



# Beyond the Classical Type I Error: Bayesian Metrics for Bayesian Designs Using Informative Priors

Nicky Best (GSK)

Acknowledgements: Simon Wandel (Novartis), Gaelle Saint-Hillary (Saryga), Maxine Ajimi (AstraZeneca), Beat Neuenschwander (Novartis)

# Context

Growing interest in use of Bayesian trial designs with informative priors in regulatory settings



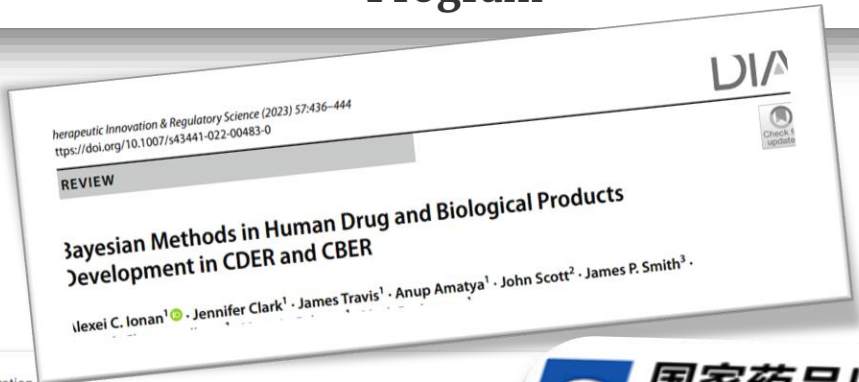
FDA U.S. FOOD & DRUG ADMINISTRATION  
← Home / Drugs / Development & Approval Process | Drugs / Development Resources / Complex Innovative Trial Design Meeting Program

## Complex Innovative Trial Design Meeting Program



ICH  
harmonisation for better health

### ICH E11A Guideline on pediatric extrapolation



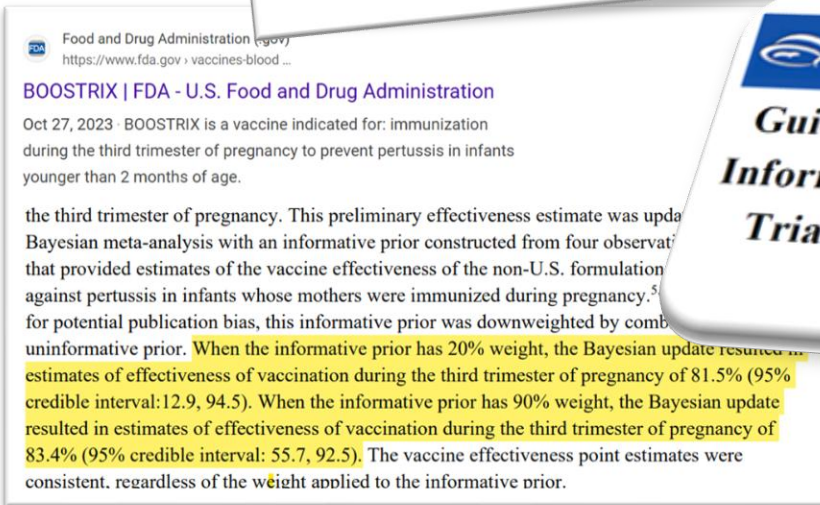
DIA

therapeutic Innovation & Regulatory Science (2023) 57:436–444  
<https://doi.org/10.1007/s43441-022-00483-0>

REVIEW

### Bayesian Methods in Human Drug and Biological Products Development in CDER and CBER

Alexei C. Ionan<sup>1</sup> · Jennifer Clark<sup>1</sup> · James Travis<sup>1</sup> · Anup Amatya<sup>1</sup> · John Scott<sup>2</sup> · James P. Smith<sup>3</sup>

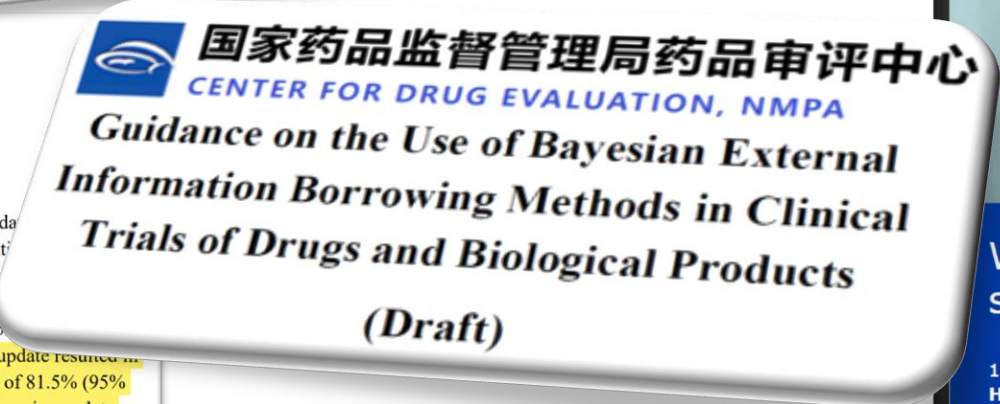


Food and Drug Administration (FDA)  
<https://www.fda.gov/vaccines-blood...>

### BOOSTRIX | FDA - U.S. Food and Drug Administration

Oct 27, 2023 · BOOSTRIX is a vaccine indicated for: immunization during the third trimester of pregnancy to prevent pertussis in infants younger than 2 months of age.

the third trimester of pregnancy. This preliminary effectiveness estimate was updated by a Bayesian meta-analysis with an informative prior constructed from four observational studies that provided estimates of the vaccine effectiveness of the non-U.S. formulation against pertussis in infants whose mothers were immunized during pregnancy.<sup>5</sup> To account for potential publication bias, this informative prior was downweighted by combining it with an uninformative prior. When the informative prior has 20% weight, the Bayesian update resulted in estimates of effectiveness of vaccination during the third trimester of pregnancy of 81.5% (95% credible interval: 12.9, 94.5). When the informative prior has 90% weight, the Bayesian update resulted in estimates of effectiveness of vaccination during the third trimester of pregnancy of 83.4% (95% credible interval: 55.7, 92.5). The vaccine effectiveness point estimates were consistent, regardless of the weight applied to the informative prior.



国家药品监督管理局药品审评中心  
CENTER FOR DRUG EVALUATION, NMPA

### Guidance on the Use of Bayesian External Information Borrowing Methods in Clinical Trials of Drugs and Biological Products (Draft)



EUROPEAN MEDICINES AGENCY  
SCIENCE · MEDICINES · HEALTH

### Workshop on the use of Bayesian statistics in clinical development

17 June 2025 , 09:00 – 17:00 (CET/CEST)  
Hybrid Meeting - Virtual/ EMA, Amsterdam, Room 1D



GSK

## Assessing risk of erroneous conclusions when borrowing prior information

In a regulatory setting, most trials are designed to control Type 1 error at 5% and select the decision rule and sample size to minimise Type 2 error (typically  $\leq 10\%$ )

- “Conventionally the probability of type I error is set at 5% ..... the precise choice may be influenced by the **prior plausibility of the hypothesis** under test and the desired impact of the results” ICH E9 (1998)
- “When the sample size is limited, the relative importance of false positive and false negative error rates **may be modified** from convention” ICH E11A (2025)
- Strict control of type 1 error is **not possible** when leveraging prior information in a Bayesian trial design (Kopp Schneider et al 2020)
- Where reliable, relevant prior information is available, to disregard this in the name of ensuring strict Type I error control **is questionable** (Psioda & Xue 2020)
- When Type I error probability is not applicable (e.g., some Bayesian designs that borrow external information), **appropriate alternative trial characteristics** should be considered FDA Guidance on Complex Innovative Designs (2020)

## Focus for this talk

We present several alternative **Bayesian (fully probabilistic) metrics** to evaluate the risk of a Bayesian trial producing **false positive conclusions**

Focus on 2 scenarios:

1. Designs borrowing prior information on the **treatment contrast**
2. Designs borrowing prior information on the **control arm**

[Full article: Best et al \(2024\) Beyond the Classical Type I Error: Bayesian Metrics for Bayesian Designs Using Informative Priors](#)



# 1. Borrowing Information on the Treatment Contrast

**GSK**

## Motivating setting: 2-arm RCT with informative prior on treatment contrast

### Notation

- $\theta_t$  = true treatment effect on active arm;  $\theta_c$  = true treatment effect on control arm
- $\delta = \theta_t - \theta_c$  = treatment contrast
- $p_A(\delta)$  = prior information on the treatment contrast
- Success rule:  $Pr(\delta > 0 \mid y, p_A(\delta)) \geq 97.5\%$
- Observed data (treatment contrast) in new trial:  $y \sim f(\delta, \sigma_n^2)$
- $\hat{\delta} = E(\delta \mid y, p_A(\delta))$  = posterior mean treatment contrast

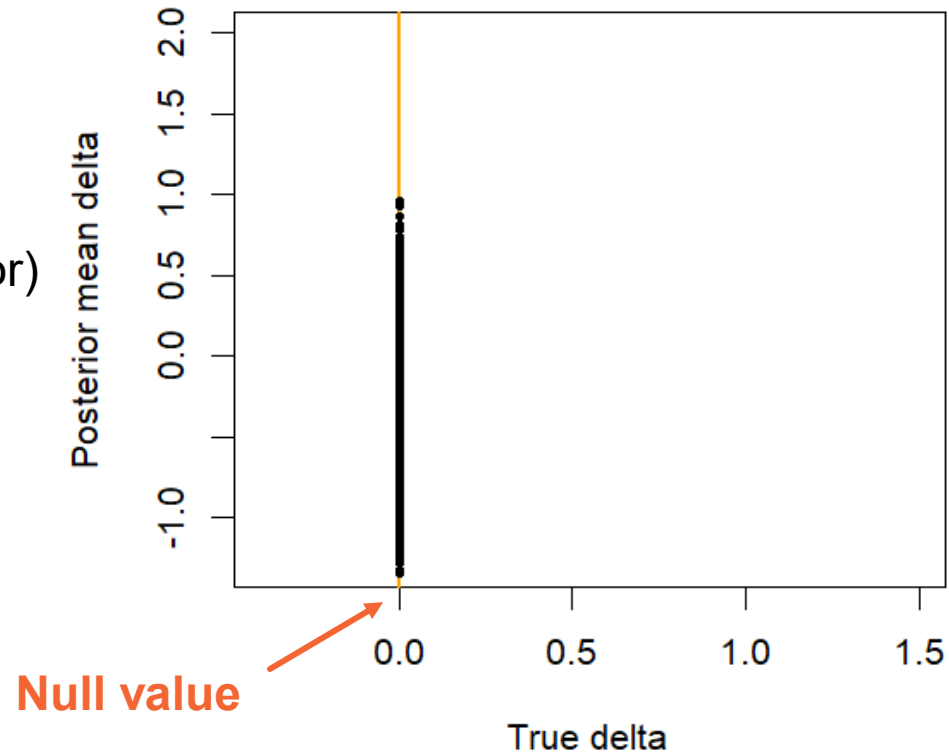
At trial design stage, can compute the **pre-posterior distribution** of  $\hat{\delta}$  (or any function of the posterior for  $\delta$ ) conditional on the assumed true value of  $\delta$  that is generating the data ( $\sigma_n^2$  assumed known for design)

# Common metrics for evaluating clinical trial designs

$$y \sim f(\delta, \sigma_n^2)$$

$$\hat{\delta} = E(\delta | y, p_A(\delta))$$

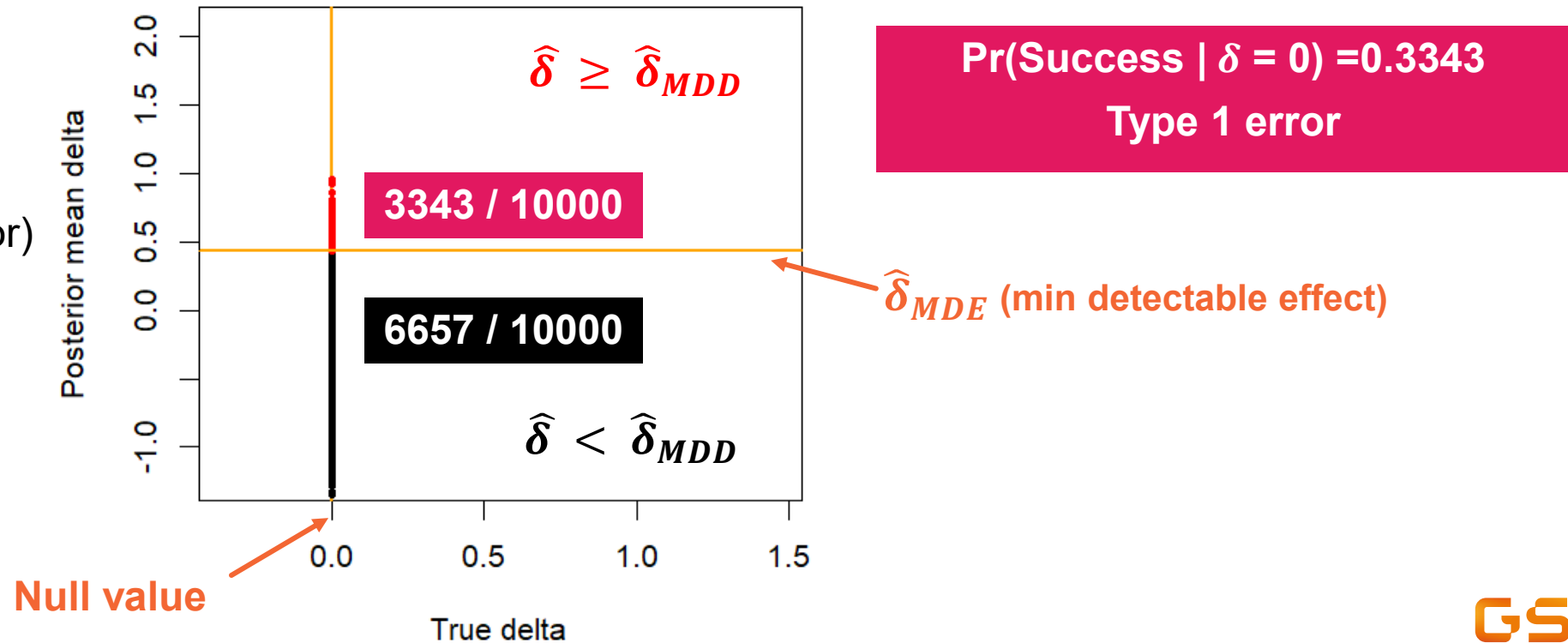
10000 samples from  
sampling (pre-posterior)  
distribution of  $\hat{\delta}$  given  
true  $\delta=0$



# Common metrics for evaluating clinical trial designs

Metric		Comments
Pr(Success   Truth = null)	Type 1 error	Hypothetical probability of making future decision <b>given fixed truth</b>

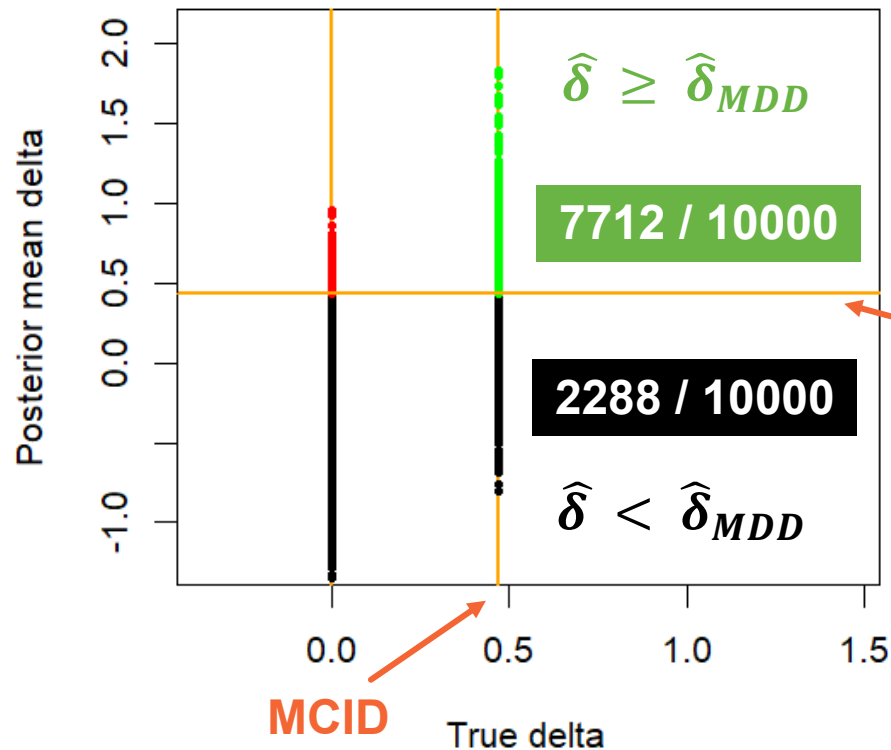
$y \sim f(\delta, \sigma_n^2)$   
 $\hat{\delta} = E(\delta | y, p_A(\delta))$   
 10000 samples from sampling (pre-posterior) distribution of  $\hat{\delta}$  given true  $\delta=0$



