Enabling Optimal Response Adaptive Designs

New Insights, Algorithms and Tests

Lukas Pin

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Randomization Working Group

August 26th, Cambridge, UK

Outline

Statistics > Methodology

[Submitted on 17 Jul 2025]

Is 1:1 Always Most Powerful? Why Unequal Allocation Merits Broader Consideration

Lukas Pin, Stef Baas, David S. Robertson, Sofía S. Villar

Is 1:1 Always Most Powerful?

- **Common belief:** 1:1 allocation maximises power and efficiency
- **Widely echoed in the literature, e.g.:**

  "balanced group sizes will maximise a study's statistical power" (Dumville et al., 2006).

  "Trials using unequal allocation will therefore either have less statistical power or will be more expensive and entail exposing more patients than necessary to a novel intervention and research procedures." (Hey and Kimmelman, 2014).

- **But:** This assumption can be misleading — unequal allocation isn't always worse

Some Common Notation

- Two treatment arms, 0 (control) and 1 (experimental).
- A study with fixed number of patients n
- Potential outcome $Y_{ki} \overset{\text{iid}}{\sim} \text{Bern}(p_k), \quad i = 1, \ldots, n_k; \quad k = 0, 1; \quad n_0 + n_1 = n$
  - p_k is success probability on arm k and $q_k = 1 - p_k$ failure probability

Note, outcome could be non-binary and we could include a random n (early stoppoing).

- Let $A_i \in \{0, 1\}$ denote the allocated treatment for patient i
- Observed outcome $Y_i = Y_{i,a_i}$ (assumes consistency/SUTVA)

Which Allocation Maximises Power?

- **Power-maximising allocation depends on:**
  - endpoint type (binary, continuous, etc.),
  - target effect measure (e.g., mean difference, odds ratio),
  - and—most importantly—**variance within each group**.

- For the Wald test comparing means:

$$Z = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{\frac{\hat{p}_0 \hat{q}_0}{n_0} + \frac{\hat{p}_1 \hat{q}_1}{n_1}}}$$

- **Neyman (1934) allocation** maximises statistical power by allocating proportionally to standard deviations

$$\frac{n_1}{n_0} = \frac{\sigma_1}{\sigma_0}$$

Binary Endpoint: Theoretical Difference between ER and Neyman

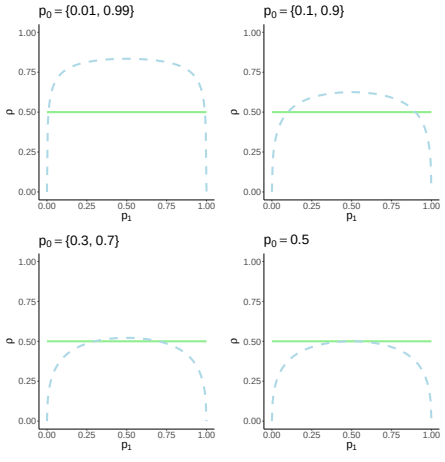


Figure 1: The horizontal axis shows the varying value of p_1 while p_0 takes a fixed value in each sub figure. For each combination of p_1 and p_0 the vertical axis shows the value of the theoretical Neyman allocation (ρ_{N_1}) and Equal Randomisation (ρ_{ER}).

Binary Trial Example

Setup: 60-patient early-phase cancer trial comparing with **binary** outcome (e.g., RECIST-based), using a Wald test at two-sided $\alpha = 0.05$ level.
Fixed Unequal Randomisation (FUR) uses a 1:2 allocation (Control:Treatment), approximating Neyman allocation ($\sigma_1/\sigma_0 \approx 0.678$).

Table: Allocation strategy comparison. Metrics: type-I error (under $p_0 = p_1 = 0.05$), power (under $p_0 = 0.05$, $p_1 = 0.3$), proportion on superior arm (n_1/n), and Expected Number of Successes (ENS). Monte Carlo error $< 0.2\%$ (Pin et al., 2025).

Design	Type-I Error	Power	n_1/n	ENS
ER	3.0%	80.0%	0.50	10.5
FUR (1:2)	5.2%	81.5%	0.67	13

- FUR **increases power by 1.5** percentage points compared to equal randomization (ER).
- Additionally, more patients allocated to the superior arm, **improving patient-benefit**.

Summary and Discussion I

- Non-binary example with larger power gain in manuscript.
- Neyman allocation is optimal in theory but power gains depend on many factors:
  1. the endpoint,
  2. the total sample size,
  3. the magnitude of the treatment effect difference,
  4. the relationship between the variances of the treatment arms.
- FUR or RAR can modestly boost power and allocate more patients to better treatments.
- Allocation decisions should weigh statistical, patient-benefit, operational, and economic factors.

1. Justify your allocation choice

Using 1:1 allocation is common, but there should be a clear reason for it — not just habit.

2. Power is not a reason for 1:1

Equal allocation does **not** maximize power in many cases, especially when arms differ.

We urge thoughtful justification of all allocation ratios, including 1:1.

Outline

What is a Response-adaptive Design?

Definition: A Response-adaptive design is an experimental design that allows for the allocation of experimental units to the different (treatment) options to dynamically change based on available *data* with the goal to optimise the experiment (*in some way*).

- Response-adaptive Randomisation (RAR) is a subclass of the above which:

  (1) is implemented through randomisation (in a probabilistic manner) and;

  (2) and where the resulting allocation probabilities to change (or adapt) based on the evidence/data collected at interim points.

- RAR is perhaps the oldest form of an adaptive design of experiments. Proposed by Thompson (1933) for *saving individuals otherwise sacrificed to an inferior treatment*

> probability of treatment by the two methods of $f_{(P)}$ and $1 - f_{(P)}$, respectively. If such a discipline were adopted, even though it were not the best possible, it seems apparent that a considerable saving of individuals otherwise sacrificed to the inferior treatment might be effected. This would be important in cases where either the rate of accumulation of data is slow or the individuals treated are valuable, or both.

Why use RAR (or not)?

RAR may be used for attaining a given experimental objective or a combination of them.

To achieve a level of statistical power with a given test at the end of the study

To assign more patients to a superior treatment when evidence suggests superiority within the trial (rare disease setting)

To select a set of promising arms (within a large set) to carry forward to a confirmatory stage (early phase setting)

...

