GMDS-CEN conference
Satellite Webinar “Long-run behaviour of Bayesian procedures“
16 September 2020

The effect of prior information on frequentist properties of Bayes test decisions

Annette Kopp-Schneider, Silvia Calderazzo and Manuel Wiesenfarth

Division of Biostatistics, German Cancer Research Center (DKFZ)
Heidelberg, Germany

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

Motivation

- Trial in adults with solid tumors harboring DNA repair deficiencies treated by targeted therapy, evaluation of response.
- DNA repair deficiencies also occur in pediatric tumors

→ investigate targeted therapy in a pediatric arm

<u>Question:</u>

Should this pediatric arm be designed as stand-alone arm

or

can power gain be expected when borrowing information from the adult trial?

Planning the pediatric arm with stand-alone evaluation

- Number of responders in children, $R_{ped} \sim \text{Bin}(n_{ped} = 40, p)$
- One-sided test $H_0\colon\ p \leq p_0$ vs. $H_1\colon\ p > p_0$, $p_0 = 0.2$
- Type I error rate $\alpha = 0.05$

Bayesian approach (1)

- Use beta-binomial model

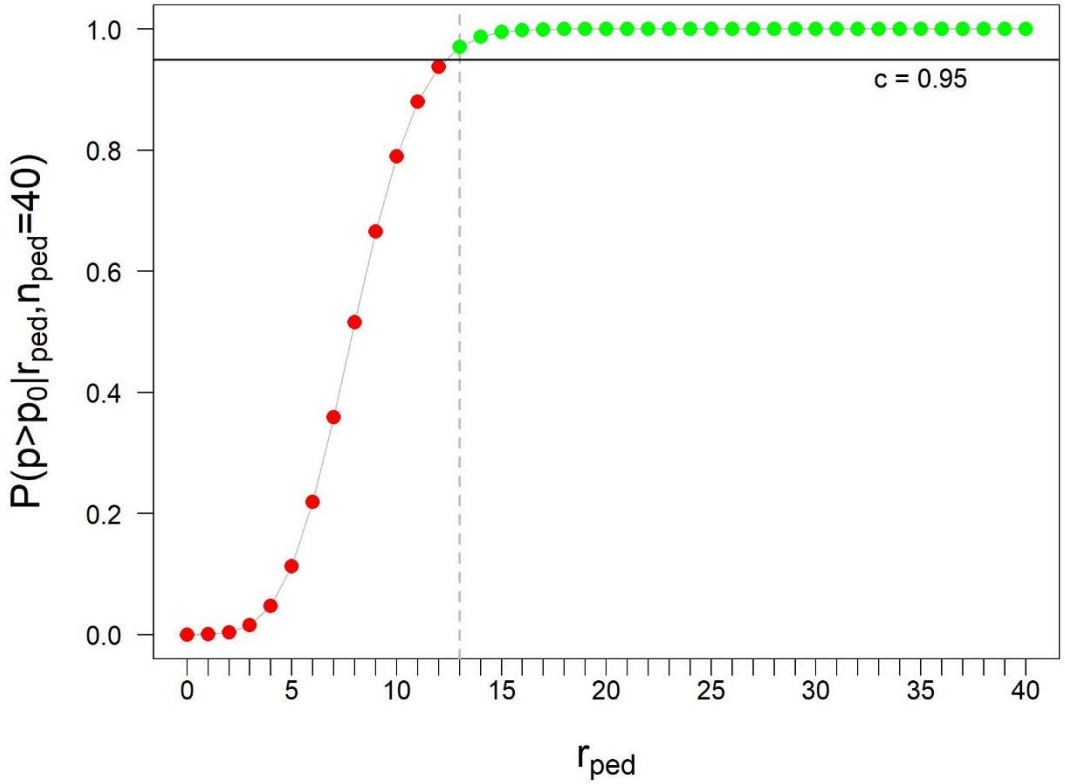
$$R_{ped} \mid p \sim \text{Bin}(n_{ped}, p),\ \pi(p) = \text{Beta}(0.5, 0.5)$$

- Evaluate efficacy based on Bayesian posterior probability:

$$\text{Reject } H_0 \ \Leftrightarrow\ P\left(p > p_0 = 0.2 | r_{ped}\right) \geq c, \text{ e.g., } c = 0.95.$$

Planning the pediatric arm with stand-alone evaluation: Bayesian approach (2)

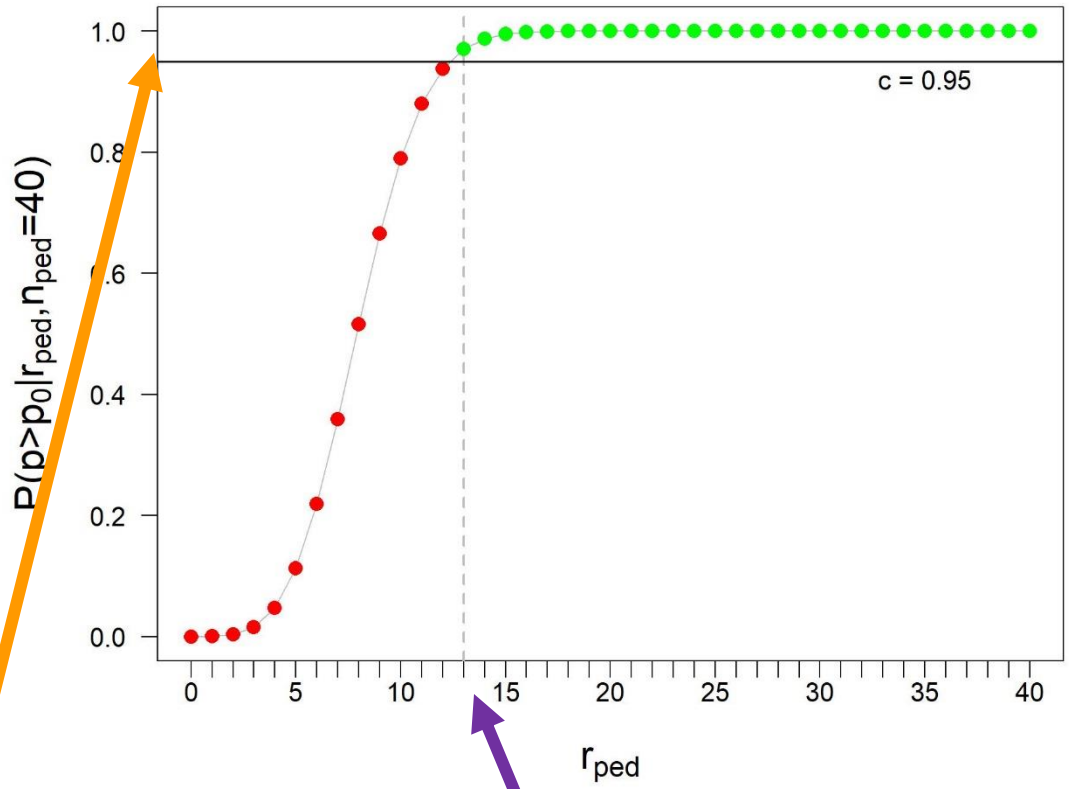
Posterior probability $P(p > p_0 | r_{ped})$ as a function of r_{ped}



