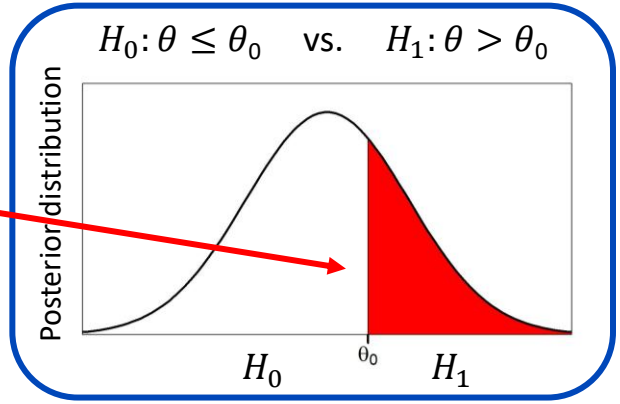
Bayesian borrowing in clinical trial test decisions: Frequentist type I error rate and power

Annette Kopp-Schneider
Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany

Joint work with Silvia Calderazzo, Vivienn Weru (Biostatistics, DKFZ),
Manuel Wiesenfarth (Cogitars)

Hypothesis testing with Bayesian methods

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



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

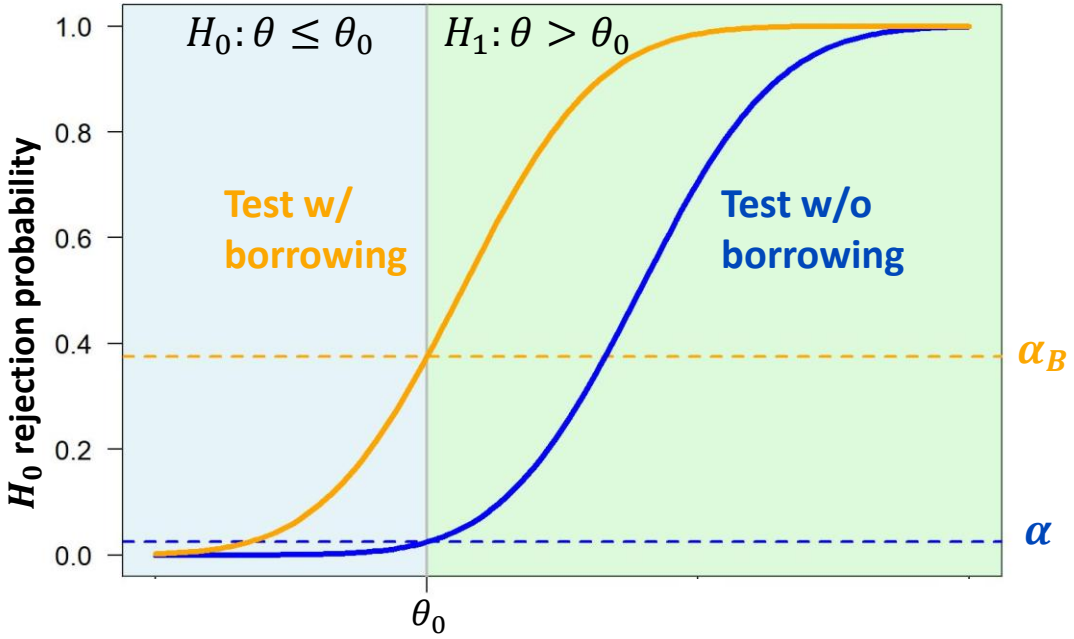
- Borrowing from external data by incorporating information into the prior.

- {current data such that P(H_1 | ...) $> 1 - \alpha$ } $\equiv$ rejection region based on current data.

Does borrowing increase power?

Problem

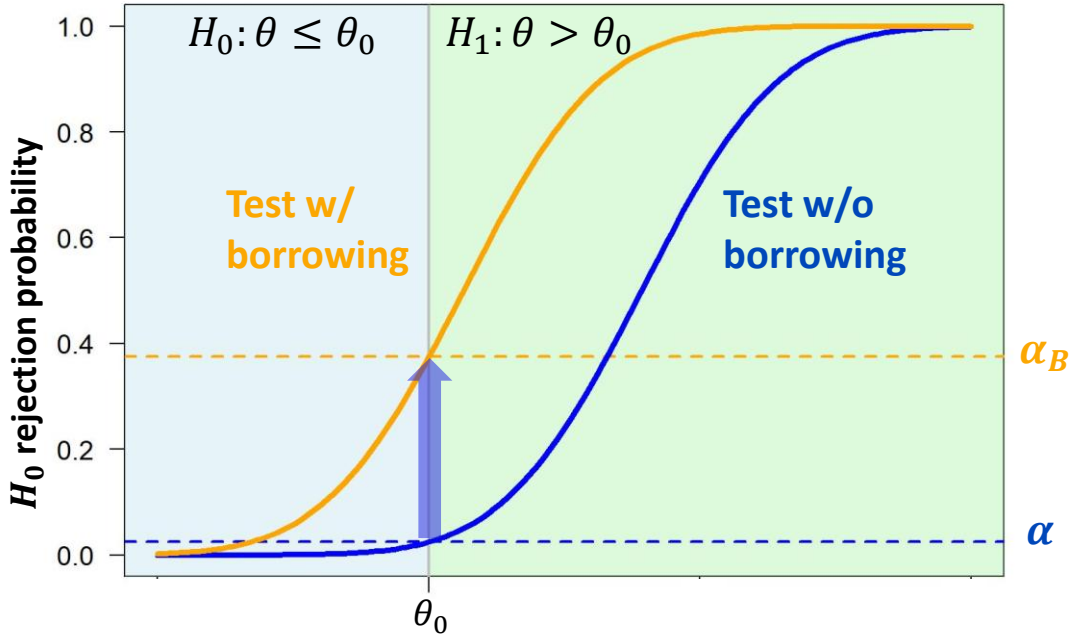
Fair comparison of
Operating Characteristics (OC)
w/ and **w/o** borrowing?



Does borrowing increase power?

Problem

Fair comparison of
Operating Characteristics (OC)
w/ and **w/o** borrowing?



