Lecture 2: Bayesian methods for (adaptive) clinical trials

David Robertson

david.robertson@mrc-bsu.cam.ac.uk

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Outline

Recap of course learning objectives

- Understand the potential benefits and drawbacks of using adaptive designs.
- Learn about the breadth and depth of the toolbox offered by adaptive designs.
- Know where to search for more information on adaptive designs.
- Understand the difference between using Bayesian methods for design and/or inference.
- Learn about the current regulatory environment regarding Bayesian methods.
- See real-life examples of using Bayesian methods for clinical trials.
- Learn about key aspects of complex innovative designs and their practical implementation.
- Understand the operation of different dose-finding methodologies and how they differ.
- Learn about key aspects of Project Optimus and the importance of Bayesian adaptive designs in early phase dose escalation/optimization trials.
- Understand the relation between design, implementation and analysis aspects of complex adaptive designs (through an example using Bayesian response-adaptive randomization).

Outline

What is a *Bayesian* adaptive design?

- Q: Does 'Bayesian' refer to
  1. The inference procedure used for the final analysis?
  2. The design of the adaptive design itself?
  3. Both?

What is a *Bayesian* adaptive design?

- Q: Does 'Bayesian' refer to
  1. The inference procedure used for the final analysis?
  2. The design of the adaptive design itself?
  3. Both?
- Choice of inferential procedure can depend on regulator preferences (more later)
- Some trials have Bayesian design aspects but inference focuses on frequentist operating characteristics
- Proposed definition of a Bayesian design: a design rule that depends recursively on the posterior probability of the parameters of interest

Areas with Emerging Bayesian Methods Use

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

- Early-phase (phase I/II) drug/biologic development
  - E.g. Dose finding/dose ranging studies
  - Bayesian model-based design and analyses becoming increasingly common
  - *See Lecture 3 for more!*
- Noninferiority (NI) trials
  - Determining the NI margin – involves synthesising data from past studies
- Pediatric Clinical Trials
  - Allows the incorporation of prior information about efficacy from adults or adolescents, *when clinically appropriate*
- Rare diseases
  - Use data to inform a prior distribution – e.g. earlier clinical trials, registry data
- Subgroup assessment
  - Bayesian hierarchical models

Outline

Relevant FDA Guidance

- **Adaptive Designs for Clinical Trials of Drugs and Biologics** (November 2019)
  `https://www.fda.gov/media/78495/download`
- **Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products** (December 2020)
  `https://www.fda.gov/media/130897/download`
- By the end of FY 2025, the FDA also anticipates publishing draft guidance on the use of Bayesian methodology in clinical trials of drugs and biologics
- See also useful concepts and discussion in FDA **Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials** (February 2010)
  `https://www.fda.gov/media/71512/download`

Other relevant FDA documents/initiatives

- Ionan et al. (2023), "Bayesian Methods in Human Drug and Biological Products Development in CDER and CBER", *Therapeutic Innovation & Regulatory Science*, 57:436–444
- *Bayesian Supplemental Analysis (BSA) Demonstration Project*, CDER
  `https://www.fda.gov/about-fda/cder-center-clinical-trial-innovation-c3ti/bayesian-supplemental-analysis-bsa-demonstration-project`
- *CID Paired Program Trial Case Studies*
  `https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program#case%20studies`

Adaptive Designs for Clinical Trials of Drugs and Biologics (2019)

FDA Guidance for industry

Highlighted comments:

- *In general, the same principles apply to Bayesian adaptive designs as to adaptive designs without Bayesian features*
- *Trial designs that use Bayesian adaptive features may rely on frequentist of Bayesian inferential procedures to support conclusions of drug effectiveness*
- *For trials that use Bayesian inference with informative prior distributions, such as trials that explicitly borrow external information, Bayesian statistical properties are more informative than Type I error probability*
- *. . . trial simulations can be particularly resource-intensive for Bayesian adaptive designs . . . it may be critical to use simulations . . . to evaluate the chance of an erroneous conclusion*

Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (2020)

FDA Guidance for industry

Highlighted comments from **Recommended Common Elements of CID Proposals**:

