BAYES 2019 - Lyon, France - 21-24 May 2019

Use of external information in clinical trials: What can be gained in terms of frequentist power?

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 children

  → investigate targeted therapy in a pediatric arm

Question:

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, $R_{ped} \sim \text{Bin}(n_{ped}, p)$
- 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 $\alpha = 0.05$
- $n_{ped} = 40$

Bayesian approach:

- Use beta-binomial model

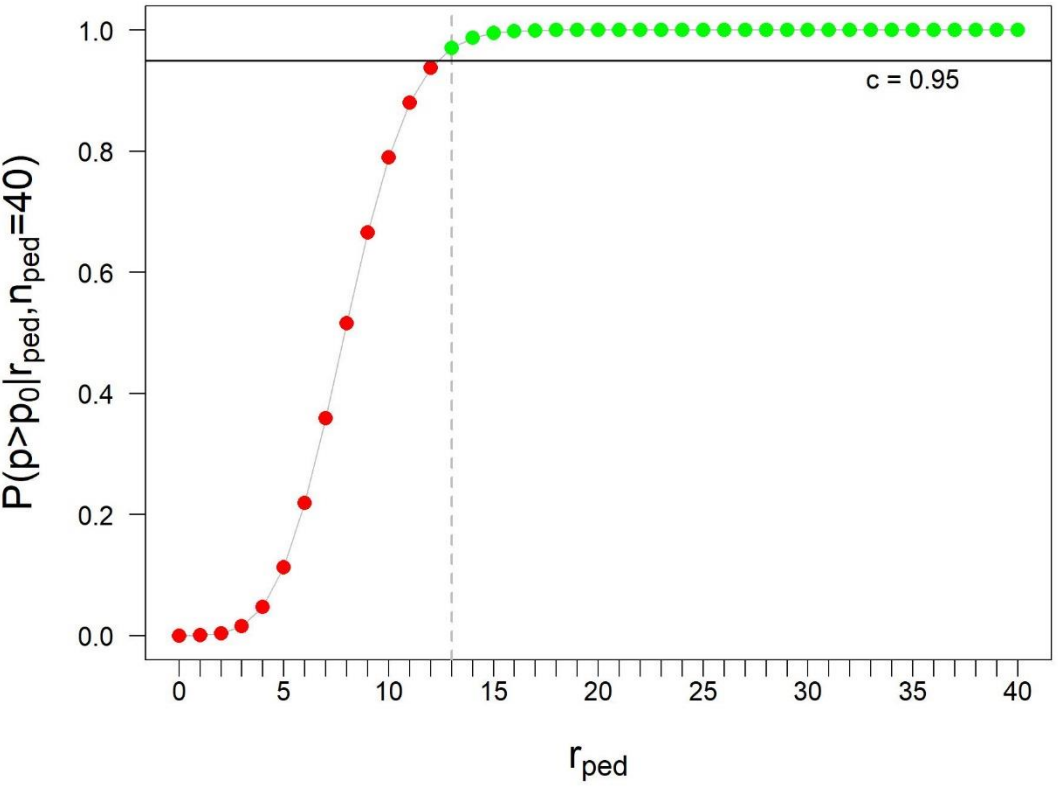
$$R_{ped} \mid p \sim \text{Bin}(n_{ped}, p),\ \pi(p) = \text{Beta}(0.5, 0.5)$$

- Evaluate efficacy based on Bayesian posterior probability:

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

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}



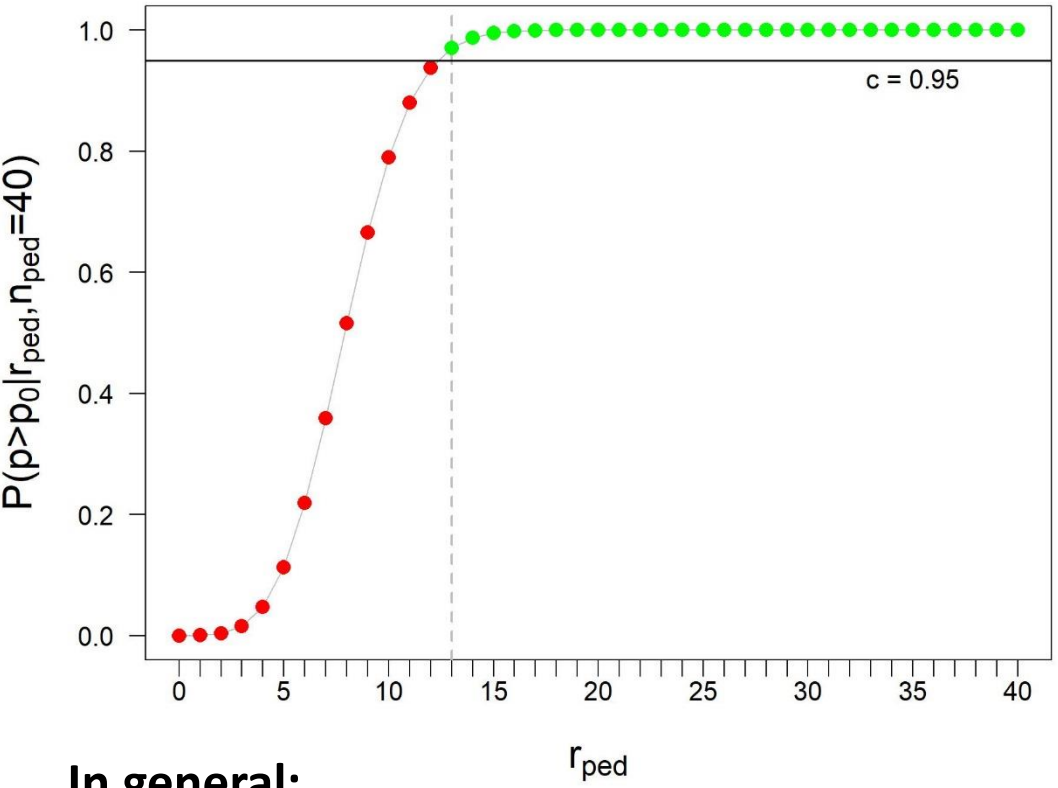
For $\boldsymbol{n_{ped} = 40}$:

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

$$r_{ped} \geq 13$$

Planning the pediatric arm with stand-alone evaluation: Bayesian approach (3)

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



For $\boldsymbol{n_{ped} = 40}$:

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

$$r_{ped} \geq 13$$

In general:

For every $c \in [0, P(p > p_0 | r_{ped} = n_{ped})]$ there exists a unique $b \in \{0,1, \ldots, n_{ped}\}$ with $P(p > p_0 | r_{ped}) \geq c \Leftrightarrow r_{ped} \geq b$

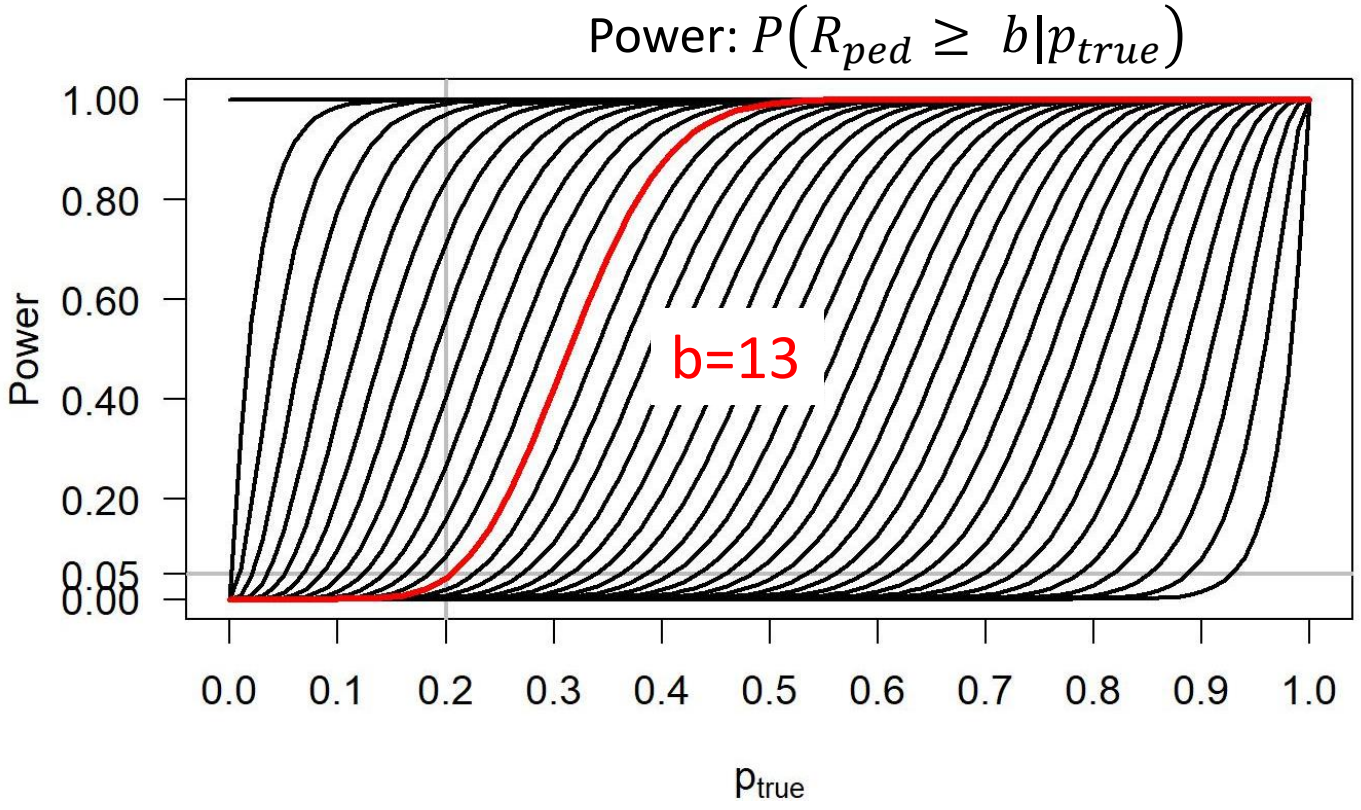
(Kopp-Schneider et al., 2019)

