POLITECNICO DI TORINO
Repository ISTITUZIONALE

On the Interplay Between Prior Weight and Variance of the Robustification Component in Robust Mixture Prior Bayesian Dynamic Borrowing Approach

Original
On the Interplay Between Prior Weight and Variance of the Robustification Component in Robust Mixture Prior Bayesian Dynamic Borrowing Approach / Ratta, M., Sainthilary, G., Gasparini, M., Mozgunov, P.. - In: STATISTICS IN MEDICINE. - ISSN 1097-0258. - 45:13-14(2026), pp. 1-14. [10.1002/sim.70616]

Availability:
This version is available at: 11583/3016822 since: 2026-10-20T10:07:44Z

Publisher:
John Wiley and Sons

Published
DOI:10.1002/sim.70616

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

15 November 2026

Check for updates

Statistics in Medicine

WILEY

Statistics in Medicine

RESEARCH ARTICLE OPEN ACCESS

On the Interplay Between Prior Weight and Variance of the Robustification Component in Robust Mixture Prior Bayesian Dynamic Borrowing Approach

Marco Ratta[1,2] | Gaëlle Saint-Hilary[2] | Mauro Gasparini[1] | Pavel Mozgunov[2,3]

[1]Department of Mathematical Sciences "G.L. Lagrange", Politecnico di Torino, Torino, Italy | [2]Department of Statistical Methodology, Saryga, Paris, France | [3]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Correspondence: Marco Ratta (marco.ratta@polito.it)

Received: 2 July 2025 | **Revised:** 15 April 2026 | **Accepted:** 13 May 2026

Keywords: Bayesian dynamic borrowing | Bayesian methods | clinical trials | Lindley's paradox | robust mixture prior

ABSTRACT

Robust Mixture Prior (RMP) is a popular Bayesian dynamic borrowing method, which combines an informative historical distribution with a less informative component (referred to as the *robustification component*) in a mixture prior to enhance the efficiency of hybrid-control randomized trials. Current practice typically focuses solely on the selection of the prior weight that governs the relative influence of these two components, often fixing the variance of the robustification component to a unit-information variance. In this study we demonstrate that the performance of RMPs critically depends on the *joint* selection of both weight and variance of the robustification component. In particular, we show that a wide range of weight-variance pairs can yield practically identical posterior inferences (in particular regions of the parameter space) and that large variance robust components may be employed without incurring the so called *Lindley's paradox*. We further show that the use of large variance robustification components leads to improved asymptotic type I error rate control and enhanced robustness of the RMP to the specification of the location parameter of the robustification component. Finally, we leverage these theoretical results to propose a novel and practical hyper-parameter elicitation routine.

1 | Introduction

Leveraging historical information in clinical trials is particularly valuable in contexts like rare diseases [1] and pediatric trials [2–4], where recruiting large patient populations is challenging. Bayesian designs are appealing as they allow incorporating available knowledge into prior distributions. However, including external data raises challenges, such as quantifying heterogeneity between external and current data, which can lead to biased estimates and poor operating characteristics if not properly addressed.

Bayesian dynamic borrowing (BDB) sets out to solve such issue by dynamically discounting the use of external information based on a measure of heterogeneity between the prior distribution and the observed data. Several borrowing strategies have been proposed over the years such as Power priors [5, 6], commensurate priors [7] and Robust Mixture Prior (RMP) [8, 9], all of them requiring the specification of a tuning parameter quantifying the amount of borrowing (called *knowledge factor* in an early non-clinical reference [8]). A thorough review of the available borrowing methods can be found in Van Rosmalen et al. [10] and Viele et al. [11]. Among them, Robust Mixture Prior (RMP) [9, 12], is acknowledged as one of the most versatile options due to its natural ability of dynamically discounting the amount of borrowed information as the prior-data conflict increases. Examples of practical use of RMP in different contexts of application can be found in literature, for example, bringing adult information to inform treatment

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Statistics in Medicine* published by John Wiley & Sons Ltd.

effect on a pediatric trial [4], exploiting expert opinion to inform a prior distribution for a treatment effect [9], borrowing historical information to predict a treatment effect on a primary endpoint based on a surrogate endpoint [13, 14] or borrowing external control data to discount sample size in the control arm [15].

The idea behind RMP is to construct a prior distribution for the parameter of interest by combining an informative component, derived from external information, and a *robustification* high-variance component in a mixture distribution. The advantage of this approach is that the information contained in the informative component of the mixture impacts the posterior inference in a dynamic way, that is, mostly in case of agreement between historical and current data, while it is progressively disregarded as the prior-data conflict increases [12].

The main object of investigation of this paper are robust mixtures of normal priors, called normal RMPs, which are vastly used in case of normally distributed (or approximately normally distributed) endpoints. In particular, we will focus on the case in which the informative component of the RMP is a single normal distribution with known mean and variance, and is combined with a robust normal component with higher variance. In this context, three parameters must be specified, namely

i. *weight* of the robustification component of the mixture prior,
ii. *location* of the robustification component and
iii. *variance* of the robustification component.

Although it has been shown that all these three factors impact the operating characteristics (see Weru et al. [16]), it is common to focus solely on the selection of the mixture weight related to the informative component (referred to as "mixture weight"), regulating the amount of information to be borrowed. The latter is commonly pre-specified based on the stakeholder degree of confidence in the historical source, while all the other parameters are commonly fixed. For the variance of the robustification component of the mixture it has been argued that extremely large variances should be avoided [9, 16, 17], as they can lead to borrowing of historical information even in case of extreme inconsistency between historical and concurrent data. To avoid this situation, robust weakly informative components have generally been preferred and unit information priors (UIP) [12] have become a common choice. Using weakly informative robustification components, however, has some drawbacks, in particular

i. it is sensitive to the choice of the location of the robustification component [16], and
ii. it causes an inflation of type I error rate in case of the major inconsistency between historical and current data.

In this work, we demonstrate that the borrowing properties of the RMP are defined by the *joint* specification of prior weight and variance of the robustification component and these two parameters should be chosen together. We theoretically demonstrate that RMP with high-variance robustification components is a viable choice, provided a jointly optimized selection of prior weight and variance of the robustification component. We argue that this approach is advantageous as

i. it practically makes the choice of the location of the robustification component impactless and
ii. it effectively prevents from the asymptotic inflation of the type I error rate, which arises—in the case of weakly informative robustification components—when major inconsistency between historical and current data is observed.

The manuscript is organized as follows: Sections 2 to 6 focus on the normal setting. Specifically, Section 2 introduces the RMP model and its application in the normal setting; Section 3 presents the motivation for this work; Section 4 details the theoretical findings for the normal setting; Section 5 provides a proof-of-concept analysis highlighting the key benefits of the proposed methodology; and Section 6 outlines a novel procedure for hyper-parameter selection. Section 7 discusses the extension to the binary case with the Beta RMP, while Section 8 presents the extension to scenarios in which the informative component of the RMP is itself a mixture. Finally, Section 9 concludes with a discussion.

2 | Methodology

2.1 | Setting

2.1.1 | Bayesian Design of a Randomized Controlled Trial (RCT)

Consider a randomized controlled trial (RCT) evaluating a novel treatment against placebo or standard of care. Let X_t and X_c denote the normally distributed mean treatment and control responses with unknown means θ_t and θ_c, and known variances $\sigma_t^2 = s^2/n_t$ and $\sigma_c^2 = s^2/n_c$, where s is the common variance of individual responses and n_j $(j = t, c)$ the arm-specific sample sizes.

The treatment effect $\delta = \theta_t - \theta_c$ is the parameter of interest, with $H_0 : \delta = 0$ tested against $H_A : \delta > 0$. Priors $\pi_t(\cdot)$ and $\pi_c(\cdot)$ are specified for θ_t and θ_c.

Trial success is declared when the posterior probability of a positive treatment effect exceeds a prespecified threshold:

$$\mathbb{P}_{\pi_c,\pi_t}\left(\delta > 0 \mid x_c, x_t\right) > 1 - \eta, \tag{1}$$

where x_c and x_t are observed mean responses. The threshold $1 - \eta$ represents the required posterior evidence for efficacy; with smaller η values imply more stringent criteria.

2.1.2 | Frequentist and Bayesian Operating Characteristics

The *type I error rate*, the probability of rejecting H_0 when $\delta = 0$, is computed by integrating the success condition over the data likelihoods:

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\alpha(H) = \iint_{\mathbb{R}^2} \mathbb{1}\left\{\mathbb{P}_{\pi_c,\pi_t}(\delta > 0 | x_c, x_t) > 1 - \eta\right\} f_{X_c}(x_c|\theta_c = H) f_{X_t}(x_t|\theta_t = H)\, dx_c\, dx_t, \quad (2)$$

where $\mathbb{1}(\cdot)$ is the indicator function, and f_{X_c}, f_{X_t} denote the sampling distributions. *Power* is obtained analogously under $\theta_t = H + \delta^*$ and $\theta_c = H$, for a target effect $\delta^* > 0$.

Type I error rate and power are frequentist quantities, as they condition on fixed parameter values. To assess Bayesian designs more comprehensively, Best et al. [18] proposed averaging α over a *design prior* Π_c, namely:

$$\alpha_{\text{avg}}^{\Pi_c} = \int_{\mathbb{R}} \alpha(t)\, \Pi_c(t)\, dt. \quad (3)$$

A *design prior* is the prior distribution used during the *planning* of the trial to reflect plausible values for the parameters, which allows evaluation of Bayesian operating characteristics such as average Type I error and power. It is not necessarily the same as the prior used in the *analysis* of the trial, which represents the formal beliefs applied to the data once observed. The design prior is primarily a tool for trial design and simulation, whereas the analysis prior is used for inference and decision-making.

2.1.3 | Posterior Estimation Metrics

Besides testing, performance is evaluated through estimation metrics. The posterior median $\hat{\delta}$ serves as point estimate, and bias, variance, and mean squared error (MSE) quantify its accuracy (see Supporting Information for formulas).