A broad(er) look at RAR: 2023 RAR review, discussion and rejoinder

STATISTICAL SCIENCE

Volume 38, Number 2 May 2023

Some Common Notation

- Two treatment arms, 0 (control) and 1 (experimental).
- A study with fixed number of patients n
- Potential outcome $Y_{ki} \overset{\text{iid}}{\sim} \text{Bern}(p_k), \quad i = 1, \ldots, n_k; \quad k = 0, 1; \quad n_0 + n_1 = n$

  Note, outcome could be non-binary and we could include a random n (early stoppoing).
- Let $A_i \in \{0, 1\}$ denote the allocated treatment for patient i
- Observed outcome $Y_i = Y_{i,a_i}$ (assumes consistency/SUTVA)
- $P(A_i = k)$ is the probability of patient i receiving the arm k.
- Equal (simple) randomization is such that $P(A_i = k) = 1/2 \ \forall i, k$.
- Assumption: Patients arrive sequentially and outcomes are immediately observable,
- RAR is defined as function that maps past outcomes and allocations to a value in $[0, 1]$:

$$\pi_i = P(A_i = k | \overline{Y_{i-1}}, \overline{a_{i-1}})$$

Neyman and RSHIR allocation proportions

- General approach: If n fixed and p_0, p_1 known. Derive optimal sampling proportions by optimizing a (linear) objective function subject to a (non linear) constraint.
- We want to test $H_0 : p_1 - p_0 = 0$ vs $H_1 : p_1 - p_0 > 0$
  with Wald test: $Z = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{s^2_{\Delta\hat{p}}(n_0, n_1)}}$ where $s^2_{\Delta\hat{p}}(n) = \frac{\hat{p}_0\hat{q}_0}{n_0} + \frac{\hat{p}_1\hat{q}_1}{n_1}$ and $\hat{q}_k := 1 - \hat{p}_k$.

Q Neyman (1934): What is **minimal total sample size** to achieve given a fixed power level? Let $\rho = \frac{n_1}{n_0+n_1}$ (and $1-\rho = \frac{n_0}{n_0+n_1}$),

$\min_\rho n_0 + n_1$ s.t. $s^2_{\Delta\hat{p}}(\rho) = C$ **Solution** Neyman allocation: $\rho_{N_1} = \frac{\sqrt{p_1 q_1}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}} = \frac{\sigma_1}{\sigma_0+\sigma_1}$

Q Rosenberger et al. (2001): What is **the minimum expected number of failures** given a power constraint (or a fixed variance level) ?

$\min_\rho q_0\, n_0 + q_1\, n_1$ s.t. $s^2_{\Delta\hat{p}}(\rho) = C$ **Solution** R[SHIR] allocation: $\rho_{R_1} = \frac{\sqrt{p_1}}{\sqrt{p_0}+\sqrt{p_1}}$

Neyman and RSHIR allocation proportions

- General approach: If n fixed and p_0, p_1 known. Derive optimal sampling proportions by optimizing a (linear) objective function subject to a (non linear) constraint.
- We want to test $H_0 : p_1 - p_0 = 0$ vs $H_1 : p_1 - p_0 > 0$
  with Wald test: $Z = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{s^2_{\Delta\hat{p}}(n_0, n_1)}}$ where $s^2_{\Delta\hat{p}}(n) = \frac{\hat{p}_0\hat{q}_0}{n_0} + \frac{\hat{p}_1\hat{q}_1}{n_1}$ and $\hat{q}_k := 1 - \hat{p}_k$.

Q Neyman (1934): What is **minimal total sample size** to achieve given a fixed power level? Let $\rho = \frac{n_1}{n_0+n_1}$ (and $1 - \rho = \frac{n_0}{n_0+n_1}$),

$\min_\rho n_0 + n_1$ s.t. $s^2_{\Delta\hat{p}}(\rho) = C$ **Solution** Neyman allocation: $\rho_{N_1} = \frac{\sqrt{p_1 q_1}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}} = \frac{\sigma_1}{\sigma_0+\sigma_1}$

Q Rosenberger et al. (2001): What is **the minimum expected number of failures** given a power constraint (or a fixed variance level) ?

$\min_\rho q_0\, n_0 + q_1\, n_1$ s.t. $s^2_{\Delta\hat{p}}(\rho) = C$ **Solution** R[SHIR] allocation: $\rho_{R_1} = \frac{\sqrt{p_1}}{\sqrt{p_0}+\sqrt{p_1}}$

Neyman and RSHIR allocation proportions

- General approach: If n fixed and p_0, p_1 known. Derive optimal sampling proportions by optimizing a (linear) objective function subject to a (non linear) constraint.
- We want to test $H_0 : p_1 - p_0 = 0$ vs $H_1 : p_1 - p_0 > 0$
  with Wald test: $Z = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{s^2_{\Delta\hat{p}}(n_0, n_1)}}$ where $s^2_{\Delta\hat{p}}(n) = \frac{\hat{p}_0\hat{q}_0}{n_0} + \frac{\hat{p}_1\hat{q}_1}{n_1}$ and $\hat{q}_k := 1 - \hat{p}_k$.

Q Neyman (1934): What is **minimal total sample size** to achieve given a fixed power level? Let $\rho = \frac{n_1}{n_0+n_1}$ (and $1 - \rho = \frac{n_0}{n_0+n_1}$),

$\min_\rho n_0 + n_1$ s.t. $s^2_{\Delta\hat{p}}(\rho) = C$ **Solution** Neyman allocation: $\rho_{N_1} = \frac{\sqrt{p_1 q_1}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}} = \frac{\sigma_1}{\sigma_0+\sigma_1}$

Q Rosenberger et al. (2001): What is **the minimum expected number of failures** given a power constraint (or a fixed variance level) ?