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

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

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

- Here: $b_{\text{UMP}}(0.05) = 13$



Planning the pediatric arm with stand-alone evaluation: Power function (1)

$$\text{Power} = f(p_{true})$$

$$= P\left(R_{ped} \geq b | p_{true}\right)$$

$$= \sum_{r_{ped}=0}^{n} P\left(R_{ped} = r_{ped} | p_{true}\right) 1_{\{r_{ped} \geq b\}}$$

Planning the pediatric arm with stand-alone evaluation: Power function (2)

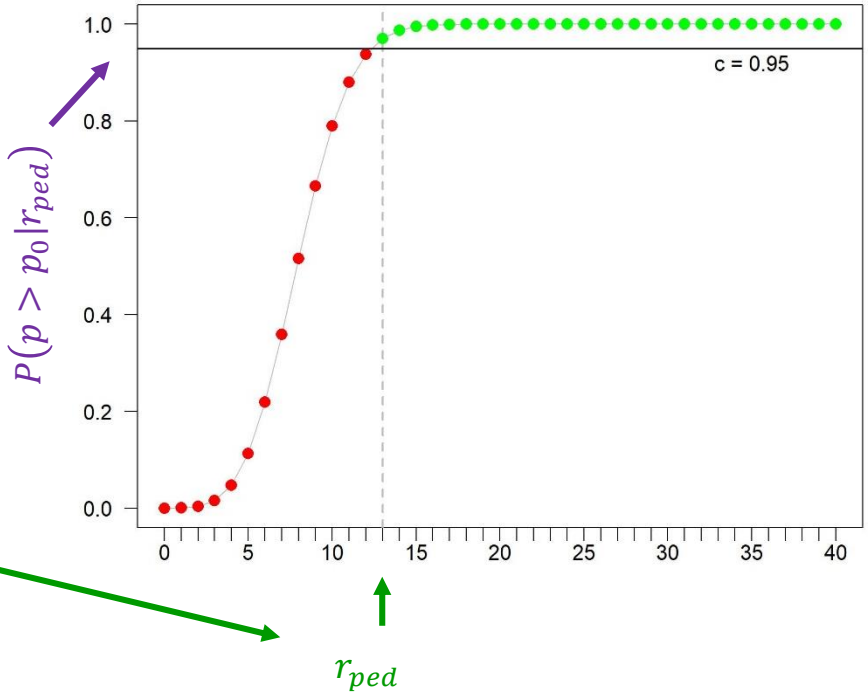
$$\text{Power} = f(p_{true})$$

$$= P(R_{ped} \geq b | p_{true})$$

$$= \sum_{r_{ped}=0}^{n} P(R_{ped} = r_{ped} | p_{true}) 1_{\{r_{ped} \geq b\}}$$

$$= \sum_{r_{ped}=0}^{n} P(R_{ped} = r_{ped} | p_{true}) 1_{\{P(p > p_0 | r_{ped}) \geq c\}}$$

(c selected appropriately)

