On the interplay between prior weight and vague variance in Robust Mixture Priors

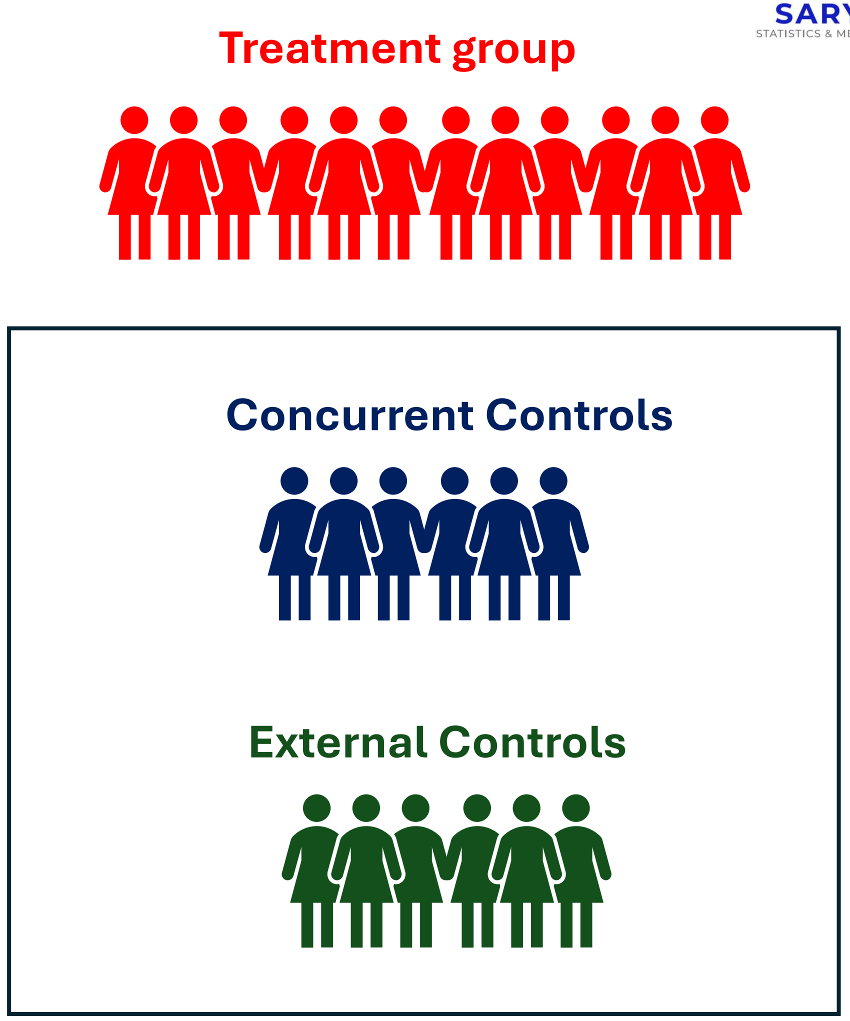
Marco Ratta[1,2], Gaëlle Saint-Hilary[1,2], Mauro Gasparini[1], Pavel Mozgunov[2,3]

[1]Politecnico di Torino, Italy
[2]Saryga, France
[3]University of Cambridge, United Kingdom

PSI Conference
Belfast, June 16 2026

Setting

- Randomized Control Trial (RCT) where a novel treatment is tested versus placebo or standard of care
- Normally distributed control and treatment data, with **unknown means $\boldsymbol{\theta_C}$ and $\boldsymbol{\theta_T}$**, and **known common standard error $\boldsymbol{s}$**.
- 200 patients are allocated with 3:1 allocation ratio (T vs C)
- A Bayesian decision rule is used $\mathbb{P}(\theta_T - \theta_C > 0) > 0.95$
- No prior information is available for treatment arm so that a **vague normal prior** is used for θ_T
- External control information is available are used to construct an **informative prior distribution** $\theta_C^I \sim N(\mu_I, \sigma_I^2)$



Hybrid-control arm

Robust Mixture Priors (RMP)

SARYGA
STATISTICS & METHODOLOGY

$$f_{\theta_C}(\cdot) = \omega f_{\theta_C}^{I}(\cdot) + (1-\omega)\, f_{\theta_C}^{rob}(\cdot)$$

Robust Mixture **Prior** (RMP)

DATA ARE COLLECTED

$$f_{\theta_C}(\cdot|X_C) = \tilde{\omega} f_{\theta_C}^{I}(\cdot|X_C) + (1-\tilde{\omega})\, f_{\theta_C}^{rob}(\cdot\,|X_C)$$

Posterior Mixture

- θ_C: unknown control parameter (i.e. effect on control)
- $f_{\theta_C}^{I}$: informative distribution for θ_C, following a $N(\mu_I, \sigma_I^2)$
- $f_{\theta_C}^{rob}$: *vague* distribution for θ_C, following a $N(\mu_{rob}, \sigma_{rob}^2)$, called **robust component**
- ω: *prior mixture weight,* reflecting the amount of borrowing, between 0 and 1
- $\tilde{\omega}$ is the *posterior mixture weight* reflecting the posterior amount of borrowing, between 0 and 1
- $X_C \sim N(\theta_C, \sigma_C^2)$ is the observed control response

The similarity between concurrent data X_C and the informative component of the RMP is represented by the so called ***drift*** or ***prior-data conflict***

$$D = X_C - \mu_I$$

Specification of the robust component

$$\theta^{rob} \sim N(\mu_{rob}, \sigma_{rob}^2)$$

- The variance of the **robust component** is commonly set so that its **effective sample size (ESS)** is equal to 1 (equivalent to the observation of a single patient). This choice is called **unit-information prior (UIP)**[4] .