$\min_\rho q_0\, n_0 + q_1\, n_1$ s.t. $s^2_{\Delta\hat{p}}(\rho) = C$ **Solution** R[SHIR] allocation: $\rho_{R_1} = \frac{\sqrt{p_1}}{\sqrt{p_0}+\sqrt{p_1}}$

Theoretical Comparison of Allocation Proportions

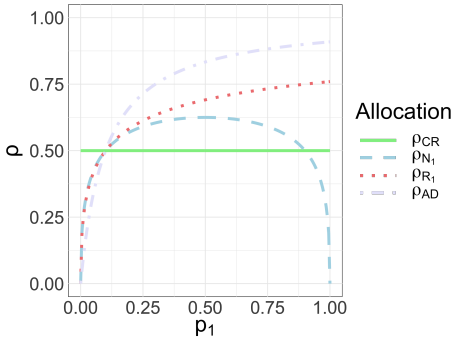


Figure 2: $p_0 = 0.1$

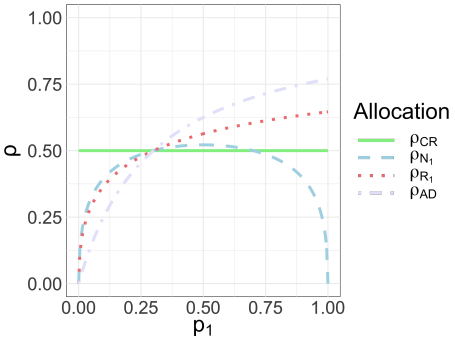


Figure 3: $p_0 = 0.3$

Estimation and Targeting

1. Estimate proportions using MLE
2. Target them using Efficient Randomized-Adaptive Design (ERADE) by Hu et al. (2009). For a parameter $\alpha \in (0,1)$, we sample patient $j+1$ towards treatment 1 with probability

$$p_1(n_1(j), \rho(j)) = \begin{cases} \alpha\rho(j), & \text{if } n_1(j)/j > \rho(j), \\ \rho(j), & \text{if } n_1(j)/j = \rho(j), \\ 1 - \alpha(1-\rho(j)), & \text{if } n_1(j)/j < \rho(j). \end{cases}$$

and then Example:

$\alpha = 0.5$, $n_0(9) = 4$ and $n_1(9) = 5$ $p_0(9) = 25\%$ and $p_1(9) = 60\%$

$$\rho_{R_1}(9) = \frac{\sqrt{0.6}}{\sqrt{0.25} + \sqrt{0.6}} = 0.608 \text{ but } n_1(j)/j = 5/9 = 0.55$$

$$p_1(n(9), \hat{\rho}(9)) = 1 - 0.5 \cdot (1 - 0.608) = 0.804$$

Outline

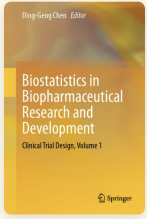
The Impact of Measures of Interest

- In the previous section: derived optimal allocation proportions for the **simple mean difference** using the Wald test.
- But: if we change the **measure of interest** (e.g., to log odds, relative risk, odds ratio, etc.), then
  - the **test statistic** changes,
  - and with it, the **optimization problem** changes as well.
- $\Rightarrow$ Different measures of interest can lead to different optimal allocations.

Response-Adaptive Randomization Designs Based on Optimal Allocation Proportions

Chapter | Open Access | First Online: 14 August 2024

pp 313–339 | Cite this chapter



Biostatistics in Biopharmaceutical Research and Development

Lukas Pin, Sofia S. Villar & William F. Rosenberger

Sections | Figures | References

Other Measures of Interest

Table 1: Overview of measures of interest with respective allocation proportions that minimize failures p_{minF} or maximize power p_N ().

	Simple difference	Relative Risk	Odds ratio	Log relative risk	Log odds ratio
θ	$p_0 - p_1$	q_1/q_0	$\frac{p_0}{q_0} / \frac{p_1}{q_1}$	$\log(q_1/q_0)$	$\log\left(\frac{p_0}{q_0} / \frac{p_1}{q_1}\right)$
p^*_{minF}	$\frac{\sqrt{p_1}}{\sqrt{p_0}+\sqrt{p_1}}$	$\frac{\sqrt{p_1 q_0}}{\sqrt{p_0 q_1}+\sqrt{p_1 q_0}}$	$\frac{\sqrt{p_0 q_0}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}}$	$\frac{\sqrt{p_1 q_0}}{\sqrt{p_0 q_1}+\sqrt{p_1 q_0}}$	$\frac{\sqrt{p_0 q_0}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}}$
p^*_N	$\frac{\sqrt{p_1 q_1}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}}$	$\frac{\sqrt{p_1 q_0}}{\sqrt{p_0 q_1}+\sqrt{p_1 q_0}}$	$\frac{\sqrt{p_0 q_0}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}}$	$\frac{\sqrt{p_1 q_0}}{\sqrt{p_0 q_1}+\sqrt{p_1 q_0}}$	$\frac{\sqrt{p_0 q_0}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}}$

