNEX	78.4	73.4	54.1	47.2	62.7
		DMPP	83.9	77.5	47.6	34.4	89.2
		RDMPP	83.5	77.1	48.6	33.0	88.3
		EBPP	62.0	60.0	51.2	46.1	57.7
		HS	84.0	76.5	44.3	32.7	84.9

One Hetero, one heterogeneity historical control.

Table A.10: Average bias of the treatment effect in the simulation study with a binary endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	−0.012	−0.012	−0.014	−0.021	−0.011
		Pooled	−0.013	−0.012	−0.014	−0.020	0.052
		EX	−0.013	−0.012	−0.014	−0.021	0.013
		EXNEX	−0.013	−0.012	−0.014	−0.021	0.013
		DMPP	−0.013	−0.012	−0.014	−0.020	0.040
		RDMPP	−0.013	−0.012	−0.014	−0.020	0.036
		EBPP	−0.013	−0.012	−0.014	−0.021	0.004
		HS	−0.013	−0.012	−0.014	−0.020	0.017
	H = 8	Current	−0.012	−0.011	−0.016	−0.022	−0.012
		Pooled	−0.013	−0.012	−0.017	−0.021	0.023
		EX	−0.013	−0.012	−0.017	−0.021	0.006
		EXNEX	−0.013	−0.012	−0.017	−0.021	0.006
		DMPP	−0.013	−0.012	−0.017	−0.021	0.019
		RDMPP	−0.013	−0.012	−0.017	−0.021	0.018
		EBPP	−0.012	−0.011	−0.017	−0.022	−0.003
		HS	−0.013	−0.012	−0.017	−0.021	0.003
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	−0.002	−0.004	−0.005	−0.008	−0.002
		Pooled	−0.002	−0.003	−0.005	−0.008	0.061
		EX	−0.002	−0.003	−0.005	−0.008	0.008
		EXNEX	−0.002	−0.003	−0.005	−0.008	0.008
		DMPP	−0.002	−0.003	−0.005	−0.008	0.049
		RDMPP	−0.002	−0.003	−0.005	−0.008	0.042
		EBPP	−0.002	−0.003	−0.005	−0.008	0.007
		HS	−0.002	−0.003	−0.005	−0.008	0.007
	H = 8	Current	−0.003	−0.003	−0.004	−0.008	−0.003
		Pooled	−0.002	−0.003	−0.004	−0.008	0.032
		EX	−0.002	−0.003	−0.004	−0.008	0.006
		EXNEX	−0.002	−0.003	−0.004	−0.008	0.006
		DMPP	−0.002	−0.003	−0.004	−0.008	0.028
		RDMPP	−0.002	−0.003	−0.004	−0.008	0.026
		EBPP	−0.002	−0.002	−0.004	−0.008	0.003
		HS	−0.002	−0.003	−0.004	−0.008	0.002

One Hetero, one heterogeneity historical control.

Table A.11: RMSD of the treatment effect in the simulation study with a binary endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	0.120	0.121	0.122	0.121	0.122
		Pooled	0.078	0.080	0.095	0.119	0.093
		EX	0.088	0.090	0.097	0.105	0.106
		EXNEX	0.089	0.090	0.097	0.105	0.106
		DMPP	0.080	0.080	0.091	0.106	0.089
		RDMPP	0.080	0.081	0.090	0.103	0.089
		EBPP	0.104	0.105	0.109	0.113	0.114
		HS	0.080	0.081	0.092	0.106	0.088
	H = 8	Current	0.122	0.122	0.122	0.121	0.123
		Pooled	0.074	0.076	0.091	0.113	0.076
		EX	0.080	0.082	0.091	0.101	0.093
		EXNEX	0.080	0.082	0.091	0.101	0.093
		DMPP	0.074	0.076	0.088	0.105	0.076
		RDMPP	0.075	0.076	0.087	0.104	0.076
		EBPP	0.109	0.110	0.113	0.116	0.116
		HS	0.074	0.076	0.090	0.106	0.075
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	0.075	0.076	0.075	0.075	0.075
		Pooled	0.049	0.054	0.080	0.107	0.079
		EX	0.055	0.058	0.067	0.071	0.072
		EXNEX	0.056	0.059	0.067	0.071	0.072
		DMPP	0.050	0.054	0.071	0.086	0.071
		RDMPP	0.051	0.054	0.069	0.081	0.068
		EBPP	0.064	0.066	0.069	0.072	0.070
		HS	0.050	0.054	0.069	0.077	0.056
	H = 8	Current	0.075	0.076	0.074	0.073	0.075
		Pooled	0.046	0.052	0.078	0.105	0.056
		EX	0.049	0.054	0.064	0.069	0.066
		EXNEX	0.049	0.054	0.064	0.069	0.066
		DMPP	0.046	0.051	0.072	0.092	0.054
		RDMPP	0.047	0.051	0.071	0.089	0.054
		EBPP	0.065	0.067	0.069	0.071	0.070
		HS	0.046	0.051	0.072	0.085	0.047

One Hetero, one heterogeneity historical control.

Table A.12: MPSD of the treatment effect in the simulation study with a binary endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{\text{CC}} = 20$, $n_{\text{CT}} = 40$, $n_h = 30$	H = 4	Current	0.123	0.123	0.122	0.120	0.123
		Pooled	0.079	0.079	0.078	0.078	0.079
		EX	0.101	0.101	0.105	0.109	0.113
		EXNEX	0.101	0.102	0.105	0.109	0.113
		DMPP	0.089	0.089	0.089	0.090	0.090
		RDMPP	0.091	0.091	0.091	0.093	0.093
		EBPP	0.086	0.086	0.087	0.087	0.085
		HS	0.087	0.087	0.090	0.094	0.096
	H = 8	Current	0.123	0.123	0.122	0.121	0.123
		Pooled	0.074	0.073	0.073	0.073	0.073
		EX	0.091	0.092	0.099	0.106	0.104
		EXNEX	0.092	0.093	0.099	0.106	0.105
		DMPP	0.081	0.081	0.081	0.081	0.081
		RDMPP	0.082	0.082	0.083	0.083	0.083
		EBPP	0.080	0.080	0.080	0.080	0.079
		HS	0.077	0.077	0.079	0.083	0.080
$n_{\text{CC}} = 60$, $n_{\text{CT}} = 120$, $n_h = 90$	H = 4	Current	0.076	0.076	0.076	0.075	0.076
		Pooled	0.050	0.050	0.049	0.049	0.049
		EX	0.062	0.063	0.068	0.071	0.074
		EXNEX	0.063	0.064	0.068	0.071	0.074
		DMPP	0.055	0.055	0.056	0.058	0.057
		RDMPP	0.056	0.056	0.058	0.060	0.060
		EBPP	0.054	0.055	0.055	0.055	0.054
		HS	0.054	0.055	0.060	0.064	0.061
	H = 8	Current	0.076	0.076	0.076	0.075	0.076
		Pooled	0.047	0.047	0.046	0.046	0.047
		EX	0.057	0.058	0.066	0.070	0.071
		EXNEX	0.057	0.058	0.066	0.070	0.071
		DMPP	0.051	0.051	0.051	0.052	0.051
		RDMPP	0.051	0.052	0.052	0.054	0.053
		EBPP	0.051	0.052	0.051	0.051	0.050
		HS	0.048	0.049	0.052	0.057	0.051

One Hetero, one heterogeneity historical control.

Table A.13: Type I error rate (%) of the treatment effect in the simulation study with a binary endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	4.78	5.52	4.73	4.70	4.85
		Pooled	4.81	5.30	8.55	14.87	15.65
		EX	3.69	4.39	4.04	4.84	4.79
		EXNEX	3.70	4.38	4.06	4.83	4.86
		DMPP	3.72	3.97	4.70	6.77	7.17
		RDMPP	3.61	4.00	4.33	5.77	6.03
		EBPP	5.80	6.57	6.49	7.49	7.73
		HS	3.82	4.23	4.37	5.54	5.29
	H = 3	Current	5.07	4.66	4.93	5.10	4.82
		Pooled	4.96	5.90	16.88	31.18	37.73
		EX	4.31	3.86	4.95	5.40	5.05
		EXNEX	4.26	3.89	4.92	5.36	5.05
		DMPP	4.02	4.25	7.21	9.22	11.99
		RDMPP	4.05	4.04	6.27	7.45	7.90
		EBPP	6.36	6.11	7.48	8.18	8.01
		HS	4.15	4.16	5.58	6.10	4.92
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	5.46	5.60	5.19	5.05	5.02
		Pooled	4.93	5.32	9.68	17.06	11.08
		EX	3.85	3.97	4.37	5.09	4.67
		EXNEX	3.86	3.96	4.41	5.09	4.68
		DMPP	3.91	4.06	5.54	8.07	6.13
		RDMPP	3.81	3.97	5.00	7.09	5.30
		EBPP	6.99	6.88	7.80	8.76	8.77
		HS	4.00	4.16	5.45	6.81	4.51
	H = 3	Current	4.57	4.95	4.96	4.76	4.67
		Pooled	4.70	6.47	17.89	33.34	25.28
		EX	3.66	4.24	4.79	4.97	4.80
		EXNEX	3.63	4.19	4.76	4.97	4.74
		DMPP	3.56	4.74	8.13	10.92	11.10
		RDMPP	3.50	4.52	7.01	8.72	7.88
		EBPP	5.99	6.83	8.05	7.85	8.91
		HS	3.57	4.62	6.61	6.57	4.24

One Hetero, one heterogeneity historical control.

