Journal of Data Science 0 (0), 1–19
??? 2026
DOI: 10.6339/26-JDS1240
Statistical Data Science

# Bayesian Dynamic Borrowing for Vaccine Efficacy Studies in Pediatric Population

Ruoyuan Qian[1], Wenlin Yuan[2], and Lei Gao[2,*]

[1]*Division of Biostatistics, The Ohio State University, Columbus, OH, USA*
[2]*Department of Biostatistics and Programming, Moderna, Cambridge, MA, USA*

**Abstract**

Pediatric vaccine efficacy studies face enrollment challenges, often resulting in underpowered studies to confirm vaccine efficacy. To address the challenge, this paper proposes a novel Two-Tier Bayesian dynamic borrowing framework that selectively incorporates two heterogeneous sources of historical data of adult populations into pediatric studies. The method employs similarity metrics based on intermediate outcomes, such as immunological markers, to dynamically determine borrowing weights via overlap coefficients, ensuring that only comparable data are utilized. Power priors are constructed with weights reflecting this similarity, while a cap on the effective sample size helps maintain control of the maximum information to borrow. Simulation studies demonstrate that the proposed approach improves power, reduces estimation bias and maintains type I error control, thus offering a robust and practical strategy for pediatric vaccine study design.

**Keywords** *effective sample size; historical borrowing; pediatric vaccine trials; power prior*

## 1 Introduction

Vaccine efficacy trials are designed to assess the protective effect of a novel vaccine by comparing disease incidence between a vaccinated group and a control group (placebo or standard of care). A major challenge arises when the target disease has low incidence, as an extremely large sample size is required to accrue enough events and achieve adequate statistical power (Chan and Bohidar, 1998; Jin et al., 2020). This challenge is amplified in pediatric studies, where strict ethical and regulatory constraints naturally limit enrollment (US Food and Drug Administration, 2023b). As a result, the tension between needing large sample sizes and facing restrictive recruitment conditions makes it particularly difficult to conduct adequately powered vaccine efficacy trials in pediatric populations.

To address these issues, researchers have increasingly turned to innovative study designs including Bayesian approaches that incorporate prior knowledge from historical data. Following the 21st Century Cures Act (US Government Publishing Office, 2016), the U.S. Food and Drug Administration (FDA) issued guidance to support the use of complex trial designs that incorporate data borrowing through Bayesian frameworks, including documents on leveraging existing clinical data (US Food and Drug Administration, 2016), the framework for real-world evidence (RWE) program and use of RWE in drug development (US Food and Drug Administration, 2018, 2023a,c, 2024), and workshop on complex innovative trial design in pediatric therapeutics

*Corresponding author. Email: lei.gao@modernatx.com.

© 2026 The Author(s). Published by the School of Statistics and the Center for Applied Statistics, Renmin University of China. Open access article under the CC BY license.
Received November 21, 2025; Accepted June 28, 2026

(US Food and Drug Administration, 2021). One prominent method is the power prior proposed by Ibrahim and Chen (2000), which is also cited by FDA for the use of Bayesian statistics in the design and analysis of medical device clinical trials (US Food and Drug Administration, 2010). It permits partial borrowing by adjusting 'power parameters' that reflect the similarity between the historical and current populations. When applied carefully, power priors can increase study power and have been successfully incorporated into a variety of Bayesian clinical trial designs (Chen et al., 2011, 2014; Jiang et al., 2015; Li et al., 2015, 2018; van Rosmalen et al., 2018; Yuan et al., 2022; Chen et al., 2024, 2025; Baumann et al., 2025).

Bayesian dynamic borrowing has been studied through several other related frameworks. Commensurate priors (Hobbs et al., 2011) provide an adaptive borrowing mechanism by introducing parameters that quantify compatibility between current and historical data, thereby reducing borrowing under prior-data conflict. Hierarchical and meta-analytic predictive (MAP) models assume partial exchangeability across studies and can be made more robust through mixture components to protect against conflict and over-borrowing (Schmidli et al., 2014). In addition, normalized power-prior (Han et al., 2023) formulations have been developed for dynamic borrowing. More broadly, Viele et al. (2014) reviewed several approaches to borrowing historical control information and emphasized that dynamic borrowing procedures should adapt to the degree of compatibility between historical and current data, since inappropriate borrowing may increase bias and inflate type I error even when power is improved.

Recent developments in historical data borrowing for vaccine trials have focused on two types of historical information: historical control (HC) data and historical vaccine efficacy. For instance, De Santis and Gubbiotti (2023) proposed a dynamic power prior method for incorporating HC data in non-inferiority trials for new vaccine. Peng et al. (2023) extended a probability-based power prior to borrow multiple sources of HC data for vaccine efficacy trials. Additionally, Callegaro et al. (2023, 2024) proposed robust mixture prior (RMP) models for dynamic HC data borrowing in vaccine efficacy trials. On the other hand, Jin et al. (2020) proposed a Bayesian framework specifically designed to borrow historical vaccine efficacy information for vaccine efficacy trials.

Another important application of Bayesian methods involves borrowing information from adult populations to support pediatric trial design and analysis. Such historical trials are typically available due to the unique drug development pathway for pediatric populations, where the same intervention is first evaluated in adults, followed by adolescents, and subsequently in pediatric populations. Therefore, for pediatric drug development, it is common that historical trials with the same intervention but different age populations are available. Borrowing information from adult data in this context is also supported by regulatory guidance (US Food and Drug Administration, 2016). Schoenfeld et al. (2009) introduced a hierarchical model in which treatment effects in adult and pediatric populations are assumed to follow a common normal distribution, with a prespecified variance controlling the extent of borrowing. Building on this approach, Jin et al. (2021) extended to accommodate binary endpoints and multiple historical adult trials, further enhancing its applicability in pediatric vaccine research. In addition, Gamalo-Siebers et al. (2017) provided a comprehensive summary and practical recommendations for applying Bayesian methods to design more efficient pediatric trials integrating adult data.

A key challenge arises when borrowing from different population to the pediatric trial. One approach is dynamic borrowing based on the similarity between historical and current trials. Similarity is commonly quantified using the overlap in propensity score (PS) distributions derived from baseline covariates. These measures can then be incorporated into Bayesian borrowing frameworks, such as the power prior (Wang et al., 2019; Qian et al., 2025; Psioda et al.,

2025) or MAP prior (Liu et al., 2021), allowing greater borrowing from external patients more similar to those in the current study. By leveraging PS overlap, these approaches help balance observed baseline covariates and improve estimation accuracy. However, a major limitation is that PS overlap-based methods require access to individual-level external data with comparable covariates, treatment definitions, and outcome measures, which are not always available in practice.

To retain the strengths of PS overlap-based approaches while relaxing the requirement for individual-level external data, we propose a Two-Tier dynamic borrowing framework based on the power prior for vaccine efficacy trials in pediatric populations. The framework leverages the strong association between antibody concentration and vaccine efficacy (Huang et al., 2023), using antibody concentration as an intermediate outcome to construct similarity metrics that dynamically determine borrowing weights when incorporating historical data. Our framework pre-specifies target maximum effective sample size (ESS) and a similarity threshold, ensuring that only the most relevant historical data are incorporated in a controlled manner, thereby balancing gains in power with control of bias and type I error rate. This adaptive data-driven borrowing strategy represents a unique contribution to the design of confirmatory vaccine efficacy studies in pediatric population, where recruitment challenges necessitate careful integration of heterogeneous external data.

The remainder of this paper is organized as follows. Section 2 details the proposed Bayesian approach, including the study design, notations, and the construction of the Two-Tier dynamic borrowing. Section 3 reports simulation results illustrating how our method controls type I error while offering substantial gains in power and estimation efficiency under a variety of scenarios. Finally, Section 4 provides a discussion of practical considerations, and possible extensions of the proposed approach.

## 2 Methods

In this section, we outline the design context and procedures and introduce the Two-Tier Bayesian dynamic borrowing framework.