Other Measures of Interest - Neyman

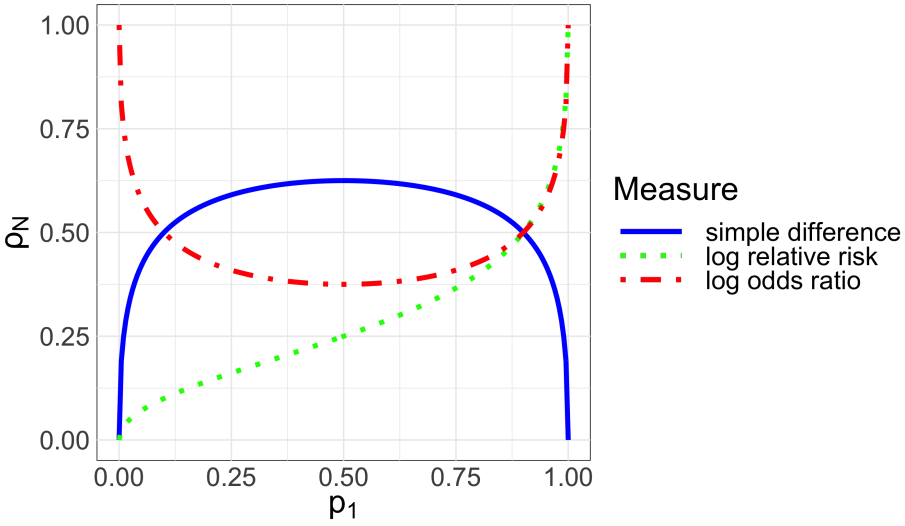


Figure 4: $p_0 = 0.9$

Other Measures of Interest - minF

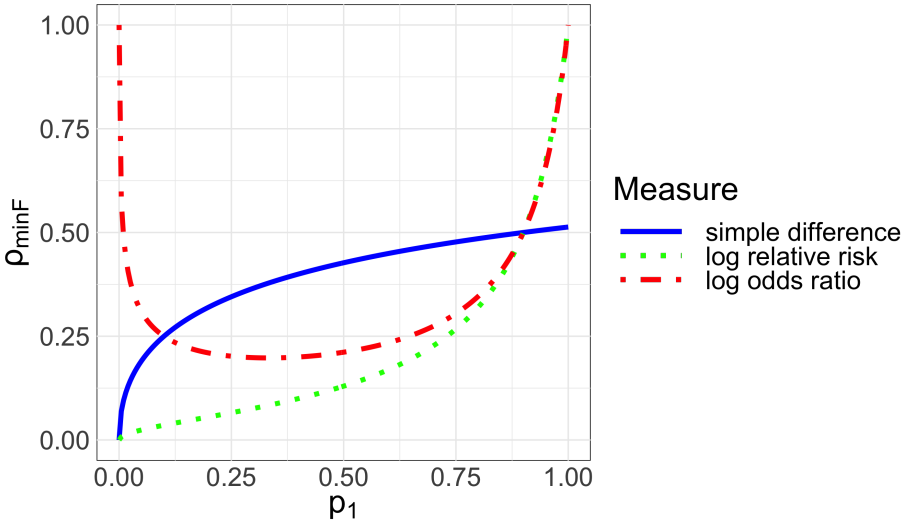


Figure 5: $p_0 = 0.9$

Other measure of interest - Conclusion

1. ER still only optimal when $p_0 = p_1$
2. Areas where efficacy and patient benefit conflict change
3. Areas where specific optimal RAR could be useful depends on measure of interest and parameter region of interest.
4. For relative risk patient-benefit and power align.

Outline

Motivation

- **A challenge connected to optimal RAR?**

  Type-I error rate control in literature on optimal RAR widely ignored

- **Why is it important?**

  Type-I error rate control requirement for (confirmatory) design **approval** by FDA or EMA

- **Solution:**

  Redefining optimization problem: different test statistic & estimators instead of unknown true values

Biometrics, 2025, 00(0), ujaf114
https://doi.org/10.1093/biomtc/ujaf114
Biometric Practice

OXFORD

Revisiting optimal allocations for binary responses: insights from considering Type-I error rate control

Lukas Pin [1,*], Sofía S. Villar [1], William F. Rosenberger [2]

[1]MRC Biostatistics Unit, University of Cambridge, Robinson Way, Cambridge, CB2 0SR, United Kingdom, [2]Department of Statistics, George Mason University, Fairfax, VA 22030, United States

CR, Neyman and RSHIR for Wald test

Testing with Z_1										
p_0	p_1	Type-I Error Rate or Power			n_1/n			ENS		
		ρ_{CR}	ρ_{N_1}	ρ_{R_1}	ρ_{CR}	ρ_{N_1}	ρ_{R_1}	ρ_{CR}	ρ_{N_1}	ρ_{R_1}
0.1	**0.1**	**5.0%**	**68.2%**	**68.1%**	0.5 (0)	0.47 (0.1477)	0.46 (0.1471)	5	5	5
0.2	**0.2**	**5.9%**	**82.2%**	**80.0%**	0.5 (0)	0.46 (0.1570)	0.48 (0.1525)	10	10	10
0.3	**0.3**	**6.3%**	**72.0%**	**66.8%**	0.5 (0)	0.47 (0.1432)	0.48 (0.1336)	15	15	15
0.4	**0.4**	**6.2%**	**64.7%**	**53.0%**	0.5 (0)	0.47 (0.132)	0.48 (0.1063)	20	20	20
0.5	**0.5**	**6.4%**	**61.9%**	**38.6%**	0.5 (0)	0.47 (0.1261)	0.49 (0.0757)	25	25	25
0.6	**0.6**	**6.0%**	**65.0%**	**26.6%**	0.5 (0)	0.47 (0.1326)	0.49 (0.0490)	30	30	30
0.7	**0.7**	**6.1%**	**71.9%**	**17.8%**	0.5 (0)	0.47 (0.1438)	0.49 (0.0295)	35	35	35
0.8	**0.8**	**6.1%**	**82.1%**	**10.6%**	0.5 (0)	0.47 (0.1567)	0.49 (0.0127)	40	40	40
0.9	**0.9**	**4.8%**	**68.3%**	**5.1%**	0.5 (0)	0.47 (0.1477)	0.49 (0.0033)	45	45	45
0.2	0.1	17.5%	84%	82.9%	0.5 (0)	0.33 (0.1308)	0.32 (0.1308)	7.5	8.3	8.4
0.2	0.3	15.3%	80%	76.7%	0.5 (0)	0.55 (0.1513)	0.57 (0.1420)	12.5	12.7	12.9
0.2	0.5	65.4%	88.3%	84.9%	0.5 (0)	0.62 (0.1342)	0.71 (0.0907)	17.5	19.3	20.7
0.2	0.7	97.1%	98.3%	97.8%	0.5 (0)	0.56 (0.1486)	0.79 (0.0476)	22.5	23.9	29.8
0.7	0.2	97.2%	98.5%	97.9%	0.5 (0)	0.38 (0.1408)	0.19 (0.0408)	22.5	25.5	30.2
0.7	0.4	62.1%	85.9%	72.0%	0.5 (0)	0.52 (0.1381)	0.33 (0.0507)	27.5	27.2	30.0
0.7	0.6	13.9%	70.4%	26.6%	0.5 (0)	0.51 (0.1394)	0.45 (0.0372)	32.5	32.5	32.8
0.7	0.8	15.0%	78.9%	22.0%	0.5 (0)	0.38 (0.1396)	0.52 (0.0213)	37.5	36.9	37.6

Table 1: Power or Type-I Error Rate, proportion allocated to the treatment arm n_1/n, and expected number of successes (ENS) for different settings of p_0 and p_1.

