The Consequences of Cluster Randomization in Phase III Clinical Trial Interim Analyses

Anna Mennenga, Grand Valley State University
Monica Ahrens, University of Iowa
Dr. Eric Foster, Assistant Professor (Clinical), Dept. of Biostatistics, University of Iowa

Overview

Our goal was to explore the consequences of using traditional interim analysis methods in the presence of cluster randomized clinical trial data.

- Background
  - Interim analyses
  - Cluster randomization
- Procedure
- Results
- Conclusions and further plans

Interim Analyses

- Data is analyzed before data collection is completed
- Performed a set number of times throughout a clinical trial
- If there is overwhelming efficacy or futility, one stops the trial early for ethical reasons

O'Brien-Fleming Boundaries

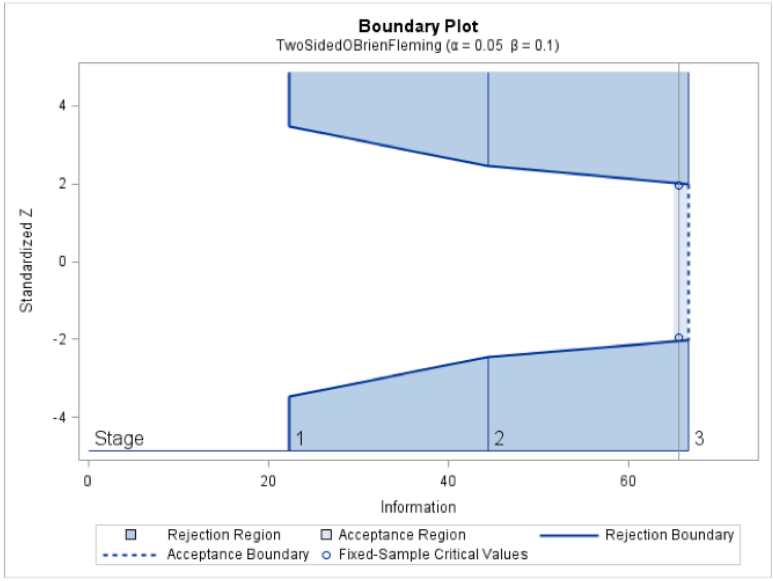
- The boundaries for rejecting the null hypothesis need to be adjusted in order to account for multiple looks
- This method results in a smaller chance of stopping early and maintains a pre-defined level of Type I error

$$|Z_k| \geq C_{OF}(K,\alpha)\sqrt{\frac{K}{k}}$$

$$|Z_1| \geq 3.471 \qquad k = 1,2,3,\ldots,K$$

$$|Z_2| \geq 2.454$$

$$|Z_3| \geq 2.004$$



Sample Size with O'Brien-Fleming Adjustment

- Multiple looks will also impact the power of the study thus there is a requirement to adjust the sample size when planning for interim monitoring

$\alpha = 0.05$

$\beta = 0.1$

$\sigma^2 = 1$

$\text{effect size} = 0.5$

$K = 3$

$$n_f \geq \frac{(Z_{(1-\alpha/2)} + Z_{(1-\beta)})^2 2\sigma^2}{(\mu_1 - \mu_2)^2} = 84.08$$

$$n \geq n_f R_{OF}(K, \alpha, \beta) = 85.43$$

$$N \geq 2n = 170.85 \approx 172$$

Cluster Randomization

- Population is split into groups, from a regional, institutional, or clinic level, and you randomize from those groups instead of the individuals.
- The goal is to recruit subjects at the same rate throughout the trial, but this isn't always the most likely outcome
- There is added correlation among subjects recruited within the same cluster, which isn't accounted for in traditional methods for interim analyses

Cluster Randomization Sample Size Adjustment

- Sample size is adjusted to account for inflated variance in cluster sampling

Sample size adjustment :

