Perspectives on Bayesian Statistics in Regulatory Decision Making

Nicky Best

Head of Statistics and Data Science Innovation Hub, GSK

Views expressed are my own and do not necessarily reflect those of GSK

Bayesian methods in regulatory decision-making

Despite ICH-E 9 mentioning Bayesian methods already in 1998:

- Why have they **not made their way** to regulatory decision-making?

Perceived barriers to using Bayesian methods

Therapeutic Innovation & Regulatory Science (2023) 57:417–425
https://doi.org/10.1007/s43441-021-00357-x

ANALYTICAL REPORT

Why are not There More Bayesian Clinical Trials? Perceived Barriers and Educational Preferences Among Medical Researchers Involved in Drug Development

The Medical Outreach Subteam of the Drug Information Association Bayesian Scientific Working Group ·
Jennifer Clark, PhD[1] · Natalia Muhlemann[2] · Fanni Natanegara[3] · Andrew Hartley[4] · Deborah Wenkert[5] ·
Fei Wang[6] · Frank E. Harrell Jr.[7] · Ross Bray[3]

1. **Knowledge**: Insufficient knowledge of Bayesian approaches
2. **Regulators**: Lack of clarity/guidance from regulators
3. **Clinical Team**: Reluctance from my internal clinical team
4. **NA**: The Bayesian approach is not applicable, and my organization sees no benefit
5. **Reg Team**: Reluctance from my internal regulatory team
6. **Stat Team**: Reluctance from my internal statistical team
7. **Mngmnt**: Reluctance from upper management
8. **Other**

- Lack of familiarity among researchers

- Absence of clear regulatory guidance
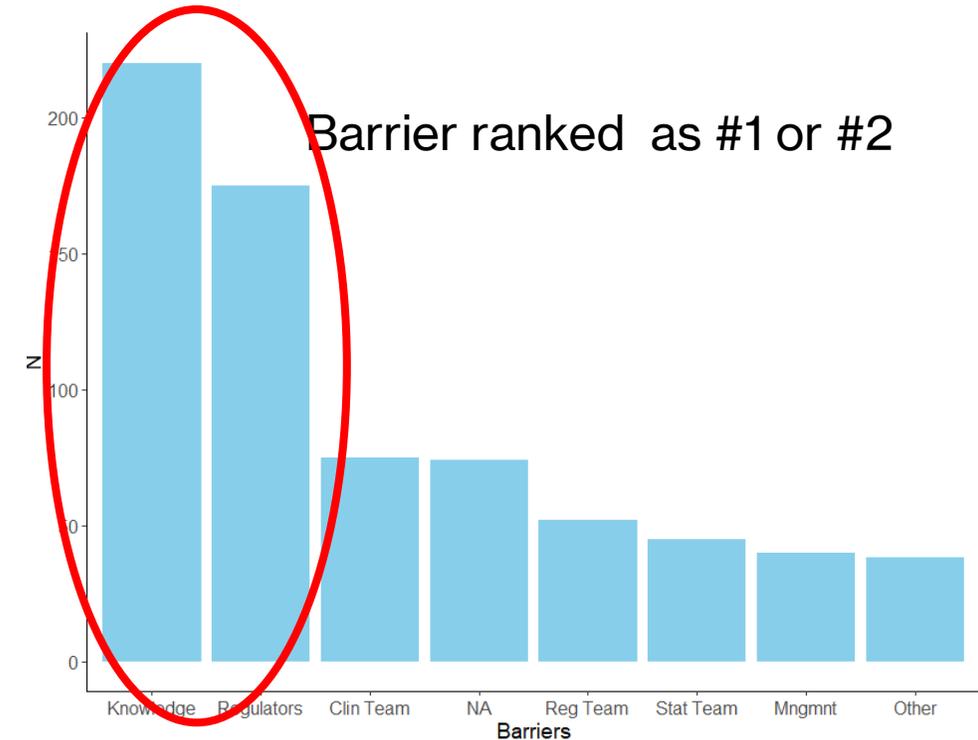
Barrier ranked as #1 or #2