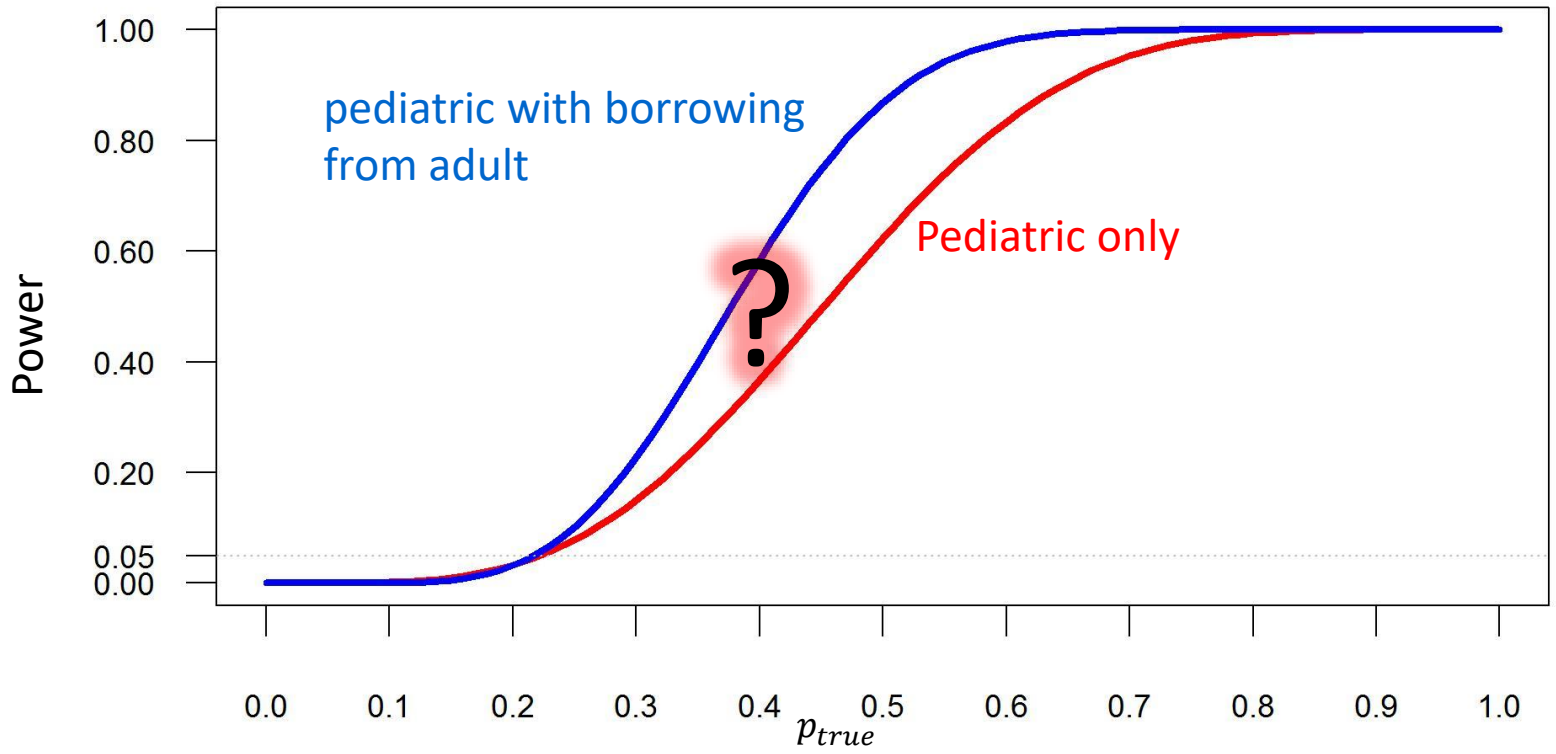


Borrowing from adult information for the pediatric arm

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 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.

Result from adult trial:

$r_{adu} = 12$ among $n_{adu} = 40$

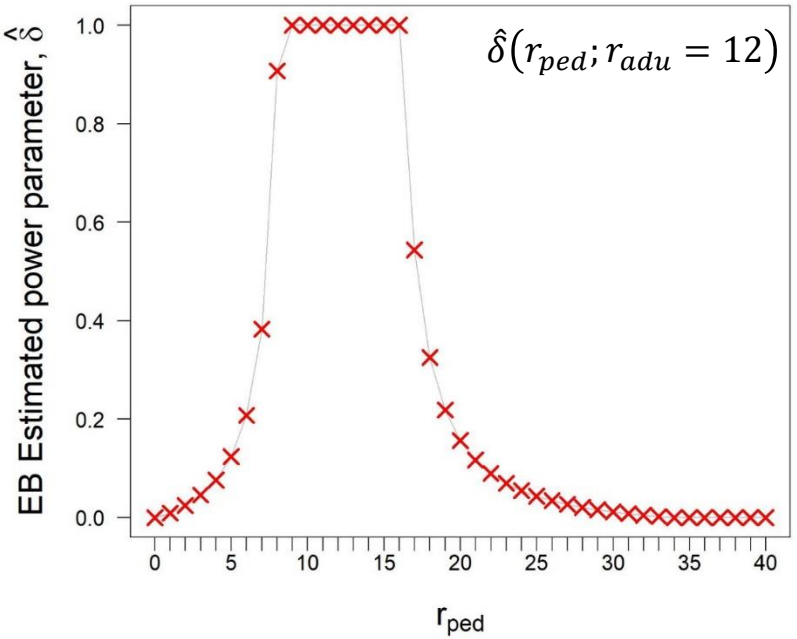
$(\hat{p}_{adu} = 0.3)$

estimate

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

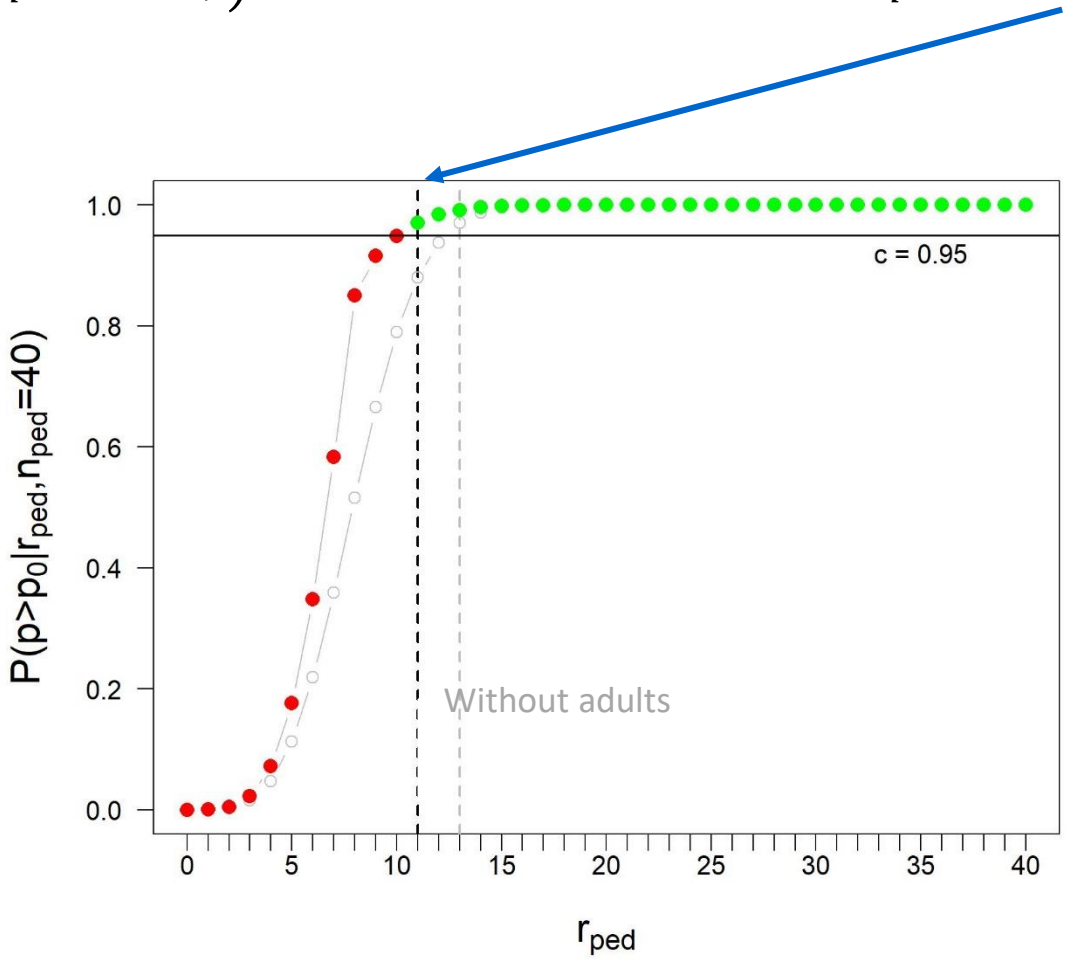
Empirical Bayes approach

(Gravestock, Held et al. 2017)