# Common metrics for evaluating clinical trial designs

Metric		Comments
$\Pr(\text{Success} \mid \text{Truth} = \text{null})$	Type 1 error	Hypothetical probability of making future decision <b>given fixed truth</b>
$\Pr(\text{Success} \mid \text{Truth} = \text{MCID})$	Power	Hypothetical probability of making future decision <b>given fixed truth</b>

$y \sim f(\delta, \sigma_n^2)$   
 $\hat{\delta} = E(\delta \mid y, p_A(\delta))$   
 10000 samples from  
 sampling distribution of  
 $\hat{\delta}$  given true  $\delta = \text{MCID}$



$\Pr(\text{Success} \mid \delta = \text{MCID}) = 0.7712$   
**Power**

$\hat{\delta}_{MDE}$  (min detectable effect)

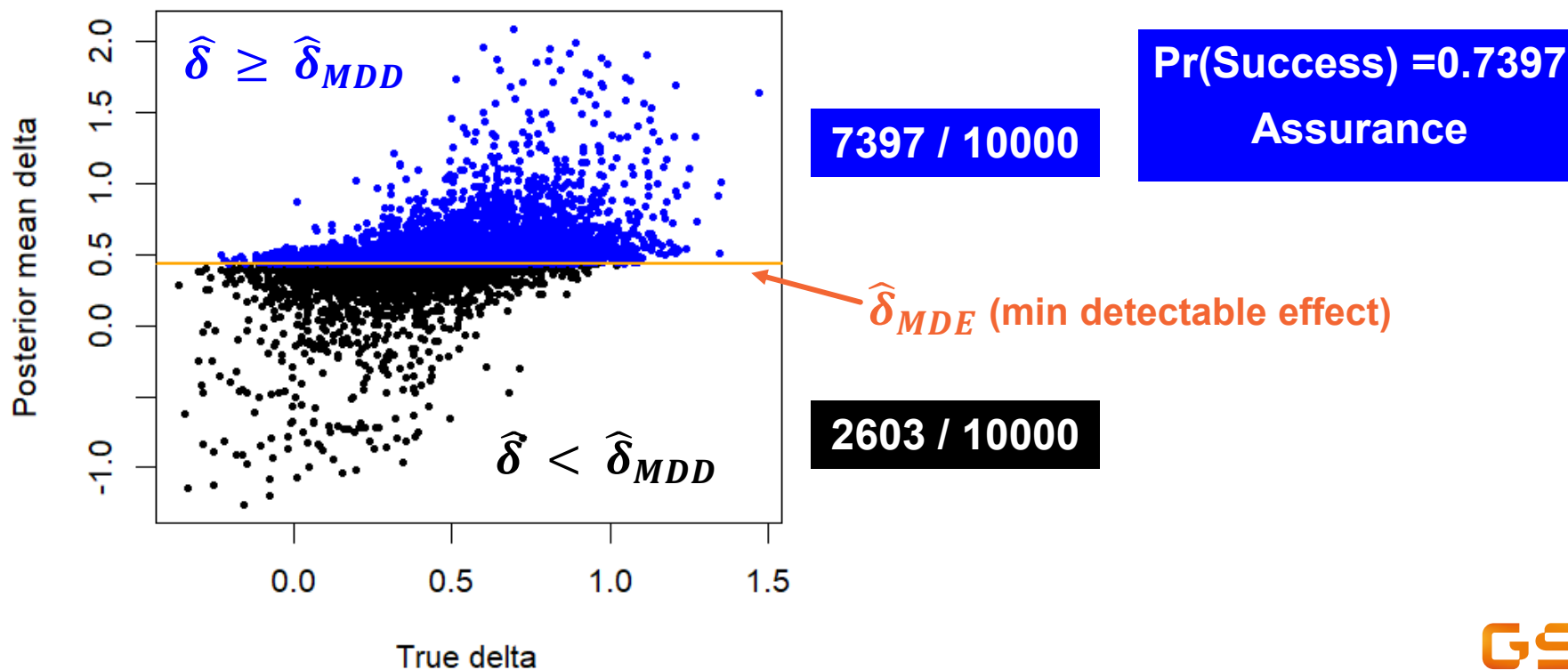
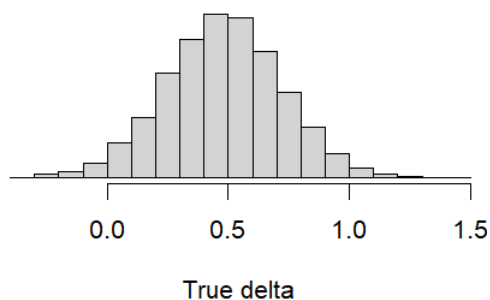
# Common metrics for evaluating clinical trial designs

Metric		Comments
$\Pr(\text{Success} \mid \text{Truth} = \text{null})$	Type 1 error	Hypothetical probability of making future decision <b>given fixed truth</b>
$\Pr(\text{Success} \mid \text{Truth} = \text{MCID})$	Power	Hypothetical probability of making future decision <b>given fixed truth</b>
$\Pr(\text{Success})$	Assurance	<b>(Unconditional) predicted probability</b> of making a positive decision Requires specification of a <b>design prior</b>

$$y \sim f(\delta, \sigma_n^2)$$

$$\hat{\delta} = E(\delta \mid y, p_A(\delta))$$

10000 samples from sampling (pre-posterior) distribution of  $\hat{\delta}$  given true  $\delta$  drawn from "design" prior  $p_D(\delta)$ :



# Common metrics for evaluating clinical trial designs

Metric		Comments
$\Pr(\text{Success} \mid \text{Truth} = \text{null})$	Type 1 error	Hypothetical probability of making future decision <b>given fixed truth</b>
$\Pr(\text{Success} \mid \text{Truth} = \text{MCID})$	Power	Hypothetical probability of making future decision <b>given fixed truth</b>
$\Pr(\text{Success})$	Assurance	<b>(Unconditional) predicted probability</b> of making a positive decision Requires specification of a <b>design prior</b>

All are special cases of the following metric:

$$M_1 = \int \Pr(\text{Study Success} \mid \delta) p(\delta) d\delta$$

- **Type 1 error:**  $p(\delta) =$  Point mass prior at  $\delta_{null}$
- **Power:**  $p(\delta) =$  Point mass prior at  $\delta_{MCID}$
- **Assurance:**  $p_D(\delta) =$  **design prior** reflecting our uncertainty around hypothesized treatment effect

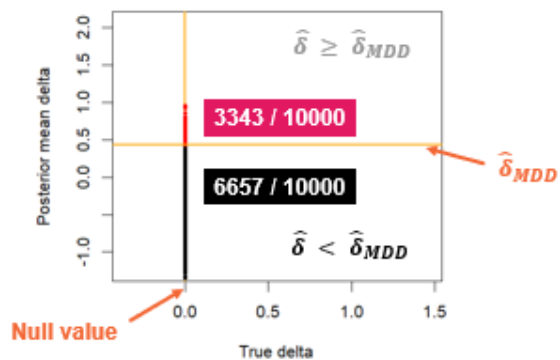
## Design (sampling) priors

**Analysis prior,  $p_A(\delta)$ :** pre-specified analysis prior for treatment contrast parameter

**Design prior ,  $p_D(\delta)$ :** design (or sampling) prior

- Mechanism for generating data scenarios to evaluate operating characteristics of trial designs
- Reflects uncertainty about the most plausible parameter values to assume *for evaluating design*
  - Shifts focus away from ‘worst’ or ‘best’ case scenarios to assessment of predictive probabilities of expected trial outcomes
  - Acknowledges that some parameter values (scenarios) are of more interest than others
- If  $p_D(\delta) \neq p_A(\delta)$  → can be used to judge accuracy of decisions in scenarios which differ from those that informed the analysis prior

# Assessing risk of false positives when borrowing prior information



## Psioda & Ibrahim (2019)

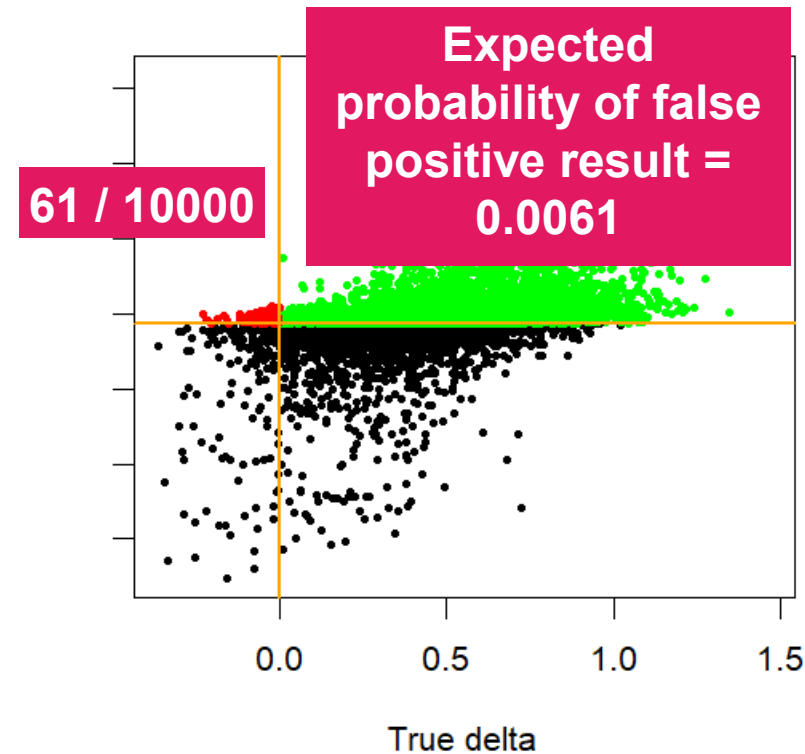
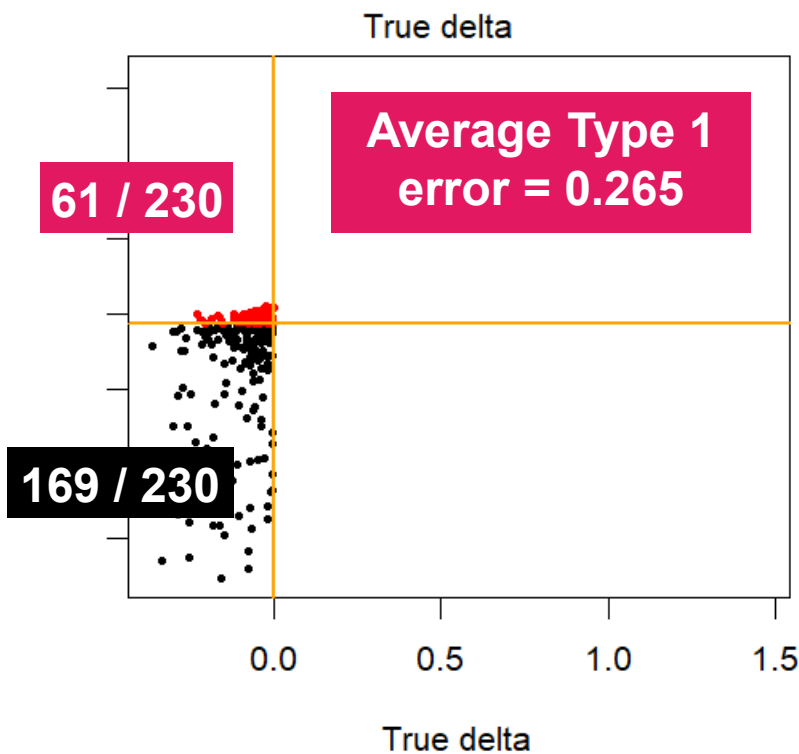
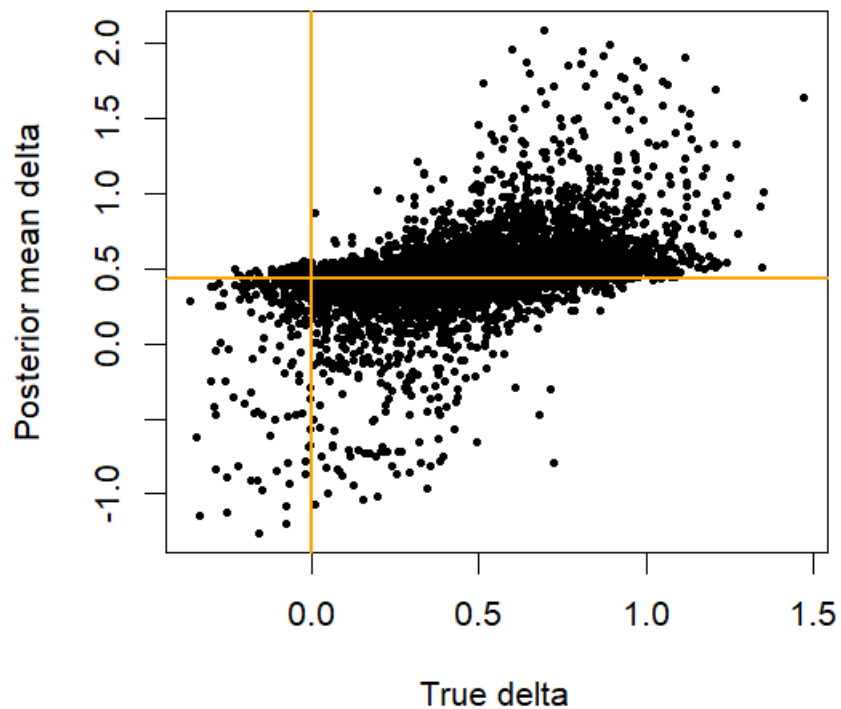
Null design prior

True delta

## Best et al (2024)

Full design prior

True delta



# Assessing risk of false positives when borrowing prior information

Average Type 1 error:

