October 22, 2026

A hybrid frequentist-Bayesian view on hypothesis testing in clinical trials borrowing external information

Silvia Calderazzo

Joint work with: Vivienn Weru, Monika Jelizarow, Manuel Wiesenfarth and Annette Kopp-Schneider

Bayes 2026 - Leiden

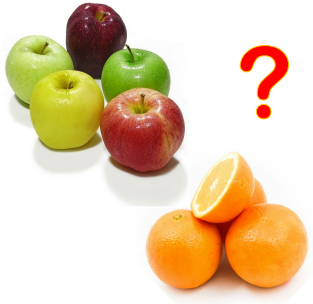
Information borrowing

- **External information** is often available when designing a trial (historical/concurrent trial, real world evidence, expert opinion...).
- Incorporation can improve **trial efficiency**.
- Bayesian paradigm offers a natural framework to borrow information

  $\rightarrow$ Specification of **informative prior distributions**.
- Potential for **heterogeneity** (prior-data conflict).

Robust borrowing

- Discount amount of borrowed information
- Various methods, e.g.:
  - (Robust) meta-analytic approaches (Neuenschwander et al., 2010; Schmidli et al., 2014)
  - Power priors (Ibrahim and Chen, 2000; Ibrahim et al., 2015),
  - Commensurate priors (Hobbs et al., 2011, 2012).
- **Choice/estimation of borrowing parameters/distributions** required



https://picryl.com/media/apples-fruit-food-food-drink-9d8ff
https://www.pexels.com/de-de/foto/orange-frucht-519058/

Here focus on frequentist operating characteristics: test error rates

No free lunch

It is not possible to gain power via information borrowing
while controlling type I error (TIE) rate
if a uniformly most powerful (UMP) test exists
(Kopp-Schneider et al., 2020).

→ True for both **dynamic and static** mechanisms.



Image by https://www.vectorportal.com" Vectorportal.com "https://creativecommons.org/licenses/by/4.0/" CC BY

Borrowing information: how to balance?

A trade-off between power and (maximum) TIE rate is always present

- One-arm trial (treatment effect borrowing): choosing a borrowing approach typically reduces to the **choice of a specific power function/TIE rate** (and H_0 rejection region in terms of data outcomes) (Whitt, 1979; Calderazzo et al., 2024).



Borrowing information: how to balance?

A trade-off between power and (maximum) TIE rate is always present

- One-arm trial (treatment effect borrowing): choosing a borrowing approach typically reduces to the **choice of a specific power function/TIE rate** (and H_0 rejection region in terms of data outcomes) (Whitt, 1979; Calderazzo et al., 2024).
- Two-arm trial (control and/or treatment arm borrowing): more complex, same maximal TIE rate can be associated with **different test decision rules** (and rejection regions).

How to select and evaluate an approach?



Rationales

Approach	TIE rate control	External information	Properties
Frequentist no borrowing	Strict	Discarded	Maximises power for chosen TIE rate level
Bayesian full borrowing	Irrelevant	Full inclusion	Optimal Bayes properties if prior is 'true'
Robust	In some form*	Adaptively downweighted	If a robust fixed prior is used and it is 'true', optimal Bayes properties

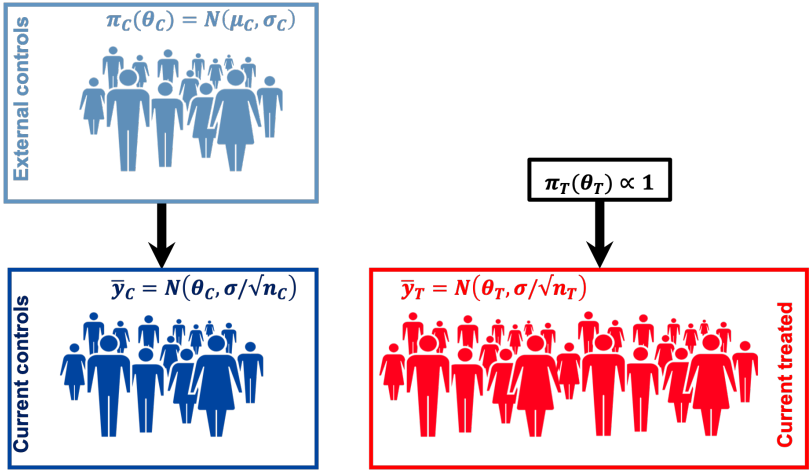
*TIE rate is controlled:

- **Globally with inflation**: e.g., heavy tailed (O'Hagan, 2010; Weru et al., 2026) or Empirical Bayes (e.g. Gravestock and Held, 2017) priors.
- **Locally** (Stallard et al., 2020; Best et al., 2025; Weru et al., 2026): over a sufficient range (additional assumptions).
- **On average** (Best et al., 2025): ideally across different prior distributions (additional assumptions).

Compromise approach

Set-up

$$H_0: \theta_T - \theta_C \leq 0 \; vs \; H_1: \theta_T - \theta_C > 0$$



Bayesian and frequentist test decisions

Aim: borrow information with pre-specified global caps on TIE rate inflation and/or deflation.

Frequentist test decision (**FD**):
reject H_0 if

$$\frac{\bar{y}_T - \bar{y}_C}{\sqrt{\frac{\sigma^2}{n_T} + \frac{\sigma^2}{n_C}}} > z_{1-\kappa},$$

Bayes test decision (**BD**):
reject H_0 if $P^{\pi}(H_0|y) \leq \gamma$, i.e.

$$\frac{\bar{y}_T - \frac{\mu_C A_C + \bar{y}_C}{1+A_C}}{\sqrt{\frac{\sigma^2}{n_T} + \frac{\sigma^2/n_C}{1+A_C}}} > z_{1-\gamma}, \quad A_C = \frac{\sigma^2/n_C}{\sigma_C^2}$$

Tune test decision threshold κ to obtain the BD under the frequentist test $\rightarrow \kappa^{BD}(\bar{y}_C)$.

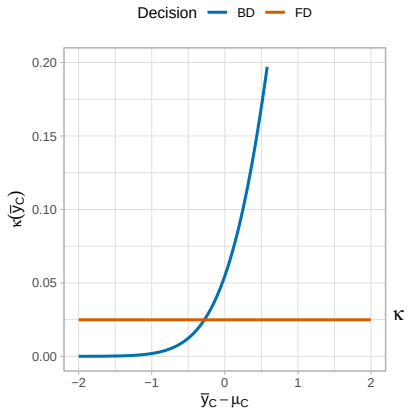
Identifying the rejection region of the Bayes test

Frequentist test decision (**FD**): reject H_0 if

$$\frac{\bar{y}_T - \bar{y}_C}{\sqrt{\frac{\sigma^2}{n_T} + \frac{\sigma^2}{n_C}}} > z_{1-\kappa},$$

Bayes test decision (**BD**): reject H_0 if

$$\frac{\bar{y}_T - \bar{y}_C}{\sqrt{\frac{\sigma^2}{n_T} + \frac{\sigma^2}{n_C}}} > z_{1-\kappa^{BD}(\bar{y}_C)},$$



A ‘compromise’ approach

We **modify the test decision threshold**:

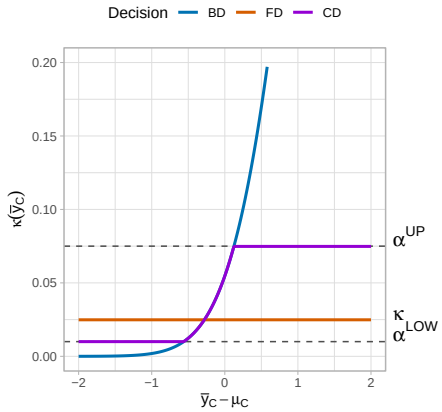
Compromise decision approach (**CD**)

$$\kappa^{CD}(\bar{y}_C) = \max(\min(\kappa^{BD}(\bar{y}_C), \alpha^{UP}), \alpha^{LOW})$$

We then reject H_0 if

$$\frac{\bar{y}_T - \bar{y}_C}{\sqrt{\frac{\sigma^2}{n_T} + \frac{\sigma^2}{n_C}}} > z_{1-\kappa^{CD}(\bar{y}_C)},$$

$\rightarrow$ Cap of TIE rate between α^{LOW} and α^{UP} guaranteed.



