

Understanding the Fundamental Problem: Resampling vs Permutation in Treatment Label Assignment

July 16, 2025

Let me break down exactly why the existing strategy for mixing up treatment labels is incorrect by examining the fundamental differences between **resampling** and **permutation**.

1 The Core Problem: Resampling vs Permutation

1.1 What the Original Code Was Doing (WRONG)

```
1 # Original stratified approach
2 all_samples <- replicate(nperm, weightedSampler(data_trimmed, treat,
  sampsiz))
3
4 # Original non-stratified approach
5 all_samples <- replicate(nperm, dplyr::slice_sample(data_trimmed, n =
  sampsiz))
```

Both approaches were doing **resampling** - they were picking random subsets of subjects from the original data.

1.2 What Should Happen: True Permutation

```
1 # Correct approach
2 permuted_data <- data
3 permuted_data[, treat] <- sample(data[, treat]) # Shuffle treatment
  labels only
```

2 Why This Difference Matters Fundamentally

2.1 1. Different Null Hypotheses Being Tested

Resampling approach tests:

- “What if I randomly picked different subjects from my population?”
- This is testing sampling variation, not treatment effect

Permutation approach tests:

- “What if the same subjects had been randomly assigned to different treatments?”
- This is testing the true null hypothesis: no treatment effect

2.2 2. Concrete Example

Let's say you have data on 1000 patients:

Original Data:

1	Subject_ID	Treatment	Outcome
2	1	A	5.2
3	2	A	4.8
4	3	B	6.1
5	4	B	5.9
6
7	1000	B	4.3

What Resampling Does (WRONG):

1	# Picks random 100 subjects, might get:		
2	Subject_ID	Treatment	Outcome
3	15	A	4.9
4	23	B	5.8
5	156	A	5.1
6	...		

- Different subjects entirely
- Introduces sampling variability
- Not testing treatment effect on same population

What Permutation Does (CORRECT):

1	# Keeps ALL 1000 subjects, shuffles treatment labels:		
2	Subject_ID	Treatment	Outcome
3	1	B	5.2 # Same subject, different treatment
4	2	B	4.8 # Same subject, different treatment
5	3	A	6.1 # Same subject, different treatment
6	4	A	5.9 # Same subject, different treatment
7
8	1000	A	4.3 # Same subject, different treatment

3 3. Statistical Implications

3.1 Resampling Problems:

1. Confounds treatment effect with sampling variation

- You're not just testing "does treatment work?"
- You're testing "does treatment work AND did I happen to sample different types of patients?"

2. Violates the permutation test assumption

- Permutation tests assume you're testing all possible ways to assign treatments to the SAME set of subjects

- Resampling changes the subjects, breaking this assumption

3. Incorrect p-values

- The null distribution will be too wide (more variable)
- This can lead to either false positives or false negatives

3.2 Permutation Advantages:

1. Pure treatment effect testing

- Same subjects, same baseline characteristics
- Only treatment assignments change

2. Correct null hypothesis

- Tests: “If treatment had no effect, how likely would we see this difference?”
- Not: “If we sampled different people AND treatment had no effect...”

4 4. Why the Stratified Logic Was Doubly Wrong

The original stratified code had two layers of problems:

```

1  # Original problematic code
2  if (stratified) {
3    # Problem 1: Limited sample size unnecessarily
4    lowest_nr_subject_by_trt <- min(nrow(data[data[treat] == trts[1],])
5                                   ,
6                                   nrow(data[data[treat] == trts[2],]))
7    sampsize <- ifelse(sampsize > lowest_nr_subject_by_trt,
8                      lowest_nr_subject_by_trt, sampsize)
9
10   # Problem 2: Still used resampling instead of permutation
11   all_samples <- replicate(nperm, weightedSampler(data_trimmed, treat
12                                                     , sampsize))

```

Problem 1: It artificially limited the analysis to the smallest treatment group size, throwing away data unnecessarily.

Problem 2: Even after “balancing,” it still resampled subjects rather than permuting treatment labels.

5 5. Mathematical Foundation

5.1 Permutation Test Theory:

The p-value in a permutation test is:

$$P = \frac{\text{Number of permutations with test statistic} \geq \text{observed statistic}}{\text{Total number of permutations}}$$

This is only valid when:

- You’re permuting treatment assignments among the SAME subjects
- Each permutation represents a possible randomization outcome under the null hypothesis

5.2 What Resampling Does:

It creates a distribution based on:

$$P_{wrong} = \frac{\text{Number of random samples with test statistic} \geq \text{observed statistic}}{\text{Total number of random samples}}$$

This tests a different hypothesis entirely and gives incorrect inference.

6 6. Real-World Impact

Consider a funnel plot analysis:

With Resampling (Wrong):

- High variability from different subject compositions
- May falsely identify publication bias when it's just sampling variation
- Or miss real publication bias because noise is too high

With Permutation (Correct):

- Clean test of whether treatment effects vary more than expected by chance
- Proper identification of publication bias patterns
- Correct statistical inference

7 Summary

The existing strategy was incorrect because it fundamentally tested the wrong hypothesis. Instead of asking “What if these same subjects had different treatment assignments?” (permutation), it asked “What if I had different subjects?” (resampling). This confounds treatment effects with sampling variation and leads to incorrect statistical conclusions.

The corrected approach using true permutation provides the proper null hypothesis test that funnel plot analysis requires.