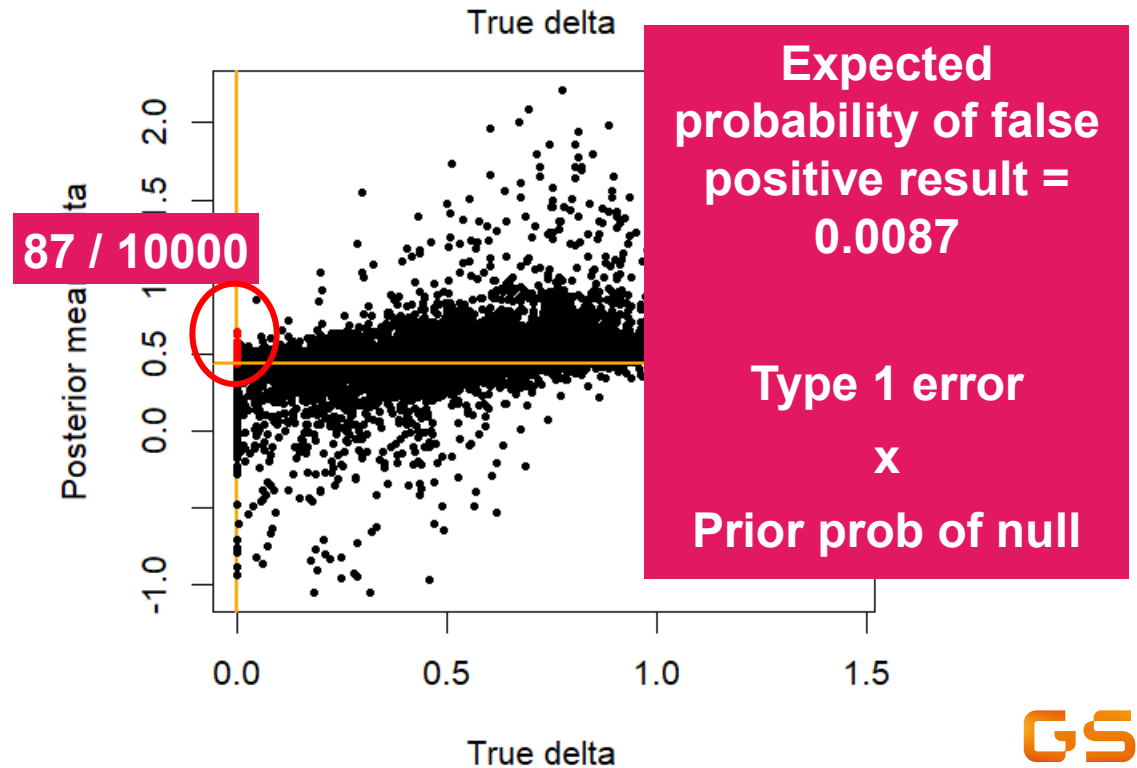
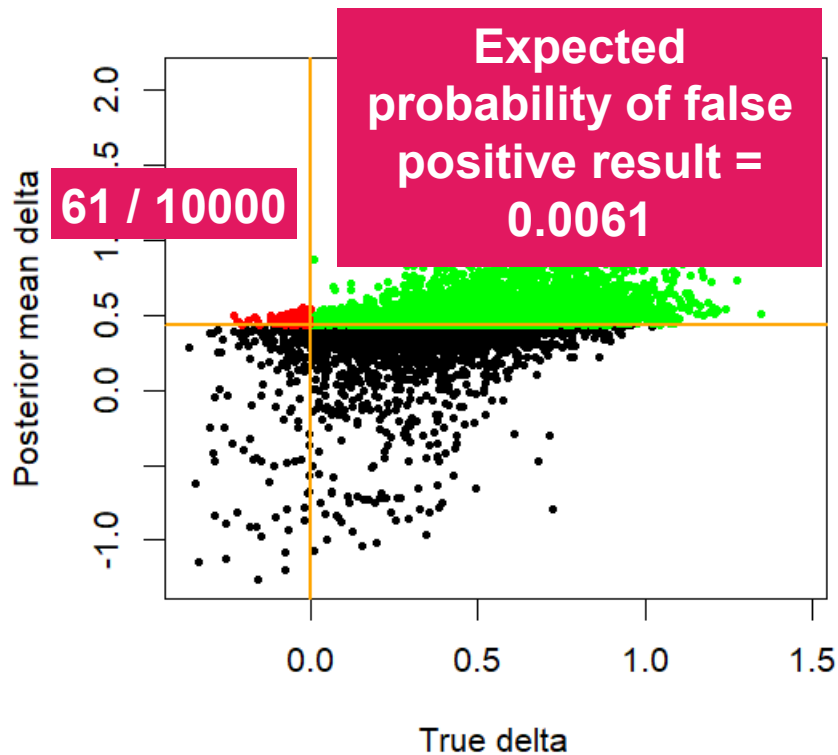
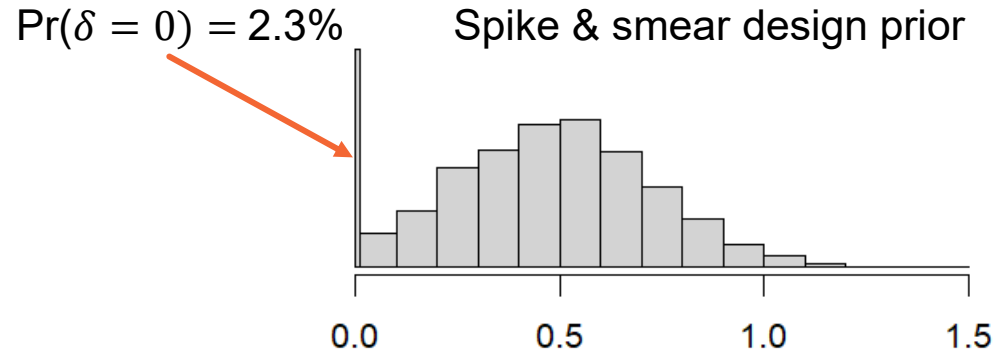
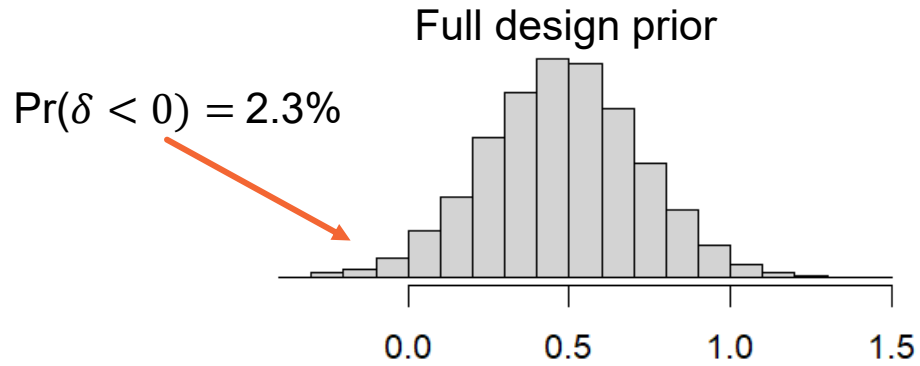
$$M_2 = \int \Pr(\textit{Study Success}|\delta) \frac{p(\delta)I\{\delta \leq \delta_{null}\}}{\Pr(\delta \leq \delta_{null})} d\delta$$

Expected probability of false positive:

$$M_3 = \int_{\delta \leq \delta_{null}} \Pr(\textit{Study Success}|\delta) p(\delta) d\delta$$

$$= \underbrace{\int \Pr(\textit{Study Success}|\delta) \frac{p(\delta)I\{\delta \leq \delta_{null}\}}{\Pr(\delta \leq \delta_{null})} d\delta}_{\textit{Average type 1 error under null (truncated) design prior}} \times \underbrace{\Pr(\delta \leq \delta_{null})}_{\textit{Prob treatment effect is null or harmful}}$$

# Assessing risk of false positives when borrowing prior information





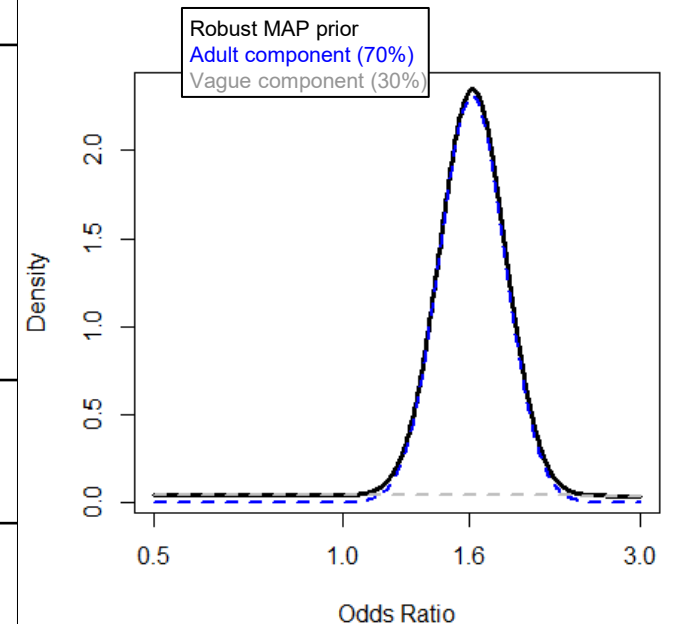
# Case Study: Borrowing Information on the Treatment Difference



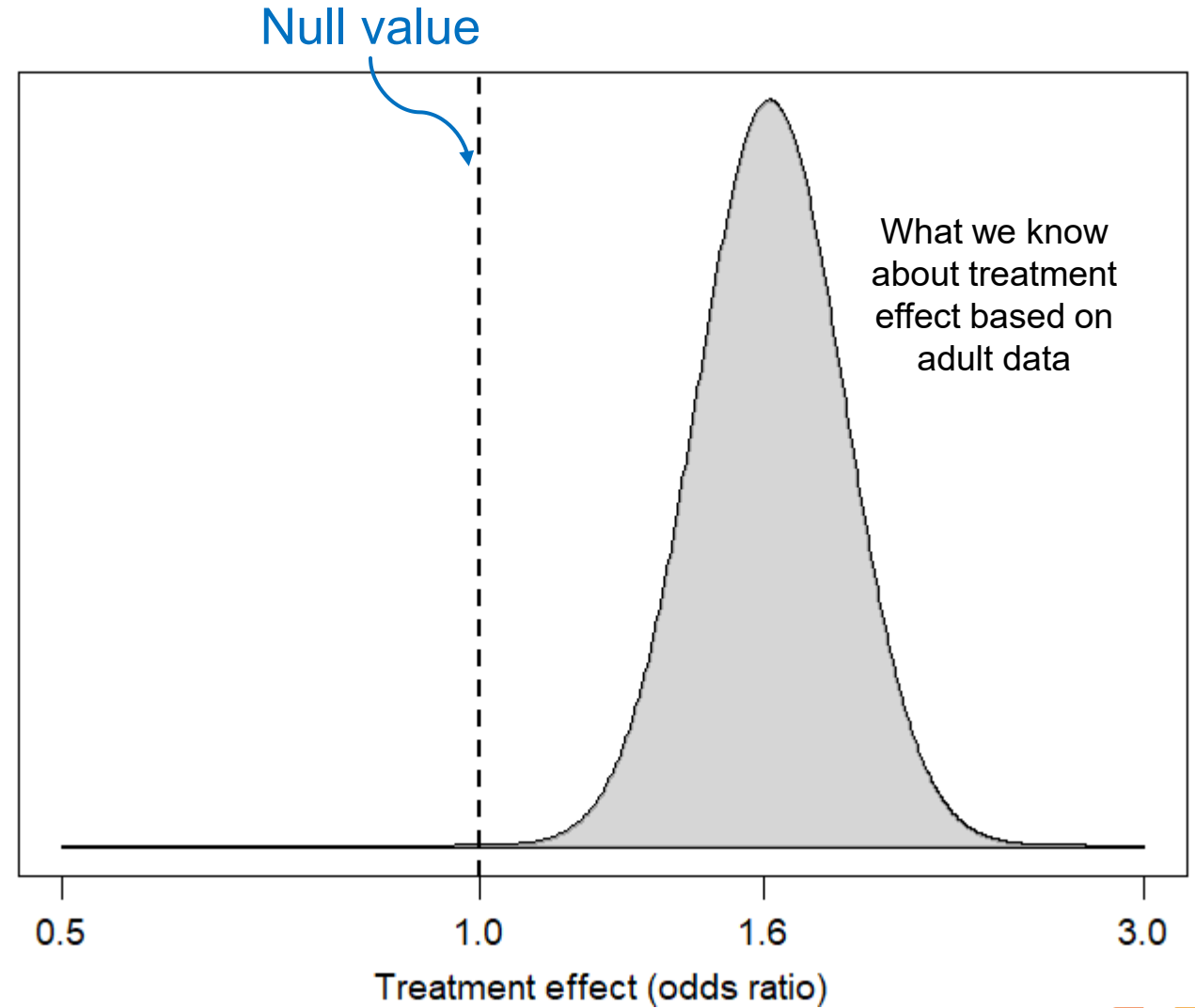
# Paediatric bridging study with partial extrapolation from adults

- Paediatric study: Double-blind RCT, experimental vs placebo
- Primary endpoint: Disease activity responder index
  - summarized by Odds Ratio for active v placebo (assumed Normally distributed on log scale)
- Historical adult data from same sponsor: Pooled efficacy results from 2 pivotal Ph3 studies

<b>Sample size in proposed paediatric study</b>	<b>50 active; 50 placebo</b>
<b>Prior distribution on true paediatric OR</b>	<b>Robust mixture prior:</b> weighted mixture of posterior distribution of treatment effect (log OR) from <b>adult study</b> and <b>vague distribution</b> centered on log OR = 0
<b>Weight on adult component of the mixture prior</b>	70%
<b>Success rule defining positive result</b>	$\Pr(\text{OR} > 1 \mid \text{paed data; adult prior}) \geq 97.5\%$



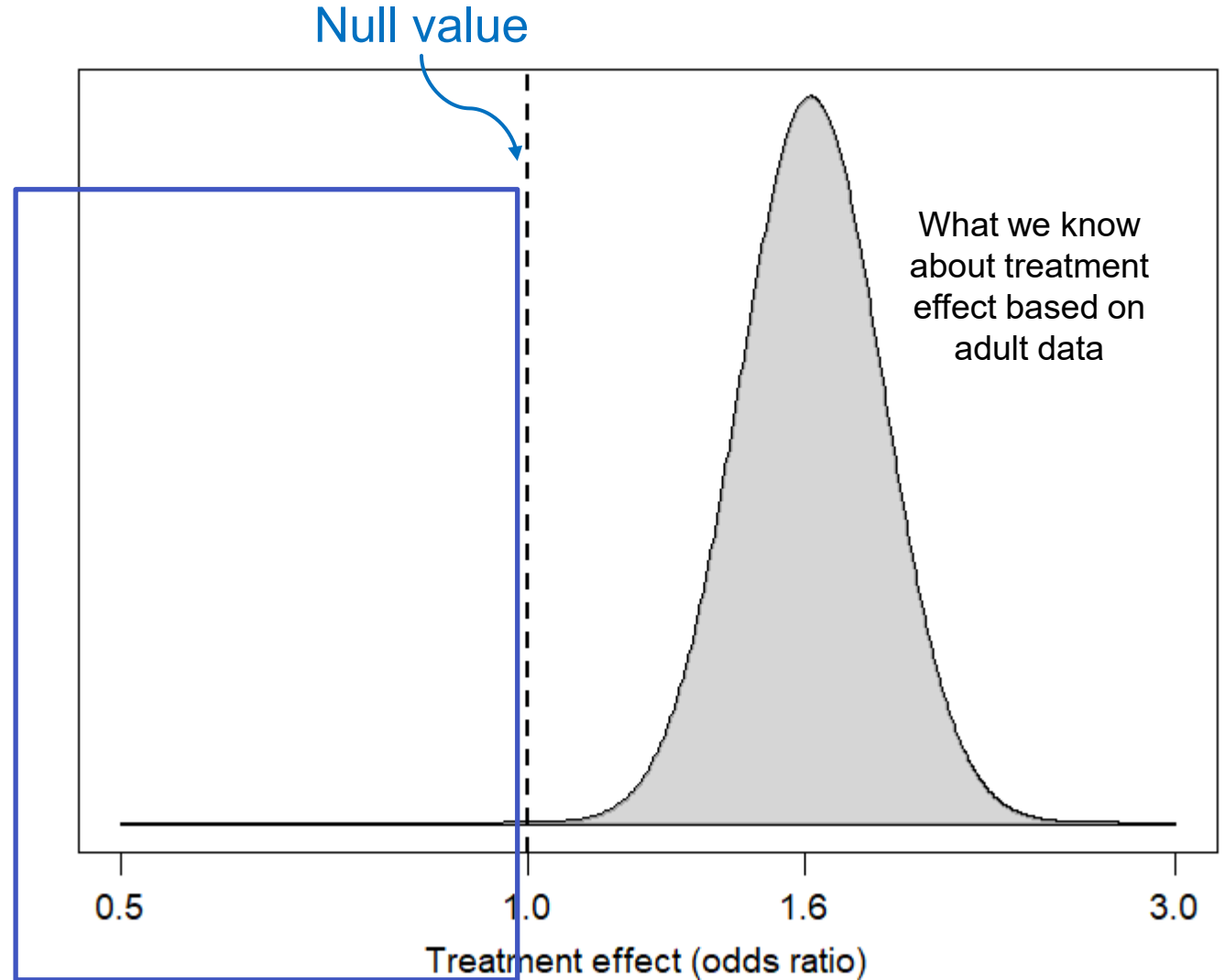
## Assessing risk of false positive result in paediatric study



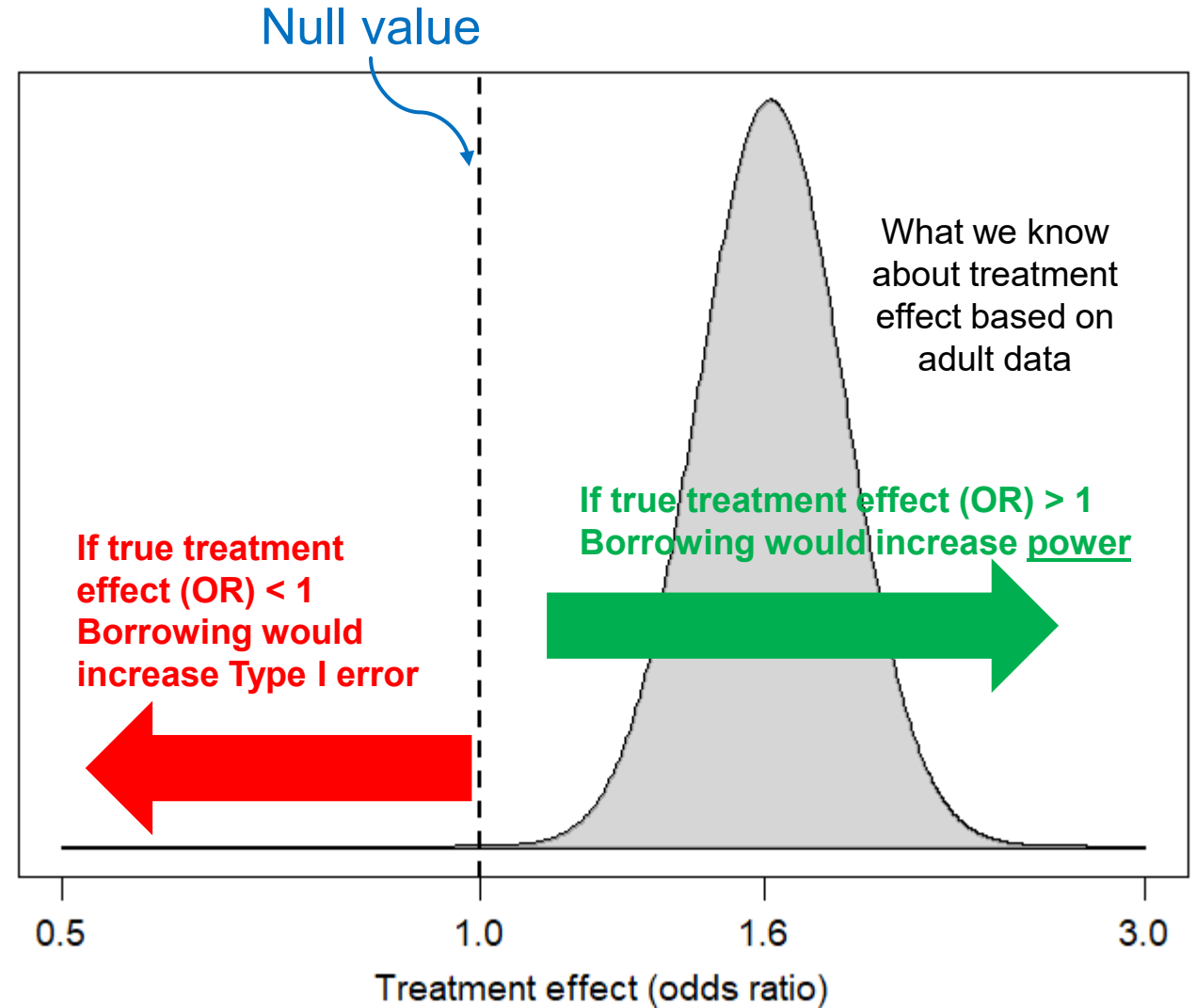
# Assessing risk of false positive result in paediatric study

