Borrowing from external data in early clinical trials using Bayesian methods

Annette Kopp-Schneider, Vivienn Weru, Manuel Wiesenfarth and Silvia Calderazzo
Division of Biostatistics, German Cancer Research Center (DKFZ)
Heidelberg, Germany

Why borrowing from external information?

- Sample sizes in precision oncology trials are often small.
- Sample sizes in precision oncology pediatric trials are even much smaller.
- External information is often available when designing a trial.

→ Can external information be used to increase trial efficiency?

Borrowing from what source?

- External trial data, e.g.:
  - Control arm (e.g., Standard of Care) of other clinical trials
    (→ Pocock criteria (1976))
  - Extrapolation: Treatment effect for adults available from clinical trial
    → use for pediatric trial?
- Real world data: Patient registry, Natural history data, other observational data collections …
- Select from external information for borrowing based on, e.g., similarity of historical patients to patients in current trial: Propensity score …
- Expert opinion

Borrowing: how?

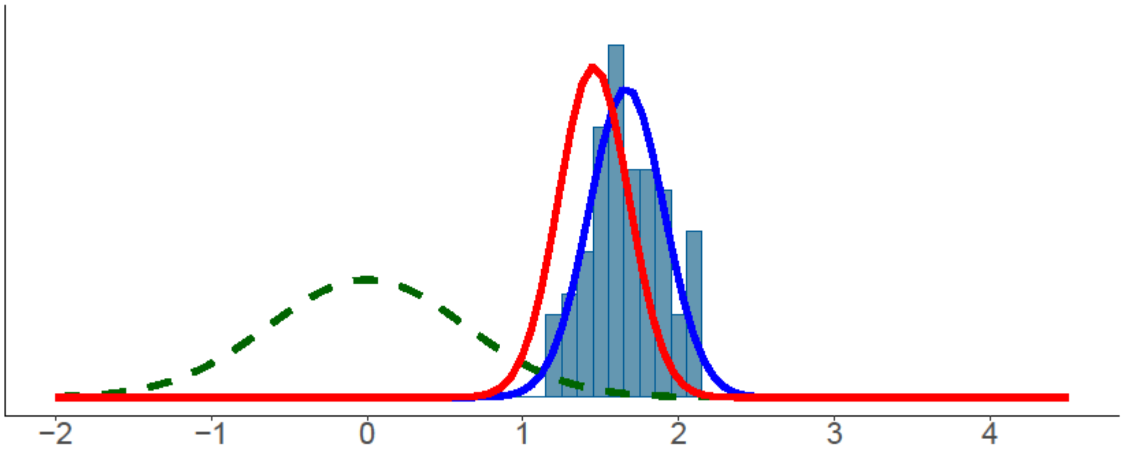
- Frequentist methods are available (see e.g. Viele et al. 2014): e.g., test-and-pool.
- Bayesian methods are ideally suited since external information can be captured in informative prior distribution.

Bayesian updating

Prior $\pi(\theta)$

Data $\pi(y|\theta)$

Posterior $\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta)$



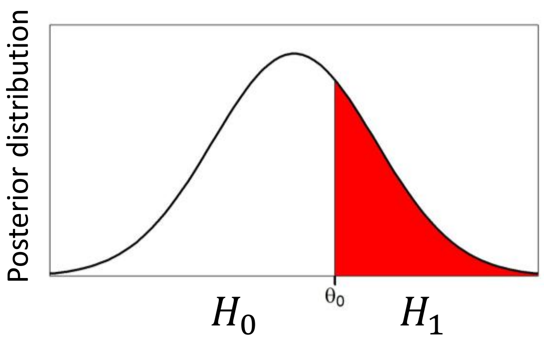
θ

dkfz.

Hypothesis testing with Bayesian methods

- Hypothesis test: $H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$

- Test decision in Bayesian framework:
  reject $H_0 \Leftrightarrow$ **P(H_1 | current data, prior)** $> 1 - \alpha$



- Bayesian decision using „non-informative prior" $\equiv$ Frequentist decision:
  reject $H_0 \Leftrightarrow$ P(H_1 | current data, non-informative prior) $> 1 - \alpha$
  has Type 1 Error probability = α.

- Borrowing from external data by incorporating information into the prior.

Frequentist operating characteristics (OC) of Bayesian hypothesis tests

- Frequentist operating characteristics (OC) of hypothesis test when borrowing from external data are of interest:
  - Type 1 error probability (T1E)
  - Power(θ) for $\theta \in H_1$

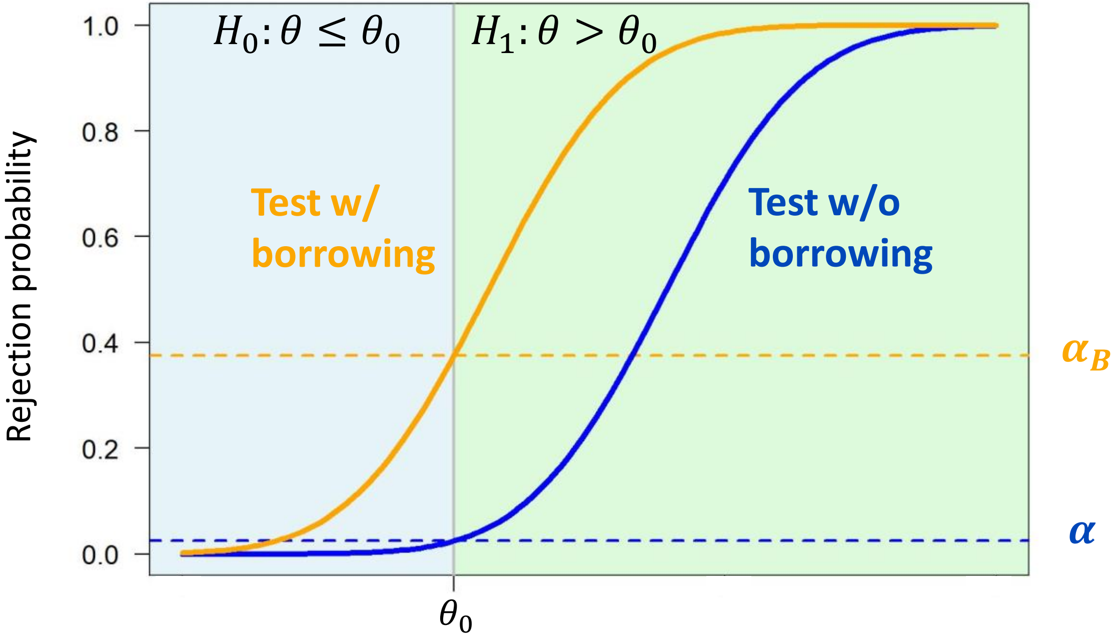
- **Problem:**
  How to assess potential power gain due to borrowing if T1E is changed by borrowing, e.g.,

  - **without (w/o)** borrowing: T1E = 0.025, power = 0.71
  - **with (w/)** borrowing: T1E = 0.046, power = 0.79

?

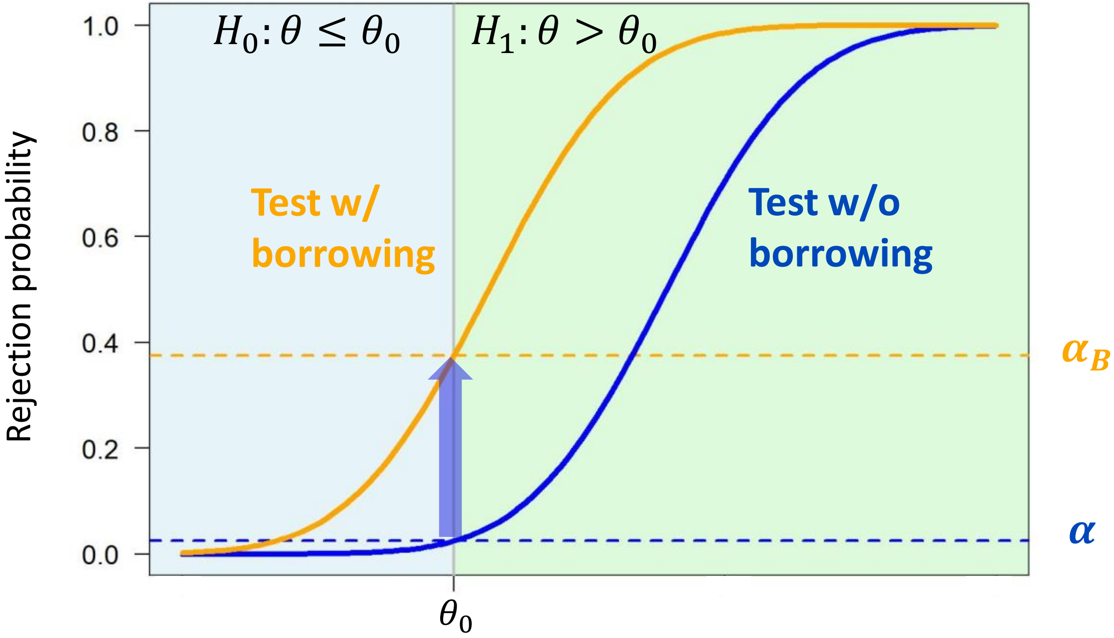
Problem

Fair comparison of
OC w/ and w/o borrowing?



Problem

Fair comparison of
OC **w/** and **w/o** borrowing?



Solution

„**test calibrated to borrowing**“ = test w/o borrowing, but T1E set to α_B instead of α

→ test calibrated to borrowing and test w/ borrowing have same T1E ($= \alpha_B$)

→ evaluate: power(test w/ borrowing) − power(test calibrated to borrowing)

(AKS et al. 2024)

Comparing OC w/ and w/o borrowing

power(test w/ borrowing) – power(test calibrated to borrowing)

Power difference = 0: No power gain by borrowing.