CR, Neyman and RSHIR for Wald test

Testing with Z_1										
p_0	p_1	Type-I Error Rate or Power			n_1/n			ENS		
		ρ_{CR}	ρ_{N_1}	ρ_{R_1}	ρ_{CR}	ρ_{N_1}	ρ_{R_1}	ρ_{CR}	ρ_{N_1}	ρ_{R_1}
0.1	**0.1**	**5.0%**	**68.2%**	**68.1%**	0.5 (0)	0.47 (0.1477)	0.46 (0.1471)	5	5	5
0.2	**0.2**	**5.9%**	**82.2%**	**80.0%**	0.5 (0)	0.46 (0.1570)	0.48 (0.1525)	10	10	10
0.3	**0.3**	**6.3%**	**72.0%**	**66.8%**	0.5 (0)	0.47 (0.1432)	0.48 (0.1336)	15	15	15
0.4	**0.4**	**6.2%**	**64.7%**	**53.0%**	0.5 (0)	0.47 (0.132)	0.48 (0.1063)	20	20	20
0.5	**0.5**	**6.4%**	**61.9%**	**38.6%**	0.5 (0)	0.47 (0.1261)	0.49 (0.0757)	25	25	25
0.6	**0.6**	**6.0%**	**65.0%**	**26.6%**	0.5 (0)	0.47 (0.1326)	0.49 (0.0490)	30	30	30
0.7	**0.7**	**6.1%**	**71.9%**	**17.8%**	0.5 (0)	0.47 (0.1438)	0.49 (0.0295)	35	35	35
0.8	**0.8**	**6.1%**	**82.1%**	**10.6%**	0.5 (0)	0.47 (0.1567)	0.49 (0.0127)	40	40	40
0.9	**0.9**	**4.8%**	**68.3%**	**5.1%**	0.5 (0)	0.47 (0.1477)	0.49 (0.0033)	45	45	45
0.2	0.1	17.5%	84%	82.9%	0.5 (0)	0.33 (0.1308)	0.32 (0.1308)	7.5	8.3	8.4
0.2	0.3	15.3%	80%	76.7%	0.5 (0)	0.55 (0.1513)	0.57 (0.1420)	12.5	12.7	12.9
0.2	0.5	65.4%	88.3%	84.9%	0.5 (0)	0.62 (0.1342)	0.71 (0.0907)	17.5	19.3	20.7
0.2	0.7	97.1%	98.3%	97.8%	0.5 (0)	0.56 (0.1486)	0.79 (0.0476)	22.5	23.9	29.8
0.7	0.2	97.2%	98.5%	97.9%	0.5 (0)	0.38 (0.1408)	0.19 (0.0408)	22.5	25.5	30.2
0.7	0.4	62.1%	85.9%	72.0%	0.5 (0)	0.52 (0.1381)	0.33 (0.0507)	27.5	27.2	30.0
0.7	0.6	13.9%	70.4%	26.6%	0.5 (0)	0.51 (0.1394)	0.45 (0.0372)	32.5	32.5	32.8
0.7	0.8	15.0%	78.9%	22.0%	0.5 (0)	0.38 (0.1396)	0.52 (0.0213)	37.5	36.9	37.6

Table 1: Power or Type-I Error Rate, proportion allocated to the treatment arm n_1/n, and expected number of successes (ENS) for different settings of p_0 and p_1.

Restricted Randomisation, Neyman and RSHIR for Wald test (n=50)

Testing with Z_1										
p_0	p_1	Type-I Error Rate or Power			n_1/n			ENS		
		ρ_{CR}	ρ_{N_1}	ρ_{R_1}	ρ_{CR}	ρ_{N_1}	ρ_{R_1}	ρ_{CR}	ρ_{N_1}	ρ_{R_1}
0.1	**0.1**	**5.0%**	**68.2%**	**68.1%**	0.5 (0)	0.47 (0.1477)	0.46 (0.1471)	5	5	5
0.2	**0.2**	**5.9%**	**82.2%**	**80.0%**	0.5 (0)	0.46 (0.1570)	0.48 (0.1525)	10	10	10
0.3	**0.3**	**6.3%**	**72.0%**	**66.8%**	0.5 (0)	0.47 (0.1432)	0.48 (0.1336)	15	15	15
0.4	**0.4**	**6.2%**	**64.7%**	**53.0%**	0.5 (0)	0.47 (0.132)	0.48 (0.1063)	20	20	20
0.5	**0.5**	**6.4%**	**61.9%**	**38.6%**	0.5 (0)	0.47 (0.1261)	0.49 (0.0757)	25	25	25
0.6	**0.6**	**6.0%**	**65.0%**	**26.6%**	0.5 (0)	0.47 (0.1326)	0.49 (0.0490)	30	30	30
0.7	**0.7**	**6.1%**	**71.9%**	**17.8%**	0.5 (0)	0.47 (0.1438)	0.49 (0.0295)	35	35	35
0.8	**0.8**	**6.1%**	**82.1%**	**10.6%**	0.5 (0)	0.47 (0.1567)	0.49 (0.0127)	40	40	40
0.9	**0.9**	**4.8%**	**68.3%**	**5.1%**	0.5 (0)	0.47 (0.1477)	0.49 (0.0033)	45	45	45
0.2	0.1	17.5%	84%	82.9%	0.5 (0)	0.33 (0.1308)	0.32 (0.1308)	7.5	8.3	8.4
0.2	0.3	15.3%	80%	76.7%	0.5 (0)	0.55 (0.1513)	0.57 (0.1420)	12.5	12.7	12.9
0.2	0.5	65.4%	88.3%	84.9%	0.5 (0)	0.62 (0.1342)	0.71 (0.0907)	17.5	19.3	20.7
0.2	0.7	97.1%	98.3%	97.8%	0.5 (0)	0.56 (0.1486)	0.79 (0.0476)	22.5	23.9	29.8
0.7	0.2	97.2%	98.5%	97.9%	0.5 (0)	0.38 (0.1408)	0.19 (0.0408)	22.5	25.5	30.2
0.7	0.4	62.1%	85.9%	72.0%	0.5 (0)	0.52 (0.1381)	0.33 (0.0507)	27.5	27.2	30.0
0.7	0.6	13.9%	70.4%	26.6%	0.5 (0)	0.51 (0.1394)	0.45 (0.0372)	32.5	32.5	32.8
0.7	0.8	15.0%	78.9%	22.0%	0.5 (0)	0.38 (0.1396)	0.52 (0.0213)	37.5	36.9	37.6

