Optimal Treatment Dispersions in Rectangular Areas

Peter Claussen
Gylling Data Management

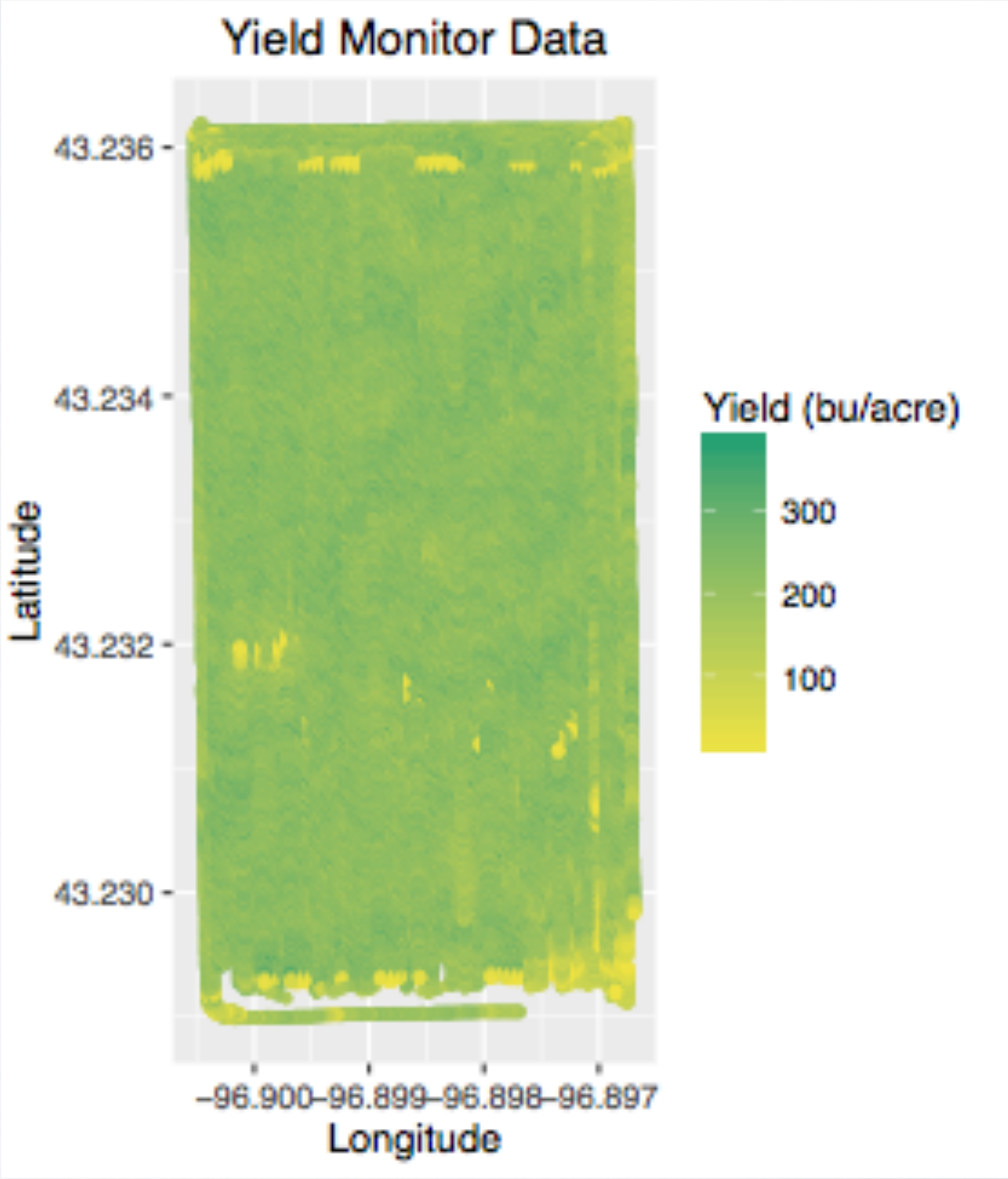
- Credit for the phrase "experiments in rectangular areas" goes to R. A. Bailey, who frequently uses the term "undesirable" layout.

    - "Experiments in rectangular areas: design and randomization." ***Journal of Agricultural, Biological, and Environmental Statistics*** 17.2 (2012): 176-191.

- What is an undesirable layout? Let us consider the effects of different layouts on simulated uniformity trials.

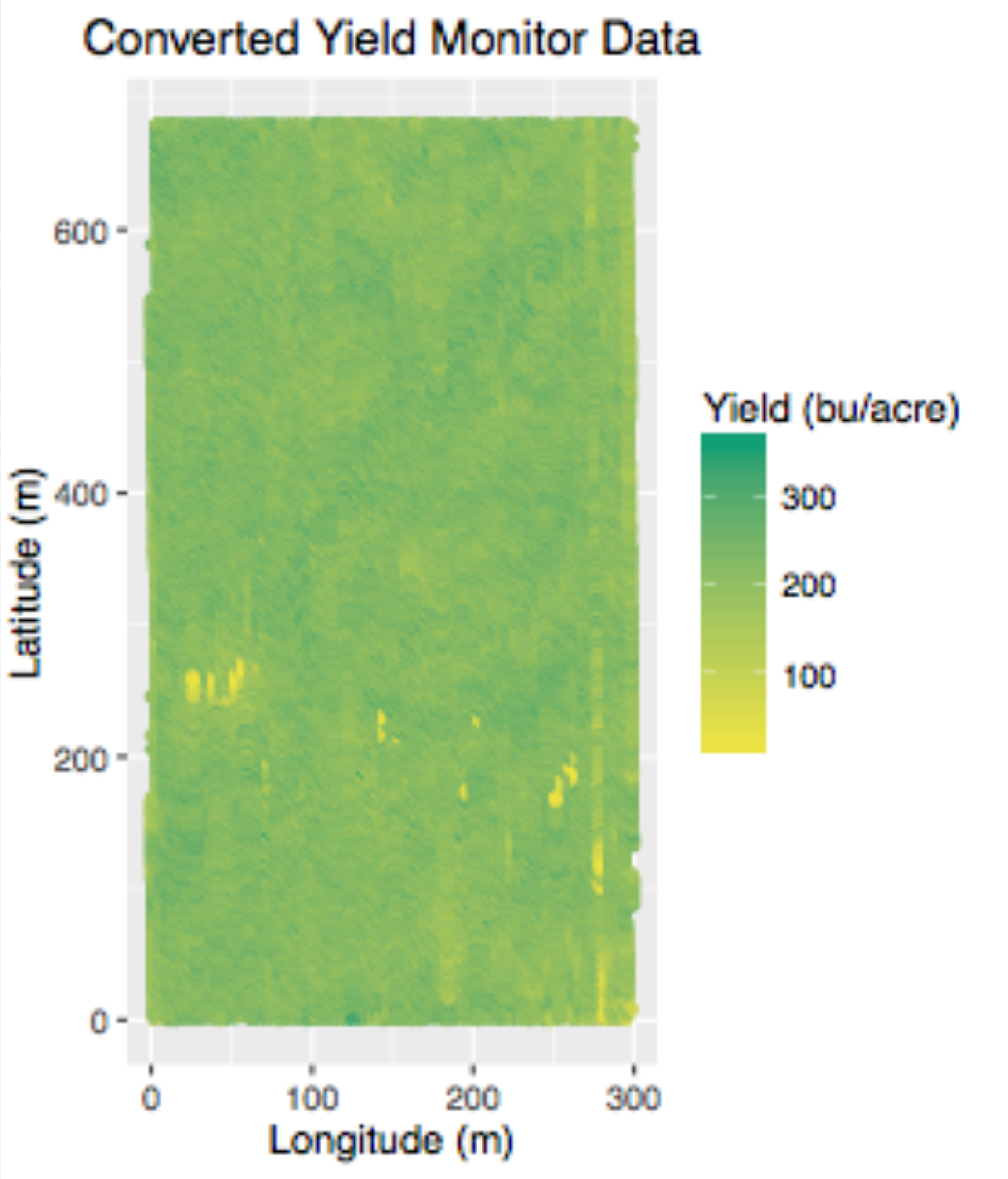
Simulated Uniformity Trials

- Start with yield monitor data.
  - South East Research Station, Beresford SD
  - 2013
  - Maize



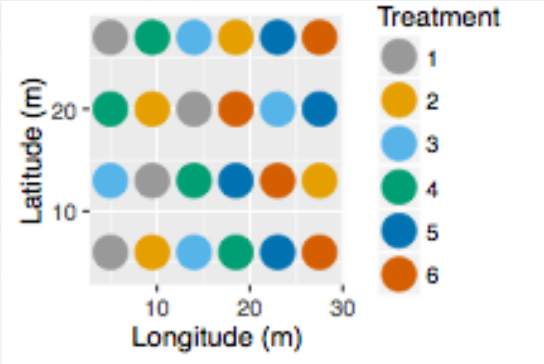
Simulated Uniformity Data

- Trim to remove end-rows and edges
- Convert longitude and latitude coordinates to meters, relative to south-west corner.