Table A.14: Power (%) of the treatment effect in the simulation study with a binary endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	49.6	49.2	49.4	48.1	50.2
		Pooled	66.3	65.7	64.1	61.6	93.1
		EX	55.3	54.7	53.3	50.3	59.0
		EXNEX	55.3	54.4	53.2	50.3	58.8
		DMPP	60.1	60.2	59.1	56.0	79.8
		RDMPP	59.0	58.8	57.7	54.7	73.4
		EBPP	62.1	61.5	60.5	57.5	65.0
		HS	58.7	58.3	56.4	52.6	63.2
	H = 3	Current	51.0	49.7	50.0	48.1	50.2
		Pooled	67.8	66.0	62.5	59.6	98.2
		EX	56.4	55.3	52.7	49.1	54.8
		EXNEX	56.4	54.9	52.9	48.9	54.7
		DMPP	62.0	60.8	58.1	53.2	81.6
		RDMPP	60.9	59.7	56.8	52.0	71.0
		EBPP	63.1	62.0	59.7	56.2	65.3
		HS	59.9	58.6	54.5	50.1	58.6
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	50.4	49.0	50.1	48.5	49.3
		Pooled	69.8	69.2	67.1	62.8	89.9
		EX	58.9	57.3	55.2	52.1	59.2
		EXNEX	58.6	57.2	55.1	52.0	59.2
		DMPP	64.6	63.2	61.5	58.2	80.0
		RDMPP	63.6	61.9	60.2	57.0	76.1
		EBPP	65.6	63.9	62.8	60.2	68.0
		HS	64.4	62.8	60.2	55.4	67.4
	H = 3	Current	50.4	50.4	48.4	47.8	49.8
		Pooled	69.9	69.1	63.8	60.6	96.6
		EX	58.2	57.9	52.4	49.3	54.7
		EXNEX	58.0	57.7	52.4	49.2	54.5
		DMPP	64.2	64.0	59.4	54.9	84.8
		RDMPP	62.9	62.8	58.3	53.8	76.2
		EBPP	64.6	64.1	60.7	58.4	67.2
		HS	63.6	62.6	56.5	51.7	62.3

One Hetero, one heterogeneity historical control.

Table A.15: Calibrated power (%) of the treatment effect in the simulation study with a binary endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	53.3	47.6	51.9	49.6	53.3
		Pooled	66.7	65.0	55.1	39.4	78.7
		EX	59.8	57.3	57.2	51.5	59.8
		EXNEX	59.8	57.4	56.9	51.6	59.4
		DMPP	65.1	63.5	60.7	51.5	73.6
		RDMPP	64.5	62.6	60.9	52.6	70.2
		EBPP	59.2	56.5	55.9	51.1	57.7
		HS	63.3	60.8	59.2	51.3	62.5
	H = 3	Current	50.6	50.9	50.2	47.6	50.9
		Pooled	67.6	62.9	37.9	NA	78.3
		EX	58.9	59.1	52.8	47.9	54.5
		EXNEX	58.9	59.3	53.6	47.9	54.5
		DMPP	65.4	63.6	51.8	43.3	67.4
		RDMPP	64.1	62.8	52.9	45.2	63.2
		EBPP	59.8	58.8	53.0	47.7	58.0
		HS	62.9	61.9	53.3	47.4	58.9
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	49.0	47.4	48.9	48.4	49.2
		Pooled	70.0	67.9	55.1	38.2	81.3
		EX	63.1	60.7	57.3	51.8	60.3
		EXNEX	63.1	60.7	57.3	51.7	60.4
		DMPP	68.7	66.4	60.1	48.9	77.9
		RDMPP	68.1	65.8	60.3	50.2	75.3
		EBPP	59.3	57.5	55.0	50.8	58.8
		HS	68.1	65.7	58.8	49.8	69.8
	H = 3	Current	51.7	50.5	48.6	48.7	50.8
		Pooled	71.2	64.2	38.9	NA	79.6
		EX	64.2	61.2	53.0	49.4	55.4
		EXNEX	64.0	61.0	53.0	49.3	55.4
		DMPP	69.1	64.8	50.8	42.2	71.6
		RDMPP	68.5	65.0	52.2	44.7	68.3
		EBPP	61.9	59.0	52.1	50.3	57.1
		HS	68.5	64.5	51.8	46.9	64.7

One Hetero, one heterogeneity historical control.

Table A.16: Average Bias of the treatment effect in the simulation study with a binary endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	−0.016	−0.018	−0.018	−0.026	−0.014
		Pooled	−0.016	−0.017	−0.018	−0.026	0.084
		EX	−0.016	−0.018	−0.018	−0.026	0.011
		EXNEX	−0.016	−0.018	−0.018	−0.026	0.011
		DMPP	−0.016	−0.017	−0.018	−0.026	0.049
		RDMPP	−0.016	−0.018	−0.018	−0.026	0.038
		EBPP	−0.016	−0.018	−0.018	−0.026	0.004
		HS	−0.016	−0.018	−0.018	−0.026	0.018
	H = 3	Current	−0.002	−0.004	−0.005	−0.011	−0.003
		Pooled	−0.002	−0.004	−0.005	−0.011	0.096
		EX	−0.002	−0.004	−0.005	−0.011	0.006
		EXNEX	−0.002	−0.004	−0.005	−0.011	0.006
		DMPP	−0.002	−0.004	−0.005	−0.011	0.052
		RDMPP	−0.002	−0.004	−0.005	−0.011	0.034
		EBPP	−0.002	−0.004	−0.005	−0.011	0.008
		HS	−0.002	−0.004	−0.005	−0.011	0.007
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	−0.014	−0.017	−0.018	−0.026	−0.016
		Pooled	−0.014	−0.017	−0.019	−0.027	0.059
		EX	−0.014	−0.017	−0.018	−0.026	0.006
		EXNEX	−0.014	−0.017	−0.018	−0.026	0.006
		DMPP	−0.014	−0.017	−0.018	−0.026	0.038
		RDMPP	−0.014	−0.017	−0.018	−0.026	0.031
		EBPP	−0.014	−0.017	−0.018	−0.026	−0.002
		HS	−0.014	−0.017	−0.018	−0.026	0.012
	H = 3	Current	−0.002	−0.003	−0.007	−0.011	−0.004
		Pooled	−0.002	−0.003	−0.006	−0.011	0.072
		EX	−0.002	−0.003	−0.007	−0.011	0.004
		EXNEX	−0.002	−0.003	−0.007	−0.011	0.004
		DMPP	−0.002	−0.003	−0.007	−0.011	0.047
		RDMPP	−0.002	−0.003	−0.007	−0.011	0.035
		EBPP	−0.002	−0.003	−0.007	−0.011	0.005
		HS	−0.002	−0.003	−0.007	−0.011	0.006

One Hetero, one heterogeneity historical control.

