Randomization in clinical trials: Time for fresh consideration?

Alex Sverdlov, PhD
alex.sverdlov@novartis.com

SCT 2024 Annual Meeting, Boston, MA
May 20-22, 2024

Disclosure

I have no financial disclosure or conflicts of interest with the material in this presentation.

Contents

Background on randomization in clinical trials

A historical perspective on randomization

R. A. Fisher (1890–1962)
“The Design of Experiments”
1935

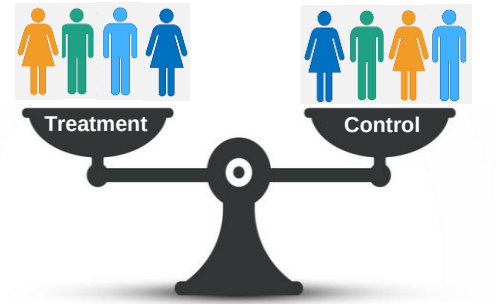
A. Bradford Hill (1897–1991)
First RCT evaluating streptomycin in treating TB
1946

Jerome Cornfield (1912–1979)
“Principles of Research”
1959

Why is randomization important in clinical trials?

1. Helps mitigate selection bias in the design, especially in open-label studies

2. Promotes similarity of treatment groups with respect to important known and unknown confounders



3. Contributes to the validity of statistical estimators and tests and can form the basis for re-randomization tests

What types of randomization designs are available?

Randomization designs

Fixed

- Randomization sequence can be generated in advance

Adaptive

- Treatment assignments are generated dynamically, based on accumulating data in the trial

Adaptive Randomization*

Allocation-adaptive	• To achieve desired treatment allocation ratio
Covariate-adaptive	• To achieve balance on important covariates
Response-adaptive	• To increase allocation to an empirically better treatment
Covariate-adjusted response-adaptive	• A combination of covariate- and response-adaptive

* Rosenberger WF et al. (2012). Adaptive randomization for clinical trials. JBS 22:4, 719-736.

This presentation: Focus on 1:1 RCT

- Throughout this presentation, I will assume:
  - Randomized, parallel group, two-arm, placebo-controlled trial design
  - Equal (1:1) target allocation
  - Randomization sequence can pre-generated before the trial starts
  - Stratified randomization is within the scope
- Adaptive randomization is another very interesting topic, but beyond the scope of this presentation

Mathematics of randomization

- Consider an RCT comparing experimental (E) vs. control (C)
- n subjects to be sequentially randomized between E and C
- A randomization sequence is a random vector:

  $\boldsymbol{\delta}_n = (\delta_1, \ldots, \delta_n)$, where $\delta_i = 1(0)$, if treatment assignment is E (C)
- The 1st patient is randomized between E and C with probability 0.5
- Thereafter, the $(j+1)$st patient is assigned to E with probability

$$\phi_{j+1} = \Pr\left(\delta_{j+1} = 1 \middle| \boldsymbol{\delta}_j\right), j \geq 1$$

(i.e., randomization probability is conditional on the past assignments)

Mathematics of randomization

- Randomization sequence: $\boldsymbol{\delta}_n = (\delta_1, \ldots, \delta_n)$
- $N_E(n) = \sum_{j=1}^{n} \delta_j$ – number of patients assigned to E
- $N_C(n) = n - N_E(n)$ – number of patients assigned to C
- Treatment imbalance: $D(n) = N_E(n) - N_C(n)$
  - $D(n)$ is a random variable that can take values in $\{-n, \ldots, n\}$
  - The distribution of $D(n)$ is determined by the distribution of $\boldsymbol{\delta}_n$
  - It is desired that $D(n) \sim 0$ (i.e., balanced allocation) with high probability

Let us consider four types of 1:1 randomization designs

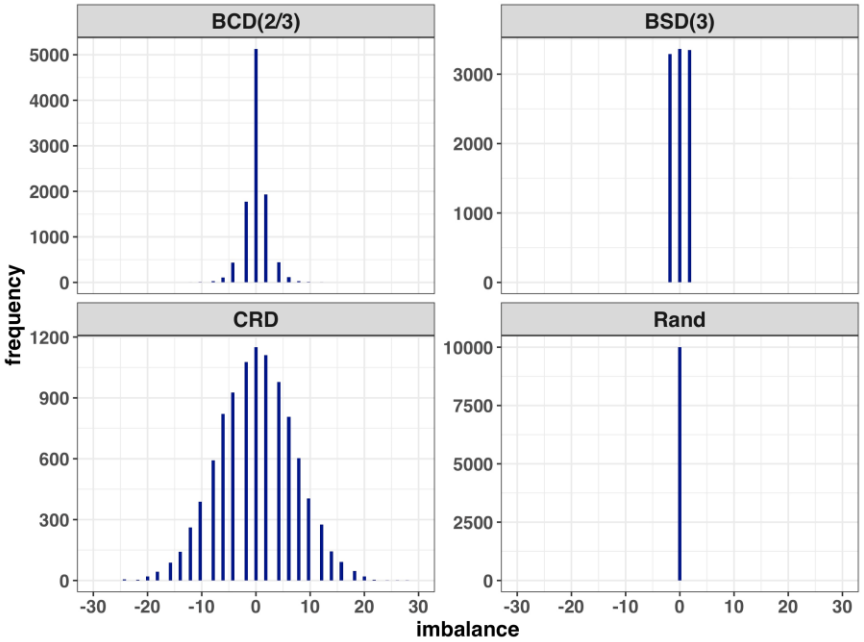
1. Complete randomization (CR)
   - Treatment assignments are made independently, by flip of a fair coin
2. Procedures that achieve exact balance at the end of the study
   - At intermediate allocation steps, it is likely to have some deviations from 1:1 ratio
3. Maximum Tolerated Imbalance (MTI) procedures
   - Treatment imbalance is maintained within user-defined limits, but final group sizes are not necessarily equal
4. Procedures that maintain near-balance throughout the study
   - Treatment imbalance is close to zero with high probability

Selecting a randomization design: balance–randomness tradeoff

Randomization designs vary in the degree of balance they induce

Imbalance is around zero with high probability

Imbalance is around zero, but variability is high



Imbalance is within pre-specified margins (±3)

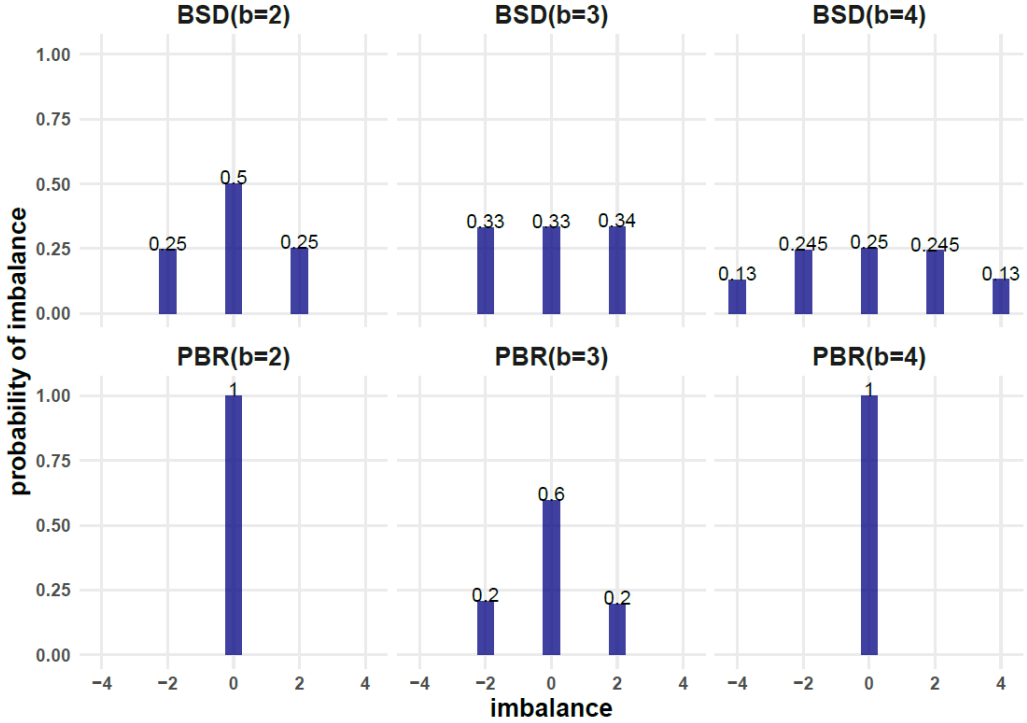
Imbalance is zero with probability 1

Permuted Block Design (PBD) vs. Big Stick Design (BSD)

- Shared properties of PBD and BSD:
  - Both procedures control imbalance within the pre-defined limits: ± *b*
  - Both procedures make the next assignment w.p. 0.5 if the current imbalance is 0
  - Both procedures are equivalent to the permuted blocks of size 2, if b=1
- Some key differences between PBD and BSD:

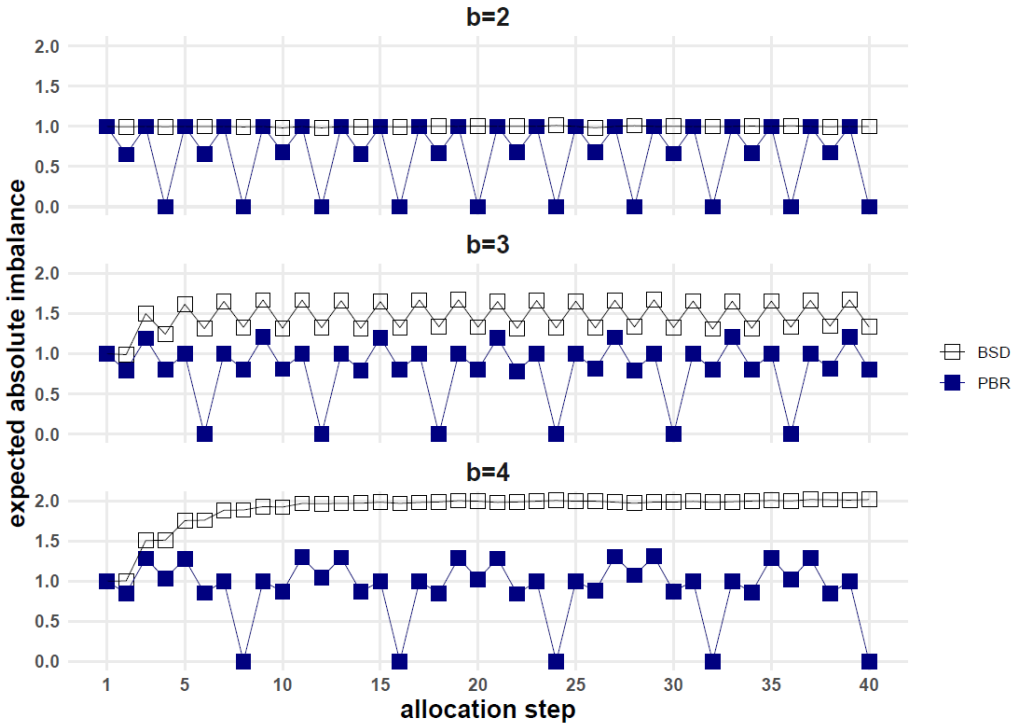
PBD	BSD
• Next assignment(s) in the block are deterministic if the absolute value of current imbalance = *b* • Next assignment(s) in the block may be deterministic even if the absolute value of current imbalance is less than *b*	• Next assignment in the sequence is deterministic (forced to reduce the imbalance by one unit), if and only if the absolute value of current imbalance is equal to *b*

PBD vs. BSD: Probability distribution of the final imbalance for a trial with n=40



Based on 10,000 simulations

PBD vs. BSD: Expected imbalance |D| for a trial with n=40



Based on 10,000 simulations

PBD vs. BSD: allocation randomness

	Expected proportion of Deterministic Assignments		Excess Correct Guess Probability	
b	**PBD**	**BSD**	**PBD**	**BSD**
1	50%	50%	25%	25%
2	33%	25%	21%	12.5%
3	25%	17%	18%	8%
4	20%	12.5%	16.5%	6%

Summary: PBD vs. BSD

- Both procedures ensure imbalance control within pre-specified limits ± *b*
- For PBD, expected absolute imbalance = 0 at every allocation step divisible by $2b$, whereas for BSD it is small (but not necessarily 0) at every step
- BSD provides stronger encryption of the randomization sequence than PBD
  - BSD has fewer deterministic assignments and lower correct guess probability => lower potential for selection bias, especially in open-label trials
- Both procedures are equally simple to implement, either through a pre-generated randomization list in an IRT system or in real-time

Selecting a randomization design: validity and efficiency

Beyond balance and randomness?