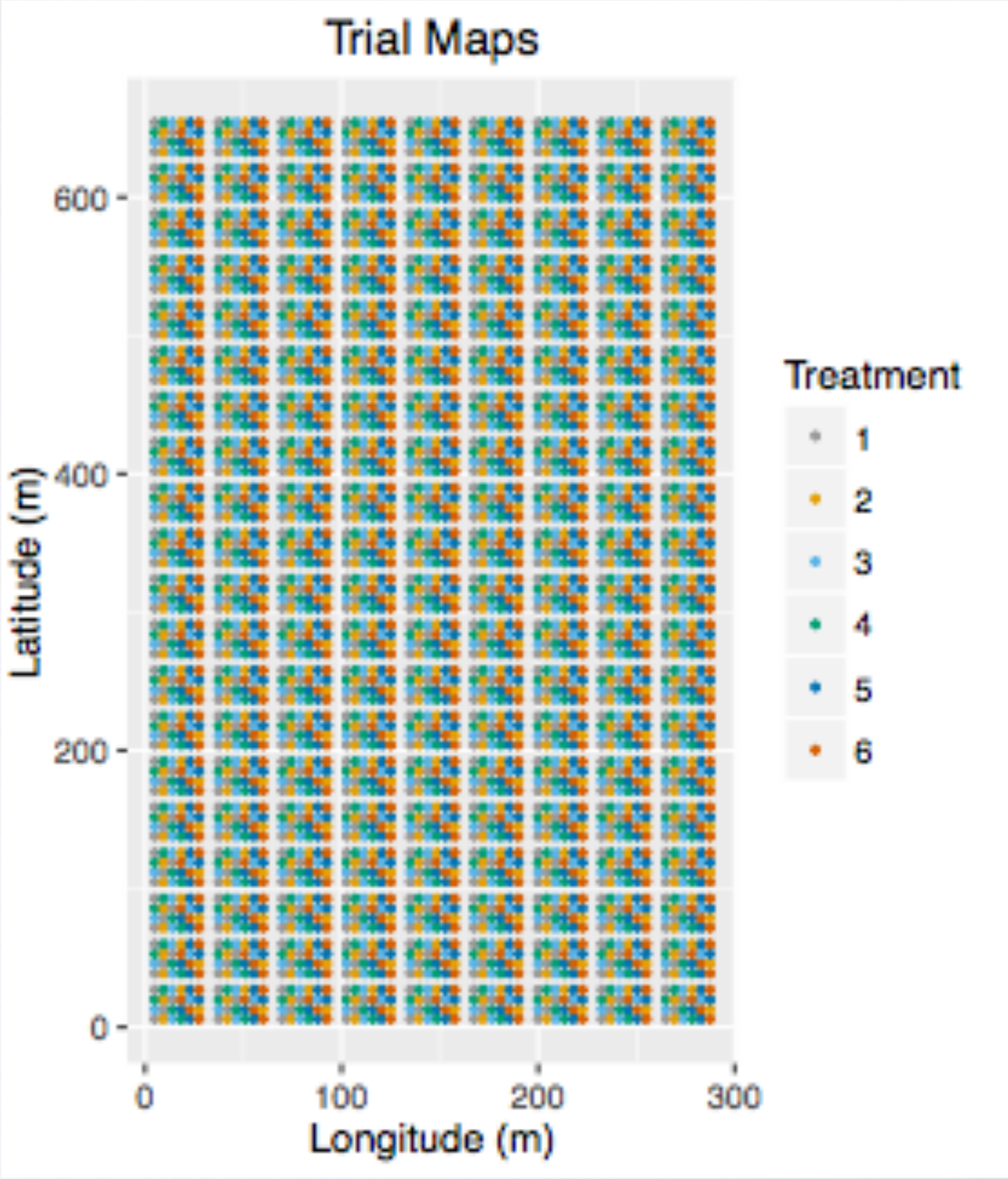


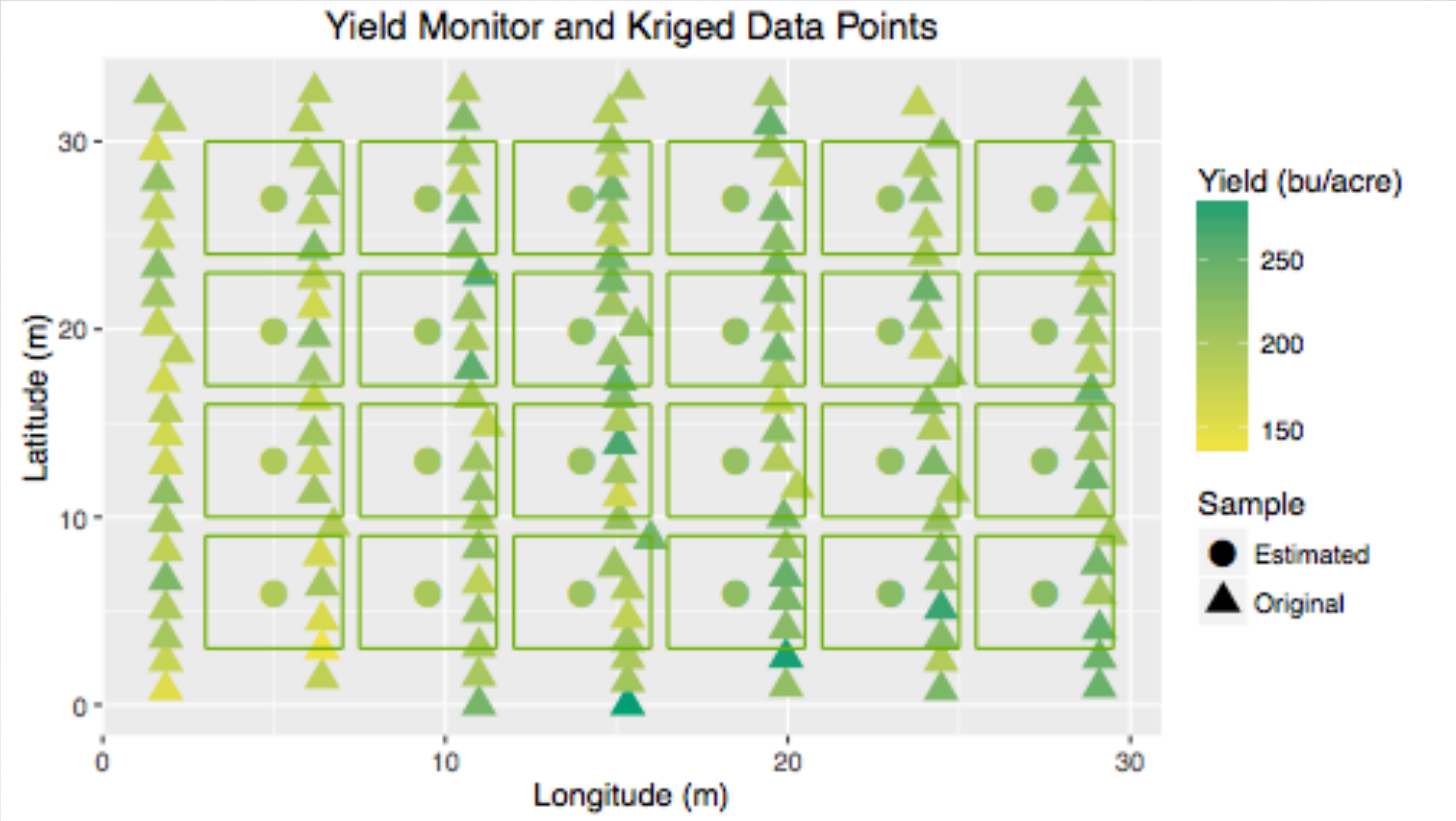
Simulated Repeated Trial Map

- Generate a trial map
  - RCB, 6 treatments and 4 replicates.



- Superimpose over field, starting in lower left corner and adding trials in rows and columns



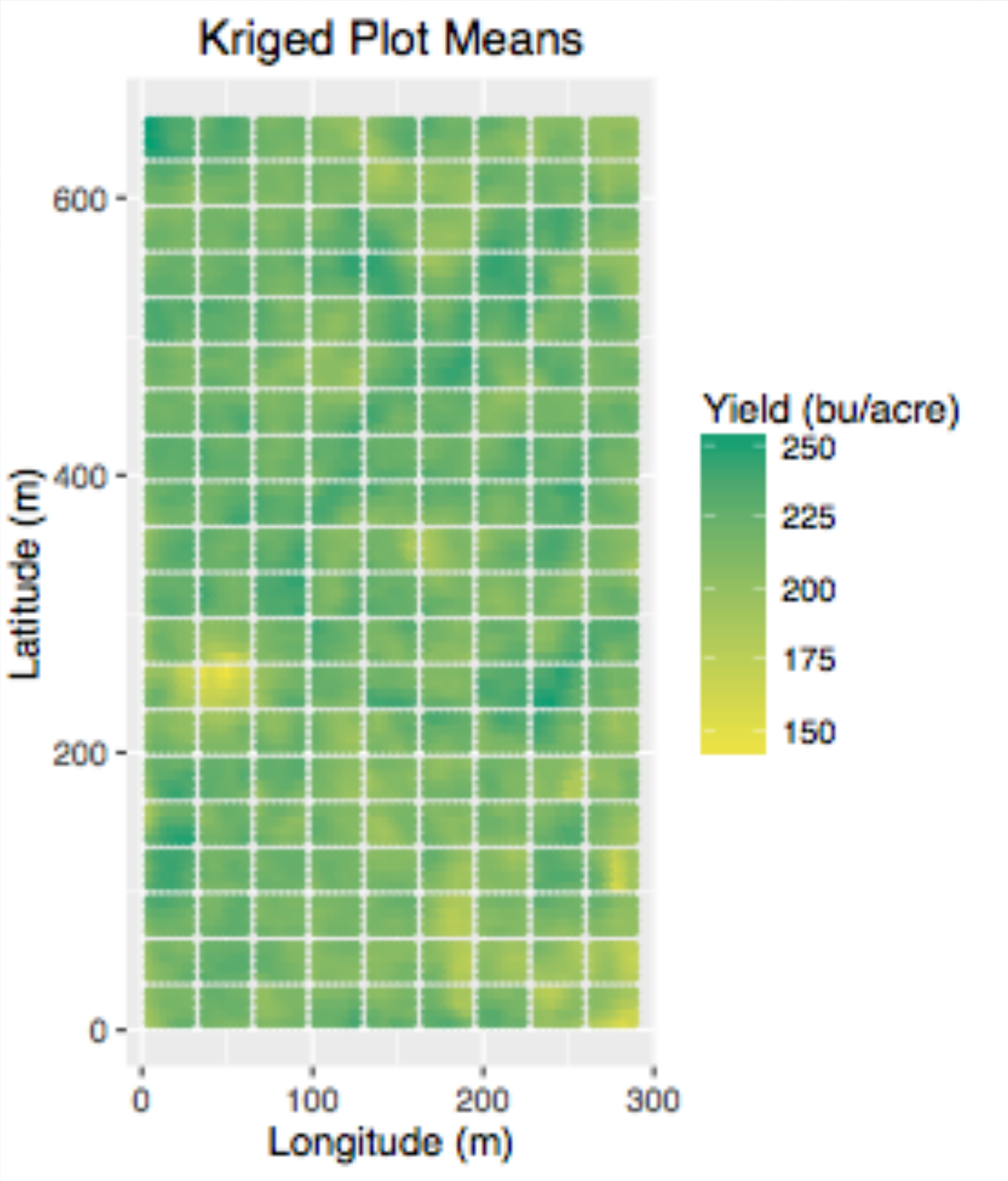


Estimated Plot Yields

- Interpolate yield at center of plot by kriging yield monitor data. Boxes represent plot borders.

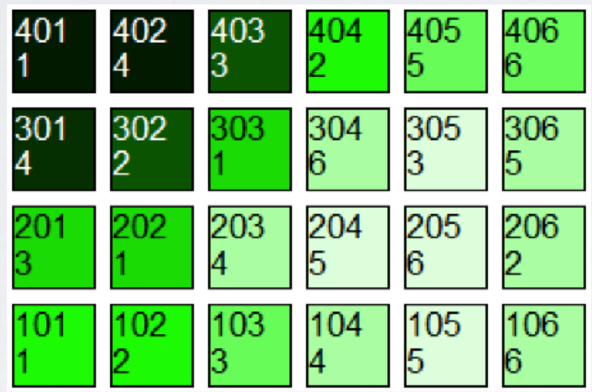
180 Simulated Uniformity Trials

- Repeat for all layouts and analyze each as a different uniformity trial.
- Each trial samples a different part of a large spatial structure.
- We can compare how different randomizations detect spatial variation.
- For our purposes, "undesirable" layouts will confound spatial variability with treatment effects.



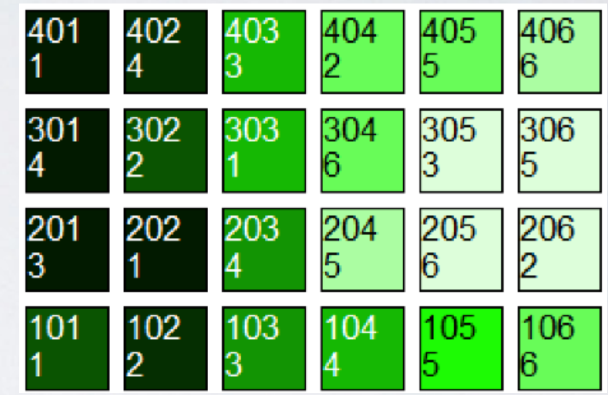
Types of Confounding

<table>
<tr><th></th><th>Significant Replicate</th><th>Non-Significant Replicate</th></tr>
<tr><td>Significant Treatment</td><td>



<table>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F</th><th>Prob(F)</th></tr>
<tr><td>Total</td><td>23</td><td>1418.823599</td><td></td><td></td><td></td></tr>
<tr><td>Replicate</td><td>3</td><td>433.208148</td><td>144.402716</td><td>4.520</td><td>0.0189</td></tr>
<tr><td>Treatment</td><td>5</td><td>506.441848</td><td>101.288370</td><td>3.171</td><td>0.0377</td></tr>
<tr><td>Error</td><td>15</td><td>479.173603</td><td>31.944907</td><td></td><td></td></tr>
</table>

</td><td>

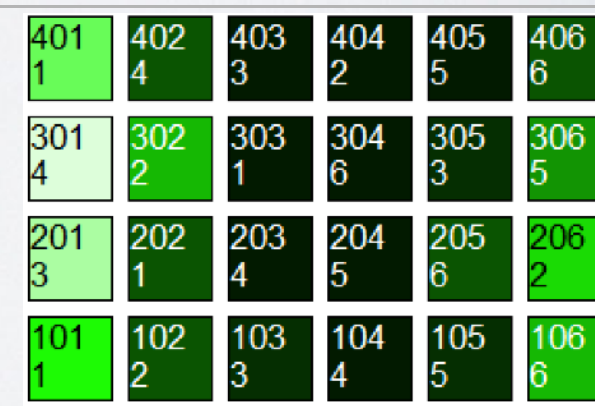


<table>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F</th><th>Prob(F)</th></tr>
<tr><td>Total</td><td>23</td><td>1394.230940</td><td></td><td></td><td></td></tr>
<tr><td>Replicate</td><td>3</td><td>32.934429</td><td>10.978143</td><td>0.268</td><td>0.8471</td></tr>
<tr><td>Treatment</td><td>5</td><td>747.858169</td><td>149.571634</td><td>3.657</td><td>0.0231</td></tr>
<tr><td>Error</td><td>15</td><td>613.438341</td><td>40.895889</td><td></td><td></td></tr>
</table>

</td></tr>
<tr><td>Non-Significant Treatment</td><td>

<table>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F</th><th>Prob(F)</th></tr>
<tr><td>Total</td><td>23</td><td>1182.769646</td><td></td><td></td><td></td></tr>
<tr><td>Replicate</td><td>3</td><td>919.308124</td><td>306.436041</td><td>18.435</td><td>0.0001</td></tr>
<tr><td>Treatment</td><td>5</td><td>14.129328</td><td>2.825866</td><td>0.170</td><td>0.9698</td></tr>
<tr><td>Error</td><td>15</td><td>249.332194</td><td>16.622146</td><td></td><td></td></tr>
</table>

</td><td>

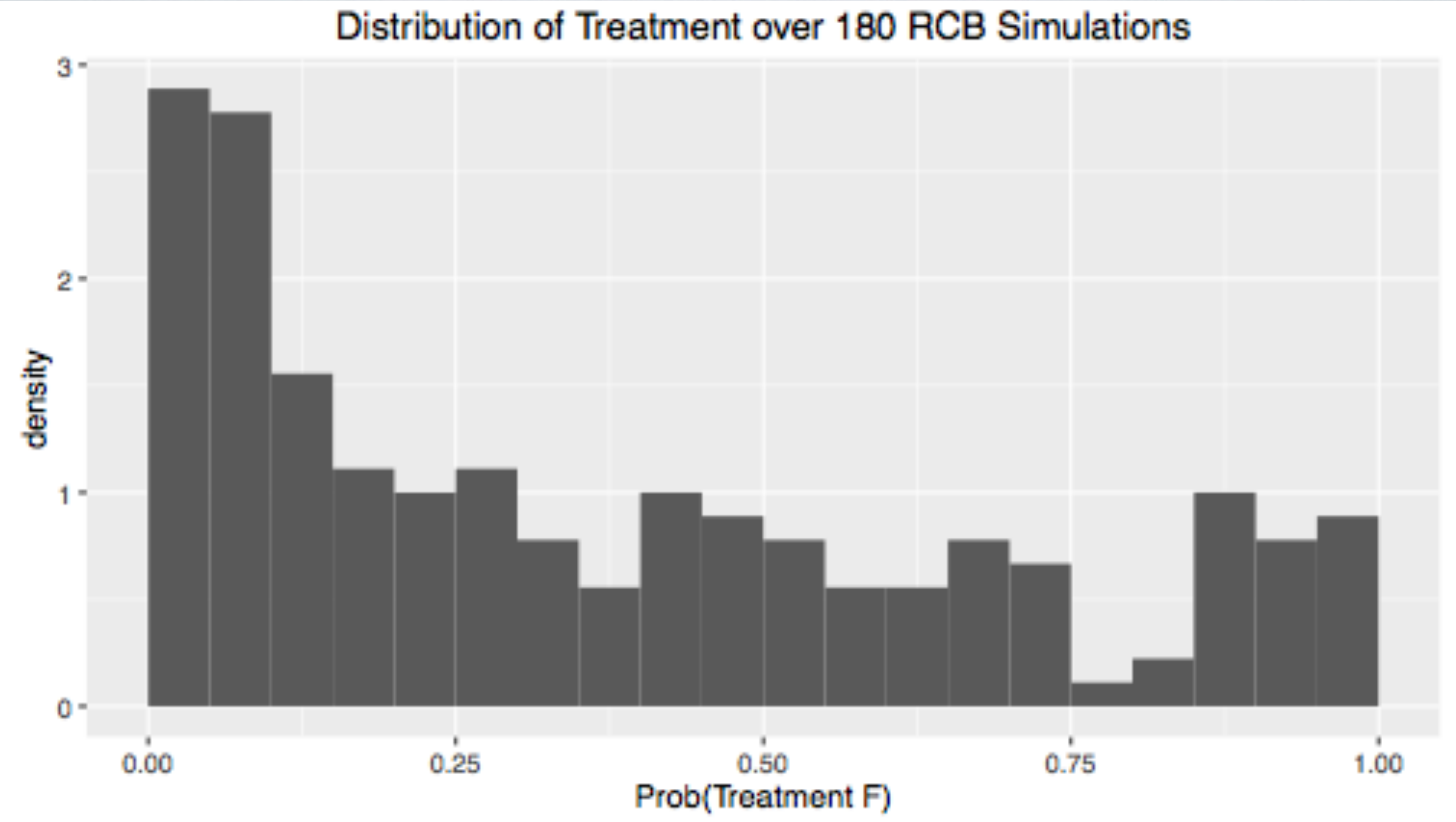


<table>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F</th><th>Prob(F)</th></tr>
<tr><td>Total</td><td>23</td><td>1252.097954</td><td></td><td></td><td></td></tr>
<tr><td>Replicate</td><td>3</td><td>26.101813</td><td>8.700604</td><td>0.117</td><td>0.9489</td></tr>
<tr><td>Treatment</td><td>5</td><td>107.615609</td><td>21.523122</td><td>0.289</td><td>0.9119</td></tr>
<tr><td>Error</td><td>15</td><td>1118.380532</td><td>74.558702</td><td></td><td></td></tr>
</table>

</td></tr>
</table>

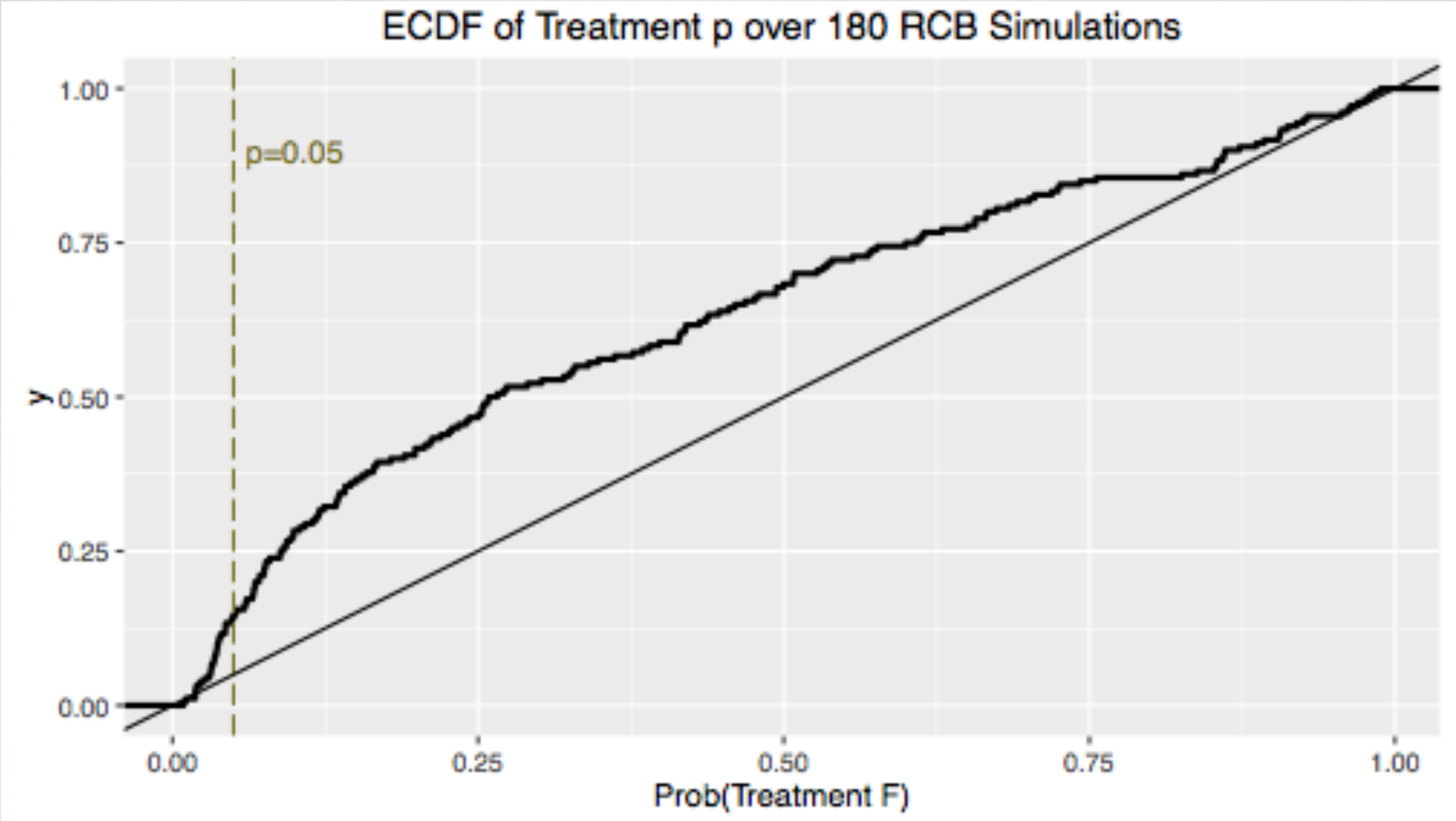
Type I Error

- Since there are no real treatment effects, analysis of variance of a uniformity trial should produce a non-significant p-value for the Treatment F statistic.
- At a nominal error rate of 5%, we can expect 1 out of 20 trials to achieve a Treatment $p<0.05$.
- In 180 simulated trials, the example RCB produced 26 trials with Treatment $p < 0.05$, for an achieved error rate of 14.44%.
- We can visualize this by plotting the distribution of p-values for these 180 trials.