$$P(p > p_0 | r_{ped}) \geq 0.95 \quad \Leftrightarrow \quad r_{ped} \geq 13$$

Planning the pediatric arm with stand-alone evaluation: Bayesian approach (2)

Posterior probability $P(p > p_0 | r_{ped})$ as a function of r_{ped}



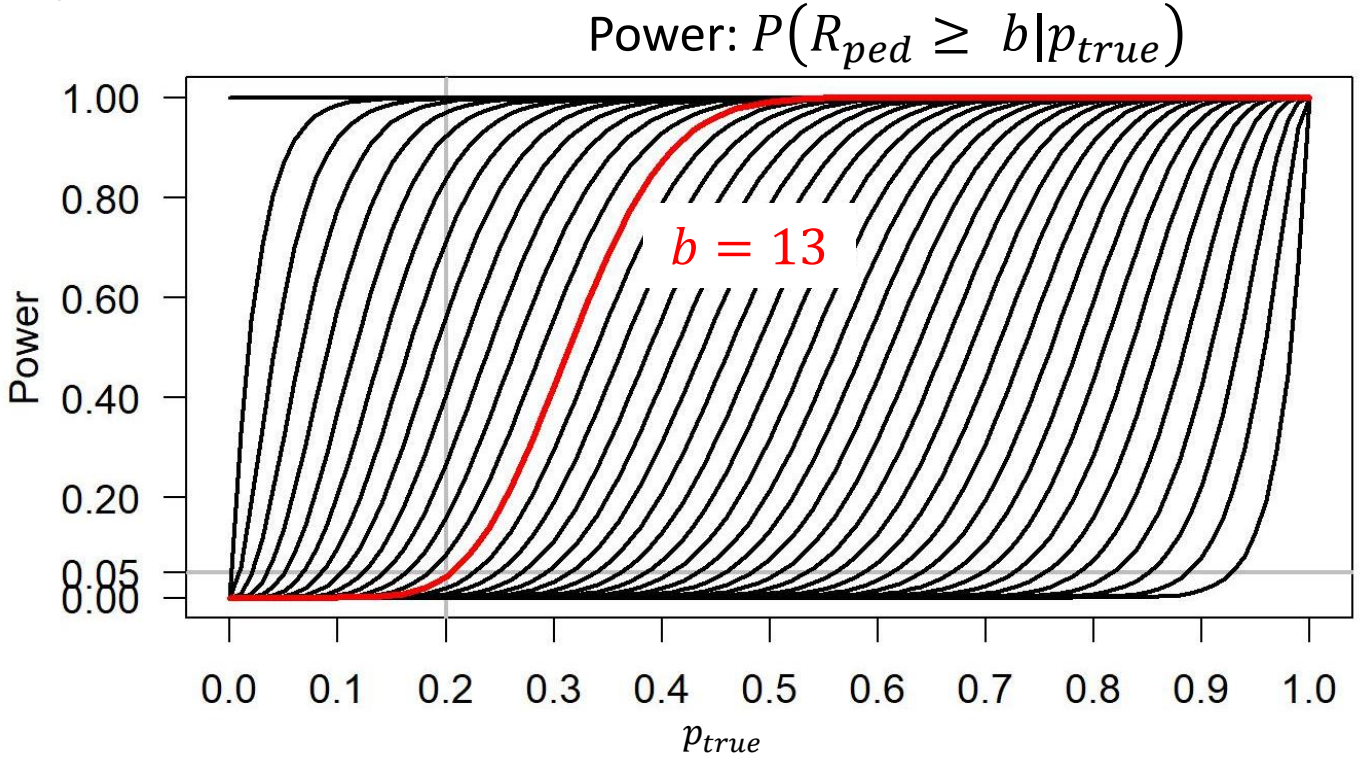
$P(p > p_0 | r_{ped}) \geq 0.95 \Leftrightarrow r_{ped} \geq 13$

Planning the pediatric arm with stand-alone evaluation: Frequentist approach

- Uniformly most powerful (UMP) level α test is given by:

  reject $H_0 \quad \Leftrightarrow \quad r_{ped} \geq b_{\text{UMP}}(\alpha)$

- Here: $b_{\text{UMP}}(0.05) = 13$
- All possible power curves for $n_{ped} = 40$ for varying threshold b (and type I error probability):