**Null or harmful treatment effect**

**Very low probability based on prior evidence** (assuming paediatric effect is exchangeable with adult effect)

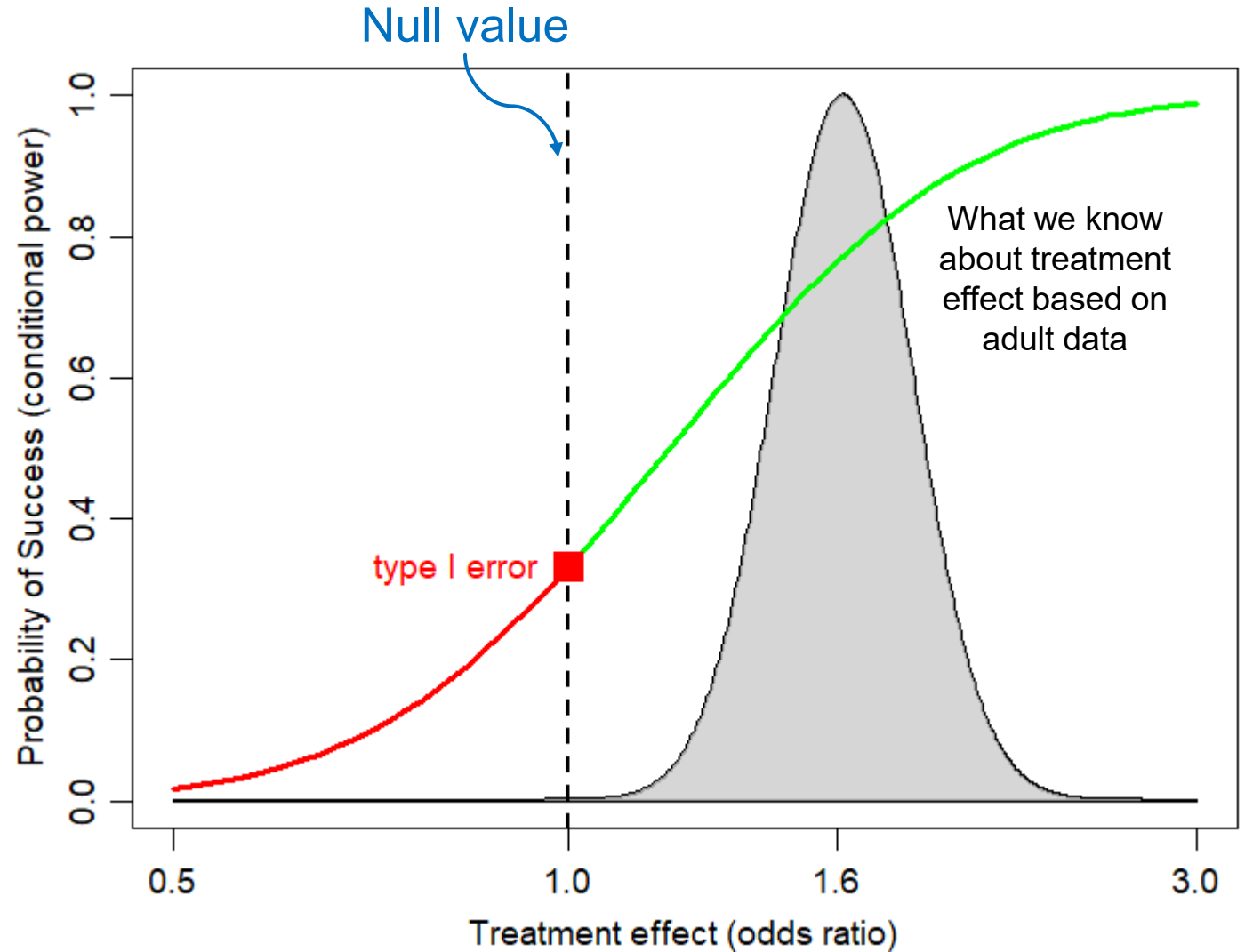


# Assessing risk of false positive result in paediatric study



## Assessing risk of false positive result in paediatric study

Type I error occurs at an improbable value of the treatment effect



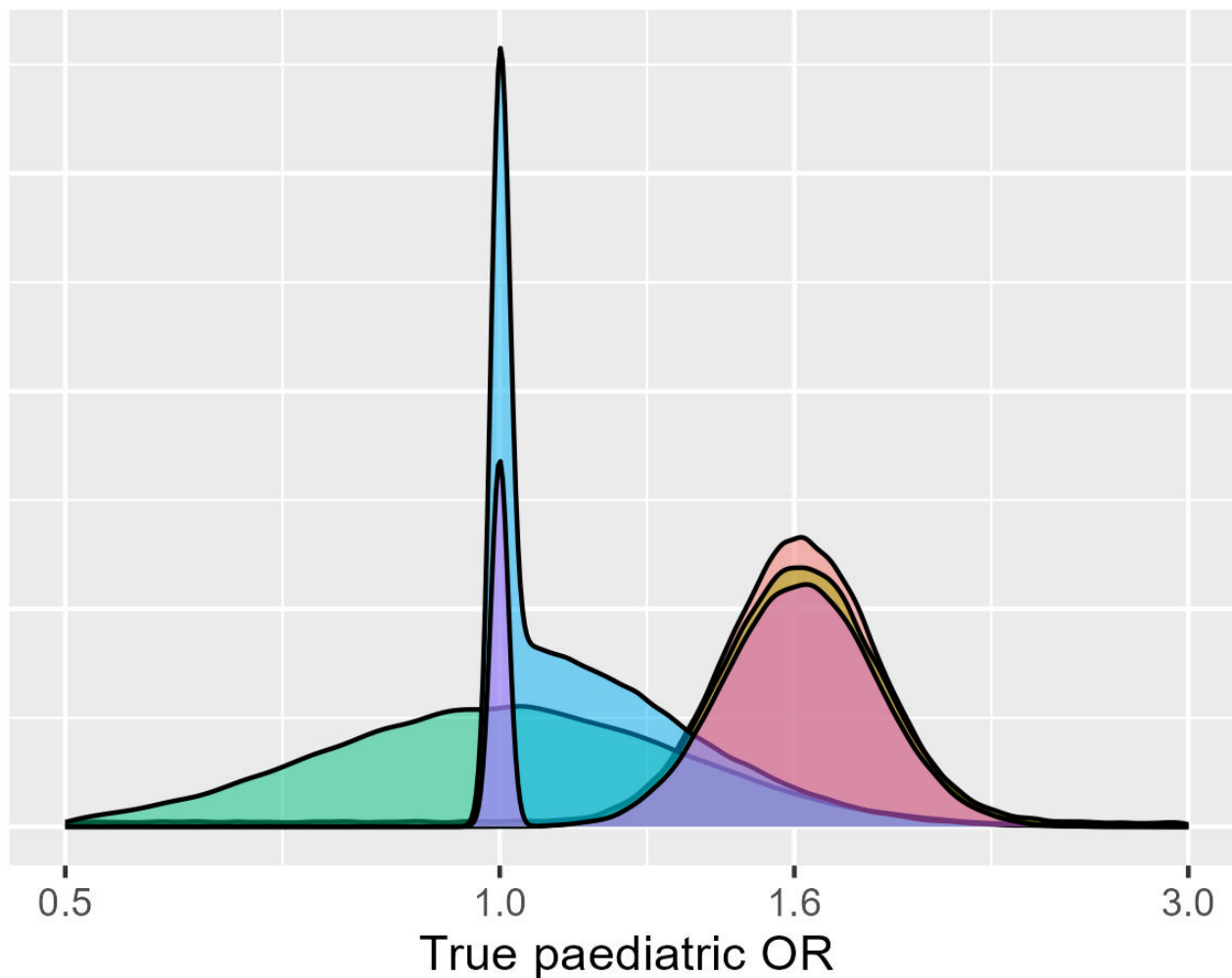
## Choosing a design prior

Recall: Analysis prior is robust mixture  $p_A(\delta) = 0.7 p_{adult} + 0.3 p_{vague}$

### Choosing $p_D(\delta)$

- If solid agreement between stakeholders about prior, then may be sufficient to choose  $p_D(\delta) = p_A(\delta)$
- In general, consider various  $p_D(\delta) \neq p_A(\delta)$  to assess how likely it is for accuracy of conclusions to be below acceptable level
- $p_D(\delta)$  often more skeptical than  $p_A(\delta)$  – can be based on:
  - Data, e.g. select least favourable previous trial, or shift mean downwards
  - Expert elicitation
  - “Reference skeptical prior” (Spiegelhalter et al 1994), e.g. mean 0, small prob of  $\delta > \text{MCID}$

# Design priors for paediatric example



## Design prior

- Adult
- RMP
- Skeptical 1
- Skeptical 2
- RMP Spike

Adult prior

Robust mixture

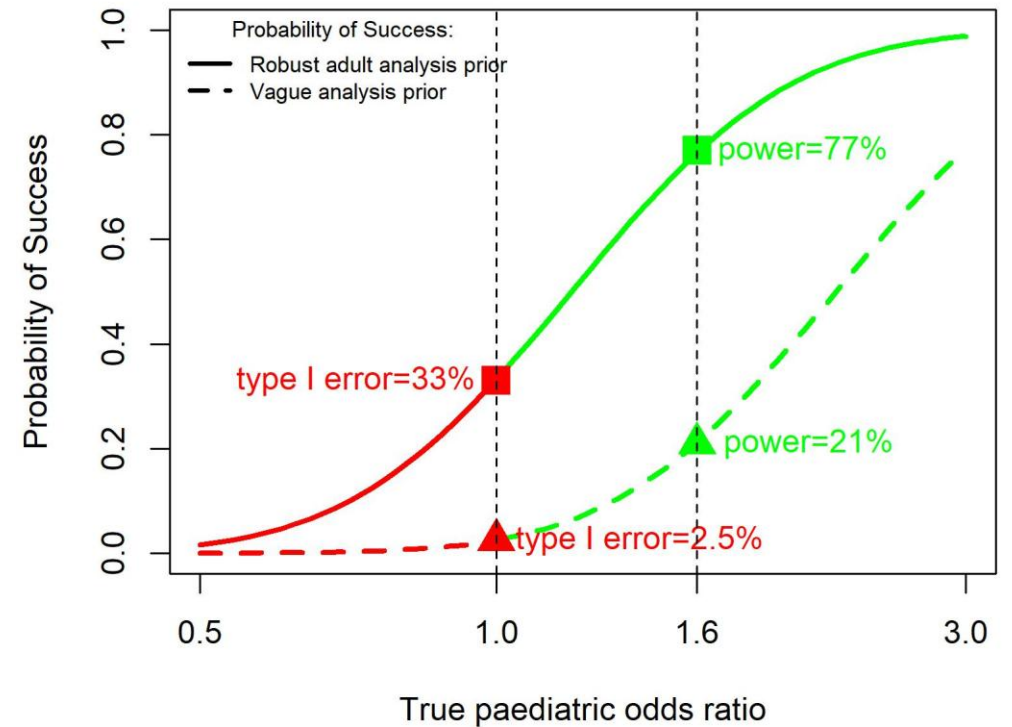
Mean 0,  $\Pr(\text{OR} > 1.6) = 0.05$

$\Pr(\text{OR}=1) = 0.3$ ,  $\Pr(\text{OR} > 1.6) = 0.05$

$\Pr(\text{OR}=1) = 0.15$ , RMP

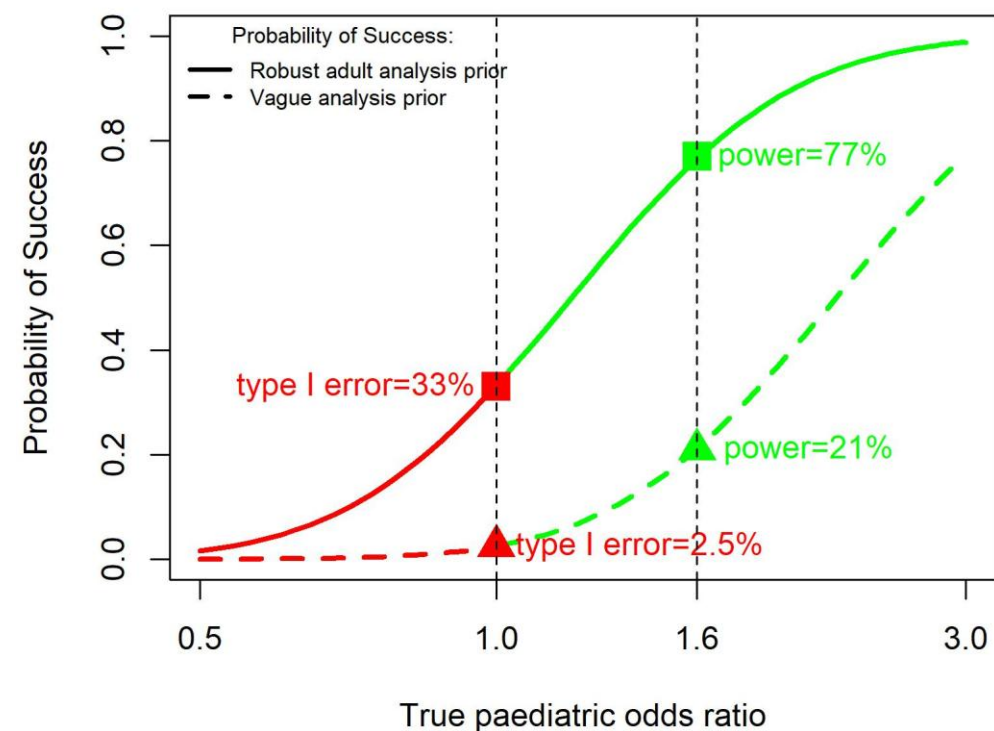
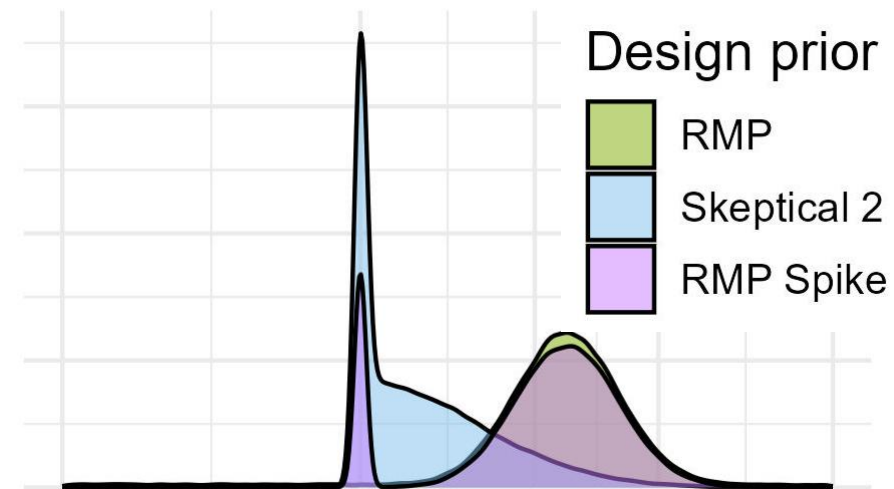
# Illustration of proposed error metrics for paediatric study

Metric	Design prior	Analysis prior	
		RMP	Vague
Type 1 error	Point mass at OR=1	33%	2.5%



