Boehringer Ingelheim

# Thinking outside the blocks

## Moving towards fit-for-purpose randomization in our clinical trials

Dr. Johannes Krisam

*Life forward*

# Outline

1. General principles of randomization
2. Balance-Randomness Tradeoff
3. Multi-center and multi-arm trials
4. Regulatory perspective on randomization
5. Summary

# 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 tuberculosis, 1946

Jerome Cornfield (1912–1979)

“Principles of Research”, 1959

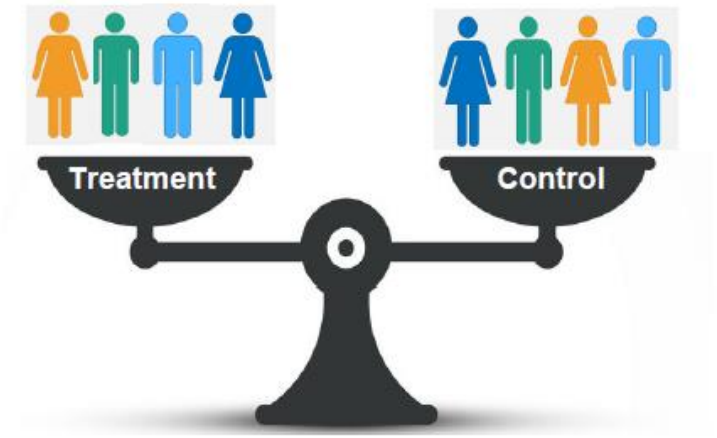
Boehringer Ingelheim

# General principles of randomization

# 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 randomization-based inference**

# What types of randomization designs are available?

**Randomization designs**

## Fixed randomization

- Randomization sequence can be generated in advance

## Adaptive randomization

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

### Allocation-adaptive

- To achieve desired treatment allocation ratio

### Covariate-adaptive

- To achieve balance on imporant 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.

Boehringer Ingelheim

# Some simplifiying assumptions

- For now, let us assume
  - ➢ Randomized parallel group two-arm controlled trial design
  - ➢ Target allocation of 1:1
  - ➢ Randomization sequence can be 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

# Balance-Randomness Tradeoff

# Let us consider three types of 1:1 randomization designs

## 1. Complete randomization (CR)

Treatment assignments are made independently, by flip of a fair coin

## 2. Permuted Block Designs

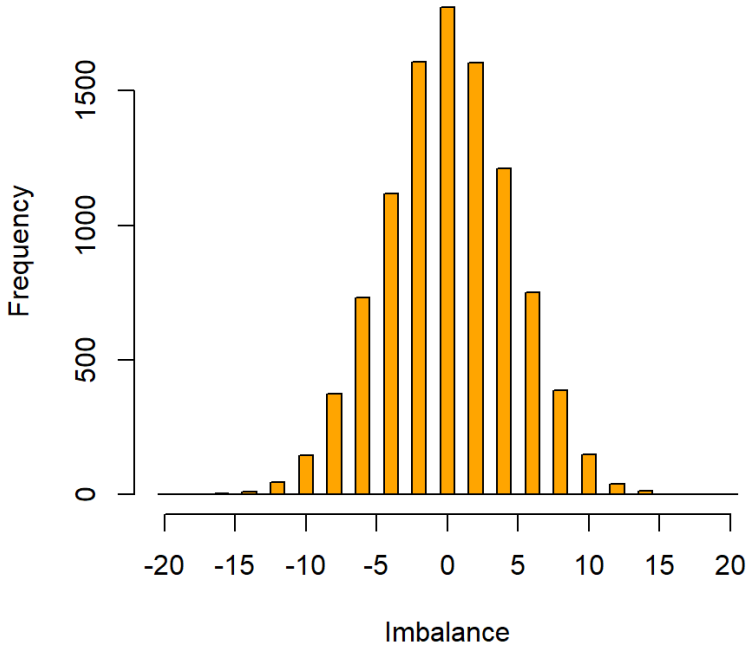
Subjects are randomized by using a sequence of permuted blocks with a prespecified block size. Treatment imbalance cannot become larger than half the block size and is zero at the end of each block

## 3. Maximum Tolerated Imbalance (MTI) procedures

Treatment imbalance is maintained within user-defined limits, but final group sizes are not necessarily equal

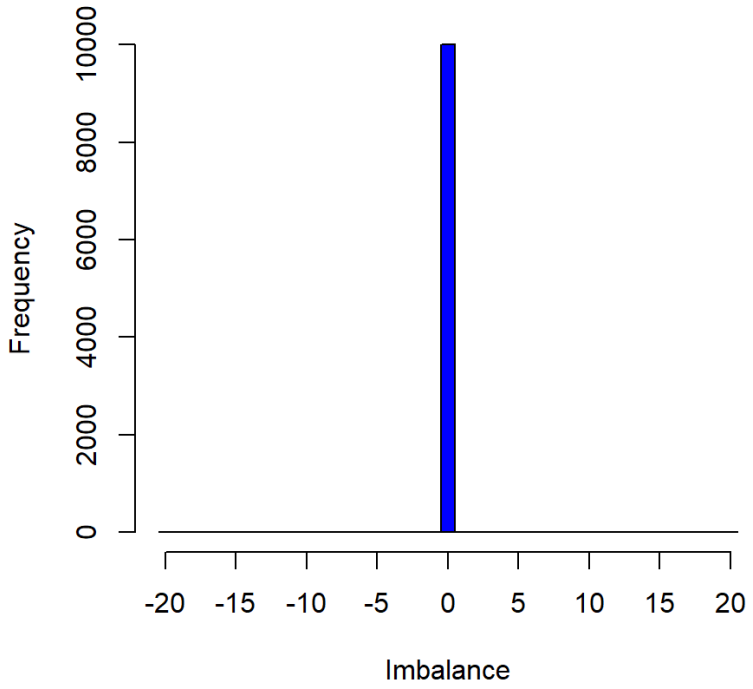
# Imbalance distribution for a RCT with N=20 patients

## Complete randomization (CR)



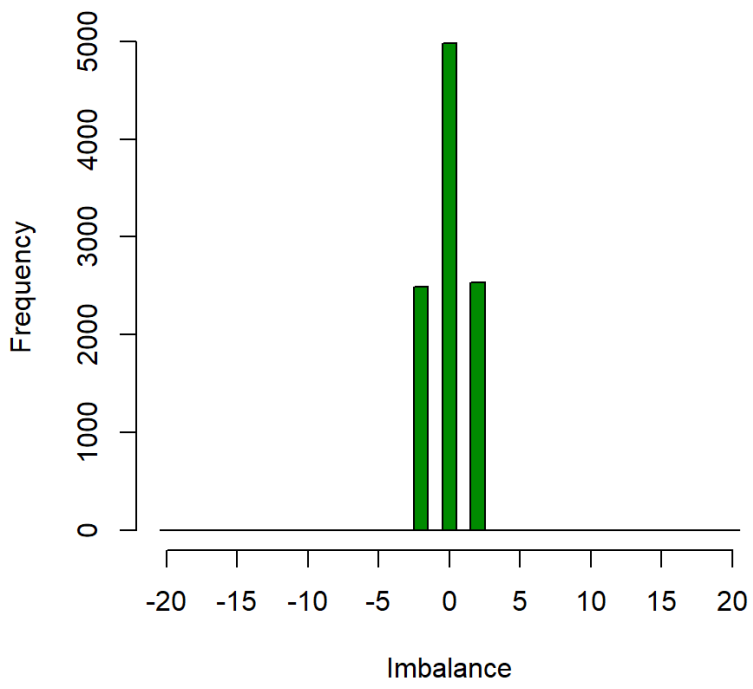

Imbalance is around zero, but variability is high

## PBD with block size 4




Imbalance is zero with probability 1 at the end of the block

## BSD with maximum tolerated imbalance of 2




Imbalance is within pre-specified margins (±2)

Boehringer Ingelheim

BSD: Big Stick Design (Soares & Wu, 1983)

# 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 with probability 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 exist:

**BSD:**

- 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***

Example: BSD with b=2

| Sequence No | Block No | Group |
|---|---|---|
| 101 | - | E |
| 102 | - | C |
| 103 | - | C |
| 104 | - | C |
| 105 | - | E |
| 106 | - | E |
| 107 | - | C |
| 108 | - | E |

# 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 with probability 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 exist:

**PBD:**

➢ 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**

**Example: PBD with b=2 (corresponding to block size 4)**

| Sequence No | Block No | Group |
|---|---|---|
| 101 | 1 | E |
| 102 | 1 | C |
| 103 | 1 | C |
| 104 | 1 | E |
| 105 | 2 | E |
| 106 | 2 | E |
| 107 | 2 | C |
| 108 | 2 | C |

Boehringer Ingelheim

# Selection bias introduced by intelligent guesses of the investigator

„In an unmasked study, or one with the potential for unmasking, the principal concern is the potential for introduction of selection bias.
From our experience, it is human nature to try to arrange for a patient whom one feels is better suited to receive treatment A ( B ) to be more likely to receive that treatment.
This could be done, for example, by scheduling the randomization visit when one thinks it is more likely that the next assignment will be A ( B )“

Rosenberger WF & Lachin J (2015), Randomization in Clinical Trials : Theory and Practice„ Chapter 8.3, p.149-150)

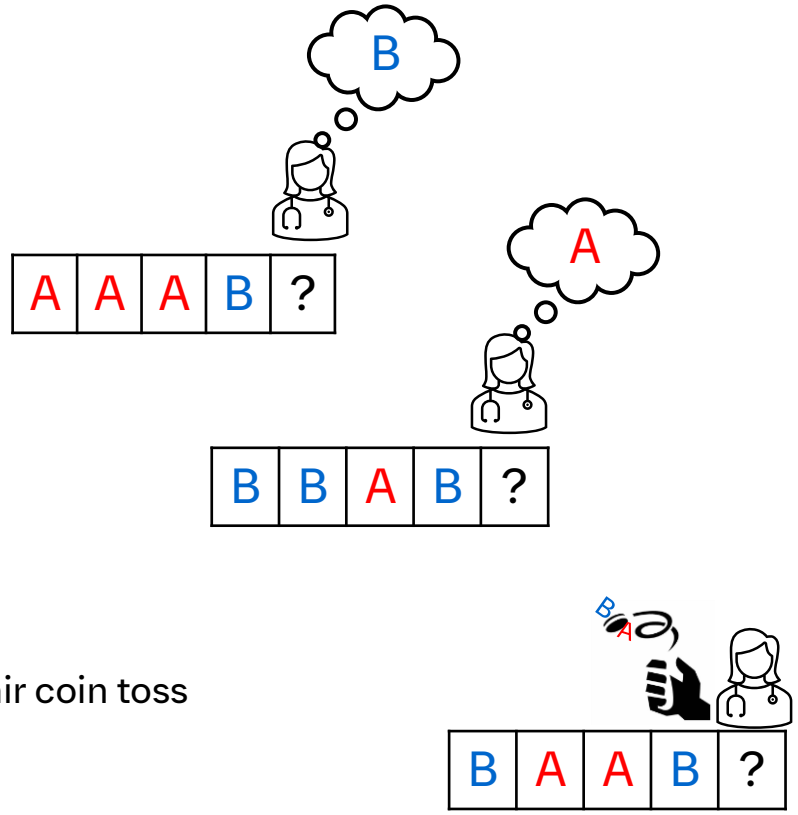
# Blackwell–Hodges convergence strategy

- Blackwell and Hodges (1957) show that, under a restricted randomization procedure with balanced allocation, the optimal strategy (also called “convergence strategy”) to guess the treatment assignment j would be

1. If, among the (j-1) previous treatment assignments, more patients were assigned to A compared to B
   → guess B

2. If, among the (j-1) previous treatment assignments, more patients were assigned to B compared to A
   → guess A

3. In case of a tie, guess the next treatment assignment at random using a fair coin toss



# Comparing block-based designs and BSD in terms of allocation randomness

| | Expected proportion of deterministic assignments | | | Excess Correct Guess Probability | | |
|---|---|---|---|---|---|---|
| b | PBD | RPBD | BSD | PBD | RPBD | BSD |
| 1 | 50% | - | 50% | 25% | - | 25% |
| 2 | 33% | 38.9% | 25% | 21% | 22.2% | 12.5% |
| 3 | 25% | 28.3% | 17% | 18% | 19.2% | 8% |
| 4 | 20% | 22.1% | 12.5% | 16.5% | 17.1% | 6% |

- Excess Correct Guess Probability = Expected Proportions of Correct Guesses according to Blackwell-Hodges Convergence Strategy – 0.5
- RPBD: Random Permuted Block Design which randomly draws block lengths of 2b and 2(b-1) to generate the allocation sequence

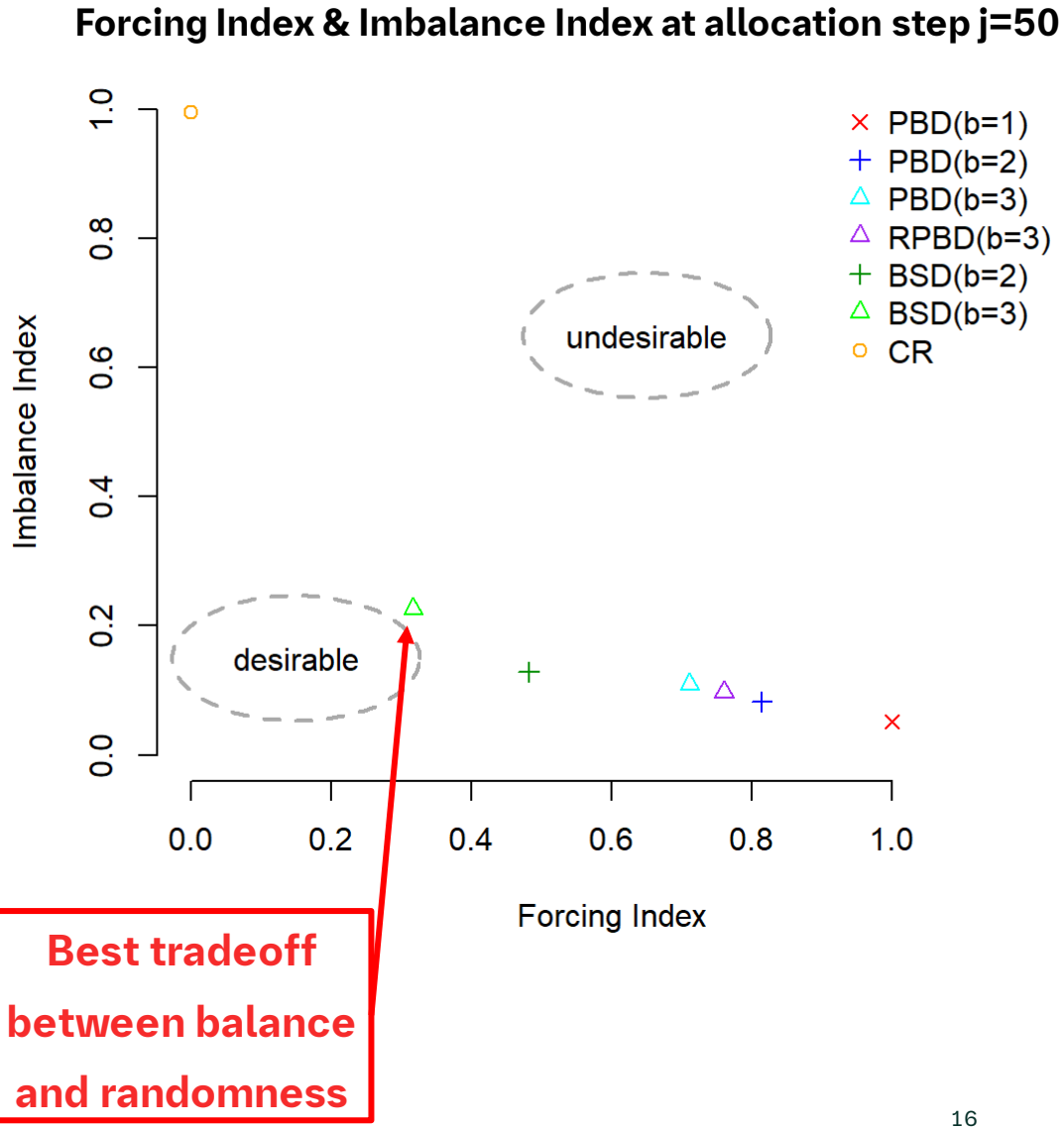
Berger V, Bour L, Carter K et al. (2021). A roadmap to using randomization in clinical trials. BMC Med Res Methodol 21, 168.