Calderazzo et al. (accepted for discussion in Biometrics)

Evaluation

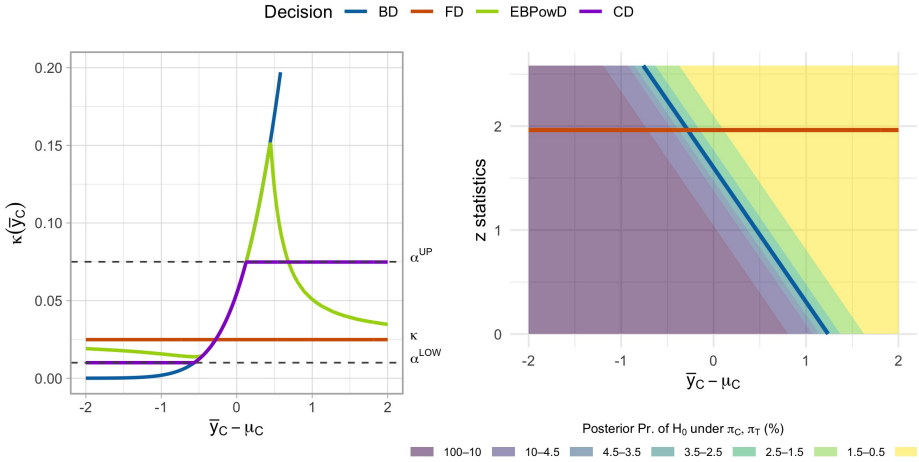
Simulation set-up

We compare **conditional** and **unconditional** properties of the following **analysis priors/approaches**:

- Frequentist approach, no borrowing (**FD**).
- Bayesian approach, full borrowing (**BD**).
- Compromise approach, $\alpha^{UP} = 0.075$, $\alpha^{LOW} = 0.01$ (**CD**)
- Empirical Bayes power prior $N(\mu_C, \sigma_C/\sqrt{\hat{\lambda}})$, $\lambda \in [0, 1]$ (**EBPowD**) (Gravestock and Held, 2017)

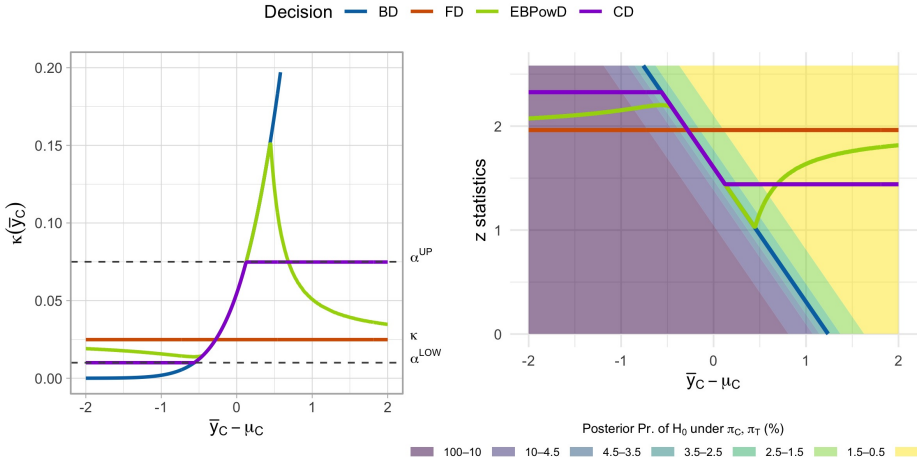
Settings:
$\bar{y}_C \sim N(\theta_C, 1/\sqrt{n_C})$, $\bar{y}_T \sim N(\theta_T, 1/\sqrt{n_T})$, $n_T = 20$, $n_C = 10$;
$\mu_C = 0$, $\sigma_C = 1/\sqrt{10}$, $\theta_T \propto 1$.
$\kappa = \gamma = 0.025$.