Table A.17: RMSD of the treatment effect in the simulation study with a binary endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	0.115	0.115	0.116	0.118	0.114
		Pooled	0.092	0.094	0.103	0.119	0.122
		EX	0.102	0.103	0.105	0.111	0.111
		EXNEX	0.102	0.103	0.106	0.111	0.111
		DMPP	0.094	0.095	0.099	0.105	0.107
		RDMPP	0.095	0.097	0.099	0.105	0.106
		EBPP	0.104	0.105	0.108	0.113	0.110
		HS	0.097	0.098	0.102	0.109	0.109
	H = 3	Current	0.072	0.072	0.071	0.071	0.072
		Pooled	0.058	0.061	0.079	0.106	0.112
		EX	0.064	0.065	0.068	0.070	0.072
		EXNEX	0.064	0.065	0.068	0.070	0.072
		DMPP	0.059	0.061	0.068	0.074	0.083
		RDMPP	0.060	0.061	0.067	0.071	0.077
		EBPP	0.064	0.065	0.068	0.070	0.069
		HS	0.061	0.062	0.067	0.070	0.068
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	0.115	0.115	0.116	0.117	0.114
		Pooled	0.088	0.089	0.100	0.119	0.104
		EX	0.097	0.097	0.102	0.108	0.106
		EXNEX	0.098	0.098	0.102	0.108	0.107
		DMPP	0.090	0.090	0.096	0.105	0.097
		RDMPP	0.091	0.091	0.096	0.103	0.097
		EBPP	0.104	0.104	0.108	0.111	0.108
		HS	0.091	0.091	0.098	0.106	0.099
	H = 3	Current	0.071	0.071	0.071	0.072	0.072
		Pooled	0.056	0.059	0.080	0.109	0.091
		EX	0.061	0.062	0.067	0.070	0.071
		EXNEX	0.061	0.062	0.067	0.070	0.071
		DMPP	0.057	0.059	0.068	0.077	0.076
		RDMPP	0.058	0.059	0.067	0.074	0.072
		EBPP	0.064	0.064	0.068	0.070	0.068
		HS	0.058	0.059	0.067	0.072	0.063

One Hetero, one heterogeneity historical control.

Table A.18: MPSD of the treatment effect in the simulation study with a binary endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	0.115	0.115	0.114	0.112	0.115
		Pooled	0.092	0.092	0.092	0.091	0.091
		EX	0.107	0.107	0.108	0.109	0.113
		EXNEX	0.107	0.107	0.108	0.109	0.113
		DMPP	0.100	0.100	0.100	0.101	0.102
		RDMPP	0.101	0.101	0.102	0.103	0.105
		EBPP	0.096	0.097	0.097	0.098	0.098
		HS	0.102	0.102	0.104	0.106	0.110
	H = 3	Current	0.071	0.071	0.071	0.070	0.071
		Pooled	0.058	0.058	0.058	0.057	0.058
		EX	0.067	0.067	0.068	0.069	0.071
		EXNEX	0.067	0.067	0.068	0.069	0.071
		DMPP	0.062	0.063	0.064	0.065	0.067
		RDMPP	0.063	0.064	0.065	0.066	0.070
		EBPP	0.061	0.061	0.062	0.062	0.062
		HS	0.064	0.064	0.067	0.068	0.069
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	0.115	0.115	0.114	0.113	0.115
		Pooled	0.088	0.088	0.088	0.087	0.088
		EX	0.104	0.104	0.106	0.107	0.111
		EXNEX	0.104	0.104	0.106	0.108	0.111
		DMPP	0.096	0.096	0.097	0.098	0.098
		RDMPP	0.098	0.098	0.099	0.100	0.100
		EBPP	0.093	0.093	0.094	0.094	0.093
		HS	0.096	0.097	0.099	0.102	0.104
	H = 3	Current	0.071	0.071	0.071	0.070	0.071
		Pooled	0.056	0.056	0.056	0.055	0.056
		EX	0.065	0.065	0.068	0.068	0.071
		EXNEX	0.065	0.065	0.068	0.068	0.071
		DMPP	0.060	0.060	0.062	0.063	0.063
		RDMPP	0.061	0.061	0.063	0.064	0.066
		EBPP	0.059	0.060	0.060	0.060	0.059
		HS	0.061	0.061	0.064	0.066	0.066

One Hetero, one heterogeneity historical control.

Table A.19: Type I error rate (%) of the treatment effect in the simulation study with a binary endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	5.53	5.05	5.16	4.58	5.26
		Pooled	5.09	5.03	10.37	18.89	21.27
		EX	3.63	3.41	4.14	4.35	5.36
		EXNEX	3.71	3.42	4.11	4.35	5.25
		DMPP	3.52	3.35	5.41	7.59	8.91
		RDMPP	3.38	3.28	4.71	6.32	6.91
		EBPP	7.41	6.74	8.58	8.86	11.12
		HS	3.69	3.58	4.85	5.73	5.80
	H = 3	Current	4.58	5.03	5.02	4.83	4.95
		Pooled	4.76	6.75	21.40	38.73	52.32
		EX	3.07	3.68	4.79	5.23	5.26
		EXNEX	3.09	3.69	4.84	5.15	5.29
		DMPP	3.10	4.18	9.05	12.72	16.25
		RDMPP	2.99	3.91	7.31	9.24	9.73
		EBPP	6.48	7.70	9.69	10.11	10.53
		HS	3.01	3.88	6.02	6.53	4.63
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	5.18	5.42	4.87	4.47	5.64
		Pooled	4.80	5.71	11.29	22.15	14.13
		EX	2.92	3.37	4.13	4.55	4.81
		EXNEX	2.94	3.29	4.16	4.54	4.87
		DMPP	3.02	3.58	5.95	9.47	7.15
		RDMPP	2.87	3.33	5.39	7.75	6.00
		EBPP	8.12	8.57	9.78	10.29	12.25
		HS	3.26	3.83	5.75	7.58	4.87
	H = 3	Current	5.15	5.22	5.22	4.97	4.94
		Pooled	5.65	7.38	23.58	41.13	33.98
		EX	3.37	3.58	5.08	5.21	4.72
		EXNEX	3.38	3.57	5.02	5.19	4.71
		DMPP	3.56	4.41	10.93	15.90	14.48
		RDMPP	3.38	4.14	8.88	11.99	9.50
		EBPP	7.93	8.56	10.79	11.41	12.02
		HS	3.65	4.29	7.76	8.35	3.80

One Hetero, one heterogeneity historical control.

Table A.20: Power (%) of the treatment effect in the simulation study with a binary endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	47.9	46.2	46.3	44.6	47.2
		Pooled	74.1	73.7	69.6	66.6	96.9
		EX	57.2	56.0	53.1	49.8	59.3
		EXNEX	57.1	55.9	52.8	49.8	59.2
		DMPP	66.3	65.6	62.6	60.0	85.1
		RDMPP	64.3	63.7	60.4	57.4	78.4
		EBPP	68.1	66.5	64.1	61.1	70.8
		HS	63.3	62.4	58.2	53.8	65.7
	H = 3	Current	45.5	46.9	46.1	43.5	46.1
		Pooled	73.5	73.0	66.8	63.1	99.7
		EX	54.9	54.9	50.4	46.5	53.3
		EXNEX	54.6	54.9	50.2	46.3	53.4
		DMPP	65.1	65.1	60.4	55.1	86.7
		RDMPP	63.0	63.3	58.2	52.6	74.7
		EBPP	66.8	66.4	62.7	58.1	70.4
		HS	61.2	60.8	54.3	48.9	58.9
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	47.4	46.9	47.0	45.0	47.0
		Pooled	78.1	78.5	73.9	69.5	95.8
		EX	60.6	59.9	57.0	51.7	60.2
		EXNEX	60.4	59.8	56.7	51.6	59.8
		DMPP	70.6	70.7	67.6	63.5	86.6
		RDMPP	68.7	68.5	65.9	61.4	82.1
		EBPP	70.4	69.8	67.6	64.1	71.7
		HS	70.0	69.4	64.9	58.7	71.5
	H = 3	Current	46.0	46.6	45.0	44.8	46.1
		Pooled	78.0	76.3	67.8	65.3	99.0
		EX	59.4	58.7	51.2	48.1	53.3
		EXNEX	59.1	58.4	51.1	47.9	53.2
		DMPP	69.9	69.2	62.6	57.8	90.4
		RDMPP	67.7	67.4	60.7	56.2	82.4
		EBPP	69.3	69.4	64.1	62.2	72.2
		HS	68.3	67.3	57.5	52.1	65.2

One Hetero, one heterogeneity historical control.