# Tradeoff assessment between balance and randomness

- Use standardized performance metrics on [0,1] range (Berger et al. 2021) for assessing tradeoff:
  - Forcing Index
  - Imbalance Index
  - $d$ corresponds to „distance from (x,y) to origin (0,0)“ .

| Method | FI | Imbalance | d |
|---|---|---|---|
| PBD(b=1) | 1 | 0.05 | 1 |
| PBD(b=2) | 0.82 | 0.08 | 0.82 |
| PBD(b=3) | 0.71 | 0.11 | 0.72 |
| RPBD(b=3) | 0.76 | 0.10 | 0.77 |
| BSD(b=2) | 0.48 | 0.13 | 0.50 |
| BSD(b=3) | 0.32 | 0.23 | 0.39 |
| CR | 0 | 1 | 1 |

Results are based on 10,000 simulations






Berger V, Bour L, Carter K et al. (2021). A roadmap to using randomization in clinical trials. BMC Med Res Methodol 21, 168.

# Multi-center and multi-arm trials

# Selection bias under central randomization

- Blackwell-Hodges convergence strategy can be directly applied by an investigator in a multi-center trial **IF** randomization is stratified by center
- In case central randomization (not stratified by center) is used, all centers will share the same list → guessing the subsequent assignment will be near impossible under a „random“ patient flow
- However, some study centers may have **„spikes“** in recruitment when multiple participants in a sequence are enrolled and randomized on the same day
- Then, there can still be a merit in using BSD over PBD, depending on the patient enrolment pattern (Krisam et al. 2024)

| **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) |

Krisam J, Ryeznik Y, Carter K, Kuznetsova O, Sverdlov O (2024). Understanding an impact of patient enrollment pattern on predictability of central randomization in a multi-center clinical trial. Statistics in Medicine, https://doi.org/10.1002/sim.10117

Boehringer Ingelheim

# Multi-arm trials with unequal allocation

„ It is possible for the randomization ratio to change in the setting of a master protocol. This can occur when products enter or exit a platform trial over time [...]. For example, one randomization scheme [...] could change the randomization ratio from $\sqrt{2}:1:1$ (control: drug A: drug B) to $\sqrt{3}:1:1:1$ (control: drug A : drug B : drug C) when a third drug, drug C, enters a trial.

FDA (2023): Master Protocols for Drug and Biological Product Development, https://www.fda.gov/media/174976/download

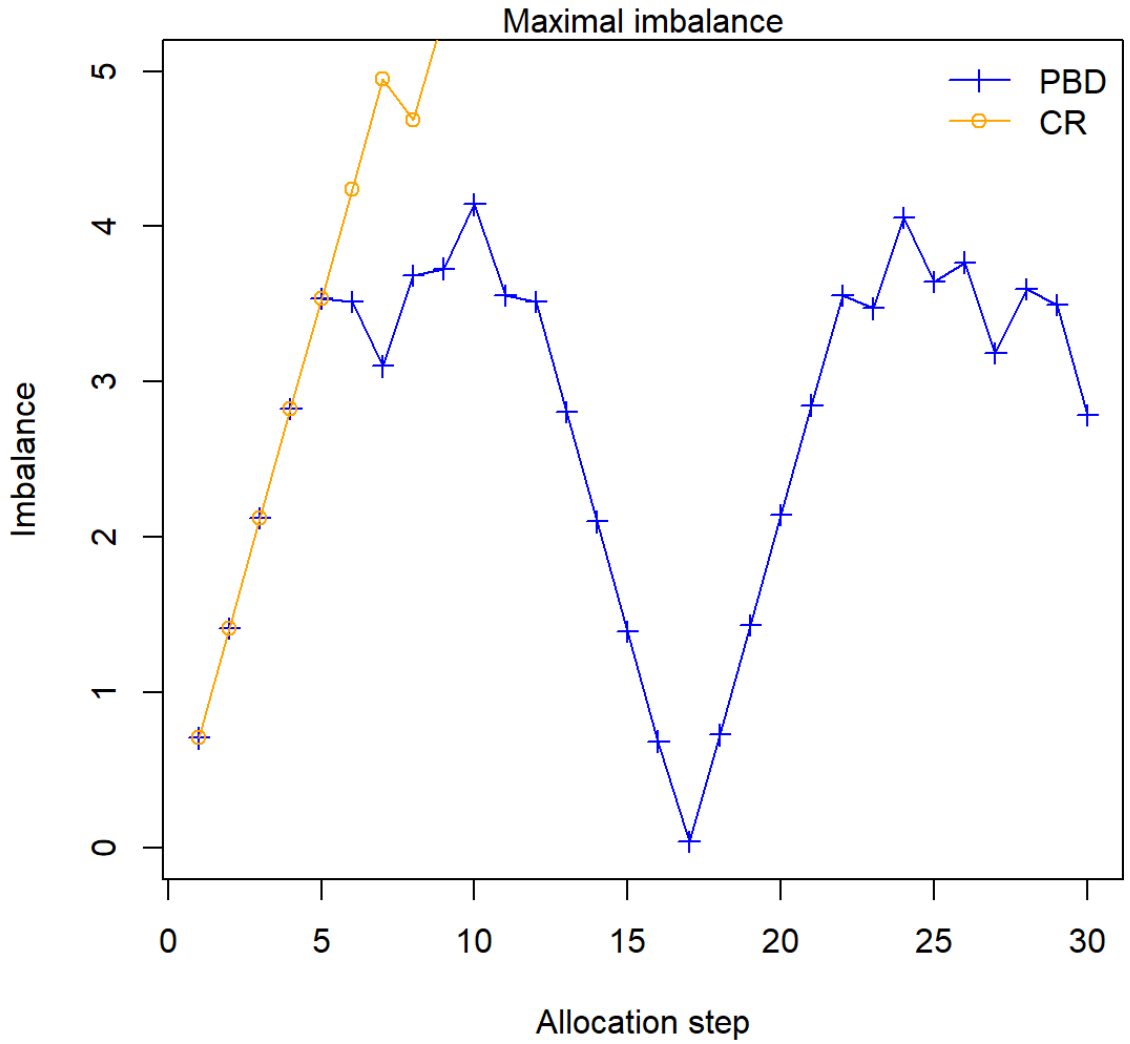
- In their recent draft guidance on Master Protocols, the FDA points out the potential benefit of using unequal allocation in multi-arm trials with a shared control arm.

**How to implement these unequal allocation ratios?**

1. We can use Complete Randomization, but this might likely lead to imbalance
2. Using PBD, we need to use rounding:
   - $\sqrt{2}:1:1 \approx (7:5:5)$ → block length of 17
   - $\sqrt{3}:1:1:1 \approx (7:4:4:4)$ → block length of 19