- Balance and randomness are ‘surrogate’ measures of goodness of the design
- Validity and efficiency of statistical inference are more relevant criteria
  - Validity = procedure provides correct statistical inference following an RCT (pre-specified probability of a type I error is achieved but not exceeded)
  - Efficiency = high statistical power for detecting meaningful treatment differences when they exist
- Both validity and efficiency are major requirements of any RCT, and they are intertwined with concepts of treatment balance and allocation randomness

Example: Which randomization designs are robust and efficient? (1 of 6)

- Data generating mechanism for the RCT outcomes:

$$Y_i = \delta_i \mu_E + (1 - \delta_i)\mu_C + u_i + \varepsilon_i, \quad i = 1, \ldots, n$$

  - μ_k = mean for treatment $k = E, C$
  - $\delta_i = 1$ (or 0), if the ith subject is randomized to E (C)
  - u_i = unknown term associated with the ith patient
  - ε_i = i.i.d. measurement errors
- We consider 4 models:
  - M1 (Normal random sampling): $u_i \equiv 0$ and $\varepsilon_i \sim$ i.i.d. N(0,1)
  - M2 (Linear time trend): $u_i = 5i/(n+1)$ and $\varepsilon_i \sim$ i.i.d. N(0,1)
  - M3 (Cauchy errors): $u_i \equiv 0$ and $\varepsilon_i \sim$ i.i.d. Cauchy(0,1)
  - M4 (Selection bias*): $u_{i+1} = \nu \cdot 1\{D(i) < 0\} - \nu \cdot 1\{D(i) > 0\}$, and $\varepsilon_i \sim$ i.i.d. N(0,1)

Berger et al. (2021). A roadmap to using randomization in clinical trials . BMC Med Res Meth 21:168

Example: Which randomization designs are robust and efficient? (2 of 6)

- 11 randomization designs to sequentially randomize $n = 50$ subjects:
  - Random allocation rule – Rand
  - Truncated binomial design – TBD
  - Permuted block design with blocks of two – PBD(2)

  Force exact balance
  - Big stick design with MTI=3 – BSD(3)
  - Biased coin design with imbalance tolerance with p=2/3 and MTI=3 – BCDWIT(2/3, 3)

  MTI procedures
  - Biased coin design with p=2/3 – BCD(2/3)
  - Adjustable biased coin design with a=2 – ABCD(2)
  - Generalized biased coin design (GBCD) with $\gamma = 1$ – GBCD(1)
  - GBCD with $\gamma = 2$ – GBCD(2)
  - GBCD with $\gamma = 5$ – GBCD(5)

  Ensure near-balance with high probability
  - Complete randomization – CR

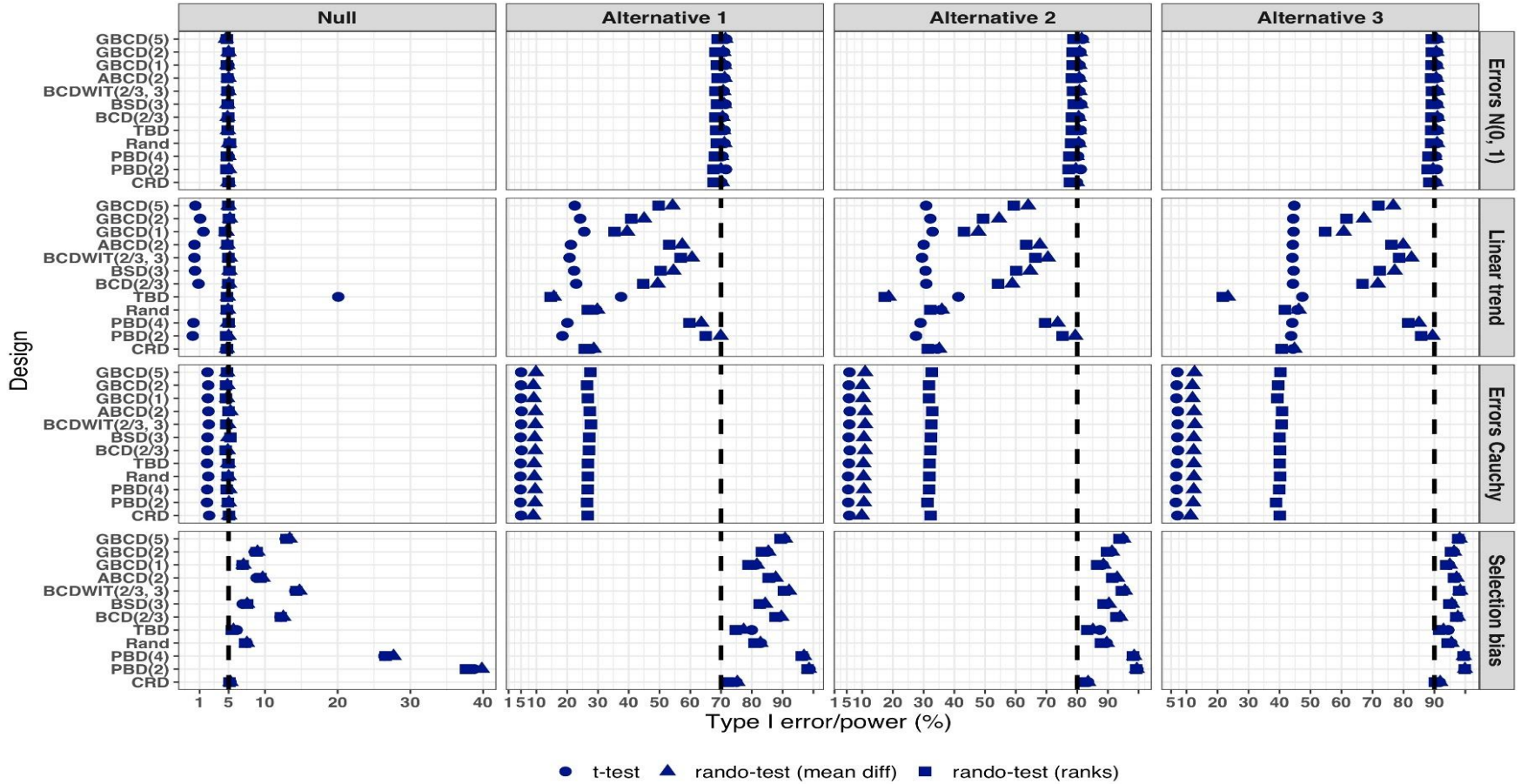
  Most random procedure

Example: Which randomization designs are robust and efficient? (3 of 6)

- 3 statistical tests (analysis strategies):
  - T1 – two-sample t-test — Likelihood-based
  - T2 – RBI test using mean difference
  - T3 – RBI test based on linear rank statistic — Randomization-based
- 4 scenarios for the true mean treatment difference $\Delta = \mu_E - \mu_C$:
  - Null ($\Delta = 0$) – to assess type I error rate
  - Alternatives 1, 2, and 3 (Δ is such that power ~70%, ~80%, and ~90%)

Example: Which randomization designs are robust and efficient? (4 of 6)

Not all designs are 'the same'!



Example: Which randomization designs are robust and efficient? (5 of 6)

- Under M1 (normal random sampling), any combination of design and analysis works well and yields reliable and consistent results
- Under M2 (linear trend), the designs have differential performance:
  - **Type I error rate (T1ER):**
    - Rand and CR maintain T1ER at 5% with all three tests
    - TBD: t-test anticonservative (T1ER ~20%); 8 other procedures: t-test conservative (T1ER in the range 0.1% – 2%)
    - Randomization-based tests: all 11 designs maintain T1ER at 5%.
  - **Power:**
    - Reduced significantly for all designs compared to the ‘normal random sampling’ case
    - t-test is most affected and the randomization-based test using ranks is most robust

Example: Which randomization designs are robust and efficient? (6 of 6)

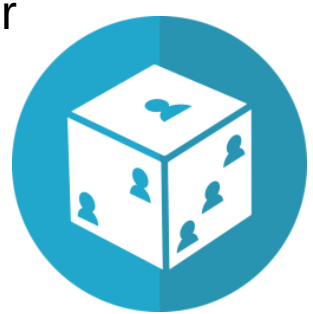
- Under M3 (Cauchy errors), all designs have similar performance:
  - Two randomization-based tests maintain T1ER at 5%; t-test deflates T1ER to 2%
  - Power: t-test is least powerful, and the randomization-based test using ranks is most powerful among the three tests
- Under M4 (Selection bias), the designs have differential performance:
  - **Type I error rate (T1ER):**
    - A bit inflated for TBD (~6%), Rand (~7%), and CR (~7%) with all three tests
    - PBD(2) is most affected: T1ER ~39%
    - In general: the more random the design, the less inflation of T1ER
  - **Power:** artificially inflated compared to the ‘normal random sampling’ case

Summary of Example

- Choice of a randomization design does matter

- Choice of a data analytic technique (parametric or nonparametric) does matter

- Decision on whether to include randomization in the analysis or not (RBI or population-based analysis) does matter

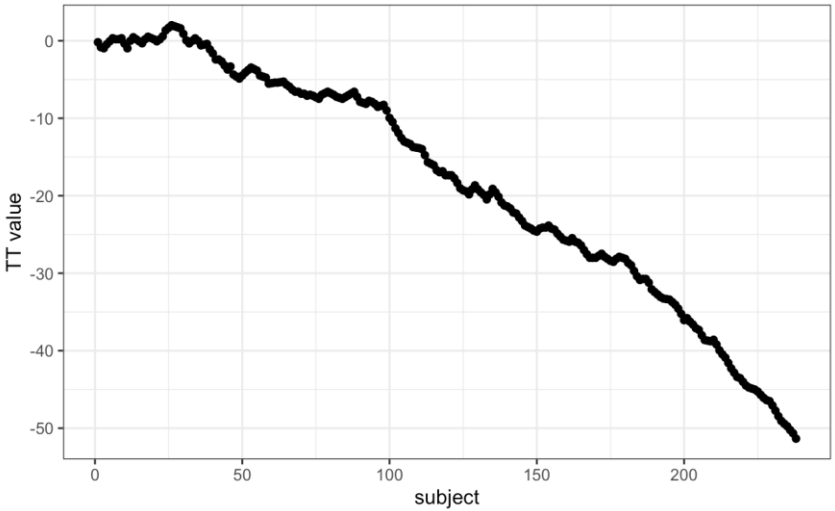


Randomization and covariate adjustment

Why is covariate adjustment important?

- Consider 1:1 randomized survival trial to compare azathioprine vs. placebo in patients with primary biliary cirrhosis*
- n=248 patients were enrolled and randomized over 7 years
- Serum bilirubin at entry was a major factor correlated with survival, and it exhibited a strong decreasing trend over time
- Analysis using Cox PH model showed significant treatment effect (p=0.01) only after adjusting for bilirubin

(reproduced from Altman and Royston, 1988)



* Altman and Royston (1988) The hidden effect of time. Statistics in Medicine 7, 629-637.

Why is covariate adjustment important?

- We simulated the given RCT with time trend in serum bilirubin (10,000 runs)
- Data generating model for the hazard:

$$h_i(t, \delta_i) = h_c(t)\exp(\delta_i \log HR + u_i), \;\; i = 1, \ldots, 248$$

  - $h_c(t) \equiv 1$ (exponential distribution)
  - $u_i =$ log serum bilirubin for the ith patient at study entry
  - $HR =$ true treatment hazard ratio
- 4 randomization designs: CR, Rand, TBD, and PBD(2)
- 2 values for the treatment effect: $HR = 1$ (Null); $HR = 0.6$ (Alternative)
- 2 analyses using Cox PH model: unadjusted and adjusted for bilirubin

Why is covariate adjustment important?

Probability of a statistically significant baseline imbalance in serum bilirubin:

	P<0.05	P<0.025	P<0.01
CR	0.1030	0.0621	0.0321
Rand	0.0091	0.0023	0.0006
TBD	0.2379	0.1677	0.1102
PBD(2)	0.0000	0.0000	0.0000

- Randomization designs differ with respect to baseline covariate imbalance under the time trend

	Type I error rate		Power	
	Unadj	Adj	Unadj	Adj
CR	0.048	0.050	0.611	0.969
Rand	0.052	0.051	0.619	0.970
TBD	0.145	0.051	0.586	0.970
PBD(2)	0.006	0.051	0.654	0.970

- Only CR and Rand are valid with unadjusted analysis
- Covariate-adjusted analyses are all valid
- Covariate-adjusted analyses are much more powerful than unadjusted

Summary

It’s good time to take a fresh look at the concept of randomization

- Existing practices on randomization – areas for improvement / innovation?
- Trials that may benefit from randomization-based inference?
- Implement novel randomization designs that are better than “standard” ones
- Develop / implement simulation software that helps judiciously select a randomization procedure for a given RCT
- Develop systematic guidelines and best practices on randomization

Thank you!

Please visit the LinkedIn page of our Randomization Working Group

Boehringer Ingelheim

Johannes Krisam

Predictability of allocation sequences under central randomization in a multi-center clinical trial

Why the patient enrollment pattern matters

Life forward

Disclosure

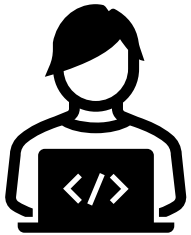
I have no financial disclosure or conflicts of interest with the material in this presentation.

Planning a randomized multi-center trial

- Imagine you‘re a trial statistician planning a **randomized multi-center clinical trial**
- One important aspect is the definition of the **randomization design**

- The **permuted block design (PBD)** is the first option that comes to your mind
- Also you‘ve recently come across the so-called **big stick design (BSD)** that can achieve the **same degree of imbalance control**, accompanied with a **higher degree of randomness**

- The trial you‘re planning is supposed to be **an open-label trial**, so **selection bias** could be an issue – maybe it might make sense to take a look into the **BSD**?

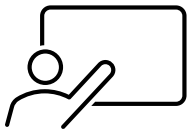


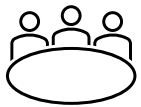
PBD vs BSD – which design to choose for your trial?

Source:
Berger *et al. BMC Med Res Methodol* (2021) 21:168

MTI	Design	Excess correct guess probability
2	PBD	20.8%
	BSD	12.5%
3	PBD	18.3%
	BSD	8.3%

- You do some literature research and find out that there is some real benefit in terms of excess correct guess probability when using **BSD** over **PBD** – this seems to be the way to go for your multi-center open-label trial!
- Enthusiastically, you propose using the **BSD** to your team of stakeholders!





PBD vs BSD – which design to choose for your trial?

Source:

Berger *et al. BMC Med Res Methodol* (2021) 21:168

MTI	Design	Excess correct guess probability
2	[illegible]	20.8%
		12.5%
3	PBD	18.3%
	BSD	8.3%

- However, to your surprise, they are not very enthusiastic about all of your arguments regarding the benefits of BSD:

What is center-stratified randomization?

Randomization list

SeqNo	RandNo	Block	Treatment
1	154	1	B
2	254	1	B
3	212	1	A
4	184	1	A
5	152	2	A
6	135	2	A
7	289	2	B
8	105	2	B
9	222	3	A
10	114	3	B
11	153	3	B
12	285	3	A

Schedule of enrolment

PatNo	Time	Center
1	7/27/2022 (9:45 AM)	Center 1 (France)

- If the randomization is stratified by center:
  - ➢ The IRT allocates complete blocks to each center.

What is center-stratified randomization?