2.2 | Robust Mixture Prior (RMP)

Let $\pi_{\text{inf}}(\cdot)$ be an informative prior for θ_c. The *Robust Mixture Prior (RMP)* combines this with a weakly informative or non-informative robustification component $\pi_{\text{rob}}(\cdot)$:

$$\pi_c(\theta_c) = \omega\, \pi_{\text{inf}}(\theta_c) + (1 - \omega)\, \pi_{\text{rob}}(\theta_c), \quad (4)$$

where $\omega \in [0, 1]$ is the prior weight on the informative component. The robustification term downweights historical information when inconsistent with current data.

After observing x_c, the posterior is again a mixture:

$$g(\theta_c|x_c) = \tilde{\omega}\, g_{\text{inf}}(\theta_c|x_c) + (1 - \tilde{\omega})\, g_{\text{rob}}(\theta_c|x_c), \quad (5)$$

where each component posterior is $g_\star(\theta_c|x_c) = f(x_c|\theta_c)\pi_\star(\theta_c)/f(x_c|\pi_\star)$, with $\star \in \{\text{inf}, \text{rob}\}$. The updated weight depends on x_c via the formula

$$\tilde{\omega}(x_c) = \frac{\omega f(x_c|\pi_{\text{inf}})}{\omega f(x_c|\pi_{\text{inf}}) + (1 - \omega) f(x_c|\pi_{\text{rob}})}. \quad (6)$$

A proof of Equation (5) and (6) is in the Supporting Information.

Equation (6) can be expressed equivalently in terms of odds as

$$\tilde{\Omega}(x_c) = \Omega \frac{f(x_c|\pi_{\text{inf}})}{f(x_c|\pi_{\text{rob}})}, \quad (7)$$

with $\Omega = \omega/(1 - \omega)$ and $\tilde{\Omega} = \tilde{\omega}/(1 - \tilde{\omega})$. It can be noticed that weights (and odds) adjust borrowing dynamically according to the data's compatibility with prior information, namely increases when the observed response x_c is compatible with the informative component of the mixture while decreases otherwise.

Note that in Equations (6) and (7), posterior weights and posterior odds are well-defined functions of the observed mean response, conditional on the specified RMP for θ_c. For simplicity, this dependence will be implicitly understood in subsequent sections and explicitly stated only when necessary.

2.3 | Normal Robust Mixture Prior

When both mixture components are Normal,

$$\pi_{\text{inf}}(\theta_c) = \mathcal{N}(\mu_{\text{inf}}, \sigma_{\text{inf}}^2), \quad \pi_{\text{rob}}(\theta_c) = \mathcal{N}(\mu_{\text{rob}}, \sigma_{\text{rob}}^2 = s^2/n_0),$$

the conjugacy ensures that the posterior remains a Normal mixture with updated parameters.

Moreover, the corresponding prior predictive distributions are also Normal:

$$f(x_c|\pi_\star) = \frac{1}{\sqrt{2\pi v_\star^2}} \exp\left[-\frac{(x_c - \mu_\star)^2}{2v_\star^2}\right], \quad v_\star^2 = \sigma_\star^2 + \sigma_c^2, \quad (8)$$

for $\star \in \{\text{inf}, \text{rob}\}$. As a consequence, letting $R = v_{\text{rob}}/v_{\text{inf}}$, then Equation (7) becomes

$$\tilde{\Omega}(x_c) = \beta(\omega, \sigma_{\text{rob}}^2) \exp\left\{-\frac{d^2}{2v_{\text{inf}}^2} + \frac{(x_c - \mu_{\text{rob}})^2}{2R^2 v_{\text{inf}}^2}\right\}. \quad (9)$$

In the latter, $\beta\left(\omega, \sigma_{\text{rob}}^2\right) = \Omega \times R$, while d represents the realization of the random variable $X_c - \mu_{\text{inf}} \sim \mathcal{N}\left(D, \sigma_c^2\right)$, with mean D representing the true *drift* parameter (also referred to as *prior-data conflict* hereinafter), indicating the level of inconsistency between concurrent data and historical information provided in the informative component of the RMP. Note that defining the function $\beta(\cdot)$ will become useful in Section 4.3.

Equation (9) shows that the posterior odds $\tilde{\Omega}$ depend on the choice of Ω (which is a deterministic function of the prior weight ω), the location parameter of the robustification component μ_{rob} and the variance of the robustification component σ_{rob}^2.

Notice that, since the robustification component must be less informative than the informative one, $R > 1$ (often $R \gg 1$ when π_{rob} is nearly non-informative).

3 | Motivation for the Work

3.1 | Background

Robust Mixture Priors (RMPs) are widely applied in randomized controlled trials (RCTs) to borrow information for the control arm [15, 18, 19]. Several approaches exist for specifying the mixture

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

weight ω [20, 21], yet the selection of hyperparameters for the robustification component has received limited attention.

Large variances for the robustification prior are often adopted to represent minimal prior knowledge; however, such weakly informative choices may retain excessive influence of the informative component even under strong priordata conflictan effect known as *Lindleys paradox* [9, 16, 17]. Schmidli et al. [12] proposed mitigating this through a *unit-information prior* (UIP), namely a distribution which effective sample size (ESS) [22] is equal to 1.

While practical and commonly used, this approach introduces two main challenges: (i) the pre-specification of the robustification mean μ_{rob}, which strongly affects posterior inference [16]; and (ii) the asymptotic inflation of the Type I error in the presence of substantial discrepancies between the historical and current control data [16, 18]. Here, the term *asymptotic inflation* refers to the progressive increase in the Type I error rate as the drift parameter D increases, such that the Type I error approaches 1 as $D \to +\infty$.

The following case study illustrates these issues in Normal RMPs within hybrid-control RCTs, providing the basis for the theoretical developments in Section 4.

3.2 | Illustration in a Hypothetical Trial

Consider a two-arm RCT comparing treatment and control (placebo or standard of care). Individual outcomes in both arms follow normal distributions with unit variance ($s = 1$), as a consequence the mean responses in the two arms are:

$$X_t \sim \mathcal{N}(\theta_t, n_t^{-1}), \quad X_c \sim \mathcal{N}(\theta_c, n_c^{-1}).$$

The trial allocates $n_t = 150$ patients to treatment and $n_c = 50$ to control (3:1 ratio). Trial success is defined by Equation (1) with $\eta = 0.05$.

No prior information is available for θ_t, so a non-informative prior $\theta_t \sim \mathcal{N}(\mu_{\text{rob}}, n_0^{-1})$ is used. For θ_c, an informative prior $\mathcal{N}(\mu_{\text{inf}}, n_{\text{inf}}^{-1})$ with effective sample size $n_{\text{inf}} = 100$ and mean $\mu_{\text{inf}} = 0$ is combined with a non-informative prior $\mathcal{N}(\mu_{\text{rob}}, n_0^{-1})$ through an RMP with weight ω.

Performance metrics include the type I error rate (Equation 2), power (for target $\delta^* = 0.31$), and the average posterior weight $\tilde{\omega}$, obtained by integrating Equation (6) over the data likelihood.

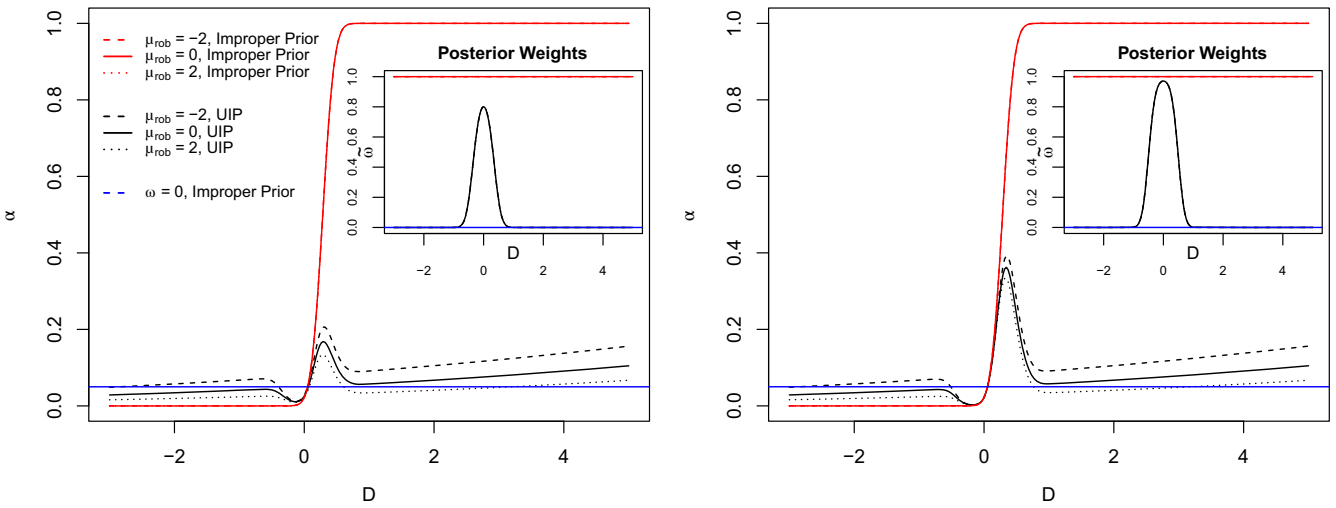
Different RMP configurations are examined, considering mixture weights $\omega \in \{0.5, 0.9\}$ to represent, respectively, moderate and strong confidence in the historical information. Six sub-scenarios are defined by varying the hyperparameters of the robustification component. Specifically, the location parameter is set to $\mu_{\text{rob}} \in \{-2, 0, 2\}$, while the variance takes values $\sigma^2_{\text{rob}} \in \{1, 10^{100}\}$, the former corresponding to a unit-information prior and the latter approximating an improper prior. A reference setting with $\omega = 0$ and $\sigma^2_{\text{rob}} = 10^{100}$ represents a standard non-informative Bayesian design. Performance metrics are assessed across a range of *drift* values D.