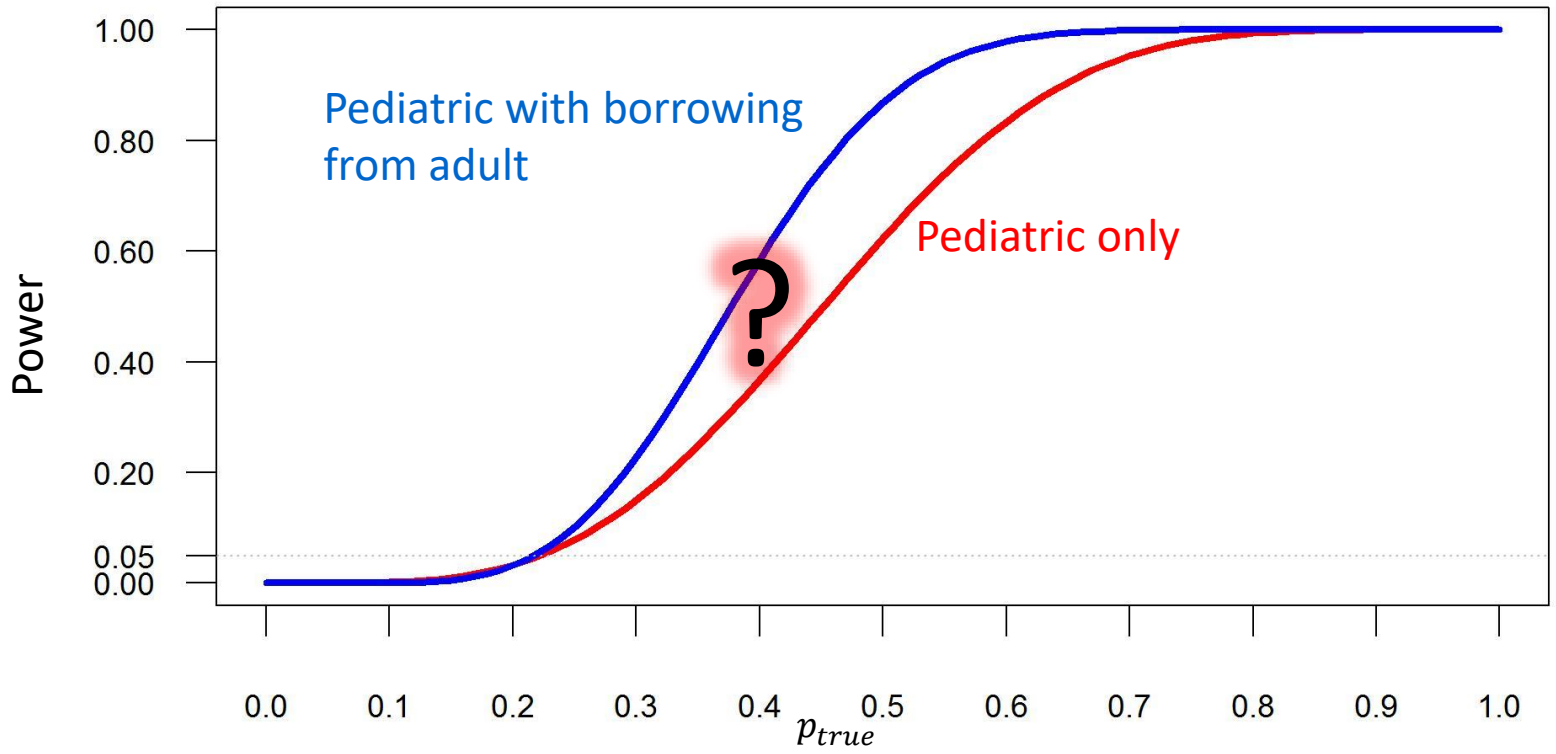


Borrowing from adult information for the pediatric arm

Use results from adult trial to inform the prior for the pediatric arm.

Hope

If treatment is successful in adults, then power is increased for pediatric arm:



Adaptive power parameter (1)

Power prior approach with power parameter $\delta \in [0, 1]$:

$$\pi(p|r_{adu}, \delta) \propto L(p; r_{adu})^{\delta} \pi(p)$$

Adapt $\delta = \delta(r_{ped}, r_{adu})$ such that information is only borrowed for similar adult and pediatric data:

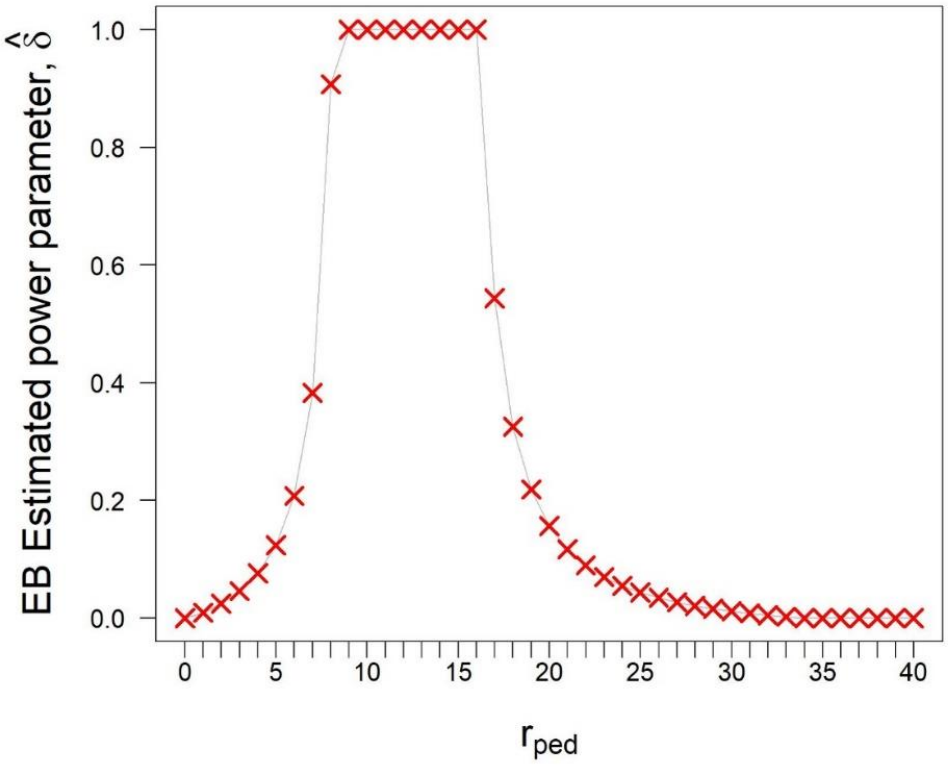
$\rightarrow \delta(r_{ped}, r_{adu})$ large when adult and children data are similar
$\rightarrow \delta(r_{ped}, r_{adu})$ small in case of prior-data conflict.

Adaptive power parameter (2)

Result from adult trial: $r_{adu} = 12$ among $n_{adu} = 40$ $(\hat{p}_{adu} = 0.3)$

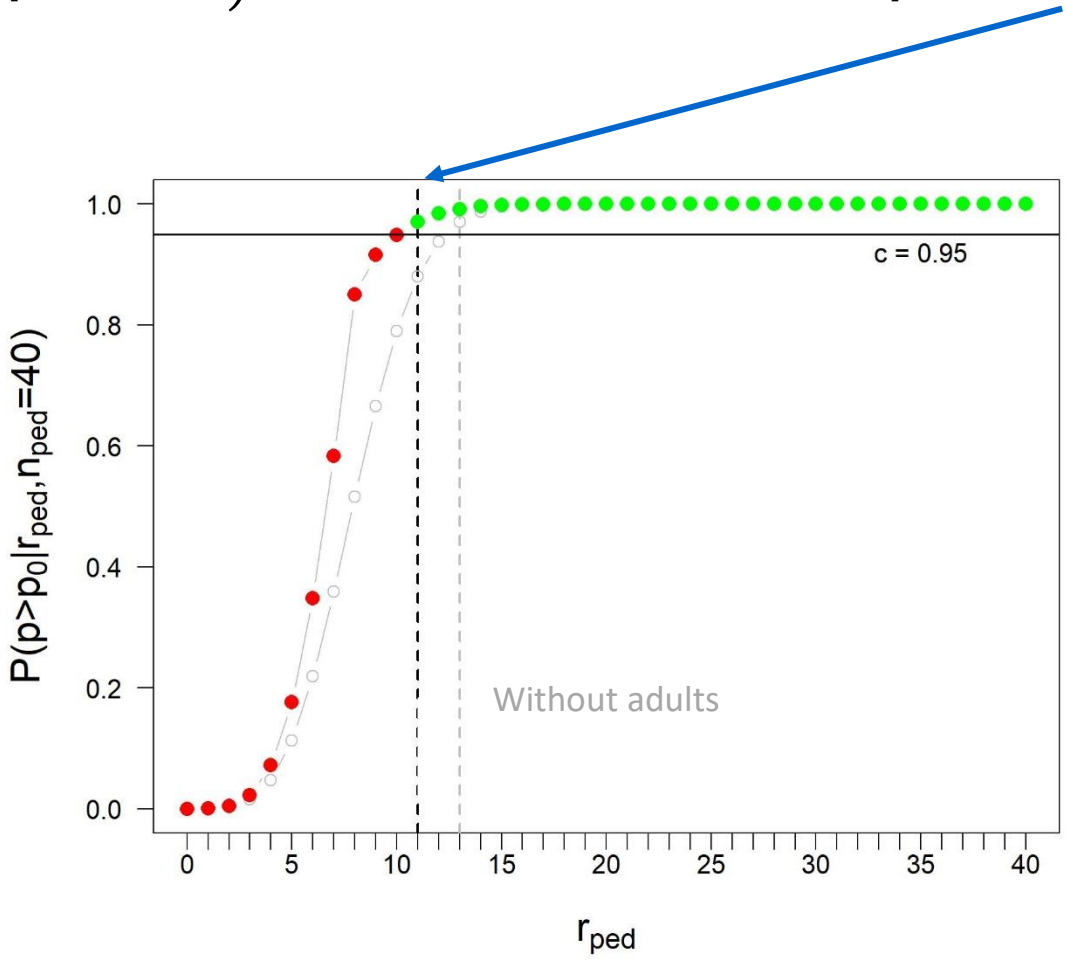
Use an Empirical Bayes power prior approach where $\hat{\delta}(r_{ped}; r_{adu} = 12)$ maximizes the marginal likelihood of δ (Gravestock, Held et al. 2017):

$\hat{\delta}(r_{ped}; r_{adu} = 12)$:



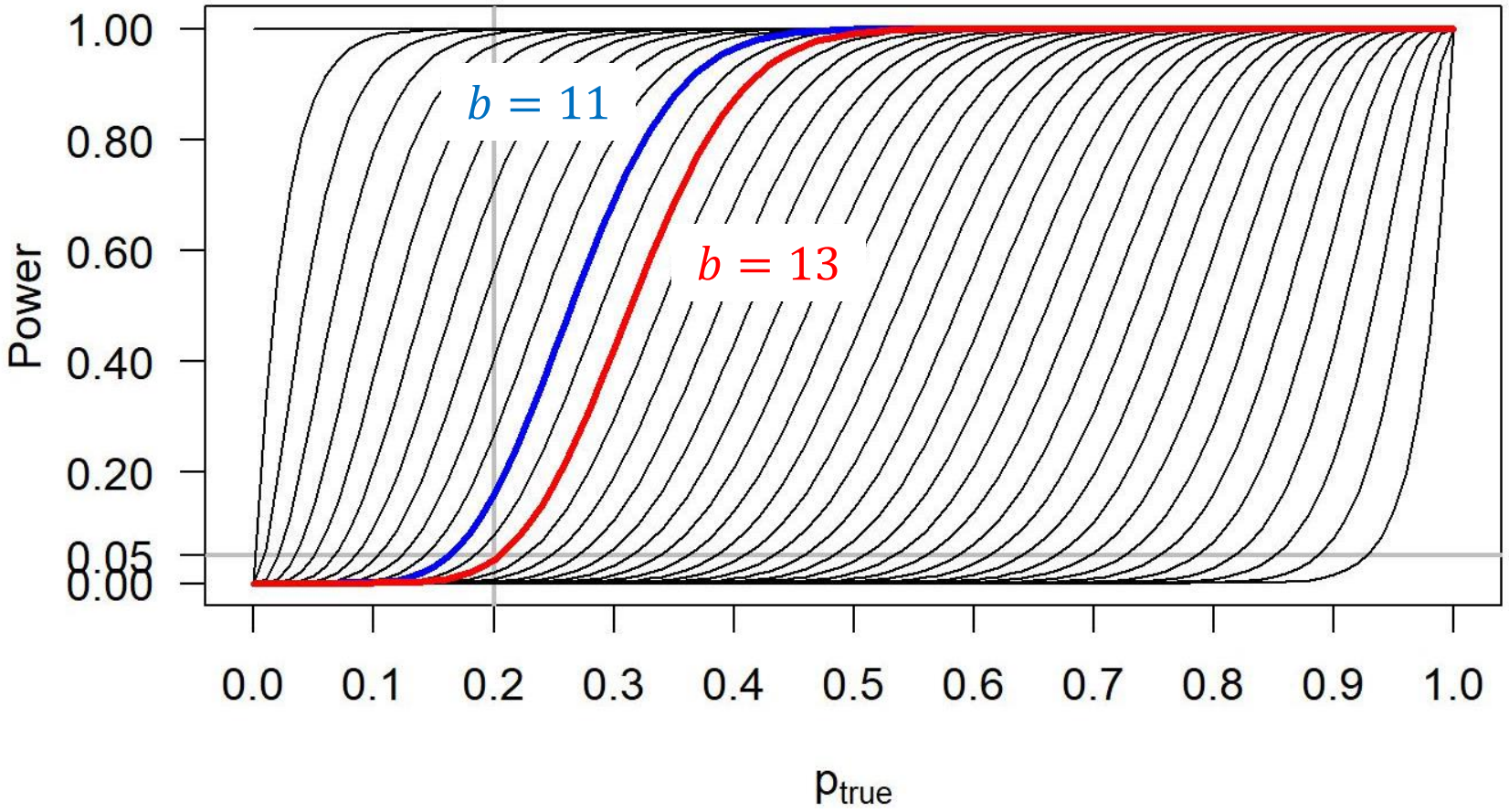
Adaptive power parameter (3)

$P\left(p > p_0 | r_{ped}, r_{adu}, \hat{\delta}\left(r_{ped}; r_{adu}\right)\right) > c = 0.95$ corresponds to $r_{ped} \geq b = 11$



Adaptive power parameter (4)

$P\left(p > p_0 | r_{ped}, r_{adu}, \hat{\delta}\left(r_{ped}; r_{adu}\right)\right) > c = 0.95$ corresponds to $r_{ped} \geq b = 11$

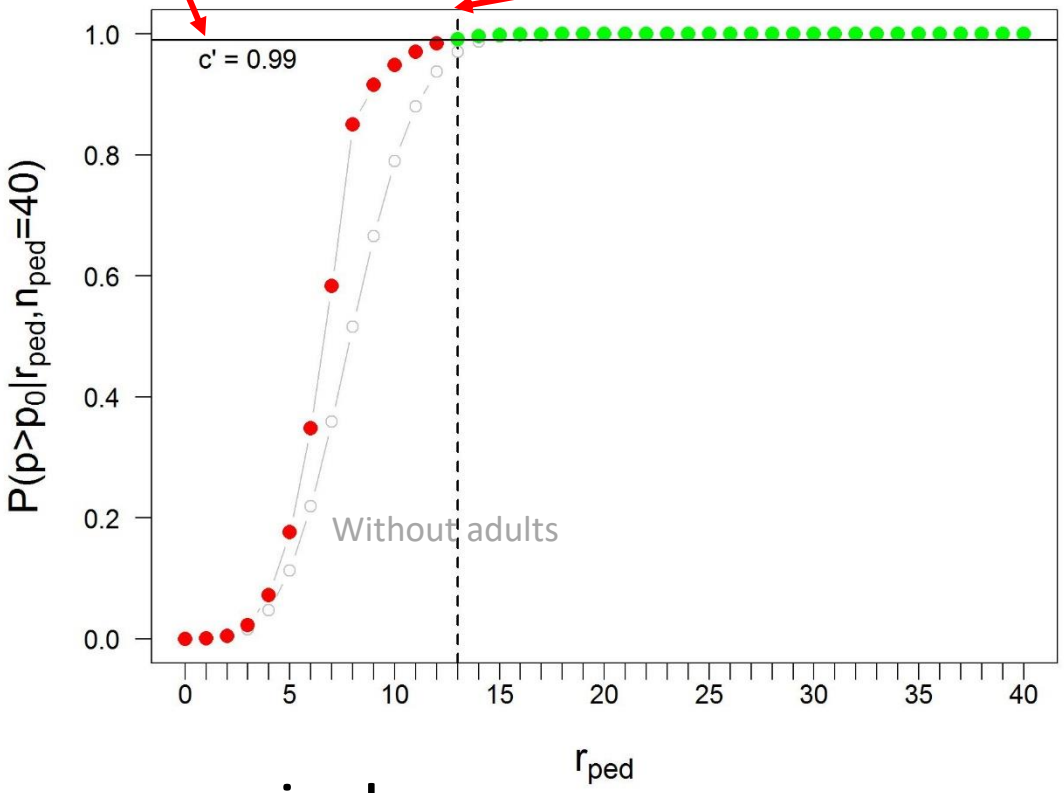


→ **power gain but type I error inflation**

Adaptive power parameter (5)

$P\left(p > p_0 | r_{ped}, r_{adu}, \hat{\delta}\left(r_{ped}, r_{adu}\right)\right)$ is monotonically increasing in r_{ped}

$\rightarrow P\left(p > p_0 | r_{ped}, r_{adu}, \hat{\delta}\right) > c' = 0.99$ corresponds to $r_{ped} \geq b = 13$



$\rightarrow$ **type I error controlled but no power gained**

“Extreme borrowing” (1)