The objective is to design a two-arm placebo-controlled phase 3 study ("the current study") to evaluate vaccine efficacy in the pediatric population. Two heterogeneous historical data sources–drawn from adult populations–are available for possible information borrowing. These adult populations can represent distinct age groups or different geographic regions. Our approach follows Jin et al. (2020)'s Bayesian design framework, which uses binomial models of vaccine efficacy parameters, power priors, and ESS. In consultation with regulatory agencies, a maximum number of cases to be borrowed is pre-specified.

The amount of information to borrow from each historical source is captured by power parameters and determined through the following key steps: 1) intermediate endpoints are used to quantify the similarity between the current study and historical source in terms of overlapping coefficients, 2) each similarity metric is evaluated against a predefined threshold to determine whether borrowing is warranted, 3) if at least one source meets the threshold, the method proceeds to borrow information, allocating power parameters in proportion to the degree of similarity, assigning greater weight to sources that are more similar to the current study.

## 2.1 Vaccine Efficacy Study Notations

Following the notation from Jin et al. (2020), we assume the allocation ratio to be $c:1$ and total sample size to be $N = N_1 + N_2$ in the current study. Here, $N_1$ and $N_2$ represent the sample sizes in the current control (CC) and current vaccine (CV) groups, respectively, such that $c = N_1/N_2$. Let $p_1$ and $p_2$ denote the incidence rates for the CC and CV groups, respectively. The number of cases in the CC group, $X$, follows a binomial distribution, $X \sim \text{Bin}(N_1, p_1)$, and the number of cases in the CV group, $Y$, is modeled as $Y \sim \text{Bin}(N_2, p_2)$.

The vaccine efficacy is measured by the relative risk reduction, defined as:

$$\pi = 1 - \frac{p_2}{p_1}.$$

This measure is often compared with a pre-specified positive lower bound, $\pi_0$. Accordingly, the hypothesis test for $\pi$ is formulated as:

$$H_0 : \pi \leqslant \pi_0 \text{ versus } H_1 : \pi > \pi_0. \tag{1}$$

The true incidence rates, $p_1$ and $p_2$, are usually small. As a result, large sample sizes, $N_1$ and $N_2$, are typically required to observe a sufficient number of cases, $X$ and $Y$, for meaningful inference. Furthermore, a binomial distribution with a large sample size and a small probability can be approximated by a Poisson distribution. Thus, letting $N_1 p_1 \approx \lambda_1$ and $N_2 p_2 \approx \lambda_2$, the number of cases $X$ and $Y$ approximately follow two independent Poisson distributions with parameters $\lambda_1$ and $\lambda_2$, respectively (Chan and Bohidar, 1998). Let $T = X + Y$ denote the total number of cases in the current study. The conditional distribution of $Y$ given $T = t$ follows:

$$Y \mid T \sim \text{Bin}(T, \theta), \tag{2}$$

where

$$\theta = \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{N_2 p_2}{N_1 p_1 + N_2 p_2} = \frac{1-\pi}{c+1-\pi}.$$

Thus, the hypothesis test for $\pi$ in (1) can be reformulated as:

$$H_0 : \theta \geqslant \theta_0 \text{ versus } H_1 : \theta < \theta_0, \tag{3}$$

where $\theta_0 = (1-\pi_0)/(c+1-\pi_0)$.

At the design stage, the total number of cases, $T$, can be determined based on the pre-specified significance level $\alpha$, the lower bound $\pi_0$, the expected efficacy $\pi_1$, and the target power $1-\beta$ (Chan and Bohidar, 1998; Jin et al., 2020). Given the total number of cases $T$, the expected efficacy $\pi_1$, the allocation ratio $c$, and the control incidence rate $p_1$, the required sample size for the vaccine group is:

$$N_2 = \frac{T}{(1 + c - \pi_1)p_1}.$$

## 2.2 Two-Tier Dynamic Borrowing

The Two-Tier method dynamically borrows data from two historical data sources by quantifying the similarity between each historical data and the current study, and calibrating the borrowing such that the total borrowed information is constrained by a pre-specified maximum ESS.

We denote two historical data sources by $A$ and $B$ and set up the power prior framework for borrowing. In historical data source $A$, let $T_A = X_A + Y_A$ represent the total number of

cases, where $X_A$ denotes the number of cases in the control group and $Y_A$ denotes the number of cases in the vaccine group. Similarly, in historical data source $B$, the total number of cases is $T_B = X_B + Y_B$, where $X_B$ is the number of cases in the control group and $Y_B$ is the number of cases in the vaccine group.

### Construct Power Prior

Following the concept of the power prior (Ibrahim and Chen, 2000), we assume that $Y_A \mid T_A$ and $Y_B \mid T_B$ each follow a binomial distribution with the same parameter $\theta$ as in the current study (2):

$$Y_A \mid T_A \sim \text{Bin}(T_A, \theta); \qquad Y_B \mid T_B \sim \text{Bin}(T_B, \theta).$$

In the power prior framework, the initial prior $\pi_0(\theta)$ is usually chosen to be non-informative. Therefore, we apply Jeffreys's Prior in our setting, $\theta \sim \text{Beta}(1/2, 1/2)$ (Jeffreys, 1998). The power prior incorporating information from historical sources $A$ and $B$ is then:

$$\begin{aligned} \pi_{\text{PP}}(\theta) &\propto L(\theta \mid Y_A, T_A)^{\alpha_A} L(\theta \mid Y_B, T_B)^{\alpha_B} \pi_0(\theta) \\ &\propto \theta^{\alpha_A Y_A + \alpha_B Y_B - 1/2} (1-\theta)^{\alpha_A X_A + \alpha_B X_B - 1/2}, \end{aligned} \tag{4}$$

equivalently,

$$\theta \sim \text{Beta}(\alpha_A Y_A + \alpha_B Y_B + 1/2, \alpha_A X_A + \alpha_B X_B + 1/2),$$

where $\alpha_A$ and $\alpha_B$ are power parameters ranging between 0 and 1 for studies $A$ and $B$, respectively, which control the degree of borrowing from each study.

The borrowed information is quantified by the prior ESS, which plays a crucial role in study design. The prior ESS represents the amount of external information contributing to the inference and, in this setting, reflects the discounted information borrowed from historical trials. In vaccine efficacy studies, this corresponds to the number of infection cases to be borrowed, denoted as $S$. The FDA draft guidance (US Food and Drug Administration, 2026) highlights the importance of quantifying both the maximum and mean ESS. Accordingly, we propose to cap the prior ESS at a pre-specified level and adjust the weights of the power parameters to enable dynamic borrowing. The pre-specified cap $S$ can be determined based on the similarity between the populations of the historical trials and the current trial, as well as the power required for the current trial. By adjusting the power parameters $\alpha_A$ and $\alpha_B$, we can cap the prior ESS at a pre-specified level $S$. For a $\text{Beta}(a, b)$ prior, the ESS is given by $\text{ESS} = a + b$ (Malec, 2001; Morita et al., 2008, 2012; Neuenschwander et al., 2020). Accordingly, the prior ESS for (4) is:

$$\text{ESS} = \alpha_A(Y_A + X_A) + 1/2 + \alpha_B(Y_B + X_B) + 1/2 = \alpha_A T_A + \alpha_B T_B + 1. \tag{5}$$

The total information from the power prior (4), quantified by (5), comprises three parts: historical source $A$ ($\alpha_A T_A$), historical source $B$ ($\alpha_B T_B$), and the initial prior $\pi_0(\theta)$ (a constant depending on the choice of initial prior).