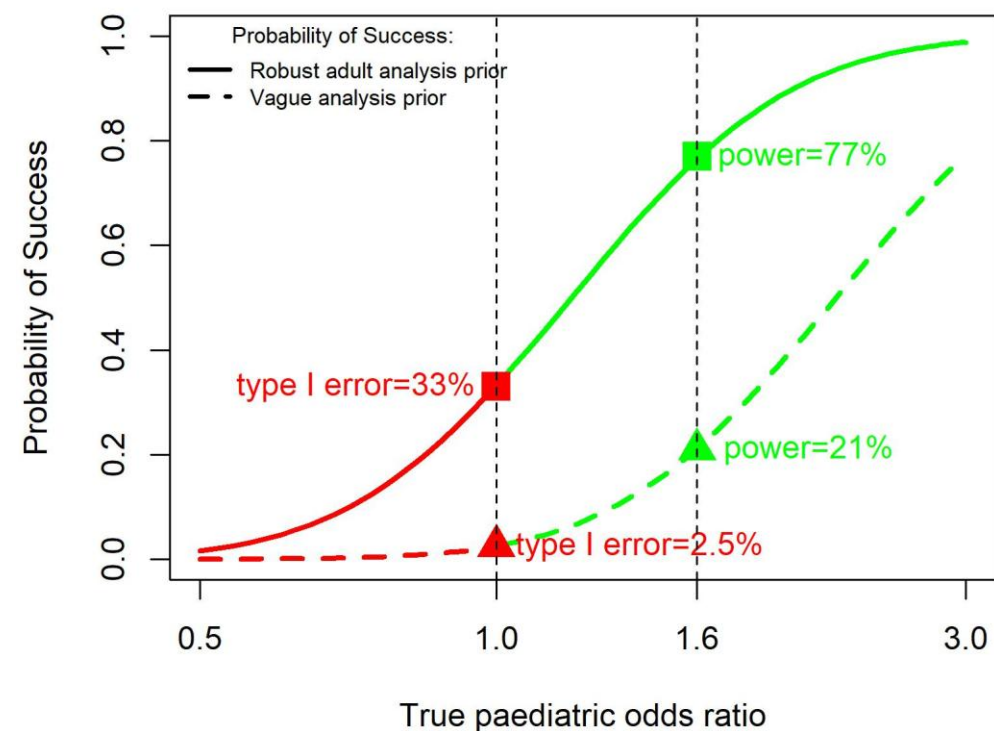
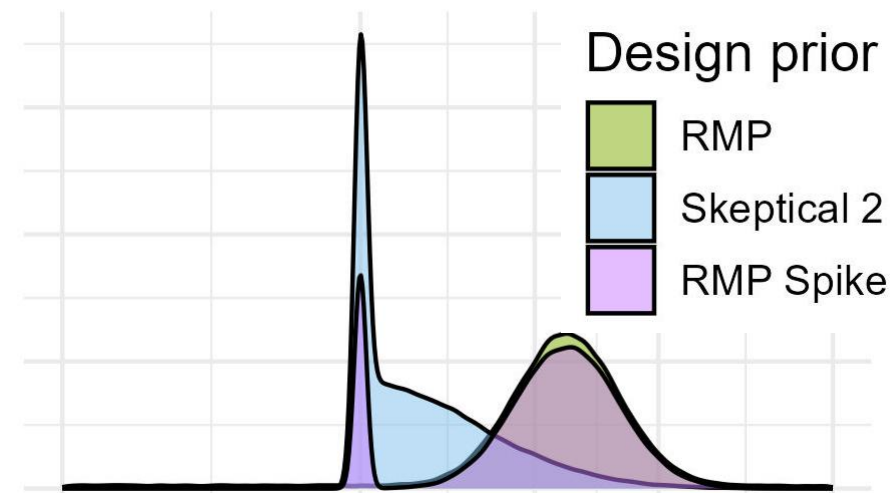
# Illustration of proposed error metrics for paediatric study

Metric	Design prior	Analysis prior	
		RMP	Vague
Type 1 error	Point mass at OR=1	33%	2.5%
Design prior probability of no treatment benefit: Pr(True OR $\leq$ 1)	Robust mixture	15%	
	Skeptical 2	30%	
	RMP Spike	15%	



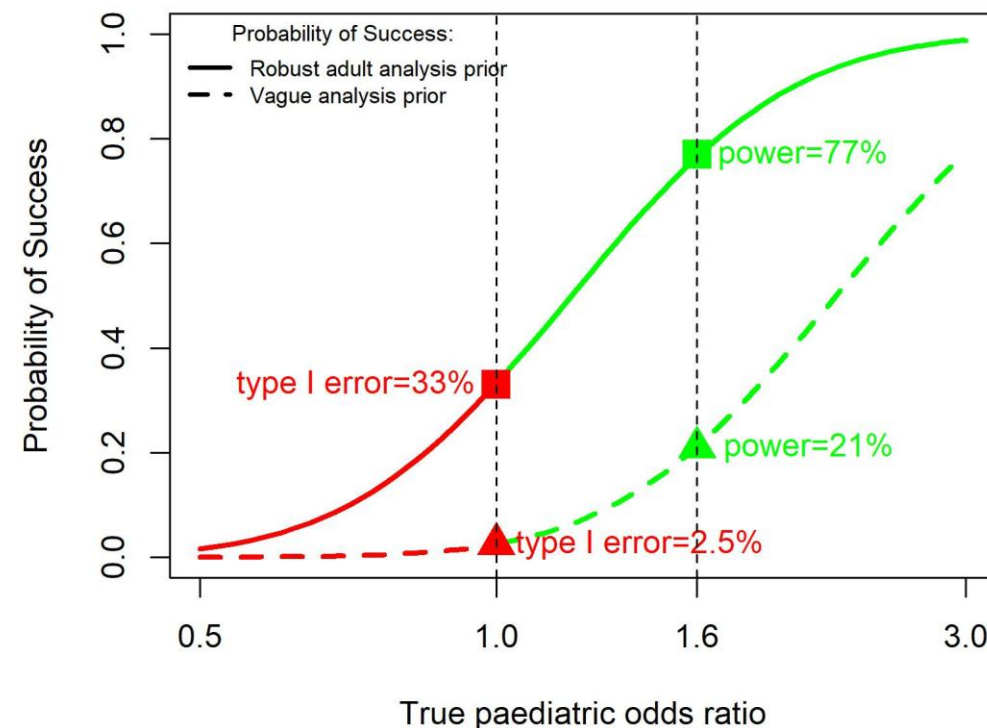
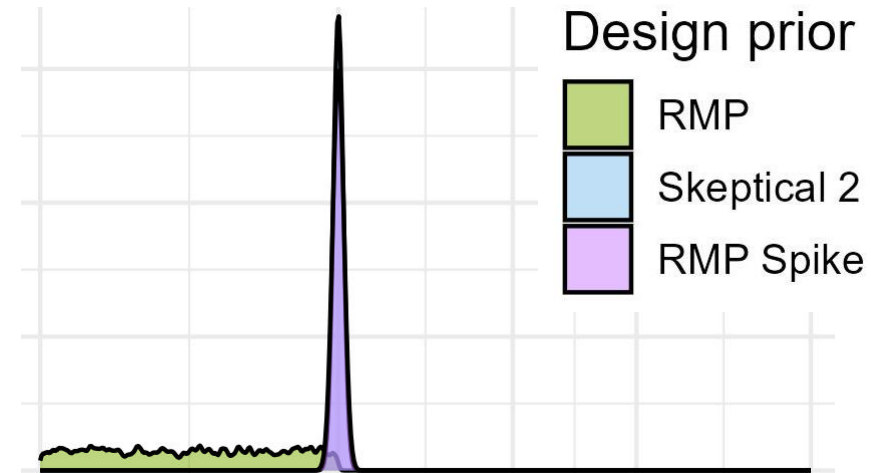
# Illustration of proposed error metrics for paediatric study

Metric	Design prior	Analysis prior	
		RMP	Vague
Type 1 error	Point mass at OR=1	33%	2.5%
Design prior probability of no treatment benefit: Pr(True OR $\leq$ 1)	Robust mixture	15%	
	Skeptical 2	30%	
	RMP Spike	15%	
<b>Predicted probability of false positive result:</b> Pr(True OR $\leq$ 1 AND success)	<b>Robust mixture</b>	<b>0.38%</b>	<b>0.015%</b>
	<b>Skeptical 2</b>	<b>10%</b>	<b>0.75%</b>
	<b>RMP Spike</b>	<b>4.98%</b>	<b>0.38%</b>



# Illustration of proposed error metrics for paediatric study

Metric	Design prior	Analysis prior	
		RMP	Vague
Type 1 error	Point mass at OR=1	33%	2.5%
Design prior probability of no treatment benefit: $\Pr(\text{True OR} \leq 1)$	Robust mixture	15%	
	Skeptical 2	30%	
	RMP Spike	15%	
Predicted probability of false positive result: $\Pr(\text{True OR} \leq 1 \text{ AND success})$	Robust mixture	0.38%	0.015%
	Skeptical 2	10%	0.75%
	RMP Spike	4.98%	0.38%
<b>Average Type 1 error</b>	<b>Robust mixture</b>	<b>2.5%</b>	<b>0.1%</b>
	<b>Skeptical 2</b>	<b>33.2%</b>	<b>2.5%</b>
	<b>RMP Spike</b>	<b>33.2%</b>	<b>2.5%</b>





## 2. Borrowing Information on the Control Effect

**GSK**

## Motivating setting: 2-arm RCT with informative prior on control

### Notation

- $\theta_t$  = true treatment effect on active arm;  $\theta_c$  = true treatment effect on control arm
- $\delta = \theta_t - \theta_c$  = treatment contrast
- $p_A(\theta_c)$  = prior information on the control arm effect
- $p_A(\theta_t)$  = flat (no prior information on active arm effect)
- Success rule:  $\Pr(\theta_t < \theta_c \mid \mathbf{y}, p_A(\theta_c)) \geq 97.5\%$  (negative effects beneficial)
- Trial data:  $\mathbf{y}_* \sim f(\theta_*, \sigma_n^2)$  for  $* = c, t$  ( $\sigma_n^2$  assumed known for design)
- $\hat{\delta} = E(\theta_t - \theta_c \mid \mathbf{y}_t - \mathbf{y}_c, p_A(\theta_c))$  = posterior mean treatment contrast

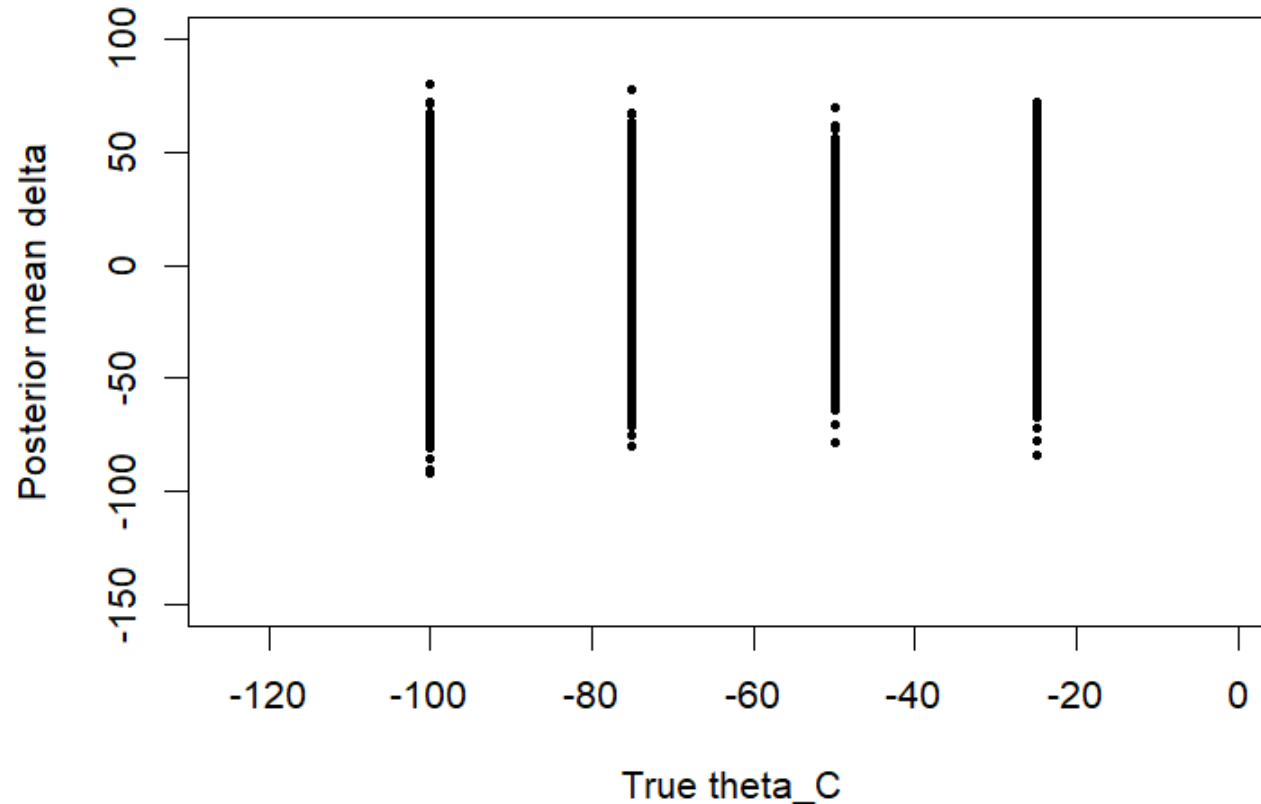
At trial design stage, the pre-posterior distribution of  $\hat{\delta}$  (or any function of the posterior for  $\delta$ ) depends on the assumed true value of  $\theta_c$  as well as the assumed true  $\delta$  that is generating the data

# Common metrics for evaluating clinical trial designs

$$y_c \sim f(\theta_c, \sigma_n^2); y_t \sim f(\theta_c, \sigma_n^2)$$

$$\hat{\delta} = E(\theta_t - \theta_c \mid y_t - y_c, p_A(\theta_c))$$

10000 samples from  
sampling (pre-posterior)  
distribution of  $\hat{\delta}$  given true  
 $\theta_t = \theta_c$  (i.e.  $\delta=0$ ) for  
different true values of  $\theta_c$



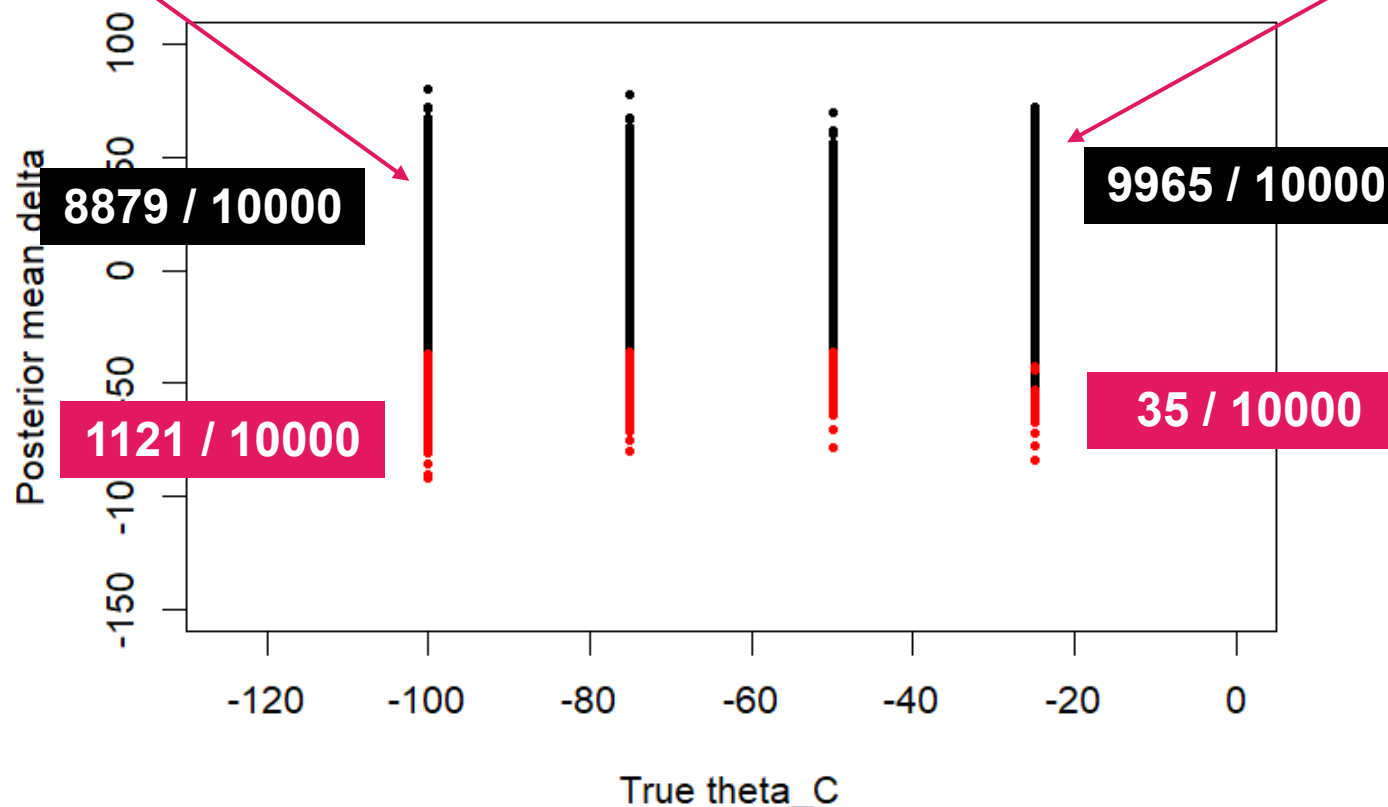
# Common metrics for evaluating clinical trial designs