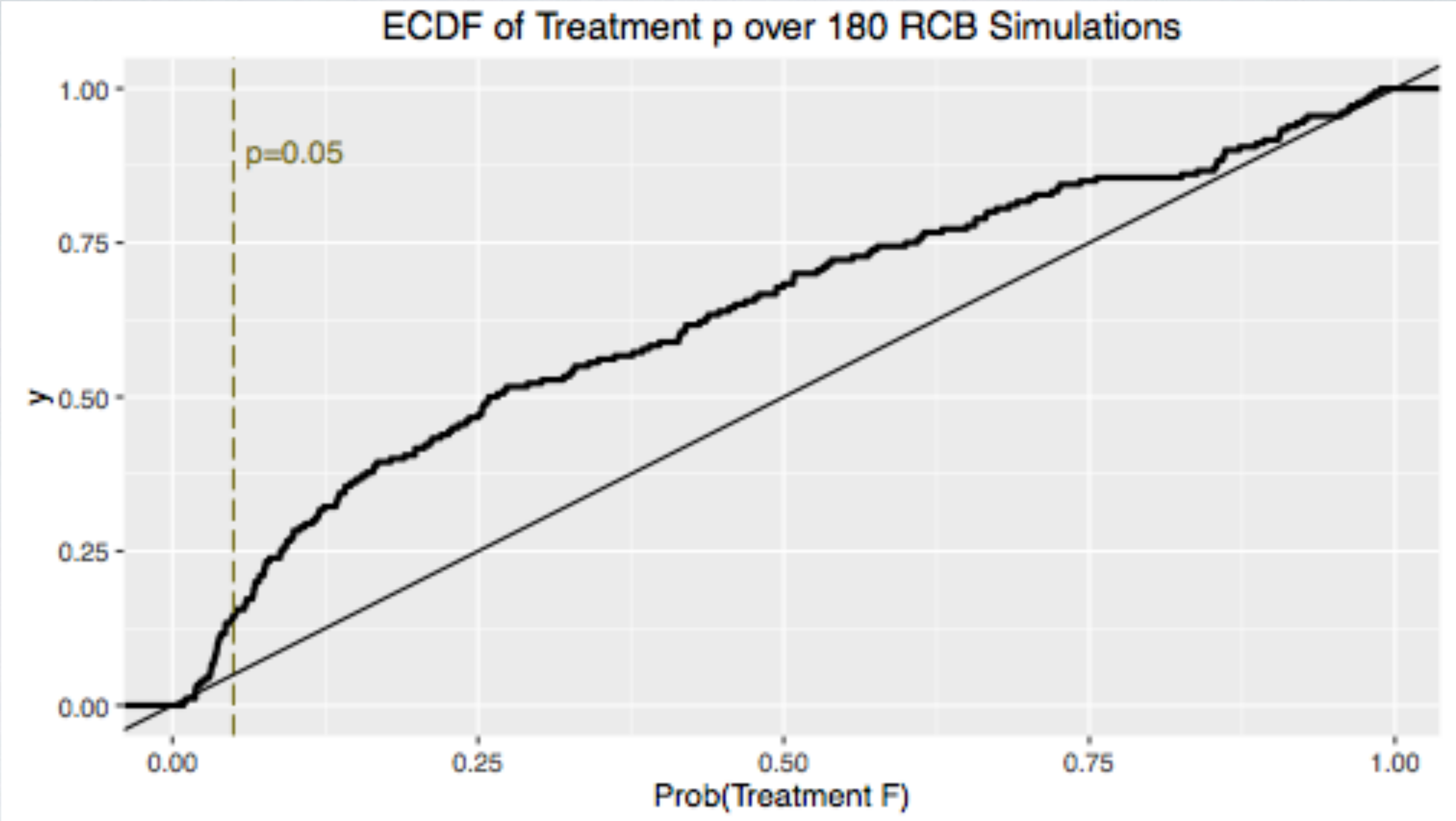
Distribution of Treatment p-values.

For an unbiased trial, we would expected the distribution of p-values to be approximately uniform. This set of trials has a greater than expected number of small p-values.



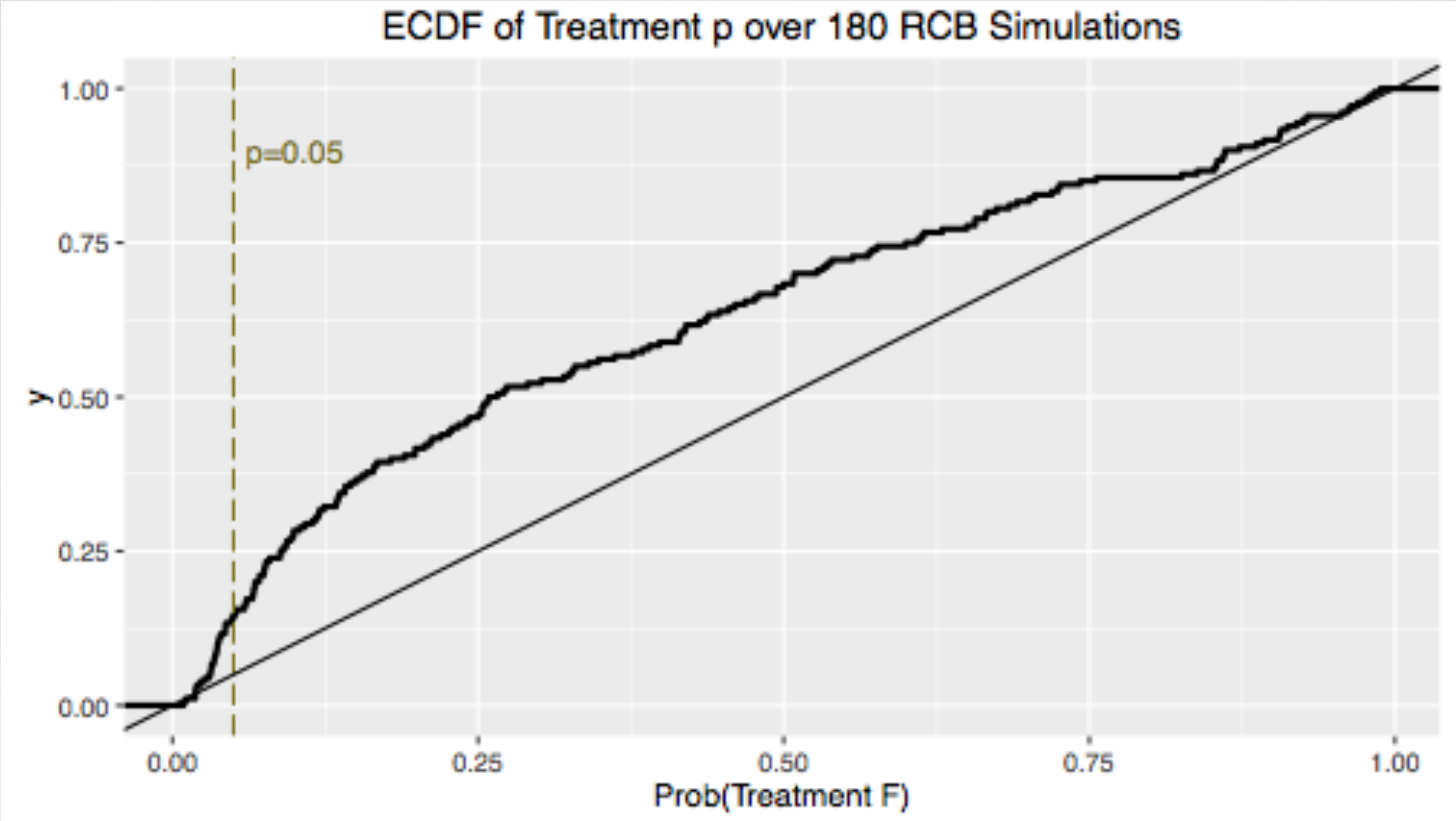
Empirical Cumulative Distribution

Same data as previous graph, but plotted as the total count of trials at or less than the nominal probability on the x-axis.



Empirical Cumulative Distribution

The diagonal line represents a uniform distribution, where the accumulated proportion of trials at a nominal probability equals that probability.

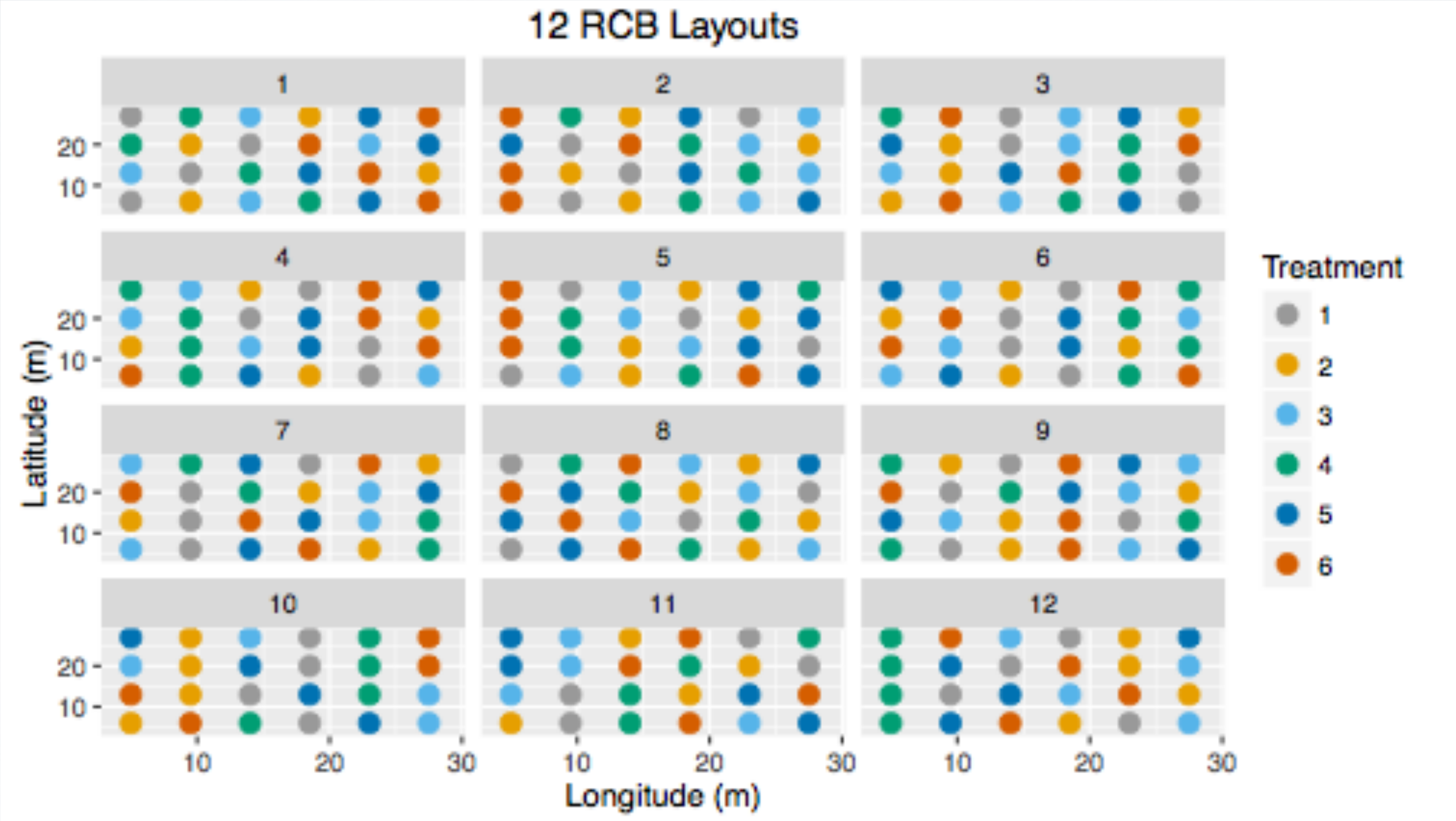


Empirical Cumulative Distribution

The ECDF at a nominal probability of 0.05 has an accumulated proportion of 0.144, implying the **achieved** Type I error rate is higher than the **nominal** error rate.