### Determine Power Parameter

The determination of the power parameters $\alpha_A$ and $\alpha_B$ aims to achieve two goals. First, the power parameters must collectively ensure that the prior ESS in (5) is consistent to $S$, the number of cases intended to be borrowed at the design stage. Second, the parameters should allocate more weight to borrowing information from the historical source that is more similar

to the current study while borrowing less, or not at all, from the less similar study. To achieve these two objectives, the power parameters are:

$$\alpha_A = \frac{S-1}{T_A} v_A; \qquad \alpha_B = \frac{S-1}{T_B} v_B, \tag{6}$$

where $v_A$ and $v_B$ range from 0 to 1 and satisfy the constraint $v_A + v_B = 1$ if at least one of historical sources is similar enough to be borrowed. This ensures that ESS $= S$, meaning the total number of borrowed cases equals $S$ when the external data qualifies for borrowing. The numerators in (6) may vary depending on the choice of the initial prior $\pi_0(\theta)$. For instance, if the initial prior of $\theta \sim \text{Beta}(1, 1)$, then $\pi_0(\theta) = 1$, resulting in ESS $= \alpha_A T_A + \alpha_B T_B + 2$. Consequently, the borrowing weights are given by $\alpha_A = (S-2)v_A/T_A$ and $\alpha_B = (S-2)v_B/T_B$. Furthermore, by making $v_A$ and $v_B$ positively associated with the similarity between the current study and historical sources $A$ and $B$, respectively, (6) assigns greater weight to the more similar historical source, thereby contributing more to the ESS in (5).

The similarity between studies can be assessed using a continuous intermediate outcome which is closely related to the primary outcome, such as antibody concentration when the correlate of protection is established. Specifically, we compute the Overlap Coefficient (OVL) of intermediate outcomes between studies to quantify their similarity (Inman and Bradley, 1989). The OVL ranges from 0 to 1, where OVL $= 1$ indicates identical distributions, and OVL $= 0$ indicates completely disjoint distributions. Let $m$ represent the intermediate outcome, and let $f_{\mathcal{C}}(m)$ and $f_A(m)$ denote the density functions of the current study ($\mathcal{C}$) and historical source $A$, respectively. The OVL between the current study and historical source $A$ is defined as:

$$\text{OVL}(\mathcal{C}, A) = \int_{-\infty}^{\infty} \min\left[f_{\mathcal{C}}(m), f_A(m)\right] \mathrm{d}m.$$

Similarly, the OVL between the current study and historical source $B$ can be computed as $\text{OVL}(\mathcal{C}, B)$.

The weights $v_A$ and $v_B$ can then be determined by normalizing the OVL values and applying a pre-specified threshold $H$ ($H \in [0, 1]$), which controls the minimum OVL required for borrowing:

$$\begin{aligned} v_A =& \mathbb{1}\{\text{OVL}(\mathcal{C}, A) \geqslant H\} \times \\ & \left[\frac{\text{OVL}(\mathcal{C}, A)}{\text{OVL}(\mathcal{C}, A) + \text{OVL}(\mathcal{C}, B)} \mathbb{1}\{\text{OVL}(\mathcal{C}, B) \geqslant H\} + 1 - \mathbb{1}\{\text{OVL}(\mathcal{C}, B) \geqslant H\}\right], \\ v_B =& \mathbb{1}\{\text{OVL}(\mathcal{C}, B) \geqslant H\} \times \\ & \left[\frac{\text{OVL}(\mathcal{C}, B)}{\text{OVL}(\mathcal{C}, A) + \text{OVL}(\mathcal{C}, B)} \mathbb{1}\{\text{OVL}(\mathcal{C}, A) \geqslant H\} + 1 - \mathbb{1}\{\text{OVL}(\mathcal{C}, A) \geqslant H\}\right]. \end{aligned} \tag{7}$$

Larger values of $H$ impose more restrictive borrowing conditions, whereas smaller values allow more permissive borrowing. The interpretation of Equation (7) can be summarized into four possible scenarios:

- **Scenario 1** $\text{OVL}(\mathcal{C}, A) < H$ and $\text{OVL}(\mathcal{C}, B) < H$:
  In this scenario, $v_A = v_B = 0$, meaning that neither historical source is similar enough to the current study, and no borrowing will occur.
- **Scenario 2** $\text{OVL}(\mathcal{C}, A) \geqslant H$ and $\text{OVL}(\mathcal{C}, B) < H$:
  Here, $v_A = 1$ and $v_B = 0$, indicating that borrowing will occur only from historical source $A$.

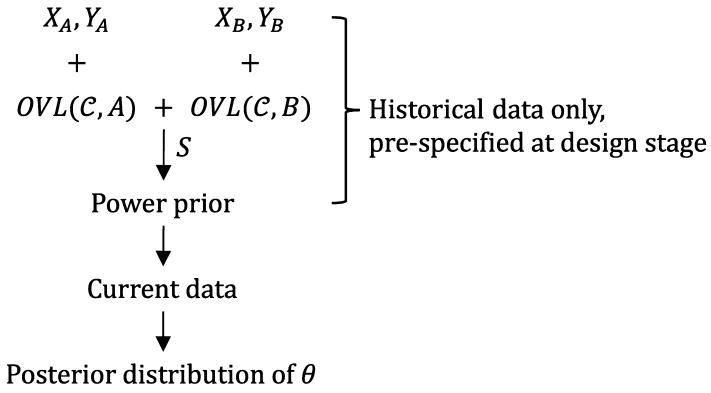


Figure 1: Flowchart of the Two-Tier method.

- **Scenario 3** $\text{OVL}(\mathcal{C}, A) < H$ and $\text{OVL}(\mathcal{C}, B) \geqslant H$:
  In this case, $v_A = 0$ and $v_B = 1$, meaning that borrowing will occur only from historical source $B$.
- **Scenario 4** $\text{OVL}(\mathcal{C}, A) \geqslant H$ and $\text{OVL}(\mathcal{C}, B) \geqslant H$:
  In this situation, $v_A = \text{OVL}(\mathcal{C}, A)/\big[\text{OVL}(\mathcal{C}, A)+\text{OVL}(\mathcal{C}, B)\big]$, $v_B = \text{OVL}(\mathcal{C}, B)/\big[\text{OVL}(\mathcal{C}, A)+\text{OVL}(\mathcal{C}, B)\big]$. Borrowing will occur from both historical sources $A$ and $B$, with more weight assigned to the study that is more similar to the current study.

The intermediate outcome data for the current study can be obtained from earlier-phase studies conducted in the same pediatric population. Therefore, the power parameters $\alpha_A$ and $\alpha_B$ are determined a priori at the design stage, without using any data from the current study. Figure 1 illustrates the logic of the proposed Two-Tier method.

### Inference and Decision Rule Based on Posterior Distribution

With the current data likelihood specified in (2) and the power prior defined in (4), Bayesian inference can be performed for $\theta$ using its posterior distribution:

$$\begin{aligned} \pi(\theta \mid Y, T) &\propto L(\theta \mid Y, T)\pi_{\text{PP}}(\theta) \\ &\propto \theta^{Y+\alpha_A Y_A+\alpha_B Y_B-1/2}(1-\theta)^{X+\alpha_A X_A+\alpha_B X_B-1/2}, \end{aligned} \tag{8}$$

equivalently,

$$\theta \mid Y, T \sim \text{Beta}(Y+\alpha_A Y_A+\alpha_B Y_B+1/2,\, X+\alpha_A X_A+\alpha_B X_B+1/2).$$

The hypothesis for $\theta$ is described in (3). To test this hypothesis, we use the marginal decision rule for rejecting the null hypothesis at a two-sided type I error rate of $\alpha$, based on the posterior distribution in (8):

$$\Pr(\theta < \theta_0 \mid Y, T) \geqslant 1 - \frac{\alpha}{2}, \tag{9}$$

we will reject the null if $\theta_U < \theta_0$, where $\theta_U$ is the $100(1-\alpha/2)\%$ quantile of the posterior distribution in (8).

Since $\theta$ is a one-to-one mapping of $\pi$, it is meaningful to obtain the point estimation and 95% credible interval for the original measure, $\pi$. To achieve this, we proceed as follows:

- **Step 1** Draw a sample $\{\theta_i, i = 1, \ldots, n\}$ from the posterior distribution of $\theta$ in (8).

- **Step 2** Convert each sampled value $\theta_i$ back to $\pi_i$ using the transformation:

$$\pi_i = \frac{1-(c+1)\theta_i}{1-\theta_i},$$

thereby forming a sample $\{\pi_i, i = 1, \ldots, n\}$.
- **Step 3** Due to asymmetric distribution of $\pi$, the point estimation of $\pi$ is computed as the median of $\{\pi_i, i = 1, \ldots, n\}$. The 95% credible interval of posterior distribution of $\pi$ is calculated using the 2.5%- and 97.5%-quantiles of $\{\pi_i, i = 1, \ldots, n\}$, denoted as $\pi_L$ and $\pi_U$.