In general:

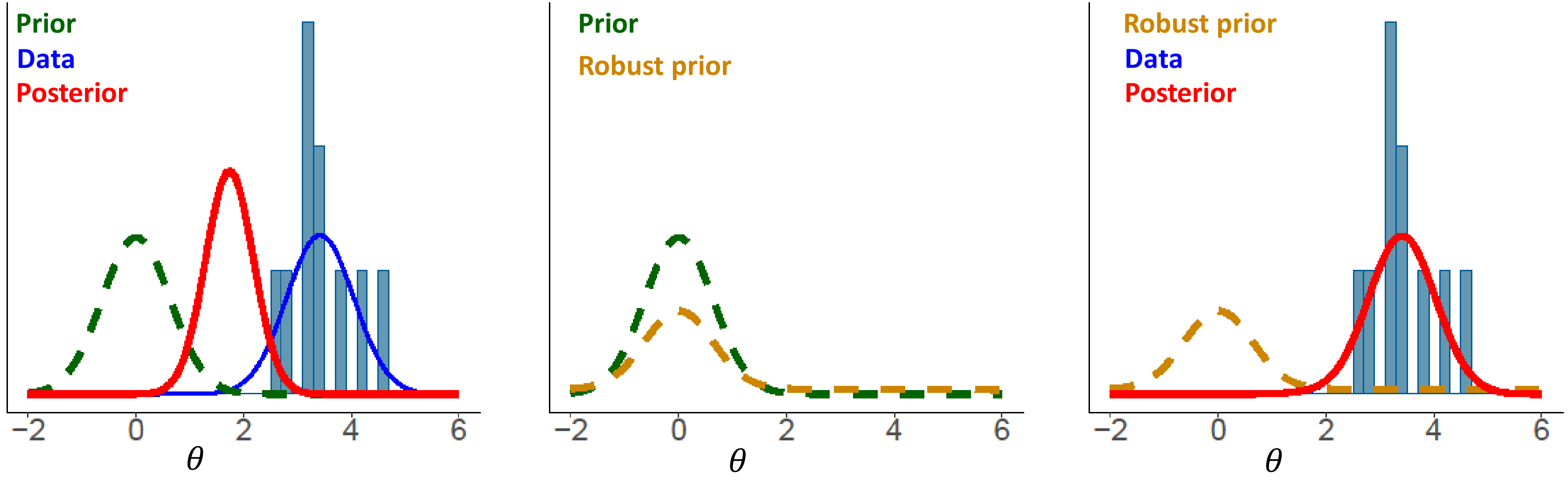
- If a uniformly most powerful (UMP) test exists in the specific hypothesis test situation

→ no test can have more power (AKS et al. 2020).

- True irrespective of borrowing approach!

Static vs. dynamic borrowing

- Static borrowing: Fix the amount of borrowing a priori.
- Dynamic borrowing:
  Adjust the amount of borrowing according to similarity of external information to current data, i.e. discount external data in case of prior data conflict:



(Dynamic) Borrowing approaches

- Empirical Bayes Power Prior approach (Gravestock, Held et al 2017)
- Robust mixture prior (Neuenschwander et al 2010)
- Compromise decision (Calderazzo et al 2024)
- …

One-arm trial with Gaussian endpoint

Example setup

$H_0: \theta \leq 0$ vs. $H_1: \theta > 0$, $\alpha = 0.025$

- Current data $D \sim \mathcal{N}\left(\theta, {}^{1}/_{\sqrt{n}}\right)$, $n = 25$
- External data $D_E \sim \mathcal{N}\left(\theta_E, {}^{1}/_{\sqrt{n_E}}\right)$, $n_E = 20$.
  Considered fixed with value d_E, e.g., $d_E = 1$

- Evaluate T1E for $\theta = 0$
- Evaluate power for $\theta = 0.5$

Empirical Bayes Power Prior (Gravestock, Held et al. 2017)

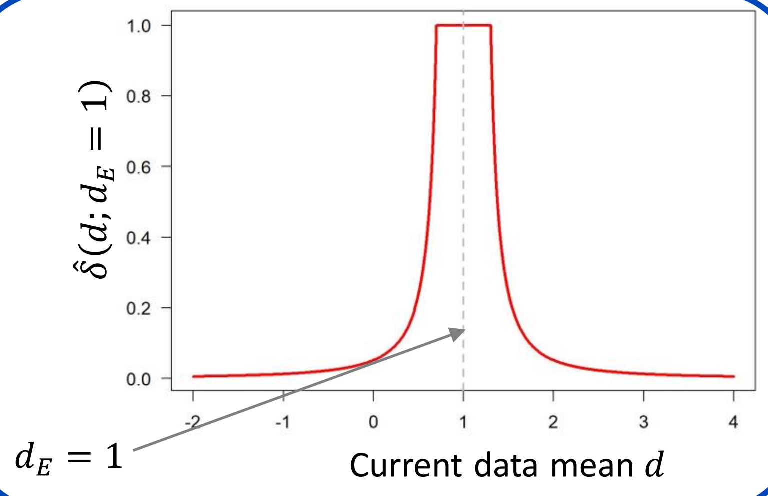
- Use Power Prior approach

$$\pi_{EB}(\theta) = \pi(\theta | d_E, \delta) \propto L(\theta; d_E)^{\delta} \pi(\theta)$$

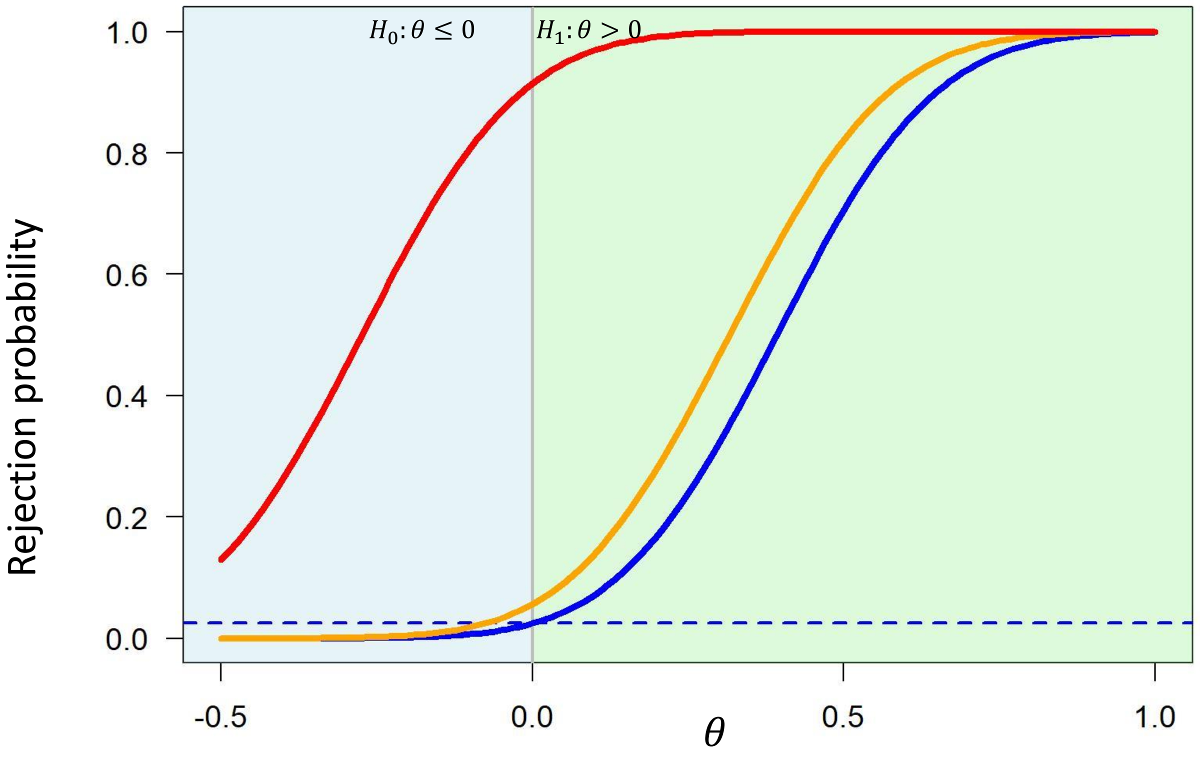
$\delta = 0$: no borrowing; prior for current trial $= \pi(\theta)$

$\delta = 1$: full borrowing; prior for current trial = posterior given external data

- Adapt $\delta = \delta(d; d_E)$ such that information is only borrowed for similar data.
- Use Empirical Bayes approach for estimating $\hat{\delta}(d; d_E)$:



Empirical Bayes Power Prior



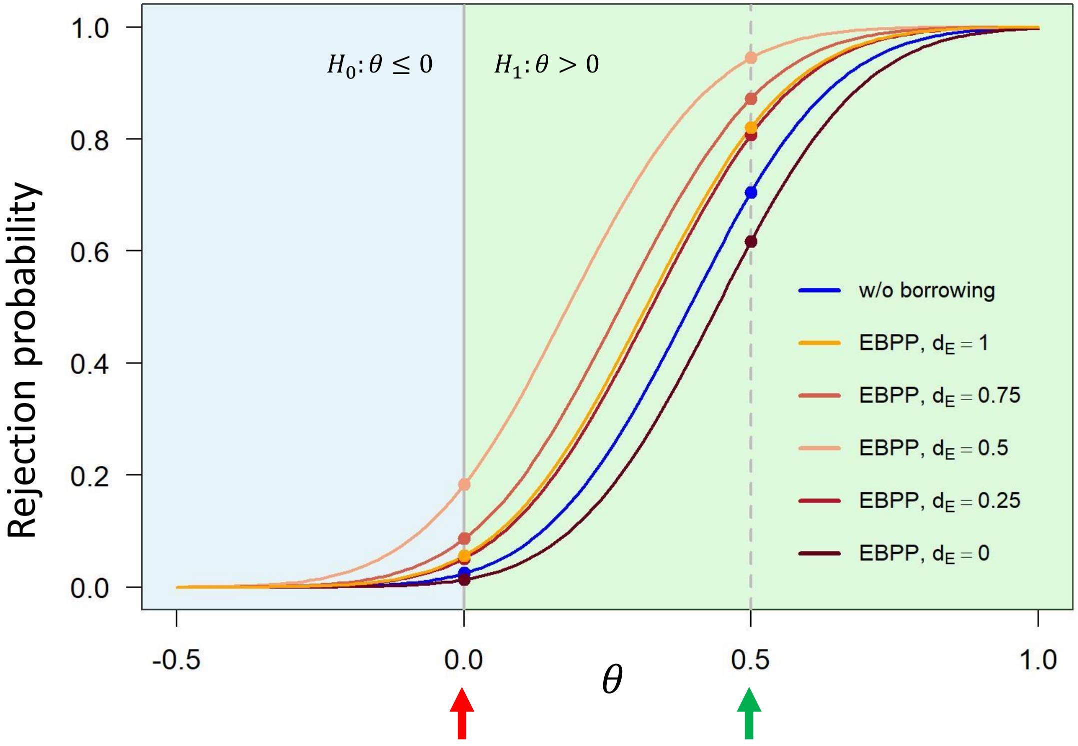
Test w/ full borrowing, $d_E = 1$

Test w/ EBPP borrowing, $d_E = 1$

Test w/o borrowing

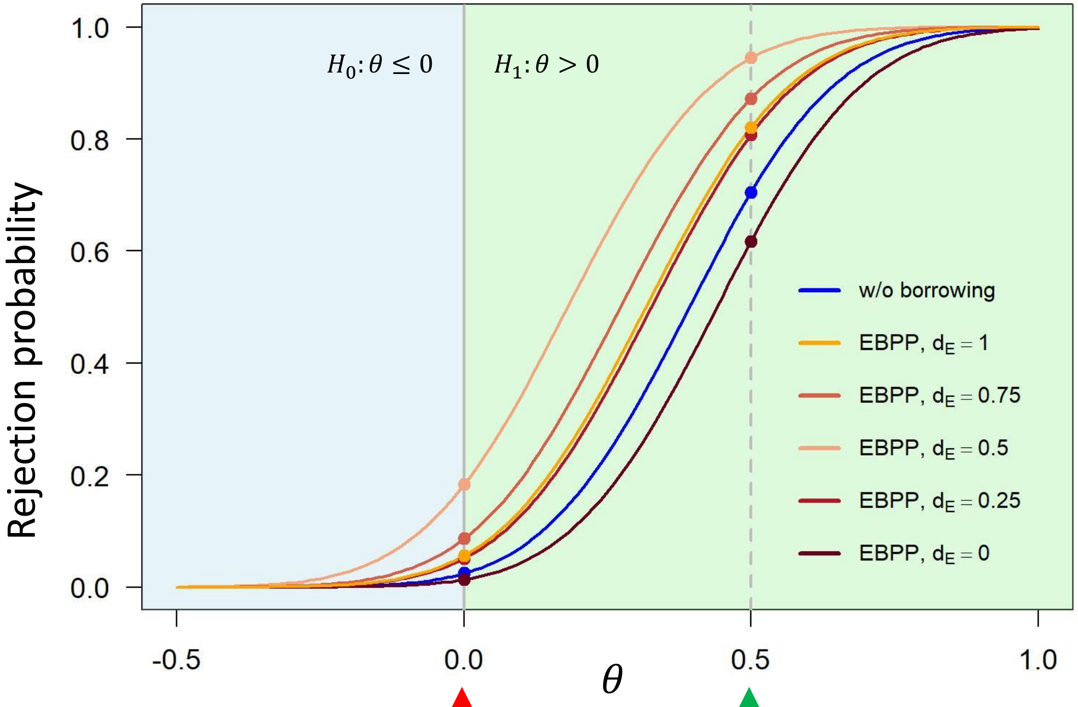
α

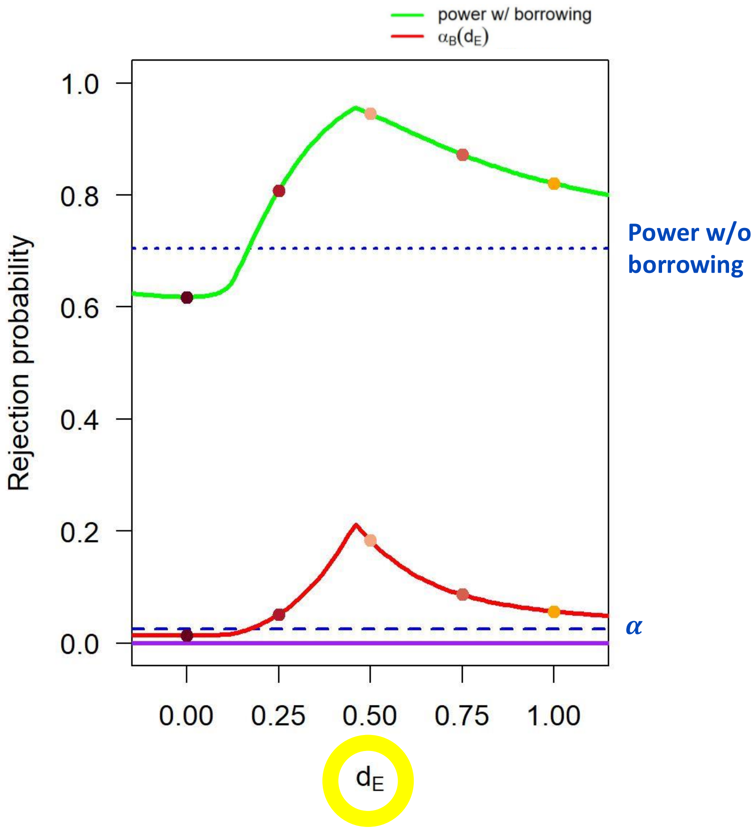
Empirical Bayes Power Prior: varying external data mean d_E



$\alpha_B(d_E)$: T1E rate **Power_B at $\theta = 0.5$**

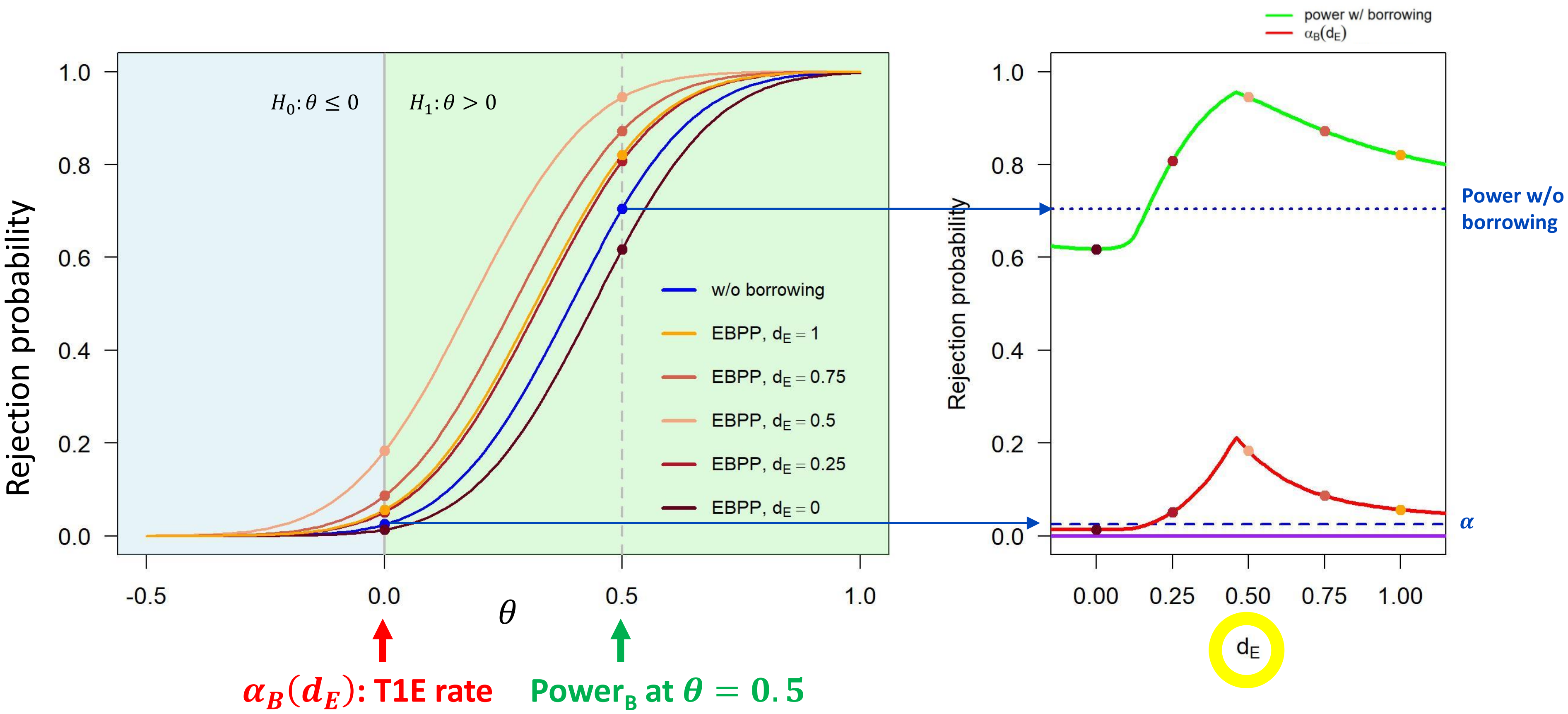
Empirical Bayes Power Prior: varying external data mean d_E