Metric	Comments
$\Pr(\text{Success}   \text{Truth} = \text{null}, \theta_c = x)$ Pointwise Type 1 error	Hypothetical probability of making future decision <b>given fixed truth for <math>\theta_c</math> AND <math>\theta_t = \theta_c \rightarrow \delta = 0</math></b>

$\Pr(\text{Success} | \delta = 0, \theta_c = -100) = 0.1121$   
‘Pointwise’ Type 1 error

$\Pr(\text{Success} | \delta = 0, \theta_c = -25) = 0.0035$   
‘Pointwise’ Type 1 error

$y_c \sim f(\theta_c, \sigma_n^2); y_t \sim f(\theta_c, \sigma_n^2)$   
 $\hat{\delta} = E(\theta_t - \theta_c | y_t - y_c, p_A(\theta_c))$   
 10000 samples from sampling (pre-posterior) distribution of  $\hat{\delta}$  given true  $\theta_t = \theta_c$  (i.e.  $\delta=0$ ) for different true values of  $\theta_c$



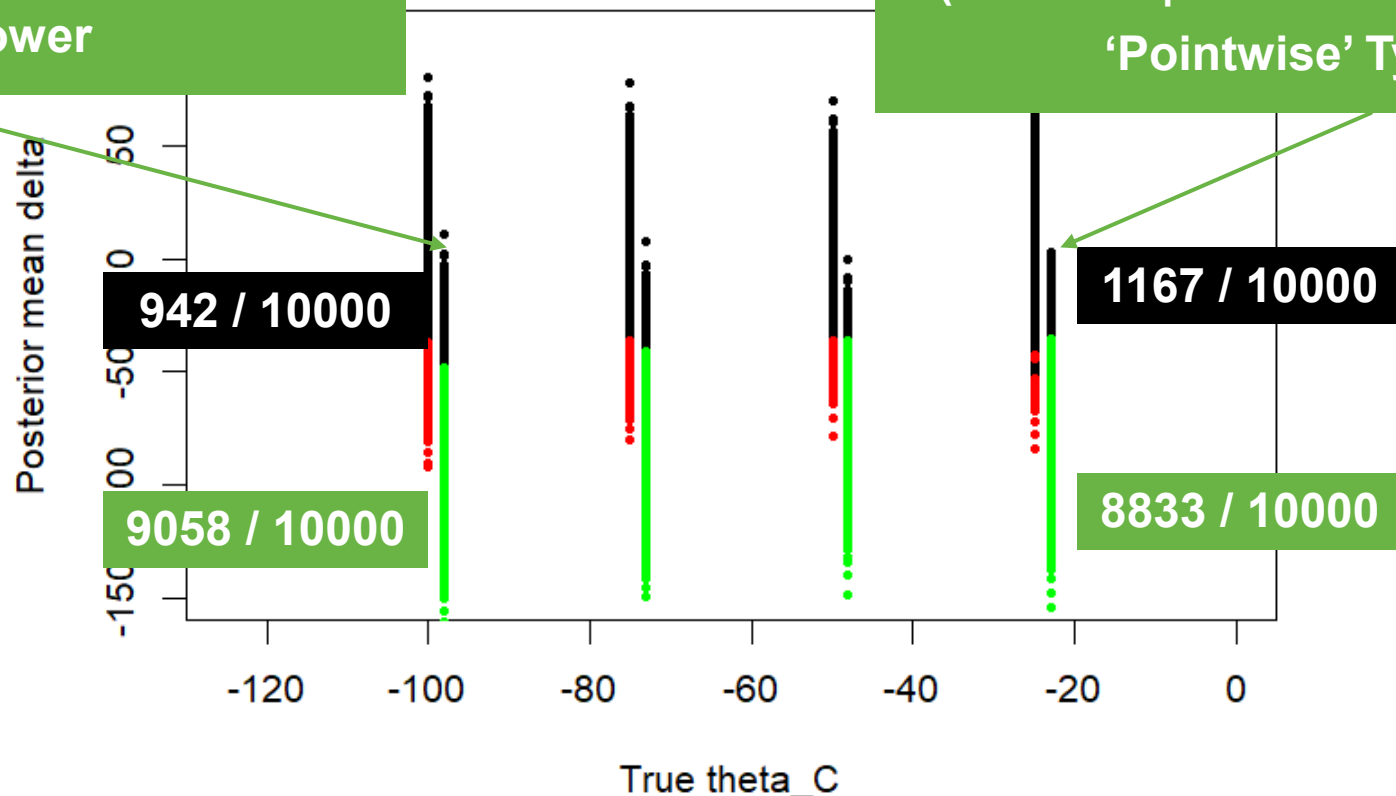
# Common metrics for evaluating clinical trial designs

Metric		Comments
$\Pr(\text{Success}   \text{Truth} = \text{null}, \theta_c = x)$	Pointwise Type 1 error	Hypothetical probability of making future decision <b>given fixed truth for <math>\theta_c</math> AND <math>\theta_t = \theta_c \rightarrow \delta = 0</math></b>
$\Pr(\text{Success}   \text{Truth} = \text{MCID}, \theta_c = x)$	Pointwise Power	Hypothetical probability of making future decision <b>given fixed truth for <math>\theta_c</math> AND <math>\theta_t = \theta_c + \text{MCID} \rightarrow \delta = \text{MCID}</math></b>

$\Pr(\text{Success} | \delta = \text{MCID}, \theta_c = -100) = 0.9058$   
 'Pointwise' Power

$\Pr(\text{Success} | \delta = \text{MCID}, \theta_c = -25) = 0.8833$   
 'Pointwise' Type 1 error

$\hat{\delta} = E(\theta_t - \theta_c | y_t - y_c, p_A(\theta_c))$   
 10000 samples from sampling (pre-posterior) distribution of  $\hat{\delta}$  given true  $\theta_t = \theta_c$  (i.e.  $\delta=0$ ) for different true values of  $\theta_c$



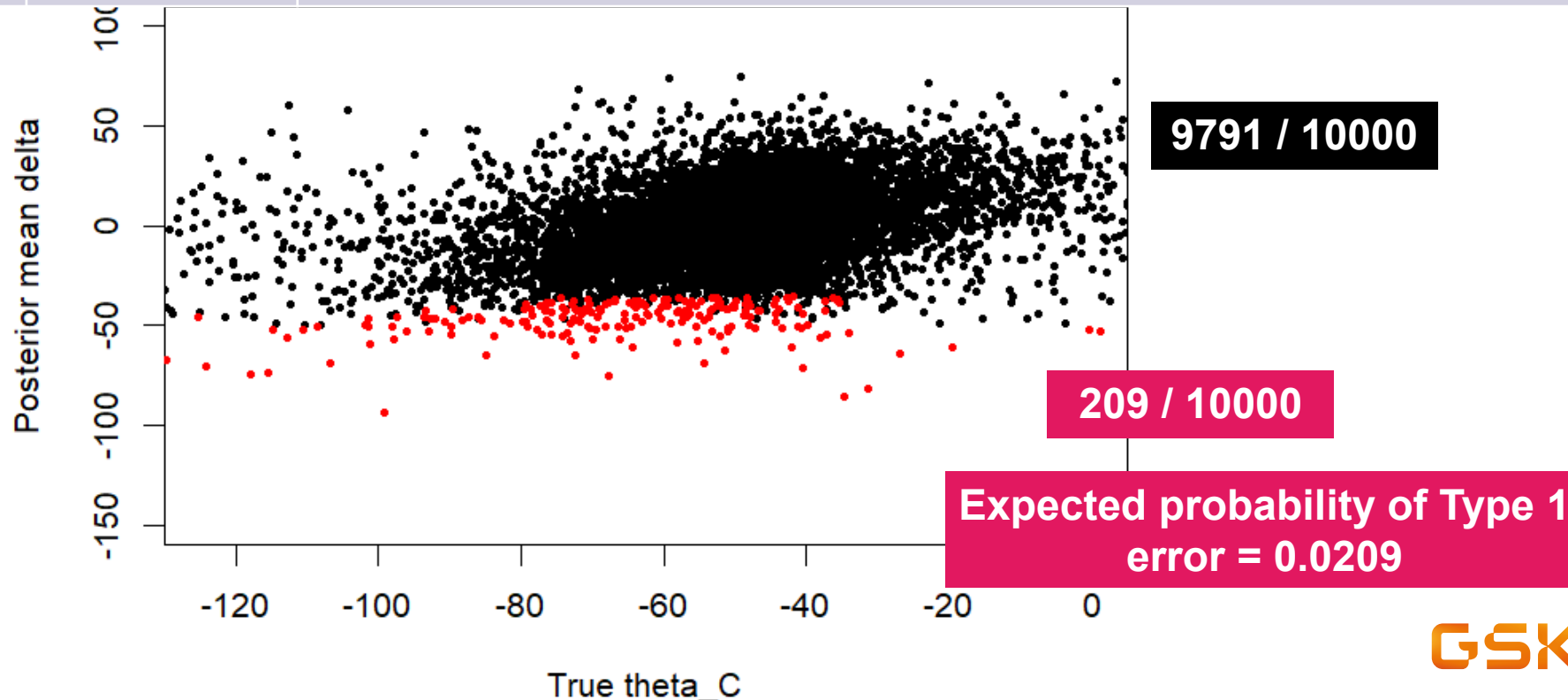
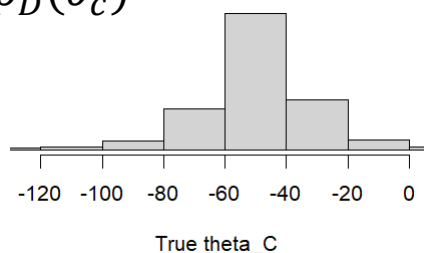
# Common metrics for evaluating clinical trial designs

Metric		Comments
$\Pr(\text{Success}   \text{Truth} = \text{null}, \theta_c = x)$	Pointwise Type 1 error	Hypothetical probability of making future decision <b>given fixed truth for <math>\theta_c</math> AND <math>\theta_t = \theta_c \rightarrow \delta = 0</math></b>
$\Pr(\text{Success}   \text{Truth} = \text{MCID}, \theta_c = x)$	Pointwise Power	Hypothetical probability of making future decision <b>given fixed truth for <math>\theta_c</math> AND <math>\theta_t = \theta_c + \text{MCID} \rightarrow \delta = \text{MCID}</math></b>
$\Pr(\text{Success}   \text{Truth} = \text{null})$	Average Type 1 error	<b>Average</b> probability of making a type 1 error accounting for uncertainty about true $\theta_c$ . Requires specification of a <b>design prior</b>

$$y_c \sim f(\theta_c, \sigma_n^2); y_t \sim f(\theta_c, \sigma_n^2)$$

$$\hat{\delta} = E(\theta_t - \theta_c | y_t - y_c, p_A(\theta_c))$$

10000 samples from sampling (pre-posterior) distribution of  $\hat{\delta}$  given true  $\theta_t = \theta_c$  (i.e.  $\delta=0$ ) for true values of  $\theta_c$  sampled from  $p_D(\theta_c)$



# Common metrics for evaluating clinical trial designs

Metric		Comments
$\Pr(\text{Success}   \text{Truth} = \text{null}, \theta_c = x)$	Pointwise Type 1 error	Hypothetical probability of making future decision <b>given fixed truth for <math>\theta_c</math> AND <math>\theta_t = \theta_c \rightarrow \delta = 0</math></b>
$\Pr(\text{Success}   \text{Truth} = \text{MCID}, \theta_c = x)$	Pointwise Power	Hypothetical probability of making future decision <b>given fixed truth for <math>\theta_c</math> AND <math>\theta_t = \theta_c + \text{MCID} \rightarrow \delta = \text{MCID}</math></b>
$\Pr(\text{Success}   \text{Truth} = \text{null})$	Average Type 1 error	<b>Average</b> hypothetical probability of making a type 1 error accounting for uncertainty about true $\theta_c$ . Requires specification of a <b>design prior</b>

All are special cases of the following metric:

$$M_4 = \int \Pr(\text{Study Success} | \theta_c, \theta_t = \theta_c + \delta^*) p(\theta_c) d\theta_c$$

- **Pointwise Type 1 error:**  $\delta^* = 0$  and  $p(\theta_c) =$  point mass prior at  $\theta_c = x$
- **Pointwise Power:**  $\delta^* = \text{MCID}$  and  $p(\theta_c) =$  point mass prior at  $\theta_c = x$
- **Average Type 1 error:**  $\delta^* = 0$  and  $p(\theta_c) =$  **design prior** reflecting our uncertainty around hypothesized control arm effect
- We don't consider joint design priors for  $\theta_c$  and  $\delta$



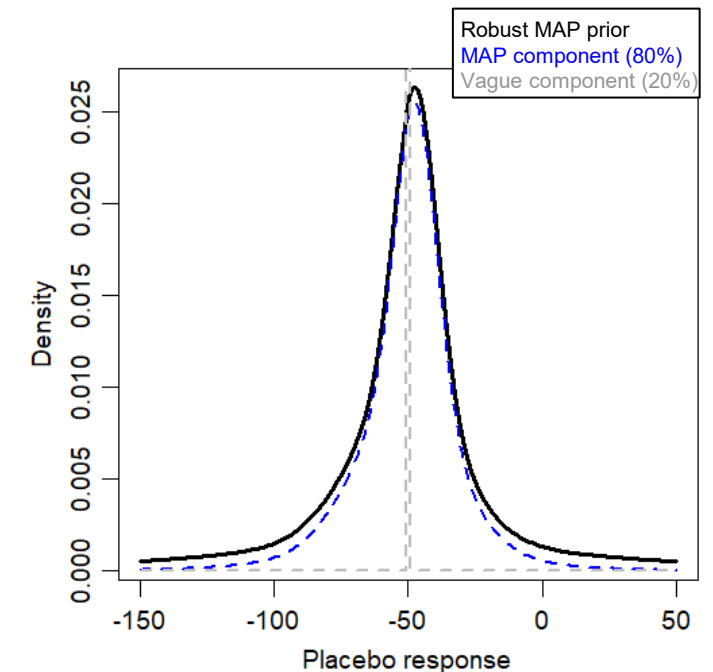
# Case Study: Borrowing Information on the Control Effect



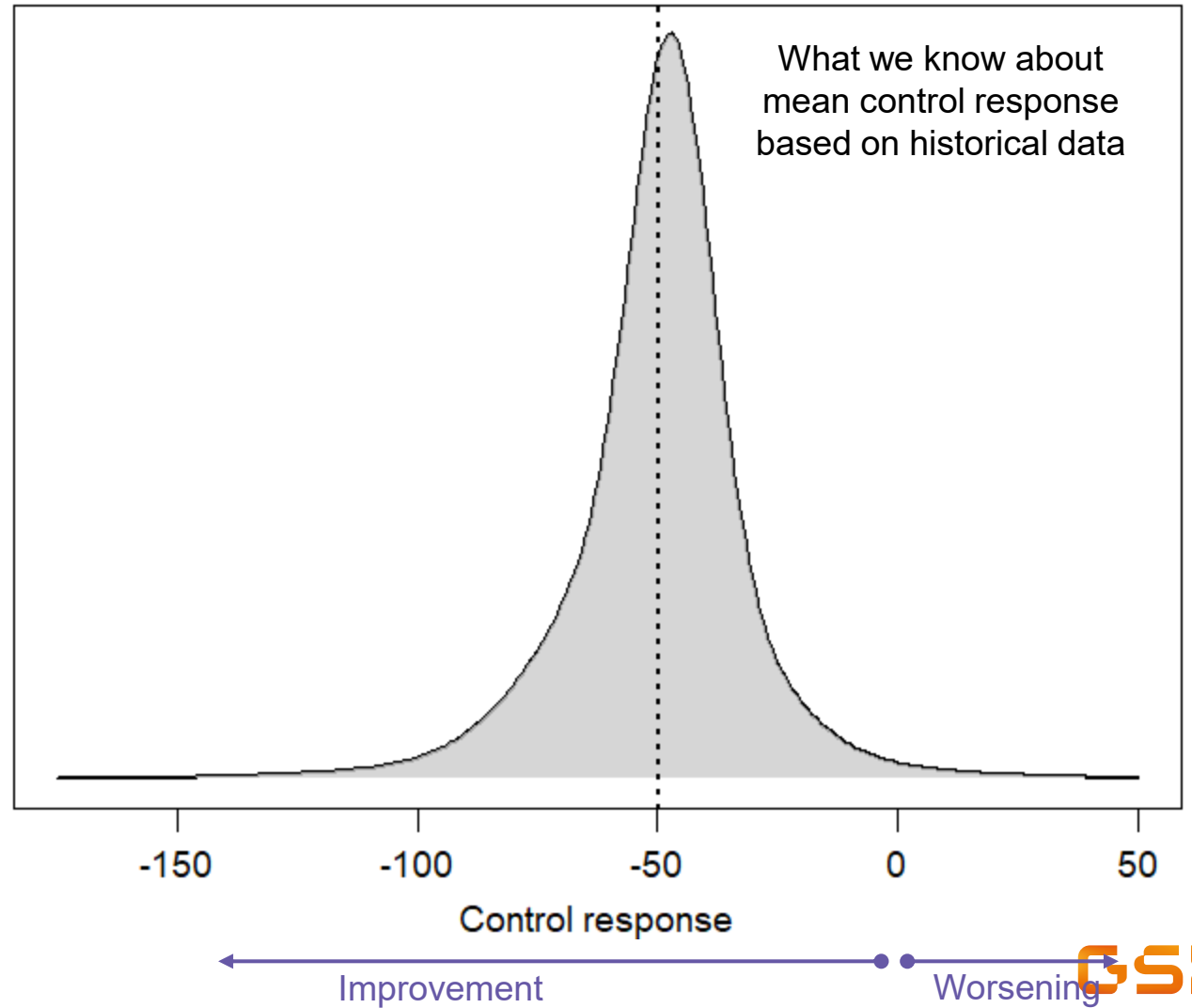
## Hybrid control arm trial

- Ph2 Double-blind RCT, experimental vs placebo (2:1 ratio)
- Primary endpoint: Continuous (Normally distributed); Improvement = negative values
- Historical placebo data: 6 studies with 671 patients in total