# 3 Simulation

## 3.1 Simulation Setting

For simulation purposes, we considered a two-arm randomized controlled study with a 1:1 allocation ratio as the current study. Therefore, the allocation ratio was $c = N_1/N_2 = 1$. We explored two different current sample sizes, $N = N_1 + N_2 = 2000, 2400$, with a fixed number of cases to be borrowed: $S = 9$. The control incidence rate was $p_1 = 0.01$, and the lower bound of vaccine efficacy for the new study was set at $\pi_0 = 0.3$, and thus $\theta_0 = 0.4118$. Therefore, we were testing the hypothesis: $H_0 : \theta \geqslant 0.4118$ versus $H_1 : \theta < 0.4118$. We explored four levels of true vaccine efficacy for the current study: $\pi = 0.3, 0.5, 0.7, 0.9$, or equivalently, $\theta = 0.4118, 0.3333, 0.2309, 0.0909$. The type I error rate was computed when $\theta = 0.4118$, and power was computed when $\theta = 0.3333, 0.2309, 0.0909$. The number of cases in the control group, denoted as $X$, and in the vaccine group, denoted as $Y$, for the current study were generated from binomial distributions:

$$X \sim \text{Bin}(N_1, p_1), \qquad Y \sim \text{Bin}(N_2, p_2).$$

Two historical sources, denoted as historical source $A$ and $B$, were considered. These sources included the same vaccine and control interventions but involved two different populations. The historical sources differed considerably in their total number of cases, with $T_A = 250$ and $T_B = 50$. We assumed that the control incidence rates in the historical sources were the same as in the current study, i.e., $p_{A1} = p_{B1} = p_1 = 0.01$. We analyzed two scenarios for vaccine efficacy in the historical sources, $\pi_A$, $\pi_B$:

- **Scenario 1** $\pi_A > \pi_B$:
  We set $\pi_A = 0.9177$ and $\pi_B = 0.6842$, corresponding to $Y_A = 19$, $X_A = T_A - Y_A = 231$, and $Y_B = 12$, $X_B = T_B - Y_B = 38$.
- **Scenario 2** $\pi_A < \pi_B$:
  We set $\pi_A = 0.6632$ and $\pi_B = 0.9130$, corresponding to $Y_A = 63$, $X_A = T_A - Y_A = 187$, and $Y_B = 4$, $X_B = T_B - Y_B = 46$.

The intermediate outcomes for each study were generated from a normal distribution related to their vaccine efficacy. Let $M_i$ denote the intermediate outcome for the $i$th subject in the vaccine group. Assuming there were $N_0 = 50$ subjects in each vaccine arm of a previous early-phase study, the intermediate outcomes associated with the vaccine efficacy $\pi$ were generated as follows:

$$\begin{aligned} M_i \sim & N(\mu, 1), i = 1, \ldots, N_0; \\ & \pi = 0.1 + 0.4\mu, \end{aligned} \tag{10}$$

Table 1: Parameter settings for current and historical sources.

| Study | $p_1$ | $p_2$ | $\pi$ | $\theta$ | $\mu$ |
|---|---|---|---|---|---|
| Current | 0.010 | 0.007 | 0.300 | 0.412 | 0.500 |
| | 0.010 | 0.005 | 0.500 | 0.333 | 1.000 |
| | 0.010 | 0.003 | 0.700 | 0.231 | 1.500 |
| | 0.010 | 0.001 | 0.900 | 0.091 | 2.000 |
| Scenario 1 | | | | | |
| A | 0.010 | 0.001 | 0.918 | 0.076 | 2.044 |
| B | 0.010 | 0.003 | 0.684 | 0.240 | 1.461 |
| Scenario 2 | | | | | |
| A | 0.010 | 0.003 | 0.663 | 0.252 | 1.408 |
| B | 0.010 | 0.001 | 0.913 | 0.080 | 2.033 |

where $\mu$ could be determined once $\pi$ was specified. The intercept was set to 0.1 to account for unobserved mechanisms between the intermediate and primary outcomes that were not fully captured by the intermediate outcome. Similarly, the intermediate outcomes for historical sources $A$ and $B$ were generated by substituting $\pi_A$ and $\pi_B$ into (10). Since the OVL of the intermediate outcome quantified similarity and was related to vaccine efficacy, multiple heterogeneity scenarios were considered by varying the vaccine efficacies for both historical and current trials. Specifically, two scenarios were considered for the two historical trials, representing different combinations of similarity with the current study. In addition, four values of vaccine efficacy were evaluated for the current study. Different combinations of vaccine efficacy between the current study and the two historical trials corresponded to different heterogeneity scenarios. The parameter settings for current and historical trials under two scenarios were summarized in Table 1. When the vaccine efficacy in the current study was low, e.g., $\pi = 0.3$, the current study was not similar to either historical trial ($\pi_A, \pi_B > 0.6$). As the vaccine efficacy in the current study increased, the similarity between the current and historical trials also increased.

The simulation settings reflected the clinical plausibility of a low-incidence pediatric phase 3 vaccine efficacy trial with limited endpoint accrual, incorporating two heterogeneous adult historical sources with unequal information sizes, as well as an immunologic intermediate endpoint that was associated with, but did not fully determine, vaccine efficacy. The heterogeneity scenarios for the historical trials were chosen to represent realistic degrees of similarity between pediatric and adult populations, ranging from no borrowing to selective borrowing from more comparable source(s). In addition, the use of a pre-specified similarity threshold and ESS cap reflected practical design constraints commonly encountered in regulatory decision-making.

For the proposed Two-Tier dynamic borrowing method, we set $H = 0.7$ as the borrowing threshold in (7), which indicated that the method would borrow data only when the OVL of intermediate outcomes between the current study and the historical source exceeded 0.7. The initial prior for $\theta$ was chosen as Jeffreys's prior, Beta$(1/2, 1/2)$, which resulted in the same posterior distribution as derived in (8). Additionally, we conducted a sensitivity analysis for the Two-Tier method by varying the threshold $H$ (0.5 and 0.9) and comparing the results with the choice $H = 0.7$ when the current sample size $N = 2000$, for both null and alternative hypotheses.

For comparison purposes, we included the "No borrow" method and the "Pool" method to

evaluate the benefits and risks of borrowing data. The "No borrow" method was identical to the proposed method but set the borrowing parameters $\alpha_A$ and $\alpha_B$ to zero, indicating no borrowing from historical data. In contrast, the "Pool" method did not distinguish between historical data sources or assess data similarity. Instead, it pooled and incorporated all available historical data indiscriminately. To make the "Pool" method comparable to the proposed method, we restricted it to borrow the same prior ESS as the proposed method. Specifically, we defined the borrowing parameter as $\alpha_p = \alpha_A = \alpha_B = S/T_p$, where $T_p = T_A + T_B$. With a Beta(1/2, 1/2) prior for $\theta$, the posterior distribution for the "Pool" method was:

$$\text{Beta}(Y + \alpha_p Y_p + 1/2, X + \alpha_p X_p + 1/2),$$

where $Y_p = Y_A + Y_B$ and $X_p = X_A + X_B$.

Since the posterior distributions of $\theta$ for all methods had closed forms following Beta distributions, Bayesian inference was conducted by directly extracting the 2.5%- and 97.5%-quantiles from the posterior Beta distribution to construct credible intervals. The hypothesis was tested based on the decision rule described in (9). To compute the point estimations for $\theta$ and $\pi$, we first drew a sample of size 10,000 from the posterior distribution of $\theta$. Each sampled value, $\{\theta_i, i = 1, \ldots, 10{,}000\}$, was then converted to $\{\pi_i = [1 - (c+1)\theta_i]/(1-\theta_i), i = 1, \ldots, 10{,}000\}$. The point estimations for $\theta$ and $\pi$ were computed as the medians of the sampled values, $\{\theta_i, i = 1, \ldots, 10{,}000\}$ and $\{\pi_i, i = 1, \ldots, 10{,}000\}$, respectively. The OVL for intermediate outcomes was estimated using the `kdensity()` function in the `kdensity` R package, which enabled flexible univariate nonparametric density estimation with asymmetric kernels (Moss and Tveten, 2019).