Conditional assessment: critical values



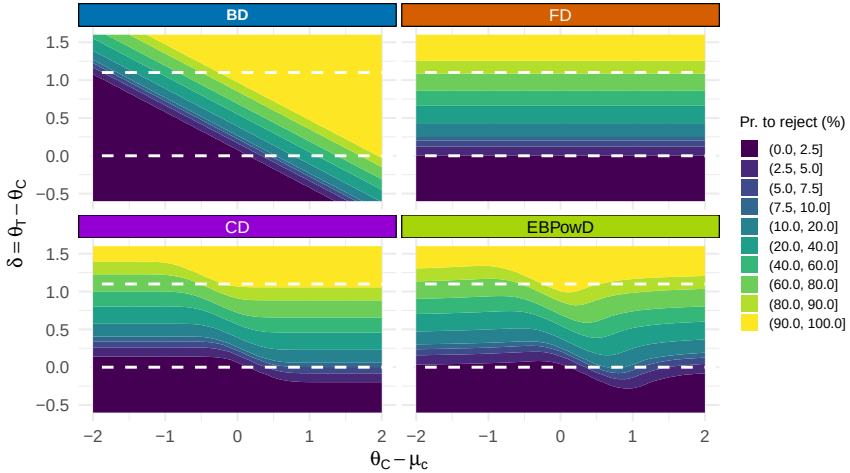
Conditional: on data outcomes $\bar{y}_C, \bar{y}_T$.

Conditional assessment: critical values



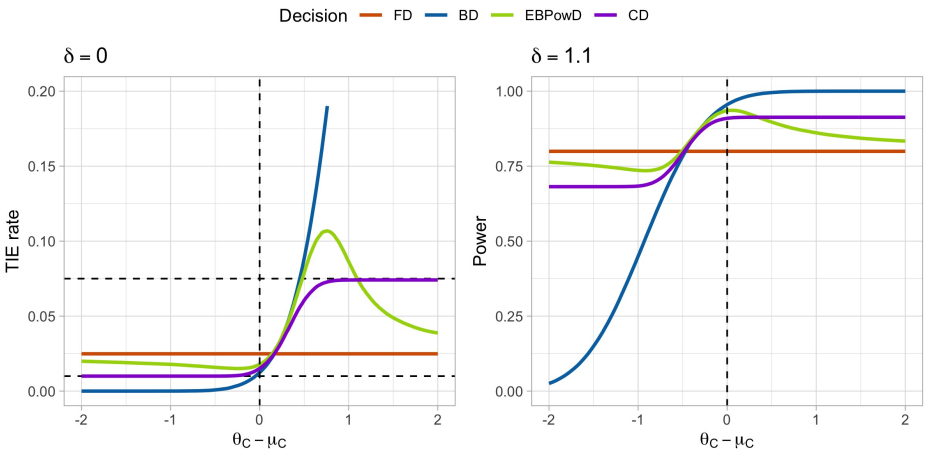
Conditional: on data outcomes $\bar{y}_C, \bar{y}_T$.

