Understanding the Sampling Process in Permutation-Based Funnel Calculation

July 15, 2025

This line of code is creating **multiple random samples** from your dataset for the permutation-based funnel calculation. Let me break it down:

1 What Each Part Does

```
all_samples <- replicate(nperm, slice_sample(data_trimmed, n = sampsize
   ))
```

1.1 replicate(nperm, ...)

- **Purpose**: Repeats the sampling operation *nperm* times (e.g., 1500 times in your case)

- **Returns**: A list or array containing all the repeated results

1.2 slice_sample(data_trimmed, n = sampsize)

- **Purpose**: Randomly samples *sampsize* rows from *data_trimmed*

- **data_trimmed**: Your dataset with only treatment and endpoint columns

- **sampsize**: The maximum subgroup size `H$results_total$N.of.subjects`

1.3 all_samples

- **Result**: Contains 1500 different random samples, each of size *sampsize*

2 Example to Illustrate

```
# If your data looks like this:
data_trimmed <- data.frame(
  trt = c("A", "B", "A", "B", "A", "B"),
  endpoint1 = c(1.2, 2.1, 1.8, 2.5, 1.1, 2.3),
  endpoint2 = c(0.8, 1.2, 0.9, 1.4, 0.7, 1.1)
)

# And sampsize = 4, nperm = 3
# Then all_samples might contain:
```

```
10
11  # Sample 1: rows 1,3,5,6 randomly selected
12  # Sample 2: rows 2,3,4,6 randomly selected
13  # Sample 3: rows 1,2,4,5 randomly selected
```

3 The Problem with This Approach

This line represents **the fundamental flaw** mentioned in the document you referenced. Here's why:

3.1 What It's Doing (WRONG)

```
1  # Just random sampling - breaks treatment/covariate relationships
2  slice_sample(data_trimmed, n = sampsize)
```

3.2 What It Should Be Doing (CORRECT)

```
1  # Proper permutation - maintains treatment balance but breaks covariate
       associations
2  permuted_data <- data_trimmed
3  permuted_data[,treat] <- sample(data_trimmed[,treat])   # Permute
      treatment assignments
```

4 Why This Matters for Funnel Plots

The funnel plot is supposed to show what happens under the **null hypothesis** (no treatment effect).

- **Current approach**: Creates random subsets → doesn't properly simulate null hypothesis

- **Correct approach**: Permutes treatment labels → properly simulates null hypothesis while maintaining data structure

5 Visual Analogy

Think of it like shuffling cards:

- **Current (wrong) method**:
    - Take random cards from the deck for each hand
    - Some hands might have all aces, others all twos

- **Correct method**:
    - Keep the same cards in each hand
    - But randomly reassign which player gets which hand
    - Maintains the balance but breaks the original associations

6 Impact on Your Results

Using the current approach means your funnel boundaries are likely **too narrow** because random sampling doesn't properly represent the null distribution of treatment effects. This could lead to:

1. **False discoveries**: Subgroups appearing significant when they're not

2. **Incorrect confidence regions**: Funnel doesn't represent true variability under null hypothesis

The fix would replace this line with proper permutation logic as described in the document's improvement strategy.