Figure adapted from Clark et al, 2023

Bayesian methods increasingly used for:

- Sponsor internal decision making

- Monitoring of safety data

- Informative priors in early-phase

- Medical device trials

- Long established use of Bayesian methods in pharmacometrics

Increased complexity of drug development:

- Rare diseases and unmet needs

- Treatment effect heterogeneity

Increased availability of high quality RWD

→ Need to move beyond traditional trial designs and analysis methods

Increased role for model-based methods and simulations to evaluate trial designs and facilitate regulatory decision-making

ICH – E9: '...when reasons for their use are clear'

No advantages	Operational advantages/ pragmatic reasons	Clearer advantages
• Bayesian t-test • Bayesian MMRM • etc all with non-informative/vague/weakly informative prior	• Estimation in small populations • Adaptive designs using predictive distributions • Pooling across small subgroups	• (Pediatric) extrapolation • Borrowing external information in rare diseases • Borrowing (external) data for underpowered secondary endpoints or small (sub)populations
No external data/ assumptions		**Incorporation or borrowing of external information**

Interpretability of Bayesian decision criteria

- Probability of clinically relevant treatment benefit e.g.

Pr(treatment effect > 10) > 90%

ICH – E9: '...when reasons for their use are clear'

No advantages	Operational advantages/ pragmatic reasons	Clearer advantages
Bayesian t-test	• Estimation in small populations	• (Pediatric) extrapolation
• Bayesian MMRM	• Adaptive designs using predictive distributions	• Borrowing external information in rare diseases
• etc	• Pooling across small subgroups	• Borrowing (external) data for underpowered secondary endpoints or small (sub)populations
all with non-informative/vague/ weakly informative prior		
No external data/ assumptions		Incorporation or borrowing of external information

Interpretability of Bayesian decision criteria

- Probability of clinically relevant treatment benefit e.g.
  Pr(treatment effect > 10) > 90%

Bayesian Central Limit Theorem guarantees good frequentist properties in large sample settings

- Including control of type 1 error
→ No disadvantages of Bayes?

Example

- Ph2 RCT in chronic pain comparing active v placebo

- Primary endpoint: CFB at wk 12 in weekly av. NRS (Numeric Rating Scale) pain score
  - ➢ Change of -1 in NRS is considered clinically meaningful

	Conventional design
Primary analysis model	Frequentist MMRM
Success rule	Statistical significance (p < 0.5) AND <u>observed</u> treatment effect is clinically meaningful (< -1)

Example

- Ph2 RCT in chronic pain comparing active v placebo

- Primary endpoint: CFB at wk 12 in weekly av. NRS (Numeric Rating Scale) pain score
  - ➢Change of -1 in NRS is considered clinically meaningful

	Conventional design	Bayesian design
Primary analysis model	Frequentist MMRM	Bayesian MMRM with vague priors on all model parameters
Success rule	Statistical significance (p < 0.5) AND <u>observed</u> treatment effect is clinically meaningful (< -1)	>97.5% certain the true treatment effect is better than placebo AND >75% certain the true treatment effect is clinically meaningful (< -1)

- Bayesian model also naturally handles imputation of missing data aligned with estimand strategy, and predicted PoS for interim decision-making

Additional complexities of Bayesian methods

ICH – E9: '…when resulting conclusions are sufficiently robust'

Additional choices (requiring justification) related to:
- Informative priors, external data sources, weight for external information
- Bayesian model, including hyperparameters

- Subjective judgments have to be made when building Bayesian models

  ➤ Requires multidisciplinary discussion between sponsors and regulators

- Assumptions also needed for frequentist designs/analyses

  ➤ ICH E9 Addendum: "The statistical assumptions that underpin the main estimator should be documented" and "Sensitivity analysis should be planned for the main estimators of all estimands that will be important for regulatory decision making and labelling in the product information"

Additional complexities of Bayesian methods

ICH – E9: '...when resulting conclusions are sufficiently robust'