Unconditional assessment: TIE rate and power



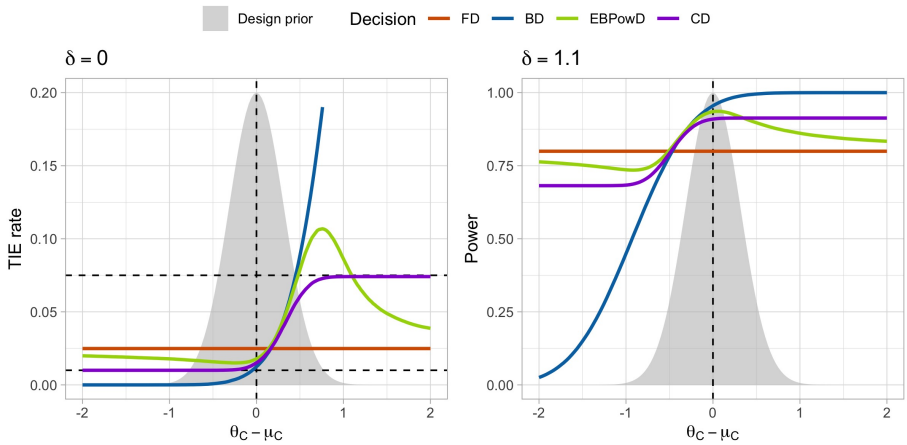
Unconditional: averaged over data outcomes $\bar{y}_C \sim N(\theta_C, \sigma/\sqrt{n_C}), \bar{y}_T \sim N(\theta_C + \delta, \sigma/\sqrt{n_T})$.

Unconditional assessment: TIE rate and power



Unconditional assessment: Average TIE rate and power

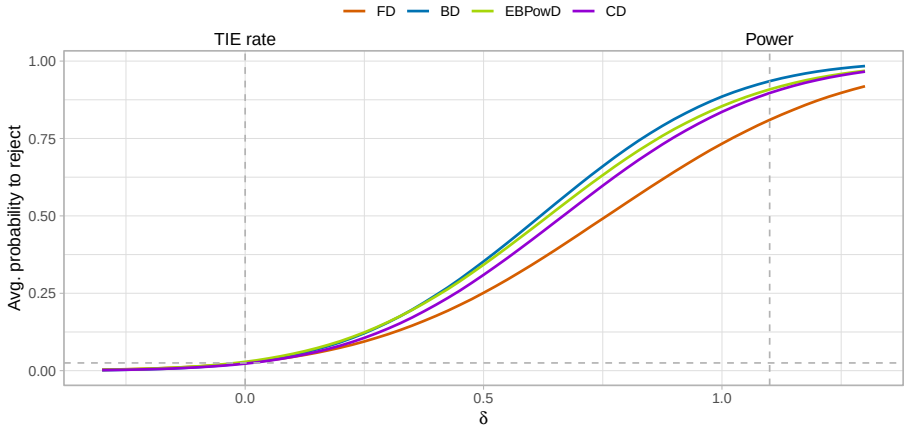
No conflict



See Best et al. (2025). **Design prior** equal to informative analysis prior: $N(\mu_C, \sigma_C)$.

Unconditional assessment: Average TIE rate and power

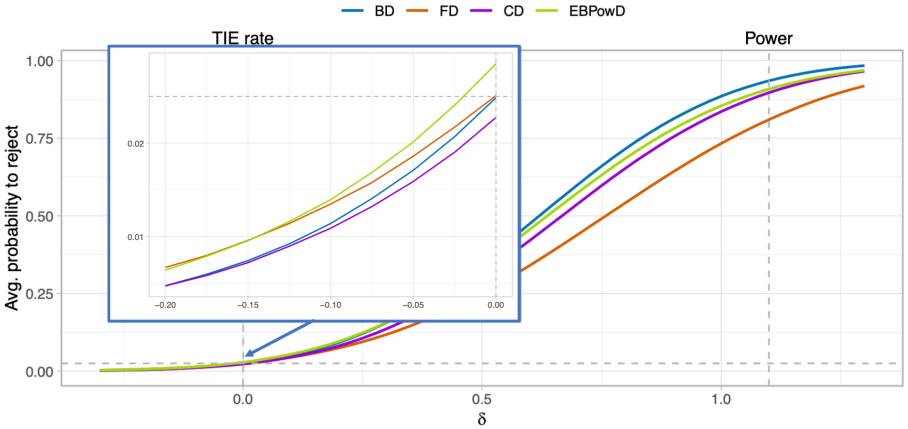
No conflict



Average with respect to the design prior equal to informative analysis prior: N(μ_C, σ_C).