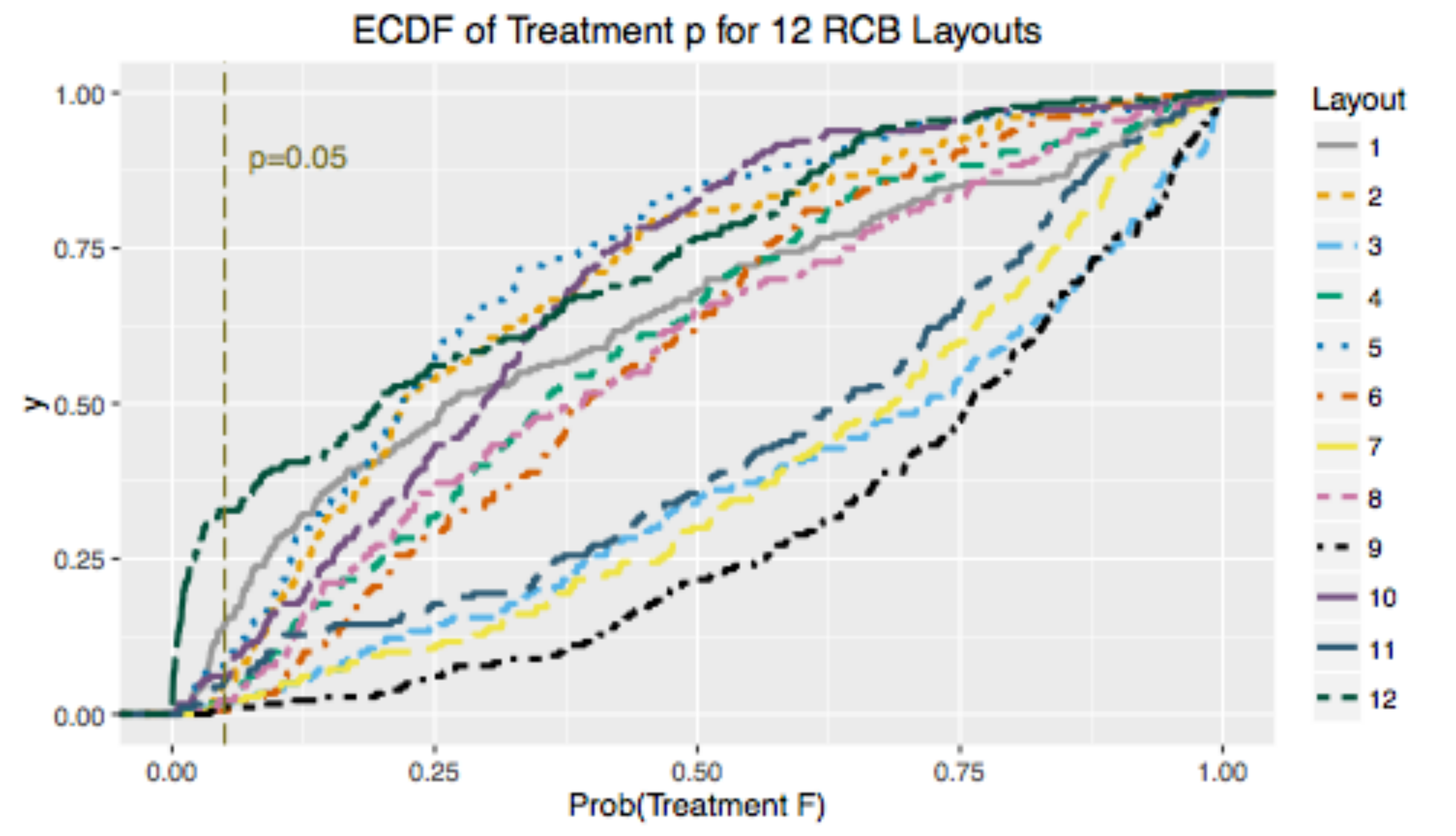
Comparing Different Randomizations

- We've focused on a single RCB experiment, repeated 180 times.
- This layout achieved a higher Type I error rate, 14.4%, than our nominal rate of 5%.
- This might simply be bad luck - the randomization we chose is "undesirable".
- We'll repeat the simulation with 11 other RCB layouts.



Twelve Proposed Layouts

Number 1 is the layout shown previously

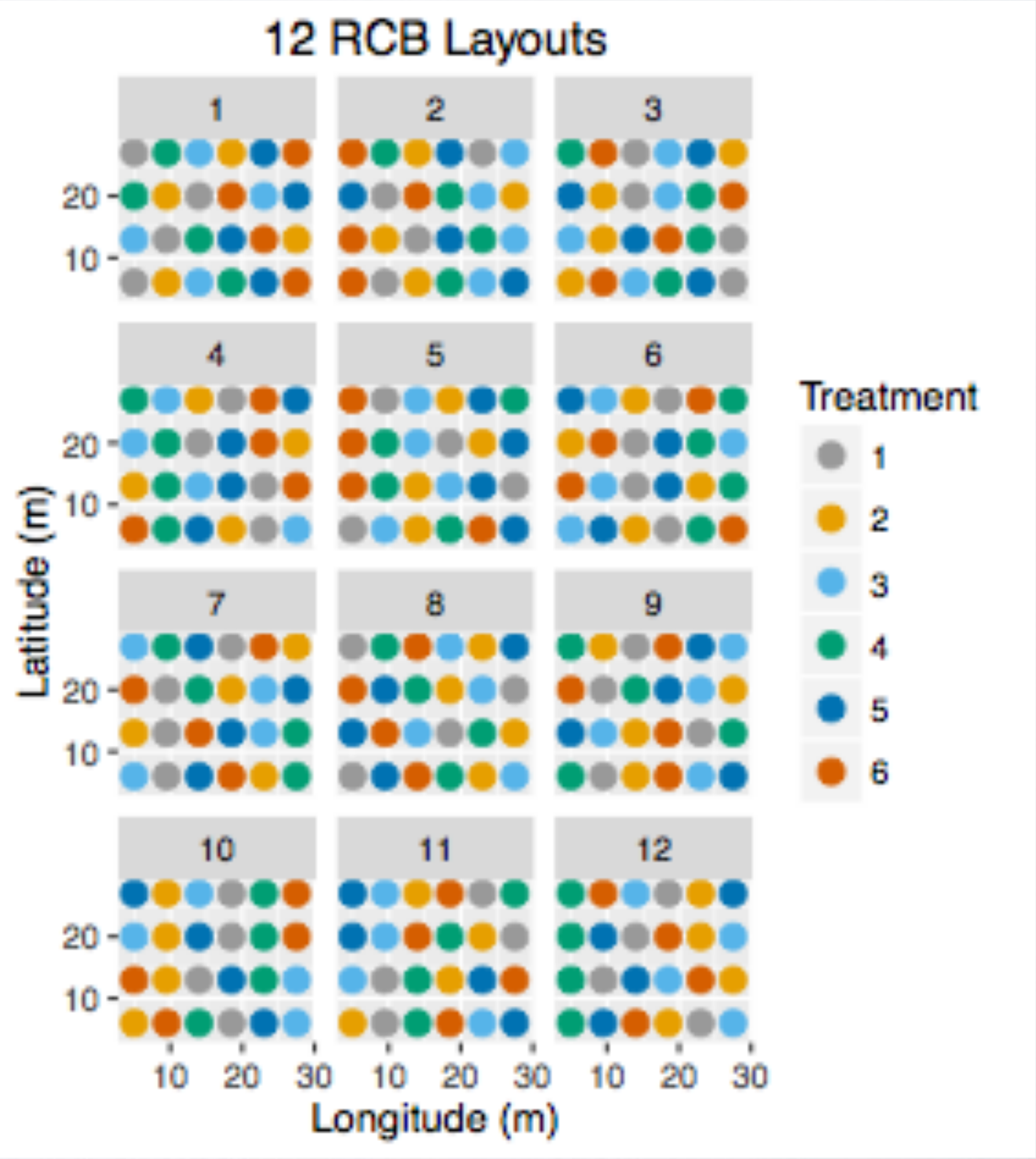


Treatment p Distributions

Some layouts are biased towards smaller p-values, some are biased towards larger p-values.

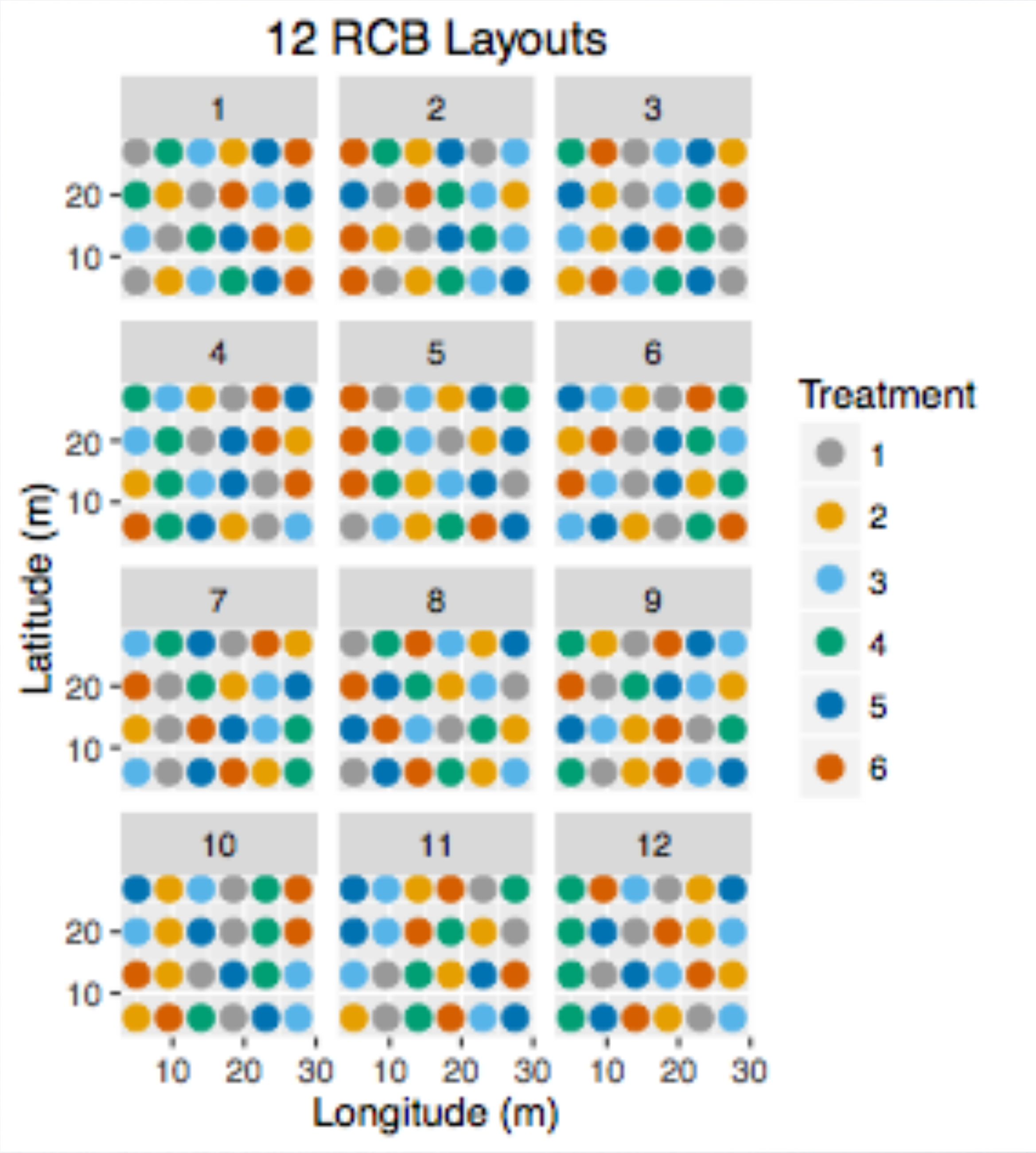
Simple Restricted Randomization

- Which of these layouts would you reject out of hand as being potentially biased?
- If you would reject any, you are practicing a form of restricted randomization.
- This can be inefficient, since it may require many randomizations to achieve a desired layout.