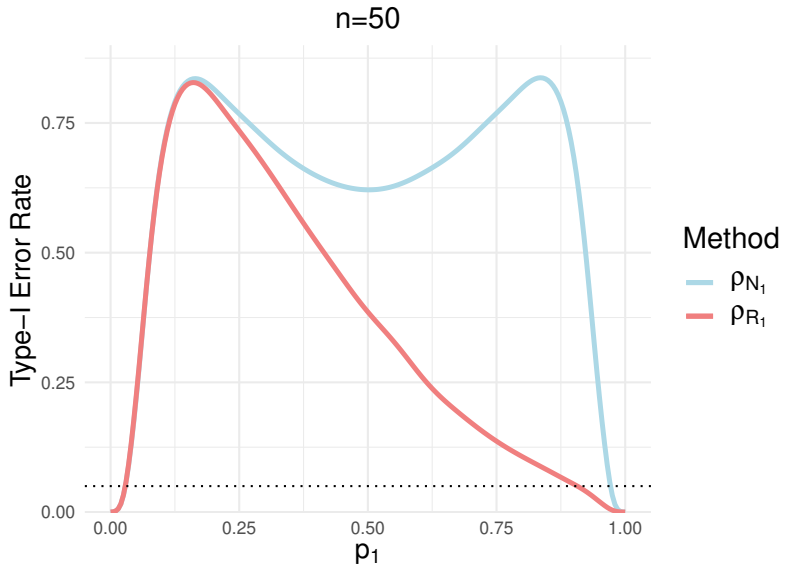
Table 1: Power or Type-I Error Rate, proportion allocated to the treatment arm n_1/n, and expected number of successes (ENS) for different settings of p_0 and p_1.

CR, Neyman and RSHIR for Wald test

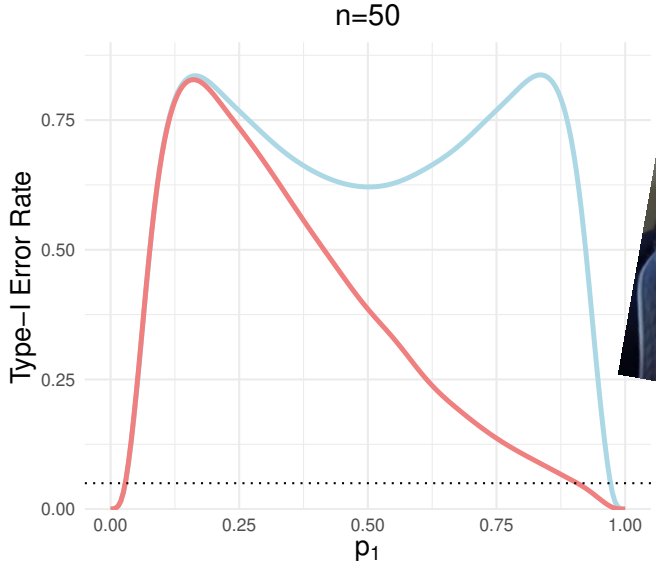
Testing with Z_1										
p_0	p_1	Type-I Error Rate or Power			n_1/n			ENS		
		ρ_{CR}	ρ_{N_1}	ρ_{R_1}	ρ_{CR}	ρ_{N_1}	ρ_{R_1}	ρ_{CR}	ρ_{N_1}	ρ_{R_1}
0.1	**0.1**	**5.0%**	**68.2%**	**68.1%**	0.5 (0)	0.47 (0.1477)	0.46 (0.1471)	5	5	5
0.2	**0.2**	**5.9%**	**82.2%**	**80.0%**	0.5 (0)	0.46 (0.1570)	0.48 (0.1525)	10	10	10
0.3	**0.3**	**6.3%**	**72.0%**	**66.8%**	0.5 (0)	0.47 (0.1432)	0.48 (0.1336)	15	15	15
0.4	**0.4**	**6.2%**	**64.7%**	**53.0%**	0.5 (0)	0.47 (0.132)	0.48 (0.1063)	20	20	20
0.5	**0.5**	**6.4%**	**61.9%**	**38.6%**	0.5 (0)	0.47 (0.1261)	0.49 (0.0757)	25	25	25
0.6	**0.6**	**6.0%**	**65.0%**	**26.6%**	0.5 (0)	0.47 (0.1326)	0.49 (0.0490)	30	30	30
0.7	**0.7**	**6.1%**	**71.9%**	**17.8%**	0.5 (0)	0.47 (0.1438)	0.49 (0.0295)	35	35	35
0.8	**0.8**	**6.1%**	**82.1%**	**10.6%**	0.5 (0)	0.47 (0.1567)	0.49 (0.0127)	40	40	40
0.9	**0.9**	**4.8%**	**68.3%**	**5.1%**	0.5 (0)	0.47 (0.1477)	0.49 (0.0033)	45	45	45
0.2	0.1	17.5%	84%	82.9%	0.5 (0)	0.33 (0.1308)	0.32 (0.1308)	7.5	8.3	8.4
0.2	0.3	15.3%	80%	76.7%	0.5 (0)	0.55 (0.1513)	0.57 (0.1420)	12.5	12.7	12.9
0.2	0.5	65.4%	88.3%	84.9%	0.5 (0)	0.62 (0.1342)	0.71 (0.0907)	17.5	19.3	20.7
0.2	0.7	97.1%	98.3%	97.8%	0.5 (0)	0.56 (0.1486)	0.79 (0.0476)	22.5	23.9	29.8
0.7	0.2	97.2%	98.5%	97.9%	0.5 (0)	0.38 (0.1408)	0.19 (0.0408)	22.5	25.5	30.2
0.7	0.4	62.1%	85.9%	72.0%	0.5 (0)	0.52 (0.1381)	0.33 (0.0507)	27.5	27.2	30.0
0.7	0.6	13.9%	70.4%	26.6%	0.5 (0)	0.51 (0.1394)	0.45 (0.0372)	32.5	32.5	32.8
0.7	0.8	15.0%	78.9%	22.0%	0.5 (0)	0.38 (0.1396)	0.52 (0.0213)	37.5	36.9	37.6

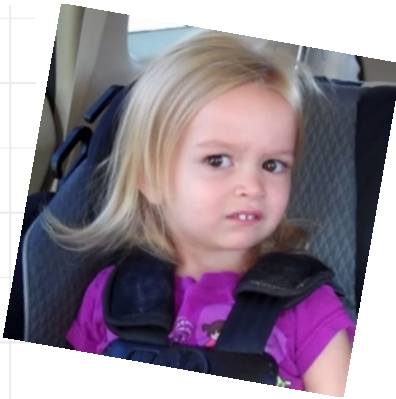
Table 1: Power or Type-I Error Rate, proportion allocated to the treatment arm n_1/n, and expected number of successes (ENS) for different settings of p_0 and p_1.