# Imbalance evaluation for √2:1:1 allocation trial with n=30 assignments

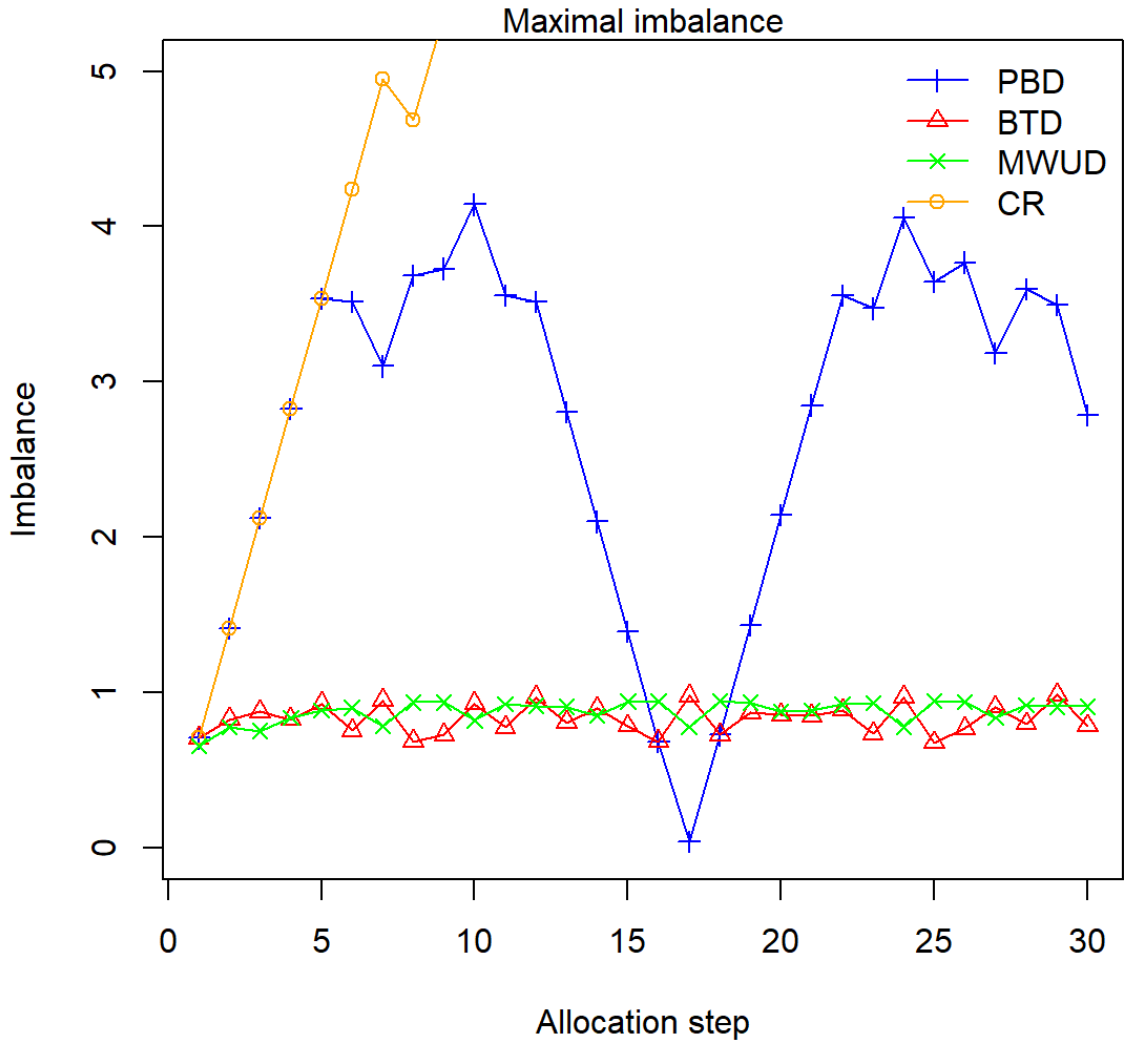
- Imbalance performance measure is the largest deviation across all arms between **expected** and **observed** treatment arm assignments.
- Figure shows maximum value of 10,000 simulations, so worst-case scenario
- **PBD** exhibits no imbalance at the end of the block – but if interim analysis needs to happen when the current randomization is in the middle of the block, considerable imbalances can occur
- This imbalance will become even more of a problem in case of perpetual changes of the randomization ratio




Boehringer Ingelheim

# Controlling imbalance via Brick Tunnel Design and Mass Weighted Urn Design

- Better design alternatives exist that can
  - Target the irrational-valued allocation ratio exactly without rounding
  - Ensure consistent imbalance control
  - Can be based on pre-generated lists just as PBD
  - ➢ **Brick Tunnel Design (BTD)** (Kuznetsova & Tymofyeyev 2011) - ensures minimal deviation from the planned allocation ratio across the whole trial
  - ➢ **Mass Weighted Urn Design (MWUD)** (Zhao 2015) – less strict than BTD with more flexibility via imbalance control parameter (α=5 here)




Kuznetsova OM, Tymofyeyev Y (2011). Brick tunnel randomization for unequal allocation to two or more treatment groups. Stat Med 30:812-24.
Zhao W (2015). Mass weighted urn design--A new randomization algorithm for unequal allocations. Contemp Clin Trials. 43:209-16.

# Regulatory perspective on randomization

# Three quotations – from ICH E9: *Statistical Principles for clinical trials* (1998)

> Although *unrestricted randomisation is an acceptable approach,* some advantages can generally be gained by *randomising subjects in blocks*. This helps to increase the *comparability of the treatment groups*, particularly when subject characteristics may change over time, as a result, for example, of changes in recruitment policy. It also provides a better guarantee that the treatment groups will be of nearly equal size.

> Care should be taken to choose block lengths that are *sufficiently short to limit possible imbalance*, but that are *long enough to avoid predictability* towards the end of the sequence in a block.

> Investigators and other relevant staff *should generally be blind to the block length*; the use of *two or more block lengths, randomly selected for each block, can achieve the same purpose*.

Randomization methods:

- CR, PBD, RPBD are mentioned

General principles:

- Importance of balance to protect against the effect of time trends
- Tradeoff between balance and randomness
- Details of randomization procedure should not be disclosed to investigators

ICH (1998): ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials E9 https://database.ich.org/sites/default/files/E9_Guideline.pdf

# FDA reviewer input to randomization paper on regulatory aspects

- For a recently published paper on regulatory guidance on randomization (Carter et al. 2023), we were allowed to publish the insights of an FDA reviewer in the manuscript:

> "We end with some general principles that we have gathered from this effort: It appears that FDA guidance documents were written to allow for flexibility. Rather than provide specific guidelines on how randomization should be conducted, the guidelines detail important principles to be considered. Sponsors are welcome to discuss proposals of specific randomization methods with the Agency."

Carter K, Scheffold AL, et al. (2023). Regulatory Guidance on Randomization and the Use of Randomization Tests in Clinical Trials: A Systematic Review. Statistics in Biopharmaceutical research, https://doi.org/10.1080/19466315.2023.2239521

# Summary

- Alternatives to PBD might provide better protection against
  - ➢ Selection bias, mostly occurring in open-label RCTs
  - ➢ Imbalances in (multi-arm) RCTs with unequal allocation ratio, e.g. in platform trials, dose-finding trials, trials with Bayesian borrowing, etc.

- While ICH E9 (1998) only mentions CR, PBD, and RPBD, this does not mean that regulatory agencies are opposing alternative methods – it is **up to us** to push for innovation in randomization to improve our RCTs

- Alternative designs are generally **not more difficult to implement** than a PBD, as long as the allocation sequence can be pre-generated

- The **Randomization Working Group** is a group of statisticians from industry, academia and regulatory working to promote the use of novel randomization methods and advance the scientific understanding of these methods in the global community.
- Some topics are:
  - ➢ Methodological research on statistical properties of randomization methods
  - ➢ Randomization-based inference
  - ➢ Development of software tools