- Artificial method for illustration of not monotonically increasing $P(p > p_0 | r_{ped}, r_{adu})$:

  borrow adult information $\Leftrightarrow$ $\hat{p}_{adu} = \hat{p}_{ped}$

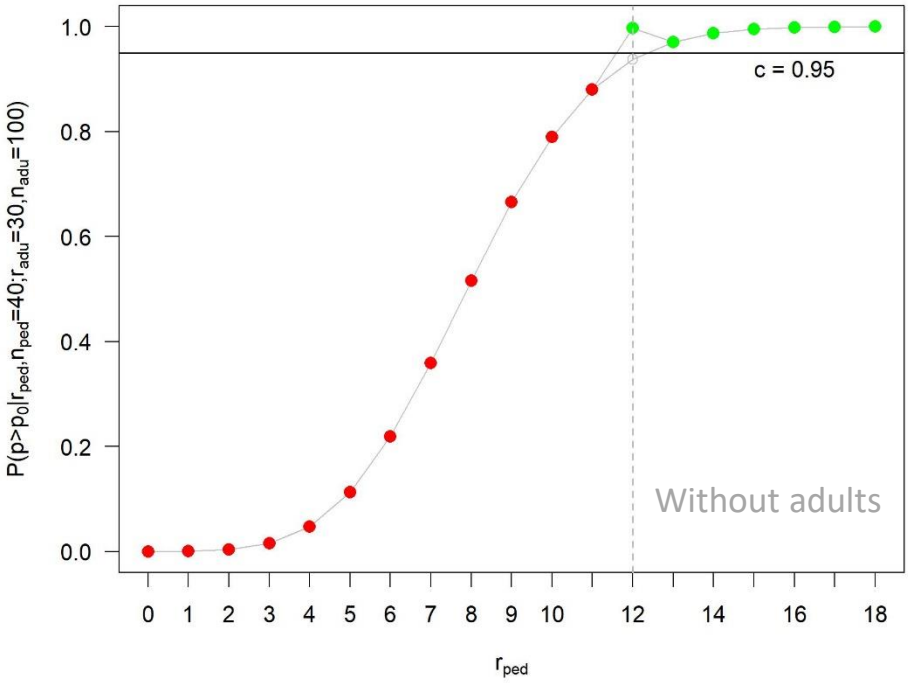
- Assume $n_{adu} = 100$, $r_{adu} = 30 \Rightarrow \hat{p}_{adu} = 0.3$

- Here:
  borrow all adult information if $\hat{p}_{ped} = 0.3$ ($r_{ped} = 12$ for $n_{ped} = 40$)
  don‘t borrow for $r_{ped} \neq 12$

“Extreme borrowing” (2)

Borrow all adult information iff r_{ped} = 12

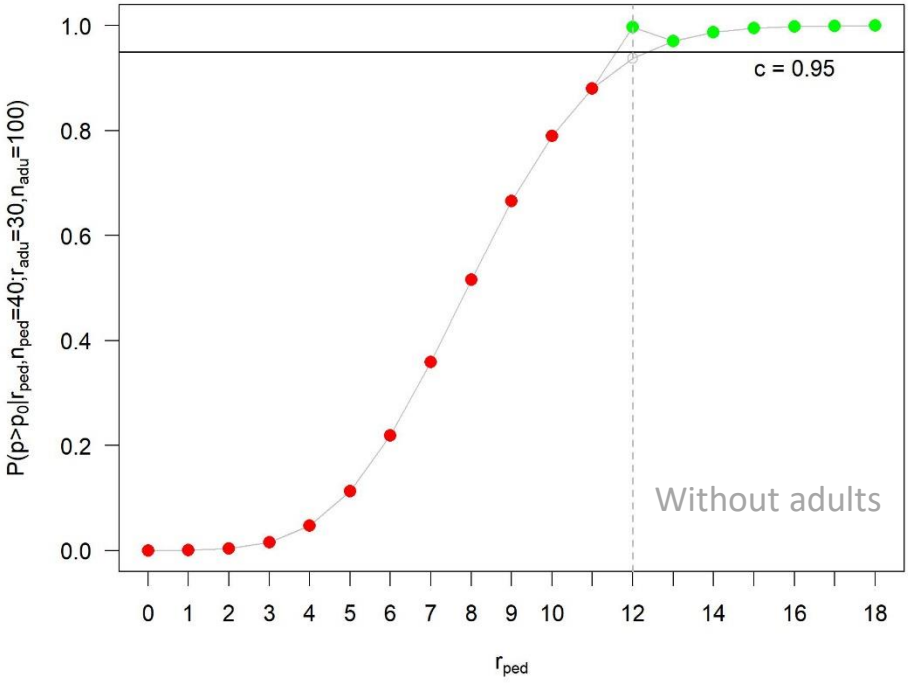
For $c = 0.95 \Rightarrow b = 12$
$\Rightarrow$ type I error rate $= 0.088$



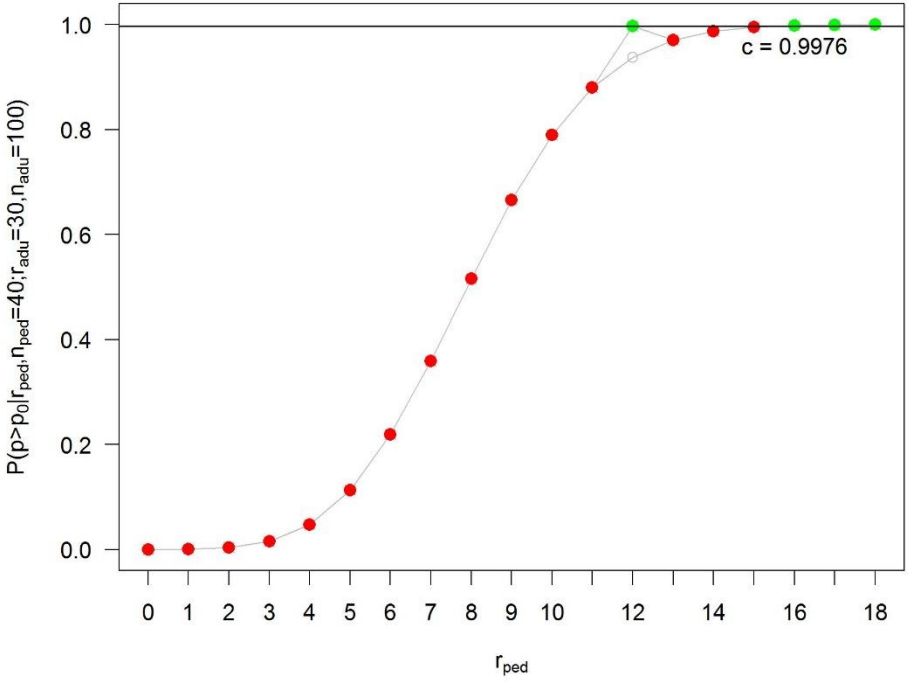
“Extreme borrowing” (3)