3.3 | Analysis

Figure 1 displays the type I error rate as a function of the drift parameter D for $\omega = 0.5$ (left) and $\omega = 0.9$ (right), under varying μ_{rob} and σ^2_{rob}.

When a UIP is used as robustification component, type I error rate decreases near $D \approx 0$, reflecting improved borrowing when historical and current data agree ($\tilde{\omega} \gg 0$). As $|D|$ increases, borrowing diminishes; however, intermediate drifts can still yield residual borrowing ($\tilde{\omega} > 0$), biasing control estimates and inflating type I error rate for positive drifts or deflating it for negative ones.

Under extreme priordata conflict ($|D|$ large), borrowing vanishes ($\tilde{\omega} \approx 0$), yet instead of stabilizing near the nominal level, type I



(a) type I error rate with $\omega = 0.5$ (b) type I error rate with $\omega = 0.9$

FIGURE 1 | Type I error rate $\alpha(D)$ under different RMP parameterizations. Red curves: improper priors ($\sigma^2_{\text{rob}} = 10^{100}$). Black curves: unit-information priors ($\sigma^2_{\text{rob}} = 1$). Line styles denote values of μ_{rob}. Panel (1a): $\omega = 0.5$; Panel (1b): $\omega = 0.9$.

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

error rate asymptotically diverges toward 1 for $D \to +\infty$ and 0 for $D \to -\infty$. This counterintuitive behavior motivates further theoretical investigation.

Figure 1 also shows that, although all UIP-based RMPs share similar asymptotic trends, the choice of μ_{rob} systematically shifts type I error rate: larger μ_{rob} increases it uniformly across D, while smaller values decrease it. This sensitivity to μ_{rob} forms a second point of interest.

When the robustification component is nearly improper ($\sigma^2_{\text{rob}} = 10^{100}$), borrowing persists regardless of how strong the priordata conflict is ($\tilde{\omega} = 1$ across all the D-space), illustrating *Lindleys paradox* [9, 16, 17]. Here, type I error rate remains near 0 for $D < -0.2$, increases sharply to 1 for $-0.2 \leq D \leq 0.5$, and stays at this level thereafter, with negligible dependence on μ_{rob}.

3.4 | Research Questions

In Section 3.3 we have shown that there are some issues related to the use of RMP in the context of hybrid control RCT. These are:

1. The asymptotic inflation of type I error for large positive values of prior-data conflict, when weakly informative robustification components are employed.
2. The sensitivity of the operating characteristics to the choice of μ_{rob}, when weakly informative robustification components are employed.
3. The apparent failure in discounting information borrowing as the prior-data conflict increases, when large variance robustification components are used (Lindley's paradox).

In the next sections the cause of these issues will be theoretically investigated, and a solution to all of them will be proposed.

4 | Analytical Results

4.1 | Asymptotic Inflation of Type I Error Rate

The cause of the asymptotic type I error rate inflation, along with the conditions under which the latter is prevented are investigated in Theorem 1. In particular, it is proven that type I error rate inflation occurs when an upwards bias is induced by the robustification component π_{rob} of the RMP on the posterior mean for the treatment difference. For a fixed value of the mixture weight ω, this bias is inversely proportional to the variance of the robustification component σ^2_{rob}, and in particular it is null if the latter diverges to $+\infty$ at least as fast as the drift parameter D. Under this condition, an asymptotic control of the type I error rate is achieved, thus making the choice of large variance robustification components in RMPs particularly attractive.

Theorem 1. *Consider a RCT where mean control and treatment responses are normal* $X_c \sim \mathcal{N}(\theta_c, \sigma_c^2)$, $X_t \sim \mathcal{N}(\theta_t, \sigma_t^2)$, *and assume* $\sigma_t^2 = K\sigma_c^2$ *(where* K^{-1} *is the randomization ratio, assumed* > 1*). Assume a RMP* $\pi_c(\theta_c) = \omega\pi_{\text{inf}}(\theta_c) + (1-\omega)\pi_{\text{rob}}(\theta_c)$ *is used for the control parameter, where* $\pi_{\text{inf}}(\theta_c)$ *and* $\pi_{\text{rob}}(\theta_c)$ *are the PDF of normally distributed random variables with parameters* μ_{inf}, σ^2_{inf} *and* μ_{rob}, σ^2_{rob} *respectively; while a normal prior distribution* $\theta_t \sim \mathcal{N}(\mu_t, \sigma^2_{\text{rob}})$ *is given to the treatment parameter. Consider the type I error rate* $\alpha(\cdot)$ *as defined in Equation (2), corresponding to the null hypothesis* $H_0 : \theta_c = \theta_t = D + \mu_{\text{inf}}$, *where* $D = \theta_c - \mu_{\text{inf}}$ *is the drift parameter. Consider moreover the following path* $\left(D, \sigma^2_{\text{rob}}(D)\right)$. *Then the following hold*:

$$\lim_{D\to+\infty} \alpha\left(D + \mu_{\text{inf}}\right) = \eta \quad \Leftrightarrow \quad \lim_{D\to+\infty} \frac{D}{\sigma^2_{\text{rob}}(D)} = 0$$

A formal proof of Theorem 1 can be found in the Supporting Information. A numerical validation of this result is shown in Section 5, while a practical use of the latter in parameter selection can be found in Section 6.

4.2 | The Impact of the Selection of μ_{rob}

The robustification component of the mixture acts to *robustly* model the tails of the informative component's prior distribution. Ideally, it represents a lack of prior knowledge, thereby hindering precise elicitation of its location parameter μ_{rob}. This choice, however, may significantly impact the posterior inference, as demonstrated by Weru et al. [16].

Theorem 2 investigates the condition under which the choice of μ_{rob} becomes impactless in the posterior inference, showing that employing robustification components with large variances effectively prevents from bias stemming from the chosen location, enabling then the use of any convenient value for μ_{rob}.

Theorem 2. *Consider a normal random variable modeling the mean control response* $X_c \sim \mathcal{N}(\theta_c, \sigma_c^2)$, *and assume two distinct RMPs are used for the underlying parameter* θ_c, *namely*

$$\pi_c^{(1)}(\theta_c) = \omega\pi_{\text{inf}}(\theta_c) + (1-\omega)\pi_{\text{rob}}^{(1)}(\theta_c)$$
$$\pi_c^{(2)}(\theta_c) = \omega\pi_{\text{inf}}(\theta_c) + (1-\omega)\pi_{\text{rob}}^{(2)}(\theta_c)$$

where $\pi_{\text{inf}}(\theta_c)$ *and* $\pi_{\text{rob}}^{(i)}(\theta_c)$ *are the PDF of normally distributed random variables with parameters* μ_{inf}, σ^2_{inf} *and* $\mu_{\text{rob}}^{(i)}$, σ^2_{rob} *respectively with* $i \in \{1, 2\}$.

Consider the posterior distributions $g(\theta_c|x_c, \pi_c^{(1)})$ *and* $g(\theta_c|x_c, \pi_c^{(2)})$, *then*

$$\lim_{\sigma^2_{\text{rob}}\to+\infty} g(\theta_c|x_c, \pi_c^{(1)}) = \lim_{\sigma^2_{\text{rob}}\to+\infty} g(\theta_c|x_c, \pi_c^{(2)}) \quad \forall x_c \in \mathbb{R}$$

A formal proof of Theorem 2 can be found in the Supporting Information. A numerical validation of this result is presented in Section 5, while a practical use of the latter in parameter selection is proposed in Section 6.

4.3 | The Lindley's Paradox

The phenomenon termed "Lindley's paradox" within the context of robust mixture priors (RMPs) describes the counterintuitive situation where full borrowing (defined as $\tilde{\omega} = 1$) occurs despite significant prior-data conflict. Literature suggests this

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

arises when the RMP's robustification component is improper [9, 16, 17]. This occurs because the prior predictive distribution for the robustification component, shown in Equation (8), becomes improper ($R \to +\infty$), leading to $\tilde{\omega} = 1$ for all observed control responses x_c according to Equation (9). In Theorem 3 we show that this behavior is due to the hidden underlying assumption that the mixture weight ω is fixed and independent on the choice of σ^2_{rob}. We find that relaxing this assumption, effectively prevents from the occurrence of Lindley's paradox.

Theorem 3. *Consider a normal random variable* $X_c \sim \mathcal{N}(\theta_c, \sigma_c^2)$, *and assume a RMP is used for the parameter* θ_c, *namely* $\pi_c(\theta_c) = \omega\pi_{\text{inf}}(\theta_c) + (1-\omega)\pi_{\text{rob}}(\theta_c)$, *where* $\pi_{\text{inf}}(\theta_c)$ *and* $\pi_{\text{rob}}(\theta_c)$ *are the PDF of normally distributed random variables with parameters*μ_{inf}, σ^2_{inf} *and* μ_{rob}, σ^2_{rob} *respectively. The following hold*:

1. *if* $\Omega < +\infty$, *then*

$$\lim_{\sigma^2_{\text{rob}} \to +\infty} \tilde{\omega}\big(x_c, \pi_{\text{inf}}(\theta_c), \pi_{\text{rob}}(\theta_c), \omega\big) = 1 \qquad \forall x_c \in (-\infty, +\infty)$$

2. *if* $\Omega \sim O(1/R)$ *for* $\sigma^2_{\text{rob}} \to +\infty$, *then*

$$\lim_{\sigma^2_{\text{rob}} \to +\infty} \tilde{\omega}\big(x_c, \pi_{\text{inf}}(\theta_c), \pi_{\text{rob}}(\theta_c), \omega\big) \neq 1 \qquad \forall x_c \in (-\infty, +\infty)$$

The preceding theorem demonstrates that Lindley's paradox arises, as $\sigma^2_{\text{rob}} \to +\infty$, when the prior weight ω (or prior odds Ω) is fixed independently of σ^2_{rob}. Conversely, if ω and σ^2_{rob} are jointly selected such that the prior odds Ω are of the same order of magnitude as R—as $\sigma^2_{\text{rob}} \to +\infty$—then Lindley's paradox is avoided. The latter holds because as $\sigma^2_{\text{rob}} \to \infty$, the posterior odds $\tilde{\Omega}$ can be written following Equation (9) as

$$\tilde{\Omega}(x_c; \omega, \sigma^2_{\text{rob}}) = \beta(\omega, \sigma^2_{\text{rob}}) \times \exp\left[-\frac{d^2}{2v^2_{\text{inf}}}\right], \tag{10}$$

where the influence of the RMP on the posterior odds is entirely captured by the function $\beta(\omega, \sigma^2_{\text{rob}})$. As a consequence, all combinations of ω and σ^2_{rob} yielding $\beta(\omega, \sigma^2_{\text{rob}}) = \beta^*$ share the same "borrowing profile", resulting in identical posterior odds (and thus, posterior weights) for any observed value x_c.