<b>Sample size in proposed study</b>	<b>40 active; 20 placebo</b>
<b>Prior distribution on true placebo response</b>	<b>Robust MAP prior:</b> weighted mixture of <b>MAP prior</b> and <b>vague distribution</b> centered at the mean of the MAP prior
<b>Weight on MAP component of the mixture prior</b>	80%
<b>Success rule defining positive result</b>	$\Pr(\theta_t < \theta_c \mid \text{data, control prior}) \geq 0.975$

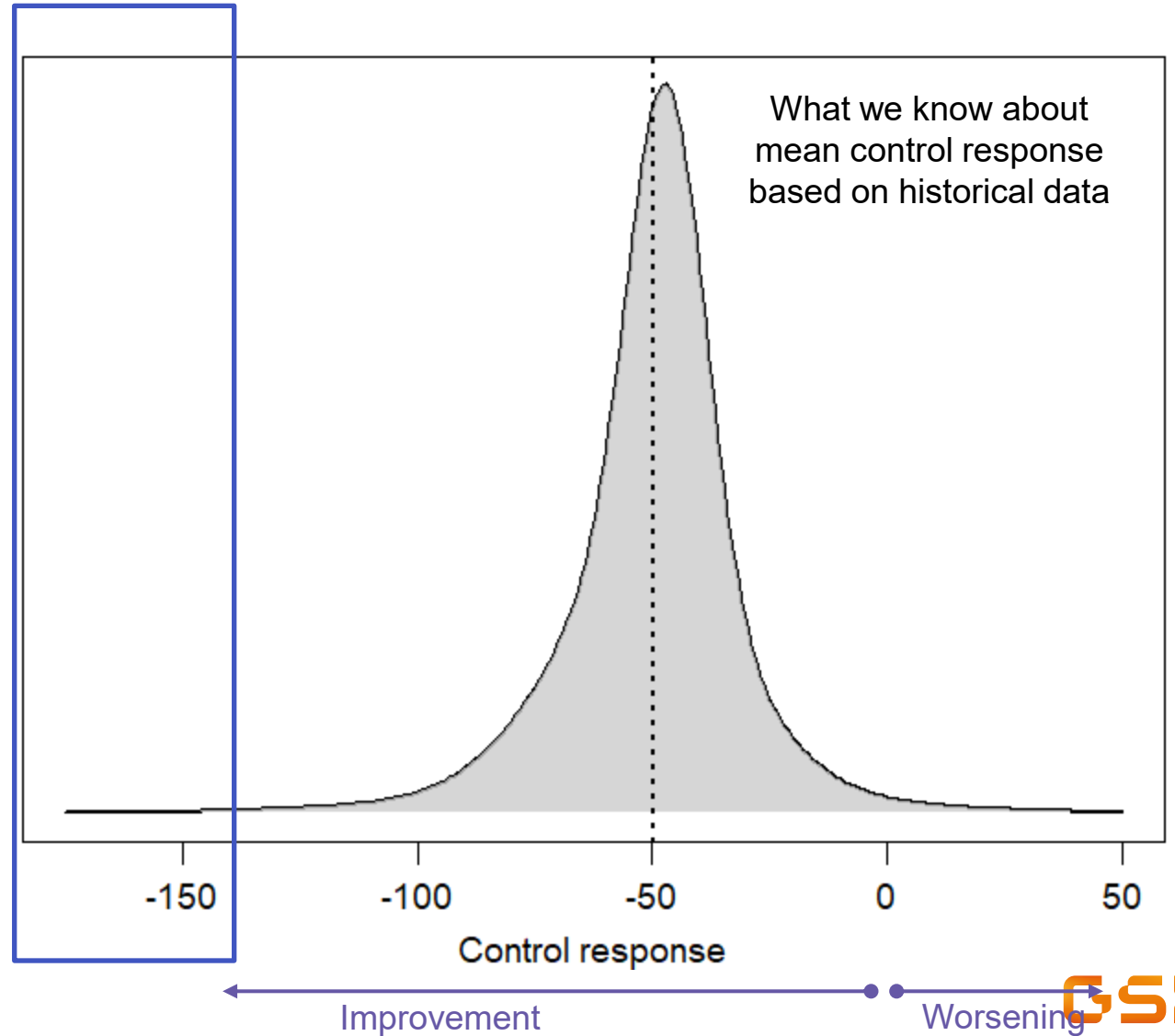


# Assessing risk of false positive result in hybrid control study

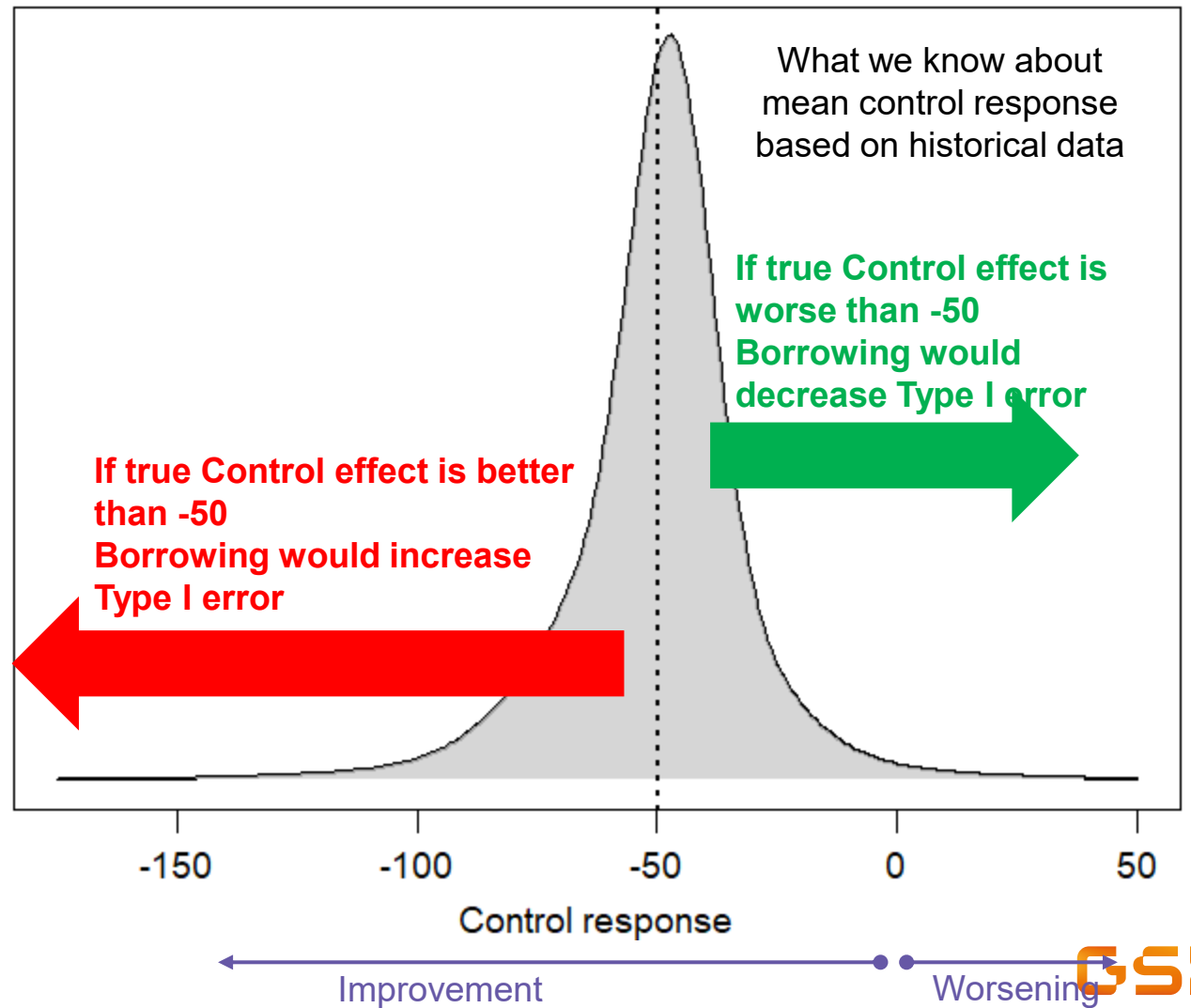


# Assessing risk of false positive result in hybrid control study

Unlikely to be interested in this region



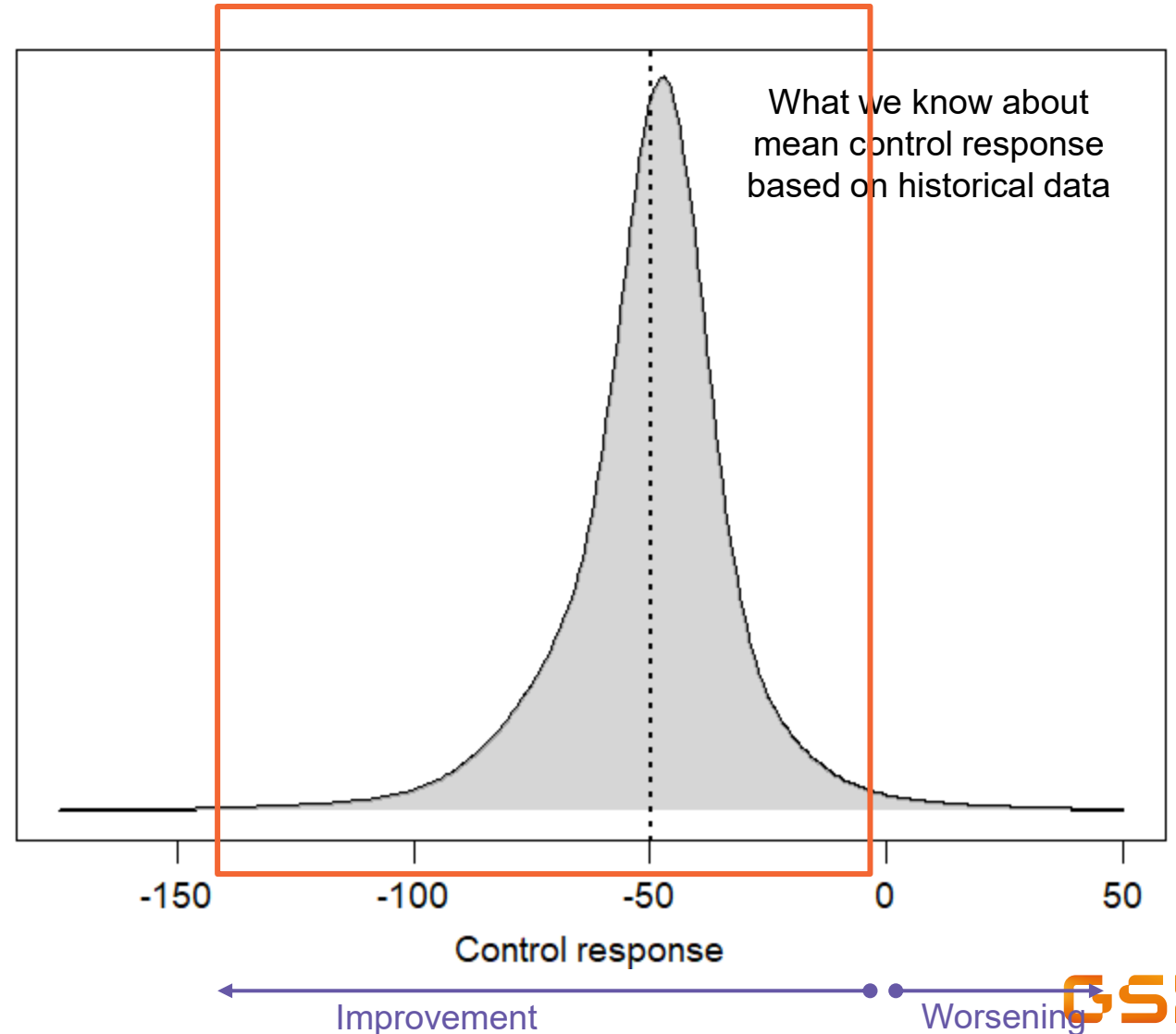
# Assessing risk of false positive result in hybrid control study



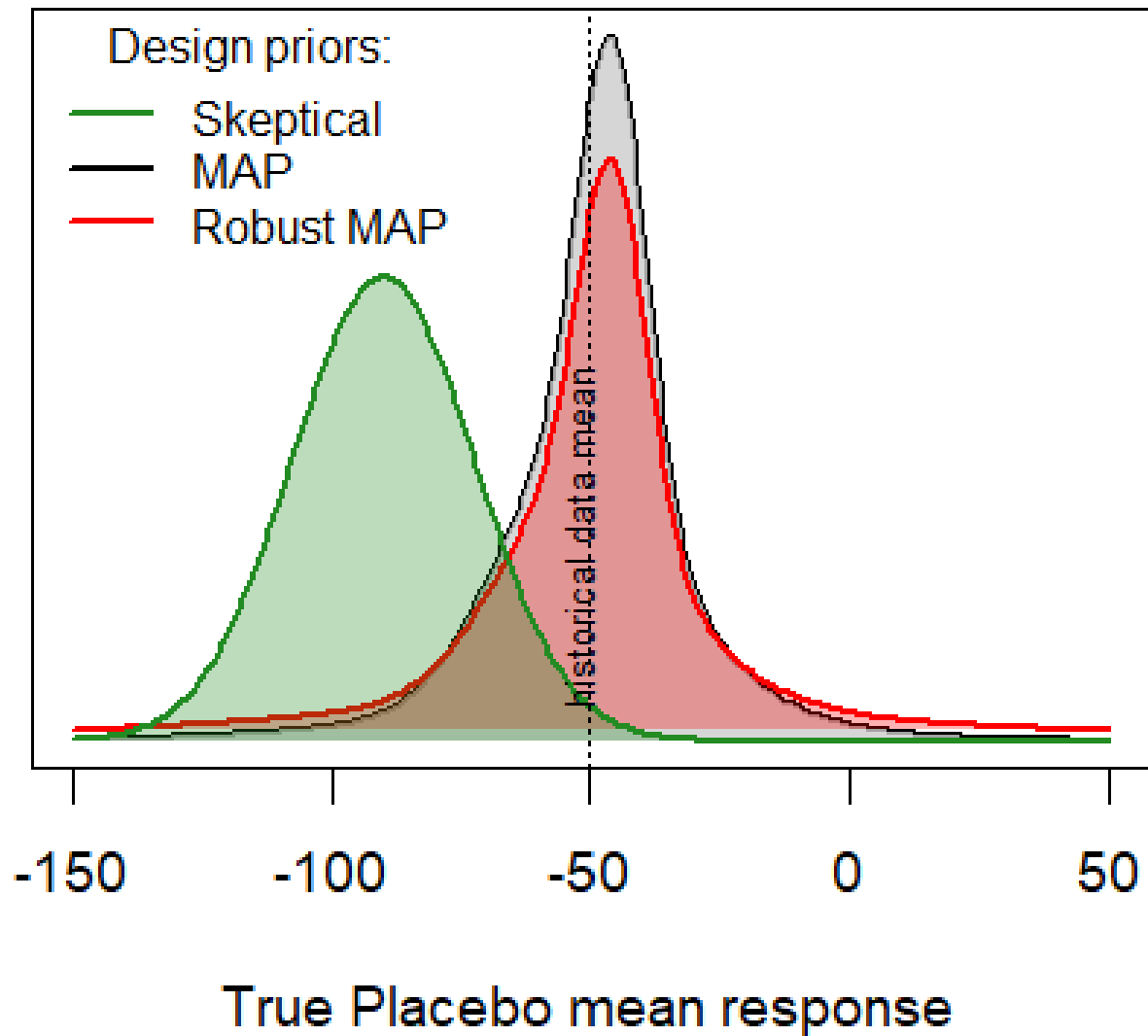
# Assessing risk of false positive result in hybrid control study

Quantify Type I error under range of relevant deviations (drift):

- Relevant drift scenarios defined by appropriate design priors for true control response
- Recommended Metrics:
  - “Averaged” Type I error
  - Maximum Type I error with corresponding probability of occurrence