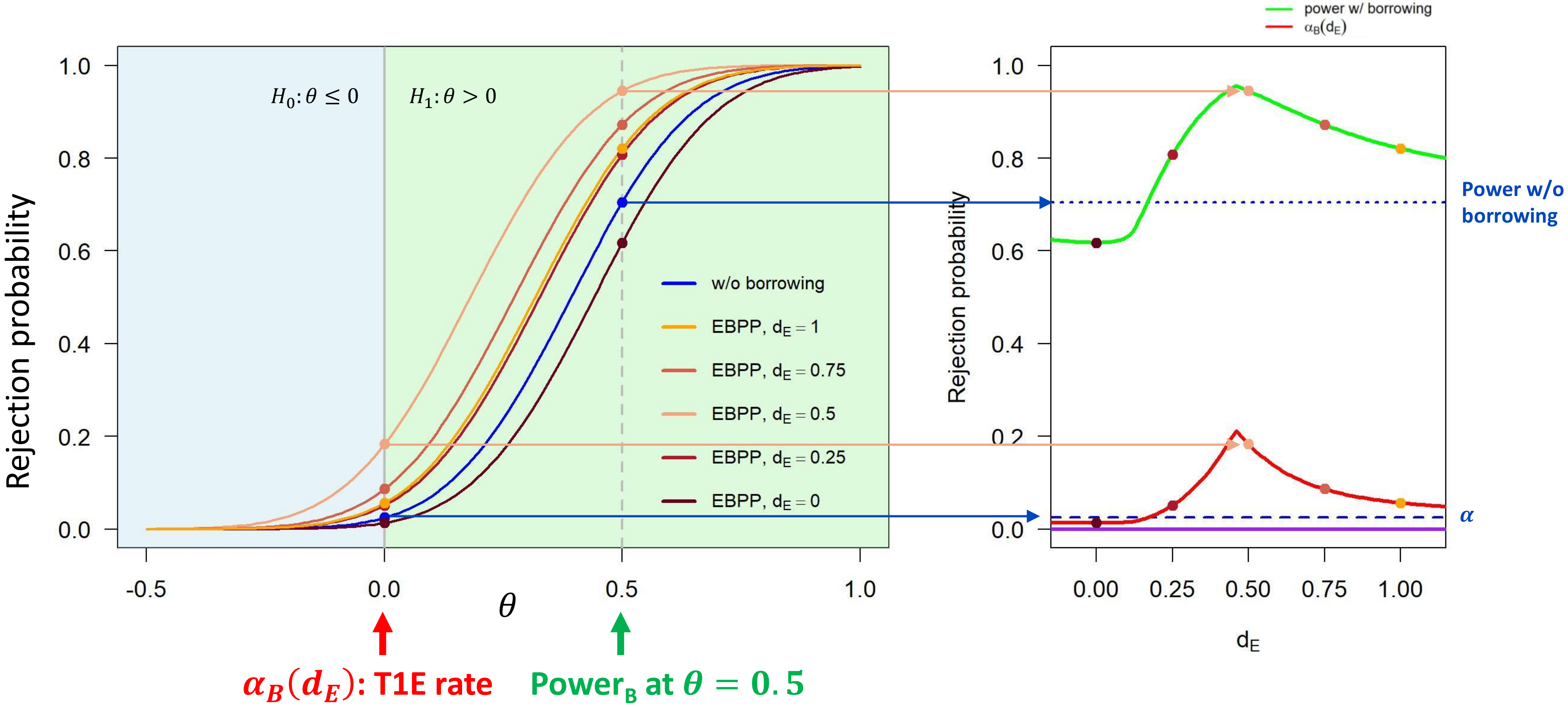
$\alpha_B(d_E)$: T1E rate Power_B at $\theta = 0.5$

Empirical Bayes Power Prior: varying external data mean d_E

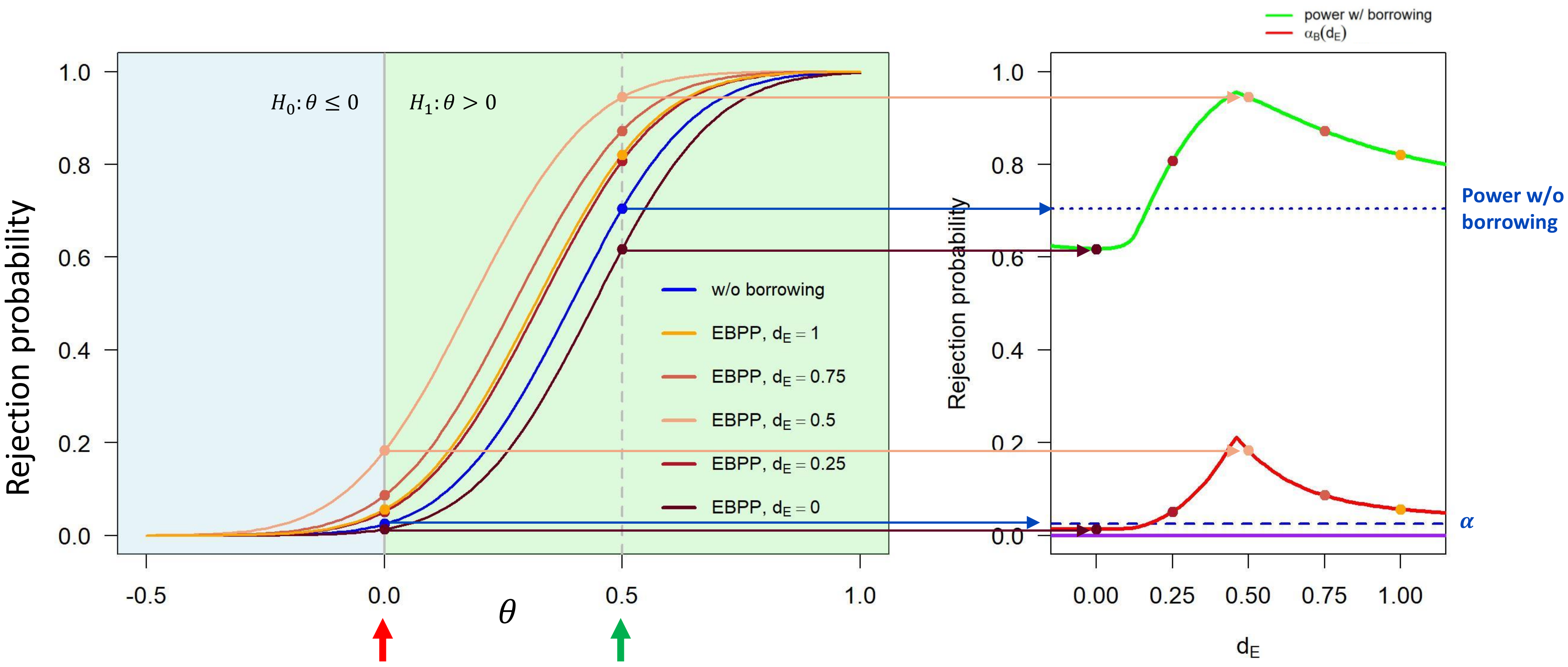


dkfz.

