

Adaptive and Complex Innovative Designs across trial phases for accelerated approval

Ayon Mukherjee, David Robertson, Sofia S. Villar, Thomas Burnett



UNIVERSITY OF
CAMBRIDGE



Agenda

- 8.30 – 8.40 Welcome and introductions
- 8.40 – 8.45 Learning objectives
- 8.45 – 9.15 Introduction to Adaptive Designs
- 9.15 – 10.00 Bayesian methods for clinical trials
- 10.00 – 10.15 Break
- 10.15 – 11.00 Project Optimus and the framework of CID
- 11.00 – 11.45 (Response) Adaptive Randomization
- 11.45 – 12.00 Recap of learning objectives and Q&A

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)

Lecture slides

https://github.com/dsrobertson/RISW_SC01



Q&As

<https://app.sli.do/event/6u2ri9P3fbviNBTx84Utu8>



Lecture 1: Introduction to Adaptive Designs

Thomas Burnett

tb292@bath.ac.uk

Department of Mathematical Sciences



Outline

Adaptive Designs

How can we use adaptive designs?

Error control and estimation

Master protocols

Bayesian methods

Outline

Adaptive Designs

How can we use adaptive designs?

Error control and estimation

Master protocols

Bayesian methods

Motivation

Attrition rates for new developments (Arrowsmith 2011a, 2011b)

- phase II: >80%
- phase III & submission: ~50%

Reasons for failure (Arrowsmith & Miller 2013)

Phase II (2011–2012)



Phase III & submission (2011–2012)



Motivation

Attrition rates for new developments (Arrowsmith 2011a, 2011b)

- phase II: >80%
- phase III & submission: ~50%

Reasons for failure (Arrowsmith & Miller 2013)

Phase II (2011–2012)



Phase III & submission (2011–2012)



Motivation

Attrition rates for new developments (Arrowsmith 2011a, 2011b)

- phase II: >80%
- phase III & submission: ~50%

Reasons for failure (Arrowsmith & Miller 2013)

Phase II (2011–2012)



Phase III & submission (2011–2012)



Problems and solutions

Likely causes for failure:

- taking forward futile treatments
- studying the wrong patient population
- poor precision (optimal dose, maximum tolerated dose, safety)

Can we do better?

- avoid going straight into large and expensive phase III
- take more care during phases I and II
- consider adaptive and Bayesian designs

Modify an ongoing trial

by design or ad hoc

based on reviewing accrued data at interim

to enhance flexibility

without undermining the study's integrity and validity.

(Chow et al. 2005)

Idea

Modify an ongoing trial
by design or ad hoc
based on reviewing accrued data at interim
to enhance flexibility

without undermining the study's integrity and validity.

(Chow et al. 2005)

Idea

Modify an ongoing trial
by design or ad hoc
based on reviewing accrued data at interim
to enhance flexibility
without undermining the study's integrity and validity.

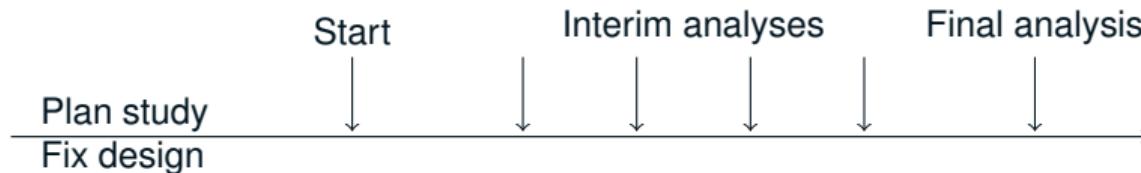
(Chow et al. 2005)

Fixed sample design



- **total** sample size known in advance
- no adjustment possible

Adaptive design



- larger **maximum** sample size
- lower **expected** sample size

At each interim:

- decide whether or not to stop
- change sample size
- drop or add a dose
- change the endpoint
- change the question

Pros and cons

- + highly flexible
- + very efficient
- + reflects medical practice
- + shorter trial and/or more accurate estimates
- + ethical
- highly flexible
- inefficient
- time-consuming to design
- post-trial estimation difficult
- simple estimates may be biased
- interim analyses may require unblinding

Outline

Adaptive Designs

How can we use adaptive designs?

Error control and estimation

Master protocols

Bayesian methods

What, which, who and does?

With so many different adaptive designs it is useful to categorise them by the questions they may help us answer.

Burnett, T., Mozgunov, P., Pallmann, P. et al. Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs. BMC Med 18, 352 (2020). <https://doi.org/10.1186/s12916-020-01808-2>

A ‘Start here’ guide to adaptive designs

PANDA (A Practical Adaptive & Novel Designs and Analysis toolkit)

<https://panda.shef.ac.uk/>

What is a safe dose?

Designs

Continual Reassessment Method (CRM)

An iterative process using all available trial data when choosing the dose for the next cohort of patients.

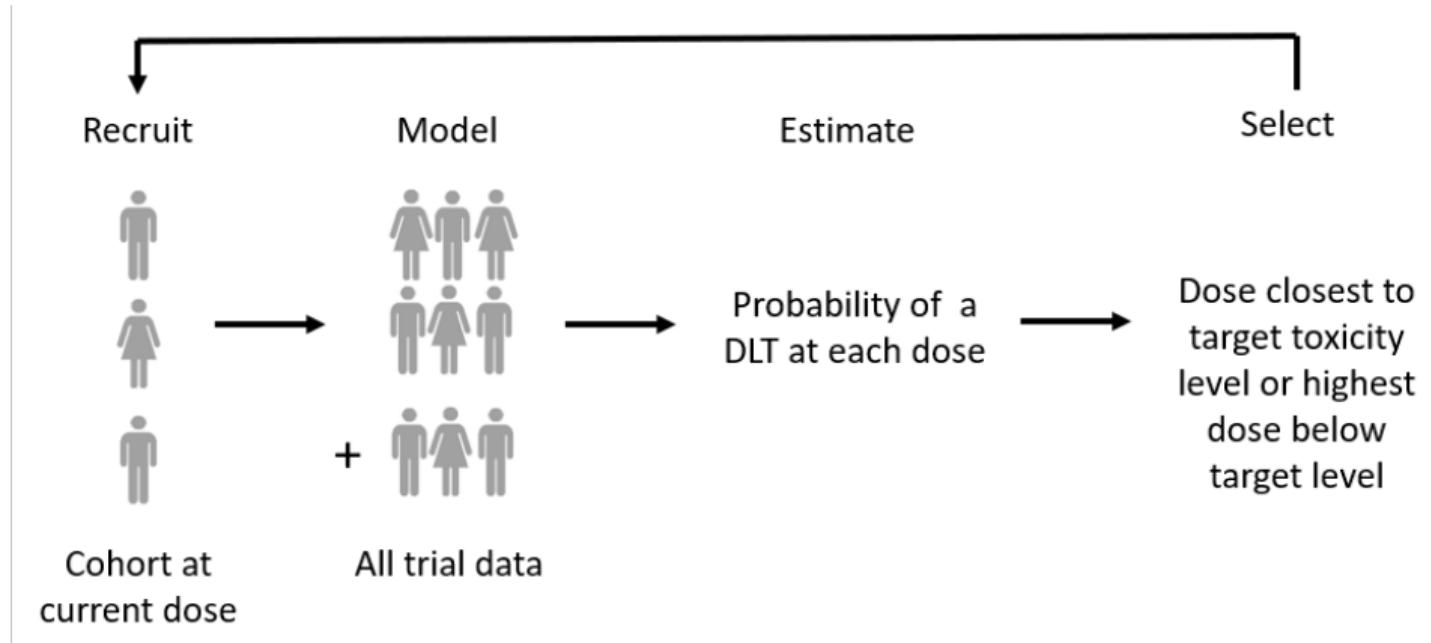
Escalation with overdose control (EWOC)

Skews the allocation criteria of CRM in favour of under-dosing patients rather than overdosing.

Bayesian Optimal Interval Design (BOIN)

Model assisted decision making for dose escalation.

CRM



The potential benefits

- + Decisions based on all trial data rather only the previous cohort.
- + Pre-specified target toxicity level.
- + Higher probability of selecting the dose nearest the maximum tolerated dose.
- + EWOC can reduce the chance of treating at excessively toxic doses.
- Additional complexity in designing the trial.
- Requires input from a suitably trained statistician.
- Can be seen as a black box.

Which is the best treatment?

Designs

Multi-Arm Multi-Stage (MAMS)

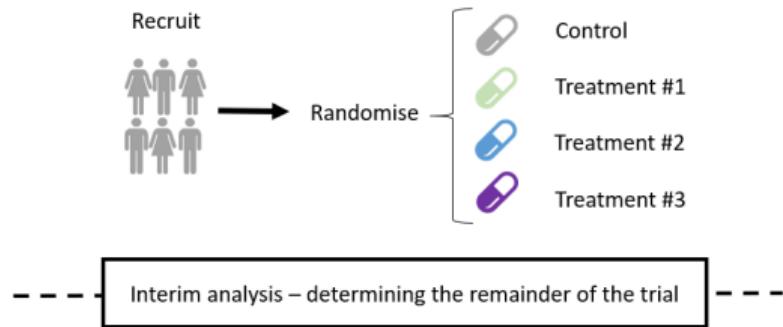
Simultaneous comparison of multiple experimental treatment arms with a common control, allowing for early stopping for futility/efficacy.

Drop the loser (DTL)

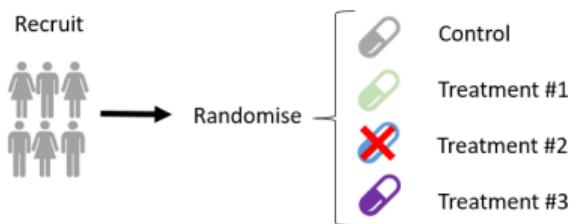
Similar to MAMS with a fixed number of treatments dropped at each stage.

MAMS

Stage 1



Stage 2



Designs

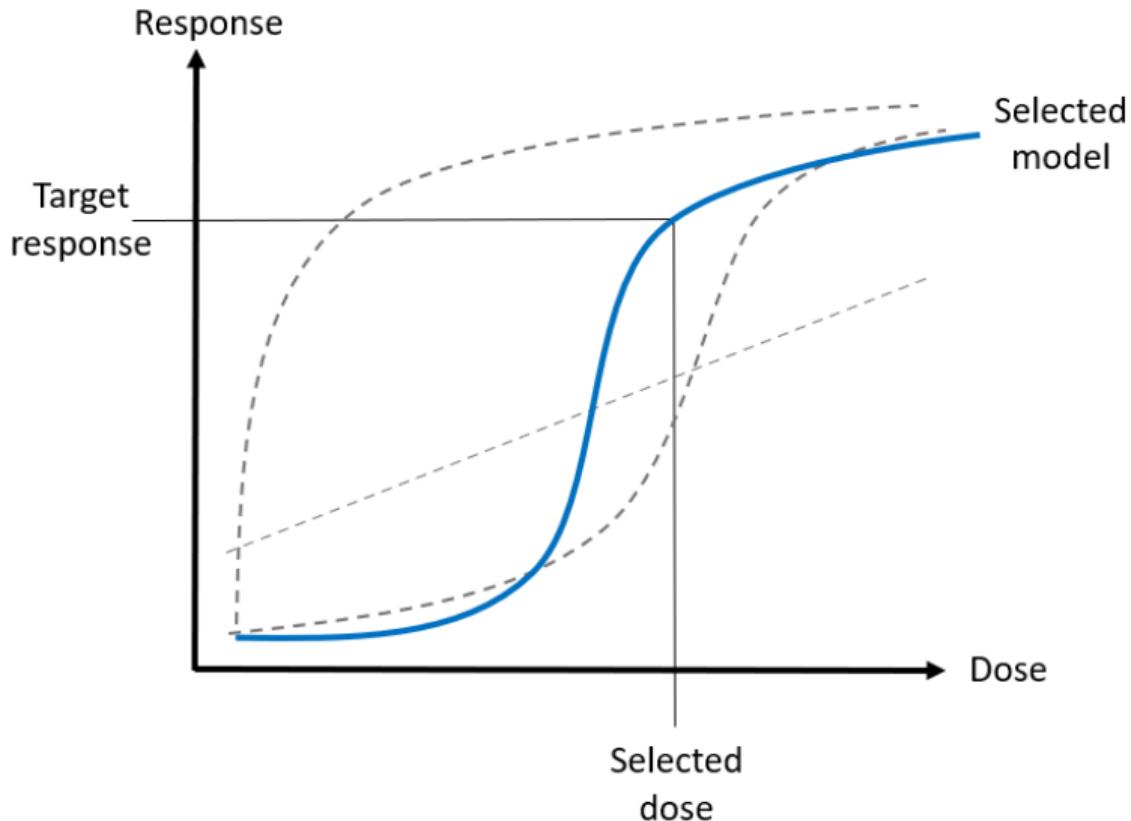
Response-adaptive randomization (RAR)

Modify the allocation probability based on accumulating data, for example based on observed efficacy of the treatment.

Multiple comparison procedures and modelling approaches (MCP-Mod)

Compares several dose-response models to detect whether there is a dose-response signal.

MCP-Mod



The potential benefits

- + Common control groups reduce the number of control patients.
- + Lower expected sample size.
- + Dropping of treatments/doses that do not benefit patients.
- + Increase the proportion of patient benefiting from treatment.
- + Efficient use of trial data through modelling.
- Additional complexity in designing the trial.
- Requires input from a suitably trained statistician.
- Uncertainty about the required number of patients and other logistical challenges.
- DTL can drop effective treatments.
- Require a quickly observed endpoint.
- MCP-Mod can be sensitive to model assumptions.

Who will benefit?

