Simulation experiment based on


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)

Feifang Hu

- 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**.

\[
 n_A = \frac{n \sqrt{P_A Q_A}}{\sqrt{P_A Q_A + P_B Q_B}},
\]
**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

**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.

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

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

\[
 n_A = \frac{n \sqrt{P_A}}{\sqrt{P_A} + \sqrt{P_B}},
\]
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.

- **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*).
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.

- **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).
Randomized procedures using urn models

- Use **urn model** to allocate treatment for each subsequent trial participant.
- Can be shown that \( \frac{N_A}{N_B} \) tends to the relative risk of failure in the two treatment groups, \( \frac{Q_B}{Q_A} \).
- 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) ~ remove balls corresponding to failing treatment group
### Table 2  Simulated power and expected treatment failures (standard deviation) for complete randomization and two response-adaptive randomization procedures

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

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

<table>
<thead>
<tr>
<th>$P_A$</th>
<th>$P_B$</th>
<th>$n$</th>
<th>reprod.</th>
<th>% diff</th>
<th>Complete Failures</th>
<th>Power</th>
<th>Failures</th>
<th>RPW Rule Failures</th>
<th>Power</th>
<th>Failures</th>
<th>DL Rule Failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.3</td>
<td>24</td>
<td>24</td>
<td>0.0</td>
<td>10 (2.4)</td>
<td>87</td>
<td>7 (2.4)</td>
<td>90 (1.8)</td>
<td>89</td>
<td>12 (2.6)</td>
<td>89 (2.6)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.5</td>
<td>50</td>
<td>50</td>
<td>0.0</td>
<td>15 (3.2)</td>
<td>87</td>
<td>12 (3.2)</td>
<td>89 (2.6)</td>
<td>89</td>
<td>12 (2.6)</td>
<td>89 (2.6)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.7</td>
<td>162</td>
<td>158</td>
<td>-2.5</td>
<td>32 (5.1)</td>
<td>88</td>
<td>28 (5.4)</td>
<td>89 (2.8)</td>
<td>89</td>
<td>27 (4.6)</td>
<td>89 (2.8)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>532</td>
<td>522</td>
<td>-1.9</td>
<td>80 (8)</td>
<td>89</td>
<td>75 (9)</td>
<td>89 (2.8)</td>
<td>89</td>
<td>73 (8)</td>
<td>89 (2.8)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3</td>
<td>62</td>
<td>62</td>
<td>0.0</td>
<td>31 (4.0)</td>
<td>88</td>
<td>28 (4.3)</td>
<td>89 (2.8)</td>
<td>89</td>
<td>27 (4.1)</td>
<td>89 (2.8)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>248</td>
<td>252</td>
<td>1.6</td>
<td>99 (7.8)</td>
<td>89</td>
<td>94 (8.2)</td>
<td>89 (2.8)</td>
<td>89</td>
<td>93 (8.0)</td>
<td>89 (2.8)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4</td>
<td>1036</td>
<td>1038</td>
<td>0.2</td>
<td>570 (16)</td>
<td>89</td>
<td>565 (16)</td>
<td>89 (2.8)</td>
<td>89</td>
<td>565 (16)</td>
<td>89 (2.8)</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>158</td>
<td>158</td>
<td>0.0</td>
<td>126 (5.1)</td>
<td>89</td>
<td>125 (5.4)</td>
<td>90 (2.8)</td>
<td>90</td>
<td>124 (5.3)</td>
<td>90 (2.8)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>532</td>
<td>518</td>
<td>-2.6</td>
<td>452 (8)</td>
<td>90</td>
<td>451 (8)</td>
<td>90 (2.8)</td>
<td>90</td>
<td>451 (8)</td>
<td>90 (2.8)</td>
</tr>
</tbody>
</table>
## Article results: power reproduced

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

<table>
<thead>
<tr>
<th>$p_A$</th>
<th>$p_B$</th>
<th>$n$</th>
<th>Complete Power</th>
<th>RPW Rule Power</th>
<th>DL Rule Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>reprod.</td>
<td>% diff</td>
<td>reprod.</td>
</tr>
<tr>
<td>0.9</td>
<td>0.3</td>
<td>24</td>
<td>90.1</td>
<td>2.3</td>
<td>89.8</td>
</tr>
<tr>
<td>0.9</td>
<td>0.5</td>
<td>50</td>
<td>90.9</td>
<td>1.0</td>
<td>90.1</td>
</tr>
<tr>
<td>0.9</td>
<td>0.7</td>
<td>162</td>
<td>90.0</td>
<td>0.0</td>
<td>88.3</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>532</td>
<td>90.7</td>
<td>0.8</td>
<td>89.0</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3</td>
<td>62</td>
<td>90.7</td>
<td>0.8</td>
<td>88.5</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>248</td>
<td>89.5</td>
<td>-0.5</td>
<td>90.2</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4</td>
<td>1036</td>
<td>90.1</td>
<td>0.1</td>
<td>89.9</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>158</td>
<td>89.8</td>
<td>-0.2</td>
<td>89.7</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>532</td>
<td>90.0</td>
<td>0.0</td>
<td>89.2</td>
</tr>
</tbody>
</table>
Table 2  Simulated power and expected treatment failures (standard deviation) for complete randomization and two response-adaptive randomization procedures

<table>
<thead>
<tr>
<th>$p_A$</th>
<th>$p_B$</th>
<th>$n$</th>
<th>Power</th>
<th>Fail</th>
<th>Fail</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>reprod.</td>
<td>% diff</td>
<td>reprod.</td>
</tr>
<tr>
<td>0.9</td>
<td>0.3</td>
<td>24</td>
<td>90</td>
<td>10</td>
<td>0.0</td>
<td>7</td>
</tr>
<tr>
<td>0.9</td>
<td>0.5</td>
<td>50</td>
<td>90</td>
<td>15</td>
<td>0.0</td>
<td>12</td>
</tr>
<tr>
<td>0.9</td>
<td>0.7</td>
<td>162</td>
<td>90</td>
<td>32</td>
<td>32.0</td>
<td>28</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>532</td>
<td>90</td>
<td>80</td>
<td>-2.5</td>
<td>75</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3</td>
<td>62</td>
<td>90</td>
<td>31</td>
<td>0.0</td>
<td>28</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>248</td>
<td>90</td>
<td>99</td>
<td>101.0</td>
<td>94</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4</td>
<td>1036</td>
<td>90</td>
<td>570</td>
<td>0.2</td>
<td>565</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>158</td>
<td>90</td>
<td>126</td>
<td>0.0</td>
<td>125</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>532</td>
<td>90</td>
<td>452</td>
<td>-2.7</td>
<td>451</td>
</tr>
</tbody>
</table>
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

- $P_a=0.2$, $P_b=0.1$, $n=518$
- $P_a=0.3$, $P_b=0.1$, $n=158$
- $P_a=0.5$, $P_b=0.4$, $n=1038$
- $P_a=0.7$, $P_b=0.3$, $n=62$
- $P_a=0.7$, $P_b=0.5$, $n=252$
- $P_a=0.9$, $P_b=0.3$, $n=24$
- $P_a=0.9$, $P_b=0.5$, $n=50$
- $P_a=0.9$, $P_b=0.7$, $n=158$
- $P_a=0.9$, $P_b=0.8$, $n=522$

Method:
- Complete
- RPW rule
- DL rule

Proportion of rejected nulls: comparison across simulation scenarios

- $P_a=0.2$, $P_b=0.1$, $n=518$
- $P_a=0.3$, $P_b=0.1$, $n=158$
- $P_a=0.5$, $P_b=0.4$, $n=1038$
- $P_a=0.7$, $P_b=0.3$, $n=62$
- $P_a=0.7$, $P_b=0.5$, $n=252$
- $P_a=0.9$, $P_b=0.3$, $n=24$
- $P_a=0.9$, $P_b=0.5$, $n=50$
- $P_a=0.9$, $P_b=0.7$, $n=158$
- $P_a=0.9$, $P_b=0.8$, $n=522$

Method:
- Complete
- RPW rule
- DL rule
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

- $P_a=0.2, P_b=0.1, n=518$
- $P_a=0.3, P_b=0.1, n=158$
- $P_a=0.5, P_b=0.4, n=1038$
- $P_a=0.7, P_b=0.3, n=62$
- $P_a=0.7, P_b=0.5, n=252$
- $P_a=0.9, P_b=0.3, n=24$
- $P_a=0.9, P_b=0.5, n=50$
- $P_a=0.9, P_b=0.7, n=158$
- $P_a=0.9, P_b=0.8, n=522$

**Method**
- Complete
- RPW rule
- DL rule
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: agreed
Reproducible simulation R code available on GitHub:


Thank you!