Empirical Bayes Power Prior: varying external data mean d_E

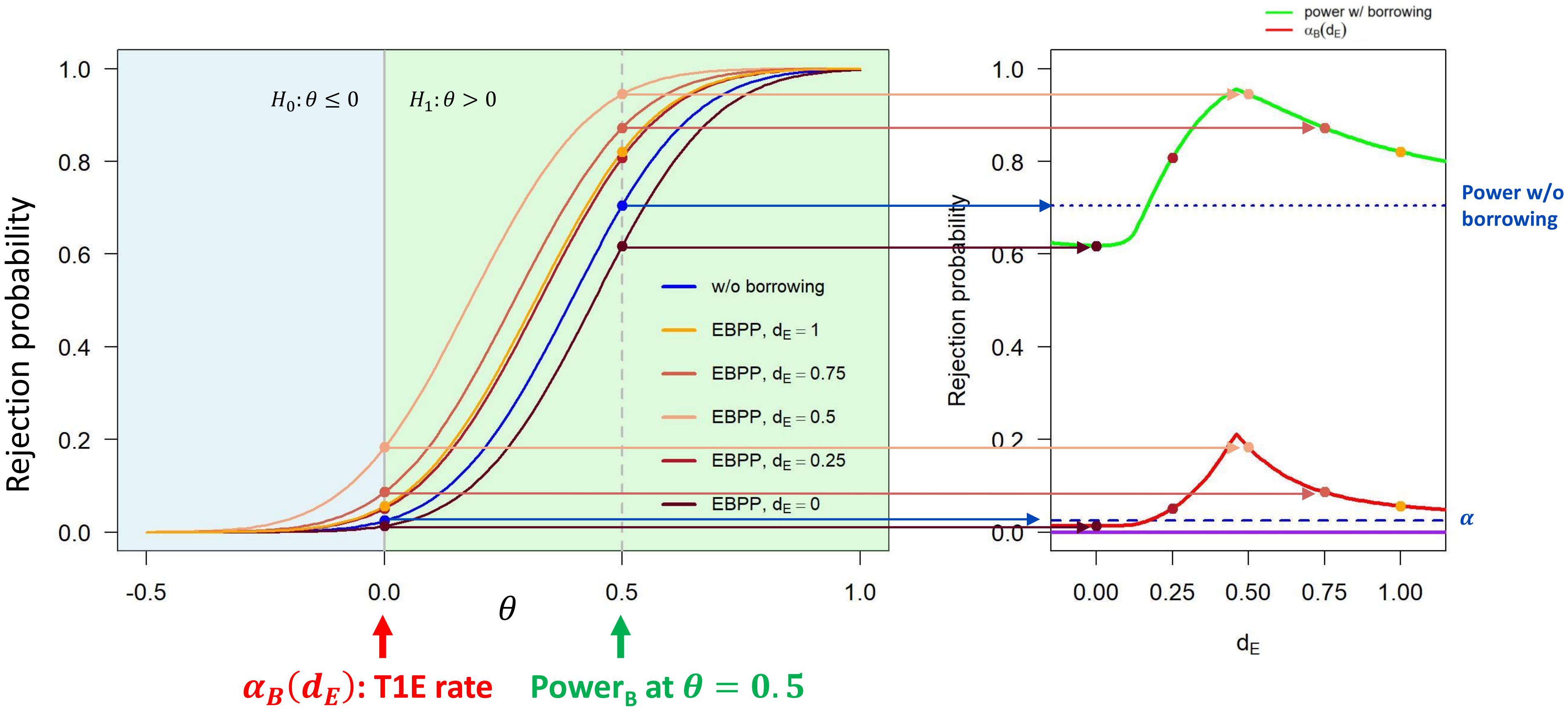


Empirical Bayes Power Prior: varying external data mean d_E

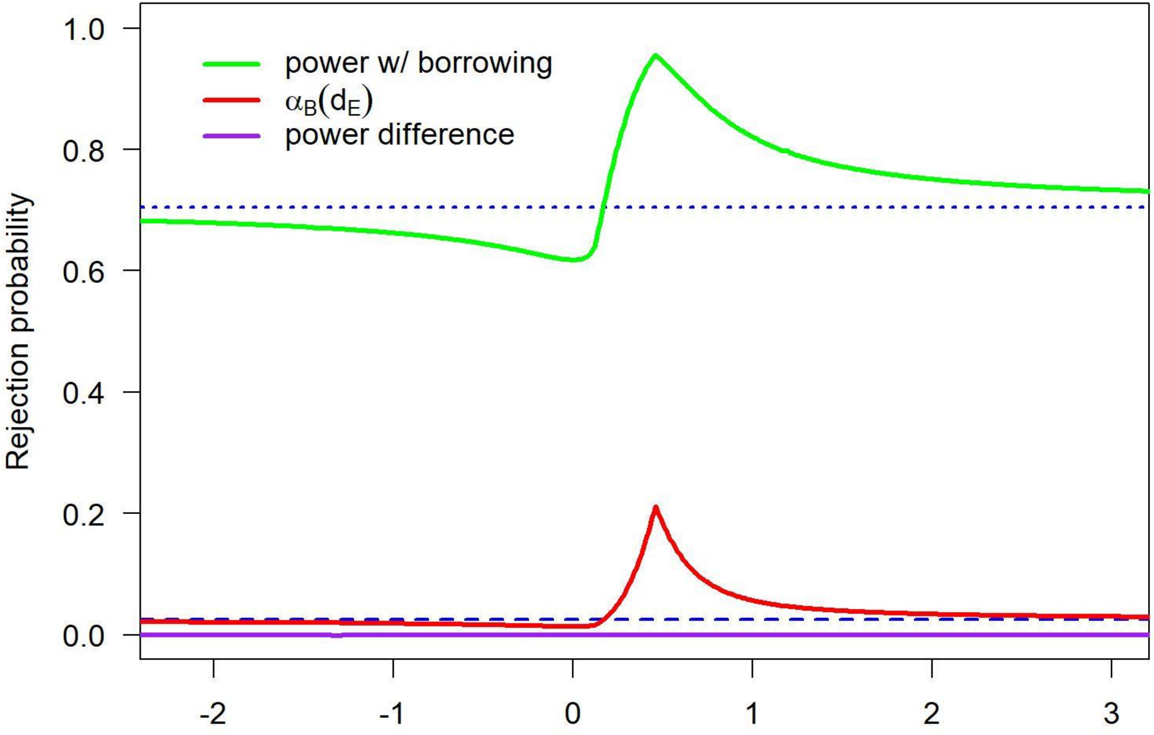


$\alpha_B(d_E)$: **T1E rate** **Power$_B$ at $\theta = 0.5$**

Empirical Bayes Power Prior: varying external data mean d_E



Empirical Bayes Power Prior: Frequentist OCs



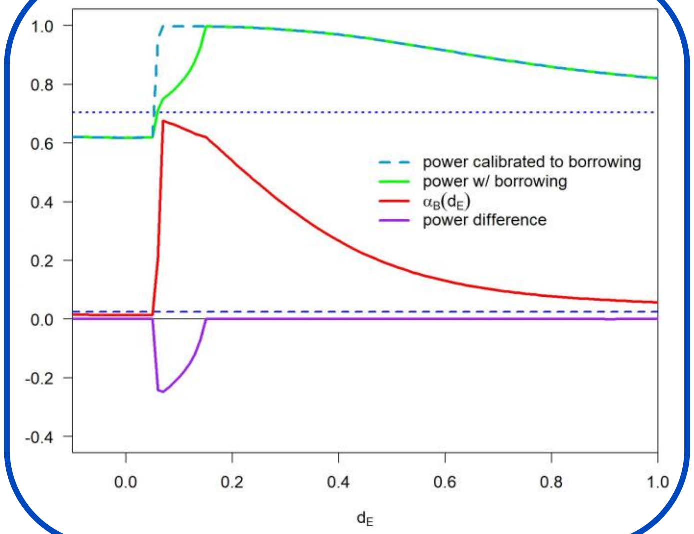
Power w/o borrowing

α

External data mean d_E

Empirical Bayes Power Prior: Properties

- Borrowing by modeling a prior that incorporates external information.
- Adapts to prior data conflict.
- Intuitive and easily interpreted.
- Easy to use: no choices to be made.
- T1E (α_B): function of external data mean d_E and α.
- But: Can be coerced to result in test inferior to UMP test (→ power loss) in (unrealistic) situation when borrowing from extremely large external data set ($n_E = 1000$).



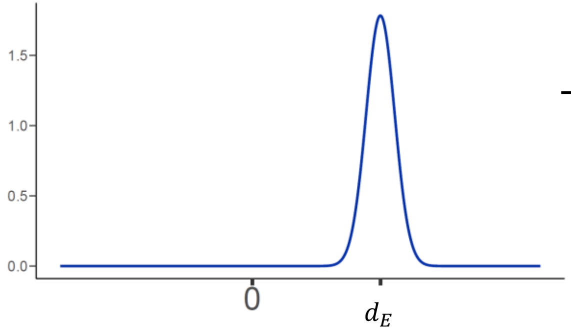
Robust Mixture Prior (Neuenschwander et al 2010)

For borrowing use prior: $\pi_{\text{mix}}(\theta) = w \cdot \pi_{\text{external}}(\theta) + (1 - w) \cdot \pi_{\text{robust}}(\theta)$, $w \in [0,1]$

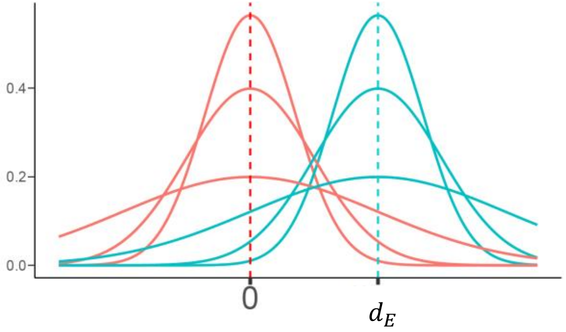
Informative component (External data)
$\pi_{\text{external}}(\theta)$

Robust component
$\pi_{\text{robust}}(\theta)$

$w \cdot$



$+ (1 - w) \cdot$



How to choose w, location and variance of π_{robust}?

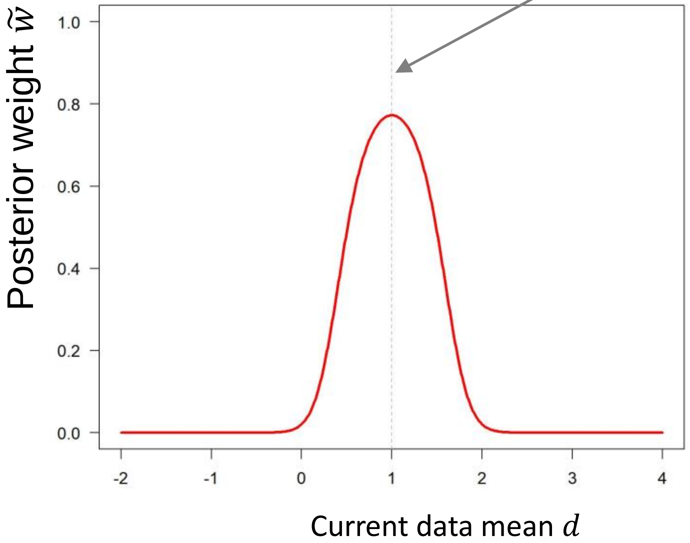
Robust Mixture Prior: Exemplary choices

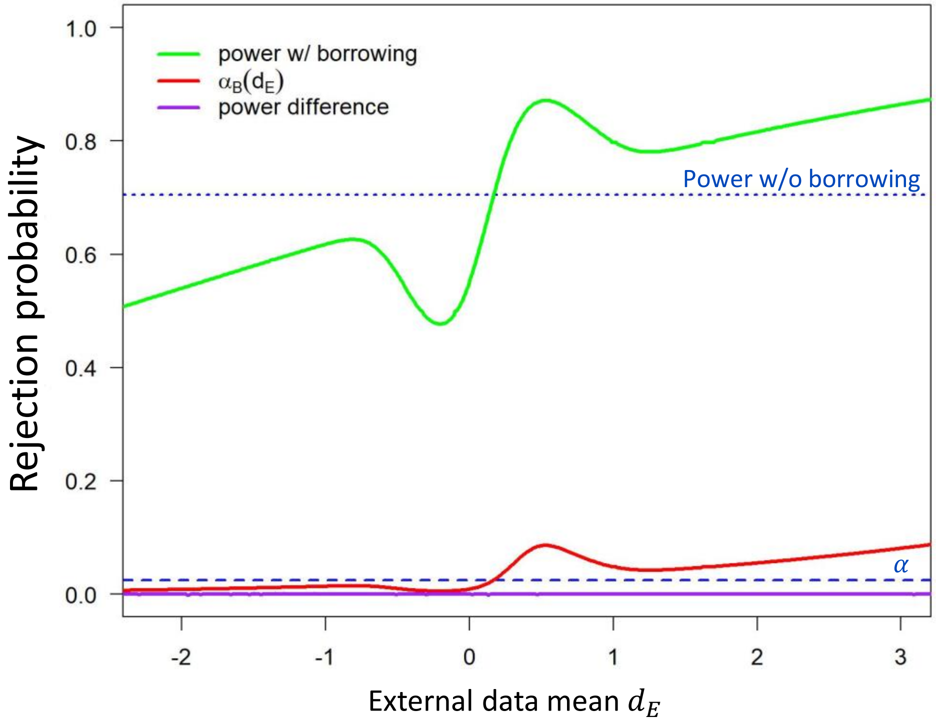
Informative component: $\pi_{\text{external}}(\theta) \sim \mathcal{N}\left(d_E, {}^1/_{\sqrt{n_E}}\right)$,