- Different definition of ESS are available in literature, e.g. the one of Morita et al.[1] and the one of Neuenschwander et al[2]. In case of normal conjugate analysis both are equal to $ESS_{rob} = s^2/\ \sigma_{rob}^2$.

- There is **no general consensus** regarding the choice of the location of the robust component μ_{rob}. Possible options are:
  - Setting $\mu_{rob} = \mu_I$
  - Setting $\mu_{rob} = x_C$ (data-dependent)

Mixture weight update

Mixture weights are updated **dynamically** depending on the **observed drift** d

$$\widetilde{\omega} = \frac{\omega h^{I}_{X_C}(d)}{\omega h^{I}_{X_C}(d) + (1-\omega) h^{rob}_{X_C}(d)} \longleftrightarrow \widetilde{\Omega} = \Omega \times \frac{h^{I}_{X_C}(d)}{h^{rob}_{X_C}(d)}$$

$h^{I}_{X_C}(\cdot)$ and $h^{rob}_{X_C}(\cdot)$ are respectively the *prior predictive distribution* related to the **informative** and the **robust** component of the RMP, reflecting the expected probability distribution of the data if the underlying parameter followed the prior distributions.

DYNAMIC BORROWING:

- If d is **small** → The posterior weight $\widetilde{\boldsymbol{\omega}}$ **is high** and the informative component is **more influent** in the posterior distribution

- If d is **large** → The posterior weight $\widetilde{\boldsymbol{\omega}}$ **is low** and the informative component is **less influent** in the posterior distribution

$$\Omega := \frac{\omega}{1-\omega}$$

$$\widetilde{\Omega} := \frac{\widetilde{\omega}}{1-\widetilde{\omega}}$$

Issues related to the current use of n-RMPs

- When the variance of the robust component is too large, the posterior weight tends to 1 for any level of drift D (**Lindley's paradox**)

  This happens because the prior predictive density of the robust component will tend to 0 for any level of drift (improper distribution), thus the ratio will tend to $+\infty$

- When the variance of the robust component is not too large (e.g. in UIP), there is an **asymptotic inflation of the type I error** for large values of D (theoretically $\alpha \to 1$)

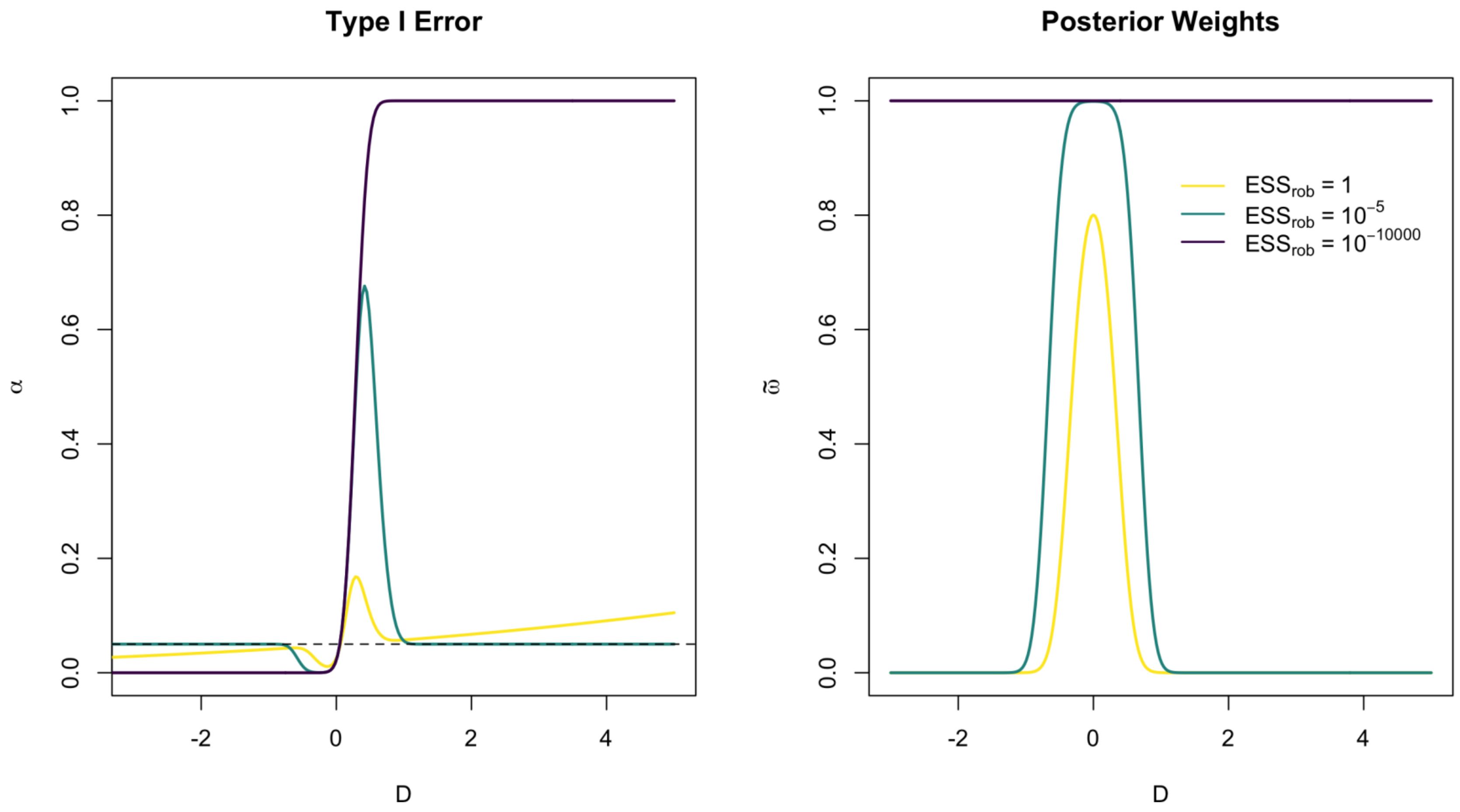
  This happens because when the drift is large, i.e. the posterior weights tend to 0, the inference is fully driven by the robust component, which becomes informative

