

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.

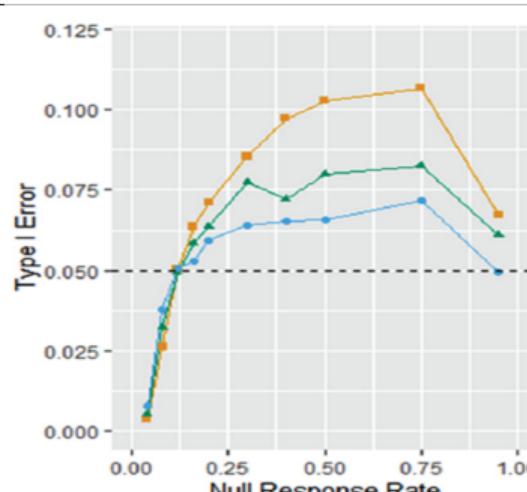
Outcome Model and initial prior to feed in the BRAR design simulations:

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

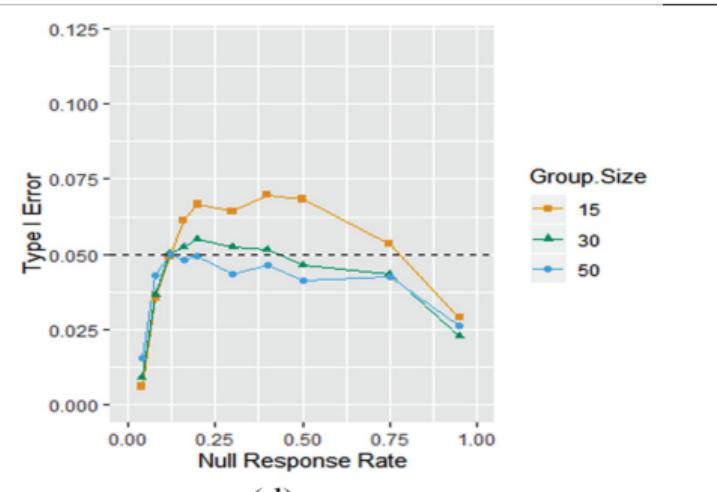
$$(\text{Alternative}) Y|p_k \stackrel{\text{ind}}{\sim} \text{Binomial}(n_k, p_k) \text{ and } \text{logit}(p_k) = \log\left(\frac{p_k}{1-p_k}\right) = \beta_0 + \beta_1(1_{k=1} - 0.5)$$

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

Does prior choice influence frequentist statistical properties? (II)



(a)



(d)

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)

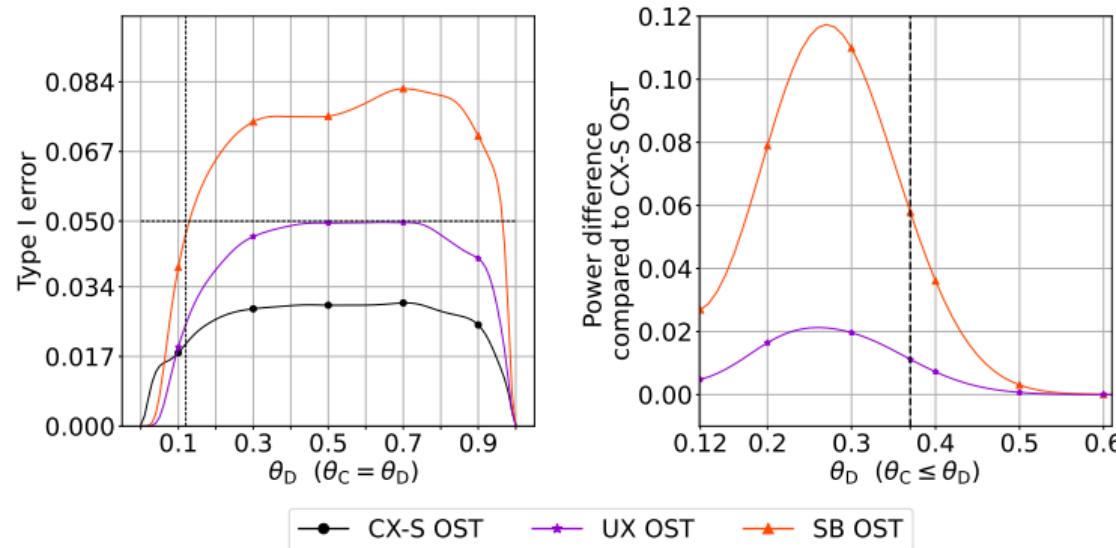


Figure 2: ARREST trial. Type I error under the SB (simulation based approach), CX-S OST (Conditional exact), and UX OST (Unconditional exact) for $p_0 = p_1$ (left), vertical line at $p_0 = p_1 = 0.12$. Power difference (right) for two tests compared to the CX-S OST, vertical line denotes $p_1 = 0.37$.

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*