The parameter β^* controls the flexibility of the RMP in borrowing information across the x_c space. In particular, it can be interpreted as the posterior odds under a scenario with no observed drift, that is, when the data perfectly align with the mean of the informative component. In this case, β^* quantifies the maximum degree of borrowing attainable by the RMP. For this reason, β^* may be referred to as the *borrowing strength*, as it naturally captures the extent to which the RMP tends to borrow information from the informative component.

It is important to note that while these pairs yield identical posterior weights, posterior inference for θ_c could differ in principle across RMPs due to variations in $g_{\text{rob}}(\theta_c | x_c, \pi_{\text{rob}})$, resulting from differing choices of μ_{rob} and σ^2_{rob}. However, as $\sigma^2_{\text{rob}} \to \infty$, the robust posterior becomes independent of μ_{rob}, leading to similar inference for θ_c across all pairs across the entire control response parameter space.

Note that the asymptotic approximation of posterior odds in Equation (10) is valid only when $R \gg 1$ ($v_{\text{rob}} \gg v_{\text{inf}}$), a reasonable assumption given the robustification component of the RMP is specifically designed for robustification.

5 | Practical Considerations

Using the same trial design considered in Section 3.2, in the following sections we will focus on the validation of the use of the RMPs with large variance robustification components in the context of unbalanced RCT with hybrid control arms.

5.1 | Overcoming Lindley's Paradox

In Section 4.3 it has been proven that different pairs $(\omega, \sigma^2_{\text{rob}})$ may induce the same posterior weights distribution on the control response space. The latter is illustrated in Figure 2.

Figure 2 presents a three-dimensional representation with parameters ω and $n_0 = \sigma^{-2}_{\text{rob}}$ on the horizontal axes and the observed control response x_c on the vertical axis. The red curve embedded in the (ω, n_0) plane delineates the set of parameter pairs (ω, n_0) satisfying $\beta(\omega, n_0) = 5.83$, each representing a distinct RMP. Notice that this value has been specifically selected so to include the pair $\omega = 0.5$, $n_0 = 1$, so that

$$\beta^* = \beta(0.5, 1) = \frac{0.5}{1-0.5} \times \sqrt{\frac{1+1/50}{1/100+1/50}}. \tag{11}$$

The figure was generated by varying the effective sample size of the robust component over the interval $(0.01, 1)$ with a step of 0.01. For each value, the prior weight ω was determined to satisfy Equation (11), and the posterior odds were computed for each

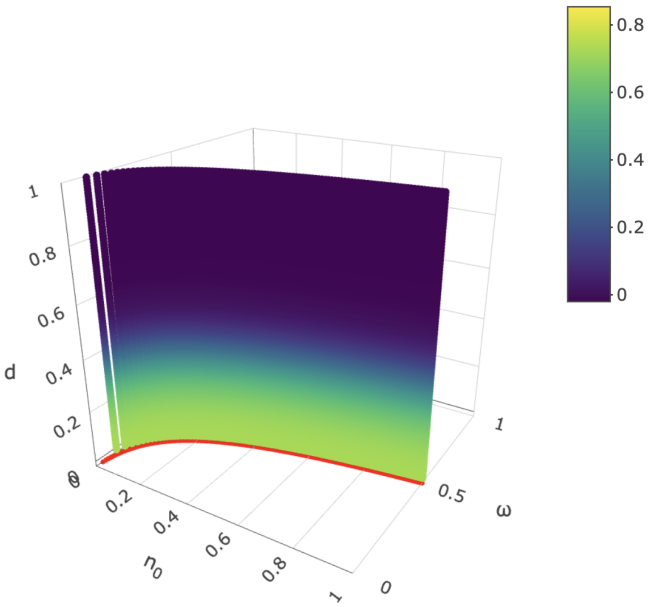


FIGURE 2 | Posterior weight $\tilde{\omega}$ as a function of effecive sample size of the robust component n_0, prior weight ω and observed control response x_c. The red curve in the (n_0, ω) represents all RMPs with $\beta^* = 5.83$.

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

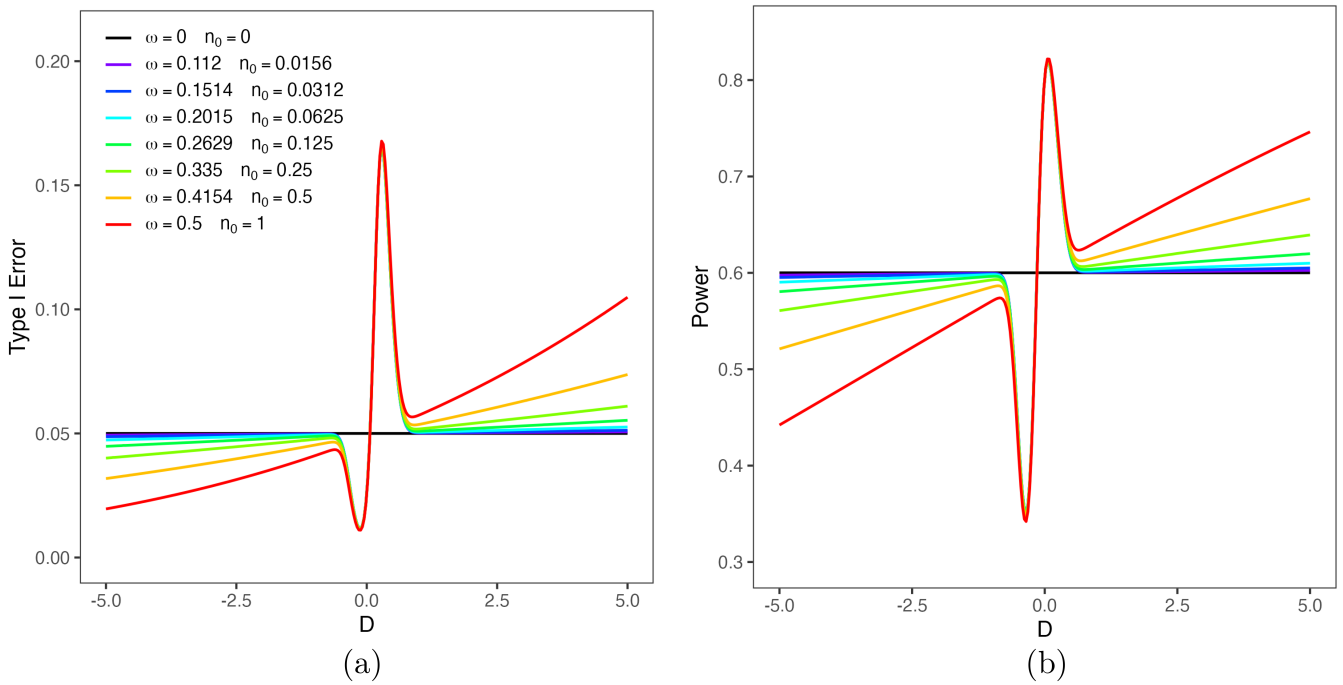


FIGURE 3 | Panel (a): type I error rate. Panel (b): power under $\delta^* = 0.31$. Colors represent different pairs of (ω, n_0), corresponding to $\beta = 5.83$.

pair (ω, n_0) using Equation (7). The posterior weights were then obtained using the formula $\tilde{\omega} = \tilde{\Omega}/(1+\tilde{\Omega})$. The vertical colored lines in the figure depict the posterior weights $\tilde{\omega}$ as a function of x_c for all RMPs considered along the red curve, the yellow color indicating a posterior weight of 1 (full borrowing) and the blue color indicating a posterior weight of 0 (no borrowing).

The vertical lines originating from each point on the red curve exhibit a continuous color gradient along the x_c axis, indicating that the posterior weights $\tilde{\omega}$, as a function of the control response x_c, depend solely on the chosen value of β^*. Consequently, all pairs (ω, n_0) yielding the same β^* correspond to identical posterior weight profiles.

These observations suggest that Lindley's paradox is effectively mitigated by a joint selection of ω and σ^2_{rob}. Specifically, the posterior weight profile characteristic of any RMP with a weakly informative robustification component (e.g, UIP) can be replicated using robustification components with arbitrarily large variance. Further visualizations of posterior weights under varying β^* values are provided in the Supporting Information.

5.2 | Overcoming Asymptotic Type I Error Rate Inflation

While the preceding analysis demonstrates that a set of RMPs share a common posterior weight profile $\tilde{\omega}$, this does not guarantee identical posterior inferences on the control parameter θ_c. Posterior inference is influenced not only by posterior weights but also by the posterior distributions of the individual RMP components, which are functions of their hyper-parameters.

In this section, an analysis of the frequentist operating characteristics is conducted, with specific attention to the problem of asymptotic type I error rate inflation. In addition, the link between the latter and the posterior inference metrics (bias, variance and MSE) is discussed.