Randomization list

SeqNo	RandNo	Block	Treatment
1	154	1	B
2	254	1	B
3	212	1	A
4	184	1	A
5	152	2	A
6	135	2	A
7	289	2	B
8	105	2	B
9	222	3	A
10	114	3	B
11	153	3	B
12	285	3	A

Schedule of enrolment

PatNo	Time	Center
1	7/27/2022 (9:45 AM)	Center 1 (France)
2	7/28/2022 (9:52 AM)	Center 2 (Italy)

- If the randomization is stratified by center:
  - The IRT allocates complete blocks to each center.

What is center-stratified randomization?

Randomization list

SeqNo	RandNo	Block	Treatment
1	154	1	B
2	254	1	B
3	212	1	A
4	184	1	A
5	152	2	A
6	135	2	A
7	289	2	B
8	105	2	B
9	222	3	A
10	114	3	B
11	153	3	B
12	285	3	A

Schedule of enrolment

PatNo	Time	Center
1	7/27/2022 (9:45 AM)	Center 1 (France)
3	7/29/2022 (10:00 AM)	Center 1 (France)
4	7/29/2022 (10:03 AM)	Center 1 (France)
5	7/29/2022 (10:04 AM)	Center 1 (France)
2	7/28/2022 (9:52 AM)	Center 2 (Italy)
6	7/29/2022 (10:08 AM)	Center 1 (France)

- If the randomization is stratified by center:
  - ➢ The IRT allocates complete blocks to each center.

What is central randomization (not stratified by center)?

Randomization list

SeqNo	RandNo	Block	Treatment
1	154	1	B
2	254	1	B
3	212	1	A
4	184	1	A
5	152	2	A
6	135	2	A
7	289	2	B
8	105	2	B
9	222	3	A
10	114	3	B
11	153	3	B
12	285	3	A

Schedule of enrolment

PatNo	Time	Center
1	7/27/2022 (9:45 AM)	Center 1 (France)

- If the randomization is not stratified by center:
  - The blocks are shared between the centers that currently enroll

What is central randomization (not stratified by center)?

Randomization list

SeqNo	RandNo	Block	Treatment
1	154	1	B
2	254	1	B
3	212	1	A
4	184	1	A
5	152	2	A
6	135	2	A
7	289	2	B
8	105	2	B
9	222	3	A
10	114	3	B
11	153	3	B
12	285	3	A

Schedule of enrolment

PatNo	Time	Center
1	7/27/2022 (9:45 AM)	Center 1 (France)
2	7/28/2022 (9:52 AM)	Center 2 (Italy)

- If the randomization is not stratified by center:
  - The blocks are shared between the centers that currently enroll

What is central randomization (not stratified by center)?

Randomization list

SeqNo	RandNo	Block	Treatment
1	154	1	B
2	254	1	B
3	212	1	A
4	184	1	A
5	152	2	A
6	135	2	A
7	289	2	B
8	105	2	B
9	222	3	A
10	114	3	B
11	153	3	B
12	285	3	A

Schedule of enrolment

PatNo	Time	Center
1	7/27/2022 (9:45 AM)	Center 1 (France)
2	7/28/2022 (9:52 AM)	Center 2 (Italy)
3	7/29/2022 (10:00 AM)	Center 1 (France)
4	7/29/2022 (10:03 AM)	Center 1 (France)
5	7/29/2022 (10:04 AM)	Center 1 (France)
6	7/29/2022 (10:08 AM)	Center 1 (France)
7	7/30/2022 (11:00 AM)	Center 2 (Italy)

- If the randomization is not stratified by center:
  - The blocks are shared between the centers that currently enroll

Advantages of central randomization over center-stratified randomization

- Central randomization ensures that, overall, the treatment assignments are practically balanced
- In theory, it could be expected that there is little potential for an investigator to guess the subsequent treatment assignment within his or her own center, as other centers also enroll patients concurrently and the investigator only knows the assignments in his or her own center.
- Thus, there should indeed be not too much benefit from using a BSD over a PBD in a multi-center RCT using central-randomization – but what if…

Randomization list			Schedule of enrolment		
SeqNo	Block	Treatment	PatNo	Time	Center
1	1	B	1	7/27/2022 (9:45 AM)	Center 1 (France)
2	1	B	2	7/28/2022 (9:52 AM)	Center 2 (Italy)
3	1	A	3	7/28/2022 (4:00 PM)	Center 3 (Belgium)
4	1	A	4	7/29/2022 (10:03 AM)	Center 4 (Italy)
5	2	A	5	7/29/2022 (10:04 AM)	Center 1 (France)
6	2	A	6	7/29/2022 (10:08 AM)	Center 4 (Italy)
7	2	B	7	7/30/2022 (11:00 AM)	Center 2 (Italy)
8	2	B	8	7/30/2022 (11:05 AM)	Center 2 (Italy)
9	3	A	9	7/31/2022 (11:12 AM)	Center 1 (France)
10	3	B	10	7/31/2022 (5:02 PM)	Center 5 (Canada)
11	3	B	11	8/1/2022 (9:44 AM)	Center 4 (Italy)
12	3	A	12	8/1/2022 (9:44 AM)	Center 1 (France)

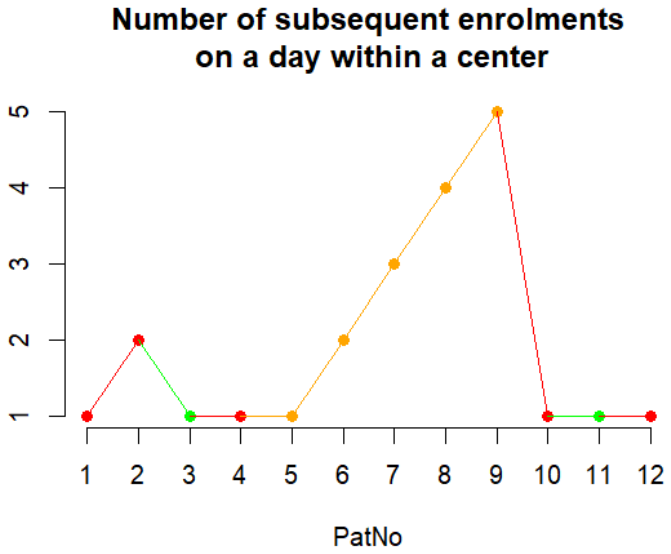
…clinical practice contradicts the assumption of a „random patient flow“

- Some study centers may have **„spikes“** in recruitment when multiple participants in a sequence are enrolled and randomized on the same day.
- Reasons:
  - ➢ Specialized institution has eligible patients waiting for a study to initiate – all of these patients are enrolled once the study goes live
  - ➢ Study may require some highly time-consuming tasks to be done at the randomization visit - center schedules the visit for their patients on the same time day to save time

Randomization list			Schedule of enrolment		
SeqNo	Block	Treatment	PatNo	Time	Center
1	1	B	1	7/27/2022 (9:45 AM)	Center 1 (France)
2	1	B	2	7/27/2022 (9:52 AM)	Center 1 (France)
3	1	A	3	7/28/2022 (11:45 AM)	Center 2 (Italy)
4	1	A	4	7/29/2022 (9:45 AM)	Center 1 (France)
5	2	A	5	7/29/2022 (10:03 AM)	Center 3 (Belgium)
6	2	A	6	7/29/2022 (10:08 AM)	Center 3 (Belgium)
7	2	B	7	7/29/2022 (10:15 AM)	Center 3 (Belgium)
8	2	B	8	7/29/2022 (10:18 AM)	Center 3 (Belgium)
9	3	A	9	7/29/2022 (10:23 AM)	Center 3 (Belgium)
10	3	B	10	7/29/2022 (11:02 PM)	Center 1 (France)
11	3	B	11	7/29/2022 (11:45 AM)	Center 2 (Italy)
12	3	A	12	8/1/2022 (9:44 AM)	Center 1 (France)

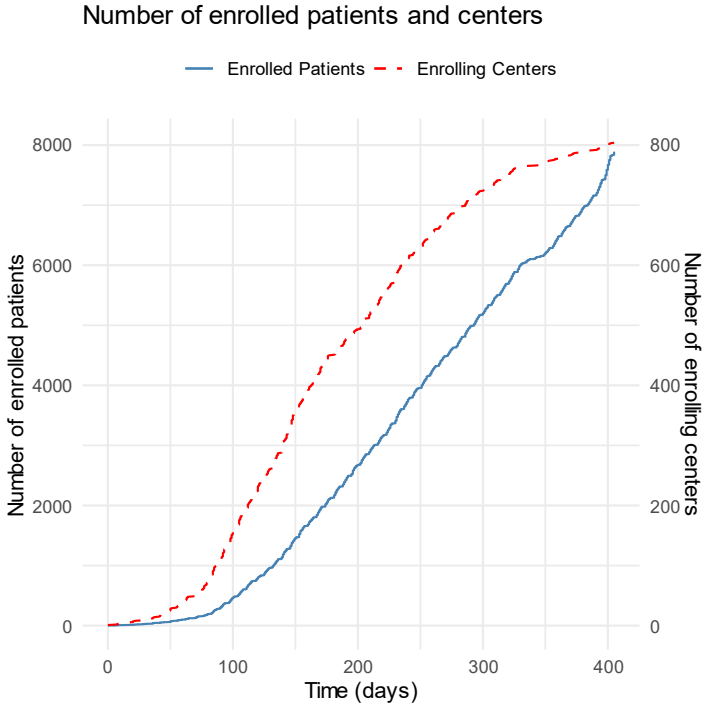
Boehringer Ingelheim

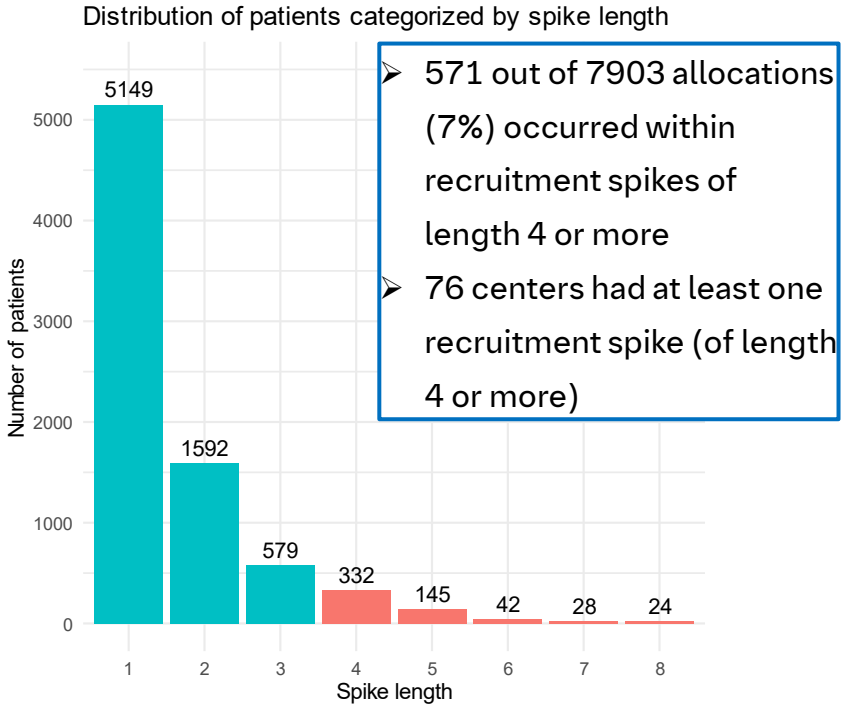
…clinical practice contradicts the assumption of a „random patient flow“



Randomization list			Schedule of enrolment		
SeqNo	Block	Treatment	PatNo	Time	Center
1	1	B	1	7/27/2022 (9:45 AM)	Center 1 (France)
2	1	B	2	7/27/2022 (9:52 AM)	Center 1 (France)
3	1	A	3	7/28/2022 (11:45 AM)	Center 2 (Italy)
4	1	A	4	7/29/2022 (9:45 AM)	Center 1 (France)
5	2	A	5	7/29/2022 (10:03 AM)	Center 3 (Belgium)
6	2	A	6	7/29/2022 (10:08 AM)	Center 3 (Belgium)
7	2	B	7	7/29/2022 (10:15 AM)	Center 3 (Belgium)
8	2	B	8	7/29/2022 (10:18 AM)	Center 3 (Belgium)
9	3	A	9	7/29/2022 (10:23 AM)	Center 3 (Belgium)
10	3	B	10	7/29/2022 (11:02 PM)	Center 1 (France)
11	3	B	11	7/29/2022 (11:45 AM)	Center 2 (Italy)
12	3	A	12	8/1/2022 (9:44 AM)	Center 1 (France)

Do these recruitment spikes really happen in clinical practice? Assessments based on a clinical trial data example





- 571 out of 7903 allocations (7%) occurred within recruitment spikes of length 4 or more
- 76 centers had at least one recruitment spike (of length 4 or more)

Source: Krisam et al. (2024): Understanding an impact of patient enrollment pattern on predictability of central (unstratified) randomization in a multi-center clinical trial. Accepted at Statistics in Medicine

Boehringer Ingelheim

PBD vs BSD: Predictability revisited unter central randomization

- Now being aware of these recruitment spikes in our clinical trial data example, let‘s assess the impact on the **excess correct guess probability**

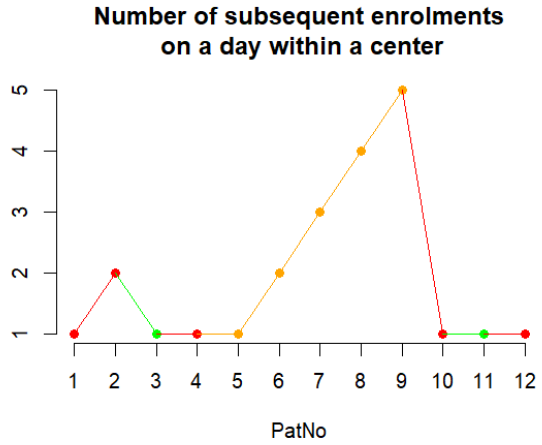
MTI	Design	Outside of recruitment spikes* (n=7332)	Within recruitment spikes* (n=571)	Overall in the study (n=7903)	Probability for monocenter trial (Berger et al. 2021)
2	PBD	1.9%	10.5%	2.6%	20.8%
	BSD	1.8%	7.9%	2.2%	12.5%
3	PBD	1.3%	7.5%	1.7%	18.3%
	BSD	1.2%	5.1%	1.5%	8.3%

*: A recruitment spike is defined as four or more patients being enrolled within one center on the same day

Note: Results are based on 10,000 simulated datasets

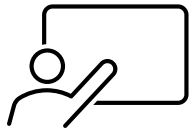
Source: Krisam et al. (2024): Understanding an impact of patient enrollment pattern on predictability of central (unstratified) randomization in a multi-center clinical trial. Accepted at Statistics in Medicine

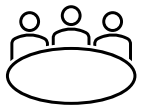
Going back to the design discussion



„We‘re using central randomization, so there‘s **no selection bias with PBD**!“

“Well that’s not true. Depending on the **patient enrollment pattern**, there still might be selection bias due to **recruitment spikes**. A BSD could still prove beneficial for our centrally randomized trial!