Designs

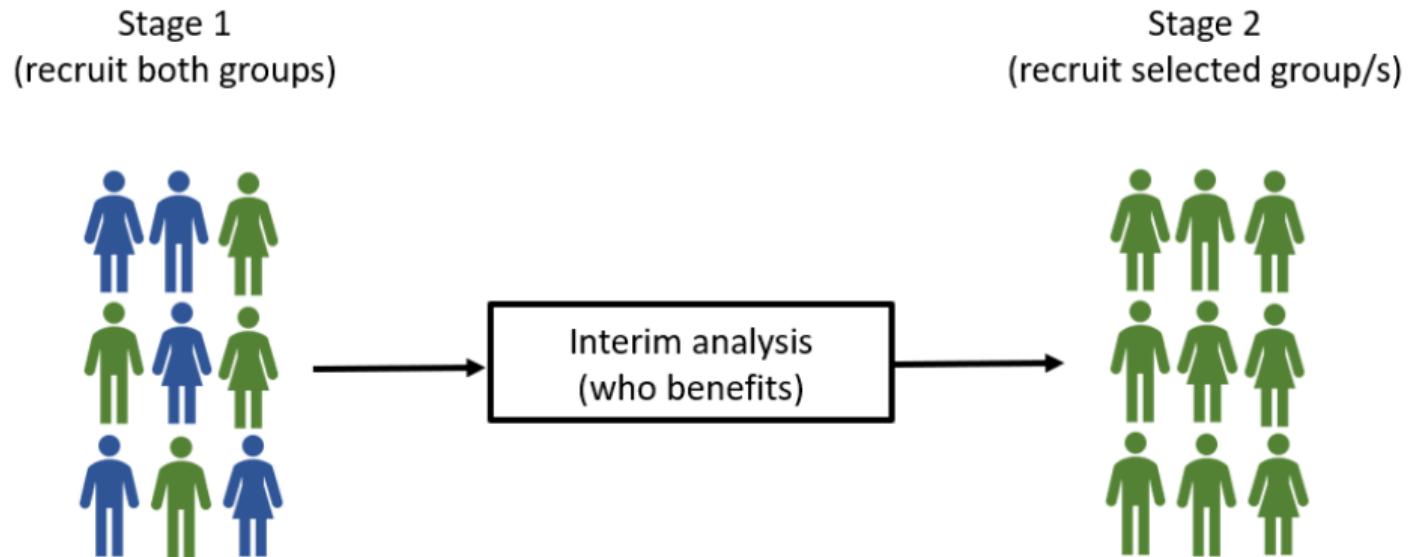
Covariate adjusted response adaptive

A form of RAR where randomization probabilities are skewed based on patients' observed bio-marker information.

Adaptive enrichment

Select bio-marker defined sub-groups to recruit from based on observed trial data at pre-planned interim analysis.

Adaptive Enrichment



The potential benefits

- + Introduce balance of efficiency and ethical goals.
- + Recruit fewer patients from patient groups that do not benefit.
- + Offers efficient decision making when uncertainty about who will benefit.
- + Compromise between fixed sample alternatives while maintaining statistical operating characteristics.
- Additional complexity in designing the trial.
- Requires input from a suitably trained statistician.
- Changing eligibility criteria.
- Sensitive to bio-marker definition.
- Computationally intensive.
- Can be inefficient when unnecessarily.

Does the treatment work?

Designs

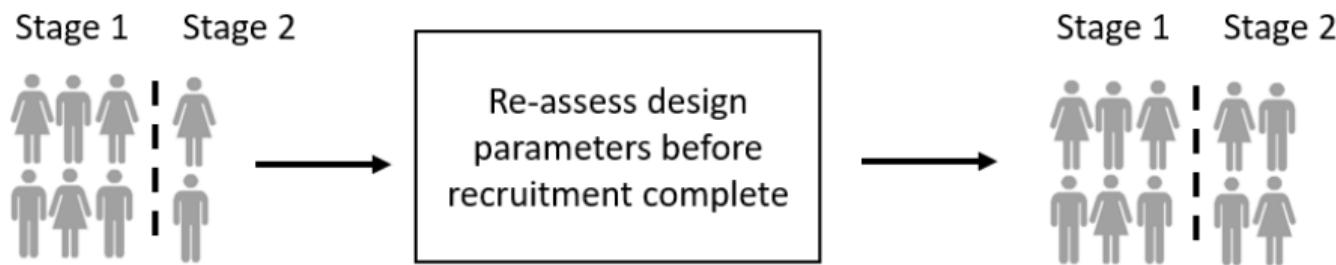
Group sequential designs

Enable early stopping of the trial while maintaining control of the type I error rate using pre-defined stopping rules.

Sample size re-assessment

Target an appropriate sample size for the trial by reassessing uncertainty about key design parameters.

Sample size re-assessment



The potential benefits

- + Reduce the expected sample size.
- + More common in practice.
- + Blinded and unblinded re-assessment possible.
- + Ensure a suitable number of patients are recruited.
- Additional complexity in designing the trial.
- Requires input from a suitably trained statistician.
- Potential increase the in the maximum sample size.
- Uncertainty about when the trial will end.
- Relatively few drawbacks to sample size re-assessment other than the practical issues.

Outline

Adaptive Designs

How can we use adaptive designs?

Error control and estimation

Master protocols

Bayesian methods

Motivation

In a late phase setting we may wish to formally test hypotheses formed about the treatments.

Through their use of interim analysis for decision making adaptive designs present several problems to consider when error control is desired.

Furthermore one may wish to consider this when performing inference, either through understanding the impact on the estimates or through adjusting the estimates accordingly.

Multiple analyses

What are combination tests?

Combination tests allow us to combine multiple trial stages.

They allow us to combine multiple stages of the trial under the assumption of conditional independence.

They yield a single test statistic.

Weighted Inverse Normal

The weighted inverse normal gives a combined p-value $P^{(c)}$ across the stages of the trial.

We find $P^{(c)}$ as follows.

$$Z^{(1)} = \Phi^{-1}(1 - P^{(1)}) \text{ and } Z^{(2)} = \Phi^{-1}(1 - P^{(2)}).$$

With pre-defined w_1 and w_2 such that $w_1^2 + w_2^2 = 1$,

$$Z^{(c)} = w_1 Z^{(1)} + w_2 Z^{(2)}$$

$$P^{(c)} = 1 - \Phi(Z^{(c)}).$$

The conditional error principle

An alternative construction to combination tests.

Given the observations from the Phase II stage of the trial, the conditional error is defined as

$$\alpha' = \mathbb{P}_{\theta=0}(\text{Reject } H_0 | Z^{(1)})$$

The probability of falsely rejecting the null hypothesis for the remainder of the trial must not exceed α' .

Discussion

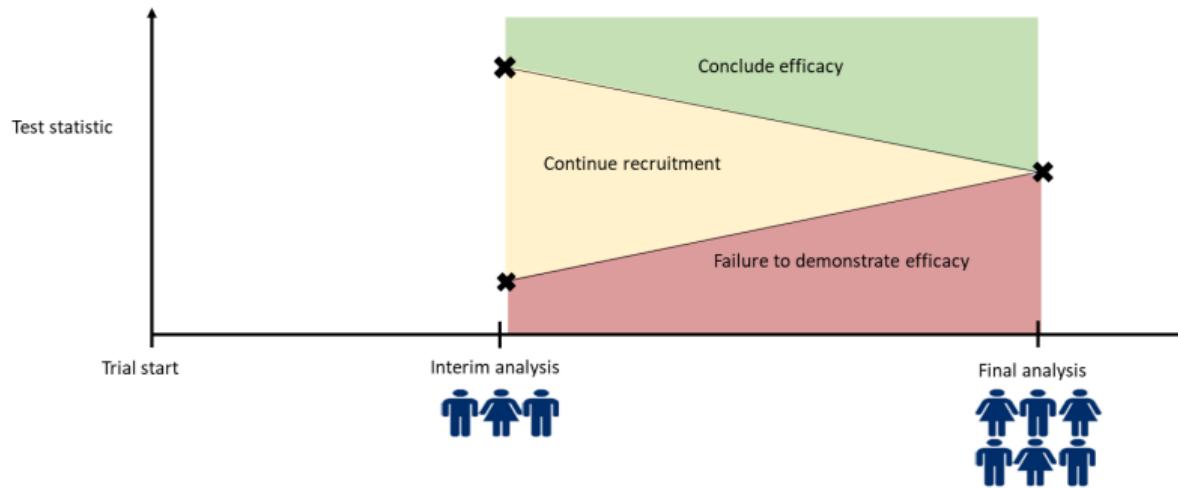
Combination tests and the conditional error principle are general approaches to combining data across a trial.

They can both guarantee error control at the pre-determined level whatever decision is made.

This allows complete freedom in how decisions are made.

Group sequential designs

Group-sequential stopping rules.



- + Stop early for efficacy
- + Drop ineffective treatments
- + Reduce the expected sample size
- higher maximum sample size
- more complex for design and analysis

Multiple hypotheses

Multiple treatments

Suppose in Phase II we wish to explore multiple experimental treatments.

For experimental treatment t we are interested in the treatment effect

$$\theta_t = \mu_t - \mu_0.$$

Hence the corresponding hypotheses are

$$H_{0,t} : \theta_t \leq 0 \text{ vs } H_{1,t} : \theta_t > 0.$$

Error control

There are several considerations when it comes to the error rate now we have multiple hypotheses.

Commonly in the confirmatory setting we require strong control of the FWER

$$\mathbb{P}_{\theta}(\text{Reject any combination of true null hypotheses}) \leq \alpha \text{ for all } \theta.$$

A Bonferroni correction is a well known method for adjusting for multiplicity.

Closed testing

Suppose we have two experimental treatments. We define a closed testing procedure as follows:

- $P_1^{(c)}$ for testing $H_{01} : \theta_1 \leq 0$,
- $P_2^{(c)}$ for testing $H_{02} : \theta_2 \leq 0$,
- $P_{12}^{(c)}$ for testing $H_{01} \cap H_{02} : \theta_1 \leq 0$ and $\theta_2 \leq 0$.

Then:

- Reject H_{01} globally if $P_1^{(c)} \leq \alpha$ and $P_{12}^{(c)} \leq \alpha$.
- Reject H_{02} globally if $P_2^{(c)} \leq \alpha$ and $P_{12}^{(c)} \leq \alpha$.

Estimation

Point estimates

One must consider the potential impacts of adaptation on estimation.

When adaptations are made one must consider the impact on estimation.

Robertson DS, Choodari-Oskooei B, Dimairo M, Flight L, Pallmann P, Jaki T. Point estimation for adaptive trial designs I: A methodological review. *Statistics in Medicine*. 2023; 42(2): 122-145. doi:10.1002/sim.9605

Robertson DS, Choodari-Oskooei B, Dimairo M, Flight L, Pallmann P, Jaki T. Point estimation for adaptive trial designs II: Practical considerations and guidance. *Statistics in Medicine*. 2023; 42(14): 2496-2520. doi: 10.1002/sim.9734

Confidence intervals

Similarly if confidence intervals are desired one must take care in their construction.

Does one desire repeated confidence intervals for each stage of the trial.

What other criteria are important?

Work in progress from the previously mentioned group will help answer these questions.

Regulatory guidance

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry. 2019.

ICH E20 is a work in progress that will further guide the implementation of adaptive designs.

Outline

Adaptive Designs

How can we use adaptive designs?

Error control and estimation

Master protocols

Bayesian methods

What is a Master Protocol?

Platform

Highly adaptable trials to evaluate multiple therapies within a single design. A well-known example of this is the RECOVERY trial.

Umbrella

Multiple biomarker-defined patient subgroups paired with specific treatments.

Basket

Examine a single experimental treatment in multiple biomarker-defined subgroups

Outline

Adaptive Designs

How can we use adaptive designs?

Error control and estimation

Master protocols

Bayesian methods

How are Bayesian methods used?

Bayesian methods represent a fundamental shift from the typically used Frequentist framework. The key difference is that the parameters such as the treatment effect are treated as random variables.

Analysis

One can plan a primary analysis using exclusively Bayesian methods.

Design

Alternatively one may construct a Frequentist analysis framework in such a way as to ensure control of the error rate of the trial while allowing complete freedom in how adaptive design decisions are made.

Lecture 2: Bayesian methods for (adaptive) clinical trials

David Robertson

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



MRC
Biostatistics
Unit



UNIVERSITY OF
CAMBRIDGE

Outline

Introduction

Regulatory guidance

Case studies

Software

Discussion

References

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

Introduction

Regulatory guidance

Case studies

Software

Discussion

References

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

Introduction

Regulatory guidance

Case studies

Software

Discussion

References

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 or 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

Introduction

Regulatory guidance

Case studies

Software

Discussion

References

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 → 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** → enrolment was a concern → frequentist approaches infeasible
- Bayesian approach specified, with information borrowed from historical data to create prior distributions

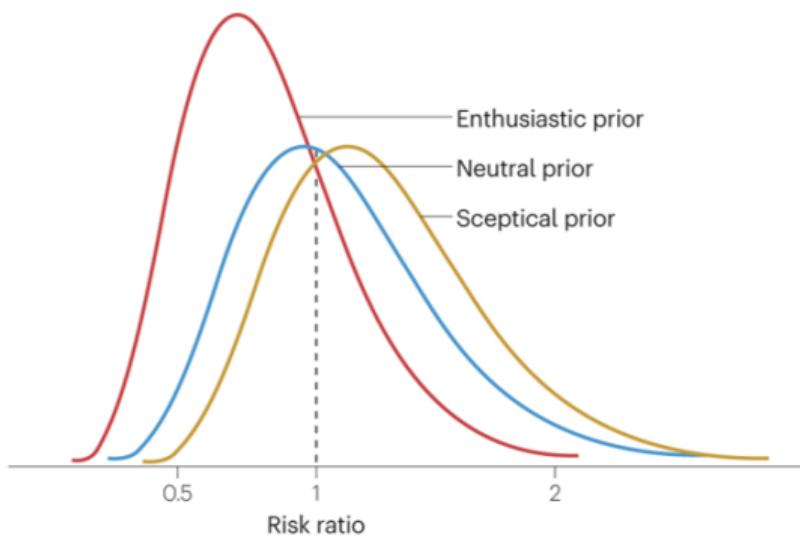
Case study 1

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

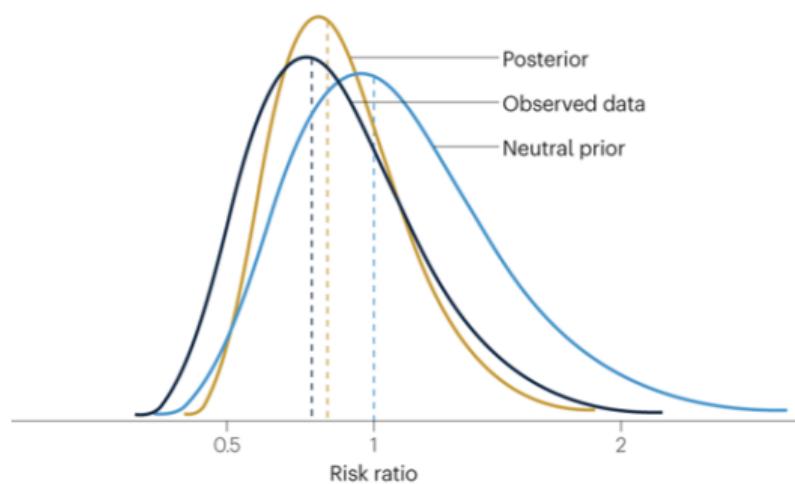
Therapeutic Hypothermia RCT

- 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 (≤ 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