Additional choices (requiring justification) related to:
- Informative priors, external data sources, weight for external information
- Bayesian model, including hyperparameters

- Subjective judgments have to be made when building Bayesian models

  ➤ Requires multidisciplinary discussion between sponsors and regulators

- Assumptions also needed for frequentist designs/analyses

  ➤ ICH E9 Addendum: "The statistical assumptions that underpin the main estimator should be documented" and "Sensitivity analysis should be planned for the main estimators of all estimands that will be important for regulatory decision making and labelling in the product information"

- Strength of Bayesian Framework is that it encourages thoughtful cross-disciplinary discussion and transparency *up front*

- Could ICH-M15 and MIDD credibility assessment framework be applicable here?

Example

- Sponsor proposed use of Bayesian Dynamic Borrowing (BDB) in rare disease

  ➢ Integration of clinical efficacy data from non-EU trial with planned multi-regional trial including EU

- Assessment of Ethnic Differences

  ➢ Characterized comparability and evaluated potential impact on efficacy and safety of treatment

  ➢ Included plans for real-world natural history and treatment patterns study

- Outlined planned *in silico* trial simulations to calibrate design and limit potential biases

- Regulatory Scientific Advice

  ➢ "Bayesian approach may not guarantee extrapolatability to EU population"

- Need for Regulatory Guidance

  ➢ What information is needed to enable up-front alignment on assumptions and parameter choices that would or would not support extrapolatability?
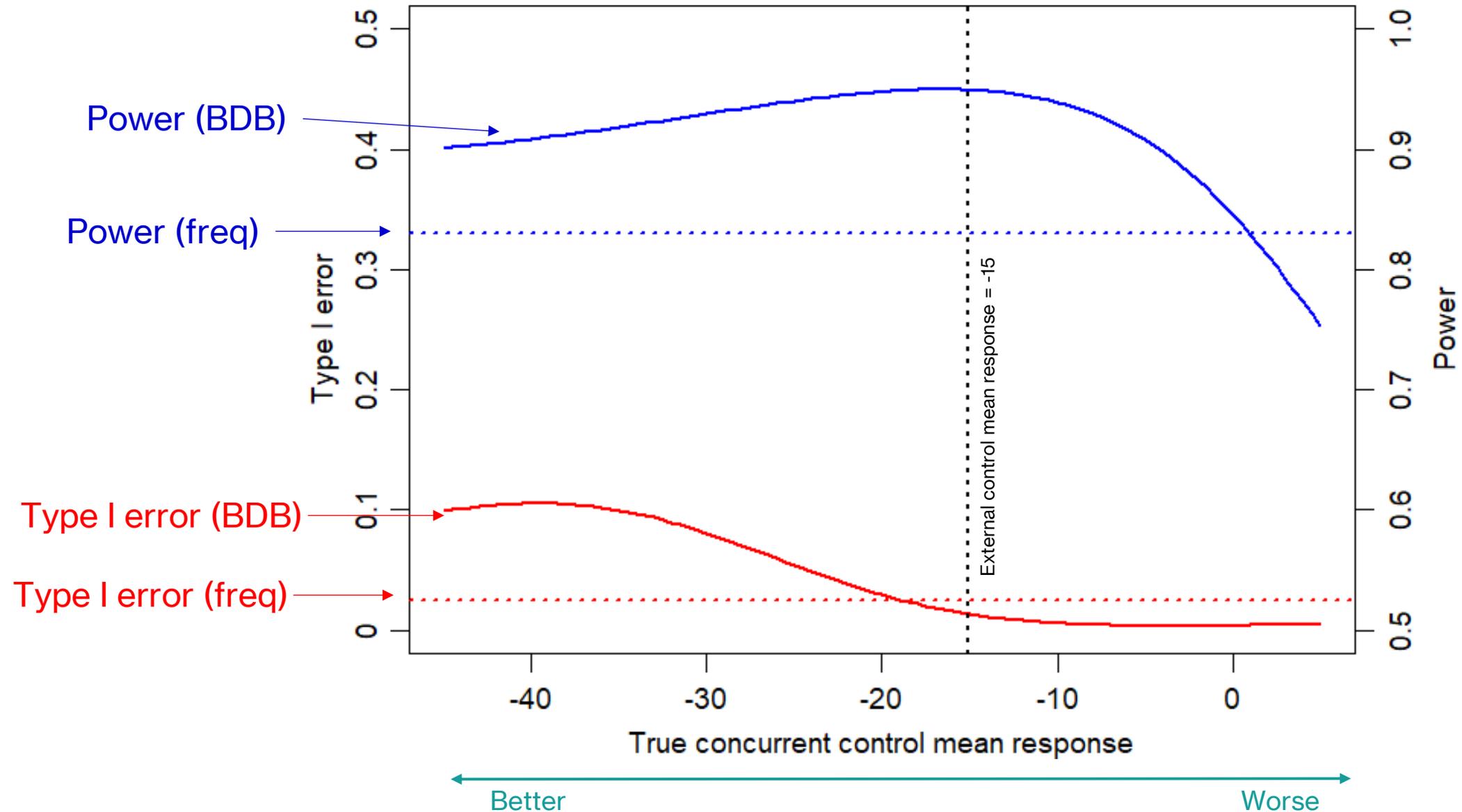
We also need to reflect on:

- How to ensure consistency of severity of testing among approaches
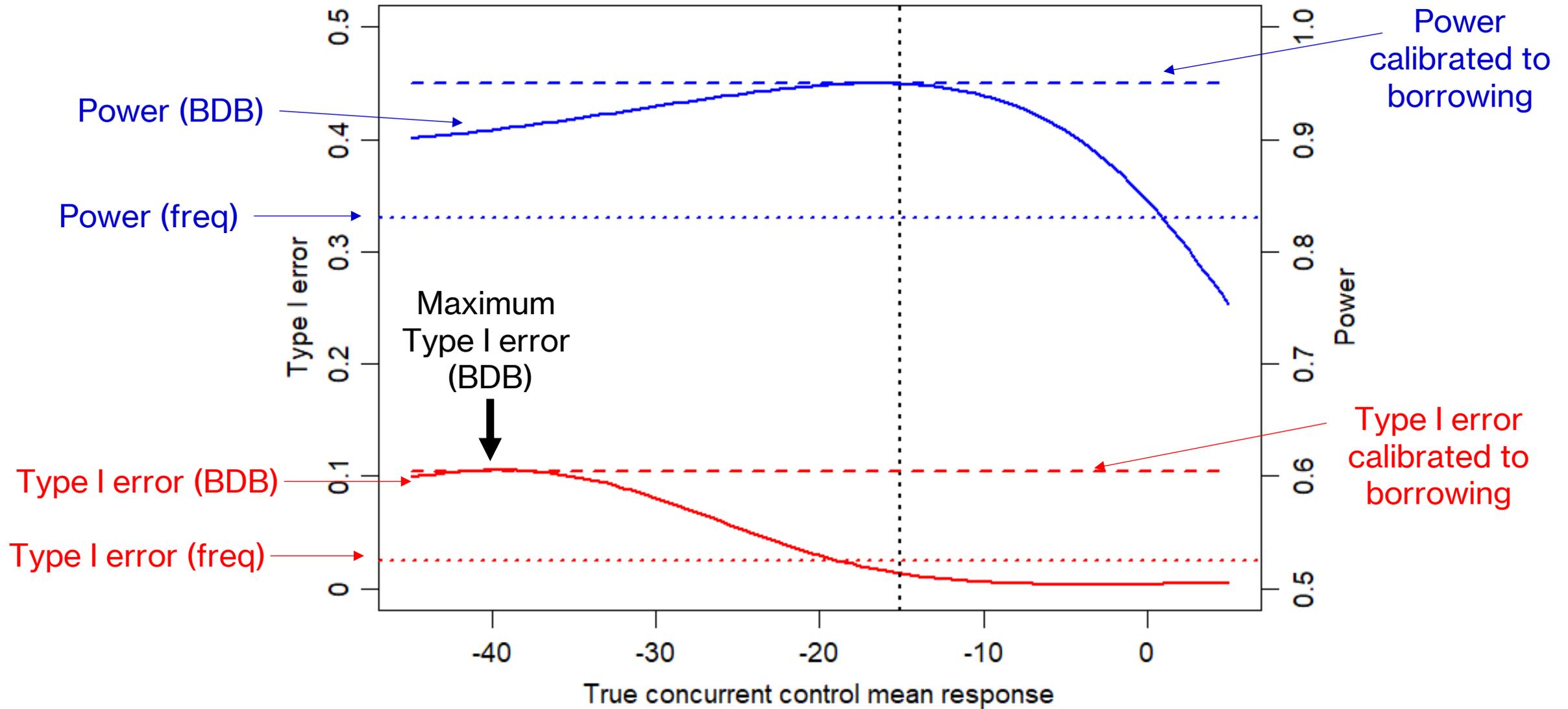
- How to deal with lack of type I error control