MTI	Design	Outside of recruitment spikes* (n=7332)	Within recruitment spikes* (n=571)	Overall in the study (n=7903)	Probability for monocenter trial (Berger et al. 2021)
2	PBD	1.9%	10.5%	2.6%	20.8%
	BSD	1.8%	7.9%	2.2%	12.5%
3	PBD	1.3%	7.5%	1.7%	18.3%
	BSD	1.2%	5.1%	1.5%	8.3%

Boehringer Ingelheim

Summary

- In a multi-center RCT using central randomization, it is possible to have so-called **recruitment spikes.**
- Spikes can occur if **multiple participants are recruited by the same center on the same day** (or over a longer time interval if other centers are not recruiting participants)
- Such spikes may open the **potential for making intelligent guesses** of treatment assignments in the sequence which may lead to **selection bias**
- If such spikes are expected, the following strategies may be useful:
  - Consider **evaluating the predictability of the chosen randomization design** through simulations at the study planning stage
  - Instead of permuted block design**, consider using MTI randomization procedures** such as the big stick design
  - **Avoid disclosure of the overall recruitment progress** to individual investigators such that an investigator from a given study center is not aware of the possible lack of recruitment activity at other centers
  - Use **scrambled allocation numbers** instead of consecutive allocation numbers to make it more difficult for an investigator to e.g. figure out whether they still are on an uninterrupted recruitment spike

References

- Berger V, Bour L, Carter K et al. (2021). A roadmap to using randomization in clinical trials. BMC Med Res Methodol 21, 168
- Krisam J, Ryeznik Y, Carter K, Kuznestova O, Sverdlov O (2024): Understanding an impact of patient enrollment pattern on predictability of central (unstratified) randomization in a multi-center clinical trial.
  Accepted at Statistics in Medicine.

Visit the LinkedIn page of our Randomization Working Group!

Allocation Ratio Preserving Randomization Procedures

Olga Kuznetsova,
Merck & Co., Inc., Rahway, NJ, USA
SCT May 2024, Boston

Disclosures

- Nothing to disclose

ACKNOWLEDGEMENTS

- THANK YOU to my great collaborators in the unequal randomization field Yevgen Tymofyeyev and Victoria Plamadeala Johnson
- THANK YOU to the Randomization Working Group for great discussions and collaboration

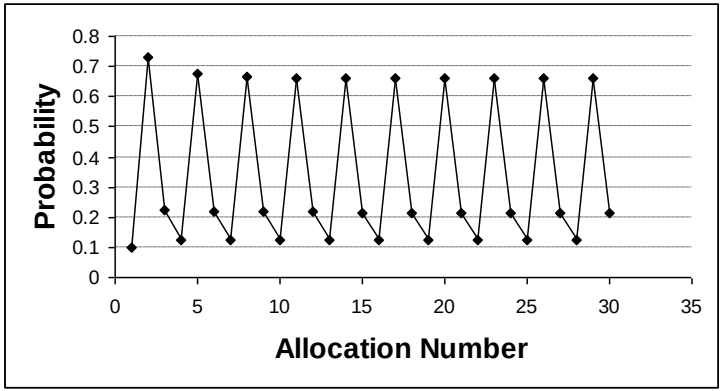
Please follow us on LinkedIn

Contents

- Variations in unconditional allocation ratio may arise with unequal allocation
- Importance of preserving the unconditional allocation ratio
- ARP unequal allocation expansion through mapping
- Other existing approaches to ARP expansion
- What is missing among the ARP tools
- What to do when no ARP procedure that meets the study need is available

Big Question: Is Unconditional Allocation Ratio Preserved at Every Allocation?

- With unequal allocation, care needs to be taken to ensure that 1st, 2nd, 3rd, … subjects are allocated at the same allocation ratio (unconditional AR)
- Called Allocation Ratio Preserving (ARP) property [Kuznetsova and Tymofyeyev 2012]
- Present in all equal allocation procedures symmetric w.r.t all treatment arms
- BUT achieving ARP is not trivial when expanding an equal allocation procedure to unequal allocation
- Example: Unconditional allocation probabilities of Treatment 1 allocation with 1:2 BCM allocation [Han et al. 2009]
- Variations in probability of Treatment 1 allocation:
  - 1st subject: 0.1
  - 2nd subject: 0.7
  - 3rd subject: 0.2
  - … pattern continues



Why is the ARP Property Important?

- Variations in the unconditional allocation ratio are undesirable:
  - Provide potential for selection and evaluation bias even in double-blind studies
    - If it is known that subject randomized first has higher than average chance to receive Active treatment, a subject with better prognosis can be selected for the 1st slot
  - Provide a potential for accidental bias confounded with the variations
    - In particular, in multi-center studies with randomization stratified by center – as was the case in Proschan et al. example.
  - Lead to a shift in the re-randomization distribution [Proschan et al. 2011, Kuznetsova and Tymofyeyev 2012]
    - Lowers the power of re-randomization test
  - Cause the treatment effect estimator to be biased from a randomization perspective (Kaiser 2012)
- ARP violation is a common problem with the expansions of fixed and dynamic allocation procedures to unequal allocation
  - Need to be watchful!
- ARP property is required for some theoretical results regarding the inference
  - As in Ye and Shao (2020)

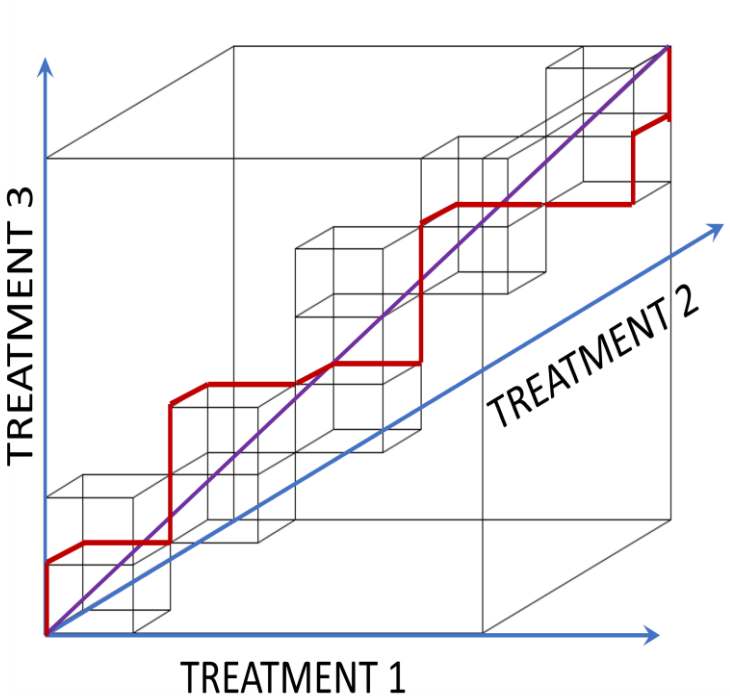
Any Multi-arm Equal Allocation Procedure Can be Expanded to an ARP Unequal Allocation Procedure Through Mapping [K&T, 2012]

- To execute an ARP allocation to $K \geq 2$ treatment groups $G_1, \ldots, G_K$ in $C_1 : \ldots : C_K$ ratio ($S = C_1 + \ldots + C_K$), where C_i are integers:
  1. Execute equal allocation to S "fake" treatment arms $F_1, \ldots, F_s$ following the equal allocation procedure
  2. Map groups of "fake" treatment arms to the actual treatment arms:
     - first C_1 "fake" treatment arms are mapped to G_1
     - next C_2 "fake" treatment arms are mapped to G_2 …
     - last C_K "fake" treatment arms are mapped to G_k.
- Due to symmetry with respect to fake treatment arms $F_1, \ldots, F_s$, such procedure provides
  - 1:1:..:1 unconditional allocation ratio at every allocation
  - Therefore, $C_1 : \ldots : C_K$ unconditional allocation ratio to actual treatment groups
- Examples of procedures obtained through mapping:
  - Permuted Block Design
  - Block Urn Design by Zhao and Weng,
  - Drop-the-Loser Urn Design by Ivanova
  - expansions of minimization and other dynamic procedures by Kuznetsova and Tymofyeyev

The Need for Tight Approximation of the Target Allocation Ratio Throughout the Enrollment

- Problem: when C_i are large integers, the allocation space is wide (wider than for PBR) and an allocation sequence can deviate a lot from the target allocation ratio
- Tighter adherence to the target AR is needed often when the allocation ratio gives rise to a large block size
  - Based on efficiency considerations
    - Draft FDA Guidance on master Protocols (2023) mentions 1:1:$\sqrt{2}$ for comparisons of two active treatments to a Control
    - When approximated with 5:5:7 allocation ratio, leads to a large block size of 17
  - In platform trials, where arms enter and exit
  - To keep close to the target AR with small cohorts
    - Response-adaptive allocation
    - Dose-ranging cohorts
- Need approaches other than mapping in this case

Brick Tunnel Randomization (BTR) [K&T 2011] Provides Tight Adherence to the Target Allocation Ratio



- Example: BTR in 5:5:7 Ratio
- An allocation path (red line) can be depicted on a 3-dimensional unitary grid
  - 3 axes represent 3 treatment arms
  - A step along the Treatment *i* axis depicts the allocation to Treatment *i*, *i*=1,2,3
- The diagonal represents the target allocation ratio
  - Achievable only after 17 allocations
- PBR allows all allocation paths within the 5×5×7 block
- BTR allows only the allocation paths that are contained in the sequence of unitary cubes pierced by the diagonal of the block (the purple line)
  - Brick Tunnel – a sequence of unitary cubes (bricks)
  - 11 bricks out of 5*5*7=175
- Thus, observed AR is always close to the target AR
- BTR allows AR expressed through irrational numbers

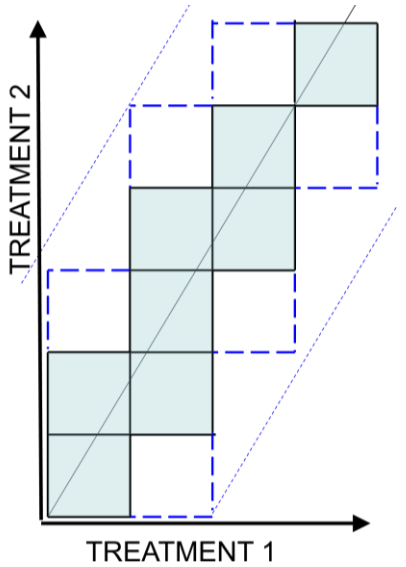
Allocation sequence: 3,2,1,3,3,2,1,2,1,3,3,2,1,1,3,2,3

Unequal Allocation Expansions of 2-arm Procedures with Maximum Tolerated Imbalance

- For single-center open-label studies selection bias is a concern
- For 2-arm equal allocation, several excellent procedures with Maximum Tolerated Imbalance (MTI) were developed (see Zhao et al. 2024)
  - Require $|N_2 - N_1| \leq b$ at any point of randomization
  - Help reduce selection bias compared to PBR while keeping the imbalance low
  - Are easily expanded to multiple arms: $|N_j - N_m| \leq b$ *for all j, m*
  - Thus easily expanded to unequal allocation through mapping
- When an MTI procedure is expanded through mapping, the allocation space is
  - The sequence of $C_1 \times C_2$ blocks (as with $C_1{:}C_2$ PBR) for b=1
  - The allocation space of the $C_1{:}C_2$ Block Urn Design (Zhao & Weng) for b>1
    - Wider than the allocation space for PBR with block size of $2(C_1 \times C_2)$
- Rather wide when block size C_1+C_2 is large

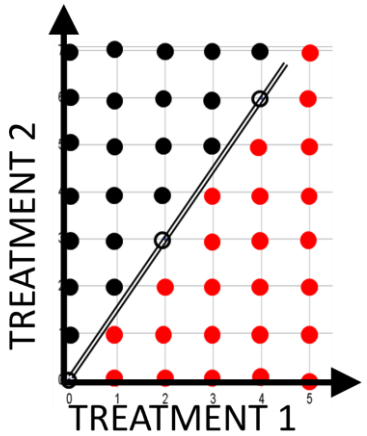
2-arm MTI Unequal Allocation Procedures with Allocation Space Wider than for BTR but Narrower than for PBR

- For C_1:C_2 randomization to Treatments 1 and 2, imbalance in treatment group totals is typically measured through weighted difference $|N_2 - N_1 \times C_2/C_1|$
  - Direct substitution of $|N_2 - N_1|$ with $|N_2 - N_1 \times C_2/C_1|$ in the randomization algorithm typically leads to a non-ARP procedure
- Wide Brick Tunnel by Kuznetsova and Tymofyeyev (2014)
  - Fills in the strip $|N_{2i} - N_{1i} \times C_2/C_1| \leq b$ in 2-dimensional grid
  - Example: the allocation space of the 3:5 BTR and Wide BTR with b=10/3
- Note: expansions of MTI procedures through mapping do not necessarily fill in the entire strip
  - This is not a problem
  - Also, when imbalance is 0, probability of allocation is not C_1:C_2
    - Except for the 1st allocation