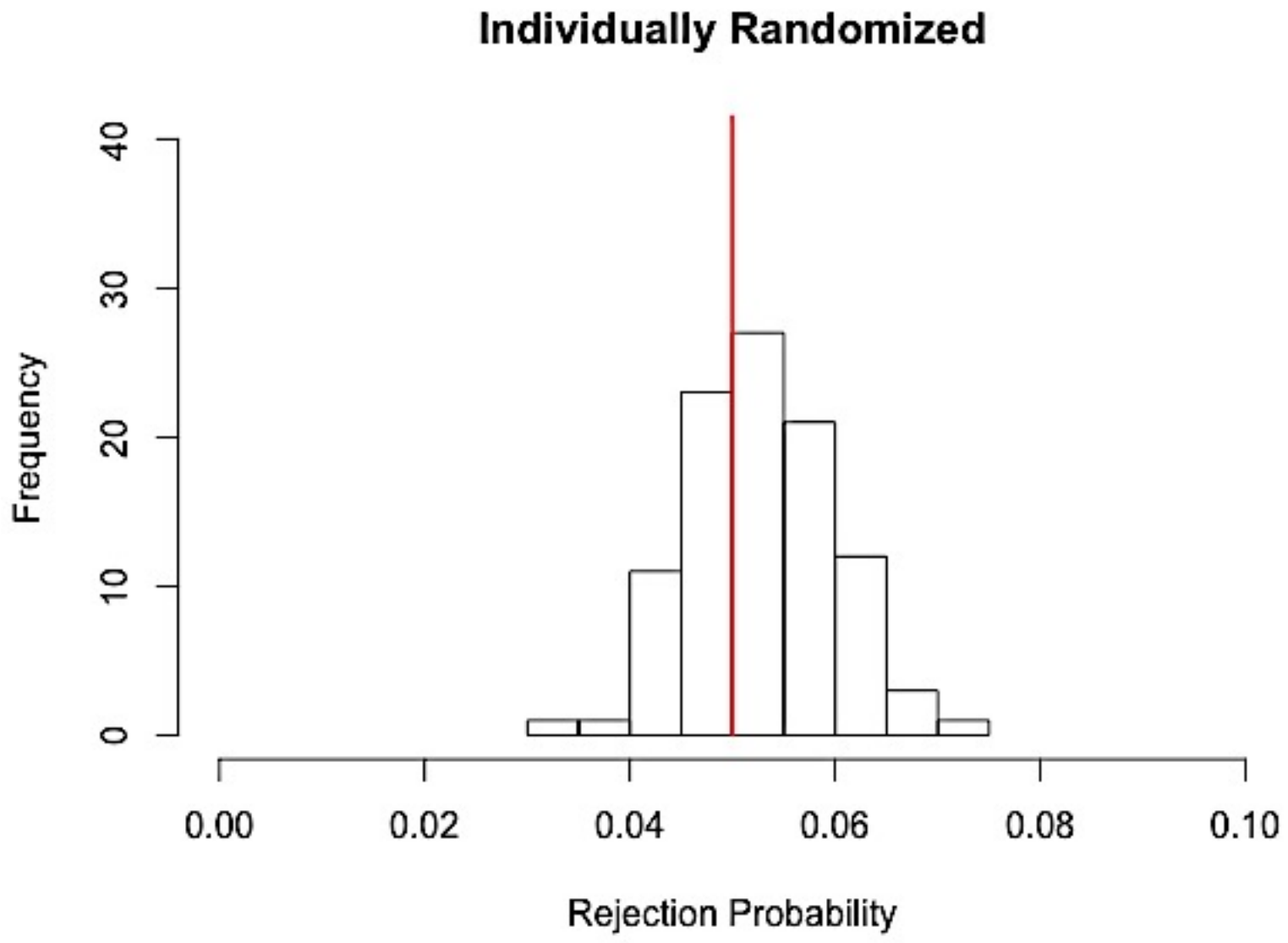
ρ = intraclass correlation coefficient = 0.005

m = number of clusters = 10

$n_c \geq n(1+(m-1)\rho) = 178.54 \approx 210$

Simulation: Randomization at the Individual Level

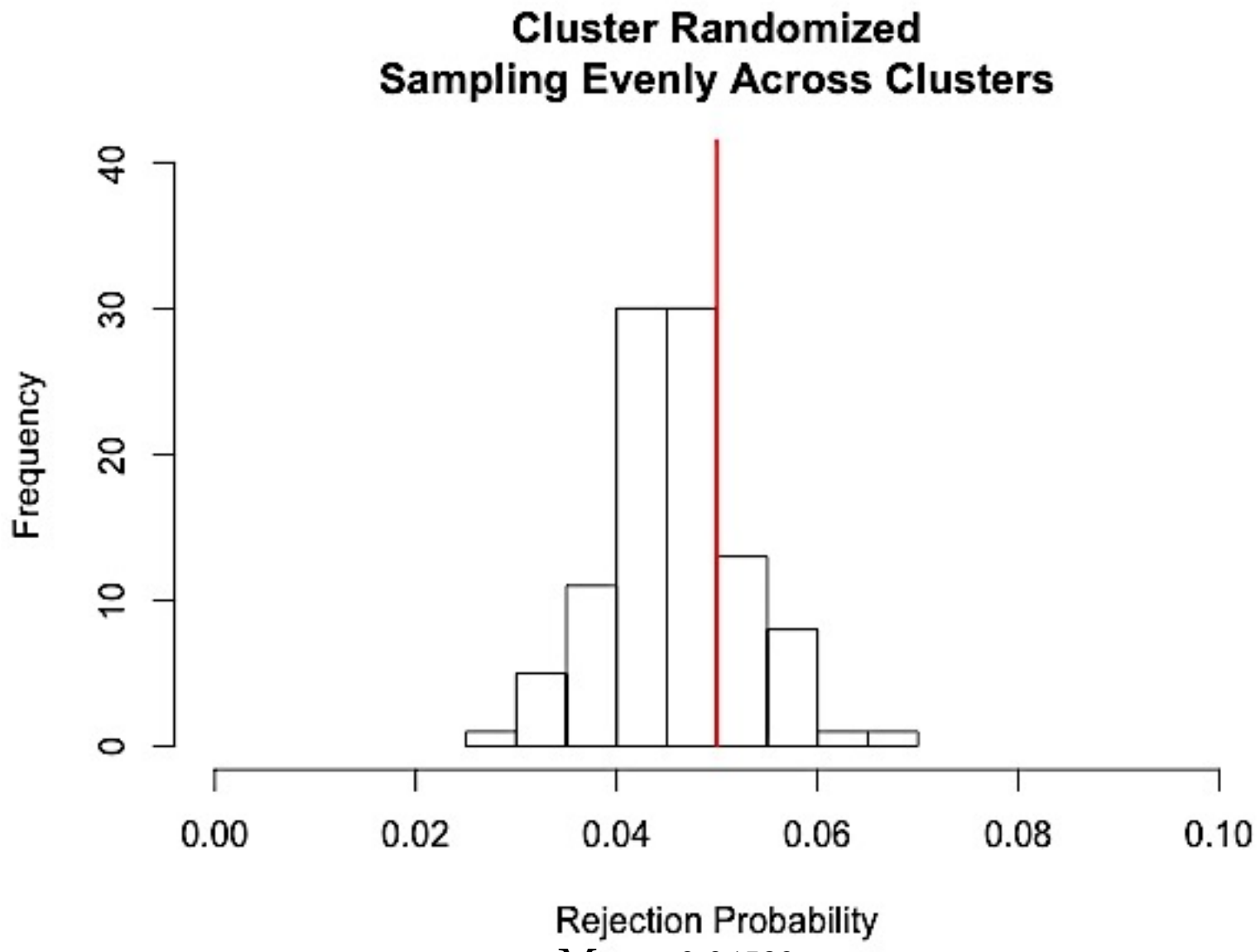
- For the case where randomization is performed at the individual level, we:
  - Recruit subjects until one third are randomized
  - Subjects are recruited evenly for both treatment groups
  - The data for both groups are simulated from a standard normal distribution
  - Therefore, we are assessing the Type I error of the standard interim analysis methodology
  - At each interim analysis, we perform a two sample t-test
- This was repeated 1,000 times and the resulting proportion of times we rejected the null hypothesis was recorded.
  - We repeated the above steps 100 times in order to build a histogram of the proportion of times the null hypothesis was rejected