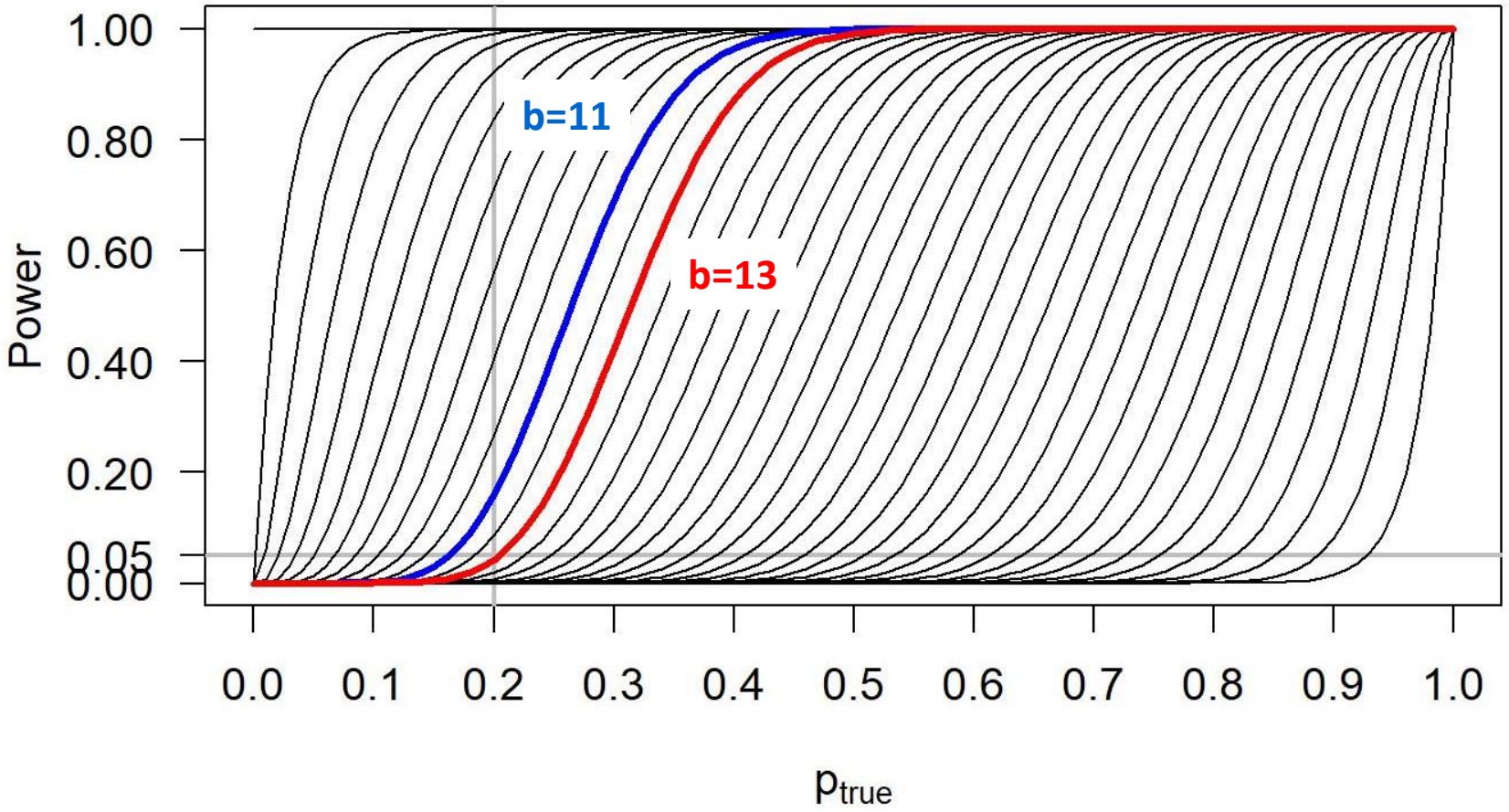
Adaptive power parameter (2)

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



Adaptive power parameter (3)