Robust component: $\pi_{\text{robust}}(\theta) \sim \mathcal{N}(d_E, 1)$ (located at external data mean, "unit information")

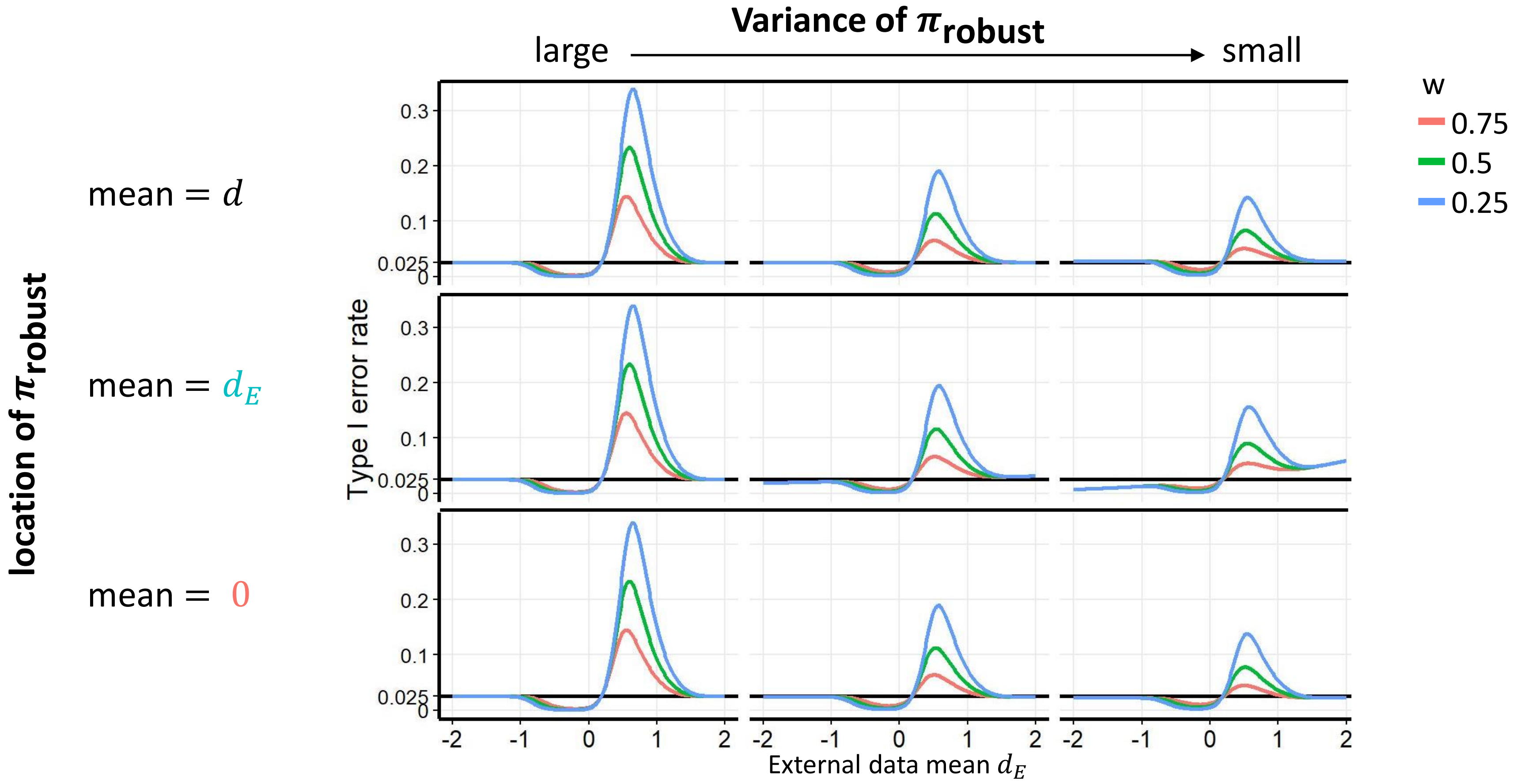
Weight: $w = 0.5$

Posterior weight $\tilde{w}$ for varying current data mean d and external data mean $d_E = 1$:





Robust Mixture Prior: Selecting parameters



Robust Mixture Prior: Properties

- Borrowing by modeling a prior that incorporates external information.
- Adapts to prior data conflict by adjusting posterior weight $\widetilde{w}$ to similarity of current data and informative component.
- Popular borrowing method.
- Requires choices of mixture prior weight w as well as location and variance of robust prior π_{robust}.
- Interpretation not straightforward: how much external information is borrowed?
- T1E (α_B): function of external data d_E, parameter choices of mixture weight and robust prior, α.

Compromise Decision (Calderazzo et al. 2024)

- Revisit:
  Bayesian decision using „non-informative prior“ $\equiv$ Frequentist decision:
  „reject H_0 if $P(H_1 |$ current data, non-informative prior$) > 1 - \alpha$“ has T1E = α.

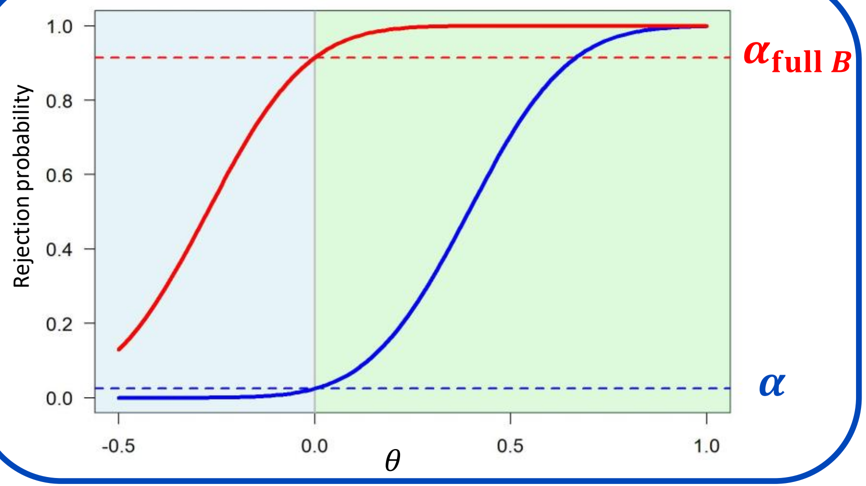
Compromise Decision (Calderazzo et al. 2024)

- <u>Revisit:</u>
  Bayesian decision using „non-informative prior" $\equiv$ Frequentist decision:
  „reject H_0 if $P(H_1 |$ current data, non-informative prior$) > 1-\alpha$" has T1E = α.

- With borrowing from external data by fully incorporating information in prior:

  Bayesian decision
  $P(H_1 | d$, full informative prior$) > 1-\alpha$
  corresponds to frequentist decision
  with T1E rate $= \alpha_{\text{full } B}$:

  $P(H_1 | d$, full informative prior$) > 1-\alpha$
  $\Leftrightarrow$
  $P(H_1 | d$, non-informative prior$) > 1-\alpha_{\text{full } B}$

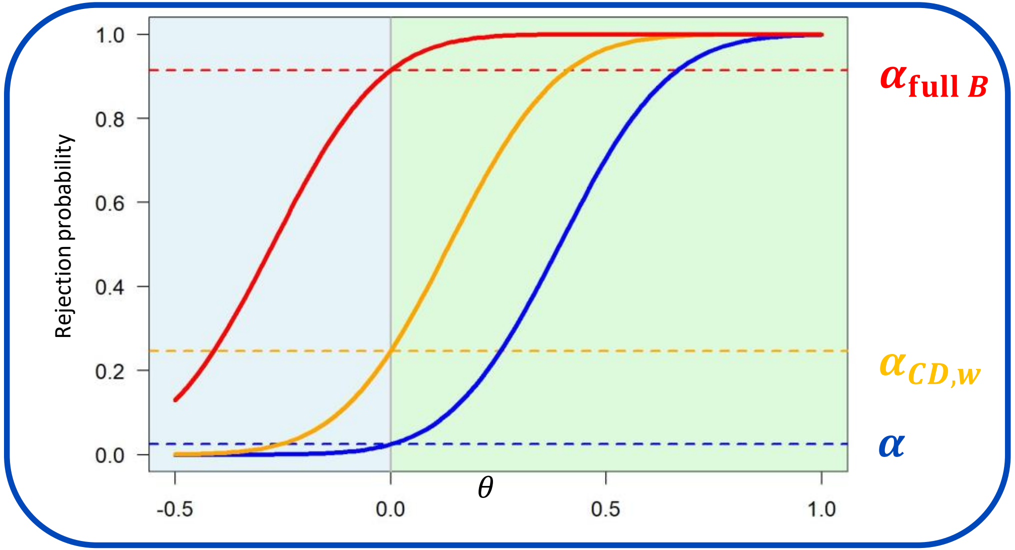


Compromise Decision (Calderazzo et al. 2024)

Compromise between w/o and w/ full borrowing:

$\alpha_{CD,w} = (1 - w) \cdot \alpha + w \cdot \alpha_{\text{full } B}, w \in [0,1]$

Here: $w = 0.25 \rightarrow \alpha_{CD,w} = 0.247$

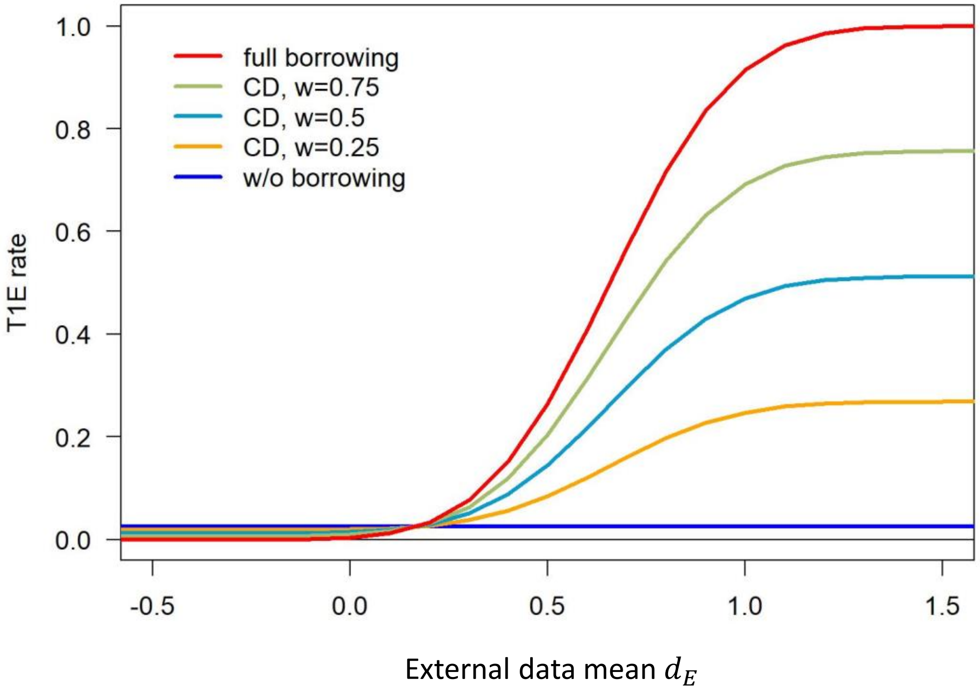


Linearly relates amount of borrowing (w) and T1E inflation.