Type-I Error Rate



Type-I Error Rate





Summary of 4 Existing Approaches

- **Agresti & Caffo correction**
  - Adding a success and failure to each arm.
- Ensuring **non-zero variance** estimators
  - Sampling with equal probability if one of the sample variances is equal to zero.
- Extending the **Burn-In** period B
  - Number of patients randomized with equal (restricted) randomization (CR) before the adaptive period starts.
- Using the **score test**
  - Wald test inherently inflates type-I error (**?**) even under CR.
  - Testing the hypothesis $H_0 : p_0 = p_1$ using

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_0} + \frac{1}{n_1}\right)}}, \tag{1}$$

Rethinking Optimal RAR

Original proportions ρ_{N_1} and ρ_{R_1} for **Wald test** derived for **true unknown values** only optimal in the limit and require MLE Instead: Redefining optimization problem for **score test** and **MLEs** directly i.e. use $\hat{p}_0$ and $\hat{p}_1$ instead of p_0 and p_1 Score test and Wald test differ in **variance** of the test statistics

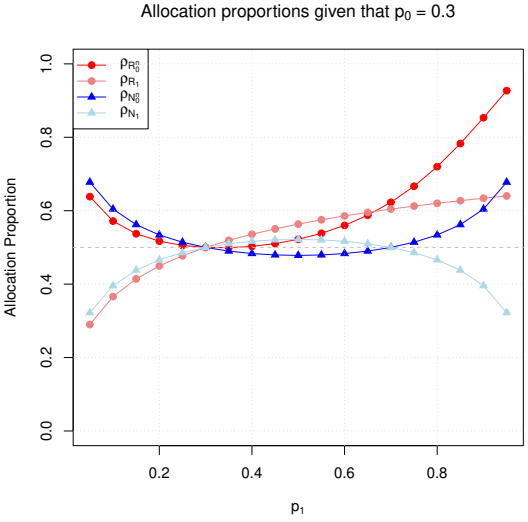
$$\hat{p}\hat{q}\left(\frac{1}{n_0}+\frac{1}{n_1}\right) \quad \text{vs.} \quad \frac{\hat{p}_0\hat{q}_0}{n_0}+\frac{\hat{p}_1\hat{q}_1}{n_1}$$

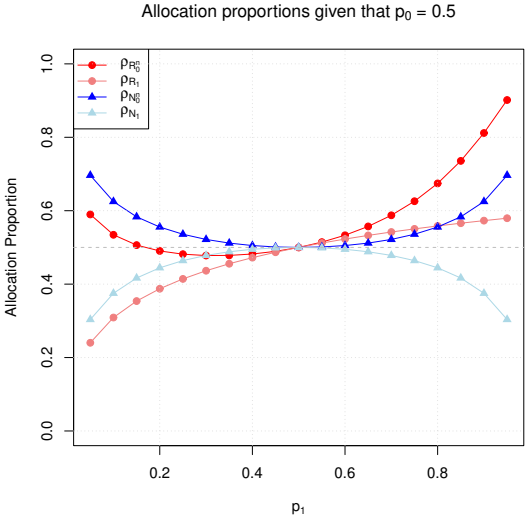
Leading to **two new proportions**:

	Wald	Score
Neyman	$\rho_{N_1}=\frac{\sqrt{p_1 q_1}}{\sqrt{p_0 q_0}+\sqrt{p_1 q_1}}$	$\rho^n_{N_0}=\frac{\sqrt{\hat{p}_0\hat{q}_0}}{\sqrt{\hat{p}_0\hat{q}_0}+\sqrt{\hat{p}_1\hat{q}_1}}$
RSHIR	$\rho_{R_1}=\frac{\sqrt{p_1}}{\sqrt{p_0}+\sqrt{p_1}}$	$\rho^n_{R_0}$ solved numerically

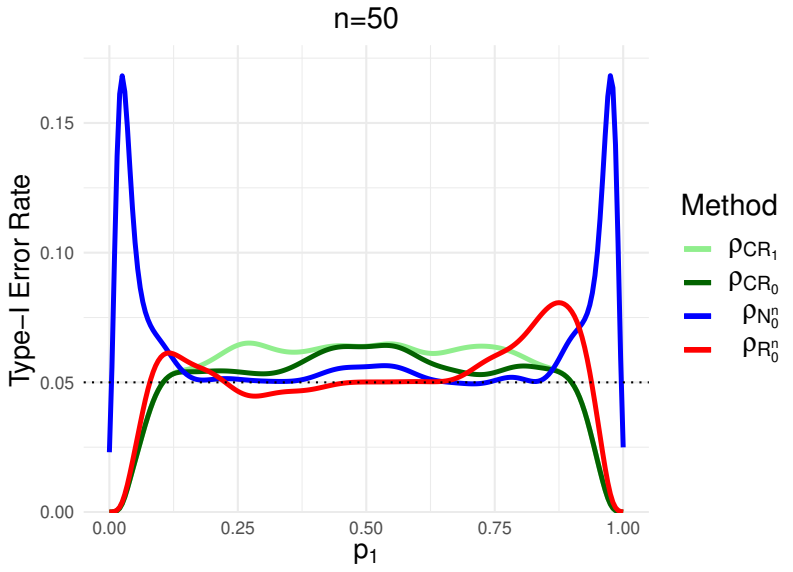
Interestingly: $\rho^n_{N_0} \to 1-\rho_{N_1}$