This application considers eight distinct RMPs, generated by varying the effective sample size of the control parameter, n_0, across the set $\{(\frac{1}{2})^k | k = 0, \ldots, 7\}$, and the prior mixture weight, ω, across the set $\{0, 0.5, 0.415, 0.335, 0.263, 0.201, 0.151, 0.112\}$. All considered pairs, excluding the first (representing an improper prior), belong to the level set $\beta(n_0, \omega) = 5.83$, thus exhibiting the shared posterior weight profile discussed in Section 5.1. For each RMP, type I error rate (Figure 3a) and power (Figure 3b), are assessed, with power calculated for a treatment difference of $\delta^* = 0.31$. Posterior inference is evaluated using bias (Figure 4a), variance (Figure 4b), and mean squared error (MSE) (Figure 4b).

For small to moderate prior-data conflicts, the power (Figure 3b) and type I error rate (Figure 3a) curves overlap for all RMPs. This occurs because both variances and bias are comparable in these regions. Consequently, the posterior distributions of the treatment difference δ are similar across pairs, centered near $\delta = 0$ (for type I error rate) and $\delta = \delta^*$ (for power). This results in highly similar null hypothesis rejection rates for all RMPs.

Conversely, significant differences among the pairs emerge under large prior-data conflicts, where RMPs with weakly informative robustification components exhibit inflation (deflation) of both type I error rate and power for large positive (negative) drifts. However, this effect is attenuated for RMPs with less informative robustification components, practically disappearing when $n_0 < (\frac{1}{2})^6$. In these regions, substantial differences in bias among the RMPs impact type I error rate and power, which deviate considerably from their nominal levels for RMPs with more informative robustification components, while remaining near their

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

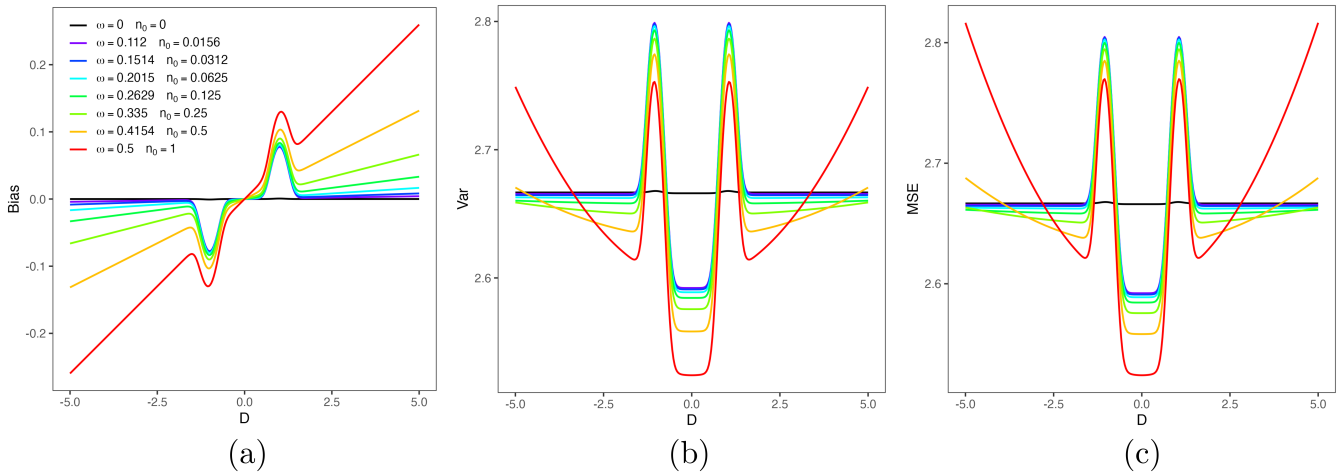


FIGURE 4 | Panel (a): bias; Panel (b): variance; Panel (c): mean squared error, all computed using the posterior mean of the treatment effect parameter δ. Colors denote different pairs of (ω, n_0), each corresponding to $\beta^* = 5.83$.

TABLE 1 | Maximum type I error rate (α_{max}), average type I error rate (α_{avg}), power gain under no data-conflict Pow(0) and width of the sweet spot for different pairs of (ω, n_0), all corresponding to $\beta^* = 5.83$.

ω	n_0	α_{max}	$\alpha(50)$	$\alpha_{avg}^{\text{VAG}}$	α_{avg}^{INF}	$\alpha_{avg}^{\text{RMP}}$	Pow(0)	Sweet spot width
0	10^{-100}	0.0500	0.0500	0.0500	0.0500	0.0500	0.600	0.000
0.500	1.000	0.168	0.9914	0.2955	0.0394	0.0492	0.803	0.207
0.415	0.500	0.167	0.6478	0.1522	0.0397	0.0496	0.803	0.206
0.335	0.250	0.166	0.2643	0.0785	0.0399	0.0498	0.802	0.207
0.263	0.125	0.166	0.1278	0.0574	0.0399	0.0499	0.802	0.207
0.201	0.062	0.166	0.0822	0.0520	0.0400	0.0499	0.802	0.207
0.151	0.031	0.165	0.0645	0.0507	0.0400	0.0500	0.802	0.207
0.112	0.016	0.165	0.0569	0.0503	0.0400	0.0500	0.802	0.207

nominal values for RMPs with less informative robustification components.

Table 1 summarizes key characteristics of the observed curves. These include the maximum type I error rate inflation, α_{max}, constrained to the interval $-5 < D < 5$ (a plausible response range); the power gain, Pow(0), when the informative component of the RMP perfectly matches the control data; the type I error rate under extreme drift, $\alpha(50)$; the average type I error rate across different design priors (an improper prior, the informative component of the RMP, and the RMP itself); and the width of the "sweet spot" region [11]. The "sweet spot" is defined as the interval of D values where type I error rate and Power are respectively below and above their nominal levels (5% and 60% in this application).

All considered (ω, n_0) pairs demonstrate comparable performance in terms of maximum type I error rate, α_{max}, power gain, Pow(0), and sweet spot width. However, a significant difference emerges when examining $\alpha(50)$. This value is notably higher for RMPs with weakly informative robustification components (approaching 100% for the UIP), progressively decreasing towards 5% as the informativeness of the robustification component increases.

Averaging type I error rate across an improper prior distribution reveals a marked inflation for RMPs with weakly informative robustification components, as consequence of the asymptotic type I error rate increase discussed previously. The type I error rate decrease observed for negative drifts does not fully compensate for the inflation because the range of increase (from 5% to 100%) is considerably larger than the range of decrease (from 5% to 0%), leading to a greater weighting of the inflation in the averaging process.

Conversely, minimal differences are observed among pairs when averaging type I error rate across more informative priors, such as the informative component of the RMP or the RMP itself. These priors are concentrated around regions of small drifts, where all RMPs have practically identical type I error rate curves. The type I error rate reduction exhibited by all RMPs in this region keeps the average type I error rate controlled at the nominal level (in the strong sense, when using the informative component or the RMP as the design prior).

In summary, RMPs with high-variance robustification components achieve comparable performance to those with weakly informative robustification components, while simultaneously mitigating type I error rate inflation. This results in average type I error rate remaining below the nominal level when the RMP or

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

its informative component are used as design priors (as demonstrated in Best et al. [18]), but also controlled just slightly above the nominal level when improper priors are used; thus guaranteeing an higher overall protection to incorrect rejections of the null hypothesis.

5.3 | Overcoming Biases Due to the Specification of μ_{rob}

Figure 5 investigate the influence of robustification component location on the type I error rate within the Robust Mixture Prior (RMP). For each of the first six (ω, n_0) pairs analyzed in Figure 3 and Table 1, five type I error rate and power curves (as functions of the drift parameter D) are presented, corresponding to variations in the robustification component location parameter, μ_{rob}, across the set $\{-2, -1, 0, 1, 2\}$.

The figures demonstrate that for large n_0 values (e.g., UIP), operating characteristics exhibit high sensitivity to the location parameter μ_{rob}. Consistently with what shown in Section 3, increasing μ_{rob} uniformly inflates both type I error rate curve, while decreasing μ_{rob} has the opposite effect. Conversely, as n_0 decreases (and accordingly σ^2_{rob} increases), the impact of μ_{rob} on posterior inference diminishes, as evidenced by the substantial overlap of the type I error rate curves when $n_0 = 0.031$. The same behavior can be appreciated in the Power analysis in Figure S3 of the Supporting Information.

6 | Hyper-Parameters Elicitation

6.1 | On the Interpretation of the Prior Weight

The use of normal RMPs in practice necessitates the pre-specification of hyper-parameters: the robustification component location μ_{rob}, the robustification component variance σ^2_{rob}, and the mixture weight ω. Current practice often prioritizes default values for the two former parameters, centering the robustification component at the informative component mean ($\mu_{rob} = \mu_{inf}$) and selecting a unit-information robust variance [12]. The mixture weight ω is then normally determined based on stakeholder or experts confidence in the data supporting the informative component.

This elicitation is typically driven by questions like "how much is the probability that historical data are relevant in the current setting?" or "how much confidence do you have in historical data being representative of the current data?". For instance, high confidence (or high probability) might lead to $\omega = 0.9$, whereas low confidence might lead to $\omega = 0.3$.

While straightforward to communicate, this interpretation may disregard the crucial interplay between ω and σ^2_{rob}, significantly influencing RMP performance as it only concerns one parameter of the RMP, while it is argued above that they should be chosen in accordance with the variance of the robustification component. Furthermore, implies that the current choice of ω is unrelated to the choice of the robustification component. In fact, following the results above, we argue that the interpretation (and as a result the elicitation) of the weight should come together with the choice of the robustification component.

We have proven in Section 4.3 that the *borrowing strength* β^* is the key parameter influencing the borrowing profile of the RMP. This suggests that an equivalent prior degree of confidence in historical data should correspond to a lower ω for RMPs with a larger robustification component variance and a higher ω for RMPs with a smaller robustification component variance. As a consequence, we posit that ω should be viewed as a *relative* confidence measure between the informative model π_{inf} and the robust model π_{rob}, which specification should then depend on how informative the robustification component itself is.

Given the suggested interpretation of ω, we propose the following procedure for its elicitation.

6.2 | An Approach for Hyper-Parameters Elicitation

A four-step elicitation approach is proposed:

1. Standard deviation of the robustification component of the RMP σ_{rob} is set to a large value. A possible option is setting it to $\sigma_{rob} = 1000 \times s$, where s represents the standard deviation of the considered endpoint (note that even higher values can be used, but as demonstrated above they will have no impact on the inference).
2. The location of the robustification component μ_{rob} is set equal to the location of the informative component μ_{inf}.
3. Clinicians are asked to determine an "equipoise drift" value d^*, representing the potentially observed control response that would induce maximum uncertainty regarding the relevance of historical data. Prompting questions could be: "At what control response value would you be 50% confident that the historical component is relevant for the current trial and 50% that it is not?" or "At what control response value would you suspect a systematic difference between historical and concurrent control data?".
4. Once specified σ_{rob} and d^*, the prior odds Ω is obtained such that $\tilde{\Omega}(d^* + \mu_{inf}) = 1$ (or equivalently $\tilde{\omega} = 0.5$), inverting equation (9) as follows:

$$\Omega = \left[R \times \exp\left\{ -\frac{d^{*2}}{2v^2_{inf}} + \frac{(x_c - \mu_{rob})^2}{2R^2 v^2_{inf}} \right\} \right]^{-1} \tag{12}$$

and accordingly the prior weight is retrieved as $\omega = \frac{\Omega}{1+\Omega}$.

Our hyper-parameter selection routine combines the benefits of RMPs with large variance robustification components and expert interaction. Moreover, while elicitation of the mixture weight ω poses challenges due to its complex interpretability, elicitation on the drift scale offers straightforward interpretation, thus justifying the approach.

7 | Beta-Binomial Case

7.1 | Beta Robust Mixture Prior

Let us now consider the setting in which a RCT is performed with a binary outcome so that the total number of responses is

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

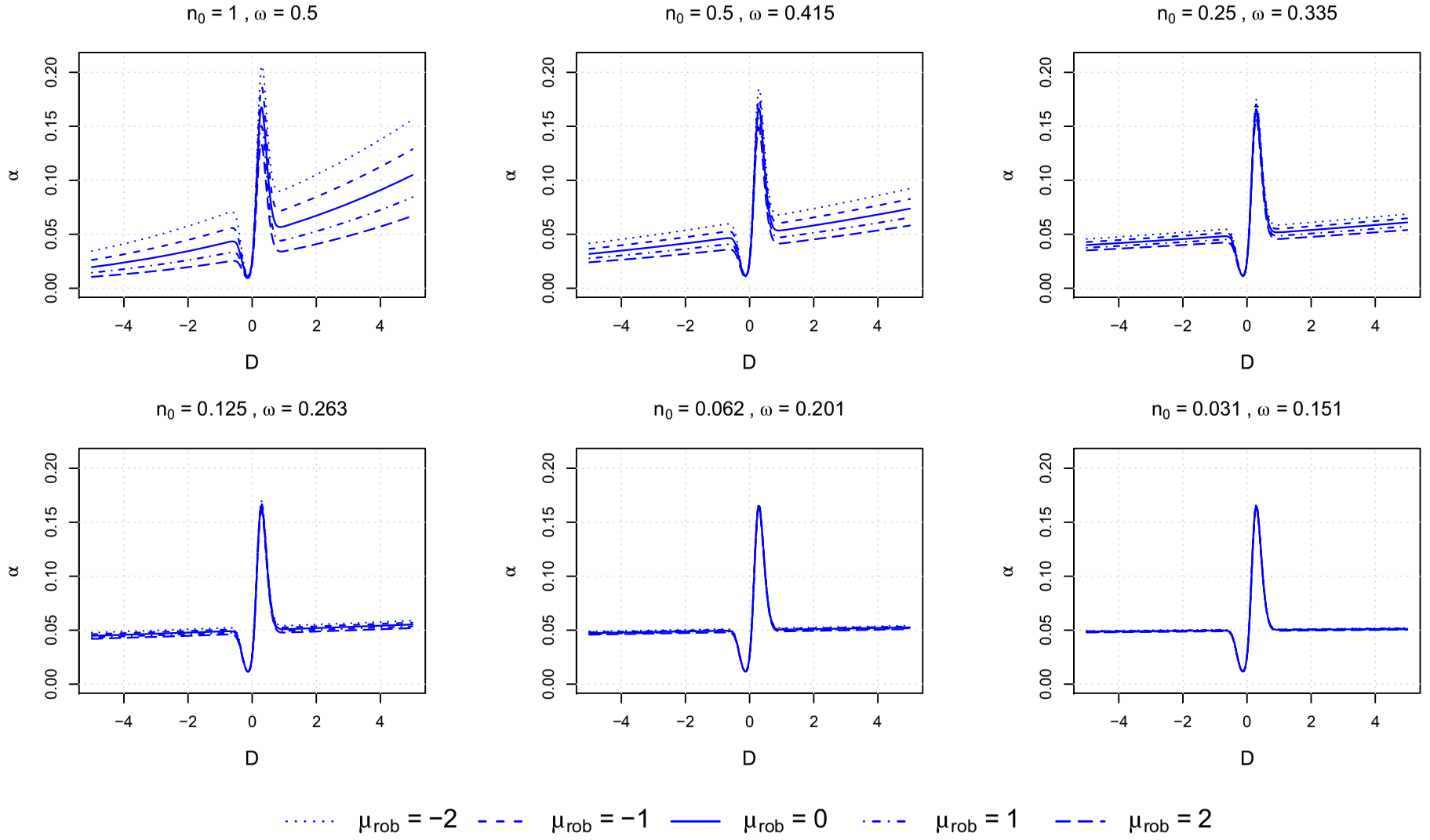


FIGURE 5 | For each panel representing a different pairs of (ω, n_0), type I error rate as a function of the prior-data conflict D is displayed for five different values of the location of the robustification component μ_{rob}.

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$X_c \sim \text{Bin}(\theta_c, n_c)$, where n_c is the number of patients allocated to the control arm and $\theta_c \in (0,1)$ represents the response parameter on the probability scale.

The Robust Mixture Prior in this case can be chosen as a mixture of two Beta distribution, namely $\text{Beta}(a_{\text{inf}}, b_{\text{inf}})$ for the informative component and $\text{Beta}(a_{\text{rob}}, b_{\text{rob}})$ for the robustification component. Then the prior predictive density of the data is a Beta-Binomial, namely

$$f(x_c|\pi_\star) = \binom{n_c}{x_c} \frac{B(a_\star + x_c, b_\star + n_c - x_c)}{B(a_\star, b_\star)} \quad \star = \{\text{inf}, \text{rob}\} \tag{13}$$

where $x_c \in (0, n_c)$ is the observed number of responders in the control arm and $B(\cdot)$ represents the Beta function. Working out with the Gamma function expression of the Beta function, it follows that the odds update of Equation (7) can be expressed in this case as

$$\tilde{\Omega}(x_c) = \beta(\omega, a_{\text{rob}}, b_{\text{rob}}) \times \frac{B(a_{\text{inf}} + x_c, b_{\text{inf}} + n_c - x_c)}{B(a_{\text{rob}} + x_c, b_{\text{rob}} + n_c - x_c)\, B(a_{\text{inf}}, b_{\text{inf}})}, \tag{14}$$

where the function $\beta(\omega, a_{\text{rob}}, b_{\text{rob}})$ can be expressed as

$$\beta(\omega, a_{\text{rob}}, b_{\text{rob}}) = \Omega \cdot B(a_{\text{rob}}, b_{\text{rob}}) \tag{15}$$

Note that although a_{rob} and b_{rob} may differ, setting them equal and small is a reasonable choice when aiming to represent limited prior knowledge. In common practice, specifications such as Beta(1, 1) or Beta(0.5, 0.5) (Jeffreys prior) are typically employed for this purpose.

7.2 | The Lindley's Paradox in the Beta-Binomial Case

Similarly to the normal case, also in the Beta-Binomial case the phenomenon of the Lindley's paradox occurs when a large variance distributions is used as a robust component of the RMP. Specifically, this happens—for a fixed ω—when the parameter of the Beta distribution related to the robust component approaches 0, because $\Gamma(0^+) \to +\infty$ and accordingly following Equation (15) the posterior odds goes to $+\infty$ and accordingly the posterior weights ω goes to 1. Similarly to what done in the normal case in Theorem 3, in Theorem 4 we show that this behavior is due to the hidden underlying assumption that the mixture weight ω is fixed and independent on the choice of a_{rob} and b_{rob}. We find that relaxing this assumption, effectively prevents from the occurring of Lindley's paradox.

Theorem 4. *Consider a binomial random variable* $X_c \sim \text{Bin}(\theta_c, n_c)$, *and assume a RMP is used for the parameter* θ_c, *namely* $\pi_c(\theta_c) = \omega\pi_{\text{inf}}(\theta_c) + (1-\omega)\pi_{\text{rob}}(\theta_c)$, *where* $\pi_{\text{inf}}(\theta_c)$ *and* $\pi_{\text{rob}}(\theta_c)$ *are the PDF of Beta distributed random variables with parameters* $a_{\text{inf}}, b_{\text{inf}}$ *and* $a_{\text{rob}} = b_{\text{rob}} = \varepsilon$, *respectively. The following hold*:

1. *if* $\Omega < +\infty$, *then*

$$\lim_{\varepsilon \to 0} \tilde{\omega}(x_c, \pi_{\text{inf}}(\theta_c), \pi_{\text{rob}}(\theta_c), \omega) = 1 \quad \forall x_c \in (0, n_c)$$

2. *if* $\Omega \sim O(\varepsilon)$ *for* $\varepsilon \to 0$, *then*

$$\lim_{\varepsilon \to 0} \tilde{\omega}(x_c, \pi_{\text{inf}}(\theta_c), \pi_{\text{rob}}(\theta_c), \omega) \neq 1 \quad \forall x_c \in (0, n_c)$$

A formal proof of Theorem 4 can be found in the Supporting Information.