- When the variance of the robust component is not too large (e.g. in UIP), the choice of the location of the robust component $\boldsymbol{\mu_{rob}}$ **impacts on the posterior inference**.

  The posterior distribution for θ_C is shrinked towards the mean of the robust component (via the standard Bayesian update rule)

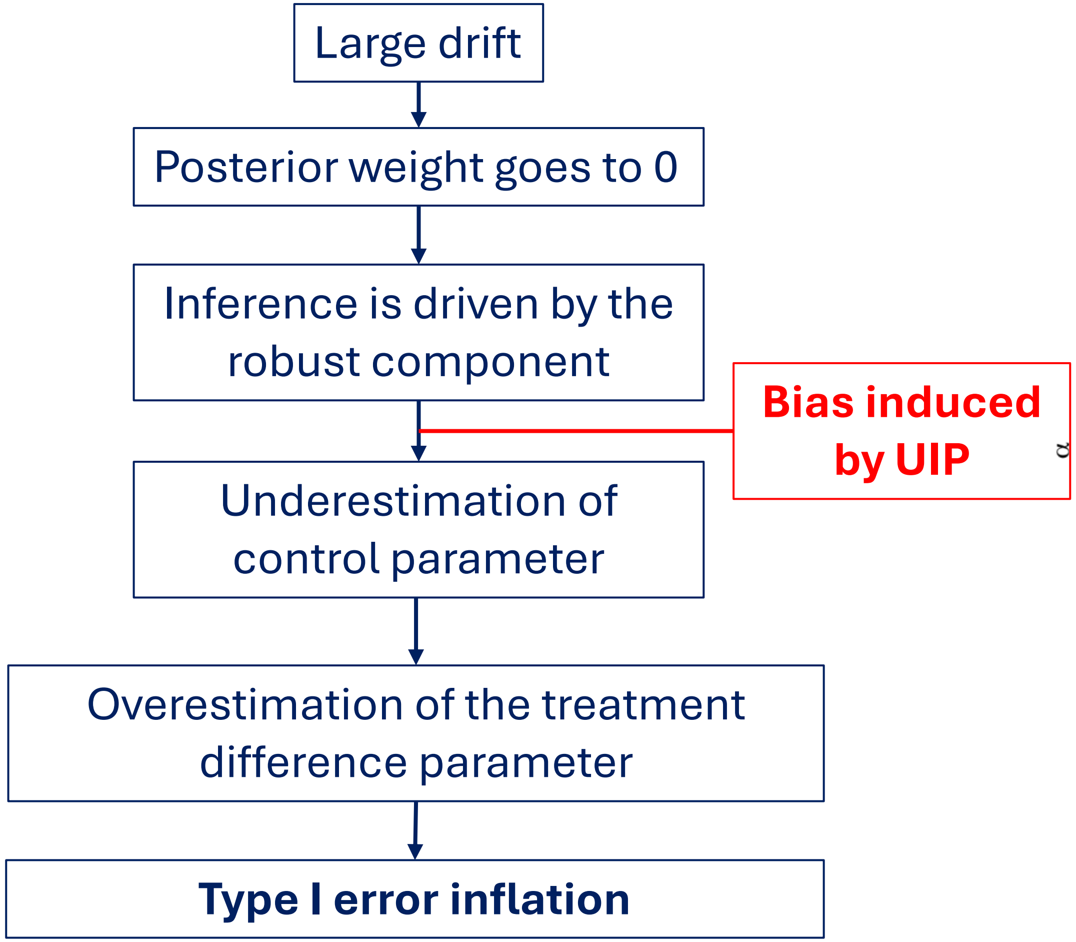
Lindley's paradox

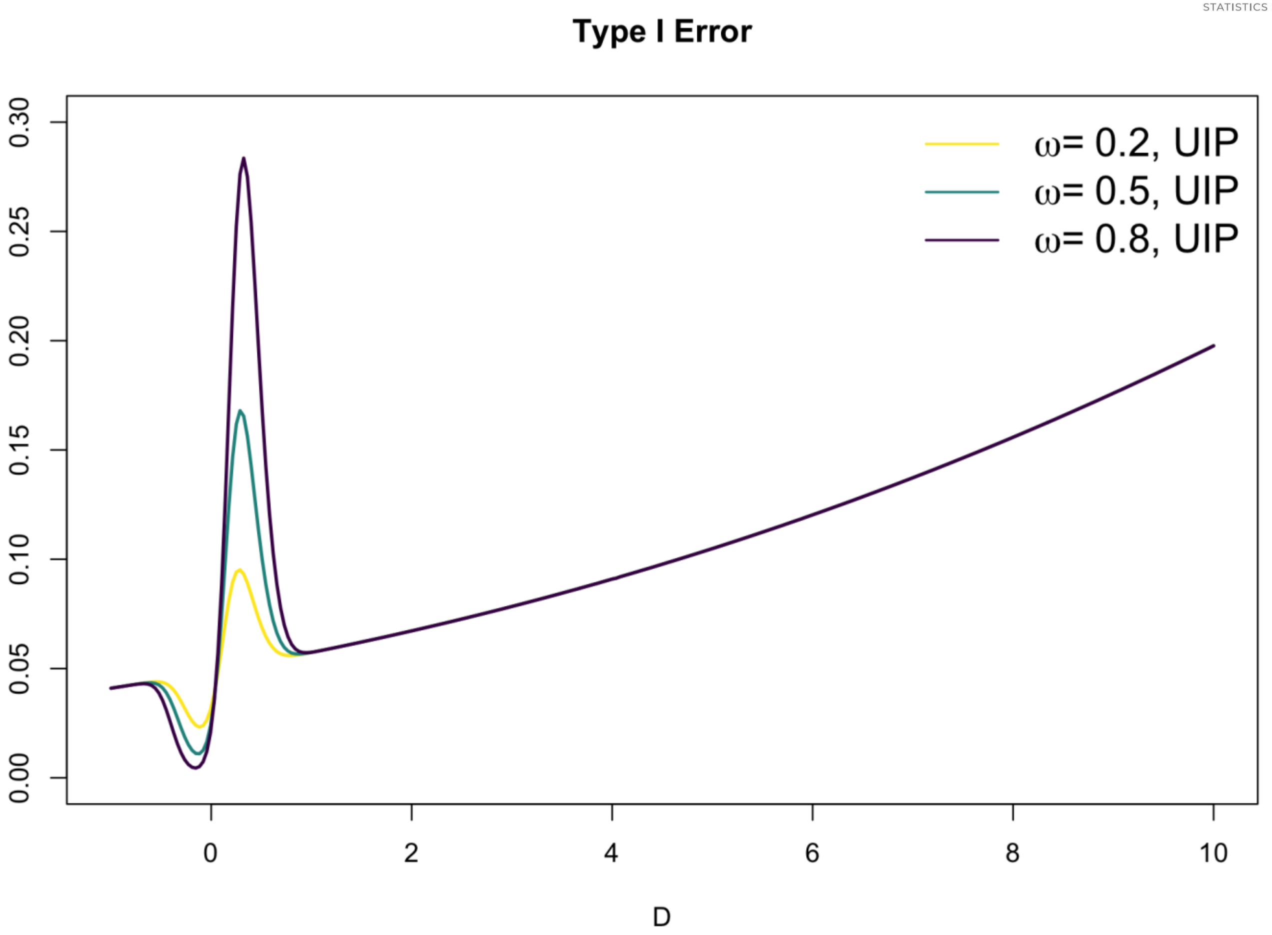
$\omega = 0.5$ **in all curves!**



When the variance of the robust component is too large, the **posterior weight tends to 1** for any level of drift, meaning that **there is no dynamic borrowing.**