Individually Randomized
Frequency
0
10
20
30
40
0.00
0.02
0.04
0.06
0.08
0.10
Rejection Probability

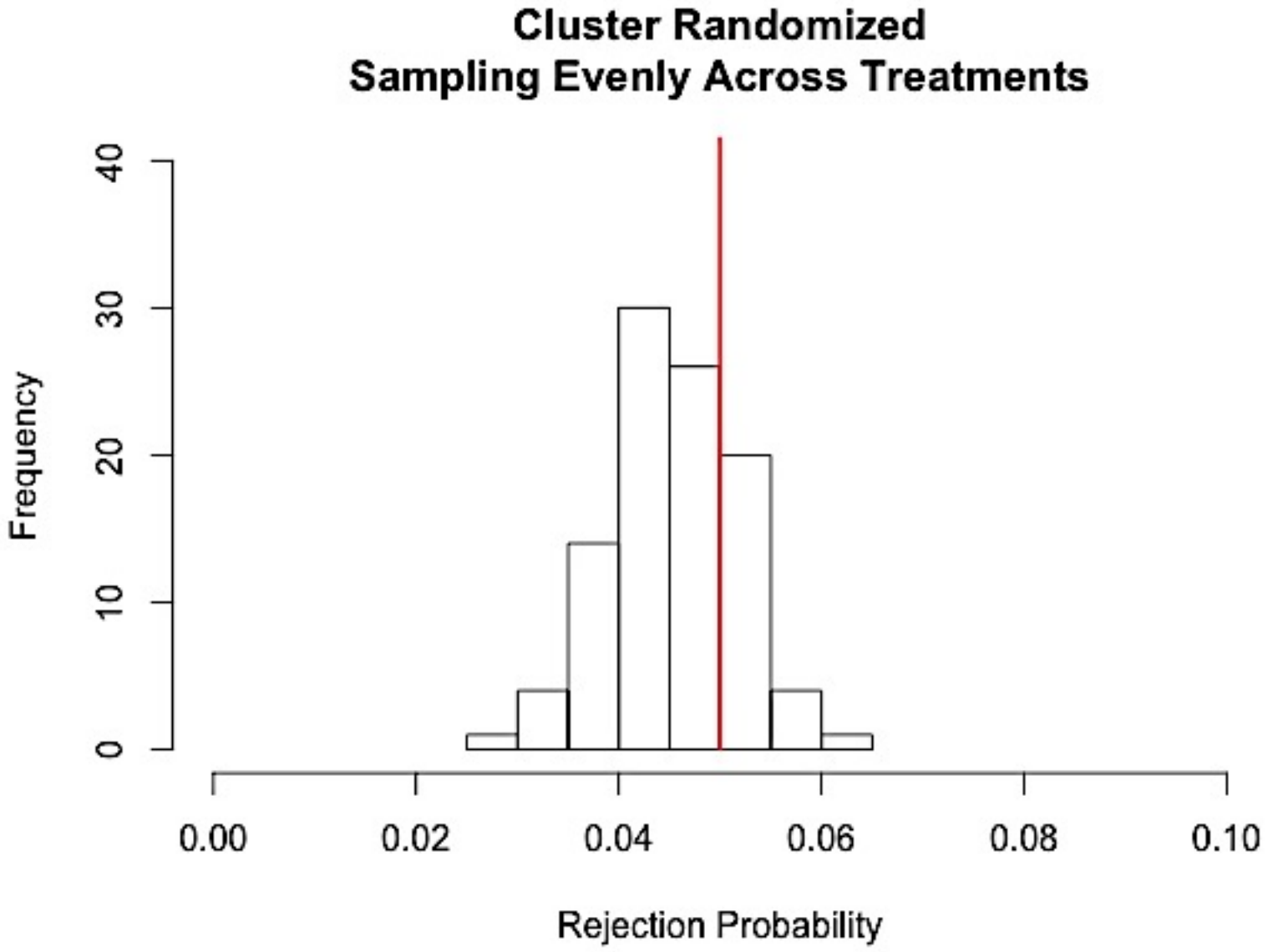
Mean: 0.05299

Simulation: Randomization at the Cluster Level

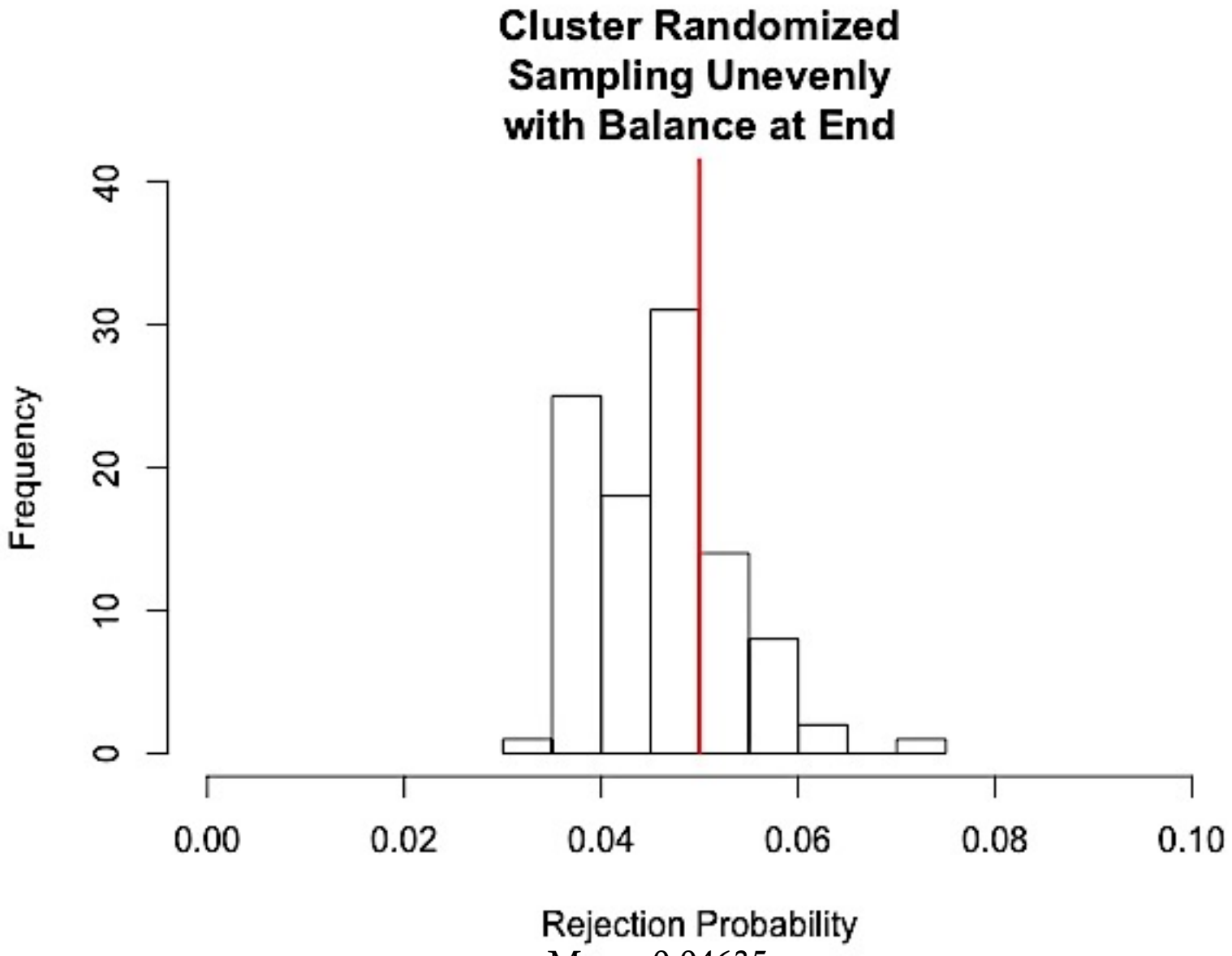
- The previous simulation parameters were again applied, but this time to data that has arisen from cluster randomized clinical trials
- We explore data accrual in three ways:
  - Sampling evenly across clusters
    - Throughout trial each cluster recruited subjects at the same rate
  - Sampling evenly across treatments
    - Throughout trial treatment and placebo groups recruited at the same rate, but clusters did not
  - Sampling unevenly, but with balance at end
    - During trial treatment groups and clusters didn't recruit at the same rate, but at the end of the trial they ended with equal amounts of subjects