Table A.21: Calibrated power (%) of the treatment effect in the simulation study with a binary endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	46.7	46.2	46.2	46.7	47.1
		Pooled	73.4	73.6	57.0	38.7	84.5
		EX	62.3	62.8	56.5	52.1	58.2
		EXNEX	62.1	62.5	56.5	52.4	58.0
		DMPP	71.6	71.8	61.5	51.9	77.3
		RDMPP	70.5	70.7	61.8	53.4	73.3
		EBPP	61.6	62.7	55.7	50.3	56.7
		HS	68.7	68.5	59.0	51.2	63.0
	H = 3	Current	47.1	46.7	46.0	44.3	46.3
		Pooled	74.5	68.5	36.6	NA	NA
		EX	62.3	59.2	51.1	45.9	52.4
		EXNEX	62.0	59.2	50.8	45.9	52.7
		DMPP	71.5	68.2	49.3	38.8	68.5
		RDMPP	70.4	67.5	50.9	41.7	63.9
		EBPP	63.6	60.3	50.5	44.5	56.2
		HS	68.1	65.3	51.2	43.6	60.4
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	46.9	45.7	47.2	47.0	45.8
		Pooled	79.0	76.8	59.3	38.6	87.0
		EX	68.7	66.5	60.3	53.0	60.7
		EXNEX	68.5	66.2	60.0	53.2	60.4
		DMPP	77.1	75.3	64.2	52.4	82.0
		RDMPP	76.4	74.1	64.6	54.2	79.3
		EBPP	64.3	61.5	56.5	52.3	56.4
		HS	75.7	73.4	62.2	51.5	71.9
	H = 3	Current	45.7	45.8	44.5	44.7	46.4
		Pooled	76.3	70.8	37.2	NA	85.2
		EX	66.0	63.8	51.2	47.2	54.6
		EXNEX	66.0	63.4	51.4	47.0	54.5
		DMPP	74.6	71.4	48.9	38.2	76.5
		RDMPP	73.8	70.8	50.1	40.1	72.1
		EBPP	62.7	60.9	49.6	46.5	56.1
		HS	73.5	69.8	50.7	42.5	69.3

One Hetero, one heterogeneity historical control.

Table A.22: Average bias of the treatment effect in the simulation study with a binary endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	−0.010	−0.013	−0.017	−0.023	−0.012
		Pooled	−0.012	−0.012	−0.017	−0.022	0.098
		EX	−0.011	−0.013	−0.017	−0.023	0.021
		EXNEX	−0.011	−0.013	−0.017	−0.023	0.020
		DMPP	−0.011	−0.013	−0.017	−0.022	0.064
		RDMPP	−0.011	−0.013	−0.017	−0.022	0.052
		EBPP	−0.010	−0.013	−0.017	−0.023	0.011
		HS	−0.011	−0.013	−0.017	−0.023	0.030
	H = 3	Current	−0.003	−0.001	−0.005	−0.011	−0.003
		Pooled	−0.003	−0.003	−0.005	−0.010	0.109
		EX	−0.003	−0.002	−0.005	−0.011	0.010
		EXNEX	−0.003	−0.002	−0.005	−0.011	0.010
		DMPP	−0.003	−0.002	−0.005	−0.010	0.068
		RDMPP	−0.003	−0.002	−0.005	−0.010	0.047
		EBPP	−0.003	−0.002	−0.005	−0.011	0.012
		HS	−0.003	−0.002	−0.005	−0.010	0.012
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	−0.011	−0.012	−0.015	−0.022	−0.012
		Pooled	−0.012	−0.012	−0.014	−0.021	0.068
		EX	−0.012	−0.012	−0.015	−0.022	0.015
		EXNEX	−0.012	−0.012	−0.015	−0.022	0.014
		DMPP	−0.012	−0.012	−0.015	−0.022	0.049
		RDMPP	−0.012	−0.012	−0.015	−0.022	0.042
		EBPP	−0.011	−0.012	−0.015	−0.022	0.005
		HS	−0.012	−0.012	−0.015	−0.022	0.021
	H = 3	Current	−0.003	−0.002	−0.006	−0.008	−0.002
		Pooled	−0.002	−0.002	−0.006	−0.010	0.079
		EX	−0.003	−0.002	−0.006	−0.008	0.009
		EXNEX	−0.003	−0.002	−0.006	−0.008	0.009
		DMPP	−0.003	−0.002	−0.006	−0.008	0.058
		RDMPP	−0.003	−0.002	−0.006	−0.008	0.046
		EBPP	−0.003	−0.002	−0.006	−0.008	0.009
		HS	−0.003	−0.002	−0.006	−0.008	0.010

One Hetero, one heterogeneity historical control.

Table A.23: RMSD of the treatment effect in the simulation study with a binary endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	0.121	0.122	0.123	0.123	0.122
		Pooled	0.087	0.087	0.102	0.122	0.128
		EX	0.101	0.102	0.107	0.112	0.117
		EXNEX	0.101	0.102	0.107	0.112	0.117
		DMPP	0.089	0.089	0.097	0.106	0.110
		RDMPP	0.090	0.091	0.097	0.105	0.109
		EBPP	0.105	0.106	0.111	0.116	0.116
		HS	0.093	0.094	0.102	0.110	0.114
	H = 3	Current	0.076	0.076	0.075	0.075	0.076
		Pooled	0.054	0.058	0.081	0.114	0.121
		EX	0.064	0.065	0.069	0.073	0.076
		EXNEX	0.064	0.065	0.069	0.072	0.076
		DMPP	0.056	0.058	0.069	0.081	0.092
		RDMPP	0.057	0.058	0.067	0.076	0.083
		EBPP	0.065	0.066	0.070	0.073	0.072
		HS	0.059	0.060	0.068	0.074	0.071
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	0.122	0.122	0.121	0.123	0.122
		Pooled	0.083	0.082	0.098	0.121	0.104
		EX	0.095	0.095	0.101	0.109	0.110
		EXNEX	0.095	0.095	0.101	0.109	0.111
		DMPP	0.085	0.084	0.093	0.106	0.097
		RDMPP	0.086	0.085	0.093	0.104	0.096
		EBPP	0.105	0.106	0.109	0.115	0.114
		HS	0.086	0.085	0.095	0.107	0.098
	H = 3	Current	0.076	0.076	0.075	0.076	0.076
		Pooled	0.052	0.055	0.081	0.114	0.094
		EX	0.059	0.061	0.068	0.073	0.074
		EXNEX	0.060	0.061	0.068	0.073	0.074
		DMPP	0.053	0.055	0.070	0.084	0.080
		RDMPP	0.054	0.055	0.068	0.079	0.075
		EBPP	0.064	0.066	0.069	0.073	0.071
		HS	0.054	0.056	0.068	0.075	0.062

One Hetero, one heterogeneity historical control.

Table A.24: MPSD of the treatment effect in the simulation study with a binary endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	0.122	0.122	0.121	0.119	0.122
		Pooled	0.086	0.086	0.086	0.085	0.085
		EX	0.109	0.110	0.111	0.113	0.118
		EXNEX	0.109	0.110	0.111	0.113	0.118
		DMPP	0.098	0.098	0.098	0.099	0.100
		RDMPP	0.100	0.100	0.101	0.103	0.106
		EBPP	0.093	0.094	0.094	0.095	0.095
		HS	0.101	0.102	0.104	0.107	0.114
	H = 3	Current	0.076	0.076	0.075	0.074	0.076
		Pooled	0.054	0.054	0.053	0.053	0.053
		EX	0.068	0.069	0.071	0.072	0.075
		EXNEX	0.068	0.069	0.071	0.072	0.075
		DMPP	0.060	0.061	0.062	0.064	0.066
		RDMPP	0.062	0.062	0.064	0.066	0.071
		EBPP	0.058	0.058	0.060	0.061	0.059
		HS	0.063	0.064	0.067	0.070	0.072
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	0.122	0.122	0.121	0.119	0.122
		Pooled	0.081	0.081	0.081	0.081	0.081
		EX	0.104	0.105	0.107	0.110	0.115
		EXNEX	0.105	0.105	0.107	0.111	0.115
		DMPP	0.092	0.092	0.093	0.094	0.094
		RDMPP	0.095	0.095	0.096	0.097	0.098
		EBPP	0.089	0.089	0.090	0.090	0.088
		HS	0.093	0.093	0.096	0.101	0.104
	H = 3	Current	0.076	0.076	0.075	0.074	0.076
		Pooled	0.051	0.051	0.051	0.050	0.051
		EX	0.065	0.066	0.069	0.071	0.075
		EXNEX	0.065	0.066	0.070	0.071	0.075
		DMPP	0.057	0.057	0.059	0.062	0.060
		RDMPP	0.059	0.059	0.061	0.064	0.065
		EBPP	0.056	0.056	0.057	0.058	0.056
		HS	0.058	0.059	0.063	0.067	0.066

One Hetero, one heterogeneity historical control.