ARP Expansions of Biased Coin Randomization (Efron 1971) to Unequal Allocation Not Based on Mapping

- Kuznetsova and Plamadeala (2017) offered different BCR expansions for C_1:C_2 randomization to Treatments 1 and 2
  - not based on mapping
- After *i* allocations,
  - If $N_{2i} = N_{1i} \times C_2/C_1$, allocate in C_1:C_2 ratio (black circles)
  - If $N_{2i} > N_{1i} \times C_2/C_1$, allocate TRT1 with probability $P_{i\_above} > C_1/(C_1+C_2)$ (black dots)
  - If $N_{2i} < N_{1i} \times C_2/C_1$, allocate TRT1 with probability $P_{i\_below} < C_1/(C_1+C_2)$ (red dots)
- Probabilities P_{i_above} and P_{i_below} vary by generation *i* and are derived to ensure ARP property
  - Forces treatment totals to stay close to C_1:C_2 ratio
- One approach: BCR With Preset Proportion of Maximal Forcing
  - P_{i_above} and P_{i_below} are derived in a way that when forcing approaches maximum, BCR approaches BTR
  - Similar to 1:1 BCR that approaches PBR with block size 2 when bias $p \rightarrow 1$,
- With no forcing, BCR reverts to Complete Randomization
  - As with 1:1 BCR



Unequal Allocation Procedures That Lack ARP Property (FYI)

- Urn design described by Rosenberger and Lachin
- Expansion of the maximal procedure by Salama et al.
- Biased coin randomization and minimization expansion by Han et al.
- Doubly adaptive biased coin design procedure by Hu and Zhang applied to fixed unequal allocation as described by Sverdlov and Zhang,
- Minimum quadratic distance constrained balance randomization by Titterington
- Adaptation of biased coin randomization by Frane
- Generalized method for adaptive randomization by Russel et al.
- Generalized multidimensional dynamic allocation method by Lebowitsch et al.
- Many other procedures
- See Sverdlov & Ryeznik 2019 for evaluation of the variations in the unconditional allocation ratio with many of these procedures

What is Missing Among the ARP Randomization Tools?

- ARP procedures for >2 arms not based on mapping – other than BTR
  - Note: if BTR space needs to be widened while preserving the ARP property, pairs of subsequent BTR allocations can be switched
- A pressing need: a dynamic covariate-adaptive allocation with inconvenient allocation ratio in small cohorts
  - To be used in response-adaptive cohort allocation or platform trials with frequent changes in allocation ratio
- A note: when an EQUAL allocation procedure is NOT symmetric with respect to all arms, ARP property is not guaranteed
  - As in Selukar et al. approach for studies with differences in eligibility across treatments

When no ARP Procedure That Meets the Study Needs is Available …

- Higher magnitude of the variations in the unconditional allocation ratio leads to higher potential biases and higher impact on the randomization test power
  - For a non-ARP procedure, higher degree of forcing towards the target AR leads to higher magnitude of variations in the unconditional allocation ratio
  - For example, with the BCM by Han et al. [2011], lower bias of the biased coin (0.7 vs 0.9) leads to lower magnitude of variations
  - Lower forcing, however, leads to higher deviations from the target ratio and potential accidental bias associated with the time trend – evaluate the trade off
- Take steps to minimize selection and evaluation bias
  - Do not reveal the randomization procedure to the investigators
  - Conceal the order of enrollment through scrambled subject IDs
- Accidental bias confounded with the variations: real danger mostly comes from stratification by study center
  - Avoid stratification by study center

Conclusions

- Wider use of the multi-arm trials, in particular umbrella and platform trials, brings attention to new unequal allocation needs
- With novel unequal allocation procedures designed and implemented, ARP property is a desired one
- When no ARP allocation procedure that meets the study needs is available, take steps to minimize potential biases
- Further development of ARP procedures is a relevant task

THANK YOU FOR YOUR ATTENTION!

References

- Rosenberger WF, Lachin JM. Randomization in Clinical Trials: Theory and Practice, 2nd Edition. John Wiley & Sons, New York, 2015.
- Proschan M, Brittain E, and Kammerman L. Minimize the use of minimization with unequal allocation. *Biometrics* 2011; **67** (3): 1135-1141, DOI: 10.1111/j.1541-0420.2010.01545.x.
- Efron B. Forcing a sequential experiment to be balanced. Biometrika 1971; **58**: 403–417.
- Zelen M. The randomization and stratification of patients to clinical trials. *Journal of Chronic Disease* 1974; **27:** 365-375.
- Kuznetsova OM. Considerations in the paper by Proschan, Brittain, and Kammerman are not an argument against minimization. In response to Vance W Berger 'Minimization: not all it's cracked up to be', Clin Trials 2011; 8: 443. Clinical Trials 2012; 9: 370
- Kuznetsova O, Tymofyeyev Y. Brick Tunnel Randomization for unequal allocation to two or more treatment groups. *Statistics in Medicine* 2011a; **30:** 812-824 (2011a). DOI: 10.1002/sim.4167.
- Kuznetsova OM, Tymofyeyev Y. Expansion of the modified Zelen's approach randomization and dynamic randomization with partial block supplies at the centers to unequal allocation. *Contemporary Clinical Trials* 2011c; **32**: 962–972. DOI:10.1016/j.cct.2011.08.006.
- Kuznetsova OM, Tymofyeyev Y. Preserving the allocation ratio at every allocation with biased coin randomization and minimization in studies with unequal allocation. *Statistics in Medicine* 2012; **31**: 701-723, DOI: 10.1002/sim.4447.
- Zhao W, Carter K, Sverdlov O, Scheffold A, Ryeznik Y, Cassarly C, Berger VW. Steady-state statistical properties and implementation of randomization designs with maximum tolerated imbalance restriction for two-arm equal allocation clinical trials. Stat Med. 2024 Mar 15;43(6):1194-1212. doi: 10.1002/sim.10013. Epub 2024 Jan 20. PMID: 38243729; PMCID: PMC10925840. https://doi.org/10.1002/sim.10013

References

- Kuznetsova OM, Tymofyeyev Y. Wide Brick Tunnel Randomization – an unequal allocation procedure that limits the imbalance in treatment totals. *Statistics in Medicine* 2014; 33: 1514–1530, DOI: 10.1002/sim.6051.
- Kuznetsova OM, Tymofyeyev Y. Hierarchical Dynamic Allocation Procedures Based on Modified Zelen's Approach in Multi-Regional Studies With Unequal Allocation. *Journal of Biopharmaceutical Statistics* 2014, 24: 1-17.
- Kuznetsova OM. Randomization challenges in adaptive design studies. In: He W, Pinheiro J, and Kuznetsova OM (eds.). Practical Considerations for Adaptive Trial Design and Implementation: pp. 157–181. Springer, New York (2014).
- Kuznetsova OM. Brick Tunnel Randomization and the Momentum of the Probability Mass. Statistics in Medicine 2015, 34(30), p.4031-4056. DOI: 10.1002/sim.6601
- Kuznetsova OM, Tymofyeyev Y. Covariate-Adaptive Procedures with Unequal Allocation. In: Sverdlov A. (ed.). Modern Adaptive Randomized Clinical Trials: Statistical, Operational, and Regulatory Aspects: pp. 171-198. CRC Press, Taylor & Francis Group (2015)
- Kuznetsova OM, Tymofyeyev Y. Brick Tunnel and Wide Brick Tunnel Randomization for Studies with Unequal Allocation. In: Sverdlov A. (ed.). Modern Adaptive Randomized Clinical Trials: Statistical, Operational, and Regulatory Aspects: pp. 83-114. CRC Press, Taylor & Francis Group (2015)
- Kuznetsova OM and Victoria Plamadeala Johnson (2017). Approaches to expanding the two-arm biased coin randomization to unequal allocation while preserving the unconditional allocation ratio. Statistics in Medicine 36(16), 2483–2498. DOI: 10.1002/sim.7290

References

- Kaiser LD. Dynamic randomization and a randomization model for clinical trials data. *Statistics in Medicine*. 2012; **31**:3858-3873. DOI: 10.1002/sim.5448.
- Pocock SJ. Allocation of Patients to Treatment in Clinical Trials. *Biometrics* 1979; **35**: 183-197.
- Abel U. Modified replacement randomization. *Statistics in Medicine* 1987; **6**: 127–135.
- Berger VW, Ivanova A, Knoll M. Minimizing Predictability while Retaining Balance through the Use of less Restrictive Randomization Procedures. *Statistics in Medicine* 2003; **22**, 3017-3028. DOI:10.1002/sim.1538
- Soares JF, Wu CF. Some restricted randomization rules in sequential designs. *Communications in Statistics - Theory and Methods* 1983;**12**:2017–2034.
- Chen YP. Biased coin design with imbalance tolerance. *Communications in Statistics. Stochastic Model* 1999;**15**:953–975.
- Chen YP. Which design is better? Ehrenfest urn versus biased coin. *Advances in Applied Probability* 2000; **32**:738–749.

References

- Baldi Antognini A. and Giovagnoli A. A new “biased coin design” for the sequential allocation of two treatments. *Journal of the Royal Statistical Society C*, 2004; **53**:651-664
- Morrissey M, McEntegart D, Lang M. Randomisation in double-blind multicentre trials with many treatments. *Contemporary Clinical Trials* 2010; **31**: 381-391. DOI: 10.1016./j/cct/2010.05.002
- Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975; 31:103–115.
- Salama I, Ivanova A, Qaqish B. Efficient generation of constrained block allocation sequences. *Statistics in Medicine* 2008; **27**: 1421-1428, DOI:10.1002/sim3014.
- Han B, Enas NH, McEntegart D. Randomization by minimization for unbalanced treatment allocation. *Statistics in Medicine* 2009; **28:** 3329-3346, DOI:10.1002/sim.3710
- Han B, Yu M, McEntegrt D. Weighted re-randomization tests for minimization with unbalanced allocation. *Pharmaceut. Statist*. 2013, 12 243–253
- Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics* 2004; **32(1)**, 268-301.

References

- Sverdlov O, Zhang Y. (2014) Statistical properties of covariate-adaptive randomization procedures for trials with unequal treatment allocation ratios. ASA Proceedings of the Joint Statistical Meetings, American Statistical Association, Boston, MA, 2014.
- Titterington DM. On constrained balance randomization for clinical trials. *Biometrics* 1983; **39(4):** 1083-1086 ()
- Frane JW. A method of biased coin randomisation, its implementation and its validation. *Drug Information Journal* 1998; **32**: 423-432.
- Russell D, Hoare ZSJ, Whitaker Rh, Whitaker CJ, Russell IT. Generalized method for adaptive randomization in clinical trials. *Statistics in Medicine* 2011; **30**: 922-934.
- Lebowitsch J, Ge Y, Young B, Hu F. Generalized multidimensional allocation method, *Statistics in Medicine* 2012; **31**: 3537-3544.
- Ivanova A. A play-the-winner-type urn design with reduced variability. *Metrika* 2003; **58**: 1-13.
- Zhao W, Weng Y. Block urn design — A new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemporary Clinical Trials*, 2011, **32**, 6, 953-961.
- Ye, T., and Shao, J. (2020), "Robust Tests for Treatment Effect in Survival Analysis under Covariate-Adaptive Randomization," *Journal of the Royal Statistical Society*, *Series* B, 82, 1301–1323.
- Sverdlov O, Ryeznik Y. Implementing unequal randomization in clinical trials with heterogeneous treatment costs. *Statistics in Medicine* 2019; **38**: 2905-2927

Randomization Designs for Bayesian Adaptive Trials

Wenle Zhao

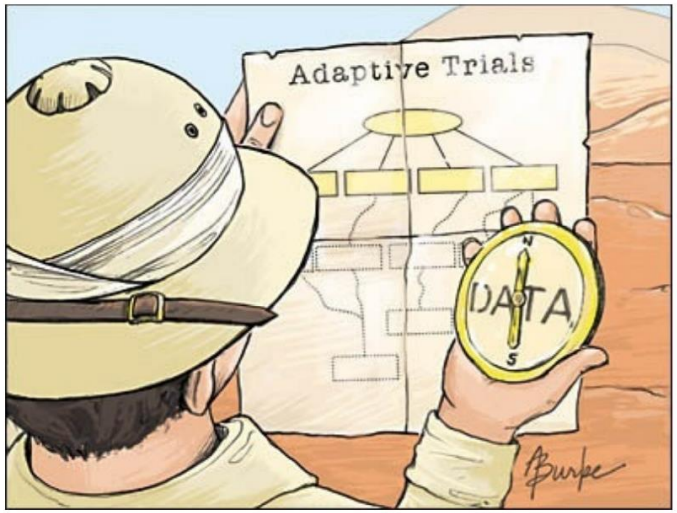
Medical University of South Carolina, Charleston, SC, USA

SCT 2024 Annual Meeting, Boston, MA
May 20-22, 2024

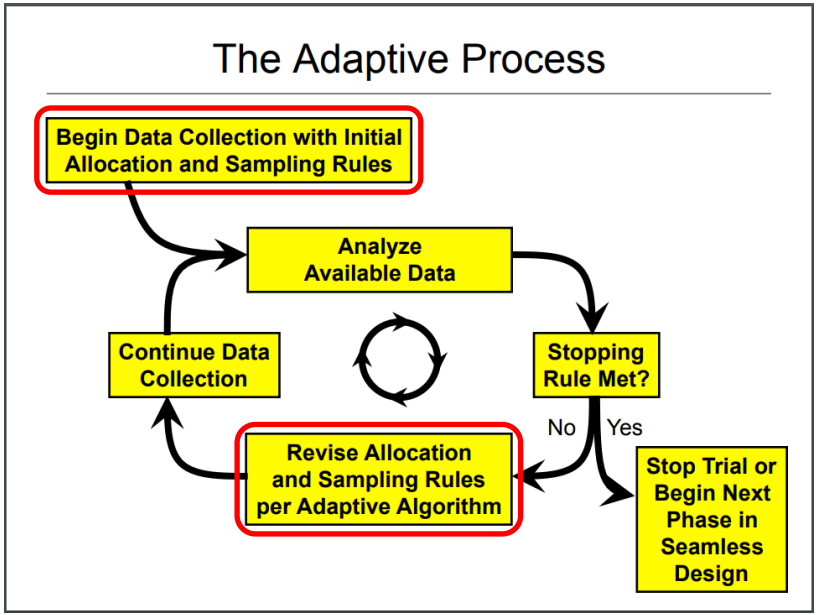
Disclosure

I have NO financial disclosure or conflicts of interest with the material in this presentation.