rwg

**Visit our LinkedIn page!**

Boehringer Ingelheim


# Visit the LinkedIn page of our Randomization Working Group!

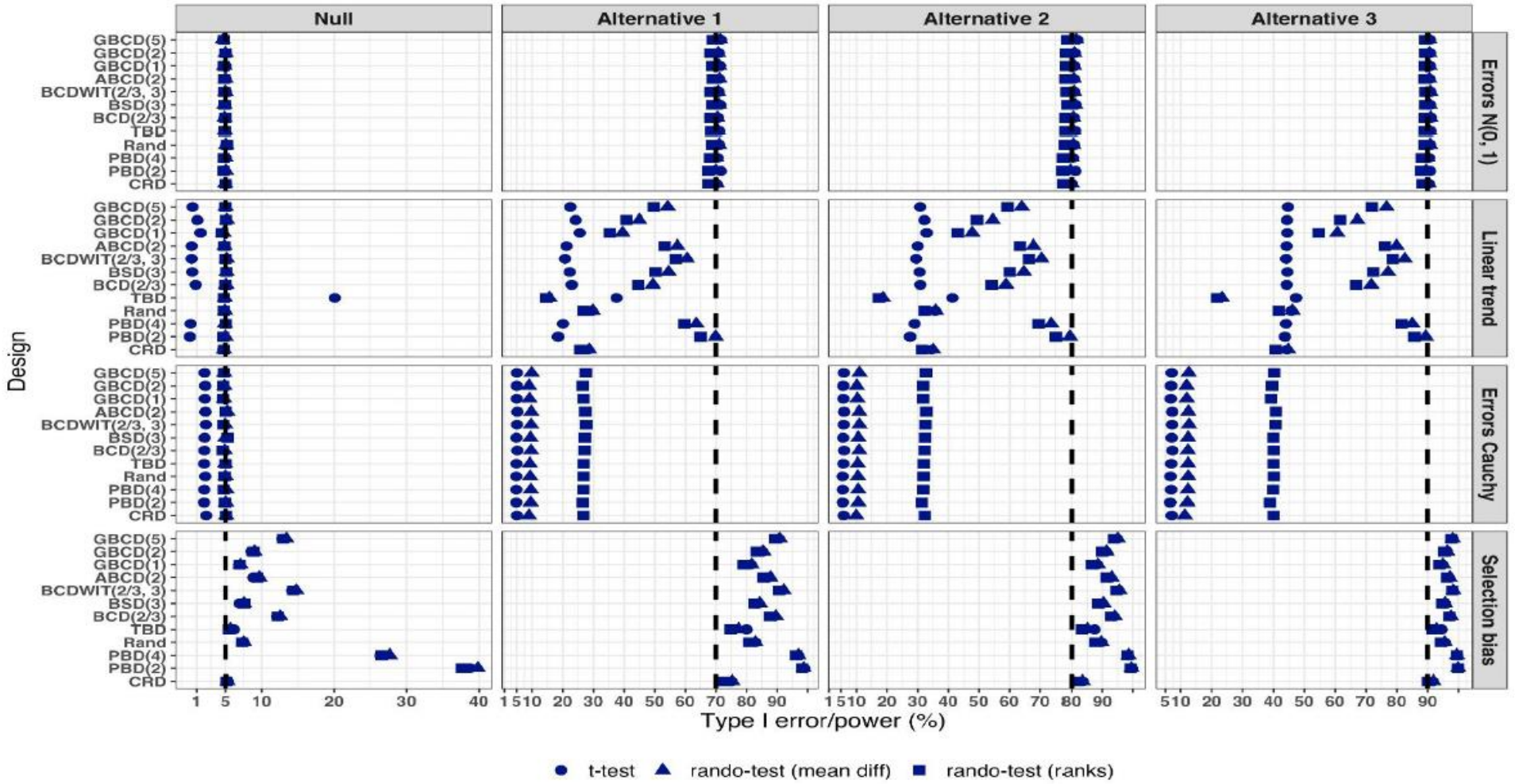
## References


- Berger V, Bour L, Carter K et al. (2021). A roadmap to using randomization in clinical trials. BMC Medical Reseach Methodology 21:168.
- Blackwell D, Hodges JL (1957). Design for the control of selection bias. Annals of Mathematical Statistics 28: 449–460.
- Carter K, Scheffold AL, et al. (2023). Regulatory Guidance on Randomization and the Use of Randomization Tests in Clinical Trials: A Systematic Review. Statistics in Biopharmaceutical Research, https://doi.org/10.1080/19466315.2023.2239521
- FDA (2023). Master Protocols for Drug and Biological Product Development, https://www.fda.gov/media/174976/download
- ICH (1998). ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials E9 https://database.ich.org/sites/default/files/E9_Guideline.pdf
- Kuznetsova OM, Tymofyeyev Y (2011). Brick tunnel randomization for unequal allocation to two or more treatment groups. Statistics in Medicine 30:812-24.
- Krisam J, Ryeznik Y, Carter K, Kuznetsova O, Sverdlov O (2024). Understanding an impact of patient enrollment pattern on predictability of central randomization in a multi-center clinical trial. Statistics in Medicine, https://doi.org/10.1002/sim.10117
- Rosenberger WF, Sverdlov O, Hu F (2012). Adaptive randomization for clinical trials. Journal of Biopharmaceutical Statistics;22(4):719-36.
- Rosenberger WF, Lachin JM (2015). Randomization in Clinical Trials : Theory and Practice, John Wiley & Sons, Incorporated
- Soares JF, Wu CFJ (1983). Some Restricted randomization rules in sequential designs. Communications in Statistics - Theory and Methods 12:17, 2017-2034.
- Zhao W (2015). Mass weighted urn design -- A new randomization algorithm for unequal allocations. Contemporary Clinical Trials. 43:209-16.


Boehringer Ingelheim

# Backup

# Impact of choice of randomization method on type I error rate and power




Simulated type I error rate and power of 12 restricted randomization procedures.

Four models for the data generating mechanism of the primary outcome (M1: Normal random sampling; M2: Linear trend; M3: Errors Cauchy; and M4: Selection bias).

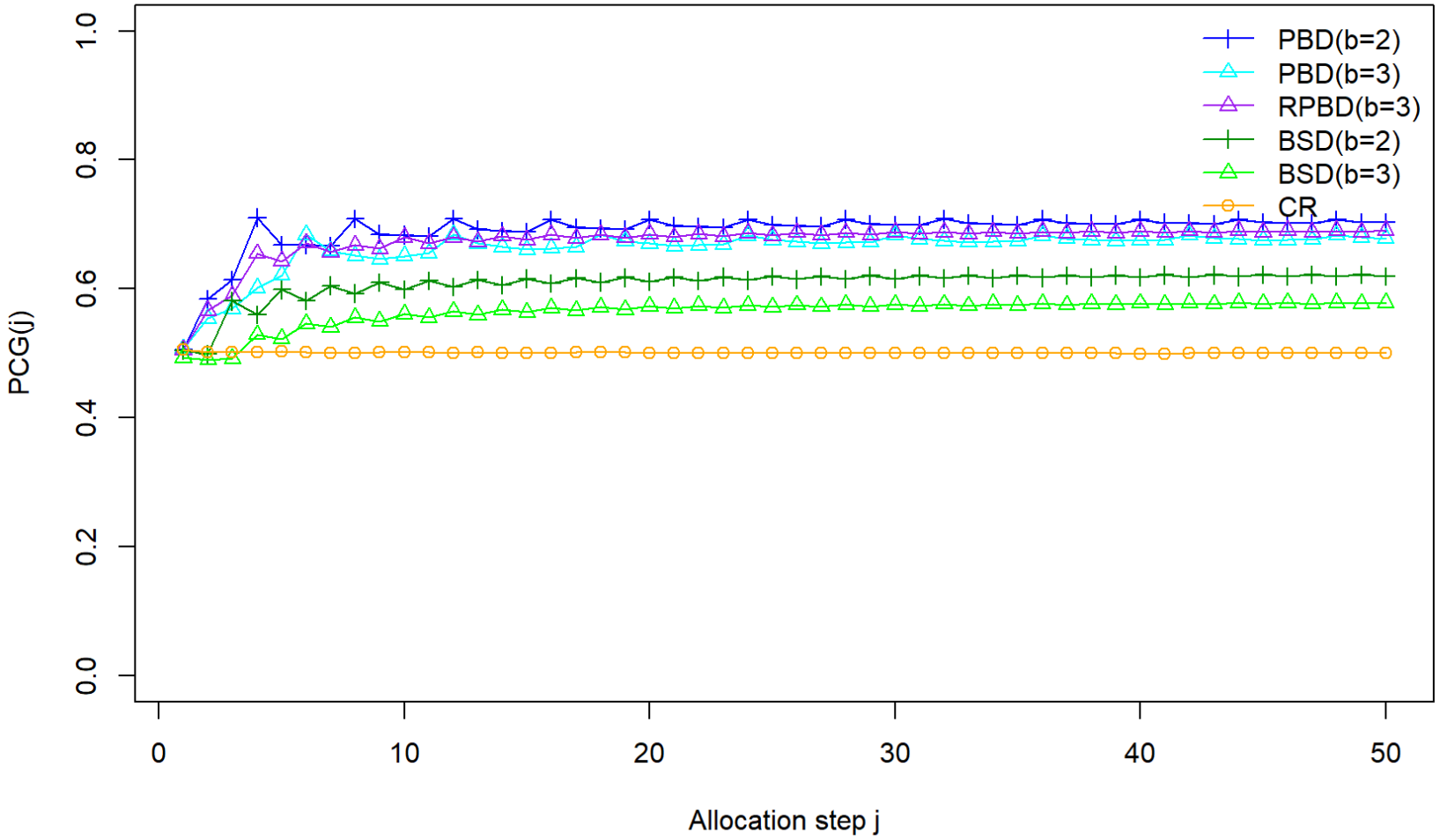
Four scenarios for the treatment mean difference (Null; Alternatives 1, 2, and 3).