Solution

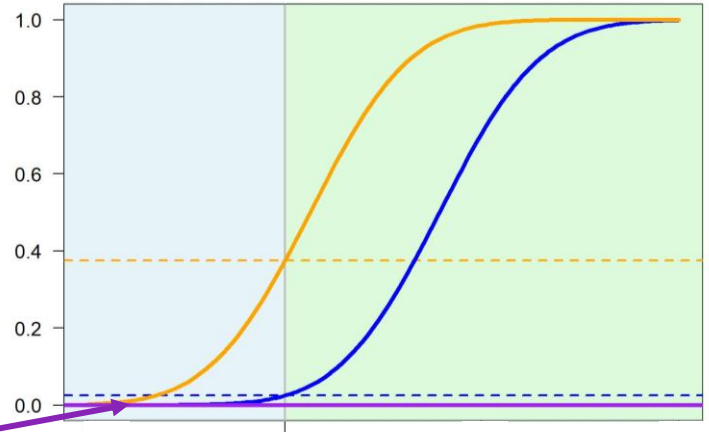
„**test calibrated to borrowing**“ = test w/o borrowing, but T1E set to α_B instead of α

→ 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 frequentist 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.

- True irrespective of borrowing approach!

(AKS et al. 2020)

Hybrid control arm trial:
Adaptive borrowing of external control data to current control data

Set-up

- Gaussian endpoint, $H_0: \theta_T - \theta_C \leq 0$ vs. $H_1: \theta_T - \theta_C > 0$
- Frequentist T1E $= \alpha = 0.025$, evaluated at $\theta_T - \theta_C = 0$; power evaluated at $\theta_T - \theta_C = 1$.

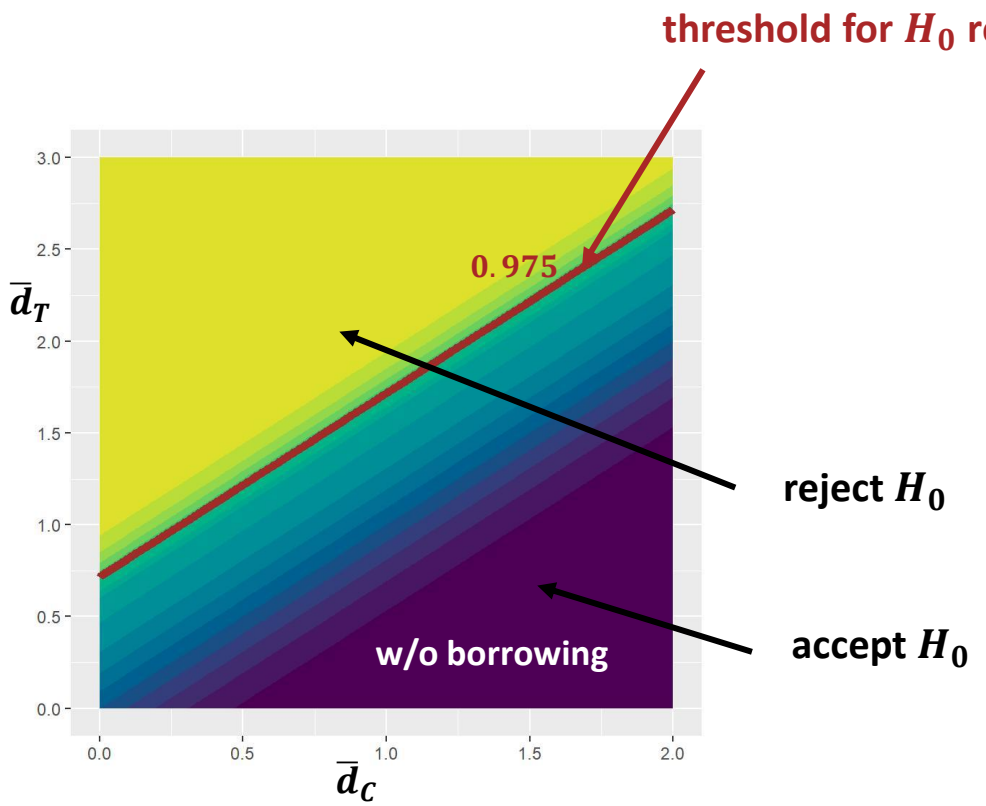
Available information

- Current control mean $\bar{\boldsymbol{d}}_{\boldsymbol{C}}$ and treatment mean $\bar{\boldsymbol{d}}_{\boldsymbol{T}}$ (with expectation θ_C and θ_T, variance known).
- External control data mean $\bar{\boldsymbol{d}}_{\boldsymbol{EC}}$.

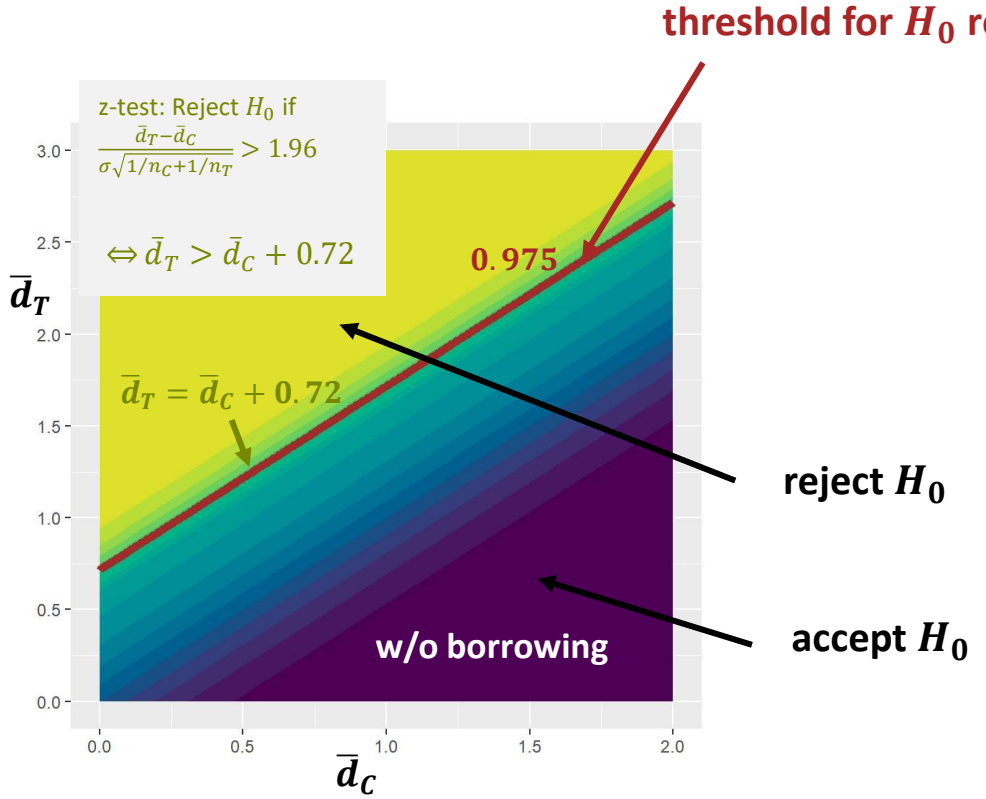
Challenge

- Potential problem: Heterogeneity between $\bar{d}_{EC}$ and θ_C (aka prior-data conflict).
- Solution: Use adaptive borrowing approach.