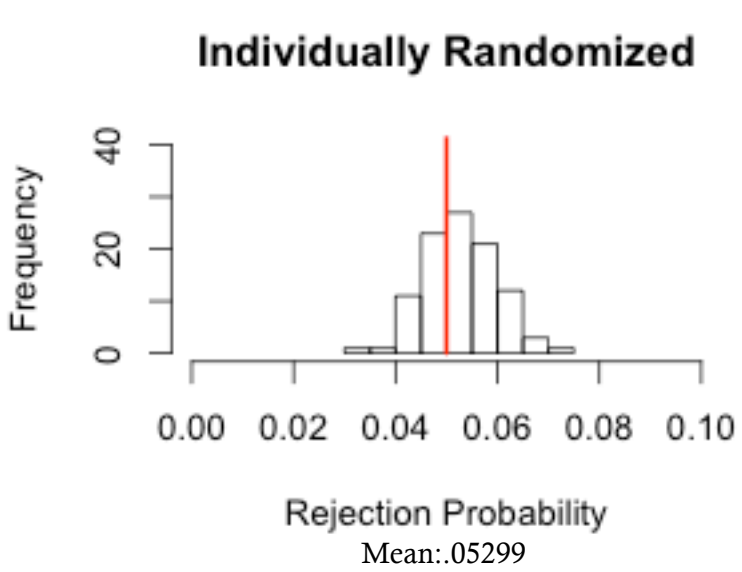

Cluster Randomized
Sampling Evenly Across Clusters
Frequency
0
10
20
30
40
0.00
0.02
0.04
0.06
0.08
0.10
Rejection Probability

Mean: 0.04592

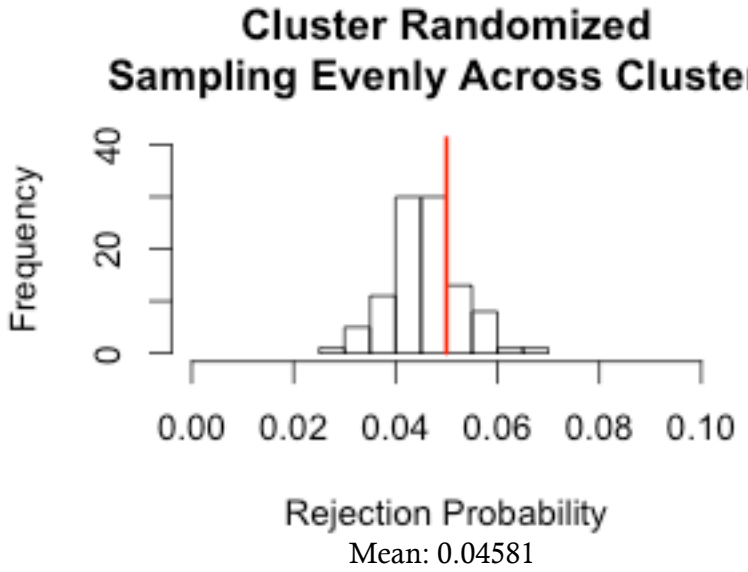

Cluster Randomized
Sampling Evenly Across Treatments
Frequency
0
10
20
30
40
0.00
0.02
0.04
0.06
0.08
0.10
Rejection Probability