Asymptotic type I error inflation





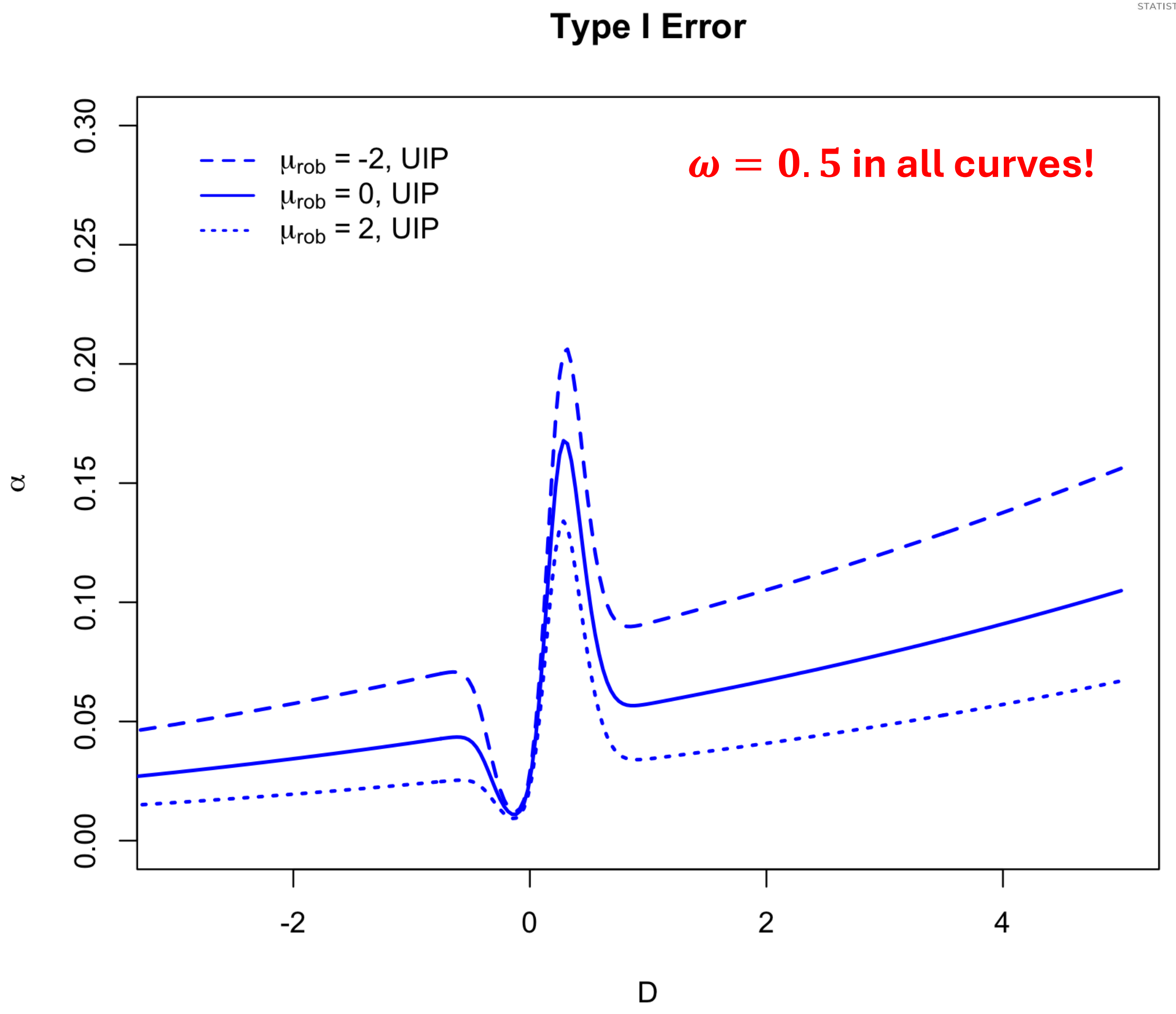
When using UIP, there is an **asymptotic inflation of the type I error** for large values of D (theoretically $\alpha \to 1$)

Impact of $\boldsymbol{\mu_{rob}}$

When using UIP as robust component of the RMP, the choice of μ_{rob} impacts the operating characteristics of the trial

This is a problem for two reasons:

- It introduces **subjectivity**
- The robust component is for making the tails bigger, so $\boldsymbol{\mu_{rob}}$ **should not impact the posterior inference**



Odds Update in n-RMPs

In the normal case the predictive prior densities takes the following form:

$$h_{X_C}^{rob}(d) = \frac{1}{\sqrt{2\pi(\sigma_{rob}^2 + \sigma_C^2)}} e^{-\frac{(d+\mu_I-\mu_{rob})^2}{2(\sigma_{rob}^2+\sigma_C^2)}} \qquad h_{X_C}^{I}(d) = \frac{1}{\sqrt{2\pi(\sigma_I^2 + \sigma_C^2)}} e^{-\frac{d^2}{2(\sigma_I^2+\sigma_C^2)}}$$

And formula for the odds update can be expressed accordingly as:

$$\widetilde{\Omega} = \Omega \times R \times e^{\frac{d^2}{2(\sigma_I^2+\sigma_C^2)} - \frac{(x_C-\mu_{rob})^2}{2R^2(\sigma_I^2+\sigma_C^2)}}$$

Vagueness Assumption

$$R^2 = \frac{\sigma_{rob}^2+\sigma_C^2}{\sigma_I^2+\sigma_C^2} >> 0$$

This decomposition is convenient because each part of the equation is related to one specific factor:

- The **green** factor is related to the choice of the mixture weight ω
- The **blue** factor is related to the choice of the variance of the robust component σ_{rob}^2
- The **red** factor is related to the observed drift d