A.2 Time-to-event endpoint

Table A.25: Type I error rate (%) of the treatment effect in the simulation study with a time-to-event endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	5.36	4.77	5.11	4.99	4.90
		Pooled	5.45	6.60	20.54	39.53	6.06
		EX	4.11	3.78	5.45	5.53	3.89
		EXNEX	4.19	3.78	5.29	5.47	3.87
		HS	4.64	4.57	8.02	8.72	4.29
	H = 8	Current	4.95	4.99	5.24	5.10	4.78
		Pooled	5.00	7.52	23.61	42.23	5.92
		EX	3.27	4.28	5.63	5.29	3.79
		EXNEX	3.35	4.14	5.19	5.15	3.87
		HS	4.32	6.27	13.60	14.93	4.94
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	5.04	5.08	4.89	5.01	5.01
		Pooled	5.63	11.41	38.19	57.76	8.29
		EX	4.01	4.65	4.95	5.07	4.36
		EXNEX	4.14	4.69	4.87	5.03	4.30
		HS	4.52	6.29	8.25	6.59	4.73
	H = 8	Current	5.20	5.42	5.15	4.91	5.01
		Pooled	4.96	11.99	42.11	62.48	6.21
		EX	3.59	5.24	5.11	5.00	3.51
		EXNEX	3.54	5.16	5.04	4.90	3.49
		HS	4.40	9.19	14.42	10.92	4.82

One Hetero, one heterogeneity historical control.

Table A.26: Power (%) of the treatment effect in the simulation study with a time-to-event endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	51.0	52.0	51.1	51.2	50.7
		Pooled	77.6	75.6	65.1	60.5	85.0
		EX	66.8	65.5	58.2	54.4	68.7
		EXNEX	65.7	64.6	57.5	54.0	67.3
		HS	73.2	71.1	62.6	57.0	76.9
	H = 8	Current	53.2	51.3	51.4	51.9	52.2
		Pooled	81.9	78.8	67.2	61.1	85.8
		EX	73.5	69.8	59.3	55.2	74.8
		EXNEX	71.9	68.1	58.4	54.7	72.8
		HS	80.5	77.2	67.1	60.6	83.5
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	50.6	50.6	50.1	50.6	50.8
		Pooled	74.5	70.5	61.1	65.2	87.4
		EX	63.4	59.6	52.4	51.7	63.5
		EXNEX	63.0	59.2	52.4	51.6	62.9
		HS	69.6	65.0	55.1	53.2	72.7
	H = 8	Current	50.7	50.4	50.3	49.4	50.6
		Pooled	78.1	72.5	62.6	66.2	85.1
		EX	68.8	62.7	53.4	50.7	67.7
		EXNEX	67.7	61.9	52.8	50.6	66.9
		HS	76.3	70.7	59.3	53.7	79.4

One Hetero, one heterogeneity historical control.

Table A.27: Calibrated power (%) of the treatment effect in the simulation study with a time-to-event endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 4	Current	50.0	52.6	51.0	52.0	51.3
		Pooled	76.5	71.3	41.9	NA	83.0
		EX	69.9	69.0	55.9	54.9	72.2
		EXNEX	68.8	68.1	56.3	55.5	70.9
		HS	74.7	71.6	54.4	48.7	78.8
	H = 8	Current	53.4	51.5	50.9	51.4	53.6
		Pooled	82.2	72.9	39.5	NA	84.2
		EX	78.6	70.9	57.0	53.7	78.3
		EXNEX	77.1	70.0	57.0	53.9	76.6
		HS	81.6	73.0	49.8	41.9	83.5
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 4	Current	50.4	50.9	50.8	50.9	50.4
		Pooled	73.0	58.7	NA	NA	81.5
		EX	66.5	61.8	51.6	51.5	64.7
		EXNEX	65.8	61.3	51.7	51.8	64.2
		HS	71.3	62.1	46.9	47.3	73.3
	H = 8	Current	50.0	49.3	50.3	49.6	50.5
		Pooled	78.1	58.6	NA	NA	82.6
		EX	73.2	62.3	53.1	51.1	71.8
		EXNEX	72.4	61.3	53.4	51.0	70.8
		HS	77.6	60.9	40.4	39.0	79.7

One Hetero, one heterogeneity historical control.

Table A.28: Average bias of the treatment effect in the simulation study with a time-to-event endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{\text{CC}} = 30$, $n_{\text{CT}} = 30$, $n_h = 30$	H = 4	Current	−0.002	0.002	−0.004	−0.004	−0.008
		Pooled	0.003	0.001	−0.031	−0.094	0.062
		EX	0.006	0.008	−0.002	−0.004	0.035
		EXNEX	0.006	0.008	−0.001	−0.004	0.032
		HS	0.006	0.006	−0.005	−0.010	0.048
	H = 8	Current	0.008	−0.003	−0.002	−0.001	0.004
		Pooled	0.011	0.001	−0.032	−0.104	0.043
		EX	0.014	0.005	0.000	−0.001	0.035
		EXNEX	0.014	0.005	0.001	0.000	0.033
		HS	0.013	0.004	−0.005	−0.017	0.041
$n_{\text{CC}} = 90$, $n_{\text{CT}} = 90$, $n_h = 90$	H = 4	Current	0.000	0.001	−0.002	0.000	−0.001
		Pooled	0.000	−0.001	−0.034	−0.084	0.061
		EX	0.002	0.002	−0.002	0.000	0.024
		EXNEX	0.002	0.002	−0.002	0.000	0.023
		HS	0.002	0.002	−0.005	0.000	0.031
	H = 8	Current	−0.002	−0.001	−0.002	−0.005	−0.001
		Pooled	0.001	−0.001	−0.035	−0.115	0.034
		EX	0.002	0.001	−0.001	−0.005	0.020
		EXNEX	0.002	0.001	−0.001	−0.005	0.019
		HS	0.002	0.002	−0.005	−0.010	0.021

One Hetero, one heterogeneity historical control.

Table A.29: RMSD of the treatment effect in the simulation study with a time-to-event endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{\text{CC}} = 30$, $n_{\text{CT}} = 30$, $n_h = 30$	H = 4	Current	0.295	0.291	0.294	0.291	0.291
		Pooled	0.228	0.241	0.347	0.514	0.233
		EX	0.243	0.246	0.277	0.288	0.247
		EXNEX	0.245	0.248	0.277	0.287	0.250
		HS	0.232	0.240	0.289	0.313	0.236
	H = 8	Current	0.291	0.292	0.295	0.293	0.289
		Pooled	0.217	0.235	0.359	0.529	0.222
		EX	0.227	0.238	0.277	0.286	0.232
		EXNEX	0.230	0.240	0.277	0.285	0.235
		HS	0.218	0.234	0.315	0.353	0.222
$n_{\text{CC}} = 90$, $n_{\text{CT}} = 90$, $n_h = 90$	H = 4	Current	0.168	0.168	0.168	0.165	0.168
		Pooled	0.132	0.159	0.293	0.461	0.144
		EX	0.141	0.150	0.165	0.164	0.152
		EXNEX	0.142	0.151	0.165	0.164	0.152
		HS	0.134	0.150	0.176	0.174	0.142
	H = 8	Current	0.169	0.170	0.168	0.168	0.167
		Pooled	0.125	0.158	0.302	0.503	0.129
		EX	0.132	0.148	0.164	0.167	0.140
		EXNEX	0.133	0.149	0.165	0.166	0.141
		HS	0.126	0.153	0.199	0.198	0.128

One Hetero, one heterogeneity historical control.

Table A.30: MPSD of the treatment effect in the simulation study with a time-to-event endpoint.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{\text{CC}} = 30$, $n_{\text{CT}} = 30$, $n_h = 30$	H = 4	Current	0.293	0.292	0.293	0.293	0.293
		Pooled	0.225	0.225	0.225	0.225	0.225
		EX	0.256	0.259	0.273	0.282	0.261
		EXNEX	0.259	0.262	0.276	0.283	0.263
		HS	0.240	0.242	0.257	0.272	0.244
	H = 8	Current	0.293	0.293	0.293	0.293	0.293
		Pooled	0.217	0.217	0.217	0.217	0.217
		EX	0.243	0.248	0.270	0.282	0.248
		EXNEX	0.247	0.252	0.273	0.283	0.251
		HS	0.223	0.225	0.237	0.256	0.225
$n_{\text{CC}} = 90$, $n_{\text{CT}} = 90$, $n_h = 90$	H = 4	Current	0.167	0.167	0.167	0.167	0.167
		Pooled	0.129	0.129	0.129	0.129	0.129
		EX	0.148	0.153	0.163	0.165	0.155
		EXNEX	0.149	0.154	0.163	0.166	0.156
		HS	0.138	0.142	0.157	0.163	0.144
	H = 8	Current	0.167	0.167	0.167	0.167	0.167
		Pooled	0.124	0.124	0.124	0.124	0.124
		EX	0.140	0.147	0.162	0.165	0.148
		EXNEX	0.141	0.149	0.163	0.166	0.149
		HS	0.128	0.131	0.148	0.159	0.130

One Hetero, one heterogeneity historical control.