Three statistical tests (T1: two-sample t-test; T2: randomization-based test using mean difference; T3:randomization-based test using ranks)

Boehringer Ingelheim

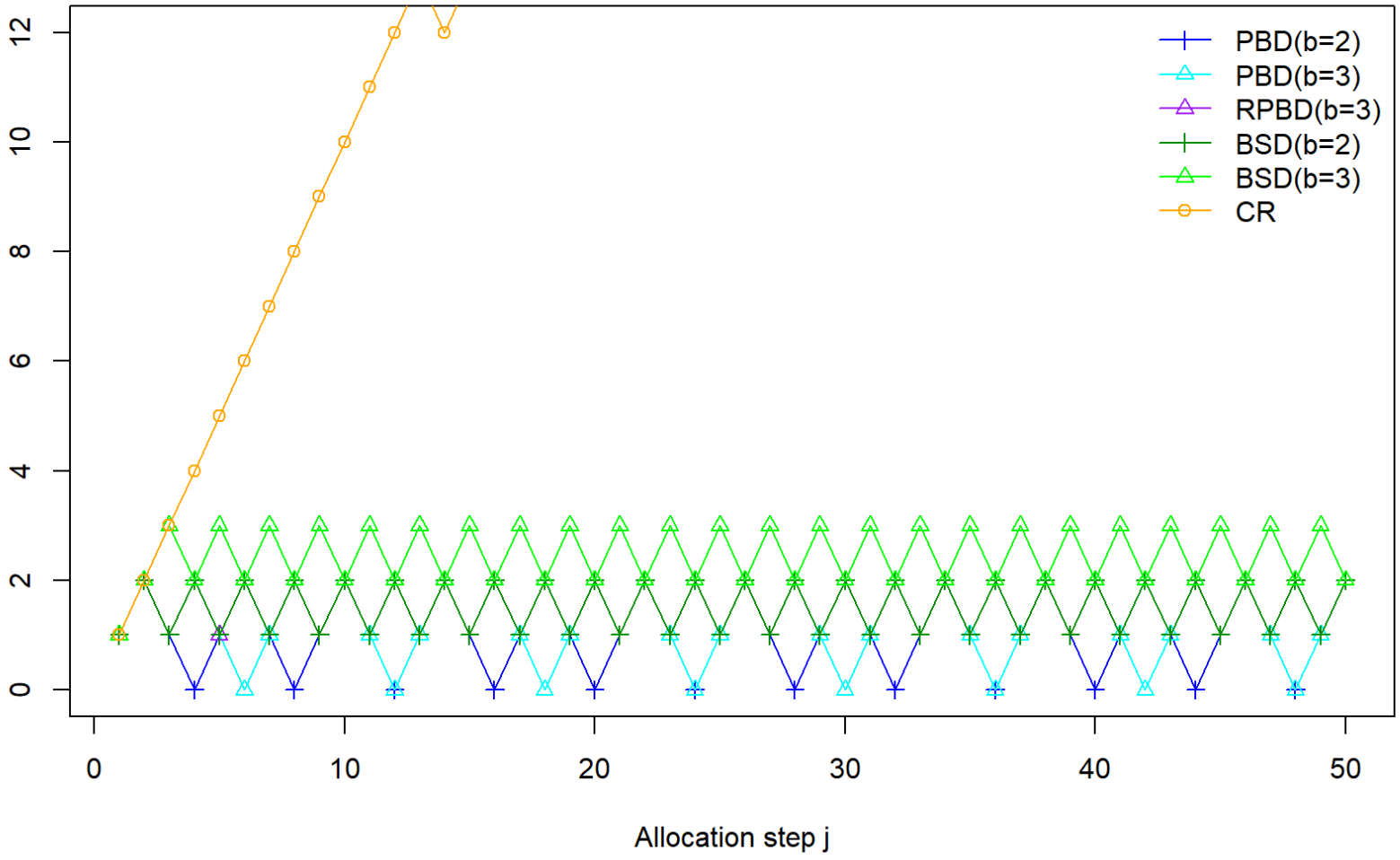
Berger V, Bour L, Carter K et al. (2021). A roadmap to using randomization in clinical trials. BMC Med Res Methodol 21, 168.