Large variance robust components

If the variance of the robust component σ^2_{rob} is large, then the vagueness assumption $(R \gg 0)$ is satisfied, and the posterior odds can be approximated as:

$$\widetilde{\Omega}(d) \approx \Omega \times R \times \underbrace{e^{\frac{d^2}{2(\sigma_I^2 + \sigma_C^2)}}}_{\text{Depends only on drift!}}$$

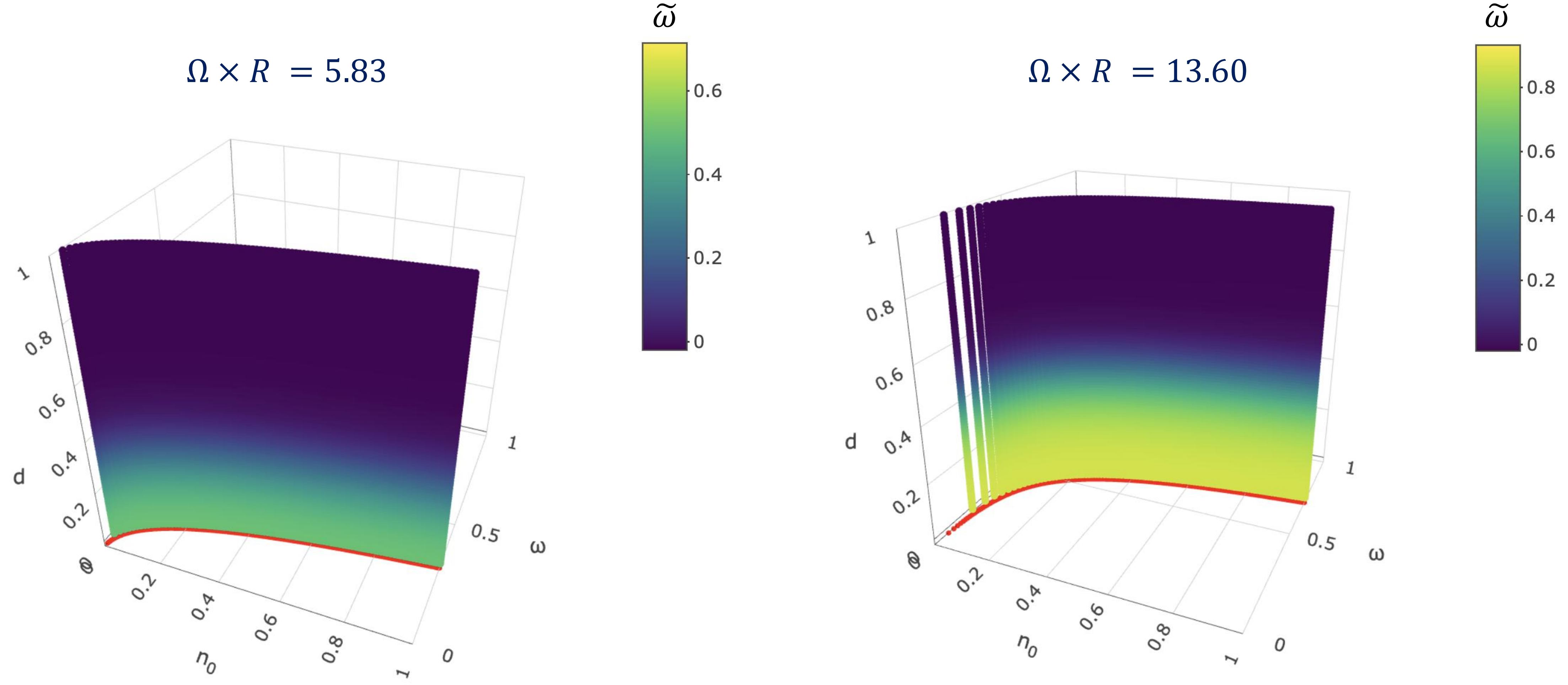
It follows that:

1) All pairs (ω, σ^2_{rob}) leading to the same value of $\Omega \times R$ have the same posterior odds $\widetilde{\Omega}(d)$ → Extremely large values of σ^2_{rob} can be used without incurring in Lindley's paradox if (ω, σ^2_{rob}) are **jointly selected**

2) If the variance of the robust component is large enough σ^2_{rob}, the posterior odds are indepentent on the choice of the location of the robust component μ_{rob} → **whatever** convenient value of μ_{rob} can be used