- *When external information is explicitly borrowed into a design, such as in a Bayesian framework, a rationale for the borrowing and an explanation of how the prior distributions were constructed from the prior information [is important]*
- *When Type I error probability is not applicable (e.g., some Bayesian designs that borrow external information), appropriate alternative trial characteristics should be considered.*
- *For Bayesian inference, it is informative to assess the sensitivity of trial operating characteristics to the choice of a prior distribution.*

Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (2020)

FDA Guidance for industry

Recommended Elements of Bayesian CID Proposals:

- *Bayesian approaches may be well-suited for some CIDs intended to provide substantial evidence of effectiveness because they can provide flexibility in the design and analysis of a trial, particularly when complex adaptations and predictive models are used.*
- *Bayesian inference may be appropriate . . . [to] systematically combine multiple sources of evidence, such as extrapolation of adult data to pediatric populations, or to borrow control data from Phase 2 trials to augment a Phase 3 trial.*
- *. . . sponsors should include a rationale for the borrowing with specific details regarding how bias was avoided in the selection of the borrowed information.*
- *FDA's evaluation of the [Bayesian CID] proposal relies on clear communication between the sponsor and FDA regarding two areas: the prior distribution and the study decision criteria for primary and key secondary endpoints.*

Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (2020)

FDA Guidance for industry

Prior distributions:

- *. . . discussions regarding the prior distribution are particularly important to FDA's evaluation of Bayesian proposals*
- *Any data or other external information used to form the prior distribution should be presented in detail for FDA to understand the source and completeness of the external information, its relevance, and the quality and reliability of the data*
- *[T]he issue of exchangeability . . . should be addressed*
- *A Bayesian proposal should also include a discussion explaining the steps the sponsor took to ensure information was not selectively obtained or used*
- *. . . where direct downweighting of the historical data or other non-data-driven features are incorporated in a prior distribution, the proposal should include a rationale for the use and magnitude of these features.*

Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (2020)

FDA Guidance for industry

Decision criteria:

- *. . . When Bayesian approaches are used, it is important to specify alternate decision criteria [than frequentist ones]*
- *Sponsors should propose decision criteria in study protocols for all primary and secondary endpoints intended to be included in product labeling if the product is approved. These proposals should include a rationale for the choice of criteria.*
- *For some Bayesian designs, it is possible to use simulations to estimate the frequentist operating characteristics of power and Type I error probability. In these cases, decision criteria can be chosen to provide Type I error control at a specified level.*

Outline

Case study 1

Therapeutic Hypothermia RCT

- Multiple RCTs had demonstrated the benefit of therapeutic hypothermia in newborns with hypoxic–ischaemic encephalopathy (HIE) when initiated within 6h of birth
- There can be practical difficulties with such a rapid intervention $\rightarrow$ interest in assessing the effect of initiating therapeutic hypothermia at time points from 6-24h after birth in a RCT (Laptook et al., 2017)
- **Rare condition** $\rightarrow$ enrolment was a concern $\rightarrow$ frequentist approaches infeasible
- Bayesian approach specified, with information borrowed from historical data to create prior distributions

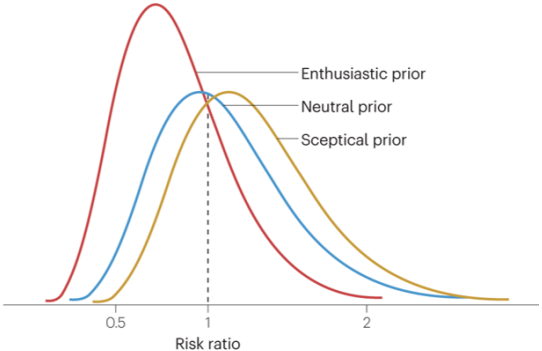
Case study 1

Therapeutic Hypothermia RCT

- Primary outcome = death or disability at 18-22 months of age
- Sample size = 168 = largest sample that could be attained in a feasible time interval (6-year enrollment)
- Primary results of the trial were expressed as an estimated Risk Ratio (RR) and a *probability* that therapeutic hypothermia initiated 6–24h after birth resulted in better outcomes at 18–22 months than the non-cooling standard of care

Case study 1

Therapeutic Hypothermia RCT



Ruberg et al. (2023)