Table A.31: Type I error rate (%) of the treatment effect in the simulation study with a time-to-event endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	4.93	5.16	5.35	4.98	4.76
		Pooled	5.05	8.50	26.20	45.11	6.39
		EX	2.72	3.75	5.85	5.73	3.54
		EXNEX	2.76	3.66	5.46	5.24	3.36
		HS	3.61	5.44	10.35	11.37	4.12
	H = 8	Current	5.21	4.93	5.45	4.86	4.74
		Pooled	5.19	7.96	28.74	49.04	5.78
		EX	3.22	3.65	6.02	5.54	3.13
		EXNEX	3.07	3.61	5.57	5.25	3.00
		HS	4.38	6.32	18.30	21.63	4.79
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	4.88	4.53	5.07	5.13	5.30
		Pooled	4.96	13.17	45.58	65.68	9.17
		EX	2.96	4.14	6.04	5.35	4.06
		EXNEX	2.92	4.12	5.68	5.35	4.11
		HS	3.42	6.70	10.77	9.25	4.33
	H = 8	Current	5.17	5.24	4.97	5.01	4.80
		Pooled	5.15	15.14	49.37	67.79	6.43
		EX	2.97	5.04	5.43	5.14	3.10
		EXNEX	2.88	4.91	5.18	4.99	3.06
		HS	4.32	11.28	21.22	16.08	4.50

One Hetero, one heterogeneity historical control.

Table A.32: Power (%) of the treatment effect in the simulation study with a time-to-event endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	44.0	44.1	45.1	44.2	43.5
		Pooled	84.9	81.7	69.5	64.1	90.9
		EX	69.0	66.6	56.9	50.5	70.5
		EXNEX	66.9	64.5	55.7	49.6	68.2
		HS	79.1	76.1	64.5	56.1	82.9
	H = 8	Current	44.0	43.9	44.0	43.1	43.8
		Pooled	88.4	85.4	70.9	64.0	91.9
		EX	77.8	74.1	58.0	50.1	77.4
		EXNEX	74.5	71.1	56.2	49.0	74.3
		HS	87.0	83.8	70.3	60.5	89.7
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	45.1	44.8	42.8	43.3	44.6
		Pooled	82.5	76.1	64.9	69.2	93.4
		EX	65.5	60.5	47.7	45.5	64.2
		EXNEX	64.3	59.4	47.1	45.3	62.9
		HS	76.1	69.8	53.4	48.9	78.7
	H = 8	Current	44.7	44.3	43.9	45.0	45.0
		Pooled	87.1	79.2	66.3	70.4	92.7
		EX	74.7	65.5	49.6	46.6	71.2
		EXNEX	72.7	63.7	48.8	46.5	69.3
		HS	85.4	77.2	60.0	52.0	87.7

One Hetero, one heterogeneity historical control.

Table A.33: Calibrated power (%) of the treatment effect in the simulation study with a time-to-event endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	44.5	43.8	44.5	44.1	44.3
		Pooled	84.9	76.6	40.8	NA	89.4
		EX	76.2	71.6	56.3	48.5	76.6
		EXNEX	74.2	70.0	56.1	48.5	74.3
		HS	82.8	75.9	53.2	42.2	84.8
	H = 8	Current	43.6	43.9	42.7	43.1	44.8
		Pooled	88.3	80.6	NA	NA	90.9
		EX	83.0	77.4	57.2	48.7	83.0
		EXNEX	80.8	75.4	56.3	48.6	80.1
		HS	88.0	81.3	48.9	38.6	90.1
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	45.3	47.0	43.1	44.0	44.7
		Pooled	82.5	61.1	NA	NA	89.1
		EX	73.5	63.5	46.3	44.6	67.5
		EXNEX	72.4	63.1	46.3	45.1	66.7
		HS	79.9	66.2	41.1	39.2	80.5
	H = 8	Current	44.1	43.5	45.2	45.1	45.6
		Pooled	86.7	64.4	NA	NA	91.0
		EX	81.1	67.5	48.2	46.4	77.0
		EXNEX	79.6	66.4	48.9	46.7	75.3
		HS	86.5	67.6	34.7	34.0	88.4

One Hetero, one heterogeneity historical control.

Table A.34: Average bias of the treatment effect in the simulation study with a time-to-event endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	−0.009	−0.009	−0.004	−0.007	−0.012
		Pooled	0.003	0.000	−0.027	−0.101	0.067
		EX	0.006	0.004	0.001	−0.008	0.043
		EXNEX	0.005	0.004	0.002	−0.005	0.039
		HS	0.005	0.004	−0.003	−0.017	0.055
	H = 8	Current	−0.008	−0.009	−0.009	−0.012	−0.013
		Pooled	0.003	−0.004	−0.042	−0.126	0.038
		EX	0.006	0.002	−0.007	−0.013	0.030
		EXNEX	0.006	0.003	−0.006	−0.012	0.027
		HS	0.005	0.001	−0.014	−0.033	0.036
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	−0.001	−0.002	−0.008	−0.007	0.001
		Pooled	0.001	−0.002	−0.038	−0.098	0.067
		EX	0.002	0.002	−0.008	−0.006	0.032
		EXNEX	0.002	0.001	−0.007	−0.006	0.030
		HS	0.002	0.001	−0.010	−0.009	0.039
	H = 8	Current	−0.001	−0.001	−0.003	0.001	−0.002
		Pooled	0.002	−0.003	−0.039	−0.125	0.037
		EX	0.003	0.001	−0.003	−0.001	0.024
		EXNEX	0.003	0.001	−0.003	−0.001	0.022
		HS	0.003	0.000	−0.010	−0.015	0.024

One Hetero, one heterogeneity historical control.

Table A.35: RMSD of the treatment effect in the simulation study with a time-to-event endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	0.309	0.311	0.313	0.312	0.309
		Pooled	0.201	0.228	0.352	0.533	0.214
		EX	0.221	0.236	0.282	0.301	0.234
		EXNEX	0.227	0.240	0.282	0.300	0.239
		HS	0.206	0.226	0.297	0.336	0.218
	H = 8	Current	0.313	0.313	0.312	0.313	0.313
		Pooled	0.193	0.214	0.354	0.542	0.197
		EX	0.204	0.219	0.275	0.298	0.212
		EXNEX	0.211	0.224	0.276	0.297	0.218
		HS	0.193	0.212	0.318	0.387	0.198
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	0.177	0.176	0.179	0.178	0.179
		Pooled	0.116	0.150	0.308	0.494	0.135
		EX	0.130	0.144	0.173	0.176	0.149
		EXNEX	0.132	0.146	0.173	0.176	0.150
		HS	0.119	0.142	0.189	0.194	0.133
	H = 8	Current	0.181	0.177	0.179	0.177	0.177
		Pooled	0.110	0.149	0.311	0.519	0.116
		EX	0.118	0.140	0.171	0.174	0.130
		EXNEX	0.122	0.141	0.171	0.174	0.132
		HS	0.111	0.144	0.220	0.226	0.114

One Hetero, one heterogeneity historical control.

Table A.36: MPSD of the treatment effect in the simulation study with a time-to-event endpoint and a 1:2 allocation ratio.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 4	Current	0.312	0.312	0.312	0.312	0.312
		Pooled	0.202	0.202	0.202	0.202	0.202
		EX	0.248	0.252	0.273	0.290	0.255
		EXNEX	0.253	0.257	0.278	0.293	0.260
		HS	0.221	0.224	0.245	0.268	0.227
	H = 8	Current	0.312	0.312	0.312	0.312	0.312
		Pooled	0.191	0.191	0.191	0.191	0.191
		EX	0.228	0.234	0.268	0.289	0.235
		EXNEX	0.236	0.242	0.274	0.292	0.242
		HS	0.198	0.200	0.215	0.241	0.200
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 4	Current	0.178	0.178	0.178	0.178	0.178
		Pooled	0.116	0.116	0.116	0.116	0.116
		EX	0.144	0.151	0.168	0.174	0.156
		EXNEX	0.146	0.153	0.169	0.174	0.157
		HS	0.128	0.133	0.156	0.168	0.137
	H = 8	Current	0.178	0.178	0.178	0.178	0.178
		Pooled	0.109	0.109	0.109	0.109	0.109
		EX	0.132	0.142	0.167	0.174	0.144
		EXNEX	0.135	0.145	0.169	0.174	0.146
		HS	0.114	0.117	0.141	0.160	0.117

One Hetero, one heterogeneity historical control.