Comparative Error Rates

Layout	Type I Error Rate
1	14.44
2	5.00
3	1.67
4	2.22
5	8.33
6	0.56
7	2.22
8	2.22
9	1.11
10	6.11
11	5.55
12	32.8
Mean	6.85
SD	9.05



Comparative Error Rates

Layout	Type I Error Rate
1	14.44
2	5.00
3	1.67
4	2.22
5	8.33
6	0.56
7	2.22
8	2.22
9	1.11
10	6.11
11	5.55
12	32.8
Mean	6.85
SD	9.05

Layout 1

401 1	402 4	403 3	404 2	405 5	406 6
301 4	302 2	303 1	304 6	305 3	306 5
201 3	202 1	203 4	204 5	205 6	206 2
101 1	102 2	103 3	104 4	105 5	106 6

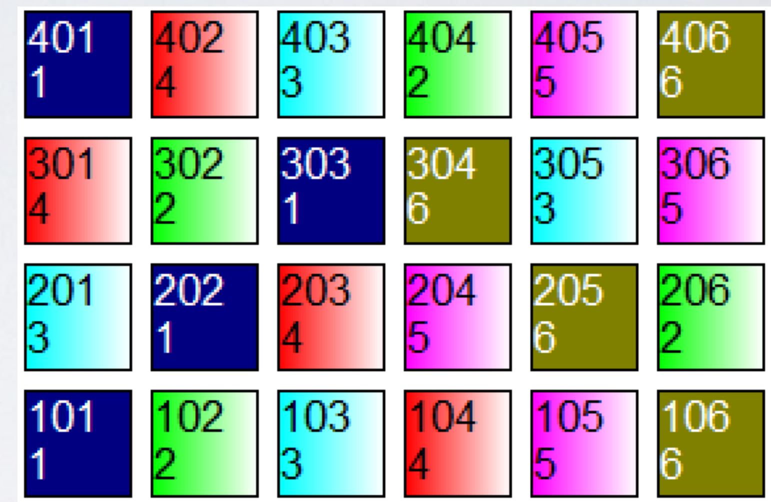
Layout 12

401 4	402 6	403 3	404 1	405 2	406 5
301 4	302 5	303 1	304 6	305 2	306 3
201 4	202 1	203 5	204 3	205 6	206 2
101 4	102 5	103 6	104 2	105 1	106 3

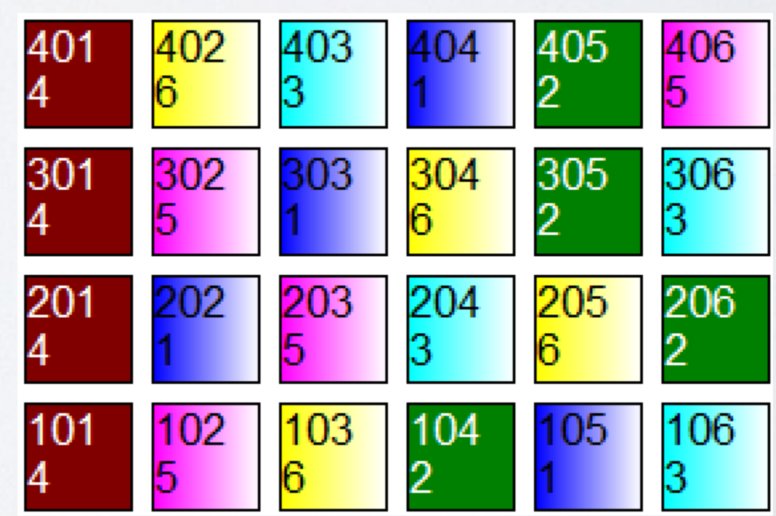
Comparative Error Rates

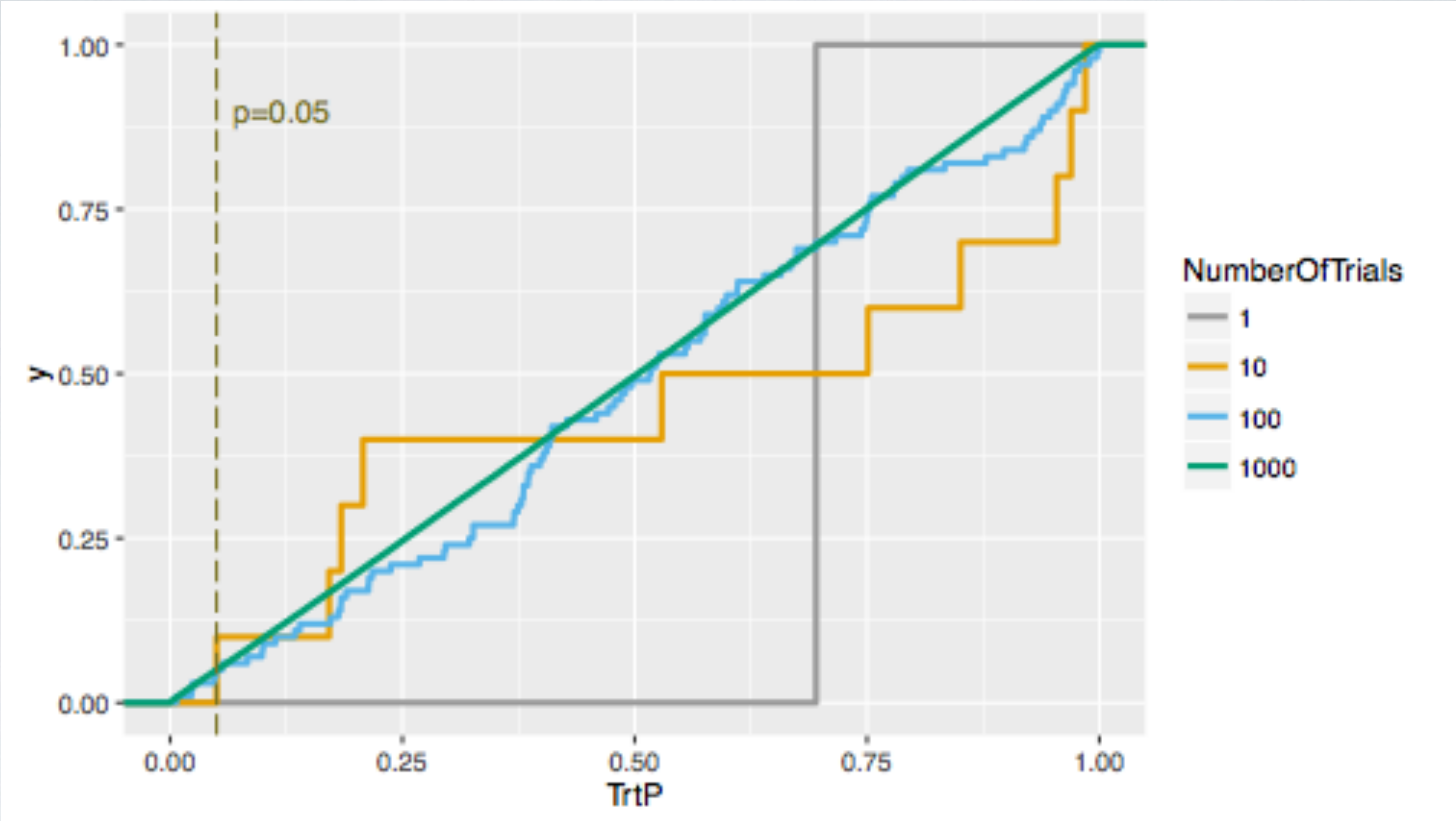
Layout	Type I Error Rate
1	14.44
2	5.00
3	1.67
4	2.22
5	8.33
6	0.56
7	2.22
8	2.22
9	1.11
10	6.11
11	5.55
12	32.8
Mean	6.85
SD	9.05

Layout 1



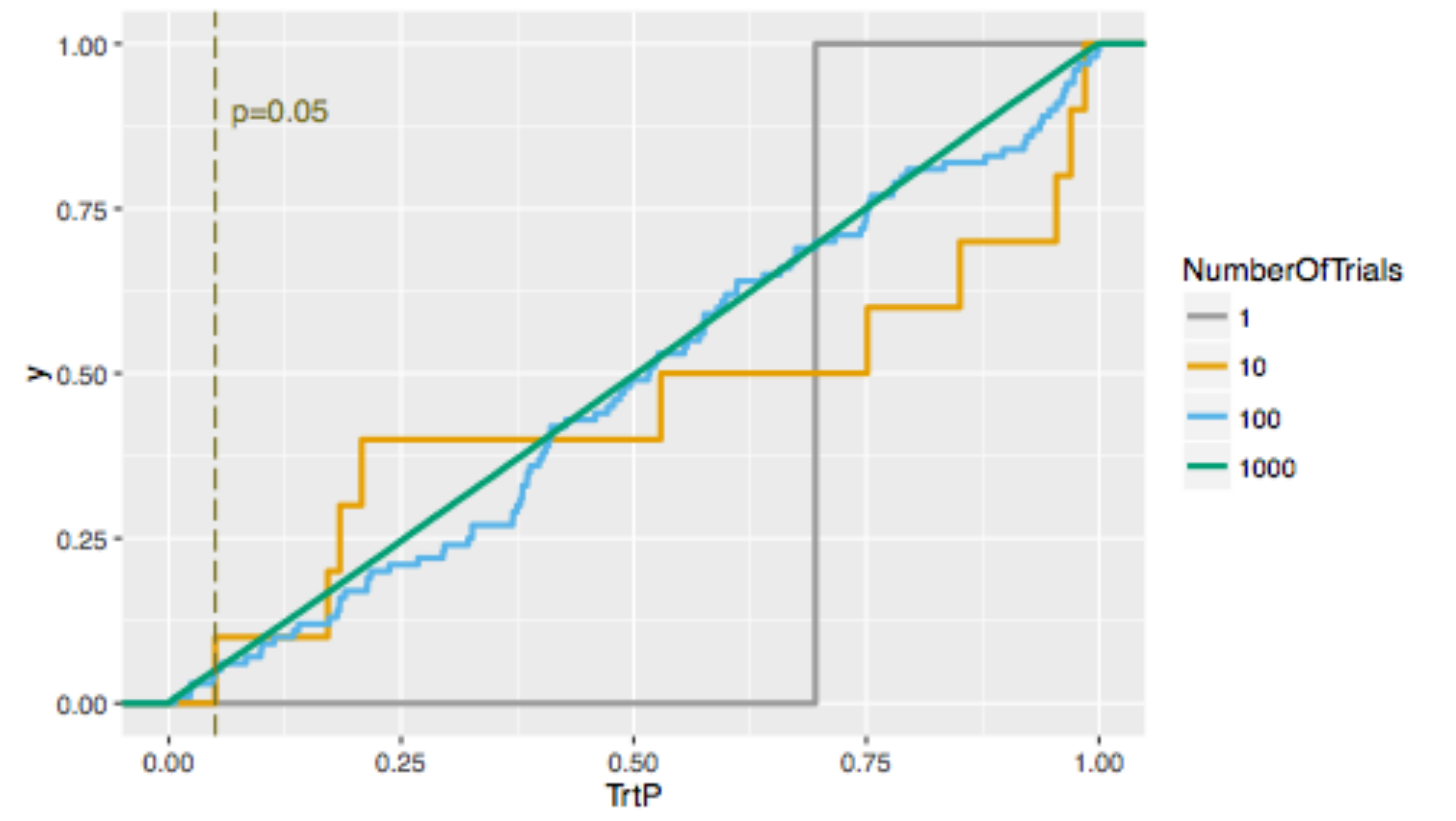
Layout 12





Large Sample Theory

Combined treatment p for increasing numbers of RCB layouts, simulated over 180 experiments. As the number of trials increases. the distribution of p becomes uniform.



Large Sample Theory

In these data, the null hypothesis is true. Uniformity implies that the probability of rejecting the null hypothesis is exactly equal to any chosen critical p-value.

Restricted Randomization

- How can we control for undesirable layouts?
- The universe of potential randomized complete block layouts includes some that place multiple plots with the same treatment in close proximity.
- Researchers might recognize and reject these designs out of hand, and re-randomize.
- This is an informal method of restricted randomization.
- Several systematic restricted randomizations have been proposed.