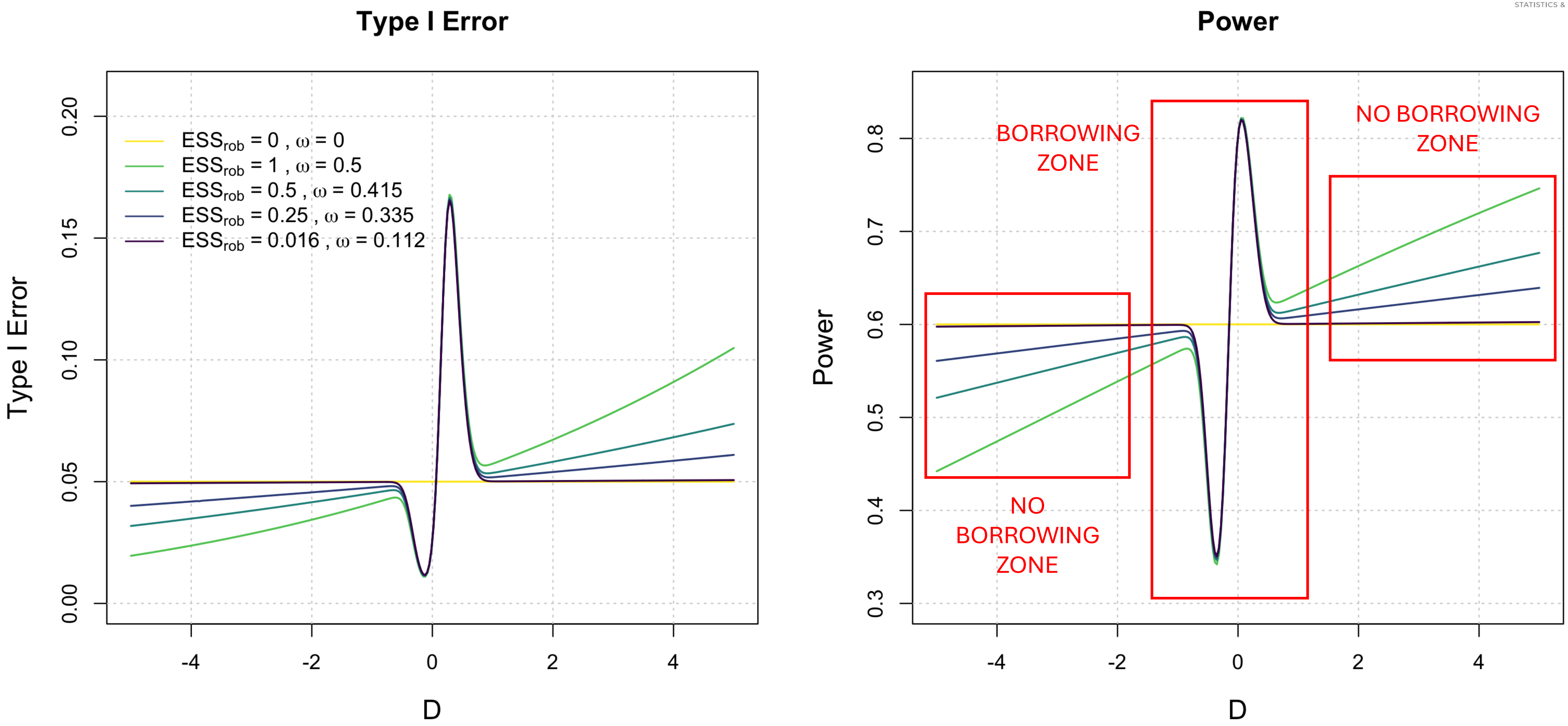
3) The larger the variance of the robust component the lower the bias introduced by the robust component in the control parameter estimation → Type I error is **asymptotically controlled** even for extremely large drifts

Overcoming Lindley's paradox



For all the pairs (ω, σ_{rob}^2) leading to the same value of $\Omega \times R$, the posterior weights $\widetilde{\omega}$ as a function of the observed mean control response x_c exhibit the same profile