Mean: 0.04581

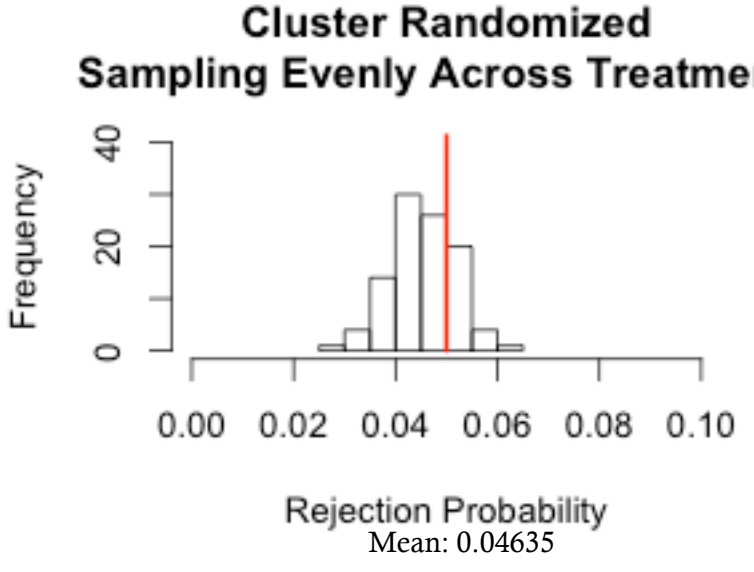

Cluster Randomized
Sampling Unevenly
with Balance at End
Frequency
0
10
20
30
40
0.00
0.02
0.04
0.06
0.08
0.10
Rejection Probability
Mean: 0.04635



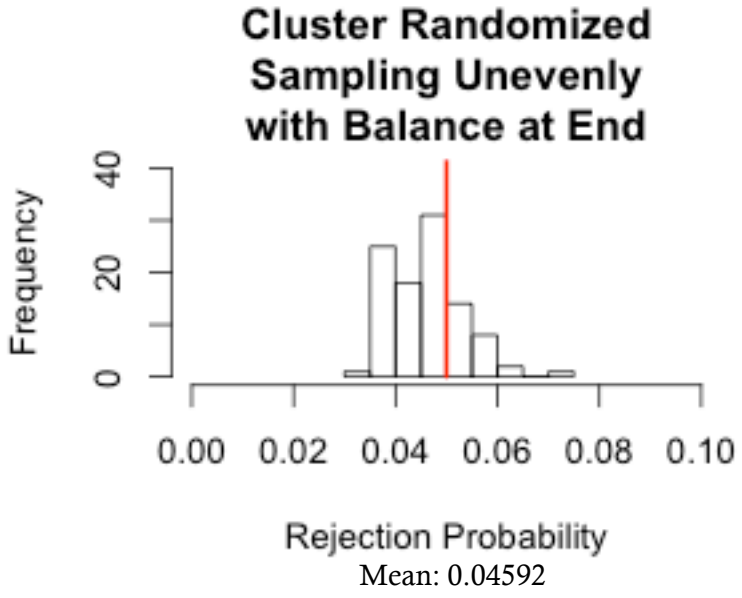
Mean:.05299



Mean: 0.04581



Mean: 0.04635



Mean: 0.04592

Results

Analysis	Individual Randomization	Cluster Randomization		
		Evenly Across Clusters	Evenly Across Treatments	Unevenly, but with Balance at End of Trial
Interim Analysis 1	.00092	0.00075	.00061	.00073
Interim Analysis 2	.01513	0.01267	.01263	.01280
Interim Analysis 3	.03964	0.0325	.03257	.03281
Overall Mean	.05299	0.04592	.04581	.04635

Conclusions

- As alpha decreases, beta increases
- As alpha decreases, power decreases
- Since the type I error for the cluster randomization was less than the individual randomization we concluded that we are conservative when testing cluster samples
- Prevented type I error, but also decreased power
- This would be seen as a more negative thing for physicians. Physicians want higher power in order to find significance in working treatments
- Statisticians want to be conservative so they're not finding significance in ineffective treatments
- **Future Research:**
  - Look into adjusting for loss of power when dealing with interim analyses on cluster randomized trials

Acknowledgements

- Eric Foster, thank you for working with us on the project, we had a lot of fun!
- Gideon Zamba
- Terry Kirk
- John VanBuren, Monelle Tamegnon and the rest of the graduate students
- Department of Biostatistics
- NIH

Questions?