For each prior, 95% of the probability of the RR distribution lies in the interval 0.5–2.0 (plausible values)

Case study 1

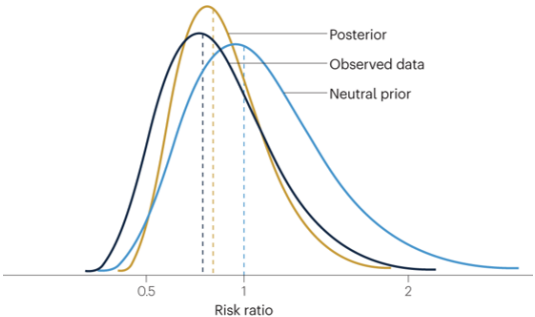
Therapeutic Hypothermia RCT

- All analyses followed intention-to-treat (ITT) principle
- A binomial model was used with a log link to estimate the posterior RR for different binary outcomes for the hypothermia group compared with the noncooled group
- Three main effects:
  1. Treatment (hypothermia or noncooling)
  2. Age at time of randomization ($\leq$ 12 hours or $>$ 12 hours)
  3. Level of encephalopathy (moderate or severe)
- There were 168 participants and 83 were randomly assigned to therapeutic hypothermia and 85 to noncooling

Case study 1

Therapeutic Hypothermia RCT

- Results with neutral prior: RR = 0.86 (95% credible interval: 0.58 – 1.29)
- 76% chance that therapeutic hypothermia reduced mortality and disability relative to the non-cooling standard of care

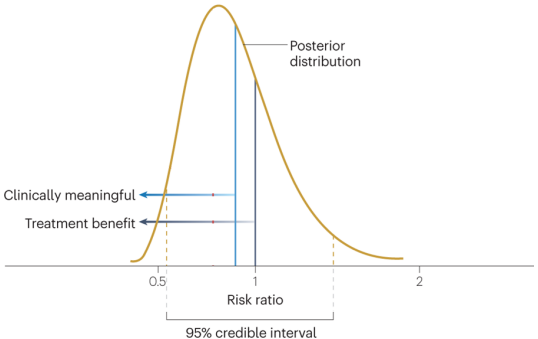


Ruberg et al. (2023)

Case study 1

Therapeutic Hypothermia RCT

- That's not all! Because a Bayesian analysis produces a posterior distribution, other clinically meaningful questions can be answered.
- A 2% reduction in mortality or disability was considered clinically meaningful
- Bayesian analysis: 64% probability that therapeutic hypothermia met that goal.



Ruberg et al. (2023)

Case study 1

Therapeutic Hypothermia RCT

- Bayesian analysis with neutral prior: RR = 0.86 (95% credible interval: 0.58–1.29)
- 76% chance that therapeutic hypothermia reduced mortality and disability relative to the non-cooling standard of care

Case study 1

Therapeutic Hypothermia RCT

- Bayesian analysis with neutral prior: RR = 0.86 (95% credible interval: 0.58–1.29)
- 76% chance that therapeutic hypothermia reduced mortality and disability relative to the non-cooling standard of care
- Frequentist analysis: RR = 0.81 (95% confidence interval of 0.44 – 1.51)
- p-value of 0.42

Case study 1

Therapeutic Hypothermia RCT

Use of Bayesian methods:

- **Advantages**
  - Use of informative prior distribution (based on historical data) $\rightarrow$ reduction in sample size, cost and time
  - Can answer clinically meaningful questions through use of posterior distribution
  - Not just a 'failed' trial with p-value > 0.05
- **Disadvantages**
  - Issues of interpretability/communication around using three different priors
  - Is a 76% chance 'high enough'?