- Type I error of Bayesian borrowing designs → area of active research

- If type 1 error is strictly controlled

  ➢ No power gains possible (Kopp-Schneider et al (2019) and others)

- Kopp-Schneider et al (2023): framework for calibrating (frequentist) test without borrowing to maximum type 1 error of a borrowing design

  ➢ No power gains are possible

  ➢ Assumes all possible values of the parameter space are equally important

  ➢ Power gains from borrowing are possible if we are willing to consider some regions of the parameter space as more important than others and restrict (or weight) operating characteristics to that region

Type I error and Bayesian dynamic borrowing (BDB)

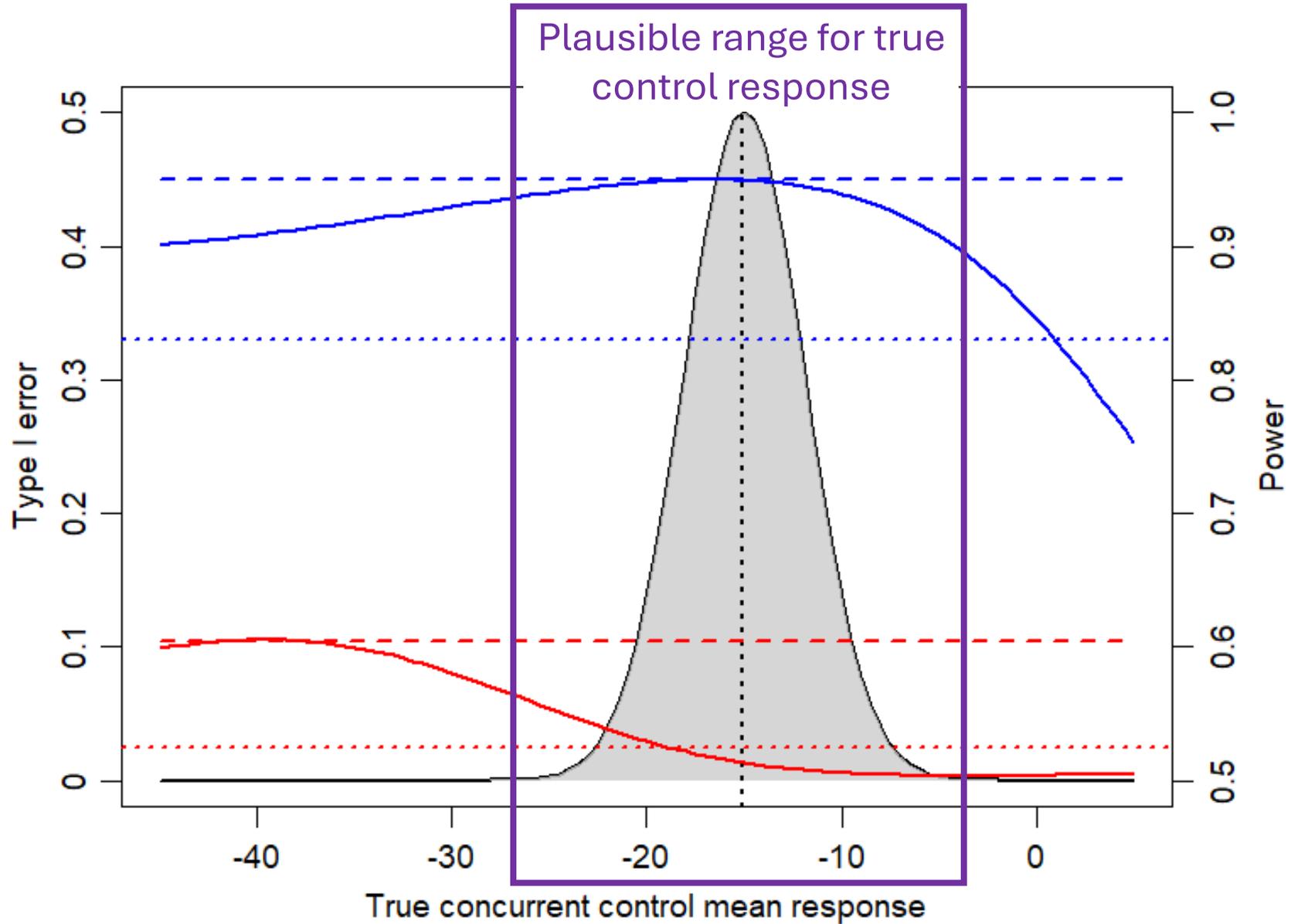## Illustrative example using BDB for hybrid control arm in 2-arm trial