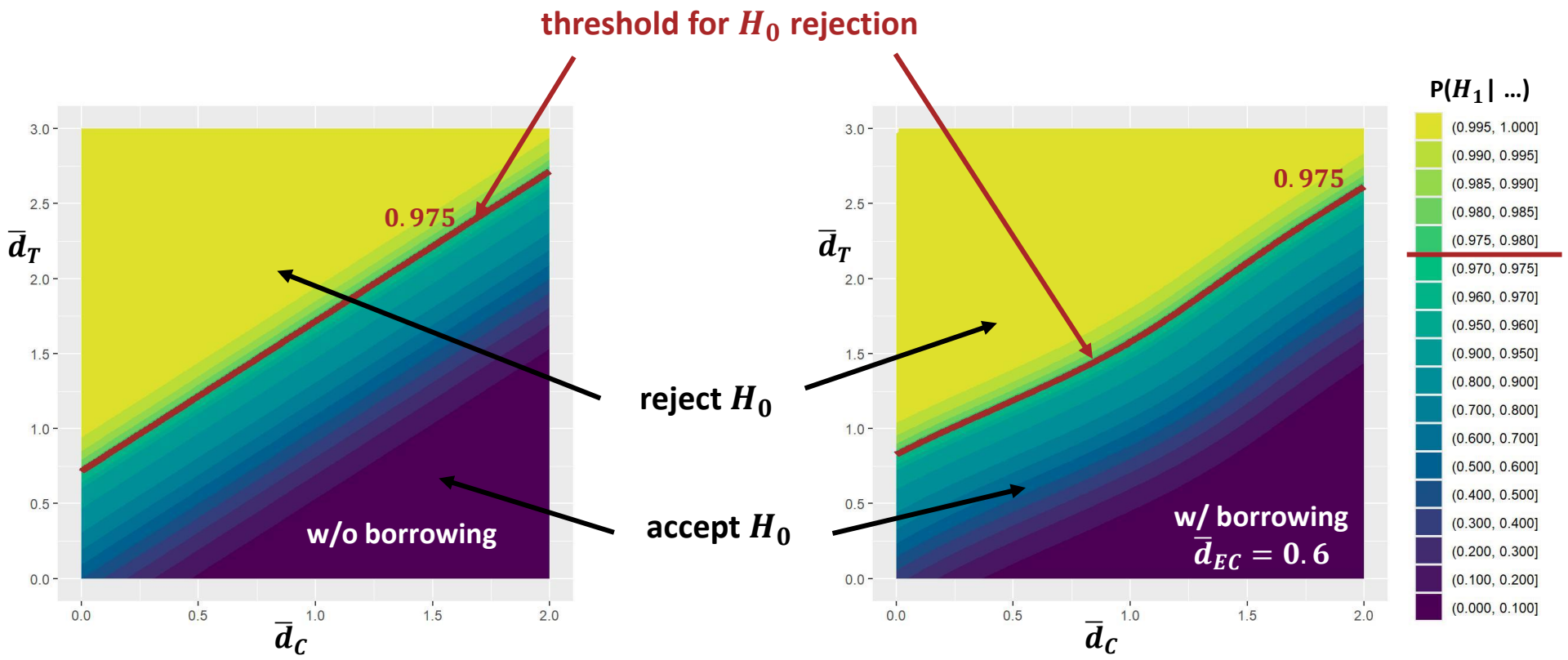
Hypothesis testing: $P(H_1 \mid \ldots) > 1 - \alpha \Leftrightarrow$ Decision based on $\bar{d}_C$ and $\bar{d}_T$



Hypothesis testing: $P(H_1 \mid \ldots) > 1 - \alpha \Leftrightarrow$ Decision based on $\bar{d}_C$ and $\bar{d}_T$

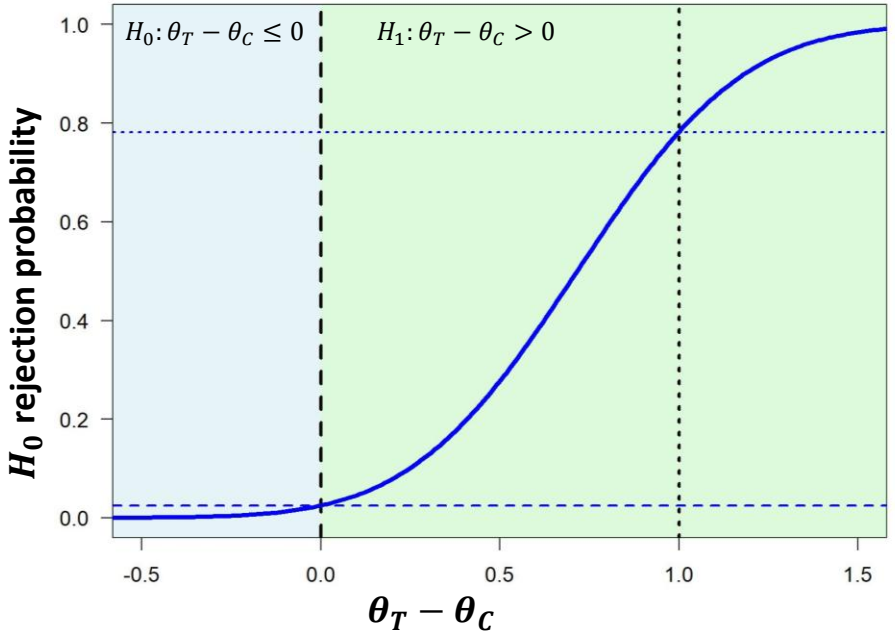


Hypothesis testing: $P(H_1 \mid \ldots) > 1 - \alpha \Leftrightarrow$ Decision based on $(\bar{d}_C, \bar{d}_T, \bar{d}_{EC})$