Case study 2

Basket trial in Oncology

- Context: multiple nonmelanoma cancers with BRAF V600 mutations
- Aim: systematically explore the efficacy of vemurafenib in nonmelanoma cancers
- Total of 122 patients enrolled $\rightarrow$ 95 entered 6 modules (Hyman et al., 2015)

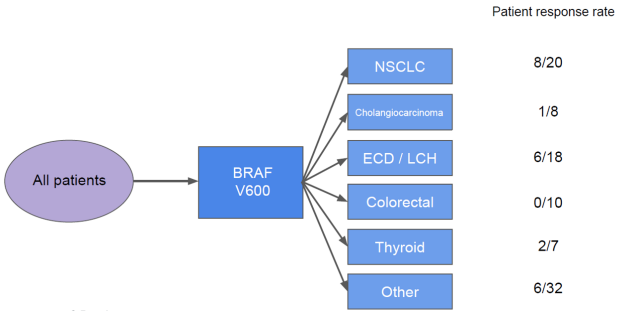


Figure from Haiyan Zheng, University of Bath

Case study 2

Basket trial in Oncology

Potential analysis strategies

- Stand-alone analyses
- Complete pooling
- Borrowing of information

Borrowing of information

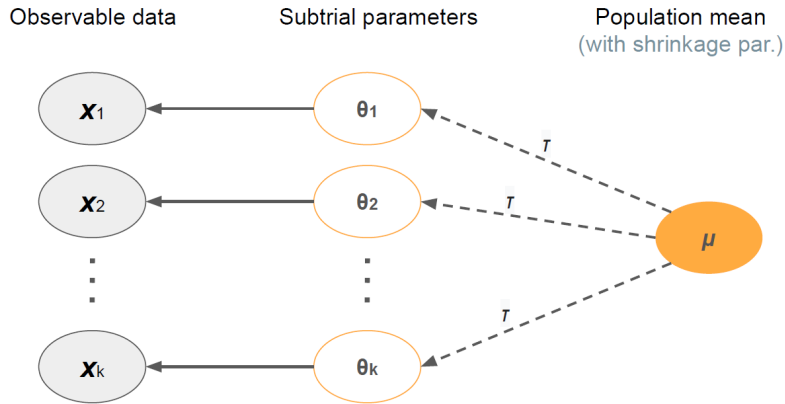
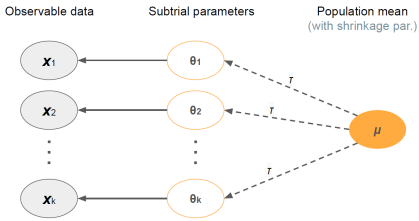


Figure from Haiyan Zheng

Borrowing of information: Bayesian hierarchical model

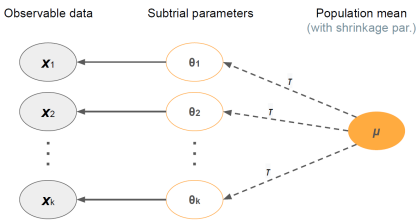


Bayesian hierarchical model for binomial data:

$$y_i|p_i, n_i \sim \text{Binomial}(n_i, p_i)$$
$$\text{logit}(p_i) = \theta_i$$
$$\theta_i|\mu, \tau \sim N(\mu, \tau^2)$$

- Assumes exchangeability (similarity) of θ_i
- Degree of borrowing determined by τ

Borrowing of information: EXNEX model

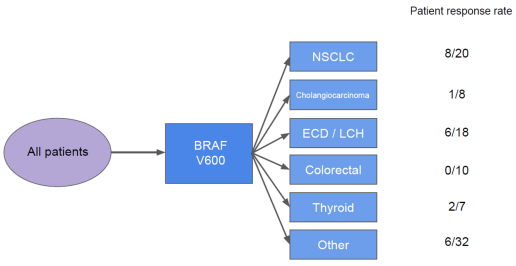


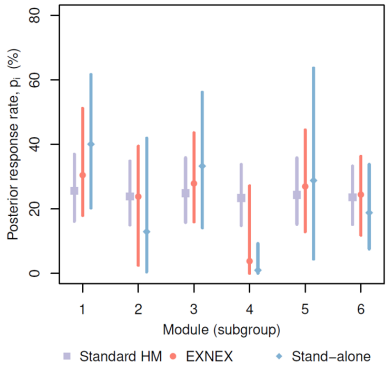
- May be too restrictive to suppose that all θ_i are exchangeable
- Neuenschwander et al. (2016) proposed extension to allow for non-exchangeability
  - **EX:** $\theta_i|\mu, \tau \sim N(\mu, \tau^2)$ with probability w_i
  - **NEX**: $\theta_i \sim N(m_i, s_i^2)$ with probability $1 - w_i$