The preceding theorem demonstrates that Lindley's paradox arises, as the parameters of the robust component of the RMP approaches zero, when the prior weight ω (or prior odds Ω) is fixed independently of the parameters of the robust component. Conversely, if ω and $a_{\text{rob}} = b_{\text{rob}} = \varepsilon$ are jointly selected such that the prior odds Ω remain of the same order of magnitude as the parameters of the robust component, namely $\Omega \sim O(\varepsilon)$, then Lindley's paradox is avoided.

This occurs because, as $\varepsilon \to 0$, the posterior odds $\tilde{\Omega}$ can be expressed following Equations (14) and (15) as

$$\tilde{\Omega}(x_c; \omega, \varepsilon) = \beta(\omega, \varepsilon) \times \frac{B(a_{\text{inf}} + x_c, b_{\text{inf}} + n_c - x_c)}{B(x_c, n_c - x_c)\, B(a_{\text{inf}}, b_{\text{inf}})}, \tag{16}$$

where the influence of the RMP on the posterior odds is entirely captured by the function $\beta(\omega, \varepsilon)$ defined in Equation (15). It follows that, similarly to what shown in the normal case, all combinations of ω and ε yielding the same $\beta(\omega, \varepsilon) = \beta^*$ share the same "borrowing profile", resulting in identical posterior odds and posterior weights $\tilde{\omega}$ for any observed number of responders x_c.

The parameter β^* governs the RMP's flexibility in borrowing information across the x_c space, determining the rate at which posterior weights decrease in the presence of prior-data conflict.

It is important to note that, while these pairs (ω, ε) yield identical posterior weights, posterior inference for θ_c could in principle differ across RMPs due to variations in the robust posterior component $g_{\text{rob}}(\theta_c|x_c, \pi_{\text{rob}})$ arising from different choices of ε. However, as $\varepsilon \to 0$, the posterior distribution related to the robust component of the RMP tends to lose its dependence on the prior parameters, thus leading to similar inference for θ_c across all such pairs.

7.3 | Practical Considerations

In the Supporting Information, the results presented in Section 7 are validated through a numerical investigation. Specifically, we considered a randomized controlled trial (RCT) in which $n_c = 100$ patients are assigned to the control arm, while $n_t = 200$ patients are allocated to the treatment arm. The number of responses in each arm follows a binomial distribution $X_* \sim \text{Bin}(\theta_*, n_*)$, $* = \{c, t\}$.

A Jeffreys prior, Beta(0.5, 0.5), is used for the treatment parameter θ_t, whereas various robust mixture priors (RMPs) are explored as prior distributions for the control parameter θ_c. The informative

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

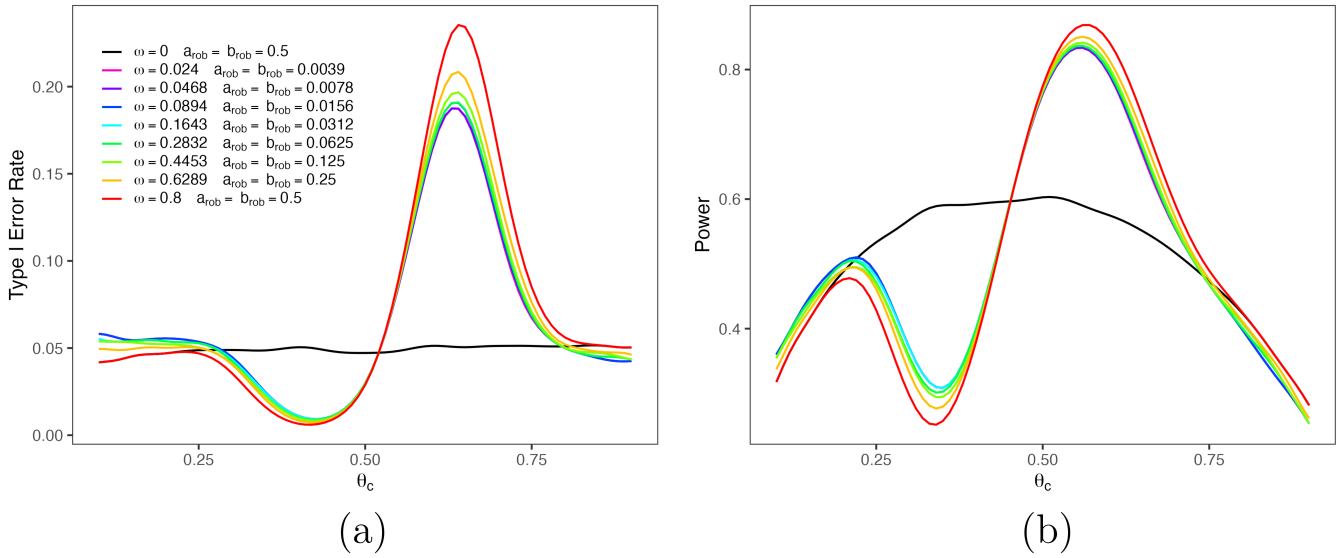


FIGURE 6 | Panel (a): Type I error rate; Panel (b): power under a target log-odds ratio $\delta^* = 0.47$, both evaluated in the Beta–Binomial setting. Colors indicate different pairs of $(\omega, a_{\text{rob}} = b_{\text{rob}})$ corresponding to $\beta^* = 12.56$.

component of the RMP is fixed to Beta(50, 50), reflecting a prior knowledge on the control parameter being close to $\theta_c = 0.5$. The success rule is the same expressed in Equation (1), where δ represents the log odds ratio corresponding to the two parameters, namely $\delta = \log\left(\frac{\theta_t(1-\theta_c)}{\theta_c(1-\theta_t)}\right)$.

Analogously to the normal case, Figure S4 illustrates how the posterior weights vary as a function of the observed number of responses in the control arm, when the prior weight ω and the parameters of the robust component of the RMP, $a_{\text{rob}} = b_{\text{rob}}$, are jointly chosen to satisfy the condition $\beta^* = 12.56$. Notice that this value has been arbitrarily selected so to include the pair $\omega = 0.8$, $a_{\text{rob}} = b_{\text{rob}} = 0.5$, so that $\beta^* = \beta(0.8, 0.5) = \frac{0.8}{1-0.8} \cdot B(0.5, 0.5)$.

The figure shows that, for all parameter pairs satisfying $\beta^* = 12.56$, the variation of the posterior weights $\tilde{\omega}$ with respect to the number of control responses x_c is closely aligned. This indicates that all such RMPs exhibit the same borrowing profile, and particularly that borrowing is possible even when a_{rob} and b_{rob} are very small, thus confirming that the Lindley's paradox can be effectively avoided provided a joint selection of the pair $(\omega, a_{\text{rob}} = b_{\text{rob}})$.

This behavior is further confirmed by examining the type I error rate and power plots in Figure 6, as well as the bias, variance, and mean squared error plots in Figure S5.

In these figures, eight pairs $(\omega, a_{\text{rob}} = b_{\text{rob}})$ satisfying $\beta^* = 12.56$ are shown, and the operating characteristics corresponding to different RMPs are displayed across the true control parameter $\theta_c \in (0.1, 0.9)$. In particular, the curves corresponding to different pairs $(\omega, a_{\text{rob}} = b_{\text{rob}})$ follow very similar trends across the θ_c range. A near-complete overlap is observed for pairs with $a_{\text{rob}} = b_{\text{rob}} < 0.1$ across the parameter space, while some deviations occur in regions of moderate prior-data conflict, that is, when more informative Beta priors are employed as the robust component of the RMP. For instance, using a Beta(0.5, 0.5) prior produces similar OCs in regions of minor drift, but the maximum type I error increases noticeably (approximately 6% higher) relative to RMPs with weaker robust components, due to higher bias in regions of intermediate conflict.

Consistent with the normal case, we conclude that employing quasi non-informative Beta distributions as the robust component in the Beta RMP is feasible without inducing Lindleys paradox, provided that the prior weight ω and the parameters of the robust component are jointly selected. Moreover, using weakly informative robust components mitigates bias in regions of the parameter space where type I error inflation is most pronounced, thus offering greater protection against potential inflation arising from moderate drift between concurrent and historical data.

Finally, it is noteworthy that, in the Beta-Binomial setting, asymptotic type I error inflation is not a concern, as the extent of prior-data conflict is inherently bounded by the domain of the parameter θ_c.

8 | Extension to a Mixture Informative Component

The framework introduced in this paper can be further extended to the case in which the informative component of the Robust Mixture Prior (RMP) is itself modeled as a mixture of distributions, such as Beta or Normal, depending on the context.

Let the informative component of the RMP be expressed as

$$\pi_{\text{inf}}(\theta_c) = \sum_{k=1}^{K} \xi_k \, \pi_{\text{inf}}^{(k)}, \tag{17}$$

where $\sum_{k=1}^{K} \xi_k = 1$. Denote by ω the weight assigned to the informative component and by $1 - \omega$ the weight assigned to the robust component. The overall RMP can then be represented as a mixture of $K + 1$ components:

$$\pi_c(\theta_c) = \sum_{k=1}^{K} \omega \, \xi_k \, \pi_{\text{inf}}^{(k)} + (1 - \omega) \, \pi_{\text{rob}}. \tag{18}$$

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Define $\eta_k = \omega\ \xi_k$ for $k = 1, \ldots, K$ and $\eta_{K+1} = 1 - \omega$. Let $\Omega_k = \eta_k/(1-\eta_k)$ denote the odds associated with the k-th component of the RMP. An extension of Equation (7) to this setting, expressed in terms of the reciprocal of the odds rather than the odds themselves (for convenience), can be written as

$$\tilde{\Omega}_h^{-1}(x_c) = \sum_{\substack{k=1 \\ k\neq h}}^{K} \frac{\xi_k\ f\left(x_c|\pi_{\text{inf}}^{(k)}\right)}{\xi_h\ f\left(x_c|\pi_{\text{inf}}^{(h)}\right)} + \frac{1}{\xi_h}\ \Omega_{K+1}^{-1}\ \frac{f\left(x_c|\pi_{\text{rob}}\right)}{f\left(x_c|\pi_{\text{inf}}^{(h)}\right)} \quad h = 1, \ldots, K \tag{19}$$

and the posterior weight related to the robust component can be retrieved as $\tilde{\eta}_{K+1} = 1 - \sum_{k=1}^{K}\tilde{\eta}_k$. Note that Equation (19) reduces to Equation (7) when $K = 1$.