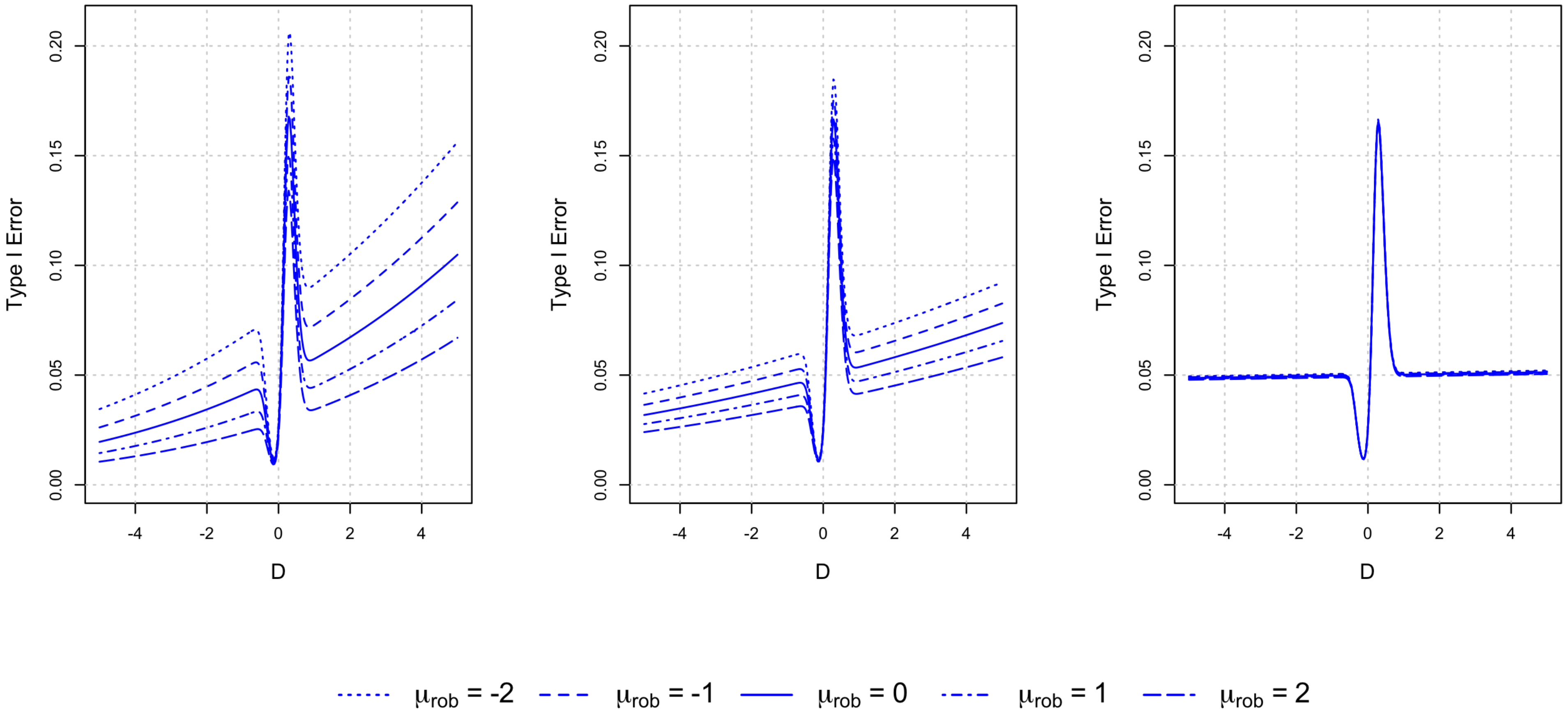
Frequentist operating characteristics



Robustness to the specification of μ_{rob}

All these pairs have the same $\Omega \times R$

$ESS_{rob} = 1$, $\omega = 0.5$ | $ESS_{rob} = 0.5$, $\omega = 0.415$ | $ESS_{rob} = 0.031$, $\omega = 0.151$



The lower the ESS of the robust component of the RMP, the lower the impact of the location of the robust component μ_{rob} in the posterior inference

Take-home messages

- The operating characteristics of the normal RMPs depend on the **joint selection** of the prior weight ω and the variance of the robust component σ_{rob}^2

- Infinite many pairs (ω, σ_{rob}^2) yield to **very similar inference** in the borrowing region of the parameter space with minor to moderate prior-data conflict

- There exists a combination of parameters $\Omega \times R$ which drives the borrowing profile of the RMP

- Using RMPs with large variance robust components is convenient because:
  - It effectively **prevents from Lindley's paradox**
  - It **prevents from asymptotic type I error inflation**
  - It makes the impact of the choice of the location of the robust component practically null

Extensions

- The same theory can be extended to the Beta-binomial case, where the RMP is a mixture of two Beta distributions, however:
  - The location is not an issue, because the non informativeness is obtained by letting the two parameters of the Beta distribution be equal
  - There is no asymptotic inflation of type I error because the domain of the probability parameter is naturally bounded

- The same theory can be extended with minor adjustments to the case where the informative component of the RMP is a mixture itself.

References

1. Morita S, Thall PF, Müller P. **Determining the effective sample size of a parametric prior.** *Biometrics*. (2008), pp 595-602.

2. Neuenschwander B, Weber S, Schmidli H, O'Hagan A. **Predictively consistent prior effective sample sizes.** *Biometrics*. (2020)

3. Joost van Rosmalen et al. “**Including historical data in the analysis of clinical trials: Is it worth the effort?**” *Statistical Methods in Medical Research* 27 (2018), pp. 3167–3182

4. Heinz Schmidli et al. “**Robust meta-analytic-predictive priors in clinical trials with historical control information**”. *Biometrics* 70 (2014), pp. 1023–1032

5. Timothy Mutsvari, Dominique Tytgat, and Rosalind Walley. “**Addressing potential prior-data conflict when using informative priors in proof-of-concept studies**”. *Pharmaceutical Statistics* 15 (2016), pp. 28–36

6. Vivienn Weru et al. “**Information borrowing in Bayesian clinical trials: choice of tuning parameters for the robust mixture prior**”. arXiv: 2412.03185 (2024).

7. Andrea Callegaro, Nicholas Galwey, and Juan J. Abellan. “**Historical controls in clinical trials: a note on linking Pocock’s model with the robust mixture priors**”. *Biostatistics* 24 (2023), pp. 443–448

8. Best, N., Ajimi, M., Neuenschwander, B., Saint-Hilary, G., & Wandel, S. (2024). **Beyond the Classical Type I Error: Bayesian Metrics for Bayesian Designs Using Informative Priors.** *Statistics in Biopharmaceutical Research* (2024) pp. 183–196

Check out the paper open access !

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, Gaëlle Saint-Hilary, Mauro Gasparini, Pavel Mozgunov

First published: 30 May 2026 | https://doi.org/10.1002/sim.70616 | VIEW METRICS