A total of 5,000 datasets were generated for each parameter combination. We assessed the methods based on three key aspects: borrowing capacity, point estimation, and inference performance. Borrowing capacity was measured by the average of ESS contributed by historical source $A$ ($\alpha_A T_A$), historical source $B$ ($\alpha_B T_B$), and the average of overall prior ESS, as defined in (5), respectively. Point estimation was evaluated for both $\theta$ and $\pi$, focusing on bias and mean squared error (MSE). Inference performance was assessed under $\theta$ by examining type I error and power. All simulations were conducted using R version 4.4.1.

### 3.2 Simulation Results

Three methods (No borrow, Two-Tier, Pool) were evaluated across two current sample sizes ($N = 2000, 2400$), four vaccine efficacy values ($\pi = 0.3, 0.5, 0.7, 0.9$) and two historical data scenarios, resulting in 16 parameter combinations. Two additional threshold values were considered in a sensitivity analysis.

Table 2 presented the simulation results under the null hypothesis, where $\pi = \pi_0 = 0.3$ (equivalently, $\theta = \theta_0 = 0.4118$). The one-sided significance level was 0.025. Overall, the Two-Tier method rarely borrowed data because the true vaccine efficacy $\pi$ for the current study was quite different from the historical sources in both scenarios. As a result, the performance of the Two-Tier method closely aligned with the "No borrow" method. In contrast, the "Pool" method, which borrowed data every time, introduced more bias since the historical data were dissimilar to the current study. This bias inflated the type I error rate, with the inflation decreasing as the sample size grows, because, as the current sample size increases, the borrowed data affect the inference less. The inflation was more pronounced in Scenario 1 than in Scenario 2. This was because the "Pool" method aggregated historical data, and in both scenarios, the number of cases in historical source $A$ was much larger than in study $B$. Consequently, historical source

Table 2: Simulation results under the null hypothesis.

| N | Method | Borrowing | | | Point Estimation | | | | Inference |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\theta$ | | $\pi$ | | $\theta$ |
| | | $\alpha_A T_A$ | $\alpha_B T_B$ | ESS | Bias | MSE | Bias | MSE | Type I Error |
| **Scenario 1** | | | | | | | | | |
| 2000 | No borrow | – | – | 1.000 | 0.002 | 0.014 | −0.090 | 0.208 | 0.027 |
| | Two-Tier | 0.000 | 0.090 | 1.090 | 0.001 | 0.014 | −0.087 | 0.206 | 0.027 |
| | Pool | 6.667 | 1.333 | 9.000 | −0.099 | 0.016 | 0.223 | 0.084 | 0.141 |
| 2400 | No borrow | – | – | 1.000 | 0.002 | 0.012 | −0.070 | 0.143 | 0.023 |
| | Two-Tier | 0.000 | 0.064 | 1.064 | 0.002 | 0.012 | −0.069 | 0.143 | 0.023 |
| | Pool | 6.667 | 1.333 | 9.000 | −0.087 | 0.014 | 0.199 | 0.072 | 0.134 |
| **Scenario 2** | | | | | | | | | |
| 2000 | No borrow | – | – | 1.000 | −0.001 | 0.014 | −0.079 | 0.195 | 0.026 |
| | Two-Tier | 0.096 | 0.000 | 1.096 | −0.002 | 0.014 | −0.076 | 0.191 | 0.027 |
| | Pool | 6.667 | 1.333 | 9.000 | −0.062 | 0.010 | 0.137 | 0.060 | 0.054 |
| 2400 | No borrow | – | – | 1.000 | −0.001 | 0.012 | −0.066 | 0.159 | 0.027 |
| | Two-Tier | 0.094 | 0.000 | 1.094 | −0.001 | 0.012 | −0.064 | 0.158 | 0.027 |
| | Pool | 6.667 | 1.333 | 9.000 | −0.054 | 0.009 | 0.120 | 0.055 | 0.058 |

$A$ dominated the analysis. In Scenario 1, the vaccine efficacy in source $A$, $\pi_A$, exceeded 0.9, resulting in worse performance for the "Pool" method compared to Scenario 2, where $\pi_A$ was around 0.6. Additionally, the "Pool" method exhibited smaller MSE, owing to the increased sample size from borrowing, which reduced uncertainty despite introducing bias.

Tables 3 and 4 presented the simulation results under the alternative hypotheses when $\pi = 0.5, 0.7, 0.9$ (equivalently, $\theta = 0.3333, 0.2309, 0.0909$) for Scenario 1 and Scenario 2, respectively. In general, the Two-Tier method dynamically borrowed from different historical sources depending on the true vaccine efficacy $\pi$ in the current study and the similarity of the available historical data. Overall, the Two-Tier method achieved smaller MSE for both $\pi$ and $\theta$, along with higher power compared to the "No borrow" method across different current sample sizes and different scenarios. This demonstrated that the Two-Tier method not only maintained the type I error around the theoretical value but also achieved higher power than the "No borrow" method. In contrast, the point estimations for the "Pool" method were heavily influenced by the difference in vaccine efficacy between historical source $A$ and the current study, making its performance highly dependent on the borrowed historical data. Regarding inference performance, the "Pool" method inflated the type I error badly under Scenario 1, rendering its power less meaningful. Under Scenario 2, the power of the Two-Tier and "Pool" methods was comparable when the current vaccine efficacy was small to moderate ($\pi = 0.5, 0.7$). However, when the current vaccine efficacy was high ($\pi = 0.9$), the power of the Two-Tier method exceeded that of the "Pool" method around 10%.

We conducted a sensitivity analysis to assess the impact of different threshold choices in the Two-Tier method. As expected, a smaller threshold ($H = 0.5$) led to more permissive borrowing, allowing the incorporation of historical information even when similarity to the current study was limited. In contrast, a larger threshold ($H = 0.9$) imposed more restrictive borrowing,

Table 3: Simulation results under the alternative hypothesis for Scenario 1.

| N | $\pi$ | Method | Borrowing | | | Point Estimation | | | | Inference |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\theta$ | | $\pi$ | | $\theta$ |
| | | | $\alpha_A T_A$ | $\alpha_B T_B$ | ESS | Bias | MSE | Bias | MSE | Power |
| 2400 | 0.5 | No borrow | – | – | 1.000 | 0.004 | 0.012 | −0.057 | 0.091 | 0.095 |
| | | Two-Tier | 0.001 | 6.687 | 7.688 | −0.021 | 0.007 | 0.023 | 0.038 | 0.149 |
| | | Pool | 6.667 | 1.333 | 9.000 | −0.069 | 0.011 | 0.125 | 0.038 | 0.344 |
| | 0.7 | No borrow | – | – | 1.000 | 0.008 | 0.012 | −0.044 | 0.052 | 0.331 |
| | | Two-Tier | 1.930 | 6.070 | 9.000 | −0.006 | 0.005 | −0.002 | 0.016 | 0.510 |
| | | Pool | 6.667 | 1.333 | 9.000 | −0.039 | 0.007 | 0.053 | 0.016 | 0.684 |
| | 0.9 | No borrow | – | – | 1.000 | 0.012 | 0.006 | −0.024 | 0.012 | 0.778 |
| | | Two-Tier | 4.151 | 3.849 | 9.000 | 0.031 | 0.003 | −0.042 | 0.006 | 0.941 |
| | | Pool | 6.667 | 1.333 | 9.000 | 0.011 | 0.002 | −0.017 | 0.004 | 0.970 |

Table 4: Simulation results under the alternative hypothesis for Scenario 2.