An Overview of Bayesian Adaptive Design from Berry Consultants



JAMA 2006;296:1955-1957.



Frequent “looks” at the data and data-driven modification of the trial.

Source: https://www.berryconsultants.com/wp-content/uploads/2012/09/An-Overview-of-Bayesian-Adaptive-Clinical-Trial-Design.pdf

The ESETT Design Paper

Title: Established Status Epilepticus Treatment Trial (ESETT)
ClinicalTrials.gov ID NCT05376267

This is a comparative effectiveness study of three medications in the emergency treatment of patients with benzodiazepine-refractory status epilepticus.

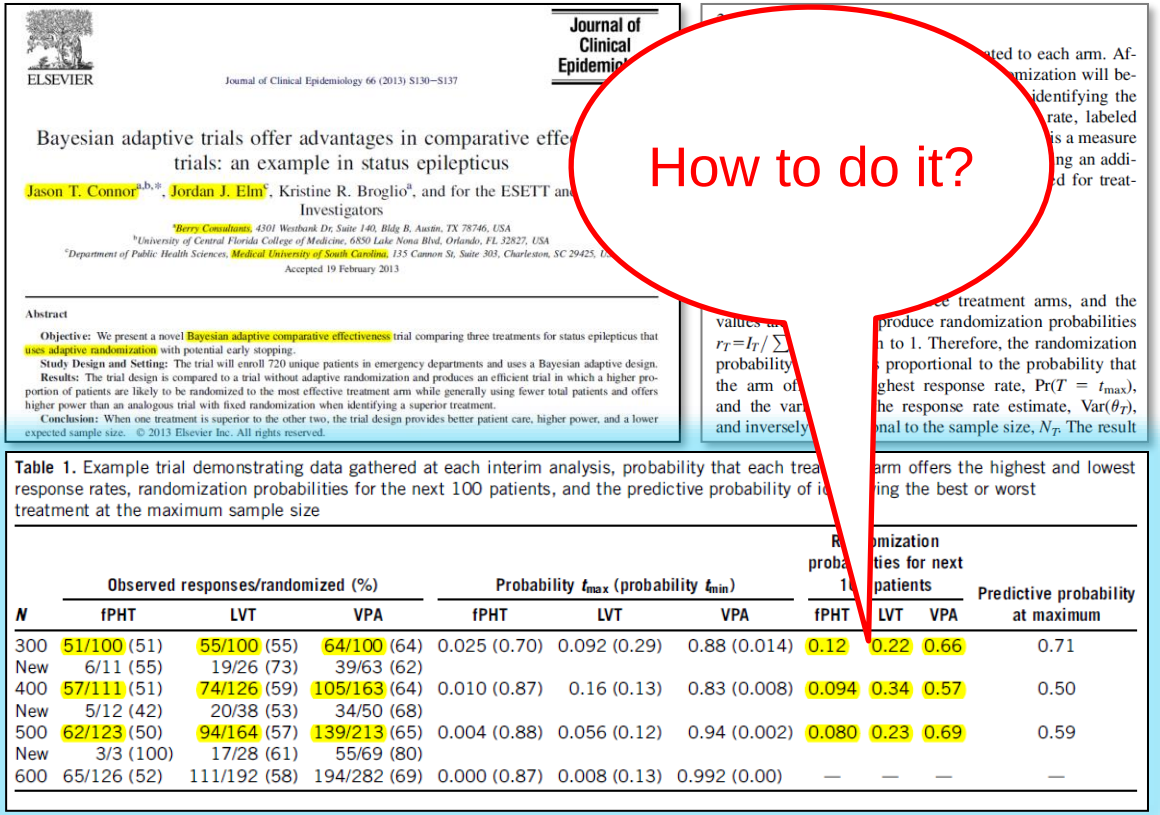
Funded by NIH and operated in the SIREN network.

Interventional Model Description: Bayesian Adaptive Design

Number of arms: 3

Sample size: up to 795

Allocation: Equal allocation among 3 arms for the first 150 subjects, response adaptive allocation updated every 100 subjects



ELSEVIER

Journal of Clinical Epidemiology 66 (2013) S130–S137

Journal of Clinical Epidemiol

Bayesian adaptive trials offer advantages in comparative effe
trials: an example in status epilepticus

Jason T. Connor[a,b,*], Jordan J. Elm[c], Kristine R. Broglio[a], and for the ESETT an
Investigators

[a]Berry Consultants, 4301 Westbank Dr, Suite 140, Bldg B, Austin, TX 78746, USA
[b]University of Central Florida College of Medicine, 6850 Lake Nona Blvd, Orlando, FL 32827, USA
[c]Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon St, Suite 303, Charleston, SC 29425, U

Accepted 19 February 2013

Abstract

Objective: We present a novel Bayesian adaptive comparative effectiveness trial comparing three treatments for status epilepticus that uses adaptive randomization with potential early stopping.

Study Design and Setting: The trial will enroll 720 unique patients in emergency departments and uses a Bayesian adaptive design.

Results: The trial design is compared to a trial without adaptive randomization and produces an efficient trial in which a higher proportion of patients are likely to be randomized to the most effective treatment arm while generally using fewer total patients and offers higher power than an analogous trial with fixed randomization when identifying a superior treatment.

Conclusion: When one treatment is superior to the other two, the trial design provides better patient care, higher power, and a lower expected sample size. © 2013 Elsevier Inc. All rights reserved.

How to do it?

ated to each arm. Af-
mization will be-
identifying the
rate, labeled
is a measure
ing an addi-
d for treat-

ce treatment arms, and the
produce randomization probabilities
to 1. Therefore, the randomization
proportional to the probability that
ghest response rate, $\Pr(T = t_{max})$,
he response rate estimate, $\mathrm{Var}(\theta_T)$,
onal to the sample size, N_T. The result

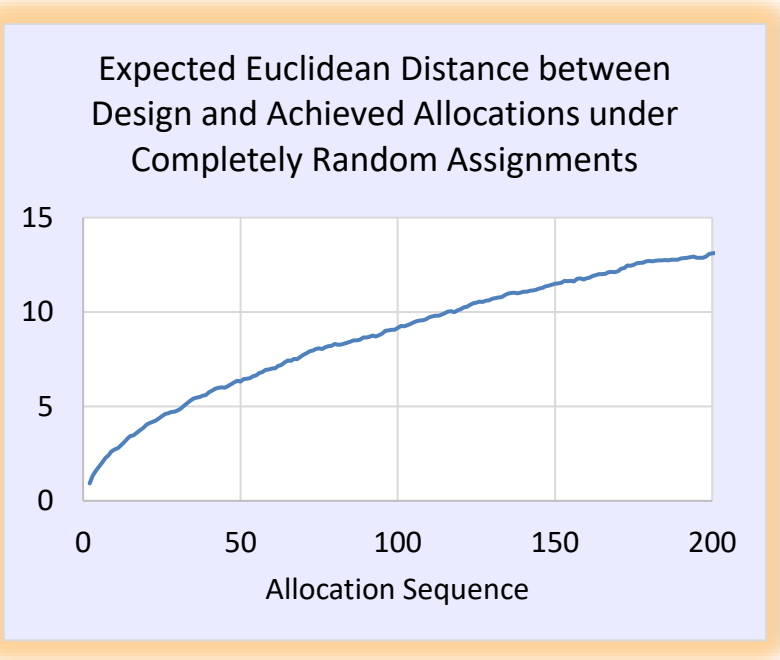
Table 1. Example trial demonstrating data gathered at each interim analysis, probability that each trea arm offers the highest and lowest response rates, randomization probabilities for the next 100 patients, and the predictive probability of i ing the best or worst treatment at the maximum sample size

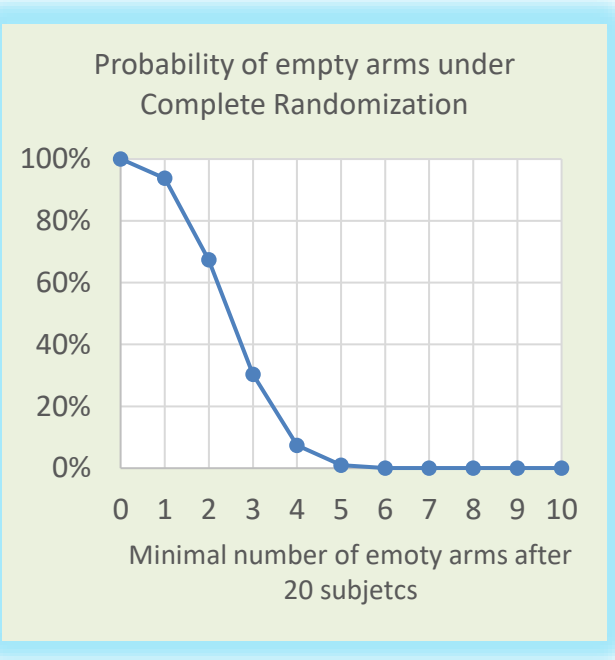
	Observed responses/randomized (%)			Probability t_{max} (probability t_{min})			R omization probabilities for next 1 patients			Predictive probability at maximum
N	fPHT	LVT	VPA	fPHT	LVT	VPA	fPHT	LVT	VPA	
300	51/100 (51)	55/100 (55)	64/100 (64)	0.025 (0.70)	0.092 (0.29)	0.88 (0.014)	0.12	0.22	0.66	0.71
New	6/11 (55)	19/26 (73)	39/63 (62)							
400	57/111 (51)	74/126 (59)	105/163 (64)	0.010 (0.87)	0.16 (0.13)	0.83 (0.008)	0.094	0.34	0.57	0.50
New	5/12 (42)	20/38 (53)	34/50 (68)							
500	62/123 (50)	94/164 (57)	139/213 (65)	0.004 (0.88)	0.056 (0.12)	0.94 (0.002)	0.080	0.23	0.69	0.59
New	3/3 (100)	17/28 (61)	55/69 (80)							
600	65/126 (52)	111/192 (58)	194/282 (69)	0.000 (0.87)	0.008 (0.13)	0.992 (0.00)	—	—	—	—

Source: Connor JT, Elm JJ, Broglio KR; ESETT and ADAPT-IT Investigators. Bayesian adaptive trials offer advantages in comparative effectiveness trials: an example in status epilepticus. J Clin Epidemiol. 2013 Aug;66(8 Suppl):S130-7. doi: 10.1016/j.jclinepi.2013.02.015. PMID: 23849147; PMCID: PMC3743558.

Option #1: Complete Randomization

Design Allocation	
Arm	Allocation Probability
1	0.25
2	0.15
3	0.13
4	0.10
5	0.09
6	0.08
7	0.05
8	0.05
9	0.05
10	0.05

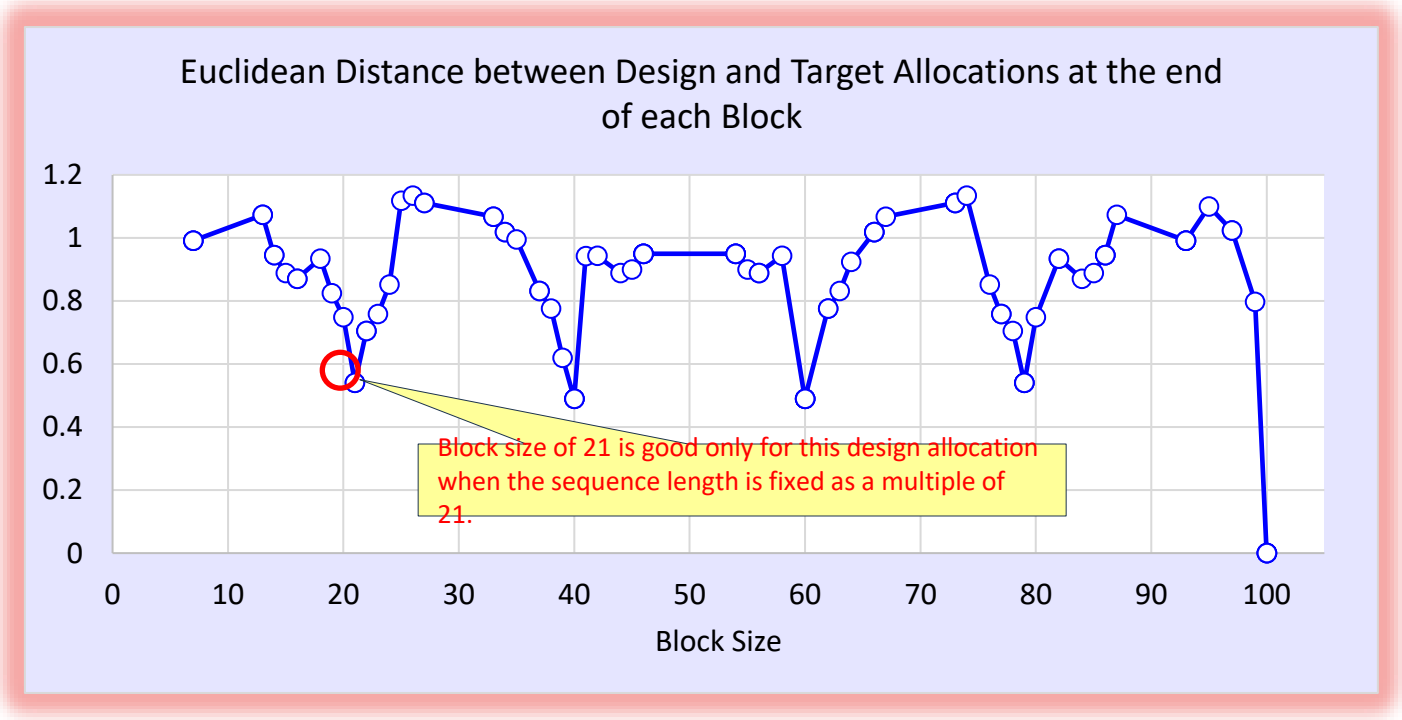




The expected treatment imbalance monotonically increases as allocation length increases.
For a short allocation sequence length, it has high chance seeing empty arms

Option #2: Permuted Block Design

Design Allocation	
Arm	Allocation Probability
1	0.25
2	0.15
3	0.13
4	0.10
5	0.09
6	0.08
7	0.05
8	0.05
9	0.05
10	0.05



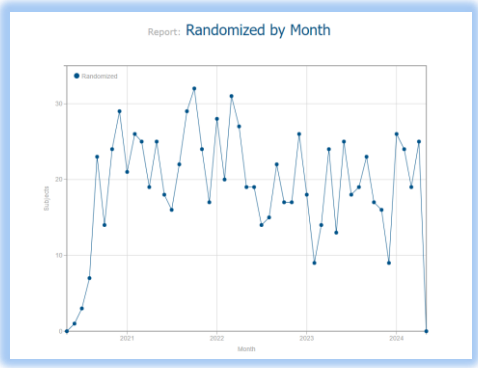
It requires a large block size to accurately target the design allocation.
The large block size may result in large treatment imbalance within the block.