Table A.37: Type I error rate (%) of the treatment effect in the simulation study with a time-to-event endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	5.23	5.41	5.38	5.11	5.05
		Pooled	5.31	6.39	17.58	32.25	7.95
		EX	4.15	4.43	5.37	5.21	4.64
		EXNEX	4.10	4.43	5.40	5.06	4.63
		HS	4.21	4.53	5.99	5.72	4.72
	H = 3	Current	5.42	5.07	5.09	4.98	5.24
		Pooled	5.16	6.52	18.83	35.93	7.01
		EX	4.00	4.09	5.10	5.91	4.55
		EXNEX	4.05	4.16	4.96	5.67	4.45
		HS	4.22	4.46	6.36	7.13	4.66
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	4.60	5.04	5.46	4.94	5.25
		Pooled	4.65	9.50	34.15	53.26	12.68
		EX	3.70	4.52	5.44	5.24	5.36
		EXNEX	3.71	4.54	5.38	5.19	5.35
		HS	3.51	4.94	5.93	5.81	5.49
	H = 3	Current	5.29	5.05	5.12	5.02	5.04
		Pooled	5.49	10.51	37.66	57.42	9.14
		EX	4.15	4.44	5.10	5.33	4.49
		EXNEX	4.27	4.35	5.05	5.26	4.45
		HS	4.23	5.38	6.37	6.71	4.47

One Hetero, one heterogeneity historical control.

Table A.38: Power (%) of the treatment effect in the simulation study with a time-to-event endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	51.4	52.0	51.3	51.6	51.6
		Pooled	72.3	71.2	63.2	59.9	84.7
		EX	61.3	61.2	56.3	54.8	66.0
		EXNEX	60.4	60.6	55.8	54.5	64.9
		HS	64.0	64.0	57.4	54.9	69.5
	H = 3	Current	51.8	51.2	51.8	51.6	52.0
		Pooled	75.2	73.6	64.8	60.9	85.3
		EX	64.5	62.6	57.8	54.9	68.5
		EXNEX	63.6	61.7	57.4	54.3	67.2
		HS	69.4	67.1	60.3	55.9	75.0
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	51.1	50.4	50.7	50.4	50.9
		Pooled	69.3	66.2	59.9	65.1	89.6
		EX	58.6	55.9	52.3	51.7	60.9
		EXNEX	58.2	55.7	52.2	51.6	60.2
		HS	61.5	58.5	53.0	51.7	65.4
	H = 3	Current	49.9	50.8	50.5	49.7	51.1
		Pooled	72.3	68.5	60.3	65.6	89.0
		EX	60.4	58.4	52.9	50.8	63.0
		EXNEX	59.7	57.9	52.7	50.6	62.3
		HS	65.4	62.3	54.7	51.1	70.1

One Hetero, one heterogeneity historical control.

Table A.39: Calibrated power (%) of the treatment effect in the simulation study with a time-to-event endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	50.7	50.7	50.1	50.9	51.6
		Pooled	71.6	67.5	41.0	NA	79.2
		EX	64.4	63.6	54.8	53.9	67.0
		EXNEX	63.5	63.2	54.9	54.4	65.9
		HS	66.6	65.3	54.4	52.3	70.4
	H = 3	Current	50.4	51.0	51.5	51.8	50.9
		Pooled	74.7	70.5	42.6	NA	82.0
		EX	67.8	65.0	57.6	51.6	69.8
		EXNEX	66.8	64.4	57.5	51.9	68.7
		HS	71.5	68.5	56.4	49.8	75.9
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	52.5	49.9	48.6	50.6	50.0
		Pooled	70.5	55.5	NA	NA	78.5
		EX	63.0	57.5	50.7	50.5	59.6
		EXNEX	62.7	56.7	50.9	50.9	58.9
		HS	66.4	58.5	50.2	49.3	63.6
	H = 3	Current	49.0	50.4	50.2	49.7	51.1
		Pooled	70.7	57.1	NA	NA	82.7
		EX	63.8	59.8	52.6	49.8	64.3
		EXNEX	63.5	59.4	52.6	50.0	64.1
		HS	68.0	61.0	50.5	45.7	71.8

One Hetero, one heterogeneity historical control.

Table A.40: Average bias of the treatment effect in the simulation study with a time-to-event endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{\text{CC}} = 30$, $n_{\text{CT}} = 30$, $n_h = 30$	H = 2	Current	0.000	0.006	−0.003	−0.003	−0.001
		Pooled	0.005	0.005	−0.026	−0.078	0.110
		EX	0.006	0.009	−0.001	−0.004	0.050
		EXNEX	0.006	0.009	−0.001	−0.003	0.046
		HS	0.006	0.009	−0.003	−0.005	0.062
	H = 3	Current	−0.001	−0.004	0.001	−0.001	0.004
		Pooled	0.006	−0.001	−0.024	−0.084	0.087
		EX	0.006	0.002	0.003	−0.001	0.048
		EXNEX	0.006	0.002	0.003	−0.001	0.045
		HS	0.006	0.002	0.000	−0.005	0.062
$n_{\text{CC}} = 90$, $n_{\text{CT}} = 90$, $n_h = 90$	H = 2	Current	0.000	−0.002	−0.001	0.001	0.001
		Pooled	0.001	−0.001	−0.023	−0.077	0.106
		EX	0.002	0.000	0.000	0.001	0.029
		EXNEX	0.002	0.000	0.000	0.001	0.028
		HS	0.002	0.000	−0.001	0.000	0.038
	H = 3	Current	−0.003	0.001	0.000	−0.003	0.001
		Pooled	0.001	−0.002	−0.034	−0.086	0.081
		EX	0.000	0.001	0.000	−0.003	0.028
		EXNEX	0.000	0.001	0.000	−0.003	0.026
		HS	0.001	0.001	−0.002	−0.004	0.037

One Hetero, one heterogeneity historical control.

Table A.41: RMSD of the treatment effect in the simulation study with a time-to-event endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	0.297	0.293	0.296	0.292	0.292
		Pooled	0.240	0.252	0.347	0.485	0.265
		EX	0.259	0.261	0.282	0.289	0.267
		EXNEX	0.261	0.262	0.283	0.288	0.268
		HS	0.252	0.255	0.285	0.294	0.265
	H = 3	Current	0.296	0.293	0.290	0.295	0.295
		Pooled	0.232	0.245	0.344	0.503	0.247
		EX	0.249	0.253	0.275	0.291	0.259
		EXNEX	0.252	0.255	0.275	0.290	0.261
		HS	0.238	0.245	0.282	0.304	0.251
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	0.164	0.167	0.169	0.168	0.168
		Pooled	0.136	0.160	0.283	0.447	0.175
		EX	0.146	0.155	0.167	0.168	0.161
		EXNEX	0.147	0.156	0.167	0.168	0.161
		HS	0.142	0.153	0.169	0.171	0.160
	H = 3	Current	0.169	0.168	0.168	0.168	0.167
		Pooled	0.134	0.159	0.294	0.465	0.153
		EX	0.144	0.152	0.164	0.168	0.154
		EXNEX	0.145	0.153	0.164	0.167	0.155
		HS	0.138	0.150	0.169	0.174	0.148

One Hetero, one heterogeneity historical control.

Table A.42: MPSD of the treatment effect in the simulation study with a time-to-event endpoint and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 30$, $n_{CT} = 30$, $n_h = 30$	H = 2	Current	0.293	0.293	0.293	0.293	0.293
		Pooled	0.238	0.238	0.238	0.238	0.238
		EX	0.269	0.270	0.278	0.284	0.272
		EXNEX	0.271	0.272	0.280	0.285	0.274
		HS	0.262	0.263	0.274	0.282	0.266
	H = 3	Current	0.293	0.293	0.293	0.293	0.293
		Pooled	0.231	0.231	0.231	0.231	0.231
		EX	0.262	0.264	0.275	0.283	0.266
		EXNEX	0.264	0.266	0.277	0.284	0.268
		HS	0.249	0.252	0.265	0.277	0.254
$n_{CC} = 90$, $n_{CT} = 90$, $n_h = 90$	H = 2	Current	0.167	0.167	0.167	0.167	0.167
		Pooled	0.137	0.137	0.137	0.137	0.137
		EX	0.156	0.158	0.164	0.166	0.161
		EXNEX	0.156	0.159	0.164	0.166	0.162
		HS	0.151	0.154	0.163	0.165	0.158
	H = 3	Current	0.167	0.167	0.167	0.167	0.167
		Pooled	0.132	0.132	0.132	0.132	0.132
		EX	0.151	0.155	0.163	0.165	0.158
		EXNEX	0.152	0.156	0.164	0.166	0.159
		HS	0.144	0.148	0.160	0.164	0.151

One Hetero, one heterogeneity historical control.