Borrow all adult information iff r_{ped} = 12

For $c = 0.95 \Rightarrow b = 12$
$\Rightarrow$ type I error rate $= 0.088$



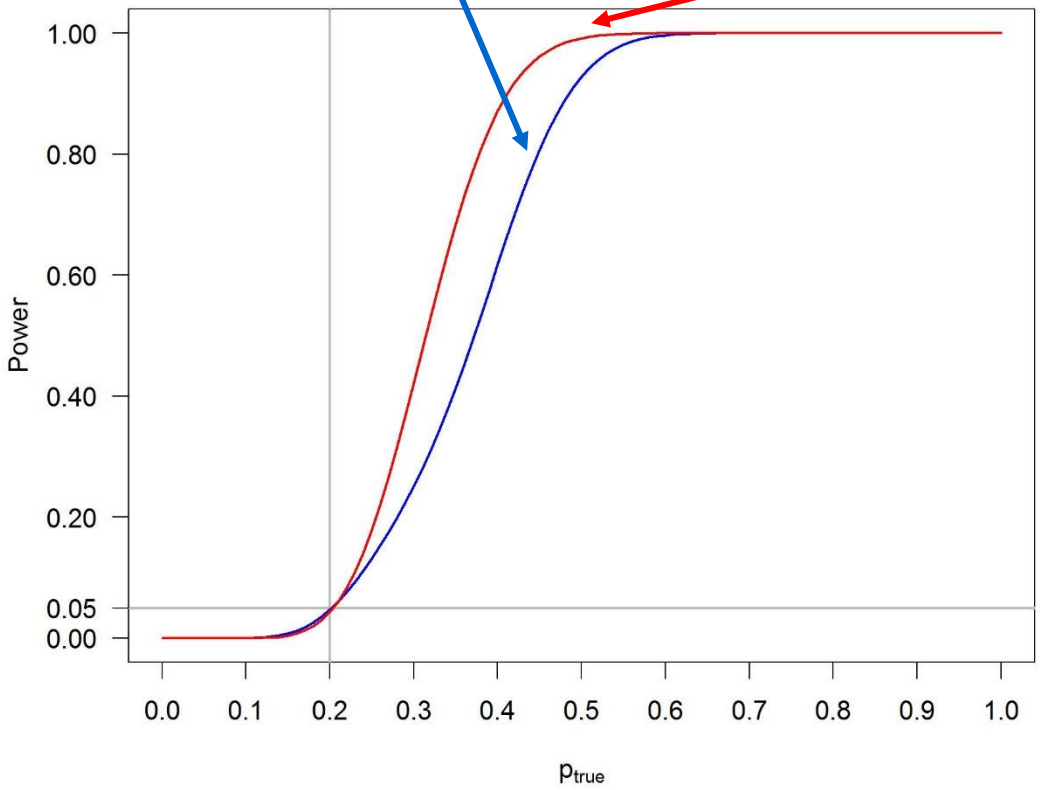
For $c = 0.9976$
$\Rightarrow$ reject H_0
if $r_{ped} = 12$ or $r_{ped} \geq 16$
$\Rightarrow$ type I error rate $= 0.047$



“Extreme borrowing” (4)

Reject H_0 if $r_{ped} \in \{12\} \cup \{16, 17, \ldots, 40\}$

Compare to: Reject H_0 if $r_{ped} \in \{13, 14, \ldots, 40\}$



→ **type I error controlled but power decreased**

Borrowing from adult information (1)

If $P(p > p_0 | r_{ped}, r_{adu})$ is **monotonically increasing** in r_{ped},

then there exists c' with

$$P(p > p_0 | r_{ped}, r_{adu}) \geq c' \Leftrightarrow r_{ped} \geq b_{\text{UMP}}(\alpha)$$

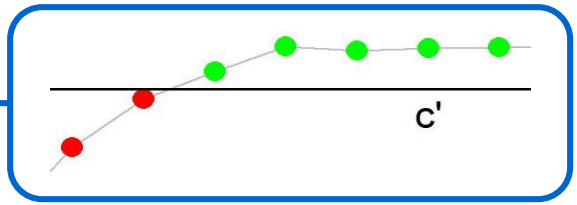
and $b_{\text{UMP}}(\alpha)$ is the level α UMP test boundary.

Borrowing from adult information (2)

If $P(p > p_0 | r_{ped}, r_{adu})$ is **not monotonically increasing** in r_{ped}, then either:

(1) a threshold c' can still be identified with

$$P(p > p_0 | r_{ped}, r_{adu}) \geq c' \Leftrightarrow r_{ped} \geq b_{\text{UMP}}(\alpha) \quad (*)$$

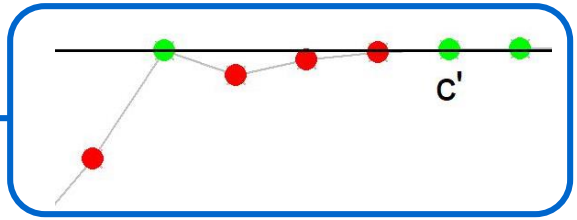


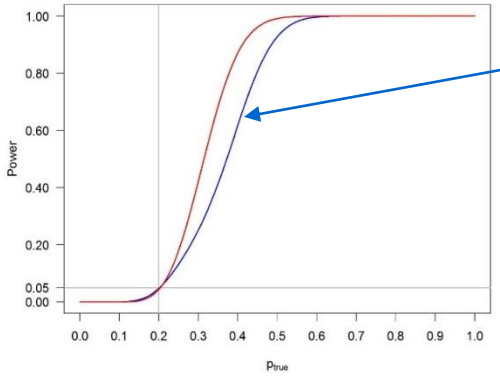
(2) if no c' with $(*)$ can be identified, then either the

- test does not control type I error

or

- test controls type I error but is not UMP.