Theoretical Comparison





Type-I Error Rate



Early Phase Example

Test	Procedure	Type-I error	Power	% sup Arm (Var)	ENS
NAC: $N = 68$, $p_0 = 0.635$, $p_1 = 0.893$					
Z_1	ρ_{CR}	5.5%	75.8%	0.5 (0)	51.9
Z_0	ρ_{CR}	**4.7%**	74.2%	0.5 (0)	51.9
Z_1	ρ_{N_1}	65.7%	94.2%	0.2216 (0.1024)	47.1
Z_1	ρ_{R_1}	23.0%	76.6%	0.5798 (0.0207)	53.4
Z_0	$\rho^n_{N_0}$	**4.6%**	73.6%	0.6064 (0.0033)	53.8
Z_0	$\rho^n_{R_0}$	**4.9%**	73.4%	0.6909 (0.0076)	55.3
CALISTO: $N = 1502$, $p_0 = 0.941$, $p_1 = 0.991$					
Z_1	ρ_{CR}	5%	100%	0.5 (0)	1450.9
Z_0	ρ_{CR}	5%	100%	0.5 (0)	1450.9
Z_1	ρ_{N_1}	96.3%	99.9%	0.1329 (0.1042)	1423.4
Z_1	ρ_{R_1}	5.4%	100%	0.5073 (0.0005)	1451.5
Z_0	$\rho^n_{N_0}$	5.2%	100%	0.7139 (0.0014)	1467
Z_0	$\rho^n_{R_0}$	5.1%	100%	0.8298 (0.0031)	1475.7

Table 5: Results for two different sample sizes. The first set corresponds to $N = 68$, $p_0 = 0.635$, $p_1 = 0.893$ with a number of simulations 10^4 and minimal burn-in of 2 patients per arm. The second set corresponds to $N = 1502$, $p_0 = 0.941$, $p_1 = 0.991$ with the same simulation settings.

Confirmatory Example

Test	Procedure	Type-I error	Power	% sup Arm (Var)	ENS
NAC: $N = 68$, $p_0 = 0.635$, $p_1 = 0.893$					
Z_1	ρ_{CR}	5.5%	75.8%	0.5 (0)	51.9
Z_0	ρ_{CR}	**4.7%**	74.2%	0.5 (0)	51.9
Z_1	ρ_{N_1}	65.7%	94.2%	0.2216 (0.1024)	47.1
Z_1	ρ_{R_1}	23.0%	76.6%	0.5798 (0.0207)	53.4
Z_0	$\rho^n_{N_0}$	**4.6%**	73.6%	0.6064 (0.0033)	53.8
Z_0	$\rho^n_{R_0}$	**4.9%**	73.4%	0.6909 (0.0076)	55.3
CALISTO: $N = 1502$, $p_0 = 0.941$, $p_1 = 0.991$					
Z_1	ρ_{CR}	5%	100%	0.5 (0)	1450.9
Z_0	ρ_{CR}	5%	100%	0.5 (0)	1450.9
Z_1	ρ_{N_1}	96.3%	99.9%	0.1329 (0.1042)	1423.4
Z_1	ρ_{R_1}	5.4%	100%	0.5073 (0.0005)	1451.5
Z_0	$\rho^n_{N_0}$	5.2%	100%	0.7139 (0.0014)	1467
Z_0	$\rho^n_{R_0}$	5.1%	100%	0.8298 (0.0031)	1475.7

Table 5: Results for two different sample sizes. The first set corresponds to $N = 68$, $p_0 = 0.635$, $p_1 = 0.893$ with a number of simulations 10^4 and minimal burn-in of 2 patients per arm. The second set corresponds to $N = 1502$, $p_0 = 0.941$, $p_1 = 0.991$ with the same simulation settings.

Conclusion for Type-I Eorror

- **Summary of Contributions**
  - **Two new optimal allocation proportions** for two-armed trials with binary outcomes.
  - Achieved **type-I error rate control** across the parametric space.
  - **Power and patient-benefit** gain (or loss) depends on region of the parametric space.
- **Limitations**
  - Does not yet extend to **multi-armed** trials.
  - **Binary** endpoints have unique variance properties, influencing optimal allocation.
- **Future Research Directions**
  - **Alternative measures of interest**: adapt approach for relative risk and odds ratios.
  - **Alternative tests**: nonparametric and exact tests.

Outline

RAR: Why? And where next?

- **Final thoughts:**
  - Fully sequential RAR $\longrightarrow$ defines a theoretical upper bound on adaptivity gains.
  - When well-designed and feasible, RAR can enhance patient benefit, personalization, and statistical efficiency, among other benefits.
  - Analytically complex, yet of significant practical importance.
  - In practice, RAR is one component of a broader adaptive design, not a standalone feature.
- **Where to next?**
  - Nonparametric RAR methods
  - Choosing the optimal burn-in period
  - Exact tests & randomisation-based inference
  - Software (design and implementation – in the making)

Acknowledgments

Thank you for listening! :)


David Robertson


Sofia S. Villar


William Rosenberger


Stef Baas

References I

Pin, L., Baas, S., Robertson, D.S. and Villar, S.S. (2025). Is 1:1 Always Most Powerful? Why Unequal Allocation Merits Broader Consideration. *arXiv preprint* arXiv:2507.13036.

Robertson, D.S., López-Kolkovska B., Lee K.M. and Villar S.S. (2023).Response-adaptive randomization in clinical trials: from myths to practical considerations. *Statistical Science* **38 (2)** 185–208.

Thompson, W.R. (1933). On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika* **25** 285–294.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, **97**(4) 558.

Rosenberger, W.F., Stallard, N., Ivanova, A., Harper, C.N. and Ricks, M.L. (2001). Optimal adaptive designs for binary response trials. *Biometrics*, **57** 909–913.

Dumville, J.C., Hahn, S., Miles, J.N.V. and Torgerson, D.J. (2006). The use of unequal randomisation ratios in clinical trials: A review. *Contemporary Clinical Trials* **27**(1), 1–12.

Hey, S.P. and Kimmelman, J. (2014). The questionable use of unequal allocation in confirmatory trials. *Neurology* **82**(1), 77–79. `https://doi.org/10.1212/WNL.0000438226.10353.1c`

References II

Hu, F., Zhang, L.X. and He, X. (2009). Efficient randomized-adaptive designs. *Annals of Statistics*, **37**(5A) 2543–2560.

Pin, L., Villar, S.S. and Rosenberger, W.F. (2024). Response-Adaptive Randomization Designs Based on Optimal Allocation Proportions. In *Biostatistics in Biopharmaceutical Research and Development* (ed. D.G. Chen). Springer. https://doi.org/10.1007/978-3-031-65948-5_12

Pin, L., Villar, S.S. and Rosenberger, W.F. (2025). Revisiting Optimal Allocations for Binary Responses: Insights from Considering Type-I Error Rate Control. *Biometrics*, accepted. doi:10.1093/biomtc/ujaf114

Questions?

Thank you for listing!

lukas.pin@mrc-bsu.cam.ac.uk

Questions?

Looking for feedback on my work: