Simulation experiment based on

William F Rosenberger, Feifang Huc (2004), **"Maximizing power and minimizing treatment failures in clinical trials"**, Clinical Trials 2004; 1: 141 -1

Marta Karas
Apr 17, 2019
JHSPH Biostat PhD Seminar on Adaptive Clinical Trials

Authors

William F Rosenberger

- University Professor and Chairman, Department of Statistics, George Mason University (Fairfax, Virginia)
- Authored 2 books: (1) Rosenberger, W. F. and Lachin, J. M. (2016). *Randomization in Clinical Trials: Theory and Practice*, (2) Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*.

Feifang Huc

- Professor of Statistics, Department of Statistics, George Washington University (Washington, D.C.)
- Areas of Expertise: Adaptive design of clinical trials; Bioinformatics; Biostatistics; Bootstrap methods; Statistical issues in personalized medicine; Statistical methods in financial econometrics; Stochastic process.

Background: strategies of treatment group allocation

Setting: The simplest clinical trial of two treatments with a binary outcome.

Question: How to allocate participants between treatment groups?

Background: strategies of treatment group allocation

Setting: The simplest clinical trial of two treatments with a binary outcome.

Question: How to allocate participants between treatment groups?

- **Idea 1**: Fix power, find n_A, n_B to minimize total sample size n.
- **Idea 2**: Fix total sample size n, find n_A, n_B to maximize power.

  Both lead to **Neyman allocation.**

$$n_A = \frac{n\sqrt{P_A Q_A}}{\sqrt{P_A Q_A} + \sqrt{P_B Q_B}},$$

Background: strategies of treatment group allocation

Setting: The simplest clinical trial of two treatments with a binary outcome.

Question: How to allocate participants between treatment groups?

- **Idea 1**: Fix power, find n_A, n_B to minimize total sample size n.
- **Idea 2**: Fix total sample size n, find n_A, n_B to maximize power.

Both lead to **Neyman allocation.** Caveat: may lead to ethical dilemma (when P_A + P_B > 1, it will assign more patients to less successful treatment).

$$n_A = \frac{n\sqrt{P_A Q_A}}{\sqrt{P_A Q_A} + \sqrt{P_B Q_B}},$$

Background: strategies of treatment group allocation

Question: What allocation will simultaneously maximize power and minimize the expected number of treatment failures?

Answer: This has no mathematical solution, but we can modify the problem as follows.

Background: strategies of treatment group allocation

<u>**Question**</u>: What allocation will simultaneously maximize power and minimize the expected number of treatment failures?

<u>**Answer**</u>: This has no mathematical solution, but we can modify the problem as follows.

- <u>**Idea 3**</u>: Fix the expected number of treatment failures, fix n_A, n_B to maximize power (leads to **optimal allocation**).

$$n_A = \frac{n\sqrt{P_A}}{\sqrt{P_A} + \sqrt{P_B}},$$

- <u>**Idea 4**</u>: Fix power, find n_A, n_B to minimize the expected number of treatment failures (leads to **urn allocation**).

Background: strategies of treatment group allocation

Question: What allocation will simultaneously maximize power and minimize the expected number of treatment failures?

Answer: This has no mathematical solution, but we can modify the problem as follows.

Ben's presentation

- **Idea 3**: Fix the expected number of treatment failures, fix n_A, n_B to maximize power (leads to **optimal allocation**).

$$n_A = \frac{n\sqrt{P_A}}{\sqrt{P_A} + \sqrt{P_B}},$$

- **Idea 4**: Fix power, find n_A, n_B to minimize the expected number of treatment failures (leads to **urn allocation**).

Background: strategies of treatment group allocation

<u>**Question**</u>: What allocation will simultaneously maximize power and minimize the expected number of treatment failures?

<u>**Answer**</u>: This has no mathematical solution, but we can modify the problem as follows.

- <u>**Idea 3**</u>: Fix the expected number of treatment failures, fix n_A, n_B to maximize power (leads to **optimal allocation**).

$$n_A = \frac{n\sqrt{P_A}}{\sqrt{P_A} + \sqrt{P_B}},$$

- <u>**Idea 4**</u>: Fix power, find n_A, n_B to minimize the expected number of treatment failures (leads to **urn allocation**).

This presentation

Randomized procedures using urn models

- Use **urn model** to allocate treatment for each subsequent trial participant
- Can be shown that **ratio N_A/N_B tends to the relative risk of failure in the two treatment groups, Q_B/Q_A**
- Two approaches considered in paper:

    (1) Randomized play-the-winner-rule,

    (2) Drop-the-loser rule

(1) Randomized play-the-winner (RPW)

- Start with fixed number of **type A balls** and **type B balls** in the urn
- To randomize a patient, a **ball is drawn**, the corresponding treatment assigned and a **ball is replaced**.
- An **additional ball of the same type** is added if the patient's response is a success, and an **additional ball of the opposite type** is added if the patient's response is a failure.

(1) Randomized play-the-winner (RPW)

- Start with fixed number of **type A balls** and **type B balls** in the urn
- To randomize a patient, a **ball is drawn**, the corresponding treatment assigned and a **ball is replaced**.
- An **additional ball of the same type** is added if the patient's response is a success, and an **additional ball of the opposite type** is added if the patient's response is a failure.

Randomized play-the-winner (RPW) ~ add balls corresponding to successful treatment group

(2) Drop-the-loser (DL)

- Urn contains balls of three types, **type A, type B**, and **type 0**.
- Ball is drawn at random. If it is **type A** or **type B,** the corresponding **treatment is assigned** and the patient's response is observed.
    - If it is a **success, the ball is replaced** and the urn remains unchanged.
    - If it is a **failure, the ball is not replaced**.
- If a **type 0** ball is drawn, no subject is treated, and the ball is returned to the urn together with one ball of type A and one ball of type B. Ensure that the urn never gets depleted.

(2) Drop-the-loser (DL)

- Urn contains balls of three types, **type A, type B**, and **type 0**.
- Ball is drawn at random. If it is **type A** or **type B,** the corresponding **treatment is assigned** and the patient's response is observed.
    - If it is a **success, the ball is replaced** and the urn remains unchanged.
    - If it is a **failure, the ball is not replaced**.
- If a **type 0** ball is drawn, no subject is treated, and the ball is returned to the urn together with one ball of type A and one ball of type B. Ensure that the urn never gets depleted.

Drop-the-loser (DL) ~ <u>remove balls</u> corresponding to failing treatment group

Article results

Table 2 Simulated power and expected treatment failures (standard deviation) for complete randomization and two response-adaptive randomization procedures

			Complete		RPW Rule		DL Rule	
P_A	P_B	n	Power	Failures	Power	Failures	Power	Failures
0.9	0.3	24	90	10 (2.4)	87	7 (2.4)	90	7 (1.8)
0.9	0.5	50	90	15 (3.2)	87	12 (3.2)	89	12 (2.6)
0.9	0.7	162	90	32 (5.1)	88	28 (5.4)	89	27 (4.6)
0.9	0.8	532	90	80 (8)	89	75 (9)	89	73 (8)
0.7	0.3	62	90	31 (4.0)	88	28 (4.3)	89	27 (4.1)
0.7	0.5	248	90	99 (7.8)	89	94 (8.2)	89	93 (8.0)
0.5	0.4	1036	90	570 (16)	89	565 (16)	89	565 (16)
0.3	0.1	158	90	126 (5.1)	89	125 (5.4)	90	124 (5.3)
0.2	0.1	532	90	452 (8)	90	451 (8)	90	451 (8)

RPW: randomized play-the-winner; DL: drop-the-loser

10 000 replications ($\alpha = 0.05$ two-sided).

The sample size was selected that yielded simulated power of approximately 90 percent under complete randomization.

Article results: sample size n reproduced

Table 2 Simulated power and expected treatment failures (standard deviation) for complete randomization and two response-adaptive randomization procedures

		Complete				RPW Rule		DL Rule	
P_A	P_B	n	reprod.	% diff	Failures	Power	Failures	Power	Failures
0.9	0.3	24	24	0.0	10 (2.4)	87	7 (2.4)	90	7 (1.8)
0.9	0.5	50	50	0.0	15 (3.2)	87	12 (3.2)	89	12 (2.6)
0.9	0.7	162	158	-2.5	32 (5.1)	88	28 (5.4)	89	27 (4.6)
0.9	0.8	532	522	-1.9	80 (8)	89	75 (9)	89	73 (8)
0.7	0.3	62	62	0.0	31 (4.0)	88	28 (4.3)	89	27 (4.1)
0.7	0.5	248	252	1.6	99 (7.8)	89	94 (8.2)	89	93 (8.0)
0.5	0.4	1036	1038	0.2	570 (16)	89	565 (16)	89	565 (16)
0.3	0.1	158	158	0.0	126 (5.1)	89	125 (5.4)	90	124 (5.3)
0.2	0.1	532	518	-2.6	452 (8)	90	451 (8)	90	451 (8)

Article results: power reproduced

Table 2 Simulated power and expected treatment failures (standard deviation) for complete randomization and two response-adaptive randomization procedures

			Complete			RPW Rule			DL Rule		
P_A	P_B	n	Power	reprod.	% diff	Power	reprod.	% diff	Power	reprod.	% diff
0.9	0.3	24	90	92.1	2.3	87	89.8	3.3	90	91.1	1.2
0.9	0.5	50	90	90.9	1.0	87	90.1	3.5	89	91.0	2.2
0.9	0.7	162	90	90.0	0.0	88	88.3	0.4	89	89.9	1.0
0.9	0.8	532	90	90.7	0.8	89	89.0	0.0	89	89.9	1.0
0.7	0.3	62	90	90.7	0.8	88	88.5	0.6	89	88.6	-0.5
0.7	0.5	248	90	89.5	-0.5	89	90.2	1.3	89	90.3	1.5
0.5	0.4	1036	90	90.1	0.1	89	89.9	1.0	89	89.8	0.8
0.3	0.1	158	90	89.8	-0.2	89	89.7	0.8	90	89.6	-0.5
0.2	0.1	532	90	90.0	0.0	90	89.2	-0.8	90	89.6	-0.5

Article results: expected # failures reproduced

Table 2 Simulated power and expected treatment failures (standard deviation) for complete randomization and two response-adaptive randomization procedures

			Complete				RPW Rule			DL Rule		
P_A	P_B	n	Power	Fail	reprod.	% diff	Fail	reprod.	% diff	Fail	reprod.	% diff
0.9	0.3	24	90	10	10	0.0	7	7	0.0	7	7	0.0
0.9	0.5	50	90	15	15	0.0	12	12	0.0	12	12	0.0
0.9	0.7	162	90	32	32	0.0	28	27	-3.6	27	27	0.0
0.9	0.8	532	90	80	78	-2.5	75	73	-2.7	73	71	-2.7
0.7	0.3	62	90	31	31	0.0	28	28	0.0	27	27	0.0
0.7	0.5	248	90	99	101	2.0	94	96	2.1	93	95	2.2
0.5	0.4	1036	90	570	571	0.2	565	566	0.2	565	566	0.2
0.3	0.1	158	90	126	126	0.0	125	125	0.0	124	124	0.0
0.2	0.1	532	90	452	440	-2.7	451	439	-2.7	451	439	-2.7

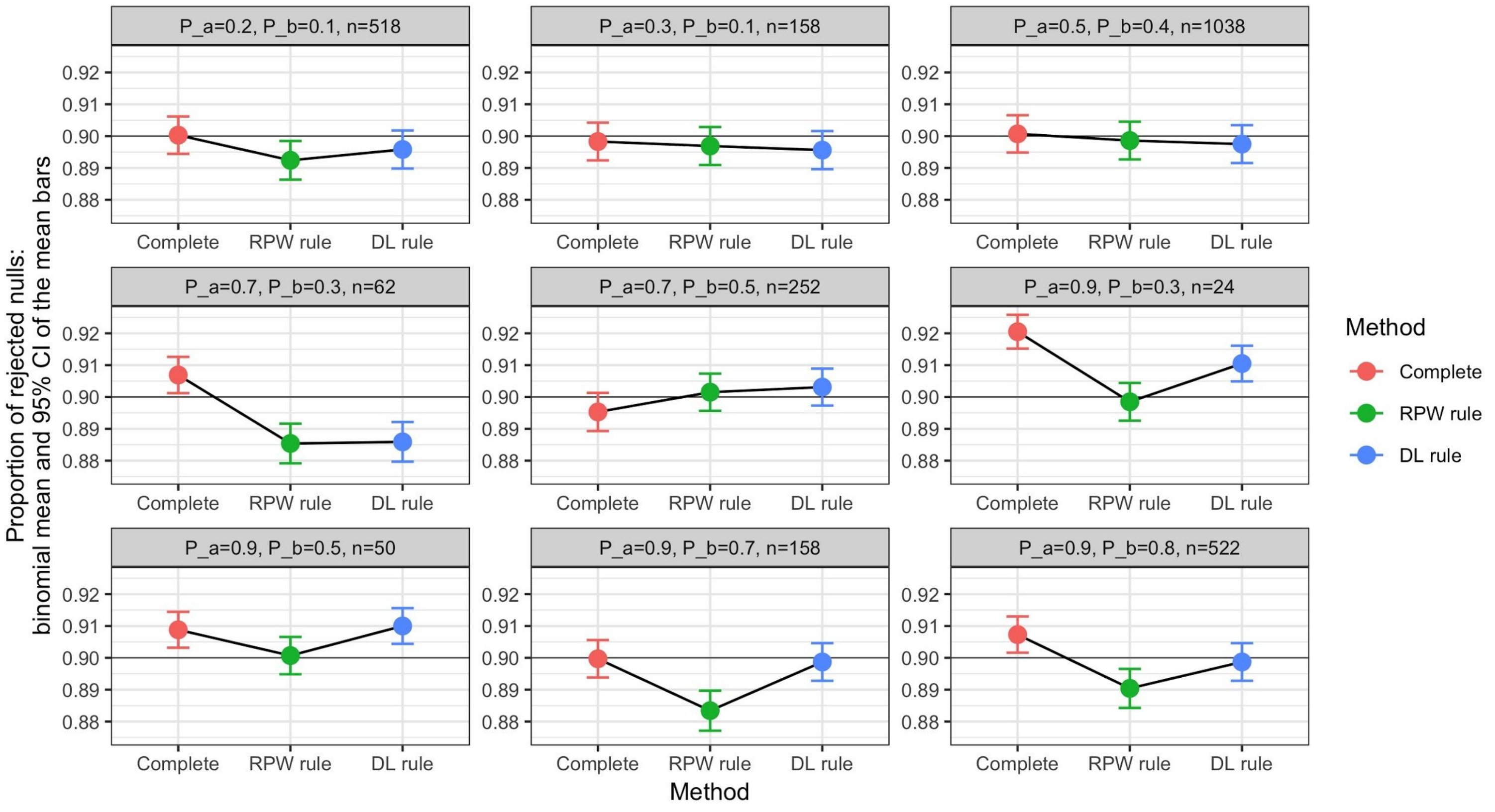
Take a closer look at simulation results

We plot

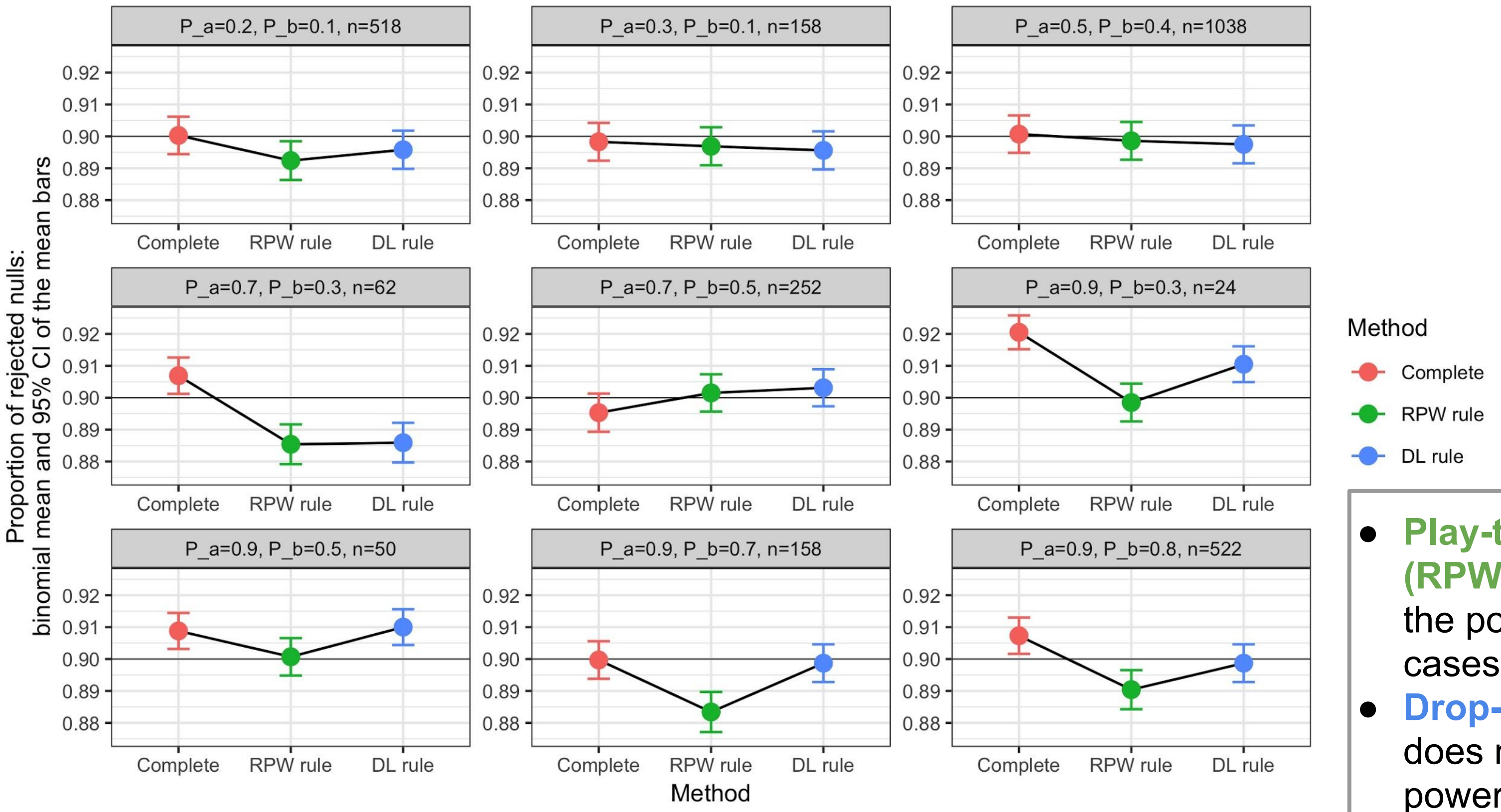
- **proportion of rejected nulls (estimator of power)**, together with 95% confidence intervals of the mean
- **mean number of failures**, together with 95% confidence intervals of the mean

across 9 simulation scenarios considered.

Proportion of rejected nulls: comparison across simulation scenarios

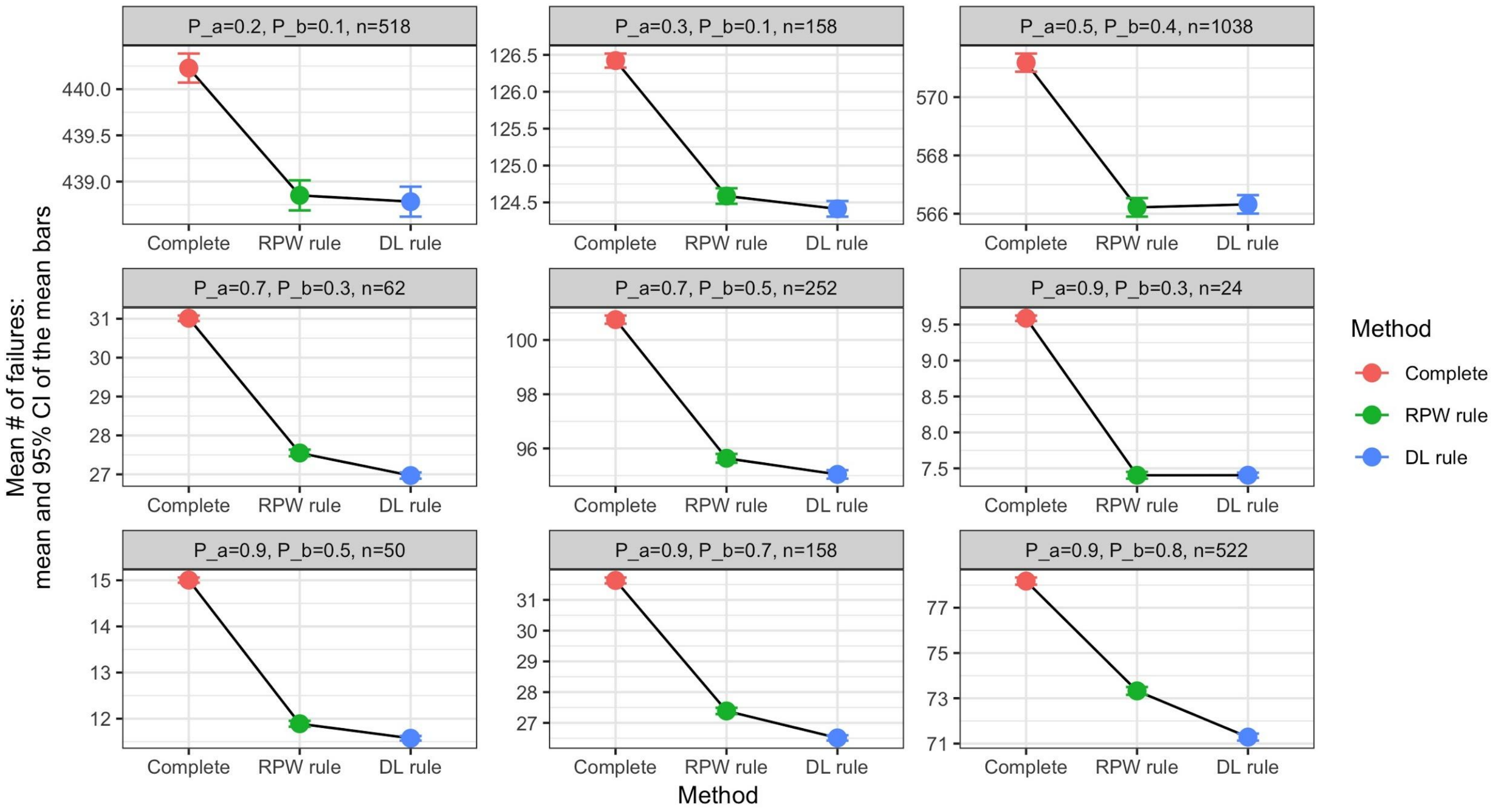


Proportion of rejected nulls: comparison across simulation scenarios

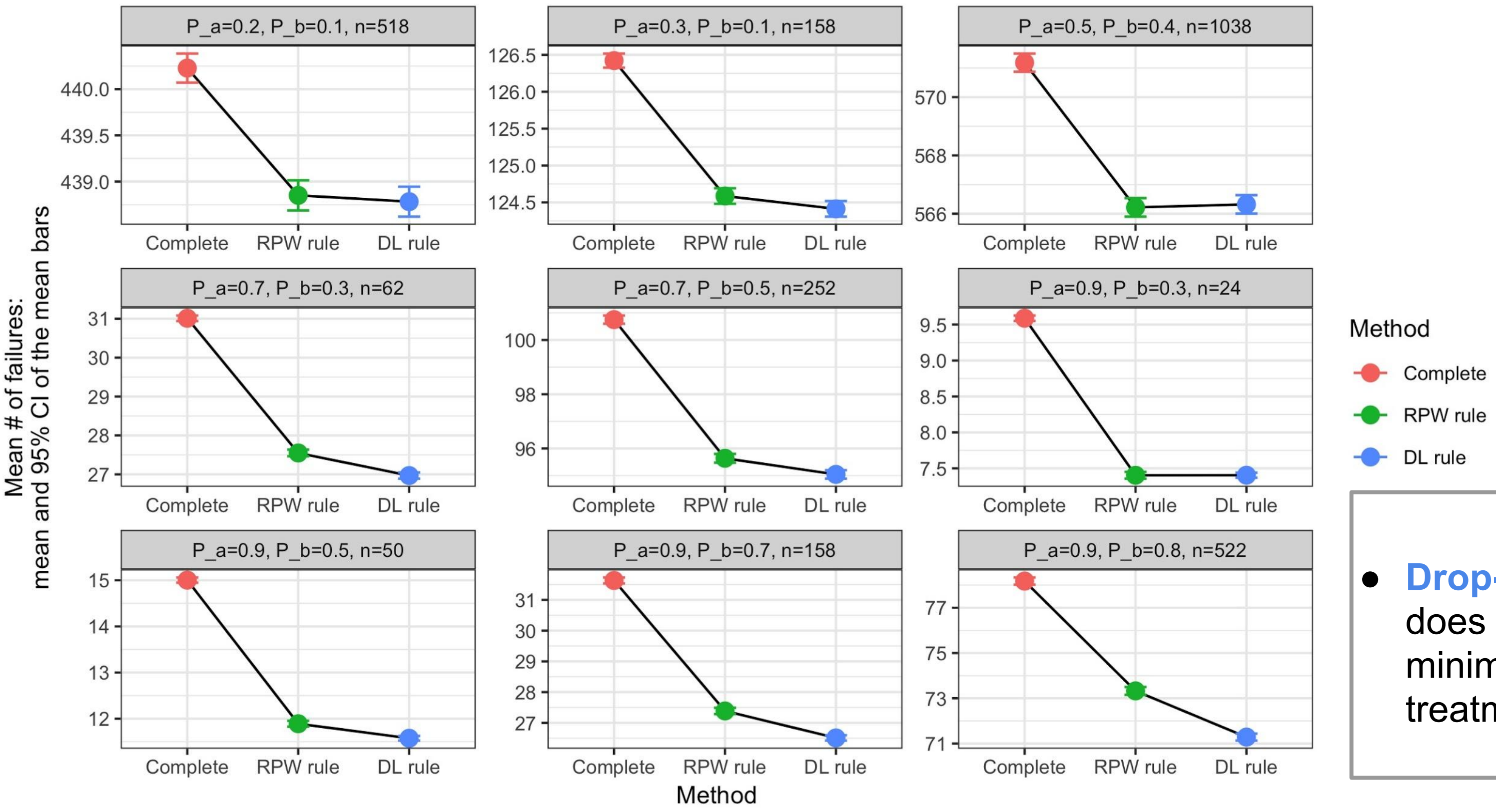


- **Play-the-winner (RPW)** does not keep the power in 4/9 cases
- **Drop-the-loser (DL)** does not keep the power in 1/9 cases

Mean # of failures: comparison across simulation scenarios



Mean # of failures: comparison across simulation scenarios



- **Drop-the-loser (DL)** does the best job in minimizing # of treatment failures

Conclusions from the article

The drop-the-loser rule is better than the randomized play-the-winner rule in every case, having slightly larger power and fewer expected treatment failures. We see that the drop-the-loser rule preserves power quite adequately over complete randomization, and in every case results in fewer expected failures, ranging from approximately one to six fewer expected failures. While these reductions may not be dramatic, such reductions are desirable in clinical trials where treatment failures are particularly undesirable. It is clear from these results that there is little reason to use the randomized play-the-winner rule when the drop-the-loser rule is available.

Conclusions from the article

The drop-the-loser rule is better than the randomized play-the-winner rule in every case, having slightly larger power and fewer expected treatment failures. We see that the drop-the-loser rule preserves power quite adequately over complete randomization, and in every case results in fewer expected failures, ranging from approximately one to six fewer expected failures. While these reductions may not be dramatic, such reductions are desirable in clinical trials where treatment failures are particularly undesirable. It is clear from these results that there is little reason to use the randomized play-the-winner rule when the drop-the-loser rule is available.

Conclusions from the article

The drop-the-loser rule is better than the randomized play-the-winner rule in every case, having slightly larger power and fewer expected treatment failures. We see that the drop-the-loser rule preserves power quite adequately over complete randomization, and in every case results in fewer expected failures, ranging from approximately one to six fewer expected failures. While these reductions may not be dramatic, such reductions are desirable in clinical trials where treatment failures are particularly undesirable. It is clear from these results that there is little reason to use the randomized play-the-winner rule when the drop-the-loser rule is available.

- **Play-the-winner (RPW)** does not keep the power in 4/9 cases
- **Drop-the-loser (DL)** does not keep the power in 1/9 cases

- **Drop-the-loser (DL)** does the best job in minimizing # of treatment failures

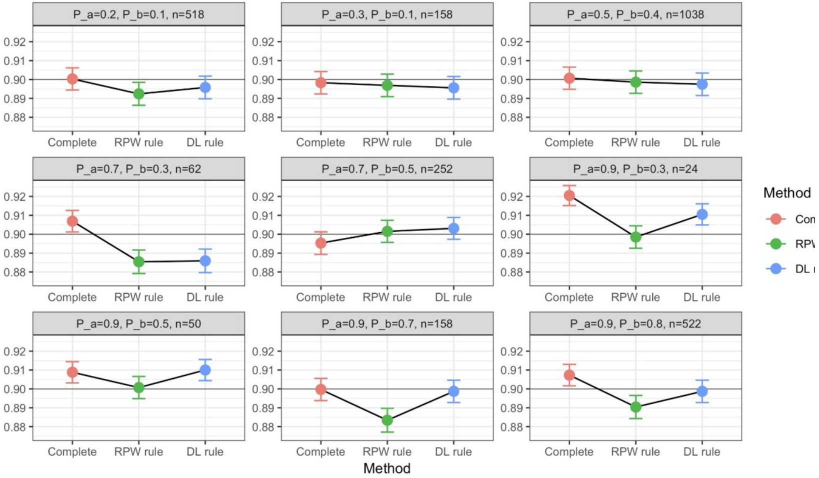
Replicated simulation: agreed

Conclusions from the article

The drop-the-loser rule is better than the randomized play-the-winner rule in every case, having slightly larger power and fewer expected treatment failures. We see that the drop-the-loser rule preserves power quite adequately over complete randomization, and in every case results in fewer expected failures, ranging from approximately one to six fewer expected failures. While these reductions may not be dramatic, such reductions are desirable in clinical trials where treatment failures are particularly undesirable. It is clear from these results that there is little reason to use the randomized play-the-winner rule when the drop-the-loser rule is available.

Conclusions from the article

The drop-the-loser rule is better than the randomized play-the-winner rule in every case, having slightly larger power and fewer expected treatment failures. We see that the drop-the-loser rule preserves power quite adequately over complete randomization, and in every case results in fewer expected failures, ranging from approximately one to six fewer expected failures. While these reductions may not be dramatic, such reductions are desirable in clinical trials where treatment failures are particularly undesirable. It is clear from these results that there is little reason to use the randomized play-the-winner rule when the drop-the-loser rule is available.



Replicated simulation: <u>quite agreed</u>

Reproducible simulation R code available on GitHub:

https://github.com/martakarass/JHU-coursework/tree/master/PH-140-850-Adaptive-Clinical-Trials/final-project

Thank you!