Borrowing from adult information: Summary

View decision rule in Bayesian approach as test function

$$\varphi(r_{ped}) = 1_{\{P(p>p_0|r_{ped},r_{adu})\geq c\}}$$

$\rightarrow$ There is nothing better than the UMP test!

- This holds for all situations in which UMP tests exist:
  exponential family distribution
  one-sided tests, two-sided tests (equivalence situation)
  one-sided comparison of two means of normal variables …
- This also holds in situations in which UMP unbiased tests exists:
  two-sided comparisons
  comparison of two proportions …
- True for any (adaptive) borrowing mechanism (power prior, mixture prior, hierarchical model, test-then-pool,…) (see Viele et al. (2014))

- Proven by Psioda and Ibrahim (2018) for one-sample one-sided normal test with borrowing using a fixed power prior.

In general

- d_C : realizations of current data D_C collected to test: $\vartheta_C \in H_0$ vs. $\vartheta_C \notin H_0$
- Without historical data:
  Lehmann (1986) notation: the UMP hypothesis test is (T sufficient test statistic)

$$\varphi(d_C) = \begin{cases} 1 & \text{if } T(d_C) \in \text{RejectionRegion} \quad (\text{reject } H_0) \\ 0 & \text{if } T(d_C) \in \text{AcceptanceRegion} \quad (\text{accept } H_0) \end{cases}$$

  → power function $E_{\vartheta_C}[\varphi(D_C)]$
  → type I error control: $E_{\vartheta_C}[\varphi(D_C)] \leq \alpha$ for all $\vartheta_C \in H_0$

- With historical data:
  Borrow from observed historical data d_H (from D_H) by:

$$\varphi_B(d_C; d_H) = \begin{cases} 1 & \text{if } T(d_C) \in \text{RejectionRegion}(d_H) \\ 0 & \text{if } T(d_C) \in \text{AcceptanceRegion}(d_H) \end{cases}$$

  → power function $E_{\vartheta_C}[\varphi_B(D_C; d_H)] = E_{\vartheta_C,\vartheta_H}[\varphi_B(D_C; D_H)|D_H = d_H]$
  → type I error: $\max_{\vartheta_C \in H_0} \{E_{\vartheta_C}[\varphi_B(D_C; d_H)]\}$ (note: ϑ_C may be multidimensional)

Simulating operating characteristics of borrowing methods (1)

- For frequentist characteristics: interest in power function

$$E_{\vartheta_C}[\varphi_B(D_C; d_H)] = E_{\vartheta_C,\vartheta_H}[\varphi_B(D_C; D_H)|D_H = d_H]$$

- But: fixing d_H may be perceived not objective enough since individual case study

- **Cave**:

  Simulating (d_C, d_H) (according to $(\vartheta_C, \vartheta_H)$) and evaluating $\varphi_B(d_C; d_H)$

  $\rightarrow E_{\vartheta_C,\vartheta_H}[\varphi_B(D_C; D_H)]$

  but $E_{\vartheta_C,\vartheta_H}[\varphi_B(D_C; D_H)] \neq E_{\vartheta_C,\vartheta_H}[\varphi_B(D_C; D_H)|D_H = d_H]$

Simulating operating characteristics of borrowing methods (2)

Proposals

A (1) Simulate d_H (according to ϑ_H)

(2) Repeatedly simulate d_C (according to ϑ_C)

→ evaluate $E_{\vartheta_C}[\varphi_B(D_C; d_H)]$

(3) Calculate type I error: $\max_{\vartheta_C \in H_0} \{E_{\vartheta_C}[\varphi_B(D_C; d_H)]\} = \alpha^{d_H}$

(4) Compare to power function of level α^{d_H} test w/o borrowing $(E_{\vartheta_C}[\varphi^{d_H}(D_C)])$:

$$E_{\vartheta_C}[\varphi_B(D_C; d_H)] - E_{\vartheta_C}[\varphi^{d_H}(D_C)]$$

(5) Repeat (1) - (4)

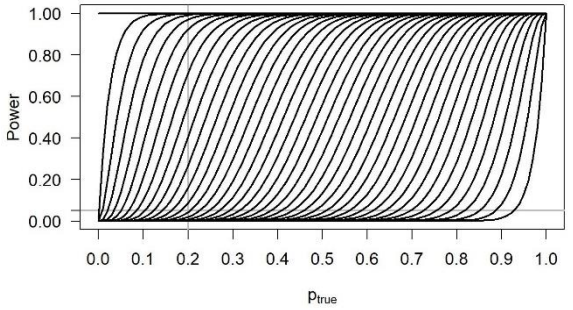
(6) Report $E_{\vartheta_H}\left[E_{\vartheta_C}[\varphi_B(D_C; d_H)] - E_{\vartheta_C}[\varphi^{d_H}(D_C)]\right]$

B Show relationship: $d_H \leftrightarrow \alpha^{d_H}$

Conclusion

- If type I error control is desired in a situation where a UMP (unbiased) test exists, external information is effectively discarded.

- For a given historical data setting,
  choose from the available power functions
  for current data.



- If prior information is reliable and consistent with the current information, the final operating characteristics of the trial can be improved: increased power or lower type I error, depending on where prior information is placed (but at expense of the other characteristic).

→ Incorporation of prior information can give a rationale for type I error inflation with benefit of a power gain,
amount of type I error inflation reflects degree of reliance on prior information.

References

- Gravestock I, Held L; COMBACTE-Net consortium (2017). Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical Statistics* 16(5): 349-360.
- 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.
- Lehmann E (1986). *Testing statistical hypotheses* (2nd ed.). Wiley Series in Probability and Statistics. New York: John Wiley & Sons.
- Psioda MA, Ibrahim JG (2018) Bayesian clinical trial design using historical data that inform the treatment effect. *Biostatistics* 20(3): 400-415.
- Viele K, Berry S, Neuenschwander B, et al. (2014) Use of historical control data for assessing treatment effects in clinical trials. *Pharm Statistics* 13(1):41-54.