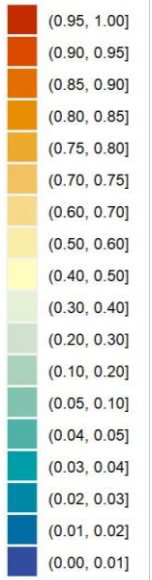
w/o borrowing: H_0 rejection probability (aka „power curve“)

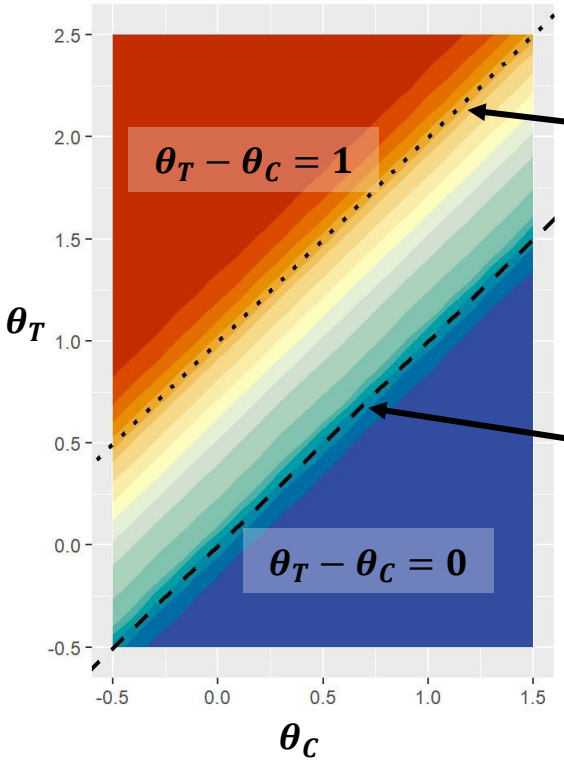
H_0 rejection probability as function of $\theta_T - \theta_C$:



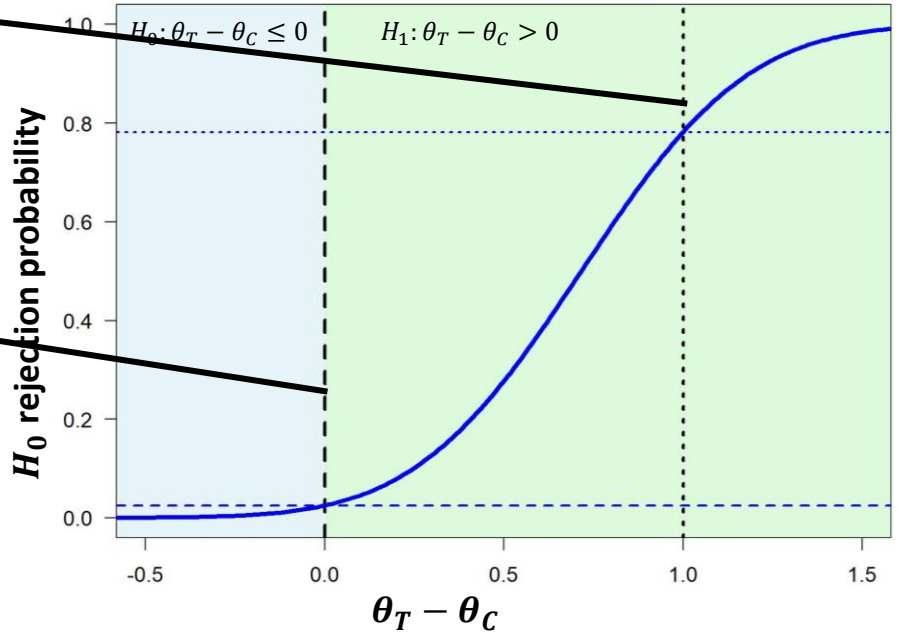
w/o borrowing: H_0 rejection probability

H_0 rejection probability as color in contour plot:

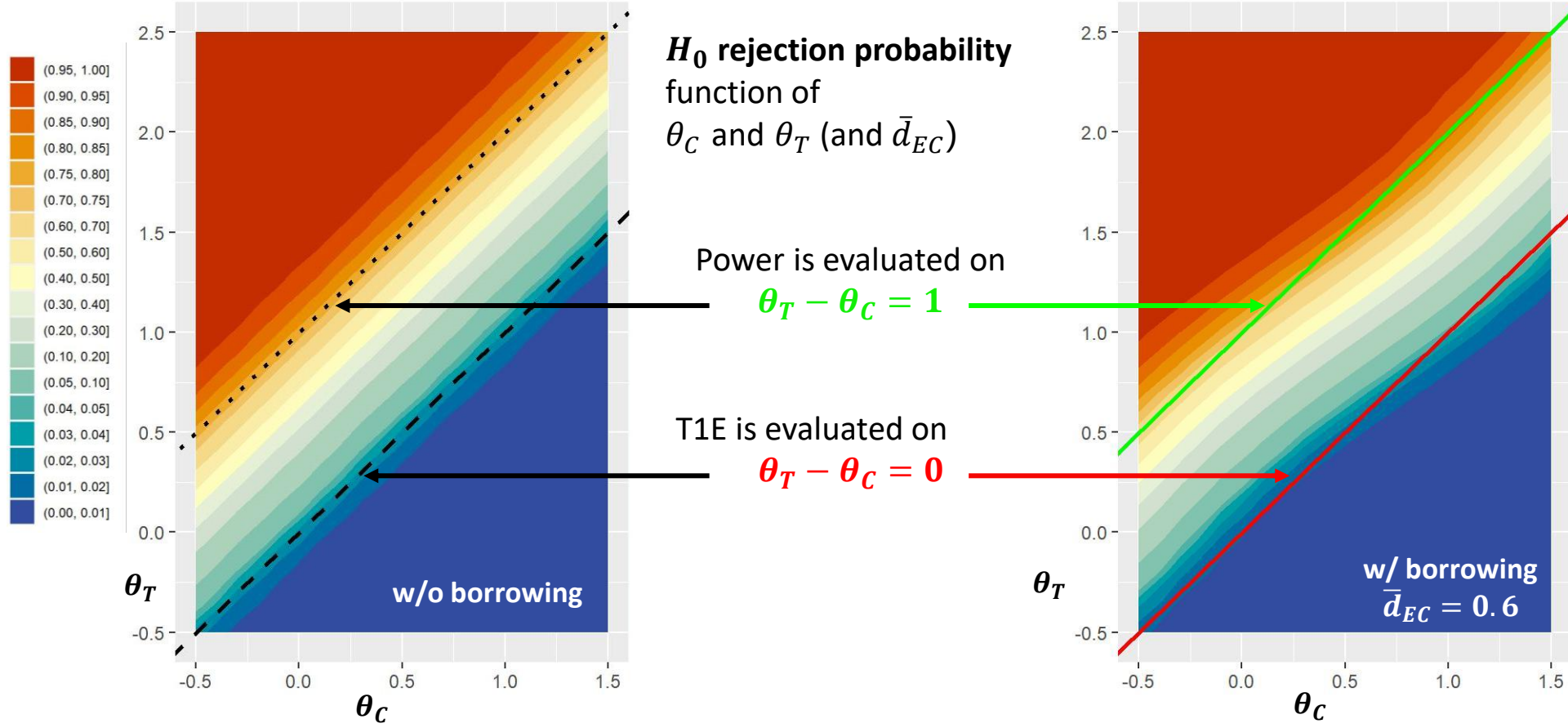




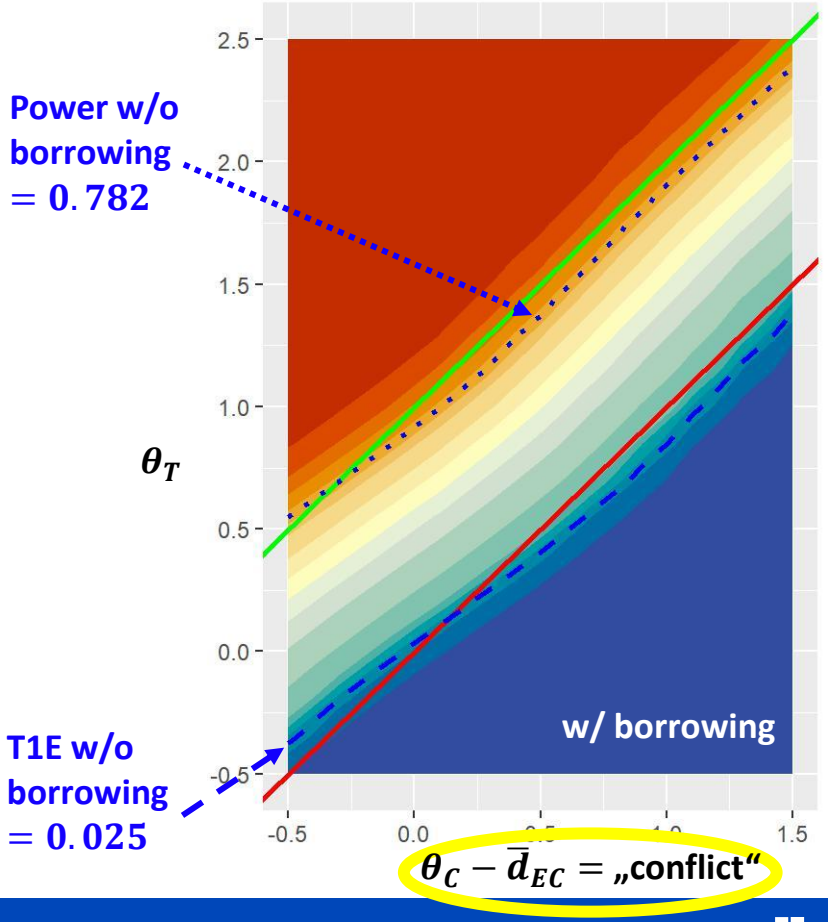
H_0 rejection probability as function of $\theta_T - \theta_C$:



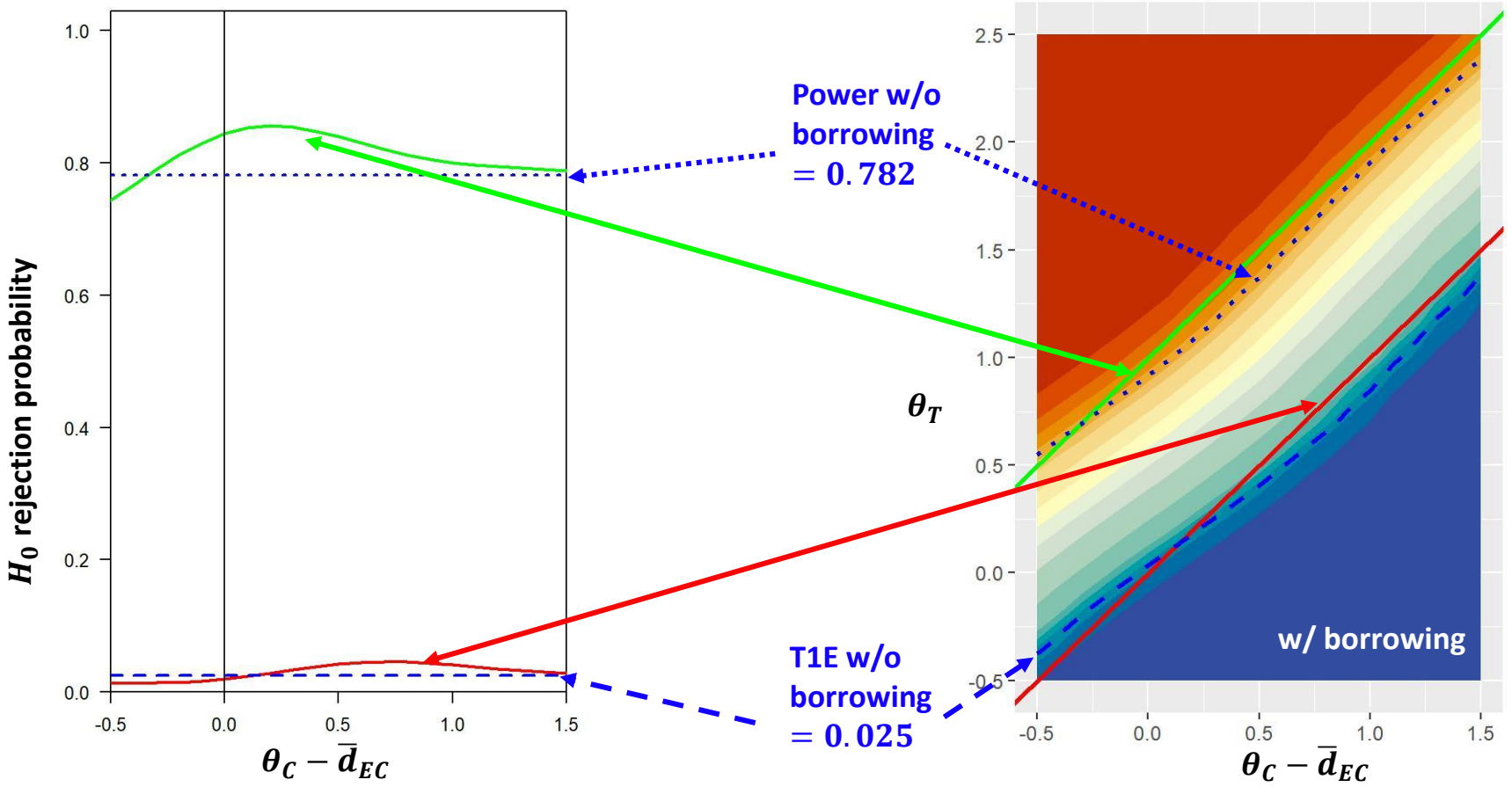
H_0 rejection probability in hybrid control trials: w/o and w borrowing