| N | $\pi$ | Method | Borrowing | | | Point Estimation | | | | Inference |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\theta$ | | $\pi$ | | $\theta$ |
| | | | $\alpha_A T_A$ | $\alpha_B T_B$ | ESS | Bias | MSE | Bias | MSE | Power |
| 2400 | 0.5 | No borrow | – | – | 1.000 | 0.005 | 0.013 | −0.062 | 0.104 | 0.104 |
| | | Two-Tier | 6.755 | 0.011 | 7.766 | −0.019 | 0.007 | 0.016 | 0.043 | 0.137 |
| | | Pool | 6.667 | 1.333 | 9.000 | −0.031 | 0.007 | 0.048 | 0.032 | 0.180 |
| | 0.7 | No borrow | – | – | 1.000 | 0.008 | 0.012 | −0.044 | 0.052 | 0.332 |
| | | Two-Tier | 5.421 | 2.578 | 8.998 | −0.007 | 0.005 | 0.000 | 0.016 | 0.510 |
| | | Pool | 6.667 | 1.333 | 9.000 | 0.002 | 0.005 | −0.016 | 0.017 | 0.456 |
| | 0.9 | No borrow | – | – | 1.000 | 0.012 | 0.006 | −0.024 | 0.013 | 0.785 |
| | | Two-Tier | 3.050 | 4.950 | 9.000 | 0.027 | 0.003 | −0.037 | 0.006 | 0.942 |
| | | Pool | 6.667 | 1.333 | 9.000 | 0.056 | 0.006 | −0.077 | 0.011 | 0.870 |

such that information was borrowed only when the historical trials were highly similar to the current study. Table 5 summarized the type I error performance of the Two-Tier method for $H = 0.5, 0.7$, and 0.9. Across both historical scenarios, when $H = 0.5$, the method tended to borrow more from the historical trial with more comparable vaccine efficacy: trial B in Scenario 1 ($\pi_B = 0.684$) and trial A in Scenario 2 ($\pi_A = 0.663$). These values were closer to that of the current study ($\pi = 0.3$ under the null hypothesis). However, because these historical trials had higher vaccine efficacy than the current study, borrowing leaded to slight type I error inflation, although performance remained superior to that of the "Pool" method. When $H = 0.9$, the Two-Tier method behaved similarly to the "No Borrow" approach. Borrowing was highly restricted, and under the null hypothesis where both historical trials were insufficiently similar to the current study, no borrowing occured.

Table 5: Sensitivity analysis over threshold $H$ for the Two-Tier method ($N = 2000$) under the null hypothesis.

| Scenario | $H$ | Borrowing | | | Point Estimation | | | | Inference |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\theta$ | | $\pi$ | | $\theta$ |
| | | $\alpha_A T_A$ | $\alpha_B T_B$ | ESS | Bias | MSE | Bias | MSE | Type I Error |
| **Scenario 1** | | | | | | | | | |
| | 0.5 | 0.021 | 7.186 | 8.206 | −0.051 | 0.010 | 0.104 | 0.072 | 0.049 |
| | 0.7 | 0.000 | 0.090 | 1.090 | 0.001 | 0.014 | −0.087 | 0.206 | 0.027 |
| | 0.9 | 0.000 | 0.000 | 1.000 | 0.002 | 0.014 | −0.092 | 0.217 | 0.027 |
| **Scenario 2** | | | | | | | | | |
| | 0.5 | 7.296 | 0.038 | 8.334 | −0.046 | 0.009 | 0.090 | 0.078 | 0.039 |
| | 0.7 | 0.096 | 0.000 | 1.096 | −0.002 | 0.014 | −0.076 | 0.191 | 0.027 |
| | 0.9 | 0.000 | 0.000 | 1.000 | 0.003 | 0.014 | −0.095 | 0.215 | 0.022 |

Table 6 presented the power performance of the Two-Tier method for $H = 0.5, 0.7$, and 0.9 under the alternative hypotheses, where the true vaccine efficacy in the current study was $\pi = 0.5, 0.7$, and 0.9. Overall, smaller values of $H$ led to larger prior ESS, reflecting greater borrowing. Consequently, when the true vaccine efficacy was relatively low ($\pi = 0.5$), the Two-Tier method with $H = 0.5$ achieved higher power than the settings with $H = 0.7$ and $H = 0.9$. When the true vaccine efficacy was higher ($\pi = 0.7$ or 0.9), the performance of $H = 0.5$ was similar to that of $H = 0.7$, as both thresholds achieved the maximum level of borrowing specified by $S = 9$. In contrast, $H = 0.9$ yielded more conservative borrowing. Although its prior ESS increased as the true vaccine efficacy increased, it did not reach the maximum borrowing level across all scenarios. As a result, it consistently exhibited lower power compared with the other two threshold choices.

In conclusion, the proposed Two-Tier method enables dynamic borrowing from historical trials based on their similarity to the current study. This approach provides improved control of type I error compared with the "Pool" method, while achieving higher power than both the "No Borrow" and "Pool" methods. Furthermore, the threshold $H$ influences the degree of borrowing restriction in the Two-Tier method. Larger values of $H$ impose more restrictive borrowing, whereas smaller values allow more permissive borrowing, resulting in different trade-offs between type I error control and efficiency.

## 4 Discussion

In this paper, we propose a Two-Tier Bayesian dynamic borrowing framework that enhances phase 3 vaccine studies by borrowing information from two historical data sources. The borrowing weights are determined by the similarity between current and historical sources, evaluated over an intermediate endpoint and gated by a pre-specified threshold. Therefore, the method can improve power and reduce bias. Moreover, by enforcing this pre-specified similarity threshold, the approach further controls the type I error. Overall, our framework serves as a valuable tool for designing confirmatory studies in pediatric populations, particularly when enrollment challenges prevent achieving the necessary sample size for confirmatory conclusions.

Table 6: Sensitivity analysis over threshold $H$ for the Two-Tier method ($N = 2000$) under the alternative hypotheses.

| Scenario | $\pi$ | $H$ | Borrowing | | | Point Estimation | | | | Inference |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\theta$ | | $\pi$ | | $\theta$ |
| | | | $\alpha_A T_A$ | $\alpha_B T_B$ | ESS | Bias | MSE | Bias | MSE | Power |
| **Scenario 1** | | | | | | | | | | |
| | 0.5 | 0.5 | 2.367 | 5.633 | 9.000 | −0.048 | 0.009 | 0.083 | 0.034 | 0.205 |
| | | 0.7 | 0.003 | 6.631 | 7.634 | −0.024 | 0.008 | 0.025 | 0.053 | 0.131 |
| | | 0.9 | 0.000 | 0.026 | 1.026 | 0.002 | 0.015 | −0.064 | 0.128 | 0.088 |
| | 0.7 | 0.5 | 3.507 | 4.493 | 9.000 | −0.020 | 0.005 | 0.021 | 0.016 | 0.525 |
| | | 0.7 | 1.954 | 6.046 | 9.000 | −0.009 | 0.005 | 0.003 | 0.017 | 0.466 |
| | | 0.9 | 0.000 | 4.397 | 5.397 | 0.007 | 0.009 | −0.036 | 0.041 | 0.334 |
| | 0.9 | 0.5 | 4.084 | 3.916 | 9.000 | 0.035 | 0.004 | −0.048 | 0.007 | 0.911 |
| | | 0.7 | 4.152 | 3.848 | 9.000 | 0.036 | 0.004 | −0.049 | 0.008 | 0.902 |
| | | 0.9 | 0.922 | 0.392 | 2.314 | 0.017 | 0.007 | −0.034 | 0.032 | 0.715 |
| **Scenario 2** | | | | | | | | | | |
| | 0.5 | 0.5 | 5.312 | 2.688 | 9.000 | −0.045 | 0.008 | 0.078 | 0.033 | 0.187 |
| | | 0.7 | 6.673 | 0.007 | 7.680 | −0.019 | 0.008 | 0.016 | 0.045 | 0.112 |
| | | 0.9 | 0.018 | 0.000 | 1.018 | 0.005 | 0.015 | −0.078 | 0.249 | 0.095 |
| | 0.7 | 0.5 | 4.332 | 3.668 | 9.000 | −0.019 | 0.005 | 0.020 | 0.015 | 0.517 |
| | | 0.7 | 5.348 | 2.652 | 9.000 | −0.010 | 0.005 | 0.005 | 0.017 | 0.476 |
| | | 0.9 | 1.237 | 0.000 | 2.237 | 0.013 | 0.012 | −0.056 | 0.060 | 0.281 |
| | 0.9 | 0.5 | 3.659 | 4.341 | 9.000 | 0.037 | 0.004 | −0.051 | 0.007 | 0.900 |
| | | 0.7 | 3.075 | 4.925 | 9.000 | 0.032 | 0.004 | −0.044 | 0.007 | 0.915 |
| | | 0.9 | 0.021 | 3.525 | 4.546 | 0.010 | 0.005 | −0.021 | 0.011 | 0.811 |

Our approach does not assume unrestricted exchangeability of vaccine efficacy between adult and pediatric populations. Instead, borrowing is allowed only when the adult and pediatric data are sufficiently similar on an intermediate endpoint that is reasonable likely to predict efficacy, and the amount of borrowing is further limited by the overlap-based weights and the pre-specified ESS cap. Thus, the method relies on a partial, biologically informed exchangeability assumption rather than full pooling across populations.