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

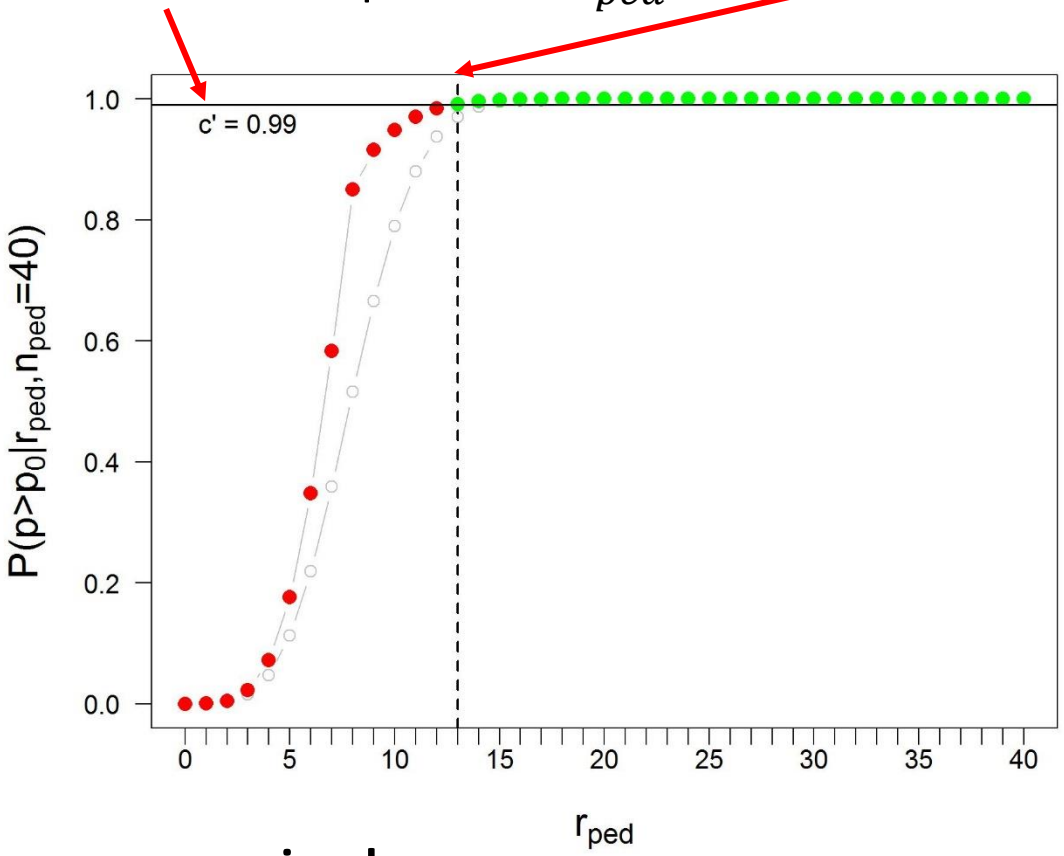


→ **power gain but type I error inflation**

Adaptive power parameter (4)

$P\left(p > p_0 | r_{ped}, r_{adu}, \hat{\delta}\left(r_{ped}, r_{adu}\right)\right)$ 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**

Borrowing from adult information in the one-sided test (1)

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

then there exists c' with

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

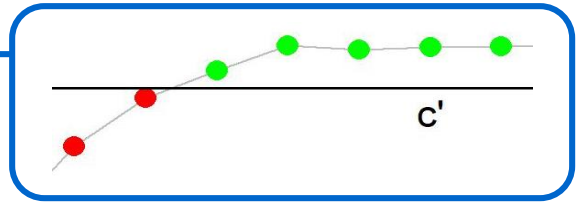
and $b_{\text{UMP}}(\alpha)$ is the level α UMP test boundary.

Borrowing from adult information in the one-sided test (2)

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

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

  can still be identified.



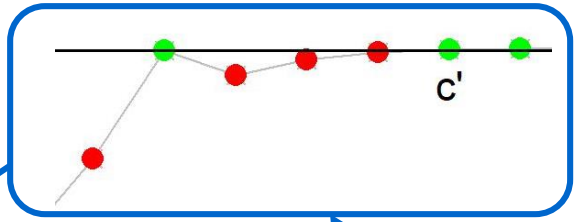
- if no c' with **(∗)** can be identified, then either the

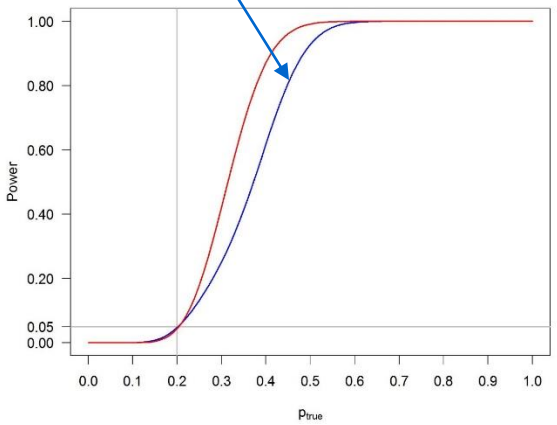
  - test does not control type I error

    or

  - test controls type I error but is not UMP.

  → The trial may be considered a success for r_{ped} responses and a failure for one more pediatric response (r_{ped} + 1)





Borrowing from adult information in the two-sided test (1)

Consider two-sided test situation

$$H_0\colon p = p_0 \text{ vs. } H_1\colon p \neq p_0\ ,\ p_0 = 0.5$$

with $n_{ped} = 100$ and $\alpha = 0.05$.