Allocation Update Schedule in Bayesian Adaptive Trials

Option #1: by calendar

Pros: Good for implementing allocation update, which may take 1 or 2 days.

Cons: Allocation sequence lengths may vary significantly, making allocation end balancing difficult.



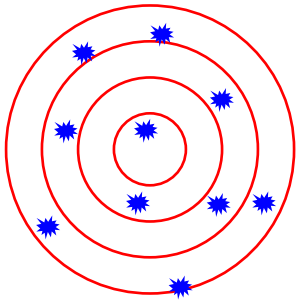
Option #2: by enrollment

Pros: Allow to select a block size fit the allocation length and to target the design allocation closely.

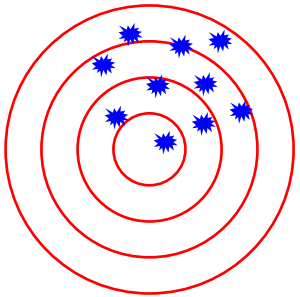
Cons: Difficult for implementation, especially when enrollment is fast.

Design Allocation		Target Allocation
Arm	Allocation Probability	Block of 21
1	0.25	**5**
2	0.15	**3**
3	0.13	2.6 → **3**
4	0.10	**2**
5	0.09	1.8 → **2**
6	0.08	1.6 → **2**
7	0.05	**1**
8	0.05	**1**
9	0.05	**1**
10	0.05	**1**

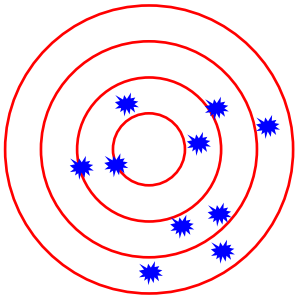
Allocation Accuracy and Precision



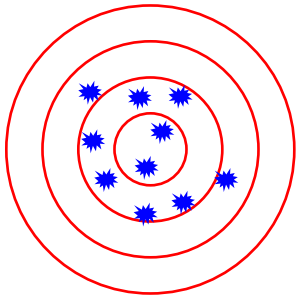
Complete Random
Highest accuracy
Lowest precision



PBR with small block size
Lower accuracy
Higher Precision



PBR with large block size
Higher accuracy
Lower Precision



New design we need
Highest accuracy
Higher Precision

Mass-weighted Urn Design

Contemporary Clinical Trials 43 (2015) 209–216

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Mass weighted urn design — A new randomization algorithm for unequal allocations

Wenle Zhao

Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon Street, Suite 305H, Charleston, SC 29425, USA

ARTICLE INFO

Article history:
Received 18 April 2015
Received in revised form 8 June 2015
Accepted 13 June 2015
Available online 17 June 2015

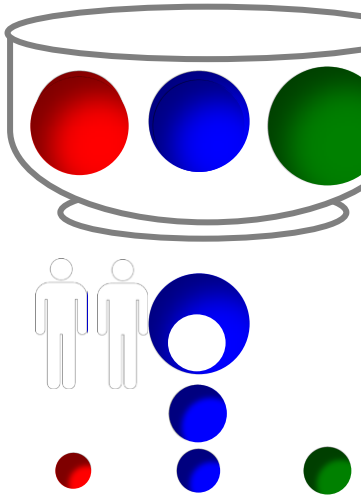
Keywords:
Mass weighted urn design
Unequal allocation
Treatment imbalance
Allocation predictability

ABSTRACT

Unequal allocations have been used in clinical trials motivated by ethical, efficiency, or feasibility concerns. Commonly used permuted block randomization faces a tradeoff between effective imbalance control with a small block size and accurate allocation target with a large block size. Few other unequal allocation randomization designs have been proposed in literature with applications in real trials hardly ever been reported, partly due to their complexity in implementation compared to the permuted block randomization. Proposed in this paper is the mass weighted urn design, in which the number of balls in the urn equals to the number of treatments, and remains unchanged during the study. The chance a ball being randomly selected is proportional to the mass of the ball. After each treatment assignment, a part of the mass of the selected ball is re-distributed to all balls based on the target allocation ratio. This design allows any desired optimal unequal allocations be accurately targeted without approximation, and provides a consistent imbalance control throughout the allocation sequence. The statistical properties of this new design is evaluated with the Euclidean distance between the observed treatment distribution and the desired treatment distribution as the treatment imbalance measure; and the Euclidean distance between the conditional allocation probability and the target allocation probability as the allocation predictability measure. Computer simulation results are presented comparing the mass weighted urn design with other randomization designs currently available for unequal allocations.

Operation Model of Mass-weighted Urn Design

Target Allocation **1 : √2 : √3 = 0.2412 : 0.3411 : 0.4177**



1. The urn starts with one ball for each arm. Mass of each ball is proportional to the target allocation ratio, in continuous format. Total mass in the urn = α, a pre-specified value.
2. Randomly draw a ball with probability proportional to the mass of each ball.
3. Assign the subject accordingly.
4. The mass of the ball is reduced by 1 unit.
5. This 1 unit mass is re-distributed to each arm based on the target allocation ratio.
6. Repeat steps 3 to 5 until the end of the study.

Conditional Allocation Probability of Mass-weighted Urn Design

$\mathbf{W} = (w_1, w_2, \cdots, w_m)$, $\sum_{j=1}^{m} w_j = 1$ as the target allocation

$\mathbf{T} = (T_1, T_2, \cdots)$ as treatment allocation sequence.

$\mathbf{N}_i = (n_{i1}, n_{i2}, \cdots, n_{im})$ as achieved allocation after T_i.

$\mathbf{X}_i = (x_{i1}, x_{i2}, \cdots, x_{im})$, $\sum_{j=1}^{m} x_{ij} = \alpha$ as mass for the balls in the urn after T_i.

$\mathbf{P}_i = (p_{i1}, p_{i2}, \cdots, p_{im})$ as conditional allocation probability for T_i.

$$x_{ij} = \alpha w_j + (i - n_{ij}) w_j - (1 - w_j) n_{ij} = \alpha w_j - n_{ij} + i w_j$$

$$p_{i+1,j} = w_j - \frac{n_{ij} - i w_j}{\alpha} \qquad (i = 0, 1, 2, \cdots; j = 1, 2, \cdots, m)$$

Conditional Allocation Probability = Target Allocation Probability $- \frac{1}{\alpha}$ (Observed Treatment Group Size $-$ Expected Treatment Group Size)

Implementation Example of Mass-weighted Urn Design

Title: **Pediatric Influence of Cooling Duration on Efficacy in Cardiac Arrest Patients (P-ICECAP)**
ClinicalTrials.gov ID NCT05376267

This is a multicenter trial to establish the efficacy of cooling and the optimal duration of induced hypothermia for neuroprotection in pediatric comatose survivors of cardiac arrest.

Funded by NIH and operated in the SIREN network.

Interventional Model Description: Bayesian Adaptive Design

Number of arms: 10

Sample size: 900

Allocation: Equal allocation among 3 arms for the first 150 subjects

Bayesian response adaptive allocation among 10 arms updated every 10 weeks

Mass-weighted Urn Design with α = 6 is used.

$$p_{i+1,j} = w_j - \frac{n_{i,j} - iw_j}{\alpha} \qquad (i = 0,1,\cdots; j = 1,2,\cdots,10)$$

Sources: https://clinicaltrials.gov/study/NCT05376267

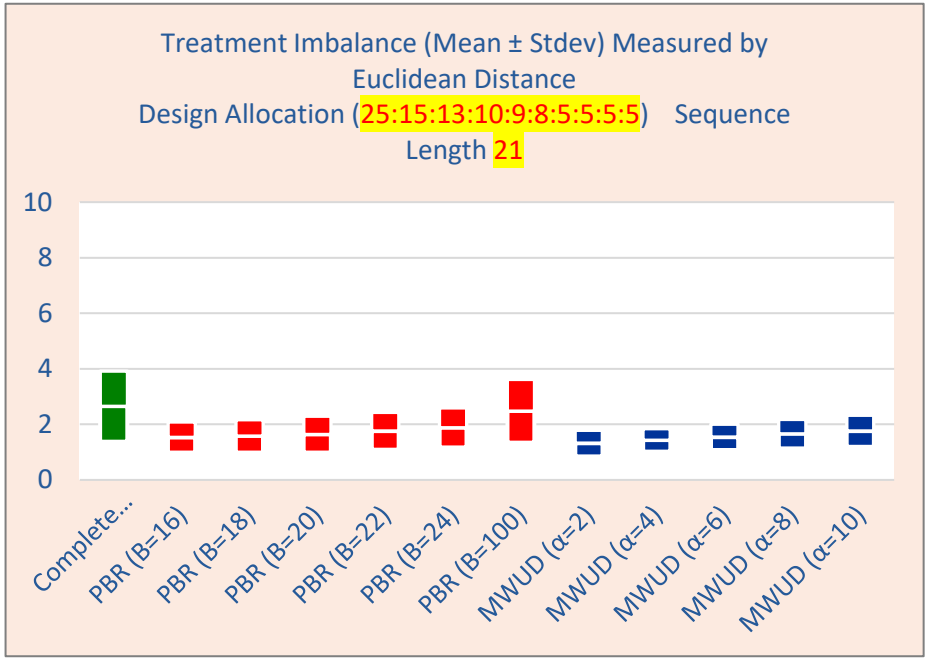
Comparison of the Three Randomization Designs

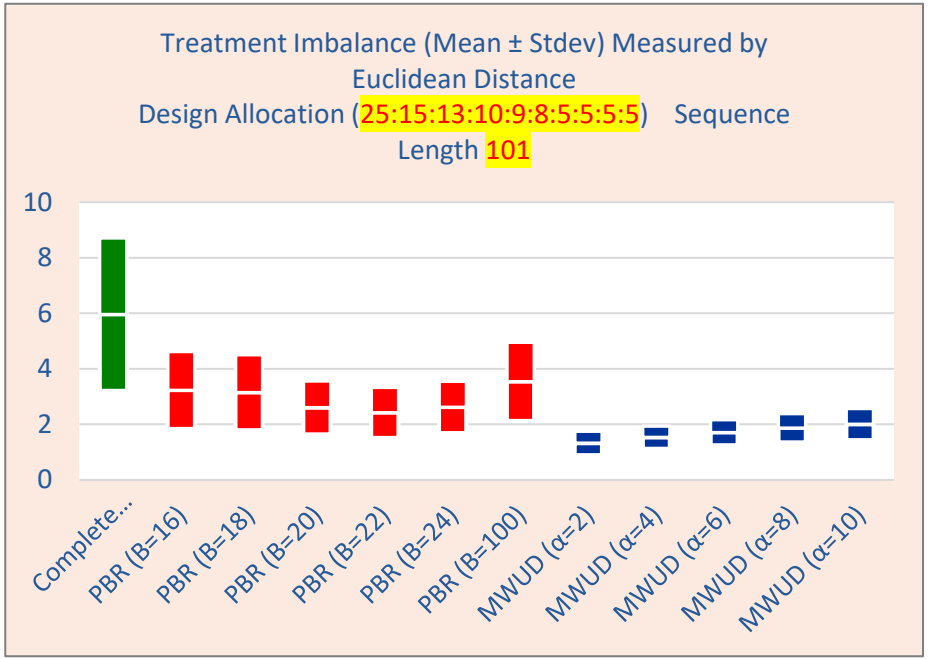
Design allocation = (25:15:13:10:9:8:5:5:5:5) Allocation Sequence Length = 21 Simulation = 2380						
	Parameter	Allocation Accuracy	Treatment Imbalance	Empty Arms	Allocation Randomness	Deterministic Assignments
Complete randomization		0	2.65 ± 1.27	1.98	0	0
Permuted Block Design	B=16 (4:2:2:2:1:1:1:1:1:1)	0.054	1.53 ± 0.55	0	0.24	5.2%
	B=18 (4:3:2:2:2:1:1:1:1:1)	0.052	1.58 ± 0.59	0	0.24	5.2%
	B=20 (5:3:2:2:2:2:1:1:1:1)	0.037	1.64 ± 0.65	0	0.25	5.2%
	B=22 (6:3:3:2:2:2:1:1:1:1)	0.032	1.76 ± 0.67	0.38	0.22	0.6%
	B=24 (6:4:3:3:2:2:1:1:1:1)	0.036	1.88 ± 0.71	0.70	0.18	0
	B=100 (25:15:13:10:9:8:5:5:5:5)	0	2.48 ± 1.14	1.72	0.03	0
Mass-weighted Urn Design	α=2	0	1.32 ± 0.43	0.05	0.31	0
	α=4	0	1.43 ± 0.47	0.15	0.25	0
	α=6	0	1.54 ± 0.53	0.32	0.21	0
	α=8	0	1.66 ± 0.59	0.47	0.19	0
	α=10	0	1.76 ± 0.64	0.59	0.16	0