In the design, we require that the maximum targeted ESS to be pre-specified. Subsequently, the weights are dynamically determined based on the established intermediate endpoints, which calibrate how much to borrow from each historical data source. This strategy differs from other Bayesian designs that rely solely on similarity, and it facilitates transparent communication with regulatory agencies regarding the maximum information that can be borrowed in the phase 3 design. The similarity assessment in our method relies on the assumption that the intermediate endpoint, such as antibody concentration, is strongly predictive for the primary endpoint, i.e., vaccine efficacy in our study. When applying the proposed Two-Tier method, determining an appropriate intermediate endpoint is essential. In practice, the intermediate endpoint is usually established as the Correlate of Protection upon completion of the adult registrational study, so that it would be available for pediatric development as proposed in the Two-Tier approach.

Another important design parameter is the threshold $H \in [0, 1]$, which influences the degree of borrowing in the Two-Tier method. Values of $H$ close to 1 impose more restrictive borrowing, whereas values close to 0 allow more permissive borrowing. The choice of $H$ can be determined during the study design stage through simulation based on historical trial information. Specifically, type I error and power can be evaluated under a fixed value of $H$, and different threshold values can be explored to identify a setting that achieves acceptable type I error control while maintaining adequate power. Although not implemented in the current paper, the similarity threshold may be determined in a decision theory framework. Under this approach, the trialist must specify a plausible parameter space for the $\pi$'s and select the optimal threshold that attains minimax (or average) power while controlling the type I error rate. The minimax (or average) rule is necessary because the threshold attaining the maximum power can vary depending on the specified $\pi$'s.

The Beta form in Equation (4) relies on the assumption that the historical studies and the current study share a common parameter $\theta$. When this assumption is relaxed and the historical parameters $\theta_A$ and $\theta_B$ differ from $\theta$, the induced prior for $\theta$ is generally no longer Beta and does not admit a simple closed form. In such cases, posterior inference would require numerical integration or Markov chain Monte Carlo methods. In this work, we adopt Equation (4) as a working power-prior specification and address potential violations of this assumption through the borrowing threshold $H$. Specifically, we quantify the similarity between the historical trials and the current study and regulate borrowing via $H$, such that lower similarity leads to reduced borrowing. This mechanism effectively discounts the assumption that $\theta_A$, $\theta_B$, and $\theta$ are identical, thereby providing robustness to departures from exchangeability.

The choice of $N_0 = 50$ in the simulation represents the sample size of the earlier phase immunogenicity data used only for estimating the similarity metric. While this sample size yielded stable operating characteristics in our numerical study, smaller $N_0$ may produce noisier OVL estimates and less stable threshold-based borrowing decisions; conversely, larger $N_0$ would improve the stability of similarity assessment. The threshold $H$ and ESS cap $S$ help mitigate the downstream impact of such variability on the analysis.

Future work could extend this methodology in several directions. First, the proposed Two-Tier framework can be generalized to other Bayesian borrowing methods, such as robust mixture priors, commensurate priors, and MAP priors. Such extensions require careful development to ensure that the total ESS is restricted to a pre-specified level and that borrowing is dynamically adjusted based on similarity. Second, the approach could be readily extended to incorporate multiple historical sources or could be tailored to borrow only from control arms. Similar to the case with two historical trials, the maximum prior ESS is capped at a pre-specified value, representing the total amount of information borrowed from all historical trials. Each historical trial is assigned a trial-specific power parameter. These trial-specific power parameters are determined based on the similarity of the intermediate outcome to that of the current study. The method borrows information only from historical trials that exceed a pre-specified similarity threshold, and greater borrowing occurs from trials that are relatively more similar to the current study. Similar to Equation (5), each historical trial contributes its infection cases with a corresponding trial-specific power parameter to the prior ESS.

Moreover, we consider only a univariate intermediate outcome for similarity evaluation in the current framework. The method can be extended to multivariate scenarios. One possible approach is to first perform dimension reduction and then compute the OVL as the similarity measure. Alternatively, the similarity measure could be estimated using other distance metrics, such as the Mahalanobis distance, which is commonly used to measure the covariates distance for

matching designs in causal inference (Rosenbaum, 2020). In addition, alternative methods can be explored to determine power parameters and adaptive weights. Another interesting direction is the introduction of a specified sequence of borrowing based on priori belief, such as recent statistical work by Han et al. (2025).

## Supplementary Material

The R code used to reproduce the simulation results is provided as online Supplementary Material.

## Acknowledgement

The authors would like to thank Qiqi Deng and Qian (Helen) Li for helpful discussions. We are also grateful to the Editor, Associate Editor, and Referee for their insightful comments and constructive feedback.

## References

Baumann L, Sauer LD, Kieser M (2025). A basket trial design based on power priors. *Statistics in Biopharmaceutical Research*, 17(3): 446–456. https://doi.org/10.1080/19466315.2024.2402275.

Callegaro A, Karkada N, Aris E, Zahaf T (2023). Vaccine clinical trials with dynamic borrowing of historical controls: Two retrospective studies. *Pharmaceutical Statistics*, 22(3): 475–491. https://doi.org/10.1002/pst.2283.

Callegaro A, Luo Y, Karkada N, Zahaf T (2024). Dynamic borrowing of historical controls adjusting for covariates in vaccine efficacy clinical trials. *Pharmaceutical Statistics*, 23(5): 630–644. https://doi.org/10.1002/pst.2384.

Chan IS, Bohidar NR (1998). Exact power and sample size for vaccine efficacy studies. *Communications in Statistics – Theory and Methods*, 27(6): 1305–1322. https://doi.org/10.1080/03610929808832160.

Chen MH, Guan Z, Lin M, Sun M (2025). Power priors for leveraging historical data: Looking back and looking forward. *Journal of Data Science*, 23(1): 1–30. https://doi.org/10.6339/24-JDS1161.

Chen MH, Guan Z, Sun M, Zhong J (2024). A unified Bayesian decision rule-based approach for Bayesian design of clinical trials using historical data. In: *Biostatistics in Biopharmaceutical Research and Development: 80-Years of Contributions and Dedications from the Deming Conference* (DG Chen, ed.), 195–217. Springer, New York, NY.

Chen MH, Ibrahim JG, Lam P, Yu A, Zhang Y (2011). Bayesian design of non-inferiority trials for medical devices using historical data. *Biometrics*, 67(3): 1163–1170. https://doi.org/10.1111/j.1541-0420.2011.01561.x.

Chen MH, Ibrahim JG, Zeng D, Hu K, Jia C (2014). Bayesian design of superiority clinical trials for recurrent events data with applications to bleeding and transfusion events in myelodyplastic syndrome. *Biometrics*, 70(4): 1003–1013. https://doi.org/10.1111/biom.12215.

De Santis F, Gubbiotti S (2023). Borrowing historical information for non-inferiority trials on COVID-19 vaccines. *The International Journal of Biostatistics*, 19(1): 177–189. https://doi.org/10.1515/ijb-2021-0120.

Gamalo-Siebers M, Savic J, Basu C, Zhao X, Gopalakrishnan M, Gao A, et al. (2017). Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharmaceutical Statistics*, 16(4): 232–249. https://doi.org/10.1002/pst.1807.