Type I error and Bayesian dynamic borrowing (BDB)

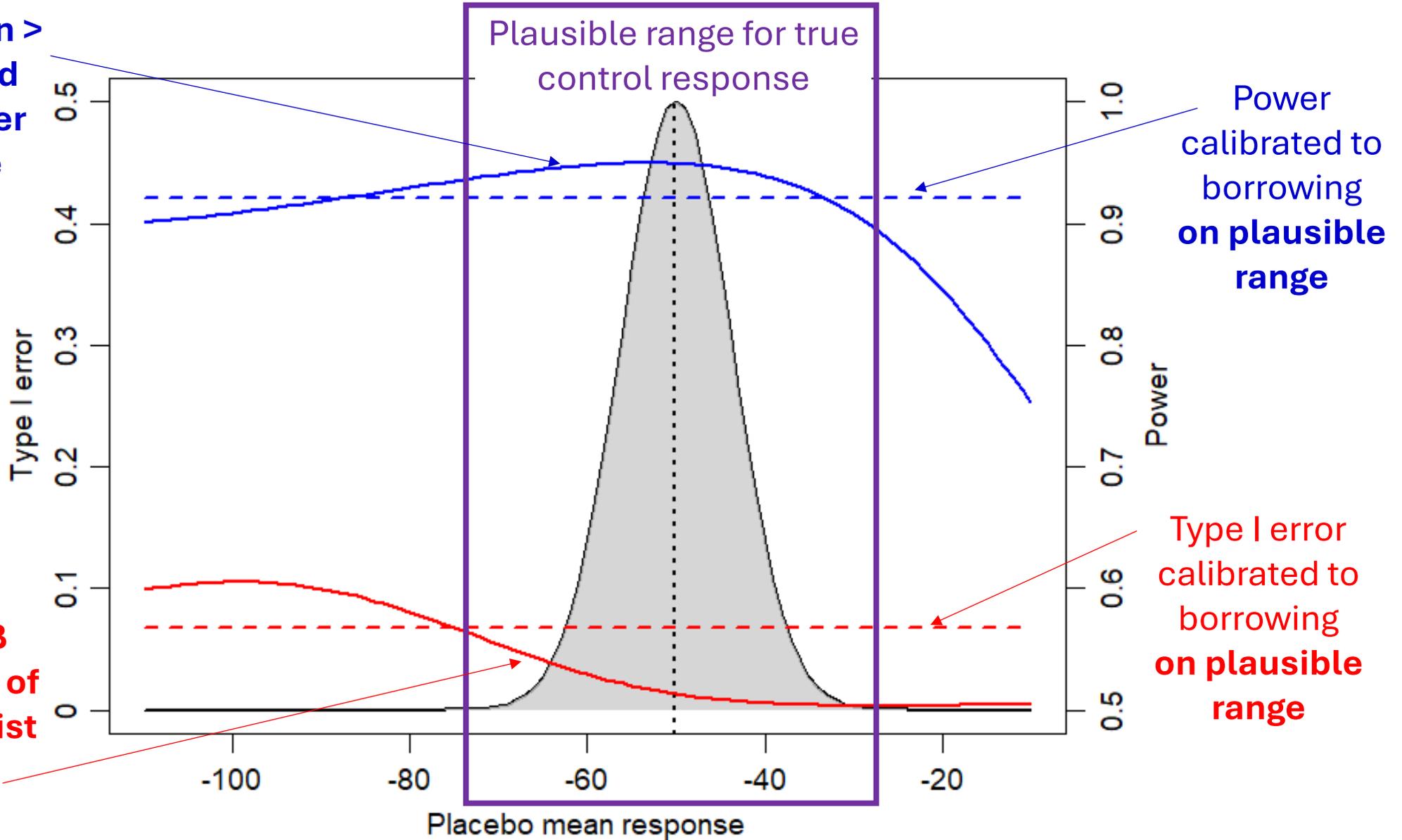# Type I error and Bayesian dynamic borrowing (BDB)

Type I error and Bayesian dynamic borrowing (BDB)

Power of BDB design > power of calibrated frequentist test over most of plausible range

Plausible range for true control response

Power calibrated to borrowing

Power calibrated to borrowing **on plausible range**

Type I error of BDB design ≤ type I error of calibrated frequentist test over entire plausible range

Type I error calibrated to borrowing

Type I error calibrated to borrowing **on plausible range**

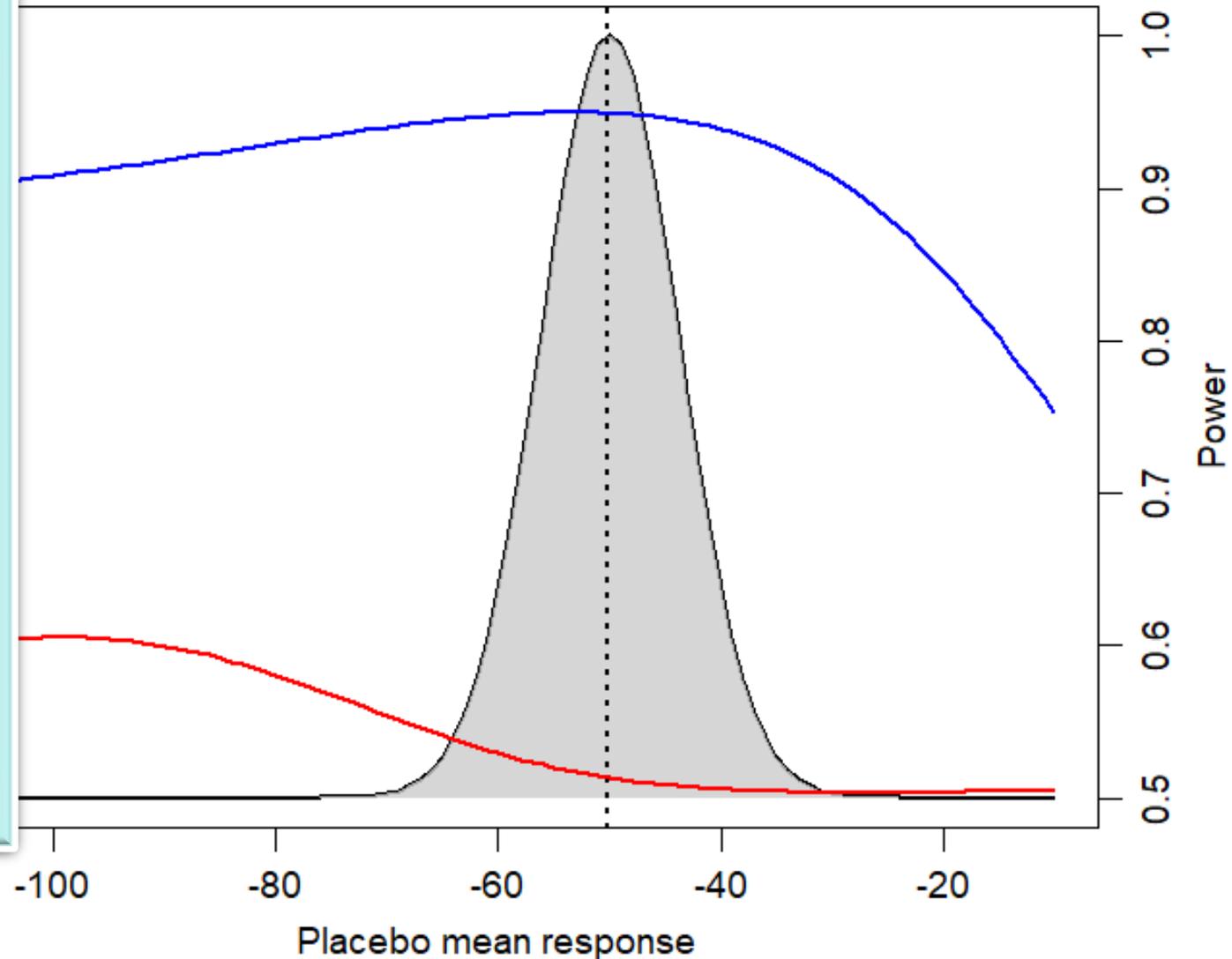Type I error

Power

Placebo mean response

Type I error and Bayesian dynamic borrowing (BDB)

Other potential metrics include:

➢ Maximum Type I error **with** corresponding probability of occurance

➢ "Averaged" Type I error

➢ Maximum or "Averaged" Type I error **over range of interest**

➢ Similar metrics for power

Best et al (2024) *Statistics in Biopharmaceutical Research*

Other considerations

- Point estimate and interval **estimation**
  - How to report treatment effect in SmPC?

PFIZER-BIONTECH COVID-19 VACCINE

FACT SHEET FOR HEALTHCARE PROVIDERS

For participants without evidence of SARS-CoV-2 infection prior to 7 days after Dose 2, vaccine efficacy against confirmed COVID-19 occurring at least 7 days after Dose 2 was 95.0% (95% credible interval: 90.3, 97.6), which met the pre-specified success criterion. The case split was 8 COVID-19 cases in the COMIRNATY group compared to 162 COVID-19 cases in the placebo group.

FDA Food and Drug Administration (.gov)
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 using 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 of BOOSTRIX against pertussis in infants whose mothers were immunized during pregnancy.[5, 6, 7, 8] 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.

Nucala approved for use in China

Jan 16, 2024

*The primary endpoint was the frequency of clinically significant exacerbations of asthma and the **Bayesian Dynamic Borrowing (BDB) analysis showed a clinically meaningful and statistically significant reduction** in the rate of clinically significant exacerbations of asthma for Nucala 100 mg SC compared with placebo over the 52week treatment period. The **frequentist analysis provides further support for the treatment benefit** of Nucala in Chinese participants, with an identified **rate ratio of 0.35 (95% CI [0.24,0.50])**. The frequency of exacerbation in this study was reduced by 65% among participants treated with SC Nucala compared with placebo.*

Concluding remarks

Bayesian methods are mathematically and scientifically rigorous and have a track record of practical application in all areas of science and society

Bayes is not a magic bullet
➢ but it is an essential tool for drug developers and regulators/decision makers

Requires investment in education & understanding of Bayes for both statisticians and clinicians

Requires commitment to rigorous upfront multi-disciplinary discussions to align on key assumptions, choice of design elements, applicability of prior data etc.