Systematic Restricted Randomization

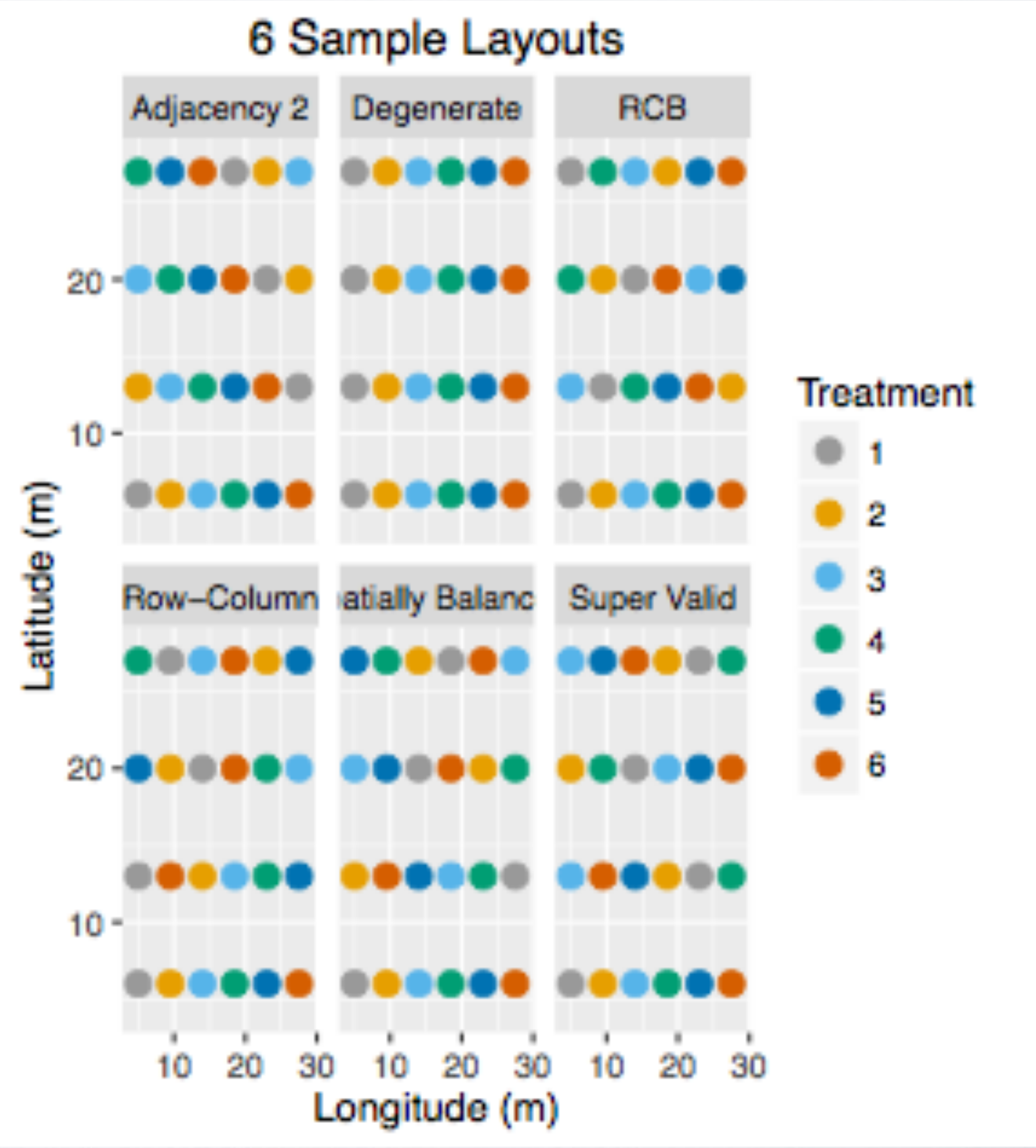
- Degenerate
  - unrandomized
- Randomized Complete Block
  - unrestricted randomization
- Restricted Adjacency
  - ARM setting = number of columns between identical treatments in adjacent blocks
- Super-valid
  - for each pair of rows, a single treatment may appear twice in the same column
- Row-column
  - Latin Rectangle
- Spatially-Balanced
  - spatial balance among treatment contrasts

Column Restrictions

- Degenerate
  - treatment number same as column number
- Randomized Complete Block
  - any treatment may be applied to any column in a block
- Restricted Adjacency
  - treatments not allowed to appear in the same column in adjacent blocks
- Super-valid
  - no treatment may appear more than twice in the same column
- Row-column
  - no treatment may appear more than once in the same column
- Spatially-Balanced
  - no treatment may appear more than once in the same column

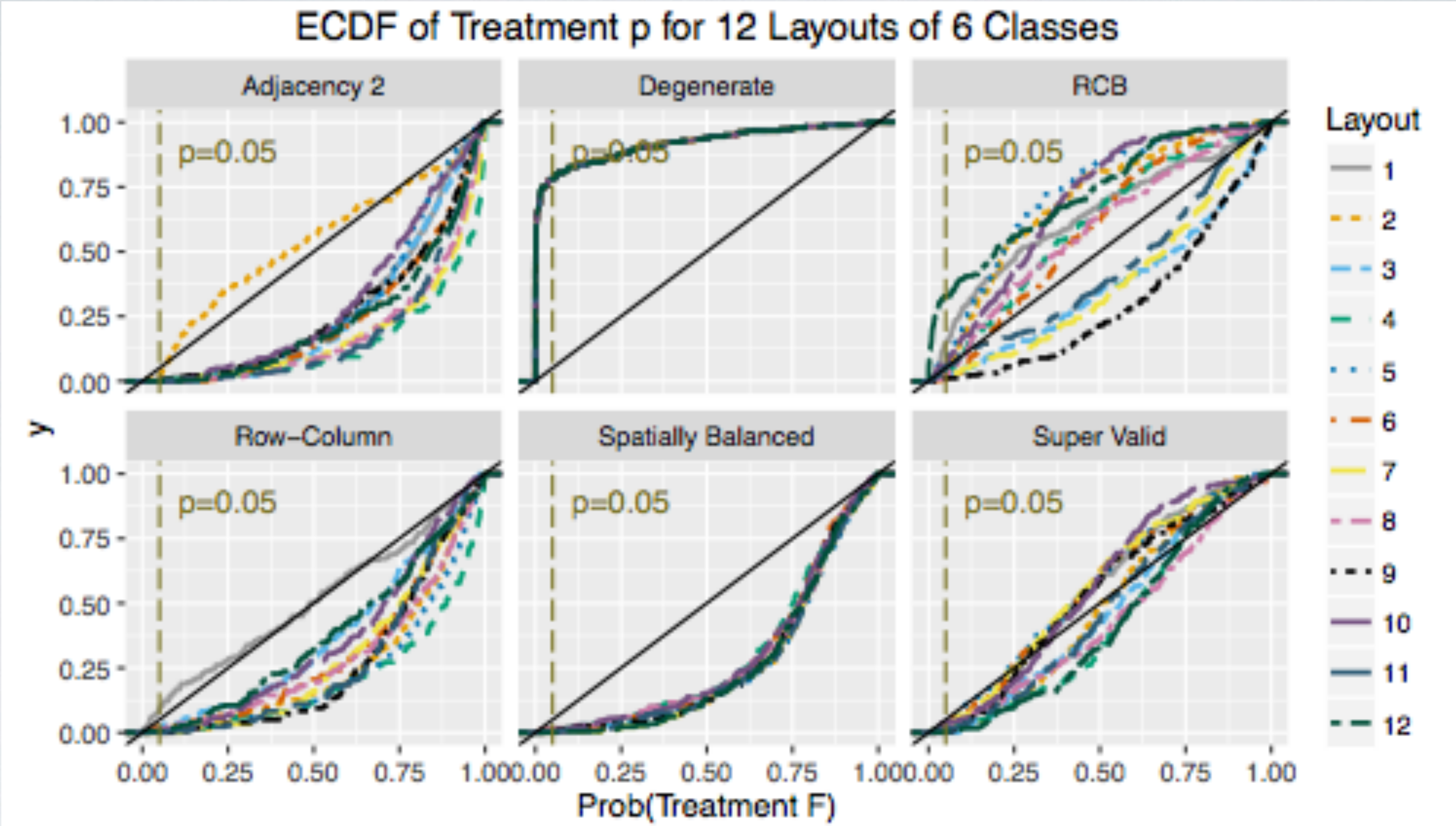
Example Randomizations

- Degenerate
- Randomized Complete Block
- Row-column
- Super-valid
- Restricted Adjacency
- Spatially-Balanced



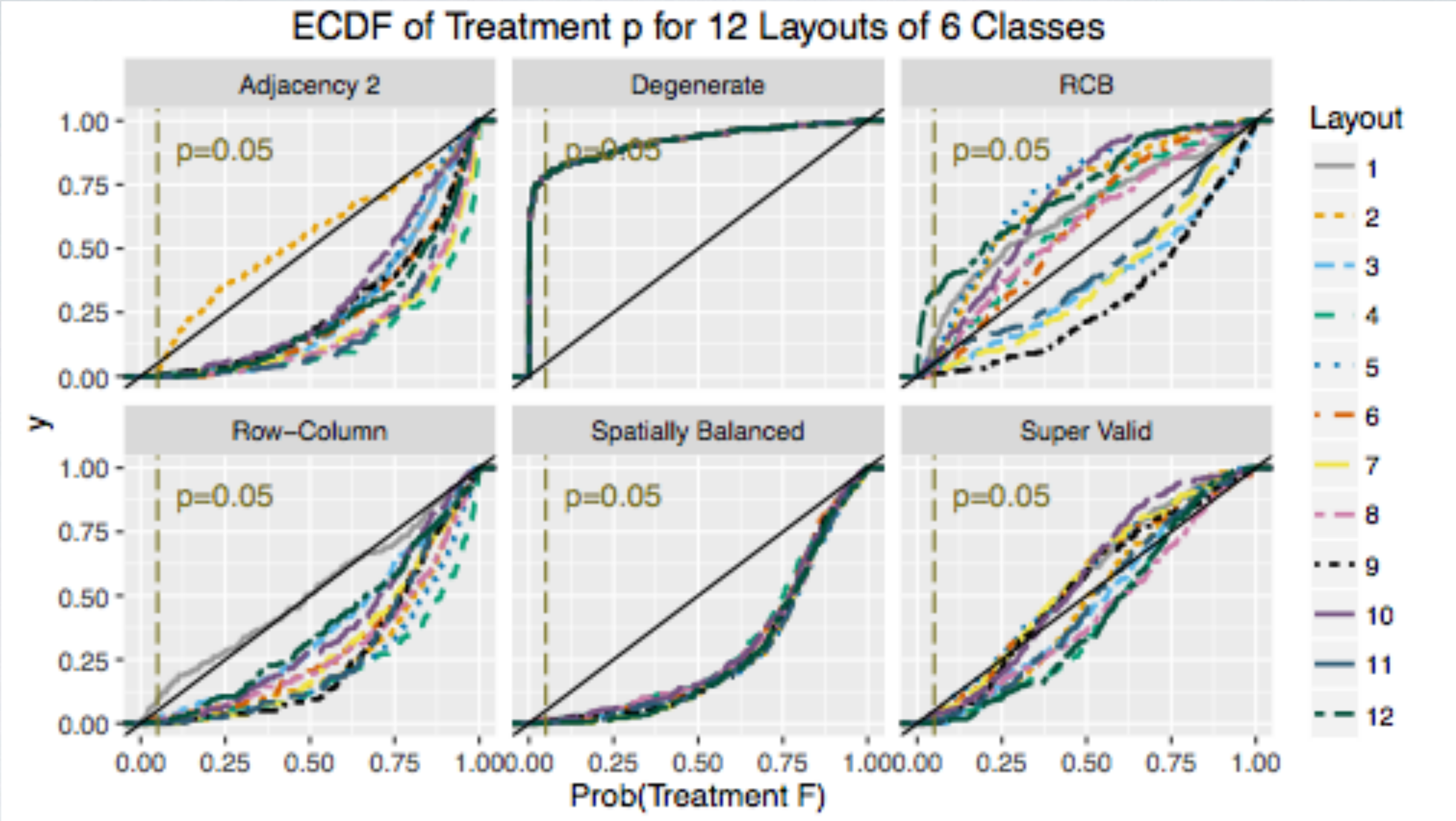
Classes of Layouts

- To compare the different classes of restricted randomization schemes, 12 instances of each were created.
- Classic Fisher randomization allows independent randomization of treatments in each block, as previously described.
- New adjacency layouts were recreated by independent randomizations with the same setting (treatment adjacency=2).
- The properties of row-column and super-valid designs allow new layouts to be produced by independently permuting (swapping) rows and columns.
- Spatially-optimal layouts are optimized for average distance between treatments in rows, so permutations are limited to swapping rows only.