Comparison of the Three Randomization Designs

Design allocation = (25:15:13:10:9:8:5:5:5:5) Allocation Sequence Length = 101 Simulation = 495						
	Parameter	Allocation Accuracy	Treatment Imbalance	Empty Arms	Allocation Randomness	Deterministic Assignments
Complete randomization		0	5.96 ± 2.75	0.03	0	0
Permuted Block Design	B=16 (4:2:2:2:1:1:1:1:1:1)	0.054	3.23 ± 1.39	0	0.28	6.5%
	B=18 (4:3:2:2:2:1:1:1:1:1)	0.052	3.15 ± 1.36	0	0.26	5.3%
	B=20 (5:3:2:2:2:2:1:1:1:1)	0.037	2.60 ± 0.96	0	0.26	5.5%
	B=22 (6:3:3:2:2:2:1:1:1:1)	0.032	2.42 ± 0.91	0	0.23	4.5%
	B=24 (6:4:3:3:2:2:1:1:1:1)	0.036	2.62 ± 0.93	0	0.23	4.4%
	B=100 (25:15:13:10:9:8:5:5:5:5)	0	3.54 ± 1.42	0	0.13	1.1%
Mass-weighted Urn Design	α=2	0	1.32 ± 0.43	0	0.33	0
	α=4	0	1.53 ± 0.41	0	0.27	0
	α=6	0	1.71 ± 0.46	0	0.24	0
	α=8	0	1.87 ± 0.52	0	0.21	0
	α=10	0	2.00 ± 0.57	0	0.19	0

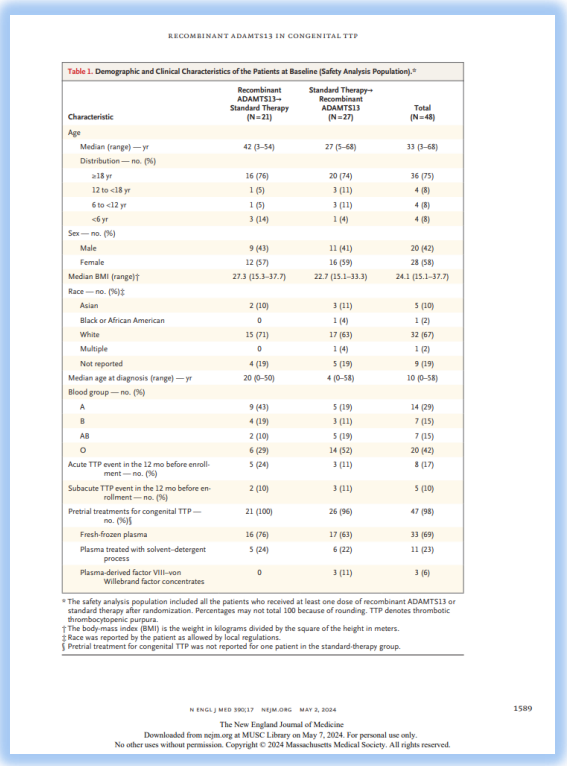
Comparison of the Three Randomization Designs





Mass-weighted Urn Design achieves the multi-arm unequal allocation with high accuracy.

Baseline Covariate Imbalance Control



RECOMBINANT ADAMTS13 IN CONGENITAL TTP

Table 1. Demographic and Clinical Characteristics of the Patients at Baseline (Safety Analysis Population).*

Characteristic	Recombinant ADAMTS13→ Standard Therapy (N=21)	Standard Therapy→ Recombinant ADAMTS13 (N=27)	Total (N=48)
Age			
Median (range) — yr	42 (3–54)	27 (5–68)	33 (3–68)
Distribution — no. (%)			
≥18 yr	16 (76)	20 (74)	36 (75)
12 to <18 yr	1 (5)	3 (11)	4 (8)
6 to <12 yr	1 (5)	3 (11)	4 (8)
<6 yr	3 (14)	1 (4)	4 (8)
Sex — no. (%)			
Male	9 (43)	11 (41)	20 (42)
Female	12 (57)	16 (59)	28 (58)
Median BMI (range)†	27.3 (15.3–37.7)	22.7 (15.1–33.3)	24.1 (15.1–37.7)
Race — no. (%)‡			
Asian	2 (10)	3 (11)	5 (10)
Black or African American	0	1 (4)	1 (2)
White	15 (71)	17 (63)	32 (67)
Multiple	0	1 (4)	1 (2)
Not reported	4 (19)	5 (19)	9 (19)
Median age at diagnosis (range) — yr	20 (0–50)	4 (0–58)	10 (0–58)
Blood group — no. (%)			
A	9 (43)	5 (19)	14 (29)
B	4 (19)	3 (11)	7 (15)
AB	2 (10)	5 (19)	7 (15)
O	6 (29)	14 (52)	20 (42)
Acute TTP event in the 12 mo before enrollment — no. (%)	5 (24)	3 (11)	8 (17)
Subacute TTP event in the 12 mo before enrollment — no. (%)	2 (10)	3 (11)	5 (10)
Pretrial treatments for congenital TTP — no. (%)§	21 (100)	26 (96)	47 (98)
Fresh-frozen plasma	16 (76)	17 (63)	33 (69)
Plasma treated with solvent–detergent process	5 (24)	6 (22)	11 (23)
Plasma-derived factor VIII–von Willebrand factor concentrates	0	3 (11)	3 (6)

* The safety analysis population included all the patients who received at least one dose of recombinant ADAMTS13 or standard therapy after randomization. Percentages may not total 100 because of rounding. TTP denotes thrombotic thrombocytopenic purpura.
† The body-mass index (BMI) is the weight in kilograms divided by the square of the height in meters.
‡ Race was reported by the patient as allowed by local regulations.
§ Pretrial treatment for congenital TTP was not reported for one patient in the standard-therapy group.

- A balanced baseline among treatment groups is desired by investigators.
- Stratified Permuted Block Randomization and Minimization are the two methods most widely used for covariate imbalance control.
- Bayesian adaptive trials often do not ask covariate imbalance control to focus on RAR implementation.

Minimal Sufficient Balance

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Statistical Methods in Medical Research
2015, Vol. 24(6) 989–1002

DOI: 10.1177/0962280212436447
smm.sagepub.com
SAGE

Minimal sufficient balance—a new strategy to balance baseline covariates and preserve randomness of treatment allocation

Wenle Zhao,[1] Michael D Hill[2] and Yuko Palesch[1]

Abstract

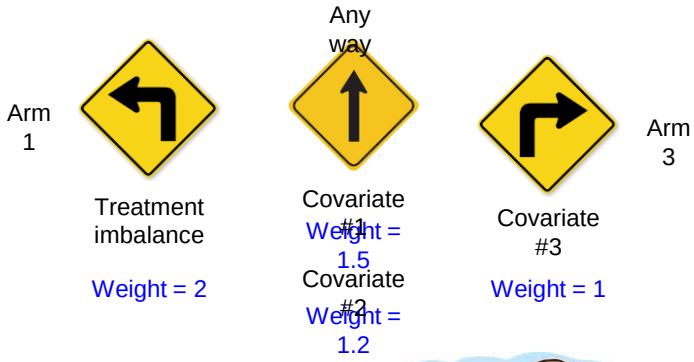
In many clinical trials, baseline covariates could affect the primary outcome. Commonly used strategies to balance baseline covariates include stratified constrained randomization and minimization. Stratification is limited to few categorical covariates. Minimization lacks the randomness of treatment allocation. Both apply only to categorical covariates. As a result, serious imbalances could occur in important baseline covariates not included in the randomization algorithm. Furthermore, randomness of treatment allocation could be significantly compromised because of the high proportion of deterministic assignments associated with stratified block randomization and minimization, potentially resulting in selection bias. Serious baseline covariate imbalances and selection biases often contribute to controversial interpretation of the trial results. The National Institute of Neurological Disorders and Stroke recombinant tissue plasminogen activator Stroke Trial and the Captopril Prevention Project are two examples. In this article, we propose a new randomization strategy, termed the minimal sufficient balance randomization, which will dually prevent serious imbalances in all important baseline covariates, including both categorical and continuous types, and preserve the randomness of treatment allocation. Computer simulations are conducted using the data from the National Institute of Neurological Disorders and Stroke recombinant tissue plasminogen activator Stroke Trial. Serious imbalances in four continuous and one categorical covariate are prevented with a small cost in treatment allocation randomness. A scenario of simultaneously balancing 11 baseline covariates is explored with similar promising results. The proposed minimal sufficient balance randomization algorithm can be easily implemented in computerized central randomization systems for large multicenter trials.

The Philosophy of Minimal Sufficient Balance

#1: Perfect balance of baseline covariates in *RCT* is neither feasible nor necessary.

#2: Measure imbalances in the way they may be challenged, such as:

- ✓ Range
- ✓ Euclidean distance
- ✓ p-value of statistical test



#3: Settle competing demands by voting.

Implementation Example of Minimal Sufficient Balance

Title: **Strategy for Improving Stroke Treatment Response (SISTER)**
ClinicalTrials.gov ID NCT05948566

A Phase II, Bayesian, adaptive, randomized, dose-finding trial of TS23 in patients with acute ischemic stroke.

Funded by NIH/NINDS and operated in the NIH StrokeNet.

Interventional Model Description: Bayesian Adaptive Design

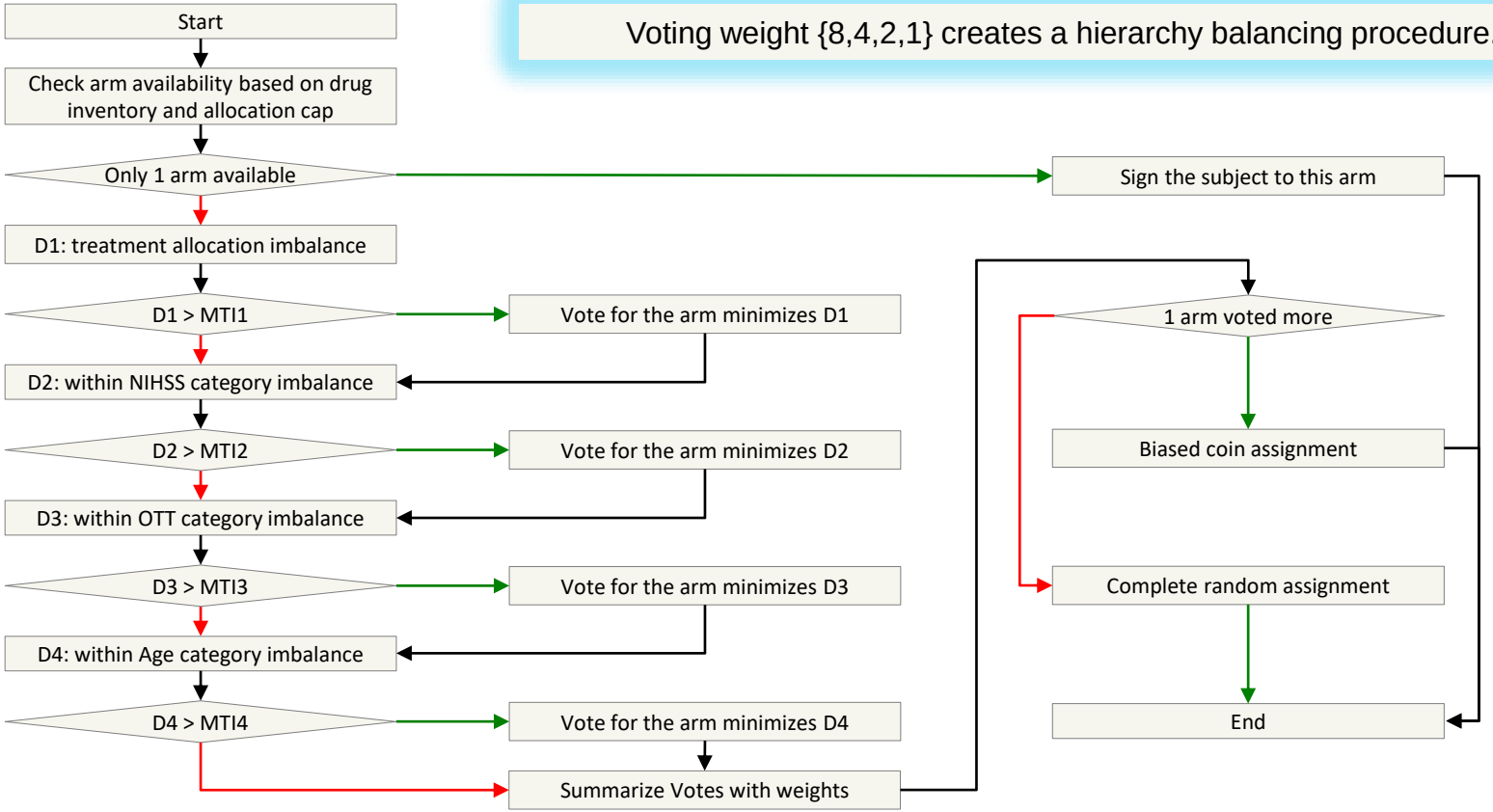
Number of arms:	5
Sample size:	300
Allocation:	Equal allocation for the first 50 subjects
	Response adaptive allocation every 50 subjects thereafter.
Baseline covariates:	1. NIHSS category (≤ 10 vs > 10)
	2. Onset to treatment time (≤ 9 hr. vs > 9 hr.)
	3. Age (< 65 yr. vs ≥ 65 yr.)

Number of arms:	5
Allocation:	Unequal
Sequence length:	50
Baseline covariates:	3 dichotomous

Sources: https://classic.clinicaltrials.gov/ct2/show/NCT05948566

Implementation Example of Minimal Sufficient Balance

Voting weight {8,4,2,1} creates a hierarchy balancing procedure.



Thank You

zhaow@musc.edu