Compromise Decision (Calderazzo et al. 2024)

T1E rate for varying external data mean d_E

$\alpha_{CD,w} = (1 - w) \cdot \alpha + w \cdot \alpha_{\text{full } B}$



Compromise Decision: Properties

- Targets the test decision instead of modeling the prior distribution.
- Linearly relates T1E inflation to amount of borrowing, i.e., interpretation directly related to T1E inflation.
- Requires choice of w.
- Extension: T1E inflation can be bounded.
- Dynamic version can be defined that uses data-dependent adaptive approach to estimate w ($\rightarrow$ no choice of w required).

Two-arm testing with borrowing to control arm: „hybrid control trial“

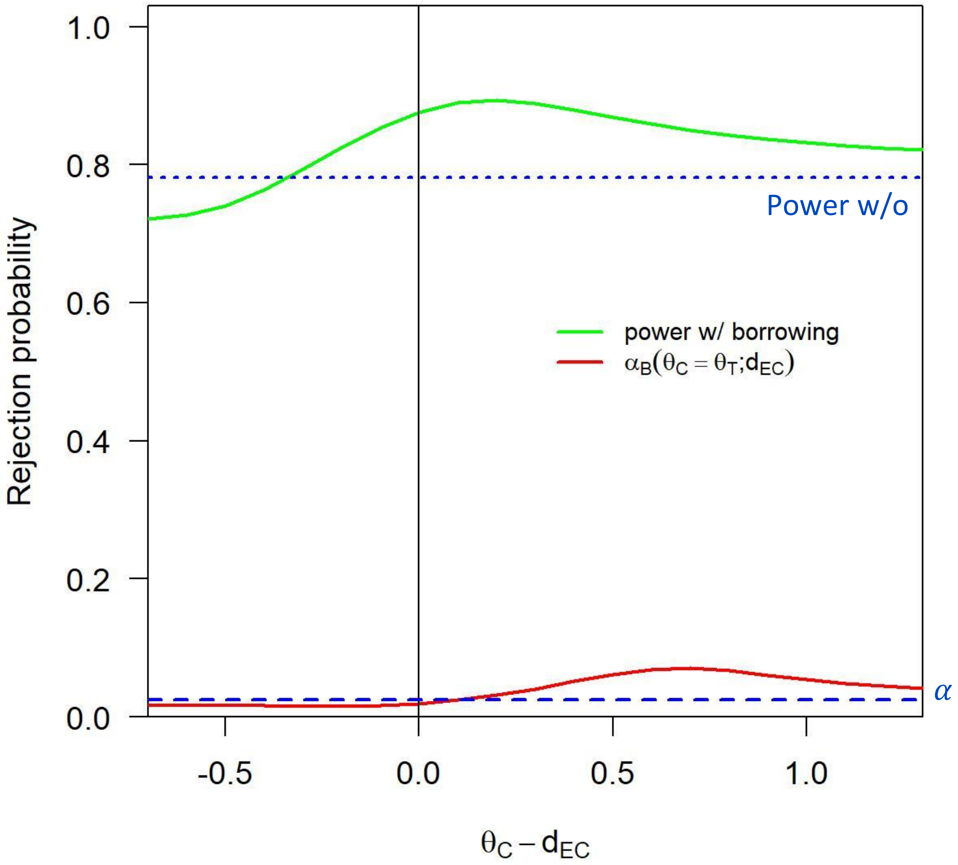
Hybrid control arm trial

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

- Treatment data $D_T \sim \mathcal{N}\left(\theta_T, {}^1/_{\sqrt{n_T}}\right), n_T = 15$
- Control data $D_C \sim \mathcal{N}\left(\theta_C, {}^1/_{\sqrt{n_C}}\right), n_C = 15$
- External control data $D_{EC} \sim \mathcal{N}\left(\theta_{EC}, {}^1/_{\sqrt{n_{EC}}}\right), n_{EC} = 10.$
  Considered fixed with value d_{EC}.

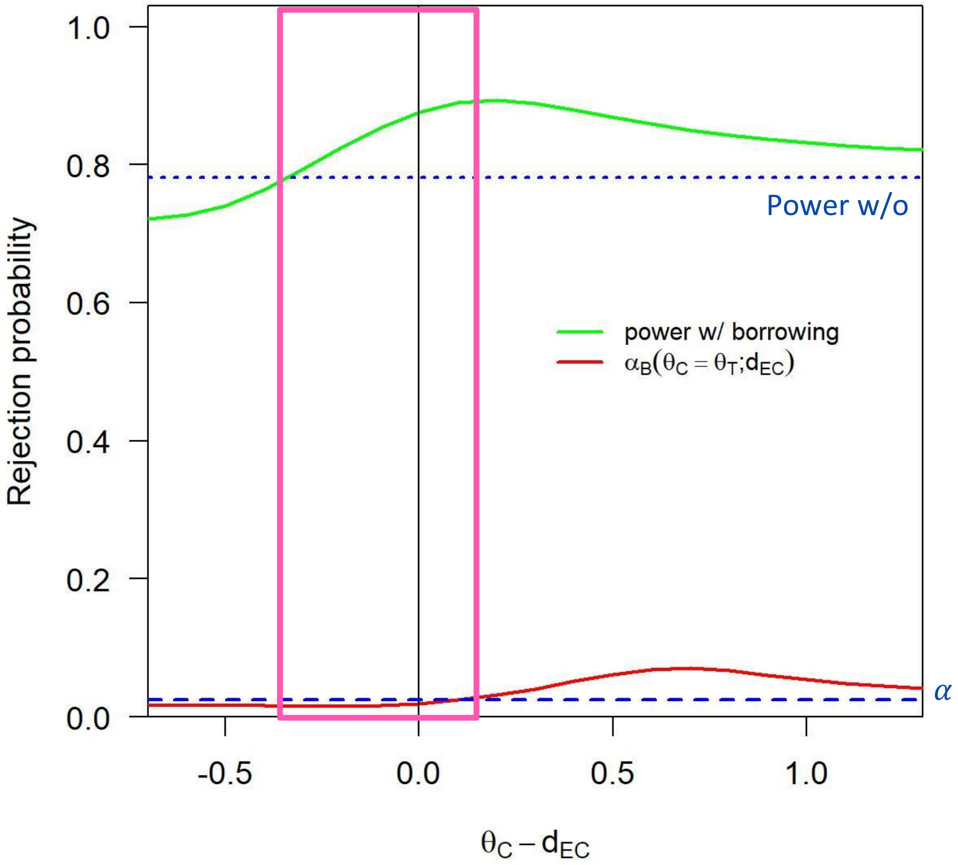
- T1E obtained for $\theta_T - \theta_C = 0$
- Power evaluated at $\theta_T - \theta_C = 1$

Hybrid control arm: EB Power Prior



dkfz.

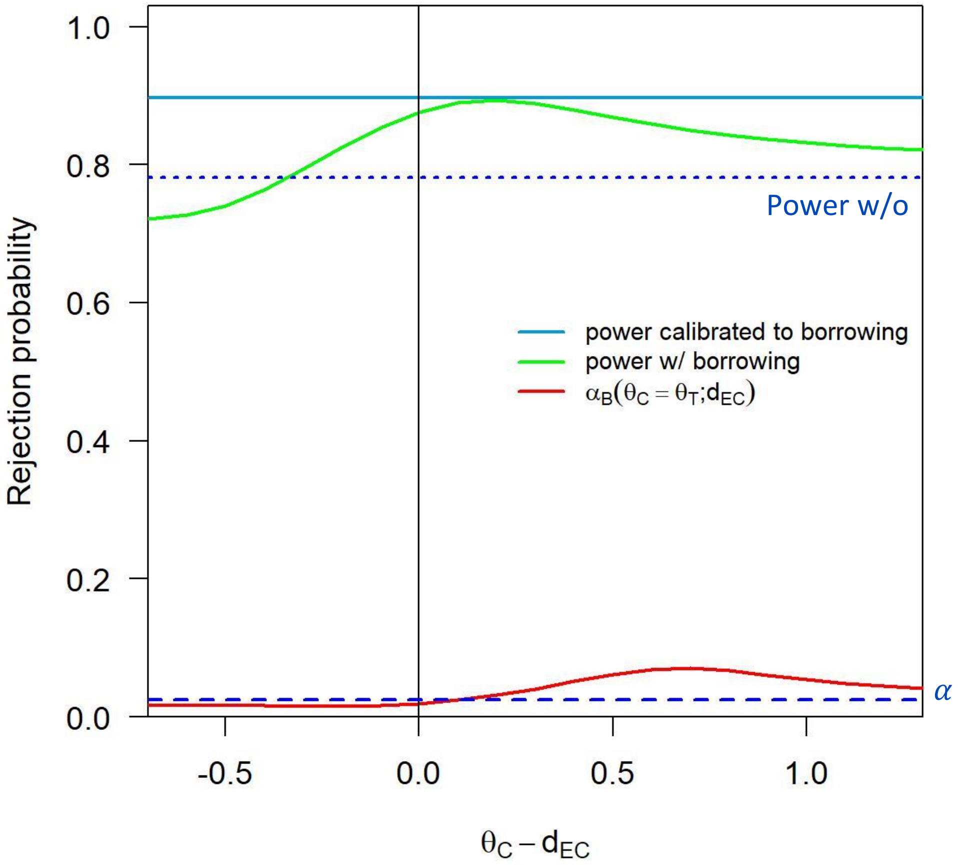
Hybrid control arm: EB Power Prior



„Sweet spot“:

(No T1E inflation) AND (power gain)

Hybrid control arm: EB Power Prior



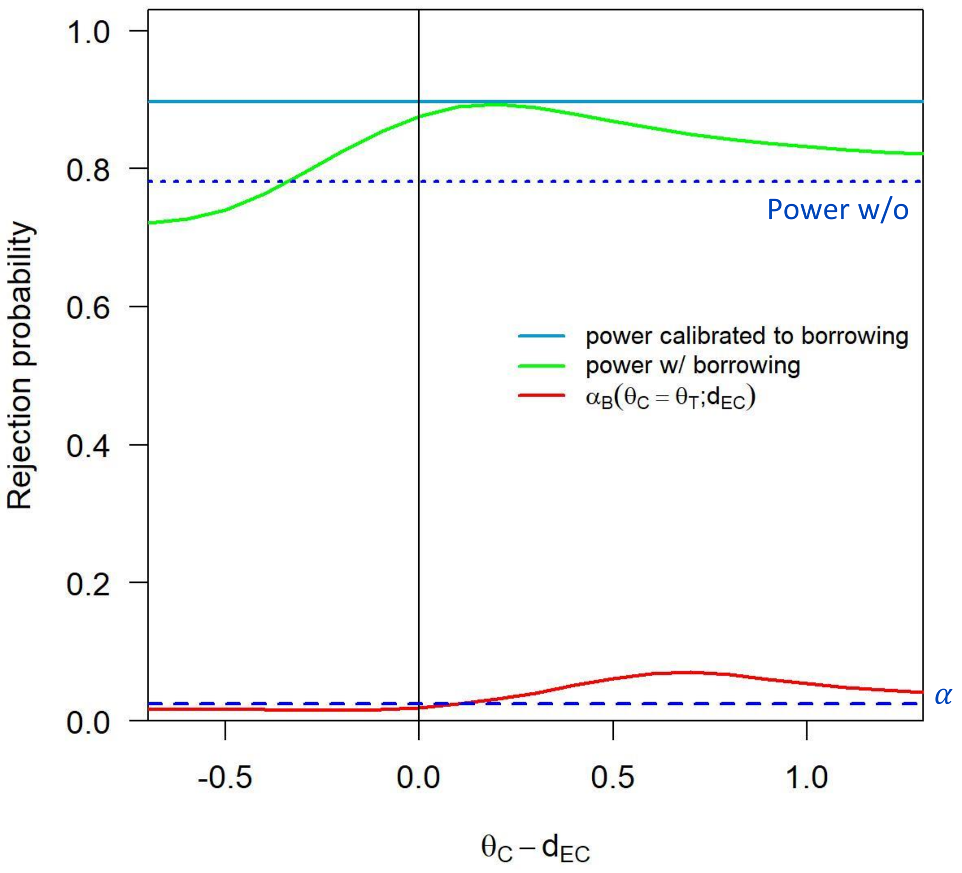
- $\alpha_B(d_{EC})$ varies with θ_C $(= \theta_T)$
- Since θ_C is unknown:

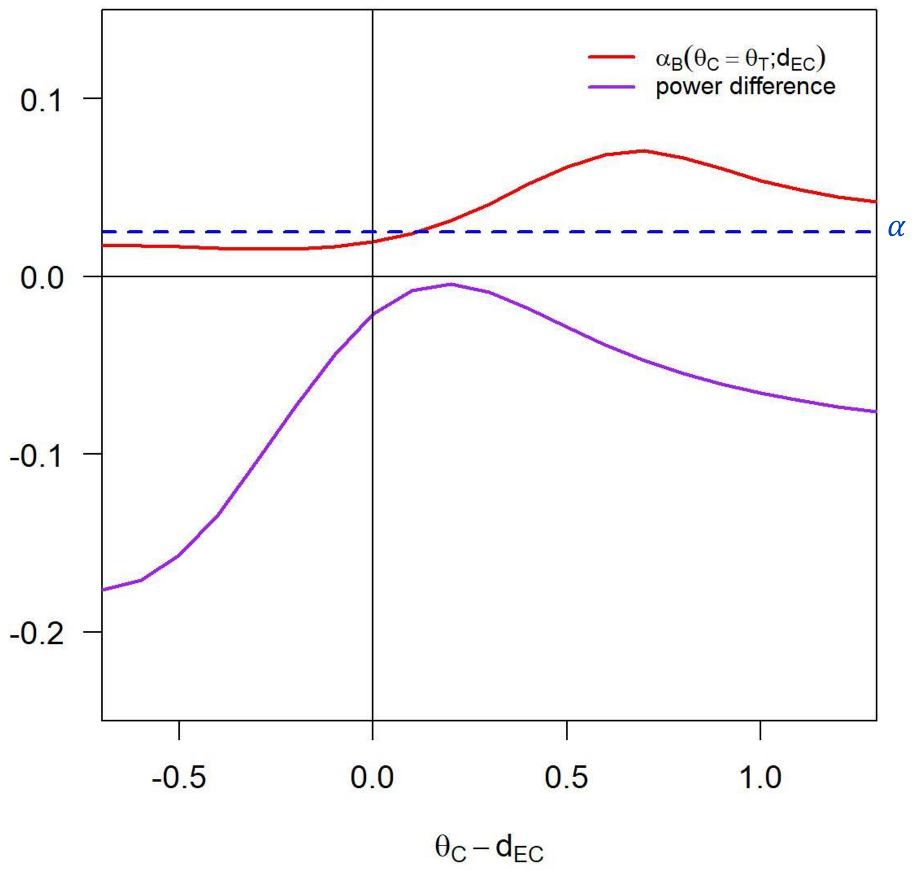
  need to calibrate test to

  $$\max_{\theta_C} \alpha_B(\theta_C = \theta_T; d_{EC}) = 0.071$$

  → **Power (at $\theta_T - \theta_C = 1$) of test calibrated to borrowing $= 0.90$**

Hybrid control arm: EB Power Prior





Hybrid control arm: EB Power Prior

- If we have no trust in similarity of external control and control data:
  need to calibrate test to $\max_{\theta_C} \alpha_B(\theta_C = \theta_T; d_{EC})$, i.e. worst case for all $\theta_C = \theta_T$.

Hybrid control arm: EB Power Prior

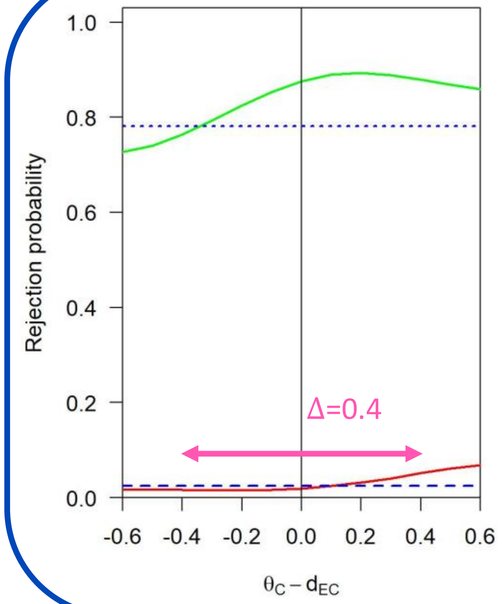
- If we have no trust in similarity of external control and control data: need to calibrate test to $\max_{\theta_C} \alpha_B(\theta_C = \theta_T; d_{EC})$, i.e. worst case for all $\theta_C = \theta_T$.

- If we trust that the maximal size of conflict is restricted by Δ:

  calibrate test to

  $\max_{\theta_C} \alpha_B(\theta_C = \theta_T; d_{EC})$

  for $|\theta_C - d_{EC}| \leq \Delta$



Δ	max α_B	max power gain
0.1	0.024	0.080
0.2	0.031	0.053
0.3	0.041	0.026
0.4	0.052	0.009
0.5	0.062	-0.001

Conclusions

- Increasing interest in using Bayesian methods for design and analysis of early clinical trials.
- Bayesian methods are natural framework for incorporation of external/historical information.
- (Adaptive) Bayesian borrowing approaches by
  - modeling the prior for the current trial

  or by
  - targeting the test decision.
- In frequentist sense: no power gains possible when T1E should be controlled.
- But: frequentist T1E is determined under worst case scenario.
- If prior information is reliable and consistent with new information, frequentists OC of the trial can be improved, e.g., if the maximal size of conflict can be trusted to be restricted.
- Cave: if borrowing from many more external data than current data, information of external data may overrule current data.

Acknowledgements

Dr. Silvia Calderazzo

Dr. Manuel Wiesenfarth

Vivienn Weru

Division of Biostatistics, DKFZ Heidelberg

References

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

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

Ibrahim JG, Chen M-H (2000) Power prior distributions for regression models. *Statistical Science* 15: 46-60.

Kopp-Schneider A, Calderazzo S, 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.

Kopp-Schneider A, Wiesenfarth M, Held L, Calderazzo S, (2024) Simulating and reporting frequentist operating characteristics of clinical trials that borrow external information: Towards a fair comparison in case of one-arm and hybrid control two-arm trials. *Pharmaceutical Statistics* 23(1): 4-19.

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

Pocock SJ (1976). The combination of randomized and historical controls in clinical trials. *J Chron Dis* 29: 175-188.

Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N....Thompson L. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics* 13(1): 41–54.