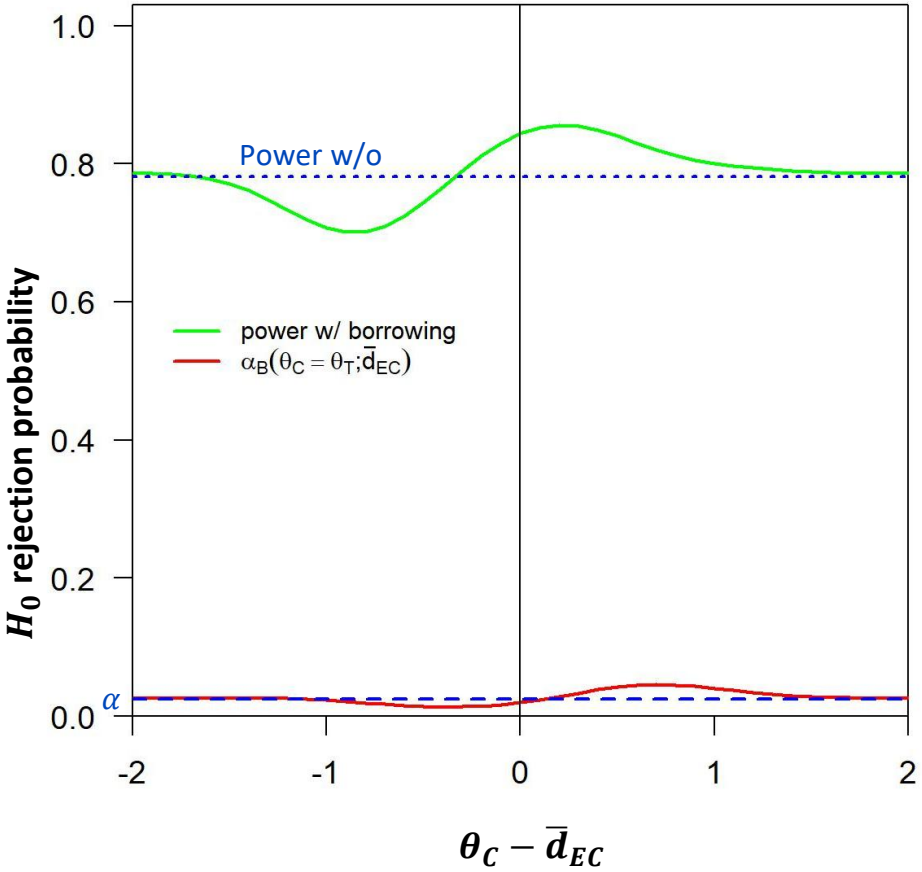
H_0 rejection probability in hybrid control trials



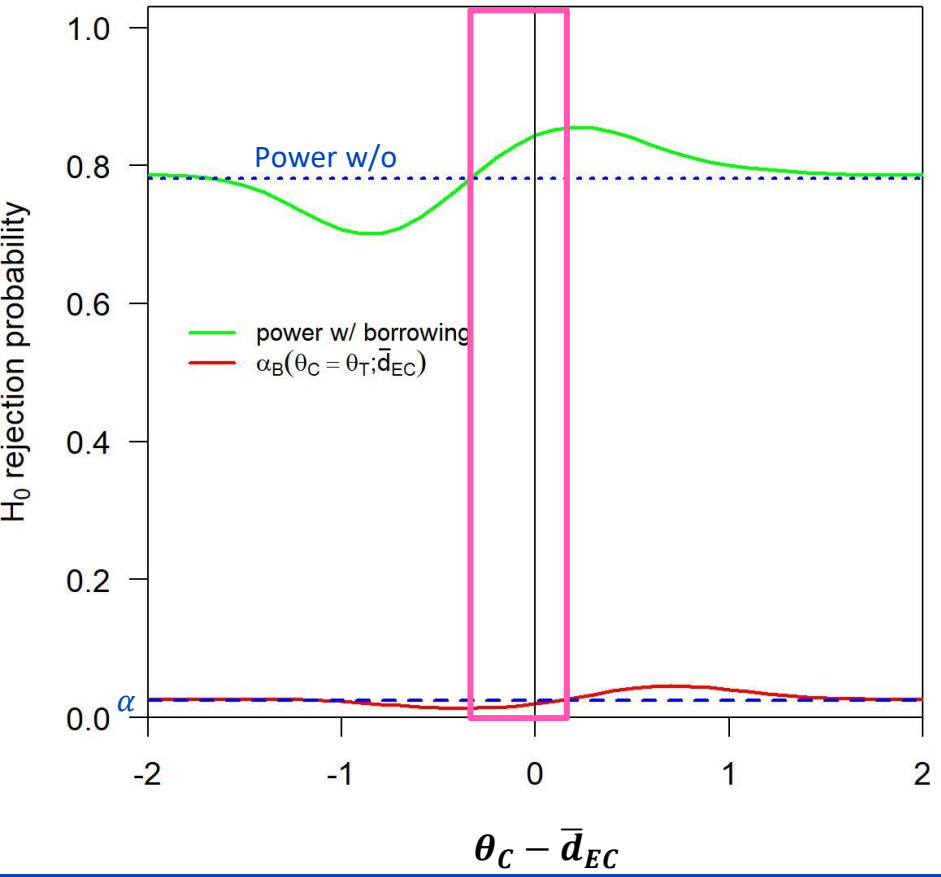
H_0 rejection probability in hybrid control trials