Unconditional assessment: Average TIE rate and power

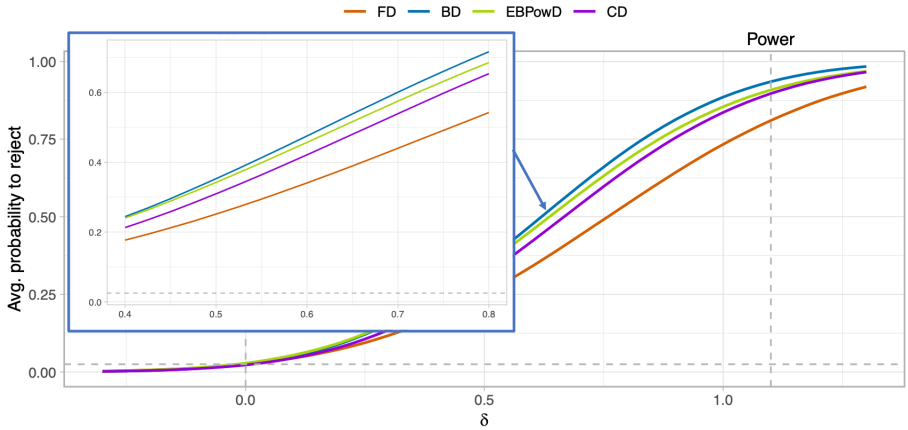
No conflict



Average with respect to the design prior equal to informative analysis prior: $N(\mu_C, \sigma_C)$.

Unconditional assessment: Average TIE rate and power

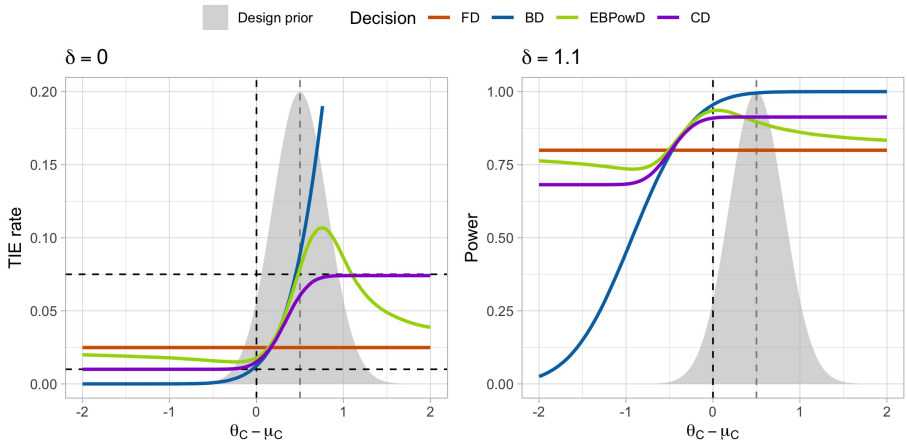
No conflict



Average with respect to the design prior equal to informative analysis prior: N(μ_C, σ_C).

Unconditional assessment: Average TIE rate and power

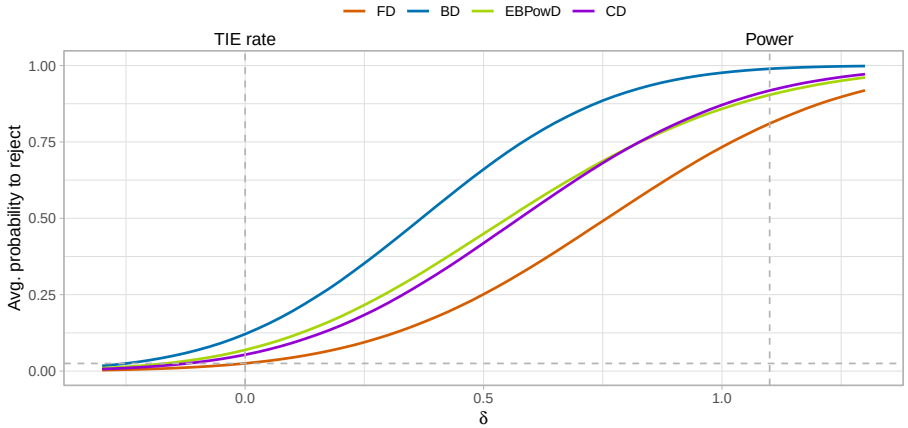
Shifted design prior



See Best et al. (2025). Design prior equal to shifted informative analysis prior: $N(\mu_C + 0.5, \sigma_C)$.

Unconditional assessment: Average TIE rate and power

Shifted design prior



Average with respect to the design prior equal to shifted informative analysis prior: $\mathrm{N}(\mu_C + 0.5, \sigma_C)$.

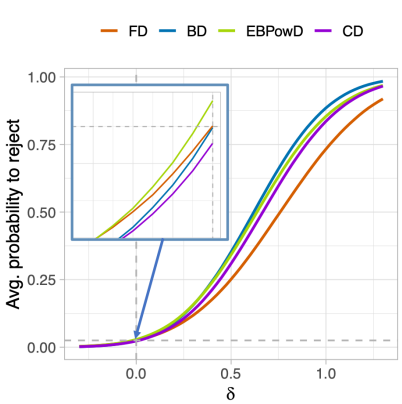
Conclusions

Which approach and when?

No assumptions on the treatment effect

- Average probability-to-reject function: integrate over θ_C.
- δ fixed, independent of θ_C:
  - If analysis $=$ design prior for θ_C:
    - Average TIE rate calibrated (Best et al., 2025),
    - Average TIE rate and Power gains compared to FD
    - $\rightarrow$ **Information borrowing improves OCs**.

Analysis prior $=$ design prior for θ_C:

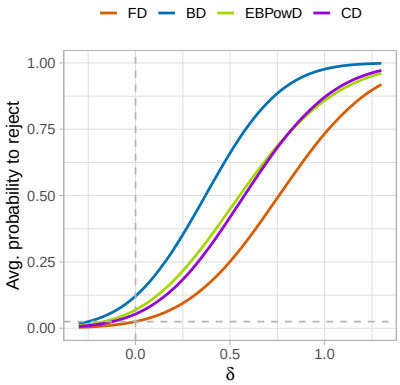


Which approach and when?

No assumptions on the treatment effect

- Average probability-to-reject function: integrate over θ_C.
- δ fixed, independent of θ_C:
  - If analysis $=$ design prior for θ_C:
    - Average TIE rate calibrated (Best et al., 2025),
    - Average TIE rate and Power gains compared to FD
    - $\rightarrow$ **Information borrowing improves OCs**.
  - If analysis $\neq$ design prior, average test error rates:
    - Are insensitive to the design prior for the FD
    - Are capped for robust approaches
    - $\rightarrow$ **FD and robust give "safety net"**.

Analysis prior $\neq$ design prior for θ_C (shift):



Outlook

Decision-theoretic view

Conditional on θ_C, δ, the frequentist risk (c_1=TIE cost, c_0= TIIE cost):

$$R(\theta_C, \delta; \mathrm{D}) = c_1 \underbrace{\beta(\theta_C, \delta; \mathrm{D})}_{\text{Pr. to reject}} \mathrm{I}_{\delta \leq \delta_0} + c_0 \left[1 - \beta(\theta_C, \delta; \mathrm{D})\right] \mathrm{I}_{\delta > \delta_0}.$$

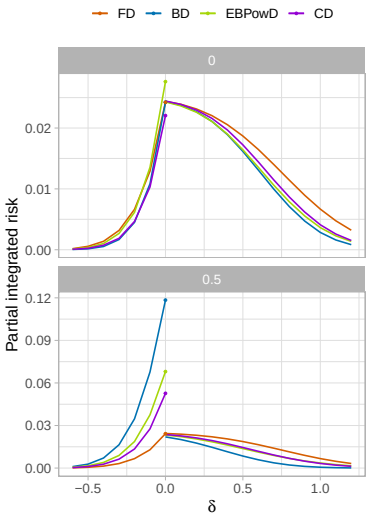
Integrate over $\theta_C|\delta$, the partial integrated risk for δ:

$$r(\delta; \mathrm{D}) \quad = \quad c_1 \underbrace{\bar{\beta}(\delta; \mathrm{D})}_{\substack{\text{Avg. Pr.} \\ \text{to reject}}} I_{\delta \leq \delta_0} + c_0 [1 - \bar{\beta}(\delta; \mathrm{D})] I_{\delta > \delta_0}.$$

What about δ?

- An informative prior on $\theta_T \rightarrow (\theta_C, \delta)$ correlated a priori
- An informative prior on δ independent of θ_C can also be assumed

Outlook: investigate the partial integrated risk & full Bayesian optimality under different analysis and design prior configurations.



Discussion

- Any borrowing approach identifies a specific **rejection region**.
- Identifying and plotting the rejection region may aid **interpretability and robustness from a frequentist standpoint**.
- It is also possible to **tune** the rejection region in order to cap TIE rate of the Bayes Decision approach $\rightarrow$ Compromise Decision:



  - Such approach **does not define a prior** (so we cannot claim Bayesian optimality)...
  - ...but it shows in practice **acceptable conditional and unconditional** behaviour
  - Capping provides a **global safeguard** in a transparent form
  - **Binomial outcomes** and **full discard in case of extreme conflict** are also addressed
  - **Theoretical properties, approximations and parameter choices** discussed in Calderazzo et al. (accepted for discussion in Biometrics, arXiv preprint: 2508.16348) .

References

Best, N., Ajimi, M., Neuenschwander, B., Saint-Hilary, G., and Wandel, S. (2025). Beyond the Classical Type I Error: Bayesian Metrics for Bayesian Designs Using Informative Priors. *Statistics in Biopharmaceutical Research*, 17(2):183–196.

Calderazzo, S., Wiesenfarth, M., and Kopp-Schneider, A. (2024). Robust incorporation of historical information with known type I error rate inflation. *Biometrical Journal*, 66(1):2200322.

Calderazzo, S., Wiesenfarth, M., Weru, V., and Kopp-Schneider, A. (accepted). Principled type I error rate inflation in two-arm clinical trial designs with external control information borrowing. *Biometrics*.

Gravestock, I. and Held, L. (2017). Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical statistics*, 16(5):349–360.

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3):1047–1056.

Hobbs, B. P., Sargent, D. J., and Carlin, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian analysis (Online)*, 7(3):639.

Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science*, 15(1):46–60.

Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015). The power prior: theory and applications. *Statistics in medicine*, 34(28):3724–3749.

Kopp-Schneider, A., Calderazzo, S., and Wiesenfarth, M. (2020). Power gains by using external information in clinical trials are typically not possible when requiring strict type I error control. *Biometrical Journal*, 62(2):361–374.

Neuenschwander, B., Capkun-Niggli, G., Branson, M., and Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials*, 7(1):5–18.

O'Hagan, A. (2010). *Kendall's Advanced Theory of Statistic 2B*. John Wiley & Sons.

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

Stallard, N., Todd, S., Ryan, E. G., and Gates, S. (2020). Comparison of bayesian and frequentist group-sequential clinical trial designs. *BMC Medical Research Methodology*, 20(1):4.

Weru, V., Kopp-Schneider, A., Wiesenfarth, M., Weber, S., and Calderazzo, S. (2026). Information borrowing in bayesian clinical trials: choice of tuning parameters for the robust mixture prior. *Statistics in Biopharmaceutical Research*, 0(0):1–25.

Whitt, W. (1979). A note on the influence of the sample on the posterior distribution. *Journal of the American Statistical Association*, 74(366):424–426.