Table A.43: Type I error rate (%) of the treatment effect in the simulation study with a time-to-event endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	5.24	4.96	4.81	4.98	5.35
		Pooled	5.01	6.92	22.29	40.24	9.58
		EX	3.36	3.81	4.93	5.87	4.38
		EXNEX	3.41	3.77	4.72	5.68	4.34
		HS	3.39	3.89	5.87	7.00	4.67
	H = 3	Current	4.89	5.40	5.27	5.08	5.14
		Pooled	4.69	7.62	25.15	43.98	7.64
		EX	2.77	3.64	5.79	5.98	3.92
		EXNEX	2.77	3.68	5.36	5.82	3.81
		HS	3.17	4.32	8.42	9.01	4.35
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	5.07	5.48	4.77	5.01	5.04
		Pooled	4.77	11.78	41.04	61.11	16.55
		EX	3.31	4.39	5.46	5.38	4.65
		EXNEX	3.34	4.43	5.38	5.21	4.65
		HS	3.34	4.63	6.45	6.18	5.19
	H = 3	Current	4.63	5.27	4.83	5.23	5.05
		Pooled	4.99	12.68	45.76	63.15	11.38
		EX	2.92	4.49	5.30	5.54	4.22
		EXNEX	2.92	4.39	5.11	5.41	4.25
		HS	3.31	5.56	8.05	7.62	4.22

One Hetero, one heterogeneity historical control.

Table A.44: Power (%) of the treatment effect in the simulation study with a time-to-event endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	44.3	43.5	42.9	43.9	44.4
		Pooled	78.0	75.3	65.7	63.9	90.3
		EX	60.0	58.9	52.7	50.0	65.4
		EXNEX	58.7	57.8	51.7	49.2	63.6
		HS	64.8	63.6	55.2	51.2	70.7
	H = 3	Current	44.3	43.8	44.6	44.1	44.1
		Pooled	83.2	80.3	68.3	63.2	91.1
		EX	65.4	63.9	55.6	50.0	68.2
		EXNEX	63.6	62.4	54.7	49.3	66.1
		HS	73.9	71.9	60.8	53.3	78.2
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	45.0	44.7	45.0	44.7	44.1
		Pooled	76.9	71.3	63.2	68.6	95.4
		EX	57.2	53.6	48.6	46.3	58.7
		EXNEX	56.2	53.2	48.4	46.1	57.7
		HS	62.3	58.1	50.2	46.6	64.8
	H = 3	Current	44.3	44.0	44.2	44.3	44.8
		Pooled	80.0	74.3	64.2	69.1	94.1
		EX	61.6	57.4	48.2	46.5	61.2
		EXNEX	60.5	56.5	47.8	46.4	60.1
		HS	70.0	64.9	52.0	48.3	72.1

One Hetero, one heterogeneity historical control.

Table A.45: Calibrated power (%) of the treatment effect in the simulation study with a time-to-event endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	43.8	43.7	43.6	44.1	43.4
		Pooled	78.1	70.6	38.3	NA	84.2
		EX	66.4	64.3	53.2	46.4	67.1
		EXNEX	64.9	62.8	52.7	46.6	65.9
		HS	71.2	68.2	51.9	44.8	71.7
	H = 3	Current	44.7	42.7	44.0	44.0	43.5
		Pooled	83.9	75.0	39.1	NA	87.6
		EX	74.2	69.3	53.6	46.8	72.3
		EXNEX	72.6	67.3	53.2	47.0	70.1
		HS	80.3	73.6	52.0	42.6	79.7
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	44.8	43.2	45.8	44.7	44.1
		Pooled	77.6	57.0	NA	NA	84.5
		EX	64.0	56.0	47.2	45.4	59.9
		EXNEX	63.0	55.4	47.1	45.3	58.7
		HS	68.8	59.7	46.4	42.3	64.6
	H = 3	Current	45.6	43.1	44.5	43.5	44.7
		Pooled	80.1	59.1	NA	NA	87.6
		EX	69.8	59.4	47.1	44.7	64.0
		EXNEX	68.6	58.5	47.5	44.9	63.0
		HS	76.0	62.9	42.8	40.6	74.3

One Hetero, one heterogeneity historical control.

Table A.46: Average bias of the treatment effect in the simulation study with a time-to-event endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	−0.009	−0.010	−0.016	−0.007	−0.005
		Pooled	0.001	−0.003	−0.032	−0.078	0.124
		EX	0.002	0.000	−0.011	−0.006	0.060
		EXNEX	0.002	0.000	−0.010	−0.006	0.056
		HS	0.002	0.000	−0.011	−0.008	0.074
	H = 3	Current	−0.008	−0.004	−0.006	−0.009	−0.006
		Pooled	0.004	0.004	−0.026	−0.094	0.090
		EX	0.005	0.007	0.000	−0.010	0.051
		EXNEX	0.005	0.007	0.000	−0.010	0.047
		HS	0.005	0.007	−0.002	−0.017	0.066
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	−0.002	−0.002	−0.002	−0.002	−0.003
		Pooled	0.003	−0.003	−0.025	−0.079	0.121
		EX	0.002	0.000	−0.001	−0.002	0.035
		EXNEX	0.002	0.000	−0.001	−0.002	0.033
		HS	0.003	0.001	−0.002	−0.004	0.048
	H = 3	Current	−0.002	−0.001	−0.002	−0.002	−0.003
		Pooled	0.001	−0.003	−0.031	−0.084	0.085
		EX	0.001	0.001	−0.001	−0.002	0.031
		EXNEX	0.001	0.001	−0.001	−0.002	0.029
		HS	0.001	0.001	−0.004	−0.004	0.042

One Hetero, one heterogeneity historical control.

Table A.47: RMSD of the treatment effect in the simulation study with a time-to-event endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	0.315	0.311	0.310	0.310	0.314
		Pooled	0.218	0.238	0.350	0.524	0.255
		EX	0.247	0.253	0.280	0.302	0.265
		EXNEX	0.251	0.256	0.281	0.301	0.268
		HS	0.236	0.245	0.284	0.312	0.260
	H = 3	Current	0.311	0.312	0.316	0.312	0.313
		Pooled	0.206	0.229	0.356	0.535	0.228
		EX	0.230	0.242	0.285	0.304	0.248
		EXNEX	0.235	0.246	0.286	0.302	0.252
		HS	0.214	0.231	0.294	0.325	0.234
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	0.180	0.180	0.178	0.177	0.177
		Pooled	0.125	0.156	0.304	0.490	0.175
		EX	0.147	0.156	0.172	0.178	0.163
		EXNEX	0.149	0.157	0.172	0.177	0.163
		HS	0.138	0.151	0.176	0.183	0.160
	H = 3	Current	0.177	0.178	0.176	0.178	0.177
		Pooled	0.119	0.153	0.314	0.490	0.147
		EX	0.136	0.150	0.171	0.177	0.154
		EXNEX	0.138	0.151	0.171	0.177	0.155
		HS	0.124	0.145	0.182	0.188	0.143

One Hetero, one heterogeneity historical control.

Table A.48: MPSD of the treatment effect in the simulation study with a time-to-event endpoint, a 1:2 allocation ratio, and two and three historical trials.

Between-trial heterogeneity τ^2			No 0	Low 0.01	Moderate 0.09	High 0.25	One-Hetero
$n_{CC} = 20$, $n_{CT} = 40$, $n_h = 30$	H = 2	Current	0.312	0.312	0.312	0.312	0.311
		Pooled	0.218	0.218	0.218	0.218	0.218
		EX	0.268	0.270	0.282	0.293	0.273
		EXNEX	0.271	0.273	0.285	0.295	0.276
		HS	0.256	0.258	0.274	0.289	0.264
	H = 3	Current	0.312	0.312	0.312	0.312	0.312
		Pooled	0.208	0.208	0.208	0.208	0.208
		EX	0.257	0.260	0.277	0.291	0.264
		EXNEX	0.261	0.264	0.281	0.294	0.268
		HS	0.235	0.238	0.259	0.279	0.243
$n_{CC} = 60$, $n_{CT} = 120$, $n_h = 90$	H = 2	Current	0.178	0.178	0.178	0.178	0.178
		Pooled	0.125	0.125	0.125	0.125	0.125
		EX	0.157	0.161	0.170	0.174	0.166
		EXNEX	0.159	0.162	0.171	0.174	0.167
		HS	0.149	0.153	0.167	0.173	0.160
	H = 3	Current	0.178	0.178	0.178	0.178	0.178
		Pooled	0.120	0.120	0.120	0.120	0.119
		EX	0.150	0.155	0.169	0.174	0.161
		EXNEX	0.152	0.157	0.170	0.174	0.162
		HS	0.137	0.142	0.163	0.170	0.148

One Hetero, one heterogeneity historical control.