Hybrid control arm trial: Frequentist OC



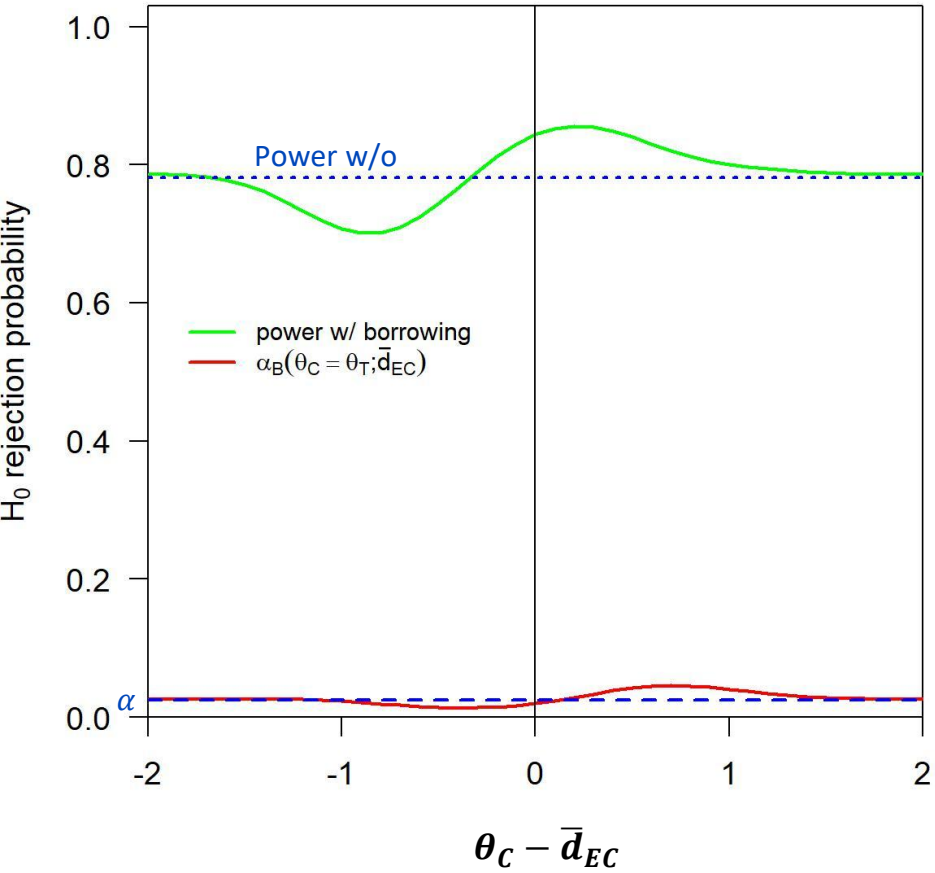
Hybrid control arm trial: Frequentist OC



„Sweet spot“:

(No T1E inflation) AND (power gain)

Hybrid control arm trial: Frequentist OC



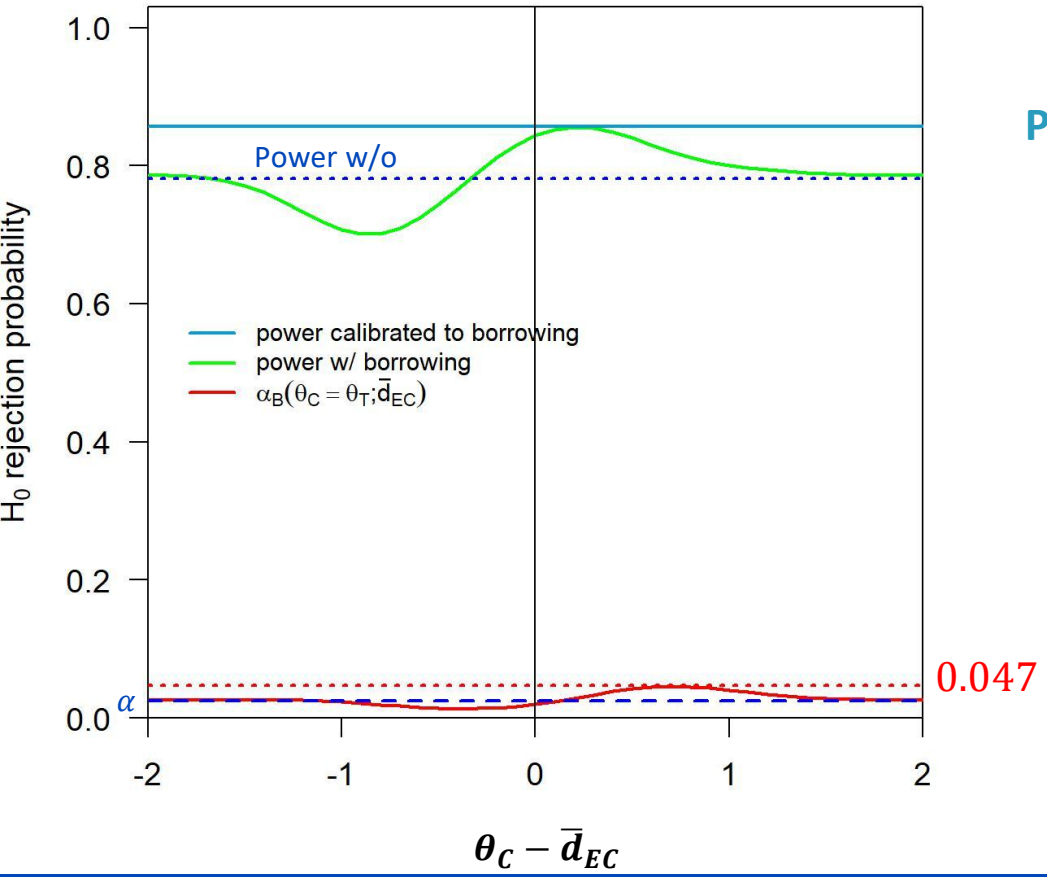
- T1E w/ borrowing, $\alpha_B(\theta_C = \theta_T; \bar{d}_{EC})$, varies with $\theta_C - \bar{d}_{EC}$
- **θ_C is unknown!**
- For fair comparison of test w/ and test w/o borrowing:

  Calibrate test w/o borrowing to have the same T1E as the test w/ borrowing (instead of $\alpha = 0.025$)

  → Since θ_C is unknown: calibrate to worst case

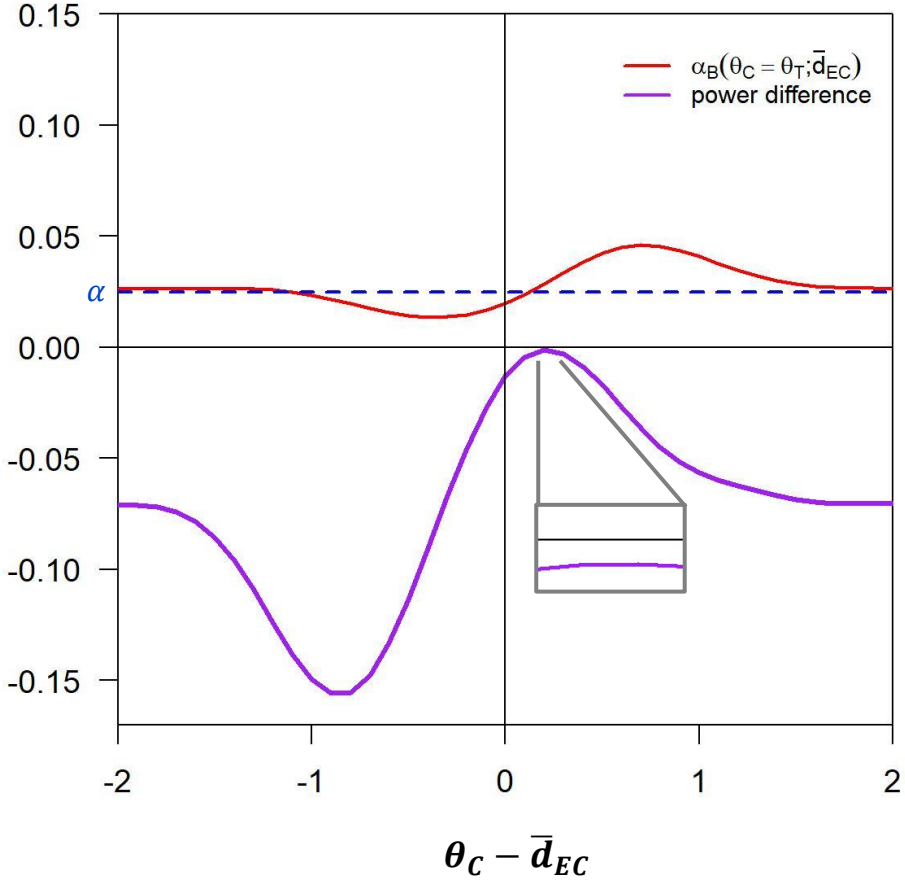
$$\max_{\theta_C} \alpha_B(\theta_C = \theta_T; \bar{d}_{EC})$$

Hybrid control arm trial: Frequentist OC



Power (at $\theta_T - \theta_C = 1$) of test calibrated to borrowing $= 0.86$

Hybrid control arm trial: Frequentist OC

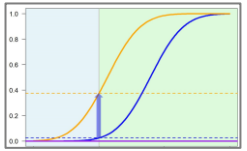


Evaluate power difference:

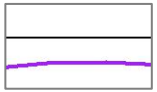
Power(test w/ borrowing) –
Power(calibrated test w/o borrowing)

Conclusions

- Whenever there is a Uniformly Most Powerful test → **No frequentist power gain possible**!
- True for **any borrowing method**, also for robust methods.
- Borrowing for 1-arm trial or to treatment effect in 2-arm trial:
  typically test w/ borrowing = test w/o borrowing (at adjusted T1E)



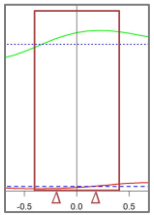
- Borrowing in hybrid control trial: typically (small) power loss



But: Power gains are possible if you trust similarity of current and external data,
i.e., leave frequentist framework

e.g.

- Instead of evaluating OCs for all $\theta_C \in (-\infty, \infty)$: **restrict $\boldsymbol{\theta_C}$** to $\left|\theta_C - \bar{d}_{EC}\right| < \Delta$



- Use Bayesian metric: Assume sampling prior for θ_C and evaluate average OCs → Nicky Best

References

- 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.

Additional work of the group:

- Calderazzo S, Wiesenfarth M, & Kopp-Schneider A (2022). A decision-theoretic approach to Bayesian clinical trial design and evaluation of robustness to prior-data conflict. *Biostatistics* 23(1), 328-344.
- Calderazzo S, Wiesenfarth M, Kopp-Schneider A (2024) Robust incorporation of historical information with known type I error rate inflation. *Biometrical Journal* 66 (1), 2200322.
- Calderazzo S, Tarima S, Reid C, Flournoy N, Friede T, Geller N, Rosenberger JL, Stallard N, Ursino M, Vandemeulebroecke M, Van Lancker K, Zohar S (2024) Coping with Information Loss and the Use of Auxiliary Sources of Data: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions. *Statistics in Biopharmaceutical Research* 16(2), 141-157.
- Kopp-Schneider A, Wiesenfarth M, Witt R, Edelmann D, Witt O, Abel U, Monitoring futility and efficacy in phase II trials with Bayesian posterior distributions - A calibration approach. *Biometrical Journal* 61, 488-502 (2019)
- Weru V, Kopp-Schneider A, Wiesenfarth M, Weber S, Calderazzo S (2024). Information borrowing in Bayesian clinical trials: choice of tuning parameters for the robust mixture prior. *arXiv:2412.03185*
- Wiesenfarth M, Calderazzo S (2020). Quantification of prior impact in terms of effective current sample size. *Biometrics 76*(1), 326-336.
- Zocholl D, Wiesenfarth M, Rauch G, Kopp-Schneider A (2022). On the feasibility of pediatric dose-finding trials in small samples with information from a preceding trial in adults. *Journal of Biopharmarmaceutical Statistics* 32(5), 652-670.