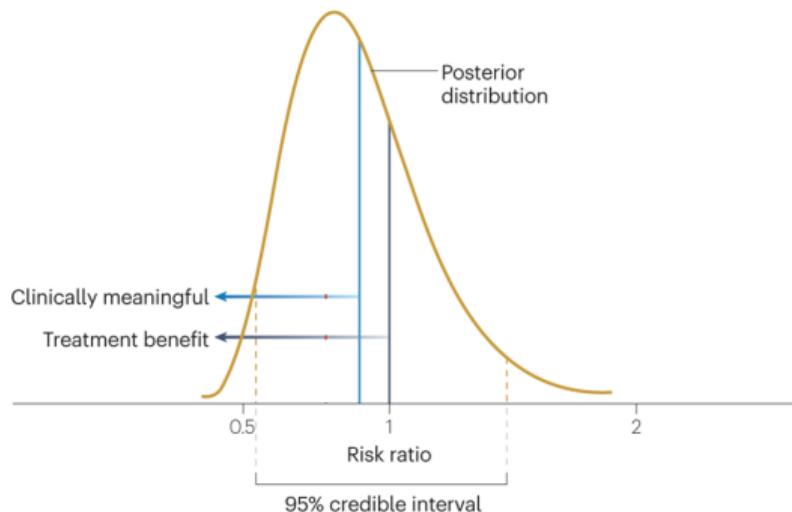
- Results with neutral prior: RR = 0.86 (95% credible interval: 0.58 – 1.29)
- 76% chance that therapeutic hypothermia reduced mortality and disability relative to the non-cooling standard of care



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.



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) → 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 → 95 entered 6 modules (Hyman et al., 2015)

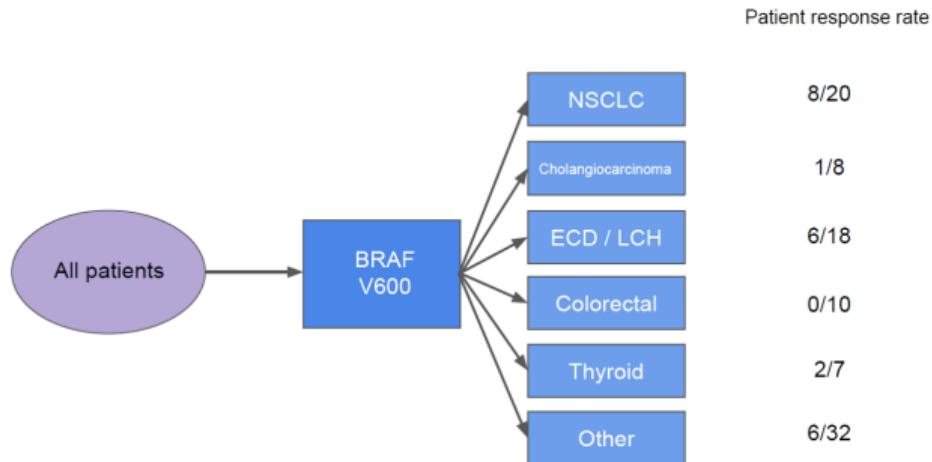


Figure from Haiyan Zheng, University of Bath

Case study 2

Basket trial in Oncology

Potential analysis strategies

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

Case study 2

Basket trial in Oncology

Borrowing of information

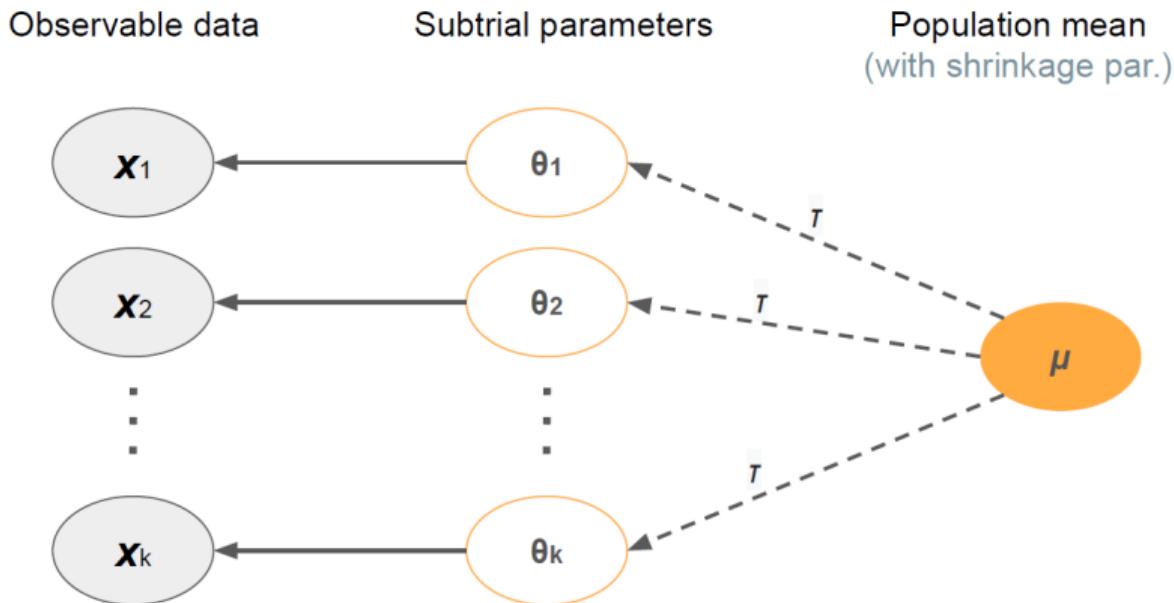


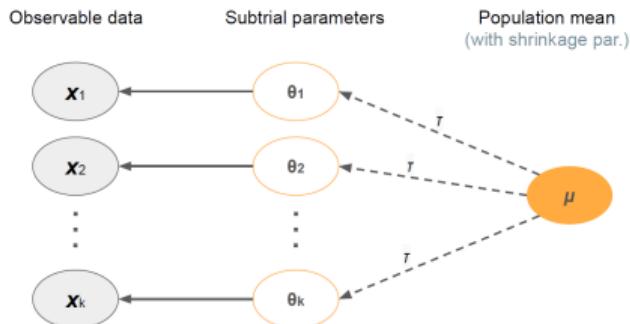
Figure from Haiyan Zheng

Case study 2

Basket trial in Oncology

Borrowing of information: Bayesian hierarchical model

Bayesian hierarchical model for binomial data:



$$y_i|p_i, n_i \sim \text{Binomial}(n_i, p_i)$$

$$\text{logit}(p_i) = \theta_i$$

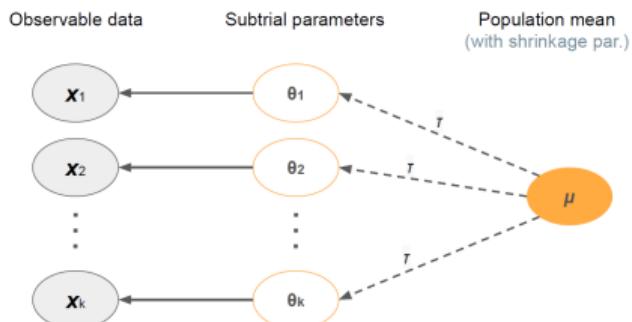
$$\theta_i|\mu, \tau \sim N(\mu, \tau^2)$$

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

Case study 2

Basket trial in Oncology

Borrowing of information: EXNEX model

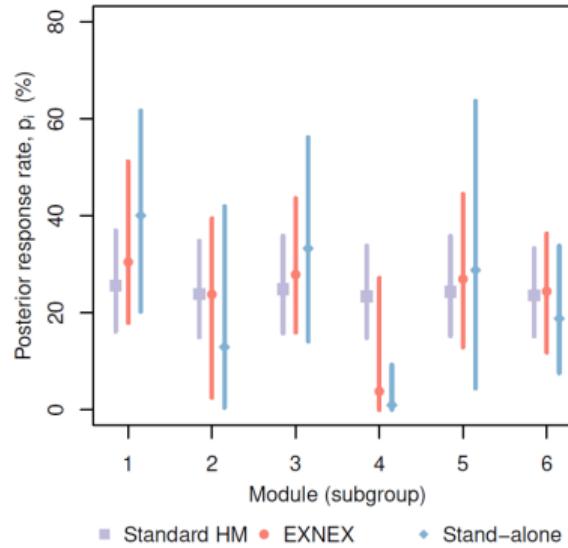
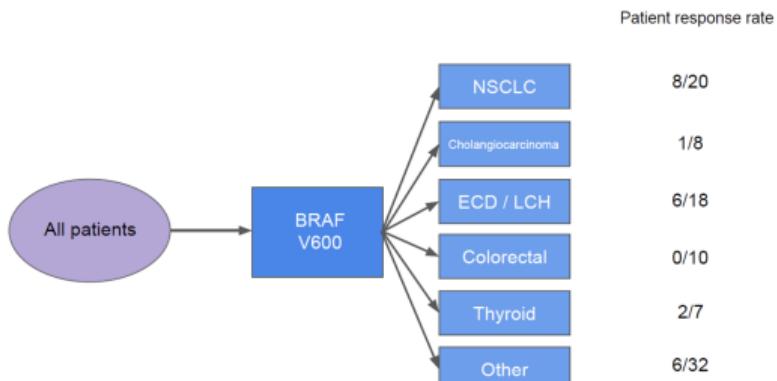


- May be too restrictive to suppose that all θ_i are exchangeable
- Neuenschwander et al. (2016) proposed extension to allow for non-exchangeability
 - **EX:** $\theta_i|\mu, \tau \sim N(\mu, \tau^2)$ with probability w_i
 - **NEX:** $\theta_i \sim N(m_i, s_i^2)$ with probability $1 - w_i$

Case study 2

Basket trial in Oncology

Results



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

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 → can decrease spurious findings + lead to more accurate treatment effect estimates
 - Explicit quantification of assumptions and priors → 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 → 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

Case study 3

FDA CID Case Study

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

Case study 3

FDA CID Case Study

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 $\theta_i|y_i, n_i \sim \text{Beta}(\alpha + y_i, \beta + n_i - y_i)$

Case study 3

FDA CID Case Study

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

Case study 3

FDA CID Case Study

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

Case study 3

FDA CID Case Study

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?

Case study 3

FDA CID Case Study

Simulations

1. Investigate a wide range of prior distributions → 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 → 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

Introduction

Regulatory guidance

Case studies

Software

Discussion

References

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

Introduction

Regulatory guidance

Case studies

Software

Discussion

References

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

Introduction

Regulatory guidance

Case studies

Software

Discussion

References

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

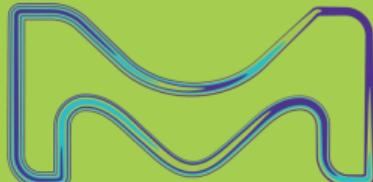
- Broglio et al. (2014), "Not too big, not too small: a goldilocks approach to sample size selection", *Journal of Biopharmaceutical Statistics*, 24(3):685–705

Lecture 3: Project Optimus and the framework of Complex Innovative Designs

Ayon Mukherjee

ayon.mukherjee@merckgroup.com

Director Statistical Methodologist, Advanced Biostatistical Sciences



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).

Complex Innovative Trial Designs

Definition: Clinical trial designs aiming to facilitate and advance the use of complex adaptive, Bayesian, and other novel clinical trial design methodologies to enhance trial efficiency.

Examples of CIDs: Master protocols including umbrella, basket and platform studies, Complex Adaptive Design including Response Adaptive Randomization or Bayesian Adaptive Design requiring design simulations.

Examples of Designs that are not CIDs: Dose escalation designs that include Bayesian design (eg: BLRM, mTPI, BOIN etc.), oncology phase Ib cohort expansions, *The Project Optimus Initiative*.

CIDs may vary but often share similar operational and design requirements:

1. Complex design, set-up, data management, patient randomization.
2. Complex study governance and decision-making.
3. Requirement for significant regulatory involvement.(CID Paired Meeting Program)
4. Need for experienced staff.

Key Criteria for Complex Study Designation

- Tests multiple hypotheses within a single study organized into sub-studies (e.g., master or core protocol)
 1. **Basket Design:** Testing a single therapy against multiple indications
 2. **Umbrella Design:** Testing multiple therapies against a single indication
 3. **Platform Design:** Testing multiple therapies and/or combinations against an indication using shared infrastructure and a common control arm

AND/OR

- Incorporates prospective adaptive elements
 1. Response Adaptive Randomization (RAR).
 2. Covariate-Adjusted Response Adaptive Randomization (CARA).
 3. Early stop for futility or efficacy.
 4. Bayesian or frequentist designs.
 5. Adaptive dose-ranging studies.



Define a core of master protocol, independent of the organization of the documents via appendices or sub-protocols

A linking document and elements that remain constant across any planned sub-study, arm, cohort, or sub-protocol for complex innovative designs

Which elements should always be included in the master protocol?

Which elements should be duplicated in the sub-study or sub-protocol?

Informational Elements

- Rationale
- Background
- Risk benefit analysis
- Disease or pathology specific information
- Definitions
- Objectives and common end points

Administrative Elements

- Registration and PI information
- Synopsis
- Study schema
- Common schedule of assessments
- Inclusion/ exclusion criteria
- sub-study organization

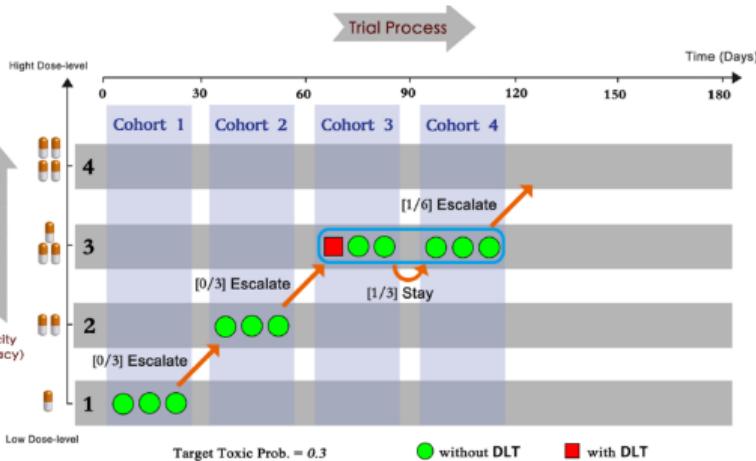
Statistical Elements

- Statistical simulations (if using)
- Power analysis
- Core SAP

Design Elements

- Key hypothesis
- Prospective adaptations
- Planned sub-studies and sub-study organization
- Common control arms (if using)
- Blinding
- Type of design

Traditional Dose Escalation Approach Targeting MTD

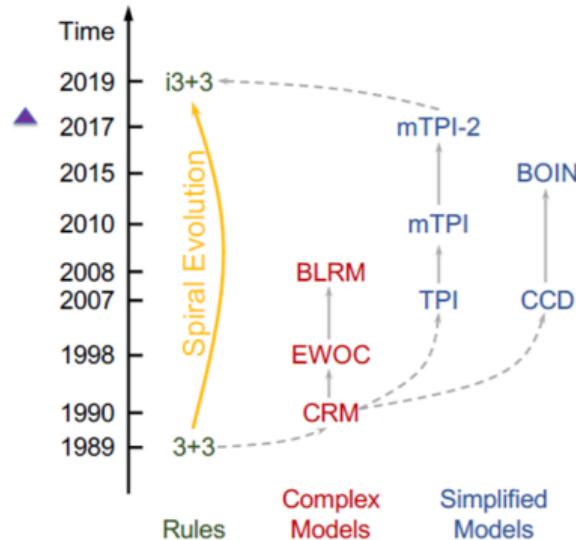
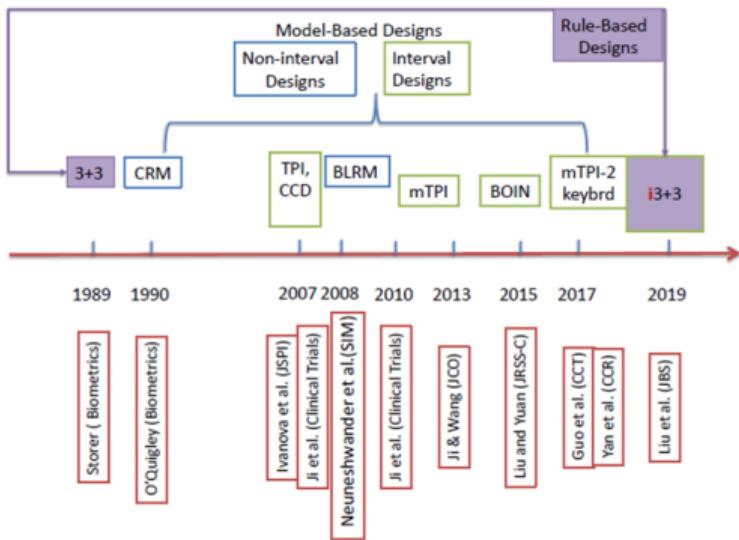


- **Goal** To establish the recommended dose and/or schedule of new drugs or drug combinations for phase II trials.
- Dose-limiting toxicities (DLTs) traditionally are defined by the occurrence of severe toxicities (grade 3 and above) during the first cycle of systemic cancer therapy.
- **Maximum Tolerated Dose (MTD)** The highest dose of a drug or treatment that does not cause unacceptable (DLTs).
- The MTD is determined in clinical trials by testing increasing doses on different groups of people until the highest dose with acceptable side effects is found.

Adaptive Dose Escalation Methods for Cytotoxic Chemotherapies

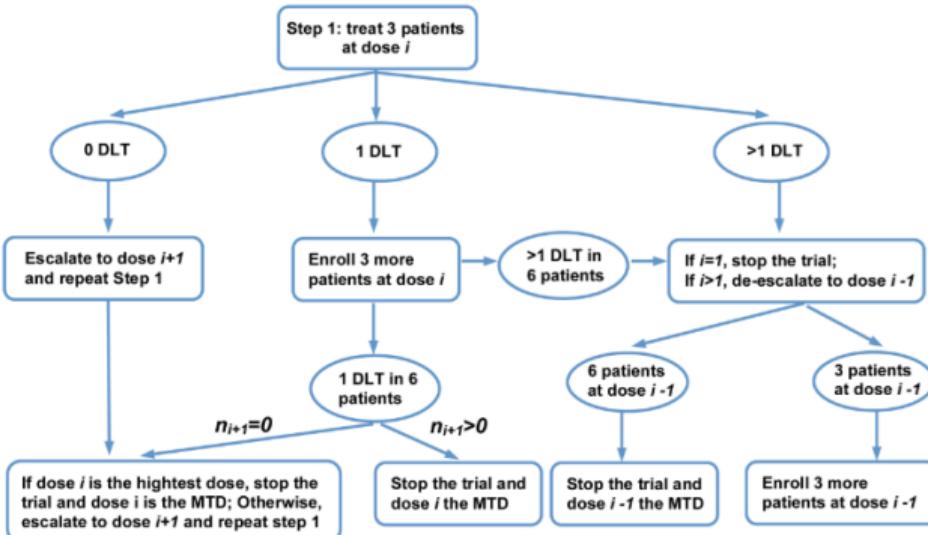
Dose-finding designs over last 30 years

- So many designs are available now.
Which one to use?



Rule Based Design : The 3+3 Method

- A Rule Based Design
- No statistical models needed to implement in practice
 - Societal Acceptance
 - Easy to implement
 - Transparent escalation rule
- Naive and Rigid rule
 - ≤ 6 patients per dose level
 - MTD wide range ($\frac{1}{6}$ to $\frac{1}{3}$)
 - Large variabilities in MTD identification
 - Often little data supporting RP2D choices
- Often identifies a subtherapeutic dose as the MTD



Model Based Dose Finding Methods as an Improvement

The CRM & BLRM designs

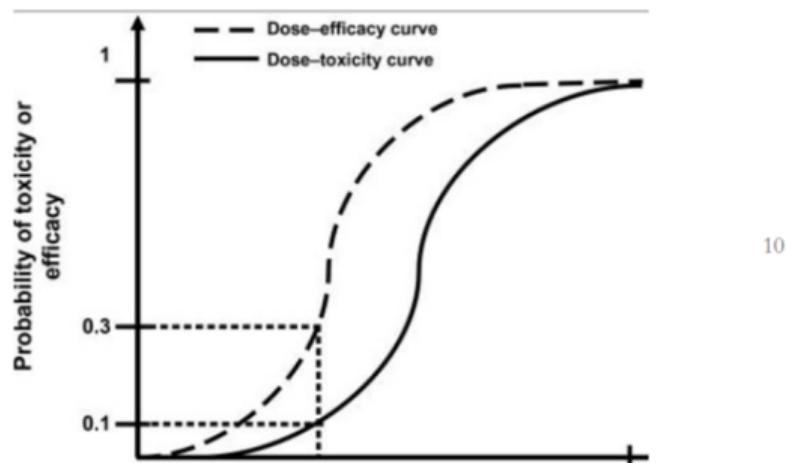
- MTD: a target rate p_T
- BLRM: probability intervals
- Dose-response curve
- $p(x) = p_0(x)\exp(\alpha)$ or $\text{logit}^{(-1)}(x\beta)$
 - $\alpha \sim N(0, 1.34)$; or $\beta \sim \text{prior}$
 - $p_0(x)$ is the “skeleton”
 - Next dose = $\text{argmin}|\hat{p}(x) - p_T|$ or based on posterior prob. of intervals
- Operation
 - Need a statistical expert for inference and decision making
 - Too complex for the clinical team
 - SMC may override dosing decision
 - Ad-hoc rules for over-dose control

- Model based

- Account for variability
- Dose response curves
- Flexible and powerful

- Lots of modifications

- Over-dose control
- Bayesian models
- # of parameters
- Black box, complex, costly



Model Assisted Design : Bayesian Optimal Interval Design (BOIN)

- At a dose j , n_j (e.g., =3, 6, 9) patients are treated, and y_j DLTs observed.
- Let the target DLT rate be ϕ , and ϕ_1 be the lowest toxicity rate below which a dose is considered sub-therapeutic, and ϕ_2 the highest toxicity rate above which a dose is considered excessively toxic. Let p_j denote the true toxicity probability of dose level j .
- Let, $H_1 : p_j = \phi$; $H_2 : p_j = \phi_1$; $H_3 : p_j = \phi_2$ and let $\{\pi_{ij} = P(H_i)\}_{i=1}^3$
- Therefore, $\lambda_e = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right) + n_j^{-1} \log\left(\frac{\pi_{2j}}{\pi_{1j}}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)}$ and $\lambda_d = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right) + n_j^{-1} \log\left(\frac{\pi_{1j}}{\pi_{3j}}\right)}{\log\left(\frac{\phi_2(1-\phi_1)}{\phi(1-\phi_2)}\right)}$
- $\lambda_e = \text{argmax}_{(y_j/n_j)}(P(H_2|n_j, m_j) > P(H_1|n_j, y_j))$ and $\lambda_d = \text{argmin}_{(y_j/n_j)}(P(H_3|n_j, m_j) > P(H_1|n_j, y_j))$.
- When setting $\pi_{1j} = \pi_{2j} = \pi_{3j} = 1/3$, the BOIN design is long-term memory coherent in the sense that the probability of dose escalation (or deescalation) is zero when the observed toxicity rate at the current dose is higher (or lower) than the target toxicity rate ϕ (Theorem 2, Liu and Yuan, 2015).

FDA Fit-for-Purpose BOIN



For FDA's Fit-for-Purpose BOIN (10th December 2021), compare $\frac{y_j}{n_j}$ with intervals

- If $\frac{y_j}{n_j} \leq \lambda_e$, Escalate
- If $\frac{y_j}{n_j} > \lambda_d$, Deescalate
- If $\lambda_e < \frac{y_j}{n_j} \leq \lambda_d$, Stay at the same dose level.

* DLT rate =
$$\frac{\text{Total number of patients who experienced DLT at the current dose}}{\text{Total number of evaluable patients treated at the current dose}}$$

The Sotorasib (Lumakras) Trial

Sotorasib (Lumakras) for NSCLC

Approved in May 2021 for patients with NSCLCs harboring KRAS p.G12C mutation (based on a phase 2 trial)

The first drug successfully targets KRAS, a historically “undruggable” and yet important cancer biomarker

However, a postmarketing trial is required by FDA to further explore lower doses than the approved one

This is due to lack of sufficient dose exploration in early-phase development (e.g., phase 1 with small sample size; dose selection under MTD-regime)

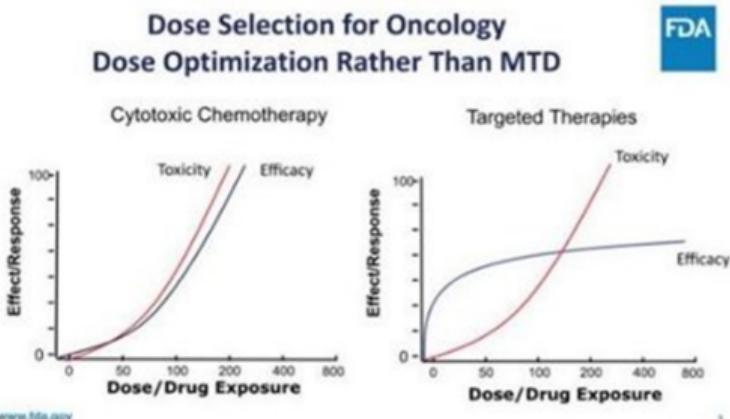
Oncology drugs with post-market dose modification

Examples of Drugs Whose Doses or Schedules Were Modified for Safety or Tolerability after Approval.*			
Small-molecule drugs			
Drug	Initial Dose and Trials	Modified or Added Dose and Trials	Reason for Modified or Added Dose
Ceritinib (Zykadia)	750 mg PO daily fasted (ASCEND-1)	450 mg PO daily with food (ASCEND-8)	Reduce gastrointestinal toxic effects
Dasatinib (Sprycel)	70 mg PO twice daily (CA180013, CA180005, CA180006, and CA180015)	100 mg PO daily (CA180034)	Reduce hematologic toxic effects and fluid retention
Niraparib (Zejula)	300 mg PO daily (NOVA)	200 mg PO daily (PRIMA)	Reduce thrombocytopenia in patients with a lower platelet count or lower body weight
Ponatinib (Iclusig)	45 mg PO daily (PACE)	45 mg PO daily, then 15 mg PO daily once $\leq 1\%$ BCR-ABL is achieved (OPTIC)	Reduce vascular occlusive events
Chemotherapy			
Cabazitaxel (Jevtana)	25 mg/m ² IV every 3 wk (TROPIC)	20 mg/m ² IV every 3 wk (PROSELICA)	Reduce hematologic toxic effects and infections
Antibody-drug conjugates			
Gemtuzumab ozogamicin (Mylotarg)	9 mg/m ² IV on days 1 and 15 (Study 201, Study 202, and Study 203)	3 mg/m ² IV on days 1, 4, and 7 (Mylofrance-1)	Reduce veno-occlusive disease and treatment-related mortality

* Adapted from the Food and Drug Administration.² IV denotes intravenous, and PO by mouth.

- All the listed drugs had to **reduce** their dose or schedule due to toxicity

Project Optimus in Oncology



- Landscape is moving away from cytotoxic chemotherapy into targeted therapies.
- Emphasis is now changing from solely identifying the MTD towards also determining an optimized dose level based on all available clinical data and the dose and exposure-response relationship.
- FDA's Project Optimus aims at finding the dose that balances the efficacy-toxicity profile.
- **Goal:** Educate, innovate, and collaborate with companies, academia, professional societies, international regulatory authorities, and patients to move forward with an unified dose optimization paradigm across oncology.

FDA Project Optimus Initiative for Dose Optimization

Collection and Interpretation of Clinical Pharmacokinetic, Pharmacodynamic, and Pharmacogenomic Data

- Integrate PK/PD/PG data with clinical data (safety and efficacy)
- Investigate effects in multiple populations when possible

Trial Designs to Compare Multiple Dosages

- Backfill patients on multiple doses before dose comparison
- Randomized dose comparison (adaptively and without considering frequentist error rates as the case for late-phase trials) – Bayesian?

Safety and Tolerability -- Endpoints

- DLT and low grade toxicity should be considered – Toxicity burden
- PRO

Subsequent Indications and Usages

- Different doses for different diseases should be considered

Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases

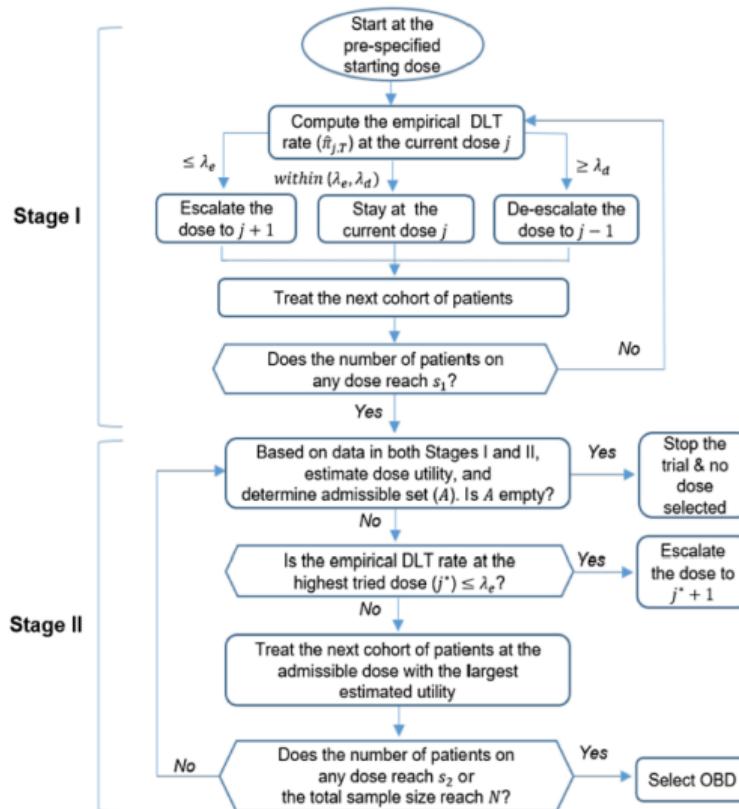
Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Oncology Center of Excellence (OCE)
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

Adaptive Design Methods in Favour of Optimus

- Rule Based Design
 - Jointi3+3 (Lin and Ji, 2020)
 - Bi3+3 (Lie *et. al.*, 2024)
- Model Based Designs
 - EffTox (Thall and Cook, 2004)
 - Backfill CRM (Dehbi, O'Quigley and Iasonos, 2021)
- Model Assisted Designs
 - **U-BOIN** (Zhou *et. al.*, 2019)
 - **BOIN12** (Lin *et. al.*, 2020)
 - BOIN-ET (Takeda *et. al.*, 2018)
 - PRINTE (Lin and Ji, 2021)
 - Comb-BOIN12 (Lu *et. al.*, 2024)

The U-BOIN Design

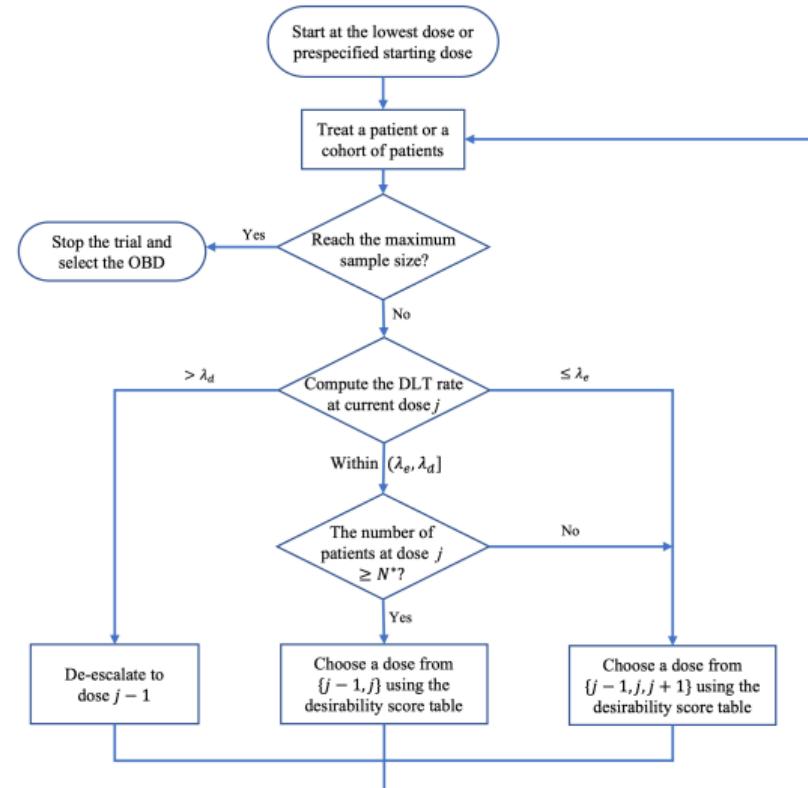


Boundaries	Target DLT rate (ϕ_T)					
	0.15	0.20	0.25	0.30	0.35	0.40
λ_e (escalation)	0.118	0.157	0.197	0.236	0.276	0.316
λ_d (de-escalation)	0.179	0.238	0.298	0.358	0.419	0.480

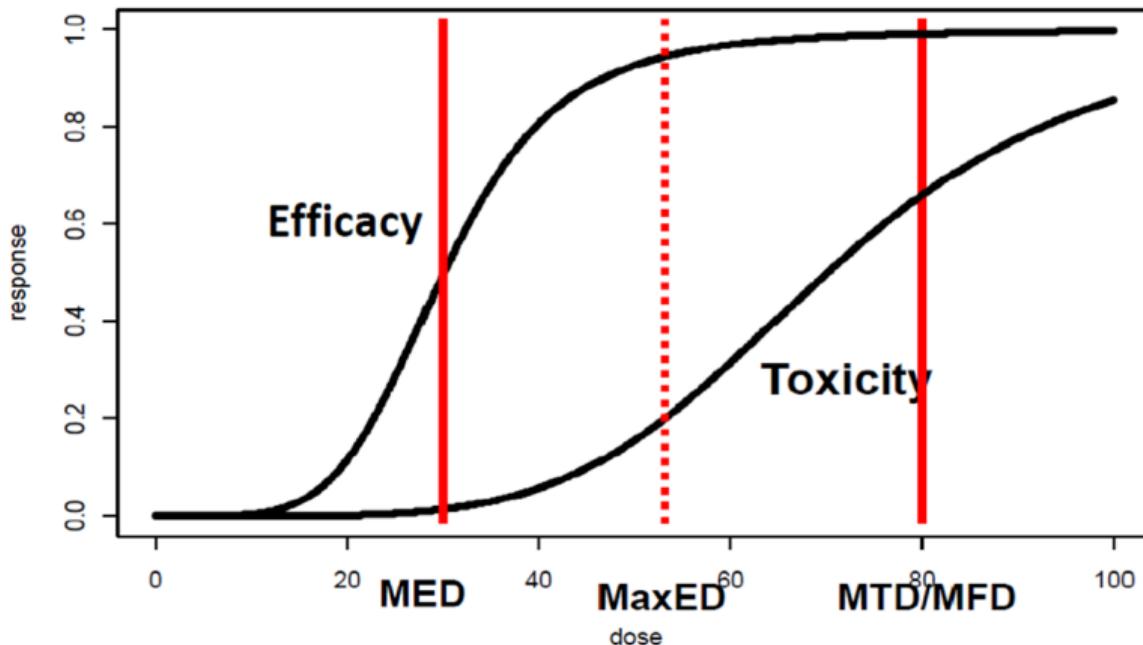
Toxicity	Efficacy	
	Yes	No
No	$u_1 = 100$	$u_2 = 40$
Yes	$u_3 = 60$	$u_4 = 0$

The BOIN12 Design

- Patients are adaptively assigned to the most desirable dose (OBD).
- $\lambda_e = 0.197$ and $\lambda_d = 0.298$. $N^* = 6$, $\phi = 0.25$, $\phi_T = 0.30$, $\phi_E = 0.35$.
- **Admissible dose** : Let $\hat{p}_j = y_j/n_j$.
 $P[\hat{p}_j > \phi_T | \text{data}] < 0.95$ and
 $P[\hat{p}_j > \phi_E | \text{data}] < 0.90$
- Reaching the maximum sample size :
 1. MTD : the dose level that has the isotonically estimated toxicity probability closest to ϕ_T
 2. The final OBD is chosen as the dose level that has the highest estimated utility among the doses that are not higher than the MTD.



Every Drug Has an Appropriate Dose



MED: Minimum Effective Dose

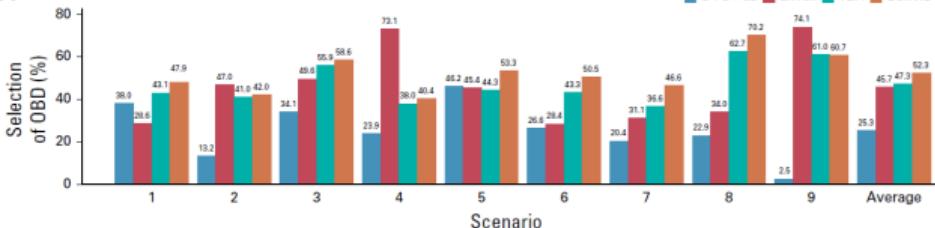
MaxED: Maximum Effective Dose

MTD: Maximum Tolerated Dose

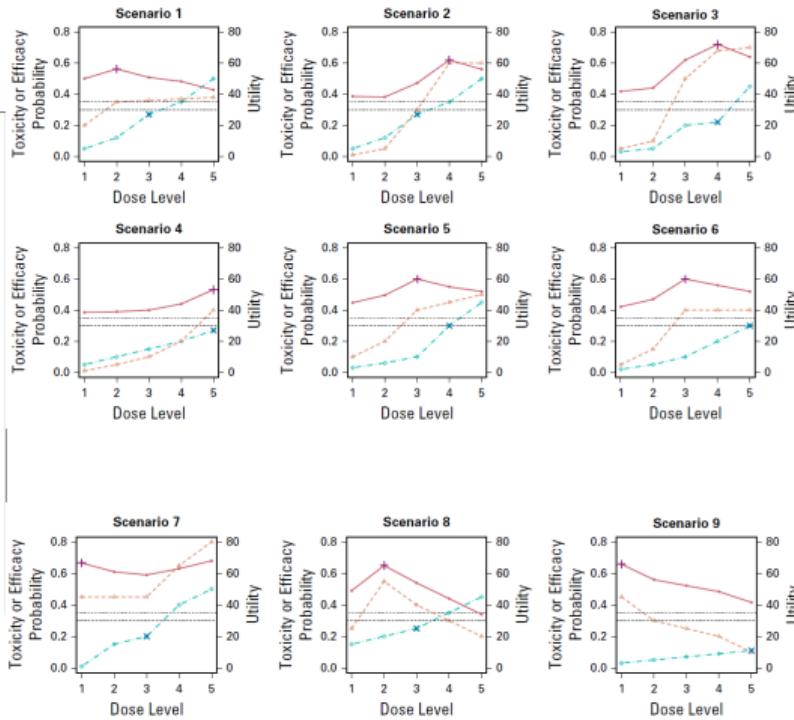
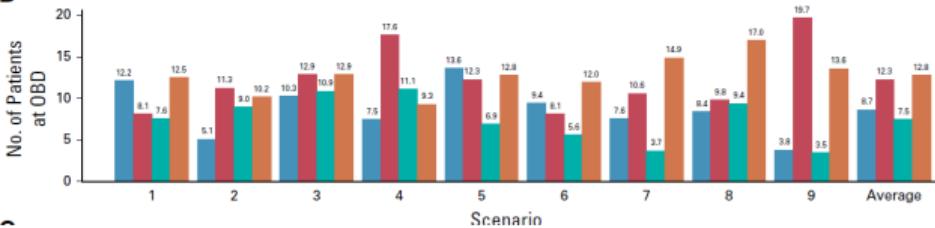
MFD: Maximum Feasible Dose

Simulation Study

A



B



A Real-Life Case Study

- Clark T, Mukherjee A, Lichtlen P, Sweeney J (2024) Bayesian Interval Based Designs for Phase I Dose-Escalation Trials: A Case Study in Oncology. *J Clin Trials.* 14:566

- The primary objective was to estimate MTD, based on the number of dose-limiting toxicities (DLTs) observed at a specific dose level.
- A BOIN design was used and comprised eight planned escalating dose levels; dose levels 1 to 8.



Journal of Clinical Trials

OPEN ACCESS Freely available online

Research Article

Bayesian Interval Based Designs for Phase I Dose-Escalation Trials: A Case Study in Oncology

Tim Clark¹, Ayon Mukherjee^{2*}, Peter Lichtlen³, James Sweeney⁴

¹Drug Development Services, ICON plc., Munich, Germany; ²Regulatory Affairs and Drug Development Solutions (RADDS), IQVIA, Frankfurt, Germany; ³Clinical Development Technology, Numab Therapeutics AG, Horgen, Switzerland; ⁴Department of Mathematics and Statistics, University of Limerick, Dublin, Ireland

ABSTRACT

The objective of phase I dose-escalation clinical trials has generally been to determine the Maximum Tolerated Dose (MTD). However, with the advent of molecular targeted therapies this approach has changed, as dose limiting toxicities are less frequently observed. For this reason, the concept of Optimal Biological Dose (OBD) has been developed, which considers efficacy and toxicity. Several Bayesian model-assisted designs have been proposed to target the MTD more accurately and/or the OBD compared to traditional rule-based approaches such as the 3+3 design. These include the Bayesian Optimal Interval (BOIN) and the BOIN phase I/II (BOIN12) design. The BOIN design targets the MTD, while the BOIN12, which takes both efficacy and toxicity into account in decisions to escalate/de-escalate the dose, targets the OBD. In this article we use a real-life case study to compare the BOIN and the BOIN12 designs under different scenarios and showcase how each of the designs perform when the compound under investigation has a benign toxicity profile. We argue that both efficacy and toxicity should be taken into consideration when designing early-phase oncology studies.

Keywords: Bayesian adaptive designs; Dose escalation; Toxicity-efficacy trade-off; Optimal biologic dose; Phase I trials

A Real-Life Case Study (Continued..)

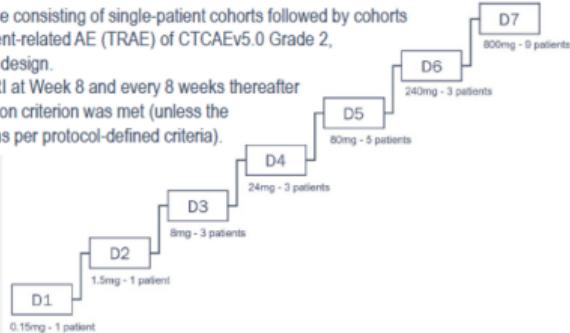
Dose escalation trial design

Treatment:

- Dose escalation started with an accelerated phase consisting of single-patient cohorts followed by cohorts of n≥3 patients upon occurrence of a first treatment-related AE (TRAE) of CTCAEv5.0 Grade 2, as guided by a Bayesian Optimal Interval (BOIN) design.
- Efficacy was assessed by on-treatment CT or MRI at Week 8 and every 8 weeks thereafter until disease progression or another discontinuation criterion was met (unless the investigator elected to treat beyond progression as per protocol-defined criteria).

Inclusion criteria

- ✓ 18 years of age or above
- ✓ Patients with metastatic/ unresectable solid tumors confirmed by pathology/fresh biopsy, with progressing disease since last therapy and for whom there is no available standard of care
- ✓ Measurable disease according to RECIST 1.1
- ✓ ECOG PS 0–1
- ✓ Adequate renal, liver, and hematologic function



Objectives:

- The primary objectives of this dose escalation part of the trial were the characterization of the safety and tolerability profile of NM21-1480, the determination of its maximum tolerated dose (MTD) and the determination of dose level(s) for further evaluation of pharmacodynamics and clinical activity in expansion cohorts.
- The secondary objectives were the establishment of a pharmacokinetic profile and the evaluation of immunogenicity.
- Exploratory objectives comprised the assessment of anti-tumoral activity of NM21-1480, based on RECIST 1.1., the characterization of the pharmacodynamic profile of the compound, and the exploration of potential biomarkers of clinical response.

Methods:

- This is a first-in-human, multicenter, open-label, phase 1/2a trial of NM21-1480 in advanced solid tumors (NCT04442126) (Figure 2).
- The trial consists of two consecutive parts: dose escalation (phase 1 – Part A) and expansion (phase 2 – Part B).
- The dose-limiting toxicity (DLT) monitoring period was 28 days, comprising two full dosing intervals.

- A newly available pharmacodynamic (PD) data suggested that efficacy (DCR) may not increase monotonically with dose and that PD activity might plateau due to the affinity-balanced design of the molecule.
- At high concentrations the target engagers for the drug may become saturated resulting in “insulating effects” that restrict drug activity.

A Real-Life Case Study (Continued..)

- Scenarios

Dose level/ Scenario	1	2	3	4	5	6	7	8	Dose level/ Efficacy Scenario	1	2	3	4	5	6	7	8
S1	0.10	0.30	0.45	0.48	0.51	0.54	0.56	0.58	EFF S1	0.05	0.10	0.20	0.30	0.40	0.35	0.30	0.20
S2	0.04	0.06	0.11	0.16	0.29	0.47	0.55	0.60	EFF S2	0.05	0.20	0.30	0.35	0.40	0.45	0.48	0.50
S3	0.03	0.06	0.10	0.15	0.30	0.46	0.68	0.80	EFF S3	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
S4	0.01	0.03	0.06	0.09	0.11	0.30	0.45	0.60									
S5	0.02	0.03	0.05	0.06	0.07	0.08	0.09	0.10									

- BOIN

TABLE 2 Sample Size Probabilities.

Sample Size/ DLT Scenario	6	9	12	15	18	21	24	27
DLT S1	0.04	0.06	0.12	0.19	0.30	0.29	0.04	0.06
DLT S2	<0.01	<0.01	0.02	0.04	0.06	0.88	<0.001	<0.001
DLT S3	<0.001	<0.01	0.01	0.03	0.07	0.88	<0.001	<0.001
DLT S4	<0.001	<0.001	<0.01	<0.01	0.02	0.96	<0.001	<0.001
DLT S5	<0.001	<0.001	0.03	0.01	0.01	0.98	<0.001	<0.001

Dose level/ DLT Scenario	1	2	3	4	5	6	7	8
DLT S1	0.14	0.65	0.17	0.03	<0.01	<0.001	<0.001	<0.001
DLT S2	<0.001	0.01	0.06	0.24	0.48	0.18	0.03	<0.001
DLT S3	<0.001	<0.001	0.05	0.25	0.49	0.19	<0.01	<0.001
DLT S4	<0.001	<0.001	<0.01	0.03	0.23	0.50	0.20	0.02
DLT S5	<0.001	<0.001	<0.01	<0.01	0.03	0.11	0.27	0.56

Sample Size/ Efficacy Scenario	3	6	9	12	15	18	21	24	27
EFF S1	<0.001	<0.01	<0.01	0.02	0.02	0.03	0.04	0.04	0.85
EFF S2	<0.001	<0.001	<0.01	0.02	0.02	0.02	0.02	0.02	0.89
EFF S3	<0.001	<0.001	<0.01	0.02	0.02	0.03	0.04	0.04	0.84

Dose level/ Efficacy Scenario	Early stop	1	2	3	4	5	6	7	8
EFF S1	0.18	0.38	0.28	0.11	0.04	0.1	<0.01	<0.001	<0.001
EFF S2	0.13	0.28	0.44	0.10	0.03	0.01	<0.01	<0.001	<0.001
EFF S3	0.18	0.39	0.27	0.10	0.03	0.02	<0.01	<0.01	<0.001
BOIN	0.99	0.14	0.65	0.17	0.03	<0.01	<0.001	<0.001	0.0

A Real-Life Case Study (Continued..)

One patient in the 80mg cohort died due to rapid disease progression.

Toxicity profile was extremely benign so the early stopping rule of a maximum of 12 patients exposed at any dose level was not met.

Table: Maximum Tolerated Dose (MTD) selection smoothed DLT rate is closest to the target DLT rate.

Dose Level	Number of Patients	Patients with DLTs	Posterior DLT Estimate	95% Credible Interval	Posterior Toxicity > 0.3
0.15 mg	1	0	0.01	(0.00 , 0.13)	0.03
1.5 mg	1	0	0.01	(0.00 , 0.13)	0.03
8 mg	3	0	0.01	(0.00 , 0.13)	0.03
24 mg	3	0	0.01	(0.00 , 0.13)	0.03
80 mg	5	1	0.01	(0.00 , 0.13)	0.09
240 mg	3	0	0.01	(0.00 , 0.13)	0.09
800 mg	9	0	0.01	(0.00 , 0.13)	0.09

The OBD was determined to be DL4 (24mg).

Table: Summary of Clinical Study Results

Dose Level/ Probabilities	DL1 0.15 mg	DL2 1.5 mg	DL3 8 mg	DL4 24 mg	DL5 80 mg	DL6 240 mg	DL7 800 mg
Pr(Toxicity=0,Efficacy=1)	0	0	0.33	1	0.4	1	0.44
Pr(Toxicity=1,Efficacy=1)	0	0	0	0	0	0	0
Pr(Toxicity=0,Efficacy=0)	1	1	0.67	0	0.4	0	0.56
Pr(Toxicity=1,Efficacy=0)	0	0	0	0	0.2	0	0
Pr(Toxicity)	0	0	0	0	0.01	0.01	0.01
Pr(Efficacy)	0	0	0.33	1	0.4	1	0.44
Mean Utility	46.67	46.67	56	80	54.29	80	63.64

Future Direction

- Delayed Efficacy
 - The measure of biological activity is often quickly observable after drug administration and correlated with the clinical response. For example, abundance of CD8+ T cells can predict response to anti-PD-1 therapy
 - Immune activity (Y_I) can be used to predict the likelihood of achieving a clinical response.

$$\text{logit} \left(\frac{\pi_{Ej}}{\gamma} \right) = \beta_0 + \beta_1 Y_I,$$

where π_{Ej} is the probability of efficacy, β_0 and β_1 are regression parameters and $0 < \gamma \leq 1$ is a plateau parameter used to reflect the probability of clinical response which often levels out after the immune activity reaches a certain level.

- Incorporating Patient Heterogeneity
- Identifying the Optimal Basket indication
- Extending the concept beyond oncology

References

- Clark T, Mukherjee A, Lichtlen P, Sweeney J (2024) Bayesian Interval Based Designs for Phase I Dose-Escalation Trials: A Case Study in Oncology. *J Clin Trials.* 14:566
- Mengyi Lu, Jingyi Zhang, Ying Yuan & Ruitao Lin (26 Jun 2024): Comb-BOIN12: A Utility-Based Bayesian Optimal Interval Design for Dose Optimization in Cancer Drug-Combination Trials, *Statistics in Biopharmaceutical Research*, DOI: 10.1080/19466315.2024.2370403
- Lin X, Ji Y. The Joint i3+3 (Ji3+3) design for phase I/II adoptive cell therapy clinical trials. *J Biopharm Stat.* 2020 Nov 1;30(6):993-1005. doi: 10.1080/10543406.2020.1818250.
- Thall P.F., Cook J.D. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics.* 2004;60(3):684–693.
- U.S. Food and Drug Administration. Drug Development Tools: Fit-for-Purpose Initiative. (<https://www.fda.gov/drugs/development-approval-process-drugs/drug-development-tools-fit-purpose-initiative>)
- Lin, R., Zhou, Y., Yan, F., Li, D. & Yuan, Y. (2020), 'Boin12: Bayesian optimal interval phase I/II trial design for utility-based dose finding in immunotherapy and targeted therapies', *JCO Precision Oncology* 4, 1393–1402.

References

- Lin, X. Ji, Y. (2021), ‘Probability intervals of toxicity and efficacy design for dose-finding clinical trials in oncology’, *Statistical Methods in Medical Research* 30(3), 843–856.
- Luke, J., Johnson, M., Gadgeel, S., Spira, A., Yang, J., Johnson, J., Losch-Beridon, T., Snell, D., Warmuth, S., Cleaver, M., vom Baur, E., Urech, D., Lichtlen, P. Hong, D. (2022), ‘732 first-in- human trial to evaluate safety, pk/pd and initial clinical activity of nm21–1480, an affinity-balanced pd-l1x4–1bbxhsa trispecific antibody: results of phase 1 dose escalation’, *Journal for ImmunoTherapy of Cancer* 10(Suppl 2), A764–A764. URL: <https://jtc.bmjjournals.org/content/10/Suppl2/A764>
- O’Quigley, J. Conaway, M. (2010), ‘Continual reassessment and related Dose-Finding designs’, *Stat Sci* 25(2), 202–216.

References

- Zhou Y, Lee JJ, Yuan Y. A utility-based Bayesian optimal interval (U-BOIN) phase I/II design to identify the optimal biological dose for targeted and immune therapies. *Statistics in Medicine*. 2019;38:S5299–S5316. <https://doi.org/10.1002/sim.8361>.
- Ji Y, Practical Dose Finding and Dose Optimization Designs in Oncology Clinical Trials. 2023; Regulatory Industry Statistics Workshop Short Courses.
- Neuenschwander B, Branson M, Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. *Stat Med*. 2008 Jun 15;27(13):2420-39. doi: 10.1002/sim.3230.
- Liu, Jiaxin, et al. "The backfill i3+ 3 design for dose-finding trials in oncology." arXiv preprint arXiv:2303.15798 (2023).
- Dehbi HM, O'Quigley J, Iasonos A. Controlled backfill in oncology dose-finding trials. *Contemp Clin Trials*. 2021 Dec;111:106605. doi: 10.1016/j.cct.2021.106605. Epub 2021 Oct 30.

Lecture 4: (Response) Adaptive Randomization (RAR)

Sofía Villar

sofia.villar@mrc-bsu.cam.ac.uk

Programme Leader



MRC
Biostatistics
Unit



UNIVERSITY OF
CAMBRIDGE

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 randomisation).**

Preamble

Introductory concepts to RAR (and CARA) through a closer look at real world examples.

Focus on an example of RAR in its simplest form (but the statistical issues analogously apply to more complex settings, including CARA).

Which stage of development for RAR (as an adaptive feature)?

Where is RAR used? Mostly used in **Phase II** studies. If used for Phase III, there are different considerations.

Where does RAR sit on the phases of methodological research? **chronological versus biological age**.



Chronological
Age



Biological
Age

RESEARCH ARTICLE

Biometrical Journal →

Phases of methodological research in biostatistics—Building the evidence base for new methods

Georg Heinze¹ ⓘ | Anne-Laure Boulesteix² ⓘ | Michael Kammer^{1,3} |
Tim P. Morris⁴ ⓘ | Ian R. White⁴ | on behalf of the Simulation Panel of the STRATOS initiative

Preamble

Introductory concepts to RAR (and CARA) through a closer look at real world examples.

Focus on an example of RAR in its simplest form (but the statistical issues analogously apply to more complex settings, including CARA).

Which stage of development for RAR (as an adaptive feature)?

Where is RAR used? Mostly used in **Phase II** studies. If used for Phase III, there are different considerations.

Where does RAR sit on the phases of methodological research? **chronological versus biological age**.



Chronological
Age



Biological
Age

RESEARCH ARTICLE

Biometrical Journal →

Phases of methodological research in biostatistics—Building the evidence base for new methods

Georg Heinze¹ | Anne-Laure Boulesteix² | Michael Kammer^{1,3} |
Tim P. Morris⁴ | Ian R. White⁴ | on behalf of the Simulation Panel of the STRATOS initiative

Bird's-eye view

Let's divide and conquer [CA]RA[R]:

R Randomisation as a design element of a clinical trial.

Think of why and how? Before we adapt on it!

RAR **What** class of designs fall under the “*response-adaptive*’ (RA) randomisation label?
[Broad definition] **Why** to use them (or not)? A few examples.

CARA(R) **What** class of designs fall under the “*covariate-adjusted* (CA) response-adaptive randomisation label? [Broad definition] **Why** to use them (or not)?

STATISTICAL SCIENCE

Volume 38, Number 2 May 2023

Response-Adaptive Randomization in Clinical Trials: From Myths to Practical Considerations	<i>David S. Robertson, Kim May Lee, Boryana C. López-Kolkovska and Sofía S. Villar</i>	185
Comment: A Quarter Century of Methodological Research in Response-Adaptive Randomization	<i>Anastasia Ivanova and William F. Rosenberger</i>	209
Comment: Response-Adaptive Randomization in Clinical Trials: From Myths to Practical Considerations	<i>Yunshan Duan, Peter Müller and Yuan Ji</i>	212
Comment: Advancing Clinical Trials with Novel Designs and Implementations	<i>Lorenzo Trippa and Yanxun Xu</i>	216
Comment: Group Sequential Designs with Response-Adaptive Randomisation	<i>Christopher Jennison</i>	219
Comment: Is Response-Adaptive Randomization a “Good Thing” or Not in Clinical Trials? Why We Cannot Take Sides	<i>Alessandra Giovagnoli</i>	224
Comment: Response Adaptive Randomization in Practice	<i>Scott M. Berry and Kert Viele</i>	229
Rejoinder: Response-Adaptive Randomization in Clinical Trials	<i>David S. Robertson, Kim May Lee, Boryana C. López-Kolkovska and Sofía S. Villar</i>	233

Outline

The ARREST trial: a case study of a 2-arm Phase II BRAR design

Step 1: Choice of prior and type I error control

Step 2: Choice of randomisation method and undesirable imbalance

Step 3: Choice of measure, statistical test and statistical power

Multi-arm trials and CARA designs

Concluding remarks

The ARREST trial design (I)

This example will help us illustrate some of the less well known challenges of *designing, running and analysing* a RAR trial through a Bayesian RAR (BRAR) Phase II case study:

The Advanced REperfusion STrategies for Refractory Cardiac (ARREST) trial
(Yannopoulos 2020a; Yannopoulos 2020b; Yannopoulos 2020c).

Population: adult patients (18-75 years old) who are transferred by emergency with refractory ventricular fibrillation/pulseless ventricular tachycardia (VF/VT) cardiac arrest

Intervention and Comparator: 2 standards of care in use/current practice.

Extracorporeal membrane oxygenation facilitated resuscitation (**ECMO**)

(Treatment 1) (The same intervention as in (Bartlett 1985) the first trial(s) using a Randomised Play-The-Winner (RPTW) design).

Standard advanced cardiac life support resuscitation (**ACLS**) (**Treatment 0**).

Outcome: Survival to hospital discharge

Time period/follow up: at 30 days from treatment

The ARREST trial design (II)

Phase II, single center, open label, intention to treat, to study safety and efficacy clinical trial of the two interventions in use.

The **primary study hypotheses** are

$$H_0 : p_1 = p_0 \text{ vs } H_A : p_1 \neq p_0.$$

p_1 the **probability** of a **positive response** in the target population under ECMO.

p_0 the **probability** of a **positive response** in the target population under ACLS.

“The PIs wanted to use RAR to minimise patient exposure to the inferior treatment, as the consequences of providing an inferior treatment were grave.” (Proper 2021)

The ARREST Bayesian design: early stopping

Prior distributions for the unknown parameters of interest p_1 and p_0 : $Beta(1, 1)$

J interim analyses are **conducted after every 30 patients** have their primary outcome observed. Hence, the **posterior distribution** for p_0 (ACLS) and p_1 (ECMO) after data from interim j are: $Beta(s_{0,j} + 1, n_{0,j} - s_{0,j} + 1)$ and $Beta(s_{1,j} + 1, n_{1,j} - s_{1,j} + 1)$.

$n_{1,j}, n_{0,j}$ denotes the number of participants treated with ECMO (treatment 1) and ACLS (treatment 0) at interim j . $s_{1,j}, s_{0,j}$ denotes the number of positive responses observed in ECMO (treatment 1) and ACLS (treatment 0) at interim j .

[**Note:** the $Beta(1, 1)$ corresponds to $n_{1,0} = s_{1,0} = s_{0,0} = n_{0,0} = 1$ and there is a closed-form for the posterior distributions.]

Calculate $P(p_1 > p_0 | s_{0,j}, n_{0,j}, s_{1,j}, n_{1,j})$ to inform interim decisions.

E.g., early stopping: stop the trial for **efficacy** or **inferiority** when:

$$P(p_1 > p_0 | s_{0,j}, n_{0,j}, s_{1,j}, n_{1,j}) \geq \xi \text{ or } P(p_1 < p_0 | s_{0,j}, n_{0,j}, s_{1,j}, n_{1,j}) \geq \xi.$$

ξ is the level needed to reject the **null hypothesis**. **Q: How do we determine ξ ?**

The ARREST trial Bayesian design: Response-adaptation

Calculate at each j (by Monte carlo simulation) $P(p_1 > p_0 | s_{0,j}, n_{0,j}, s_{1,j}, n_{1,j})$

E.g., determine randomisation probability: favour the most promising arm so far.

A form of **Thompson sampling** (or BRAR) was used to randomize the next 30 patients to treatment 1 (ECMO) with probability equal to $P(p_1 > p_0 | s_{0,j}, n_{0,j}, s_{1,j}, n_{1,j})$ as long as this probability does not exceed 75% or goes below 25%.

If it falls outside those thresholds and early stopping has not been triggered then the randomisation probability is truncated to 75% or 25% respectively.

Note: Because of the initial uniform prior for both treatments, the first group of patients acts as a **run-in is size 30** with equal allocation probability for each arm.

The ARREST Trial - Type I Error and Power

The design used the critical value of $\xi^* = 0.986$ to achieve a **Type I error** rate of 5% two sided (accounting for the multiple interim analysis) using 10^4 simulations (for a null $p_0 = p_1 = 0.12$). This design was approved by FDA, DSMB statisticians, and NHLBI leadership.

The sample size ($N = 148$) was determined so that **power** of 90% was expected when using $\xi = 0.986$ assuming success rates of 12% vs. 37% in the 2 groups. Inflated to $N = 174$ for **15% expected drop out**. But only the **first 150 patients** are evaluated.

Scenario	Prob reject null	$E(N)$	$E(N_1)$	$E(N_0)$
Null	0.048	148.5	74.2	74.3
Alternative	0.905	81.6	52.5	29.2

Table 1: Simulated operating characteristics of the adaptive trial design as given in (Yannopoulos 2020a). $E(N)$ expected total sample size. $E(N_1)$ expected sample size for ECMO. $E(N_0)$ expected sample size for ACLS.

Outline

The ARREST trial: a case study of a 2-arm Phase II BRAR design

Step 1: Choice of prior and type I error control

Step 2: Choice of randomisation method and undesirable imbalance

Step 3: Choice of measure, statistical test and statistical power

Multi-arm trials and CARA designs

Concluding remarks

Does prior choice influence frequentist statistical properties?

The ARREST trial: simulations to determine ξ^* and sample size were based on a **Beta-Binomial (and uniform initial priors)** paired with a simple randomisation model (i.e. independent coin flip with probabilities that reflect the current target allocation).

In (Proper 2021) the authors noted that:

'sensitivity' in the type I error rate to the underlying response probabilities and prior specification → they suggested an alternative probability model and prior specification to temper type I error inflation.

Outcome Model and initial prior to feed in the BRAR design simulations:

$$Y|p_k \stackrel{\text{ind}}{\sim} \text{Binomial}(n_k, p_k) \text{ and } p_k \stackrel{\text{ind}}{\sim} \text{Beta}(p_k^* n_{k,0}, (1-p_k^*) n_{k,0}) \forall k, (p_k^* = 0.5, n_{k,0} = 2)$$

$$(\text{Alternative}) Y|p_k \stackrel{\text{ind}}{\sim} \text{Binomial}(n_k, p_k) \text{ and } \text{logit}(p_k) = \log\left(\frac{p_k}{1-p_k}\right) = \beta_0 + \beta_1(1_{k=1} - 0.5)$$

$$\beta_0 \sim t_7(\log \frac{0.12}{0.88}, 2.5), \beta_1 \sim t_7(0, 2.5), \text{ arises from solving for } \beta_0 \text{ when } p_z = 0.12, \beta_1 = 0$$

Does prior choice influence frequentist statistical properties? (II)

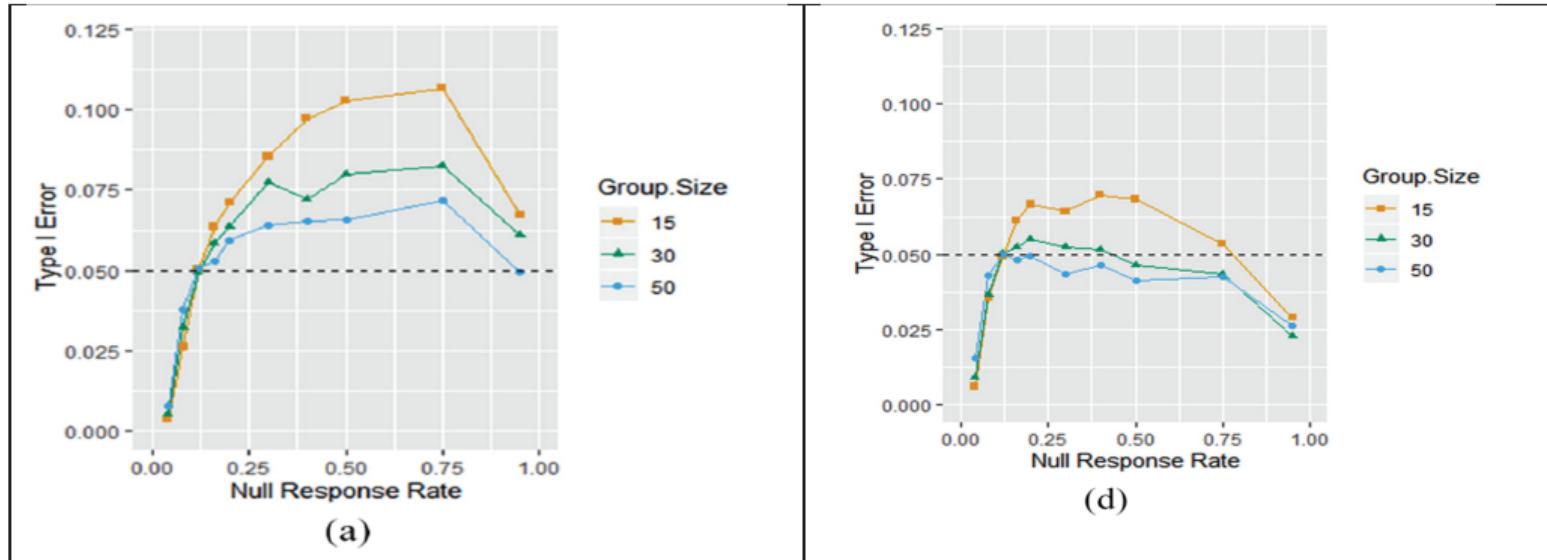


Figure 1: Type I error at various null response rates and sequential group sizes by probability model (independent coin flip (weighted) randomization method): (a) independent beta-binomial model with prior mean 0.50 and (d) logistic regression model with location = $\log(0.12/0.88)$

Does prior choice influence frequentist statistical properties? (III)

Alternatively, use an *exact test for RA* approach – work in progress at BSU (with Baas & Jacko)

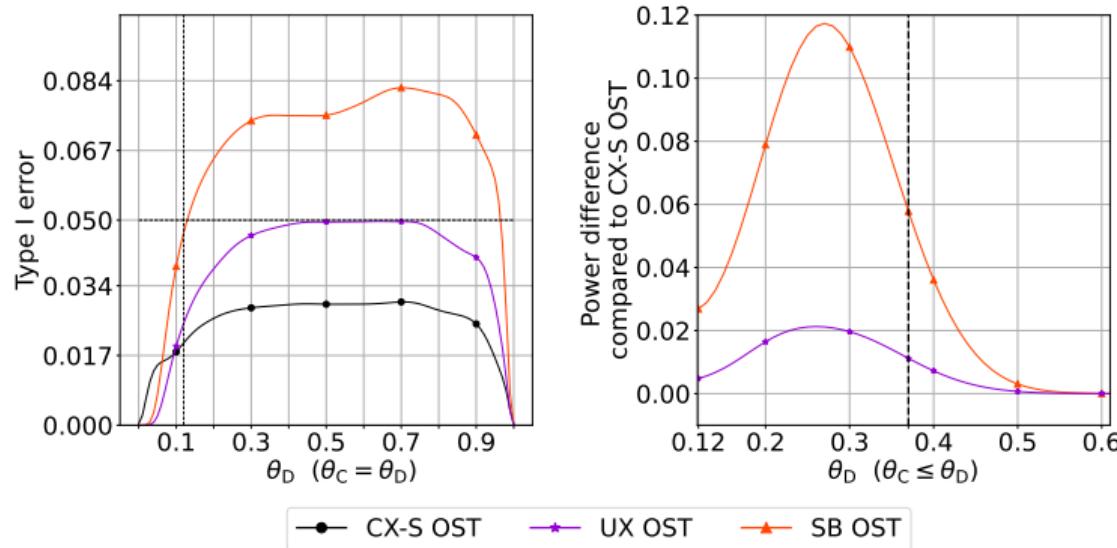


Figure 2: ARREST trial. Type I error under the SB (simulation based approach), CX-S OST (Conditional exact), and UX OST (Unconditional exact) for $p_0 = p_1$ (left), vertical line at $p_0 = p_1 = 0.12$. Power difference (right) for two tests compared to the CX-S OST, vertical line denotes $p_1 = 0.37$.

Does refining prior choice ensure better statistical properties?

Summary

Prior choice can impact type I error (and power) - through the selection of ξ^* value.

$\xi^* = 0.986$ controls the type I error rate at 5% for the case of $p_0 = p_1 = 0.12$

The choice of prior *may* control type I error over a larger subset of the parameter space but will remain **too conservative** (large power loss) at p_k values close to the boundaries of the parameter space

The alternative prior definition **loses the conjugacy** and requires a less straightforward computation for the $P(p_1 > p_0 | \mathbf{A}^{(j)}, Y^{(j)})$.

For type I error control an alternative solution is to use a conditional or unconditional exact test incorporating the BRAR design (Begg 1990). Valid outside of ARREST context. Retains conjugacy of the priors and closed-form for outcome posterior distribution. Paper and codes available upon request. QR code provided (final slides).

Outline

The ARREST trial: a case study of a 2-arm Phase II BRAR design

Step 1: Choice of prior and type I error control

Step 2: Choice of randomisation method and undesirable imbalance

Step 3: Choice of measure, statistical test and statistical power

Multi-arm trials and CARA designs

Concluding remarks

Does the randomisation method affect the statistical properties?

The ARREST trial uses a Beta-Binomial (uniform initial priors) paired with a **simple randomisation model** (i.e. independent coin flip with *adapted* probabilities)

In (Proper 2021) the authors noted that:

"high variability of allocation probabilities with small sample sizes, using the Thompson sampling which can reduce power and engender a considerable risk of allocating more subjects to the interior treatment" → $b = 15, 30, 50$ and $0.25 \leq P(a_{i,j}^k = 1) \leq 0.75$

Unacceptably large chance of imbalance in the "wrong direction" (e.g., as in Thall and Wathen $n_0 - n_1 > 0.1 n$ when $p_0 < p_1$) → suggest a controlled randomisation method may alleviate this.

One complication with RAR is that it may not be possible to achieve the target allocation within each group. One may want to avoid "unacceptable" allocations.

*E.g., for $b = 15$ and $P(a_{i,j}^k = 1) = 0.5$, $b * (P(a_{i,j}^k = 1)) = 7.5$ participants is not possible. Solution: allocate 7.5 participants to each arm **in expectation***

Does the randomisation method affect the statistical properties? (II)

Results from (Proper 2021) comparing the simple Weighted Coin (WC) flip to their Modified Permuted Block (MPB) design (with $b = 30$).

Design	Randomisation method	Power	$E(N)$	$E(N_1 - N_0)$	$P(N_1 < N_0)$
ER	WC	0.91	85.1	0.03	0.45
ER	MPB	0.92	81.8	0.00	<u>0.00</u>
BRAR	WC	0.87	88.6	26.6	<u>0.10</u>
BRAR	MPB	0.87	87.6	26.3	<u>0.00</u>

Table 2: Power and Imbalance achieved by two randomization methods using 1:1 or BRAR for the logistic regression model with prior intercept location = $\log(0.12/0.88)$. WC = Weighted Coin, MPB = Modified Permuted Block.

Does the randomisation method affect the statistical properties? (III)

Our current work for a BRAR implementation for STRATOSPHERE (Deliu 2023) comparing simple coin flip (WC) to a *mapped* randomisation method (Mapped).

Map from a probability to acceptable allocation sequences in the pre-specified blocks.

Design + Rand. Method	Frequentist properties		Empirical Allocation		
	$(1 - \beta)$	α	Arm C	Arm T_1	Arm T_2
BRAR WC	0.751	0.092	0.24 (0.12)	0.23 (0.12)	0.53 (0.17)
BRAR 'Mapped'	0.788	<i>0.086</i>	0.30 (0.00)	0.20 (0.11)	0.50 (0.11)

Table 3: Operating characteristics of the evaluated BRAR designs. Values are averaged across 10,000 independent replicas; results are reported in terms of mean (standard deviation).

Note: Potential impact on **type I error** and not just imbalance and power.

Does the randomisation method affect the statistical properties?

Summary

A simple weighted coin design (individual randomisation) is practically the only randomization method studied/reported/used for BRAR.

Although the simple weighted coin method maximizes randomness, its lack of imbalance control may result in arbitrarily large deviations from the target allocation (or “undesirable allocations”) as well as unacceptably large variability in allocations.

An important aspect of any randomization method is the trade-off between **allocation randomness and treatment balance**.

The choice of the randomisation method is not always explicitly detailed in simulation studies and protocols. **Its impact on the operating characteristics remains unclear/understudied.** This is more important for designs with RAR.

Outline

The ARREST trial: a case study of a 2-arm Phase II BRAR design

Step 1: Choice of prior and type I error control

Step 2: Choice of randomisation method and undesirable imbalance

Step 3: Choice of measure, statistical test and statistical power

Multi-arm trials and CARA designs

Concluding remarks

The ARREST Trial - statistical test and power considerations

Does the treatment effect of interest (or measure of interest) influence the impact of RAR in terms of power? If so, how much can this impact power?

Illustrated through a design with BRAR as the sole adaptation (fixed $n = 150$).

Shiny app results: <http://shiny.mrc-bsu.cam.ac.uk/RAR/> (thanks to Lukas Pin).

Design	Scenario	Measure	Prob reject null (Using Wald Test)	$E(N_1)$	$E(N_0)$
ER vs. BRAR	$p_1 = 0.37, p_0 = 0.12$	$p_1 - p_0$	0.9576 vs. 0.7975	75 vs. 132	75 vs. 18
ER vs. BRAR	$p_1 = 0.37, p_0 = 0.12$	$\frac{(1-p_1)}{(1-p_0)}$	0.9640 vs. 0.8721	75 vs. 132	75 vs. 18
ER vs. BRAR	$p_1 = 0.95, p_0 = 0.8$	$p_1 - p_0$	0.8341 vs. 0.3371	75 vs. 120	75 vs. 30
ER vs. BRAR	$p_1 = 0.95, p_0 = 0.8$	$\frac{(1-p_1)}{(1-p_0)}$	0.9343 vs. 0.8188	75 vs. 120	75 vs. 30

Table 4: Simulated operating characteristics of the different trial designs, scenarios and measure.

Outline

The ARREST trial: a case study of a 2-arm Phase II BRAR design

Step 1: Choice of prior and type I error control

Step 2: Choice of randomisation method and undesirable imbalance

Step 3: Choice of measure, statistical test and statistical power

Multi-arm trials and CARA designs

Concluding remarks

What about more complex settings?

Most issues discussed so far remain valid in other (more complex settings) and need attention.

The RAR toolkit allows for dealing with additional aspects of design (particulalry within patient outcome) and potential achieve additional advantages (efficiencies).

Multi-arm settings offer particular efficiencies (for scenarios where there is a clear winner or only 1 arm can be carried forward).

CARA can offer ways to deal with covariate-treatment interactions in an patient-oriented and efficient manner (personalisation of treatment)

The EndTb trial

Multi-arm and non inferiority setting

Randomized, controlled, open-label, non-inferiority, **Phase III** trial. Uses a type of BRAR algorithm to randomise patients to treatments.

Run-in: 180 patients. Then, randomization probabilities adapted according to treatment response on interim endpoints

Compares each of 5 experimental regimens to control: Efficacy and Safety.

Detect as many non-inferior (NI) regimens as possible

NI established at 73 weeks in modified Intention to Treat (mITT) and Per-Protocol (PP) populations but adaptation is done using an early response (8 and 39 weeks). **Note** The posterior is defined to reflect a NI margin (rather than the usual superiority one).

Note The BRAR design relies on assumption of positive correlation between early (8w) and late endpoints (39w). The weaker the correlation the less adaptation (and efficiency gain). This requires assessment by simulations to quantify effects of delay/lack of correlation.

The I-Spy 2 trial

Multi-arm and CARA setting

I-SPY 2: Platform Trial in Neoadjuvant Breast Cancer.

Randomized, controlled, open-label, non-inferiority, **Phase II** trial. Uses a BRAR algorithm to assign patients to treatments.

Probabilities depended on patient profile. 3 biomakers: Her+/-, MP+/-, HR+/- . So each groups of patients, each randomised to different options with an different process.

The probability to assign any of the patients to the standard of care is kept fixed at 20%. The remaining 80% is assigned among experimental arms per type and given outcome data so far.

Paired with early stopping rules (efficacy/futility) per patient subgroup.

Able to detect interactions.

Principle used in many other platforms.

Outline

The ARREST trial: a case study of a 2-arm Phase II BRAR design

Step 1: Choice of prior and type I error control

Step 2: Choice of randomisation method and undesirable imbalance

Step 3: Choice of measure, statistical test and statistical power

Multi-arm trials and CARA designs

Concluding remarks

RAR and CARA and the CID programme of FDA

<https://www.fda.gov/media/155404/download?attachment> (FDA website link)

CID Case Study: A Study in Patients with Systemic Lupus Erythematosus

Study Design:

The proposed study is a randomized, double-blind, Phase 2 study in patients with systemic lupus erythematosus (SLE), a rare disease with a high unmet need. Patients are to be randomized to one of four treatment groups: three doses of investigational product (IP) or placebo. The primary endpoint of the study is Systemic Lupus Erythematosus Responder Index 4 (SRI-4) response at 52 weeks, a dichotomous outcome where response indicates success. This composite endpoint incorporates a hybrid Safety of Estrogens in Systemic Lupus Erythematosus National Assessment Systemic Lupus Erythematosus Disease Activity Index (SELENA-SLEDAI) score, British Isles Lupus Assessment Group index (BILAG) 2004 domain scores, and the Physician's Global Assessment (PGA). This endpoint will be evaluated using a Bayesian Hierarchical Model (BHM) with non-informative priors. Interim analyses will occur at 8 prespecified time points. The initial randomization ratio will be 1:1:1:1.

At each interim analysis except for the last one, a response adaptive randomization (RAR) procedure will take place where the randomization allocation probabilities for each of the three IP arms may be modified moving forward. The randomization allocation probability for the placebo arm will remain fixed at 25% throughout the study.

At each interim analysis except for the first one, a futility determination will be made, based on the performance of the three IP arms compared to placebo.

Case study

RAR and CARA and the CID programme of FDA

<https://www.fda.gov/media/155404/download?attachment> (FDA website link)

Considerations for the Proposed Design:

- What impact does RAR have on the comparability of treatment groups with respect to baseline measurements and characteristics?
- Does this study design have adequate operating characteristics (e.g., power, type I error rate, reliability of point estimates, probability of selecting the best dose, etc.) across different true dose-response relationships and across the multidimensional range of plausible values for the nuisance parameters?
- How does the use of RAR perform against arm-dropping approaches?
- For the primary analysis, how does the BHM perform compared to competitive conventional methods?
- What firewalls and other procedures will be put in place to ensure that interim analysis results will remain confidential and study integrity will be maintained?

Simulations:

The Sponsor conducted simulations to assess the operating characteristics of the proposed model under different true dose-response relationships and under a multidimensional range of plausible values for the nuisance parameters. Nuisance parameters included the true underlying placebo response rate, the patient enrollment rate, and the within-patient correlation of response status at adjacent visits.

Simulations evaluated important operating characteristics such as type I error rate, power, reliability of point estimates, futility determination probabilities, and probabilities of selecting the most effective dose. The set of combinations of values for the nuisance parameters (under which simulations were to be performed) was expanded several times as a result of iterative feedback.

Questions

Final thoughts

RAR/CARA as a trial adaptation is very versatile (broad): it can go from early phases to later phases. It can also be used in combination with other adaptations (early stopping).

Its appropriate use requires carefully tailoring it to (best) meet the trial objectives.

Despite its "chronological" age, it is "biological" age is of a young method.

Factors that should be carefully considered when evaluating RAR/CARA as a design component (but often ignored) include: the RAR type, the trial phase (and main goal), the treatment effect definition (or measure of interest), the relevant parameter values for a study and the statistical test to be used for final analysis and the possibilities of temporal trends.

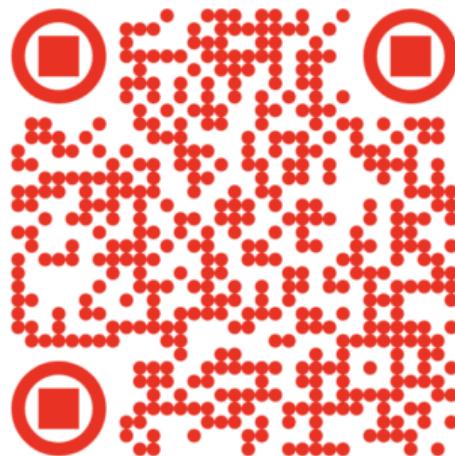
Delivering a trial with a RAR/CARA element requires carefully thinking of how to choose a randomisation method, how this can impact the statistical performance and (very importantly) how this would be logistically delivered (possibly by an external providers).

RAR Short Course

23–24 October 2024

Cambridge, UK

<https://www.mrc-bsu.cam.ac.uk/short-courses>



Acknowledgments

With special thanks to:



Stef Baas



Rajenki Das



Lukas Pin



...among others working in RAR methods and practice with me now

References

- [Yannopoulos 2020a] Yannopoulos, D and others (2020) Advanced REperfusion STrategies for Refractory Cardiac Arrest (The ARREST Trial) Protocol version 1.4 (2020a)
- [Yannopoulos 2020b] Yannopoulos, D and others (2020b) Rationale and methods of the Advanced R2Eperfusion STrategies for refractory cardiac arrest (ARREST) trial. *American Heart Journal*. 229 29–39.
- [Yannopoulos 2020c] Yannopoulos, D and others (2020c) Advanced reperfusion strategies for patients with out-of-hospital cardiac arrest and refractory ventricular fibrillation (ARREST): a phase 2, single centre, open-label, randomised controlled trial. *The lancet*. 396 1807–1816.
- [Bartlett 1985] Bartlett, R and others (1985) Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study. *Pediatrics Journal*. 76(4) 479–487.
- [Proper 2021] Proper and others (2021) Alternative models and randomization techniques for Bayesian response-adaptive randomization with binary outcomes. *Clinical Trials*, 18(4) 417–426.
- [Begg 1990] Begg, C. B. (1990), On inferences from Wei's biased coin design for clinical trials *Biometrika*, 77, 467–473
- [Deliu 2023] Deliu and others (2023) StratosPHere 2: A response-adaptive randomised placebo-controlled Phase II trial to evaluate hydroxychloroquine and phenylbutyrate in pulmonary arterial hypertension caused by mutations in BMPR2 Under review for *Trials*

Adaptive and Complex Innovative Designs across trial phases for accelerated approval

Ayon Mukherjee, David Robertson, Sofia S. Villar, Thomas Burnett



UNIVERSITY OF
CAMBRIDGE



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)

Q&A

Any questions?