Han Z, Zhang Q, Tiwari R, Bai T (2025). Bayesian borrowing with multiple heterogeneous historical studies using order restricted normalized power prior. *Statistics in Medicine*, 44(3–4): e10302. https://doi.org/10.1002/sim.10302.

Han Z, Zhang Q, Wang M, Ye K, Chen MH (2023). On efficient posterior inference in normalized power prior Bayesian analysis. *Biometrical Journal*, 65(5): 2200194 https://doi.org/10.1002/bimj.202200194.

Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3): 1047–1056. https://doi.org/10.1111/j.1541-0420.2011.01564.x.

Huang Y, Hejazi NS, Blette B, Carpp LN, Benkeser D, ..., Gilbert PB (2023). Stochastic interventional vaccine efficacy and principal surrogate analyses of antibody markers as correlates of protection against symptomatic COVID-19 in the COVE mRNA-1273 trial. *Viruses*, 15(10): 2029. https://doi.org/10.3390/v15102029.

Ibrahim JG, Chen MH (2000). Power prior distributions for regression models. *Statistical Science*, 15(1): 46–60. https://doi.org/10.1214/ss/1009212673.

Inman HF, Bradley JEL (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics-Theory and Methods*, 18(10): 3851–3874. https://doi.org/10.1080/03610928908830127.

Jeffreys H (1998). *Theory of Probability*. Oxford University Press.

Jiang Y, Simon S, Mayo MS, Gajewski BJ (2015). Modeling and validating Bayesian accrual models on clinical data and simulations using adaptive priors. *Statistics in Medicine*, 34(4): 613–629. https://doi.org/10.1002/sim.6359.

Jin M, Feng D, Liu G (2020). Bayesian approaches on borrowing historical data for vaccine efficacy trials. *Statistics in Biopharmaceutical Research*, 12(3): 284–292. https://doi.org/10.1080/19466315.2020.1736617.

Jin M, Li Q, Kaur A (2021). Bayesian design for pediatric clinical trials with binary endpoints when borrowing historical information of treatment effect. *Therapeutic Innovation & Regulatory Science*, 55(2): 360–369. https://doi.org/10.1007/s43441-020-00220-5.

Li W, Chen MH, Tan H, Dey DK (2015). Bayesian design of non-inferiority clinical trials with co-primary endpoints and multiple dose comparison. In: *Applied Statistics in Biomedicine and Clinical Trials Design – Selected Papers from 2013 ICSA/ISBS Joint Statistical Meetings* (Z Chen, A Liu, Y Qu, L Tang, N Ting, Y Tsong, eds.), 17–33. Springer, New York, NY.

Li W, Chen MH, Wang X, Dey DK (2018). Bayesian design of non-inferiority clinical trials via the Bayes factor. *Statistics in Biosciences*, 10(2): 439–459. https://doi.org/10.1007/s12561-017-9200-5.

Liu M, Bunn V, Hupf B, Lin J, Lin J (2021). Propensity-score-based meta-analytic predictive prior for incorporating real-world and historical data. *Statistics in Medicine*, 40(22): 4794–4808. https://doi.org/10.1002/sim.9095.

Malec D (2001). A closer look at combining data among a small number of binomial experiments. *Statistics in Medicine*, 20(12): 1811–1824. https://doi.org/10.1002/sim.782.

Morita S, Thall PF, Müller P (2008). Determining the effective sample size of a parametric prior. *Biometrics*, 64(2): 595–602. https://doi.org/10.1111/j.1541-0420.2007.00888.x.

Morita S, Thall PF, Müller P (2012). Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis*, 7(3): 591–614. https://doi.org/10.1214/12-BA720.

Moss J, Tveten M (2019). kdensity: An R package for kernel density estimation with parametric starts and asymmetric kernels. *Journal of Open Source Software*, 4(42): 1566. https://doi.org/10.21105/joss.01566.

Neuenschwander B, Weber S, Schmidli H, O'Hagan A (2020). Predictively consistent prior effective sample sizes. *Biometrics*, 76(2): 578–587. https://doi.org/10.1111/biom.13252.

Peng L, Jin J, Chambonneau L, Zhao X, Zhang J (2023). Bayesian borrowing from historical control data in a vaccine efficacy trial. *Pharmaceutical Statistics*, 22(5): 815–835. https://doi.org/10.1002/pst.2313.

Psioda MA, Bean NW, Wright BA, Lu Y, Mantero A, Majumdar A (2025). Inverse probability weighted Bayesian dynamic borrowing for estimation of marginal treatment effects with application to hybrid control arm oncology studies. *Journal of Biopharmaceutical Statistics*, 35(6): 1–23. https://doi.org/10.1080/10543406.2025.2489285.

Qian R, Yang B, Xu X, Lu B (2025). Matching-assisted power prior for incorporating real-world data in randomized clinical trial analysis. *Statistics in Medicine*, 44(3–4): e10342.

Rosenbaum PR (2020). *Design of Observational Studies Springer Series in Statistics*. Springer, New York.

Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4): 1023–1032. https://doi.org/10.1111/biom.12242.

Schoenfeld DA, Zheng H, Finkelstein DM (2009). Bayesian design using adult data to augment pediatric trials. *Clinical Trials*, 6(4): 297–304. https://doi.org/10.1177/1740774509339238.

US Food and Drug Administration (2010). Guidance for the use of Bayesian statistics in medical device clinical trials. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials. Online. Accessed October 20, 2025.

US Food and Drug Administration (2016). Leveraging existing clinical data for extrapolation to pediatric uses of medical devices. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/leveraging-existing-clinical-data-extrapolation-pediatric-uses-medical-devices. Online. Accessed October 20, 2025.

US Food and Drug Administration (2018). Framework for FDA's real-world evidence program. https://www.fda.gov/media/120060/download. Online. Accessed October 20, 2025.

US Food and Drug Administration (2021). Advancing the development of pediatric therapeutics complex innovative trial design public workshop. https://www.fda.gov/drugs/news-events-human-drugs/fda-m-cersi-advancing-development-pediatric-therapeutics-complex-innovative-trial-design-public. Online. Accessed October 20, 2025.

US Food and Drug Administration (2023a). Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug. Online. Accessed October 20, 2025.

US Food and Drug Administration (2023b). Pediatric drug development: Regulatory considerations – complying with the pediatric research equity act and qualifying for pediatric exclusivity under the best pharmaceuticals for children act. https://www.fda.gov/media/168201/download. Online. Accessed October 20, 2025.

US Food and Drug Administration (2023c). Real-world data: Assessing registries to support regulatory decision-making for drug and biological products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-registries-support-regulatory-decision-making-drug-and-biological-products. Online. Accessed October 20, 2025.

US Food and Drug Administration (2024). Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory. Online. Accessed October 20, 2025.

US Food and Drug Administration (2026). Use of Bayesian methodology in clinical trials of drug and biological products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-bayesian-methodology-clinical-trials-drug-and-biological-products. Online. Accessed March 11, 2026.

US Government Publishing Office (2016). 21st Century Cures Act. https://www.govinfo.gov/content/pkg/PLAW-114publ255/html/PLAW-114publ255.htm. Online. Accessed October 20, 2025.

van Rosmalen J, Dejardin D, van Norden Y, Löwenberg B, Lesaffre E (2018). Including historical data in the analysis of clinical trials: Is it worth the effort? *Statistical Methods in Medical Research*, 27(10): 3167–3182. https://doi.org/10.1177/0962280217694506.

Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, 13(1): 41–54. https://doi.org/10.1002/pst.1589.

Wang C, Li H, Chen WC, Lu N, Tiwari R, ..., Yue LQ (2019). Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. *Journal of Biopharmaceutical Statistics*, 29(5): 731–748. https://doi.org/10.1080/10543406.2019.1657133.

Yuan W, Chen MH, Zhong J (2022). Flexible conditional borrowing approaches for leveraging historical data in the Bayesian design of superiority trials. *Statistics in Biosciences*, 14(2): 197–215. https://doi.org/10.1007/s12561-021-09321-7.