Case study 2

Basket trial in Oncology

Results





Complete pooling gives response rate $\approx 24\%$

Figures from Haiyan Zheng

Case study 2

Basket trial in Oncology

Use of Bayesian methods:

- **Advantages**
  - Bayesian hierarchical models account for individual differences in the subgroups of interest and borrow strength from the full model $\rightarrow$ can decrease spurious findings + lead to more accurate treatment effect estimates
  - Explicit quantification of assumptions and priors $\rightarrow$ this increases transparency compared with models focused on a single level of analysis
- **Disadvantages**
  - The assumptions made must be plausible
  - Care needed in making assumptions about consistency across subgroups based on insufficient information
  - Sample size calculations can be tricky

Case study 3

FDA CID Case Study

- Context: Children and adolescents with epilepsy with myoclonic-atonic seizures (EMAS)
- Primary endpoint is EMAS-associated seizure frequency over the treatment period
- Proposed study is a multisite, double-blind, randomized, placebo-controlled, parallel group study
- **Rare population** + **Consistent treatment effect across related indications** $\rightarrow$ **Bayesian methods** to formally incorporate previous study results
- Bayesian hierarchical model used (as in Case Study 2)
- **Additional feature**: sample size updated via unblinded interim analyses based on Bayesian predictive probabilities ('Goldilocks methodology')

Case study 3

FDA CID Case Study

Goldilocks method for sample size re-estimation (Broglio et al., 2014)

- *"Is the sample size too big, too small, or just right?"*
- Based on *predictive probabilities* = chance of trial success if the trial continues
- Includes two sources of variability:
  1. The natural variability in the data that has not yet been observed
  2. The variability around the estimate of the treatment effect.
- Predictive probabilities average the trial's probability of success over the posterior distribution of the treatment effect
- Goldilocks design allows for early futility stopping *and* early stopping for predicted success
- Allows decisions based on all (current *and* future) *enrolled* patients, i.e. including those whose outcomes have not yet been observed/follow-up not completed

Goldilocks method for sample size re-estimation (Broglio et al., 2014)

- Let P_n = predictive probability of trial success at the current sample size if the trial were to stop accrual immediately *and* all enrolled patients complete follow-up
- Let P_{max} = predictive probability of trial success if the trial continues to the maximum sample size n_{max}
- At an interim analysis with n patients enrolled:
  - Stop the trial early for predicted success if $P_n > S_n$
  - Stop the trial early for futility if $P_{max} < F_n$
  - If neither criteria satisfied, the trial continues

Goldilocks method for sample size re-estimation (Broglio et al., 2014)

- Binary outcome model for number of observed successes: $y_i \sim \text{Binomial}(n_i, \theta_i)$ where $i = 0, 1$ and $n_i =$ number of currently enrolled patients on arm i
- Prior $\theta_i \sim \text{Beta}(\alpha, \beta)$
- Resulting posterior for θ_i is $\theta_i | y_i, n_i \sim \text{Beta}(\alpha + y_i, \beta + n_i - y_i)$

Predictive Probability of Success at Current Sample Size

- At an interim analysis with n patients enrolled let:
  - z_i = number of observed patients without success on arm i
  - n_i^* = number of currently enrolled patients with an unknown outcome on arm i
  - y_i^* = number among the n_i^* who will ultimately achieve success on arm i
- Hence $y_i^* | y_i, z_i, n_i^* \sim \text{Beta-Binomial}(n_i^*, \alpha + y_i, \beta + z_i)$
- Each possible value of y_i^* has associated probability $\Pr(y_i^*)$
- Predictive probability of success is then

$$P_n = \sum_{y_0^*=0}^{n_0^*} \sum_{y_1^*=0}^{n_1^*} I(y_0^*, y_1^*) \Pr(y_0^*) \Pr(y_1^*)$$

where $I(y_0^*, y_1^*)$ is an indicator function that equals 1 for values of (y_0^*, y_1^*) that would result in trial success

Predictive Probability of Success at the Maximum Sample Size

- Same calculations as before, except that n_i^* = number of currently enrolled patients with an unknown outcome on arm i **and** the $n_{max}/2 - n_i$ patients yet to be enrolled to arm i

Considerations for the proposed design

- What is the impact of the borrowed information?
- Is the proposed approach for borrowing appropriate and interpretable?
- Is the proposed design robust to deviations from the model assumptions?
- What are the statistical properties and performance of the design under various plausible deviations from these model assumptions?
- What is the impact of the Goldilocks adaptation on the operating characteristics in the setting of borrowing?