# Two-arm trial: Expected Proportion of Correct Guesses

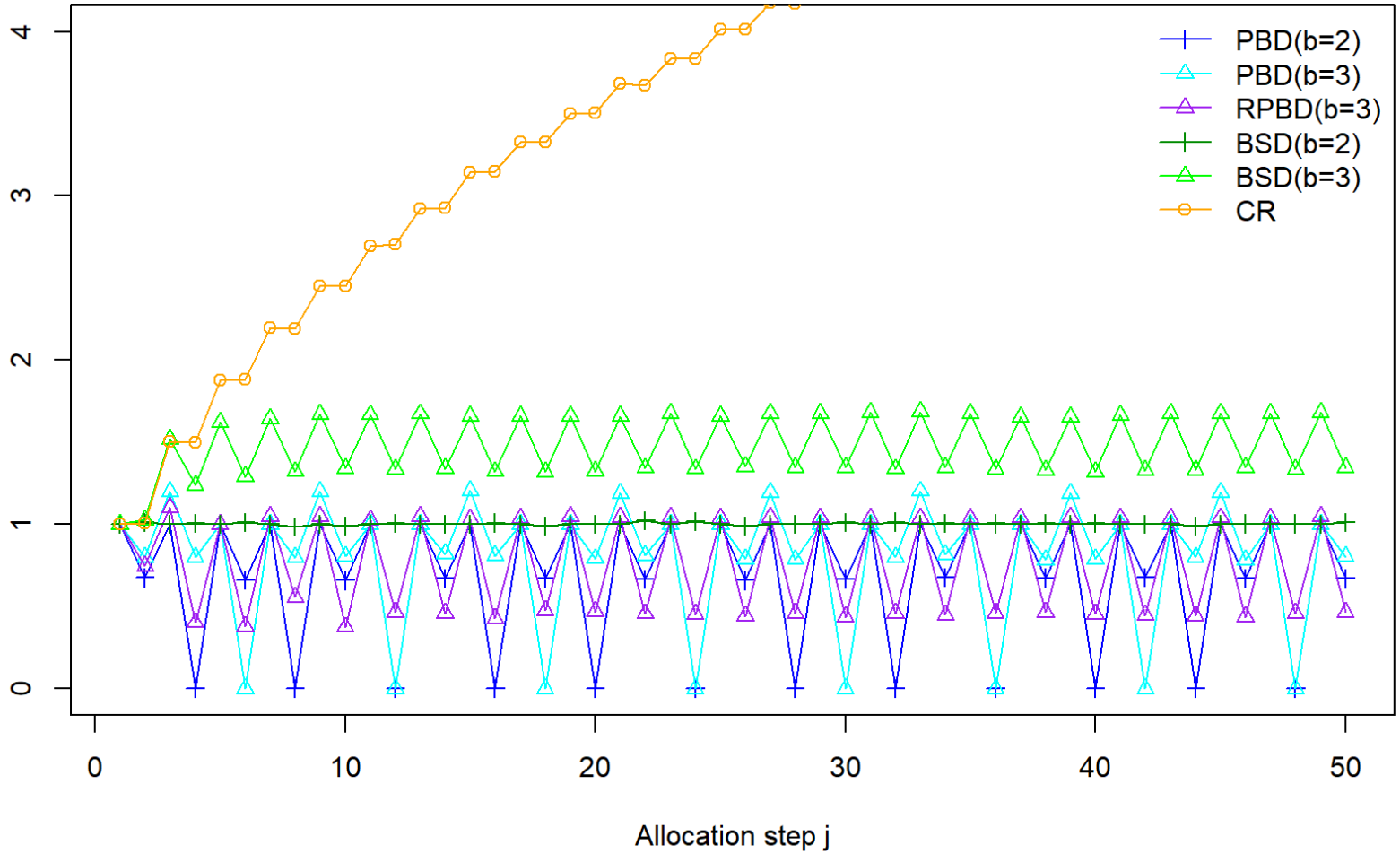



Boehringer Ingelheim

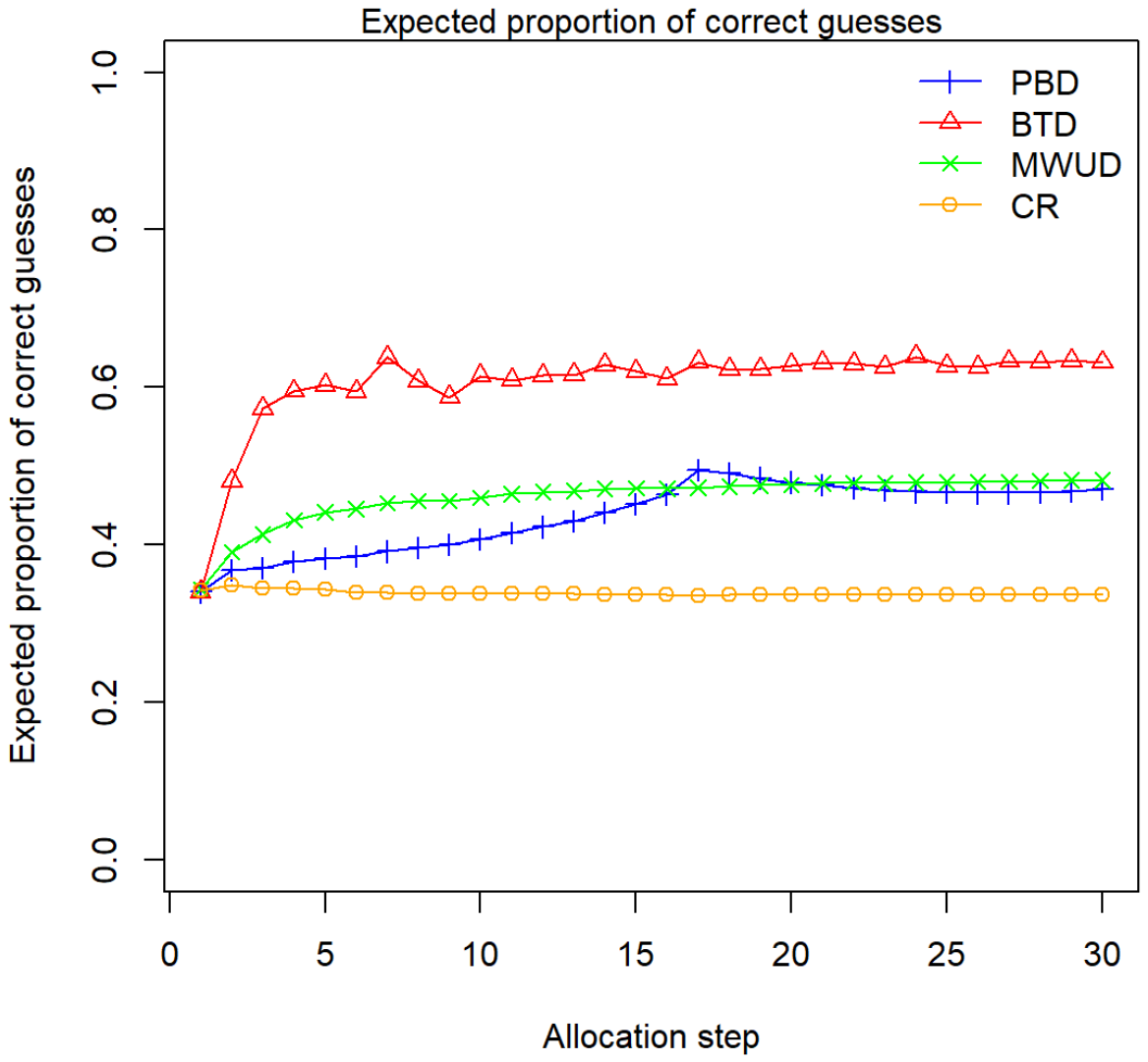
# Two-arm trial: Maximal imbalance $|D(j)|$



# Two-arm trial: Mean imbalance $|D(j)|$

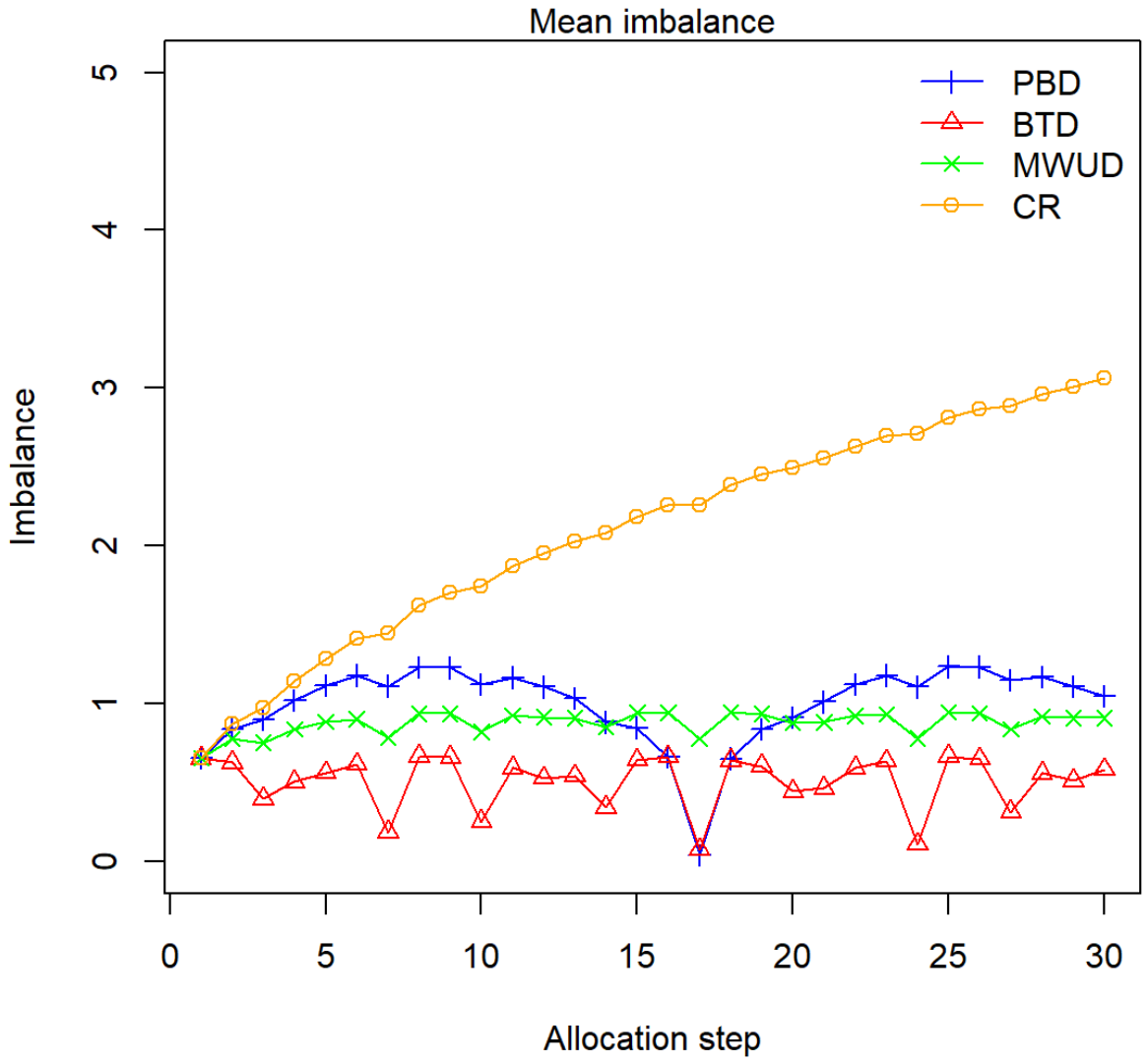


# Multi-arm trial: Expected Proportion of Correct Guesses




Boehringer Ingelheim

# Multi-arm trial: Mean imbalance




Boehringer Ingelheim

# Mass weighted urn design (MWUD)

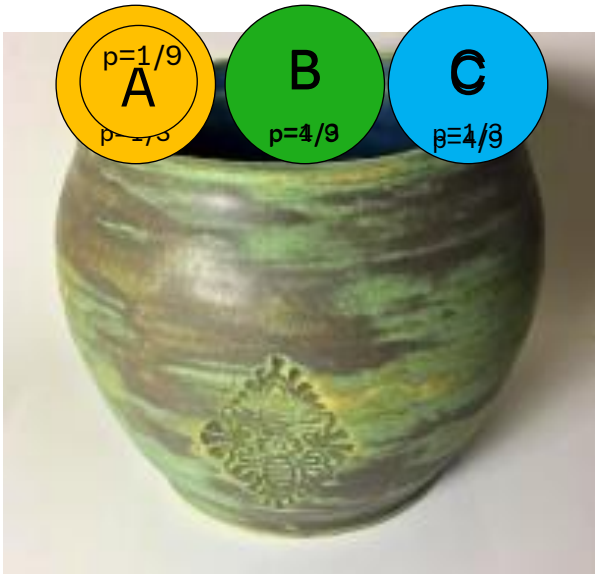
Can be used for unequal allocations and more than 2 arms

- Uses one urn with one ball for each treatment
- The balls can have different weights → balls with more weight are likelier to be picked
- Does not rely on a blocking structure, but has a tuning parameter α for imbalance control

**How does it work?**

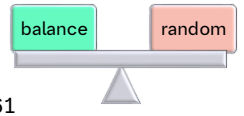
➢ Pick from urn with probability corresponding to weight of the ball → record treatment

➢ decrease weight of selected ball → increase weight of unselected ball




***Properties:***

- Balanced and random
- Can achieve better balance control than PBD in case of many arms and unequal allocation ratios




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 Nov;32(6):953-961

Boehringer Ingelheim

# Brick tunnel design (BTD)

Can be used for unequal allocations and more than 2 arms

- BTD ensures minimal deviation from the planned allocation ratio across the whole trial
- This is done via eliminating any sequence that deviates too far from the planned allocation ratio