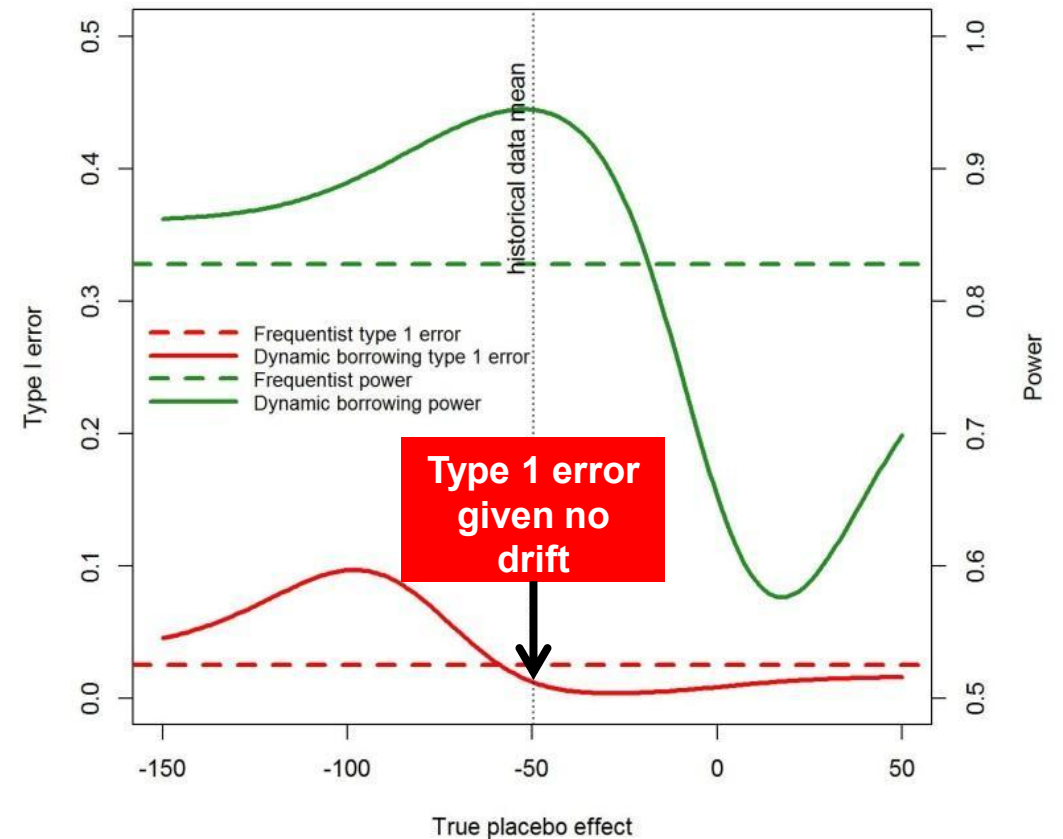
## Design priors for hybrid control example



**Skeptical prior:**  
posterior distribution of a stand-alone analysis of the historical study with the “most extreme” placebo effect

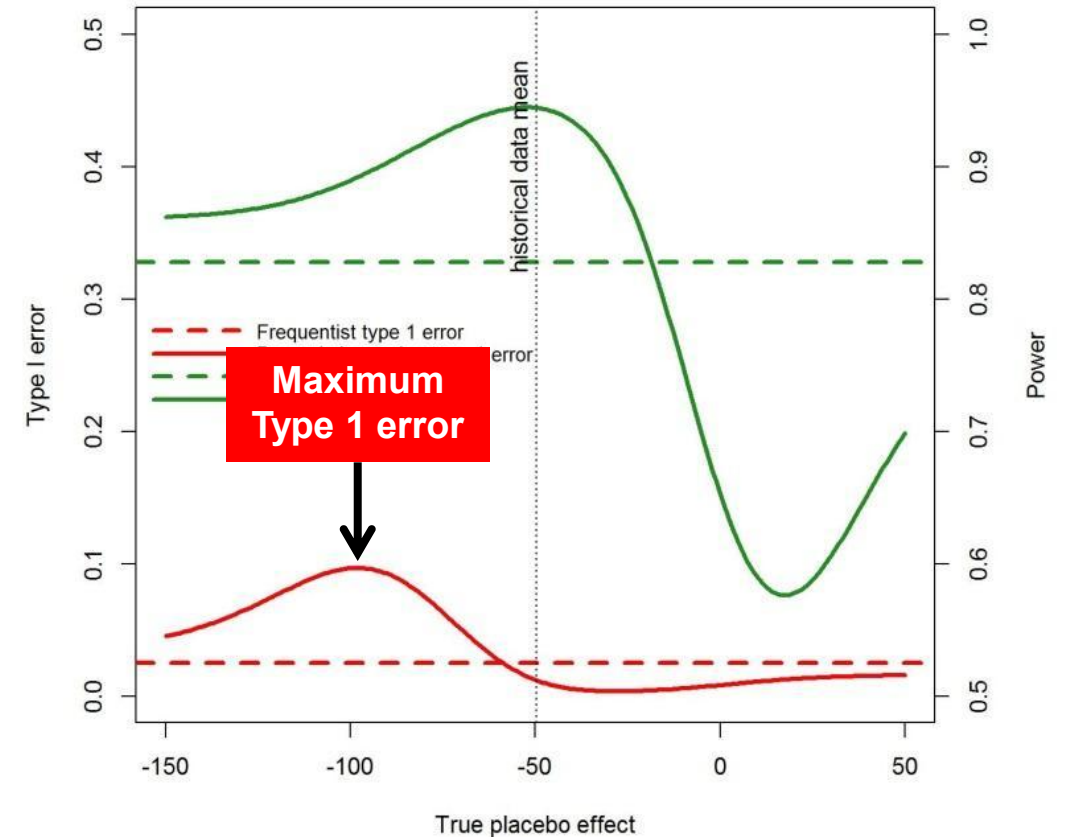
# Illustration of proposed metrics for hybrid control study

Metric	Design prior	Analysis prior	
		RMP	Vague
Pointwise Type 1 error	Point mass $\theta_c = -50$ (no drift)	1.3%	2.5%



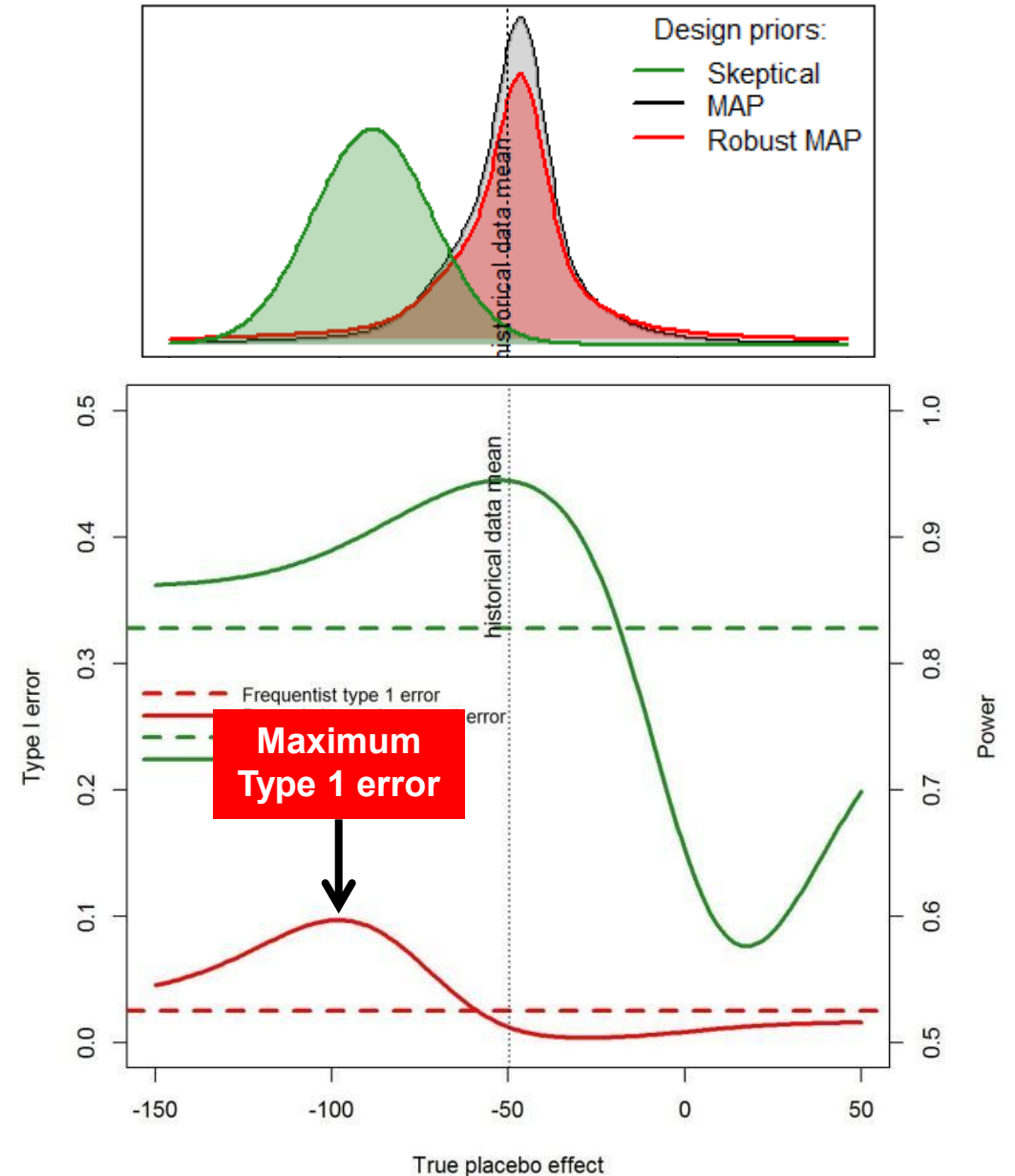
# Illustration of proposed metrics for hybrid control study

Metric	Design prior	Analysis prior	
		RMP	Vague
Pointwise Type 1 error	Point mass $\theta_c = -50$ (no drift)	1.3%	2.5%
<b>Maximum Type 1 error</b>	<b>Point mass <math>\theta_c = -99</math></b>	<b>10%</b>	<b>2.5%</b>



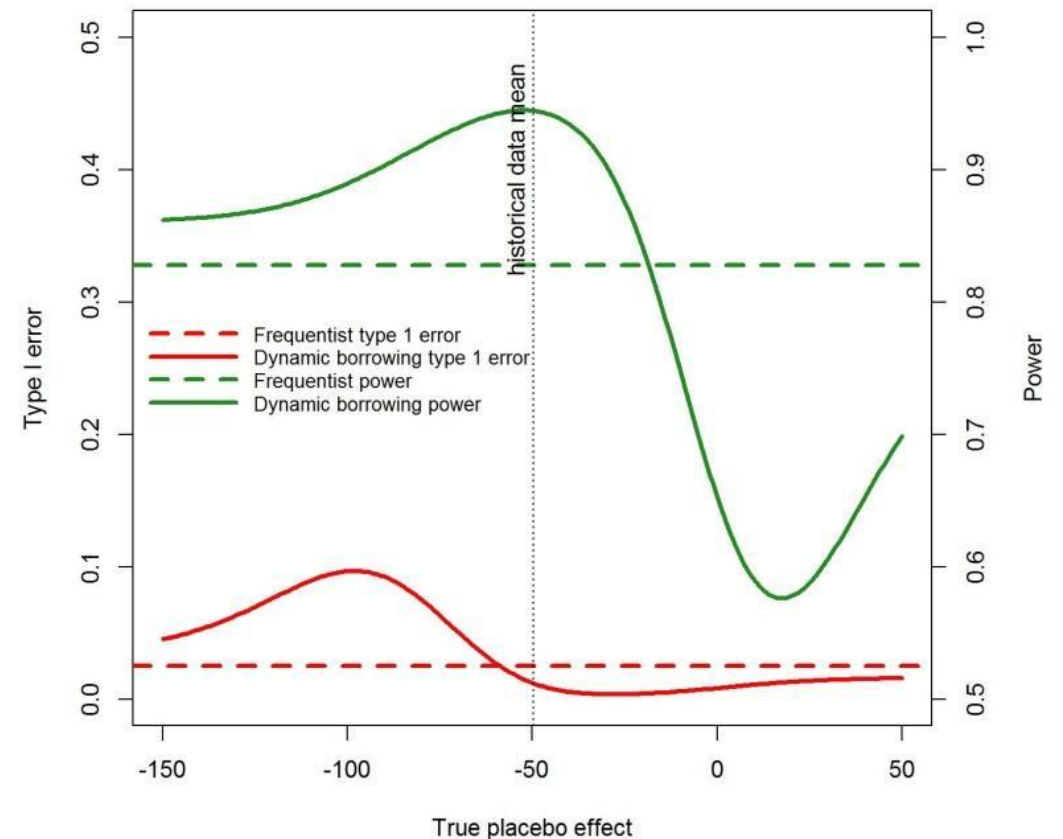
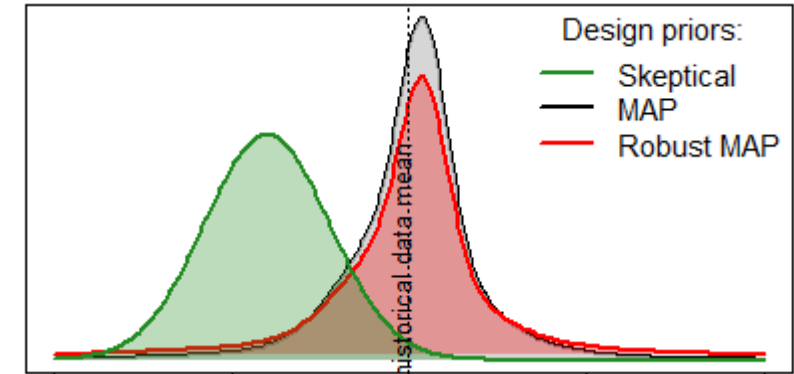
# Illustration of proposed metrics for hybrid control study

Metric	Design prior	Analysis prior	
		RMP	Vague
Pointwise Type 1 error	Point mass $\theta_c = -50$ (no drift)	1.3%	2.5%
Maximum Type 1 error	Point mass $\theta_c = -99$	10%	2.5%
$\Pr(\theta_c \leq -99)$	MAP	1.8%	
	Robust MAP	7.2%	
	Skeptical	30.5%	



# Illustration of proposed metrics for hybrid control study

Metric	Design prior	Analysis prior	
		RMP	Vague
Pointwise Type 1 error	Point mass $\theta_c = -50$ (no drift)	1.3%	2.5%
Maximum Type 1 error	Point mass $\theta_c = -99$	10%	2.5%
$\Pr(\theta_c \leq -99)$	MAP	1.8%	
	Robust MAP	7.2%	
	Skeptical	30.5%	
Average Type 1 error	MAP	2.0%	2.5%
	Robust MAP	2.5%	2.5%
	Skeptical	8.8%	2.5%



## Concluding remarks

Conventional Type 1 error not always appropriate for trial designs that borrow prior information

Bayesian metrics enable evaluation of **predicted probabilities of trial outcomes** given a design prior for the true parameters generating the data

- ✓ Considers **likelihood of true parameter value being in particular region** when assessing risk of false positive results
- ✓ Enables assessment design characteristics under “**drift**” **scenarios** where prior information used for analysis may differ from true process generating the data
- ✓ Under consistency assumption (i.e. design prior = analysis prior) the **average type error is controlled at the nominal level** (Best et al 2024; Walley & Grieve 2021)

These metrics **complement traditional Type 1 error calculations** and can help facilitate consistent evaluation of trial designs with informative priors

- ✓ **Choice of design prior(s)** requires careful consideration and discussion by the trial sponsor and regulatory agency at the design stage

## References

Best, N., Ajimi, M., Neuenschwander, B., Saint-Hilary, G., & Wandel, S. (2024). Beyond the Classical Type I Error: Bayesian Metrics for Bayesian Designs Using Informative Priors. *Statistics in Biopharmaceutical Research*, 17(2), 183–196. <https://doi.org/10.1080/19466315.2024.2342817>

Kunzmann K, Grayling MJ, Lee KM, Robertson DS, Rufibach K, Wason JMS. (2021). A Review of Bayesian Perspectives on Sample Size Derivation for Confirmatory Trials. *Am Stat*. 2021;75(4):424-432. doi: 10.1080/00031305.2021.1901782. Epub 2021 Apr 22.

Psioda, M., and Ibrahim, J. (2019), “Bayesian Clinical Trial Design Using Historical Data that Inform the Treatment Effect,” *Biostatistics*, 20, 400–415. DOI:10.1093/biostatistics/kxy009

Spiegelhalter, David J., et al. “Bayesian Approaches to Randomized Trials.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, vol. 157, no. 3, 1994, pp. 357–416. *JSTOR*, <https://doi.org/10.2307/2983527>

Walley RJ, Grieve AP. (2021). Optimising the trade-off between type I and II error rates in the Bayesian context. *Pharmaceutical Statistics*. 2021; 20: 710–720. <https://doi.org/10.1002/pst.2102>