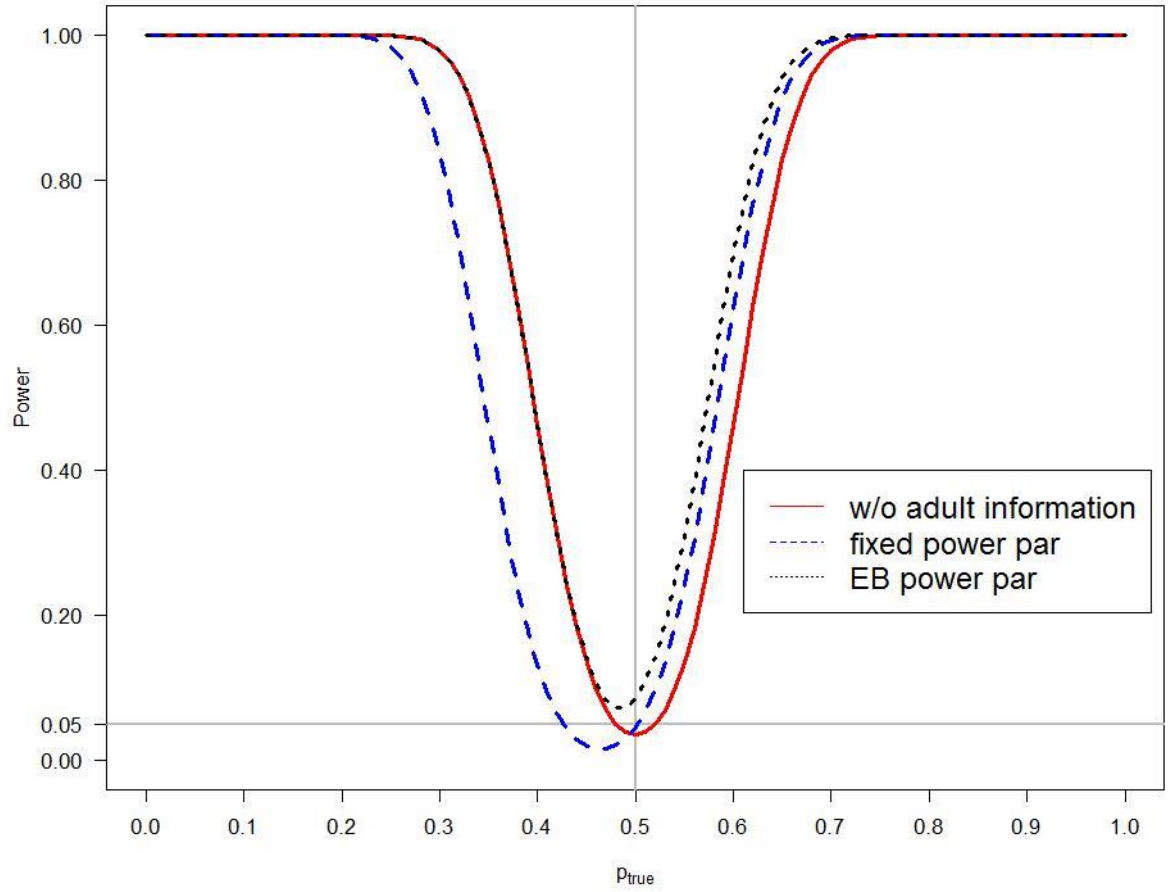
- Without adults: accept H_0 if $r_{ped} \in \{40, \ldots, 60\}$
  equivalent to $P(p > p_0 | r_{ped}) \in [0.022, 1 - 0.022]$.
- Borrowing from $r_{adu} = 57$ responders among $n_{adu} = 100$ adults:

  - using power parameter approach with fixed $\delta = 0.5$:
    accept H_0 if $r_{ped} \in \{35, \ldots, 58\}$.

  - using EB power parameter approach:
    accept H_0 if $r_{ped} \in \{40, \ldots, 58\}$.

Borrowing from adult information in the two-sided test (2)

Power curves:



- Fixed borrowing: type I error controlled but biased test, power increased for $p > p_0$, power decreased for $p < p_0$
- Adaptive borrowing: type I error inflated, power increased for $p > p_0$, but mostly identical for $p < p_0$

Summary

View decision rule 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 groups 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,…)
- Proven by Psioda and Ibrahim (2018) for one-sample one-sided normal test with borrowing using a fixed power prior.

Conclusion

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

- However, if prior information is reliable and consistent with the new 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).

  $\longrightarrow$ Incorporation of prior information can give a rationale for type I error inflation with benefit of a power gain.

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, Wiesenfarth M, Witt R, Edelmann D, Witt O, Abel U. (2019) Monitoring futility and efficacy in phase II trials with Bayesian posterior distributions – A calibration approach. *Biometrical Journal* 61(3): 488-502.
- Kopp-Schneider A, Calderazzo S, Wiesenfarth M. Power gains by using external information in clinical trials are typically not possible when requiring strict type I error control. *Biometrical Journal* (accepted).
- Psioda MA, Ibrahim JG (2018) Bayesian clinical trial design using historical data that inform the treatment effect. *Biostatistics* online.