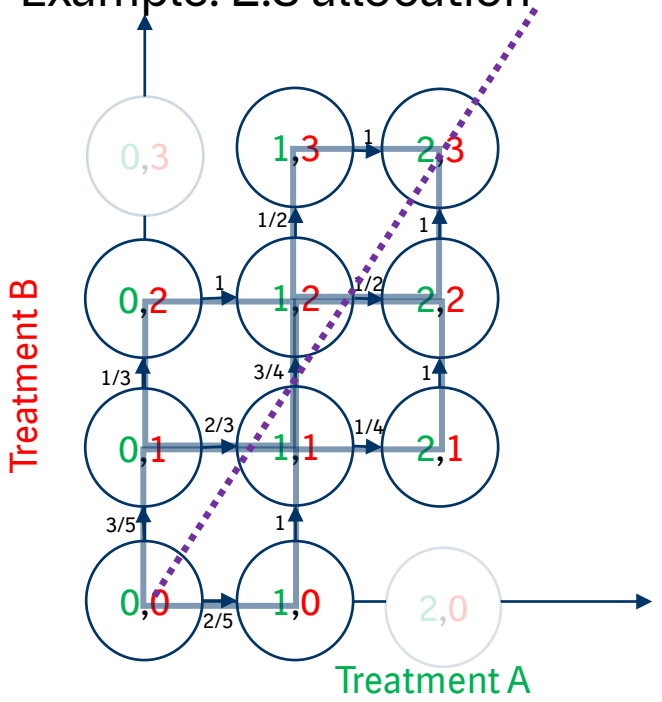
**How does it work?**

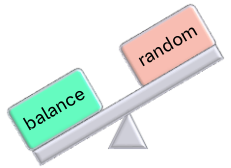
- Eliminate any sequence that deviates too far from the planned allocation ratio
- Transition probabilities are determined such that the unconditional allocation ratio is preserved at each step

***Properties:***

- Very balanced, even under weird allocation ratios and many treatment arms
- Predictable due to strong imbalance control

Example: 2:3 allocation

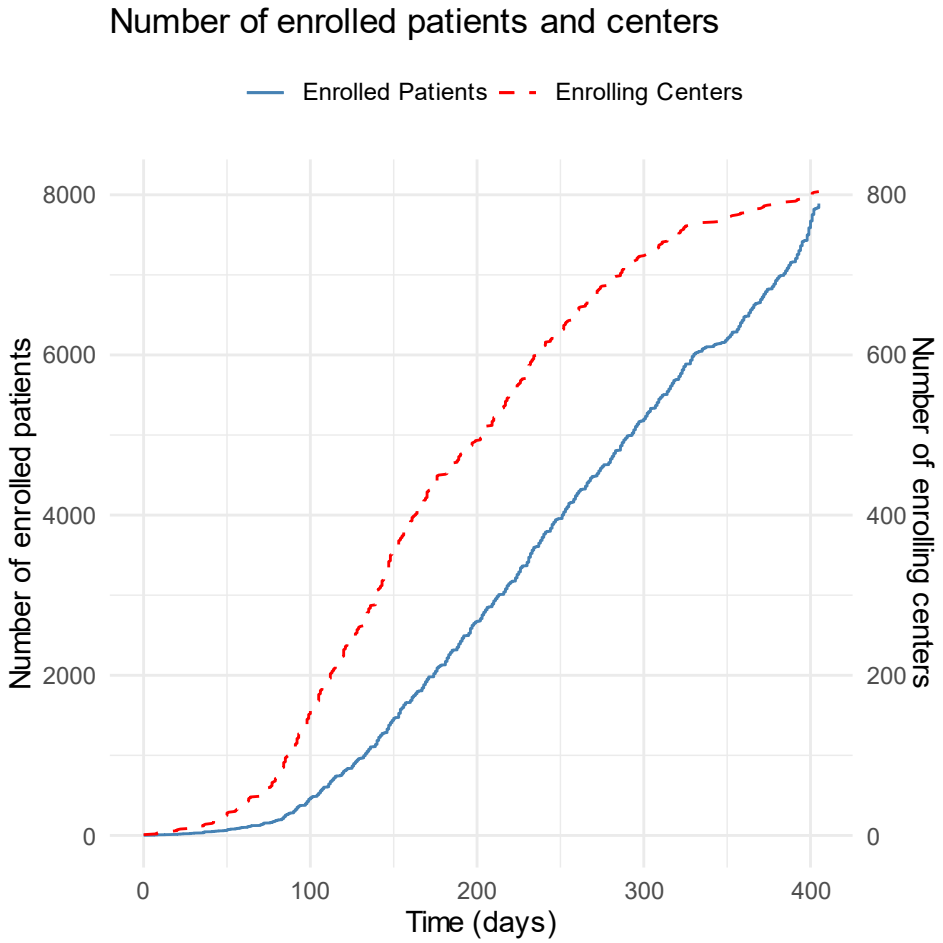


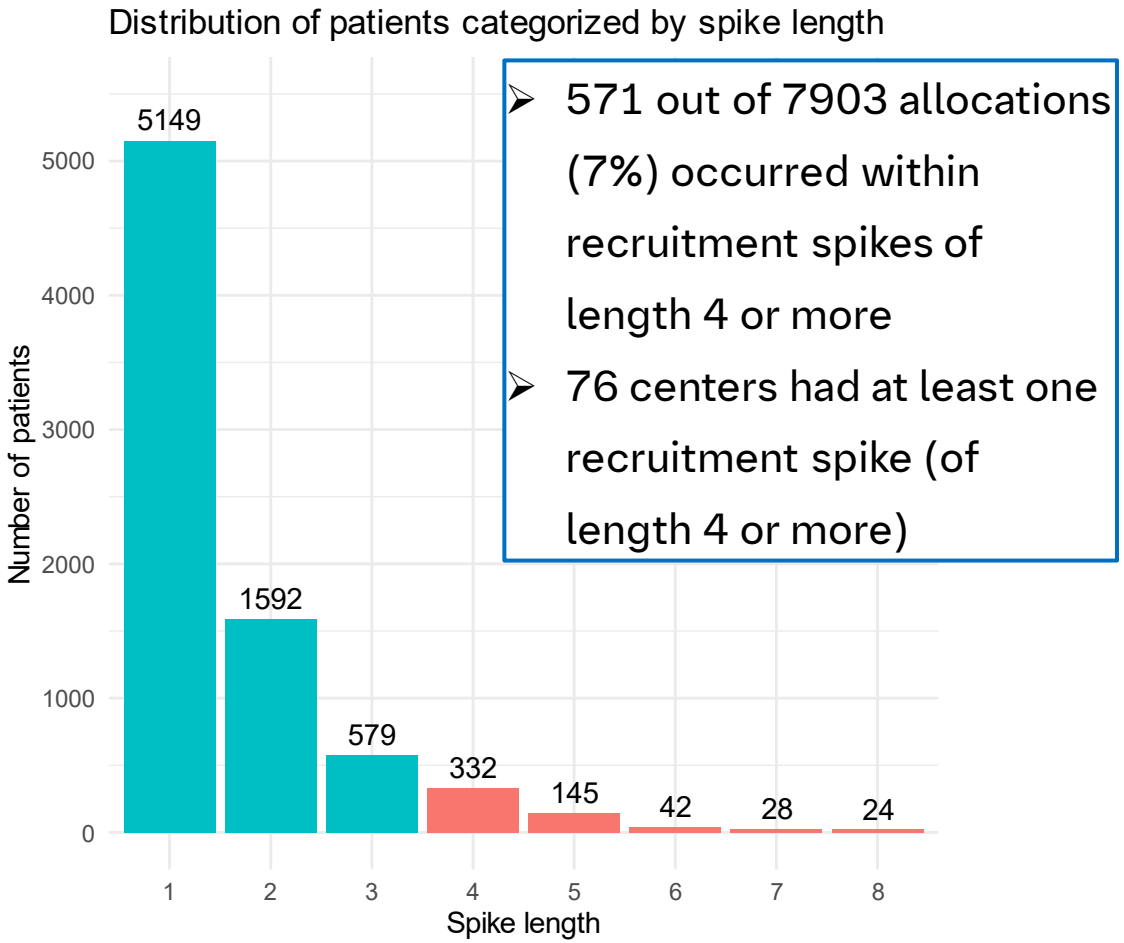




Kuznetsova OM, Tymofyeyev Y (2011). Brick tunnel randomization for unequal allocation to two or more treatment groups. Stat Med 30:812-24.

Boehringer Ingelheim

# Do 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 (= „expected proportion of correct guesses” – 50%)**

| 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