It is worth noting that the first summation term in the above expression does not depend on the prior weights assigned to the informative and non-informative components, but only on the fixed weights ξ_k associated with each element of the informative part of the RMP. Moreover, it is independent of the specification of the robust component of the RMP. The reciprocal of the second term, in contrast, coincides with Equation (7), rescaled by a component-specific factor ξ_h. Consequently, the asymptotic decomposition derived in the previous sections (for both the continuous and binary cases) remains valid, and the proposed methodology can be seamlessly extended to the mixture-based framework.

9 | Discussion

Robust Mixture Priors (RMPs) are a prominent dynamic borrowing approach used to incorporate historical control data in the analysis of a current randomized trial. However, specifying parameters for the RMP components, particularly the robustification component and mixture weights, presents a challenge, as these parameters strongly influence posterior inferences. While improper normal distributions may seem intuitive for the robustification component, their use has been discouraged due to the potential for Lindley's paradox, prompting a preference for weakly informative priors. Employing the unit-information prior (UIP) [12] has become common; nevertheless, this choice remains somewhat arbitrary and context-dependent [17]. Specifically, concerns have been raised regarding the UIP's potential over-informativeness in trials with limited sample sizes [16], as well as the theoretical unbounded type I error rate in unbalanced trials using UIP [18].

In this article, we demonstrate, for both normal and binary endpoints, that jointly eliciting the mixture weight and the hyperparameters of the robustification component within a Robust Mixture Prior (RMP) framework effectively mitigates Lindley's paradox, even when using arbitrarily large variances.

This approach offers several practical advantages. In the normal case, it practically eliminates the impact of the location of the robustification component and prevents asymptotic type I error rate inflation in unbalanced trials, which is a critical regulatory consideration. While asymptotic inflation does not occur in balanced trials, these scenarios are of limited practical interest, as the main goal of borrowing is to reduce sample size on the control arm.

For binary endpoints, asymptotic type I error inflation does not occur due to the natural bounds of the probability parameter (0 to 1). Nevertheless, employing a large-variance robustification component (i.e., a Beta distribution with parameters approaching 0) has been shown to reduce the maximum type I error inflation compared to the commonly used Jeffreys prior.

We illustrate these properties through a proof-of-concept case study. Additionally, we propose a novel routine for selecting hyperparameters that combines a large-variance robustification component with an expert opinion-driven prior weight, ω.

We further extend the methodology to the setting where the informative component of the RMP itself is a mixture of normal distributions, enhancing the flexibility of the approach.

Importantly, the insights derived from this work are general and extend to any framework employing a Robust Mixture Prior (RMP). The demonstrated interplay between the prior weight ω and the robustification component π_{rob} is not limited to the specific implementation proposed here but is also relevant to other approaches that rely on RMPs, including those based on empirical Bayes formulations such as the EB-rMAP [20] and the SAM prior [21]. Consequently, our findings provide a unifying perspective that can inform the specification and calibration of RMP-based borrowing mechanisms across diverse methodological frameworks.

Although the mathematical results could, in principle, be extended to one-arm trials where borrowing is performed on the treatment effect scale, exploring this application is beyond the scope of the current study. We leave the investigation of one-arm trial extensions and the evaluation of whether similar advantages hold in practice as future work.

Acknowledgments

This work was supported by Institut de Recherches Internationales Servier. The results reported herein are part of a collaboration between Servier, Saryga, and P. Mozgunov whose research is supported by the National Institute for Health and Care Research (NIHR Advanced Fellowship, Dr Pavel Mozgunov, NIHR300576). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health and Care Research or the Department of Health and Social Care (DHCS). P. Mozgunov received funding from UK Medical Research Council (MC UU 00040/03). M. Gasparini received funding from PNRR (MUR-M4C2-Investimento 1.5) funded by the European Union — NextGenerationEU (Grant agreement no. ECS00000036). Open access publishing facilitated by Politecnico di Torino, as part of the Wiley - CRUI-CARE agreement.

Funding

This work was supported by the National Institute for Health and Care Research (Grant No. NIHR300576), the Medical Research Council (Grant No. MC UU 00040/03), and the PNRR (MUR-M4C2-Investimento 1.5) (Grant No. ECS00000036).

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

References

1. M. Dunoyer, "Accelerating Access to Treatments for Rare Diseases," *Nature Reviews Drug Discovery* 10, no. 7 (2011): 475–476.

2. J. Dunne, W. J. Rodriguez, M. Dianne Murphy, et al., "Extrapolation of Adult Data and Other Data in Pediatric Drug-Development Programs," *Pediatrics* 128, no. 5 (2011): e1242–e1249.

3. D. A. Schoenfeld, H. Zheng, and D. M. Finkelstein, "Bayesian Design Using Adult Data to Augment Pediatric Trials," *Clinical Trials* 6, no. 4 (2009): 297–304.

4. C. Röver, S. Wandel, and T. Friede, "Model Averaging for Robust Extrapolation in Evidence Synthesis," *Statistics in Medicine* 38, no. 2 (2019): 674–694.

5. B. P. Hobbs, B. P. Carlin, S. J. Mandrekar, and D. J. Sargent, "Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials," *Biometrics* 67, no. 9 (2011): 1047–1056.

6. J. G. Ibrahim, M.-H. Chen, Y. Gwon, and F. Chen, "The Power Prior: Theory and Applications," *Statistics in Medicine* 34, no. 12 (2015): 3724–3749.

7. S. J. Pocock, "The Combination of Randomized and Historical Controls in Clinical Trials," *Journal of Chronic Diseases* 29, no. 3 (1976): 175–188.

8. A. Kleyner, S. Bhagath, M. Gasparini, J. Robinson, and M. Bender, "Bayesian Techniques to Reduce the Sample Size in Automotive Electronics Attribute Testing," *Microelectronics and Reliability* 37, no. 6 (1997): 879–883.

9. T. Mutsvari, D. Tytgat, and R. Walley, "Addressing Potential Prior? Data Conflict When Using Informative Priors in Proof of Concept Studies," *Pharmaceutical Statistics* 15, no. 1 (2016): 28–36.

10. J. van Rosmalen, D. Dejardin, Y. van Norden, B. Löwenberg, and E. Lesaffre, "Including Historical Data in the Analysis of Clinical Trials: Is It Worth the Effort?," *Statistical Methods in Medical Research* 27, no. 10 (2018): 3167–3182.

11. K. Viele, S. Berry, B. Neuenschwander, et al., "Use of Historical Control Data for Assessing Treatment Effects in Clinical Trials," *Pharmaceutical Statistics* 13, no. 1 (2014): 41–54.

12. H. Schmidli, S. Gsteiger, S. Roychoudhury, A. O'Hagan, D. Spiegelhalter, and B. Neuenschwander, "Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information," *Biometrics* 70, no. 4 (2014): 1023–1032.

13. R. Fougeray, L. Vidot, M. Ratta, Z. Teng, D. Skanji, and G. Saint-Hilary, "Futility Interim Analysis Based on Probability of Success Using a Surrogate Endpoint," *Pharmaceutical Statistics* 7 (2024): 971–983.

14. G. Saint-Hilary, V. Barboux, M. Pannaux, M. Gasparini, V. Robert, and G. Mastrantonio, "Predictive Probability of Success Using Surrogate Endpoints," *Statistics in Medicine* 38, no. 5 (2019): 1753–1774.

15. S. Roychoudhury and B. Neuenschwander, "Bayesian Leveraging of Historical Control Data for a Clinical Trial With Time-To-Event Endpoint," *Statistics in Medicine* 39 (2020): 984–995.

16. V. Weru, A. Kopp-Schneider, M. Wiesenfarth, S. Weber, and S. Calderazzo, "Information Borrowing in Bayesian Clinical Trials: Choice of Tuning Parameters for the Robust Mixture Prior," *Statistics in Biopharmaceutical Research* (2024): 1–25. https://doi.org/10.1080/19466315.2026.2646537.

17. A. Callegaro, N. Galwey, and J. J. Abellan, "Historical Controls in Clinical Trials: A Note on Linking Pocock's Model With the Robust Mixture Priors," *Biostatistics* 24, no. 4 (2023): 443–448.

18. N. Best, M. Ajimi, B. Neuenschwander, G. Saint-Hilary, and S. Wandel, "Beyond the Classical Type I Error: Bayesian Metrics for Bayesian Designs Using Informative Priors," *Statistics in Biopharmaceutical Research* 17, no. 4 (2025): 183–196.

19. A. Callegaro, N. Karkada, E. Aris, and T. Zahaf, "Vaccine Clinical Trials With Dynamic Borrowing of Historical Controls: Two Retrospective Studies," *Pharmaceutical Statistics* 22, no. 5 (2023): 475–491.

20. H. Zhang, Y. Shen, J. Li, H. Ye, and A. Y. Chiang, "Adaptively Leveraging External Data With Robust Meta-Analytical-Predictive Prior Using Empirical Bayes," *Pharmaceutical Statistics* 22, no. 9 (2023): 846–860.

21. P. Yang, Y. Zhao, L. Nie, J. Vallejo, and Y. Yuan, "Sam: Self-Adapting Mixture Prior to Dynamically Borrow Information From Historical Data in Clinical Trials," *Biometrics* 79 (2023): 2857–2868.

22. S. Morita, P. F. Thall, and P. Müller, "Determining the Effective Sample Size of a Parametric Prior," *Biometrics* 64, no. 2 (2008): 595–602.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1**: sim70616-sup-0001-Supinfo.pdf.

10970258, 2026, 13-14, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.70616 by Mauro Gasparini - Politecnico Di Torino Sist. Bibl Del Polit Di Torino , Wiley Online Library on [20/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License