Simulations

1. Investigate a wide range of prior distributions $\rightarrow$ choose a single prior distribution prospectively that provides acceptable operating characteristics
2. Evaluate the operating characteristics under a wide set of treatment effects and endpoint variability
3. Investigate the behaviour of Bayesian borrowing for single virtual studies (that is, understand how much influence borrowing can have on results of a single observed EMAS data set).
4. Evaluate the performance of the proposed Goldilocks approach.

Case study 3

FDA CID Case Study

- **Advantages**
  - Early stopping $\rightarrow$ potential savings in sample size, cost and time
  - Explicitly accounts for complete follow-up of (all) patients before the primary analysis is conducted.
- **Disadvantages**
  - Questions around robustness of proposed design, especially given trial complexity!
  - More complex trial has practical implications

Further case studies

- Dose finding: Bayesian Optimal Interval Design (BOIN) – see Lecture 3
- Bayesian Response-Adaptive Randomization (RAR) – see Lecture 4

Outline

General software for Bayesian computation

- JAGS (Just Another Gibbs Sampler) – uses Markov Chain Monte Carlo (MCMC)
  - `rjags` package for use in R
- Stan (named after Stanislaw Ulam, inventor of MCMC) – also uses MCMC
  - `rstan` package for use in R
- R INLA package – uses integrated nested Laplace approximation (INLA)
  - Alternative to MCMC with (potentially large) speed advantages

Software for Bayesian adaptive designs

- Dose-finding – see Lecture 3
- BATSS (Bayesian Adaptive Trials Simulator Software) R package
  `https://batss-dev.github.io/BATSS/`
- FACTS (Fixed and Adaptive Clinical Trial Simulator) – Berry Consultants
  `https://www.berryconsultants.com/software/facts/`
- East Horizon Platform – Cytel
  `https://www.cytel.com/bayesian-trial-designs/`
- Some additional software for specific types of (Bayesian) adaptive designs can be found in PANDA, see `https://panda.shef.ac.uk/`

Outline

Framework for deciding on Bayesian (or frequentist) analysis

Figure: Ruberg et al. (2023)

(Perceived) Barriers to the use of Bayesian methods in clinical trials

- Historically, Bayesian methods were computationally intensive/intractable – now *largely resolved* with advances in statistical theory and computational technology
- (Perceived) lack of acceptance and/or familiarity among regulators and industry sponsors
- Lack of experience and (regulatory) guidance about how to use them, especially in confirmatory phase III trials
- More upfront planning is required, e.g. discussion of prior knowledge and external data, selection of a prior distribution and definition of decision rules
- These additional discussions between sponsors and regulators are more time-consuming and require resources.
- No established convention for what constitutes "substantial evidence" of a treatment effect based on Bayesian posterior probabilities or distributions

The future for Bayesian methods?

- Better communication and knowledge exchange
  - Starting step: Bayesian analysis as supplementary?
- Training and software
- Upcoming FDA guidance (2025)

Outline

Additional References

- Ruberg et al. (2023), "Application of Bayesian approaches in drug development: starting a virtuous cycle", *Nature Reviews Drug Discovery*, 22:235–250

Case study 1

- Laptook et al. (2017), "Effect of therapeutic hypothermia initiated after 6 hours of age on death or disability among newborns with hypoxic-ischemic encephalopathy", *JAMA*, 318:1550

Case study 2

- Hyman et al. (2015), "Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations", *NEJM*, 373(8):726–736
- Neuenschwander et al. (2016), "Robust exchangeability designs for early phase clinical trials with multiple strata", *Pharmaceutical Statistics*, 15(2):123–34

Case study 3

- Broglio et al. (2014), "Not too big, not too small: a goldilocks approach to sample size selection", *Journal of Biopharmaceutical Statistics*, 24(3):685–705