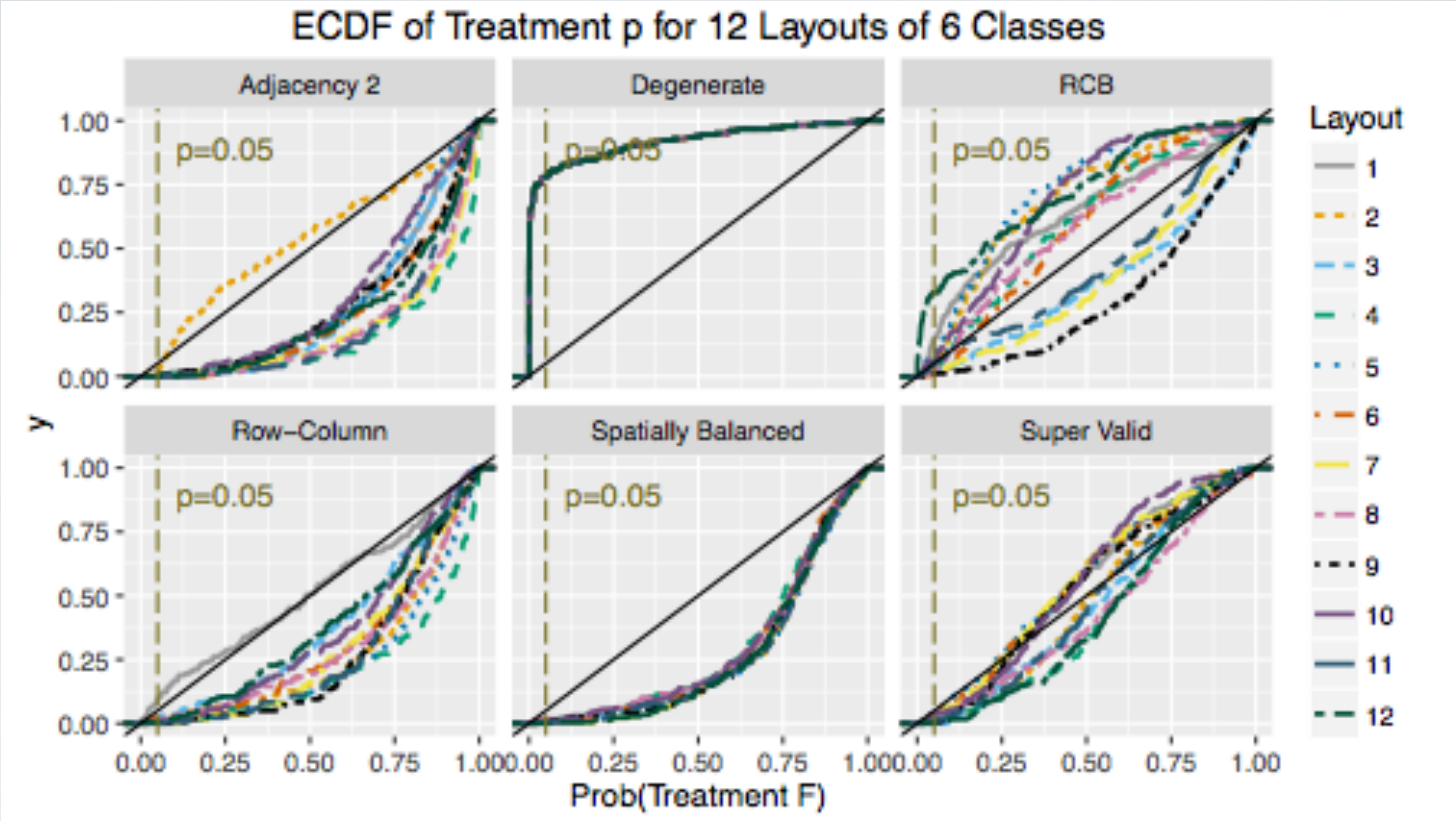
Probability Distributions, 12 Layouts

Restricted randomization tends to exclude designs that produce left-skewed distributions.



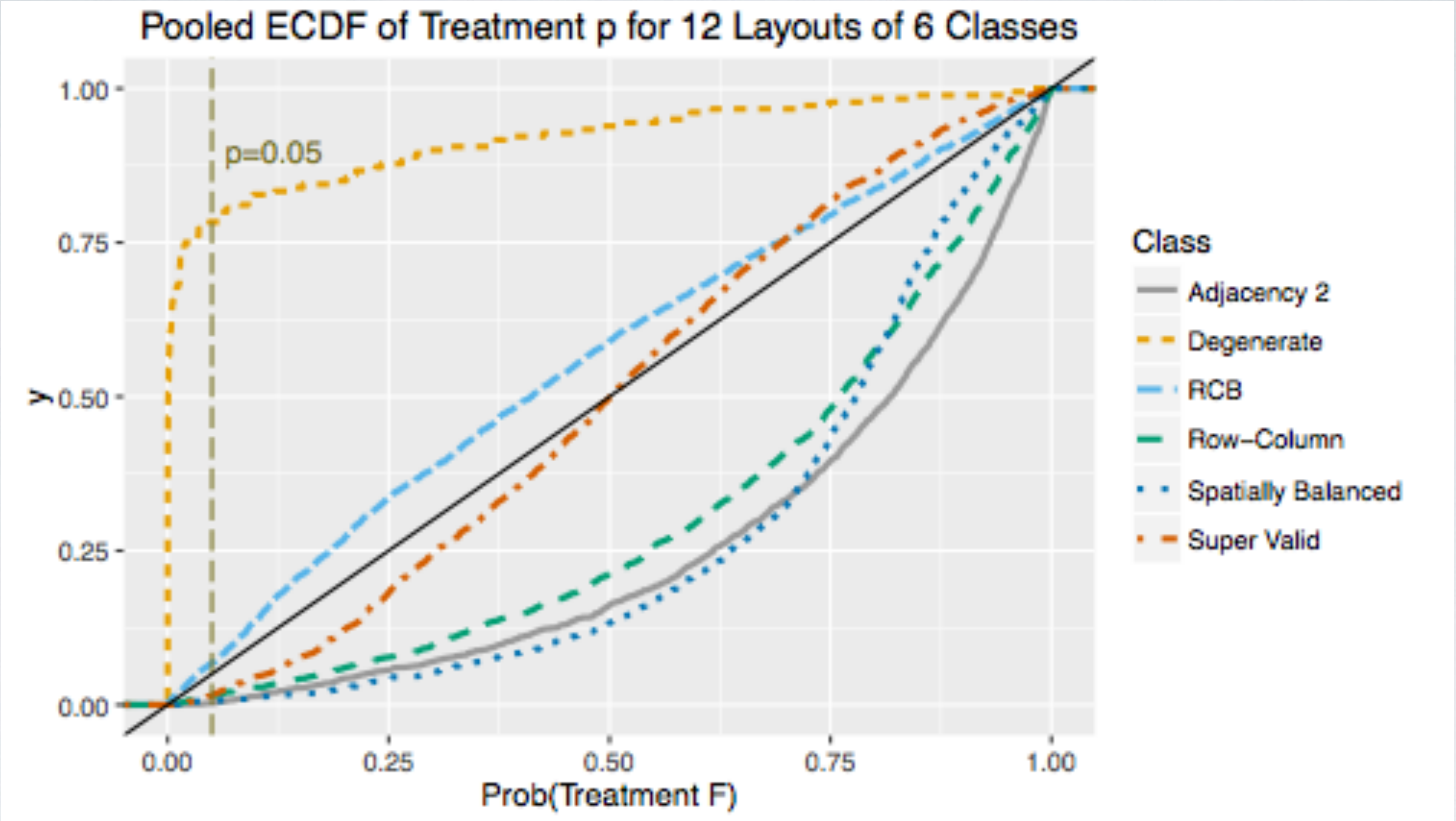
Probability Distributions, 12 Layouts

Super-valid designs, which allow two treatments to appear in the same columns, tend to be centrally distributed.



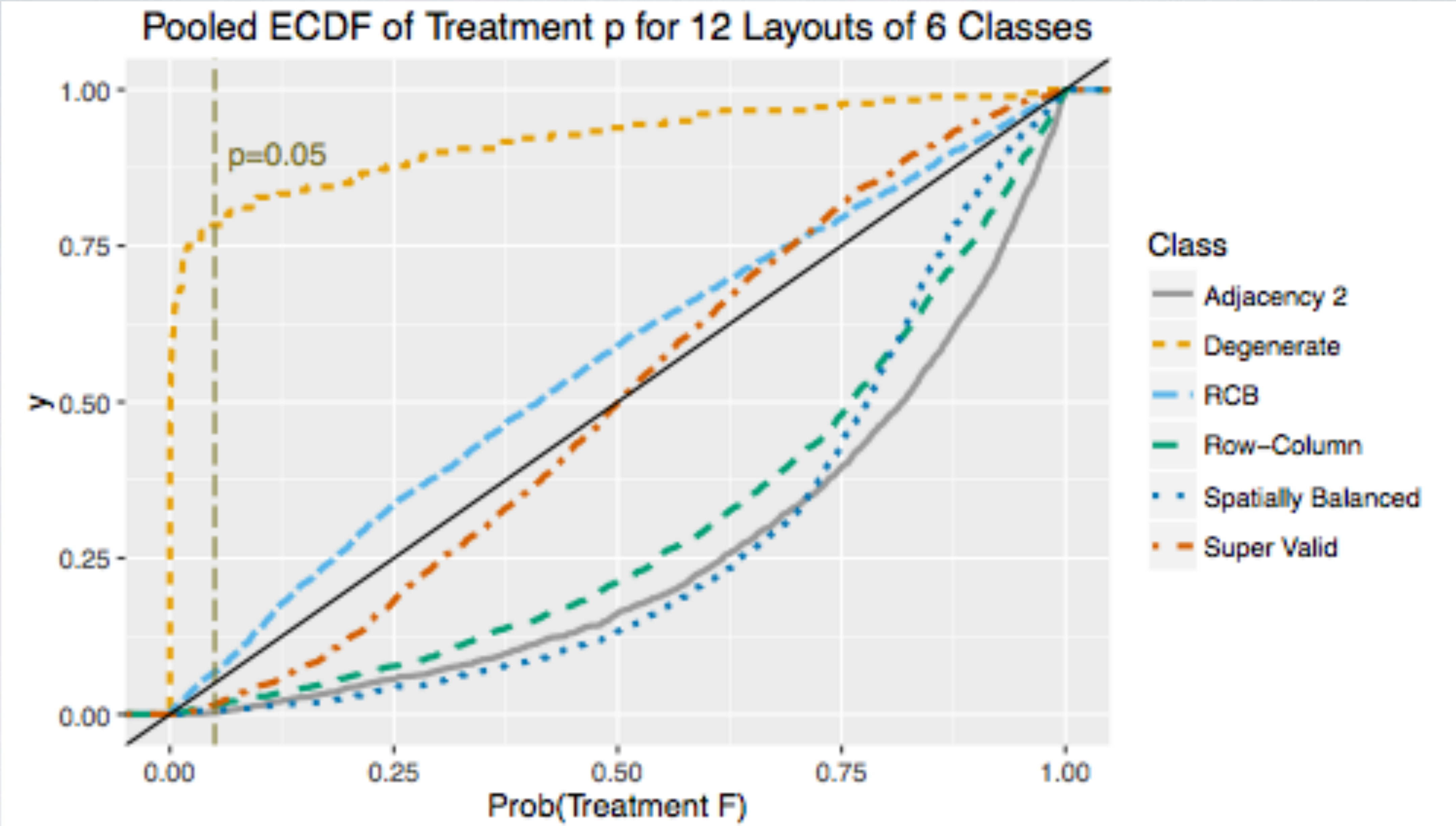
Probability Distributions, 12 Layouts

Spatially balanced designs, which are generated by row-permutations only, are a more homogeneous class of randomization.



Pooled Distributions, 12 layouts Each

Row-column, spatially balanced and restricted adjacency show similar tendencies away from smaller p-values



Pooled Distributions, 12 Layouts Each

At a nominal probability of 0.50, the super-valid layouts tend toward an achieved rate of 0.50.

Power Analysis

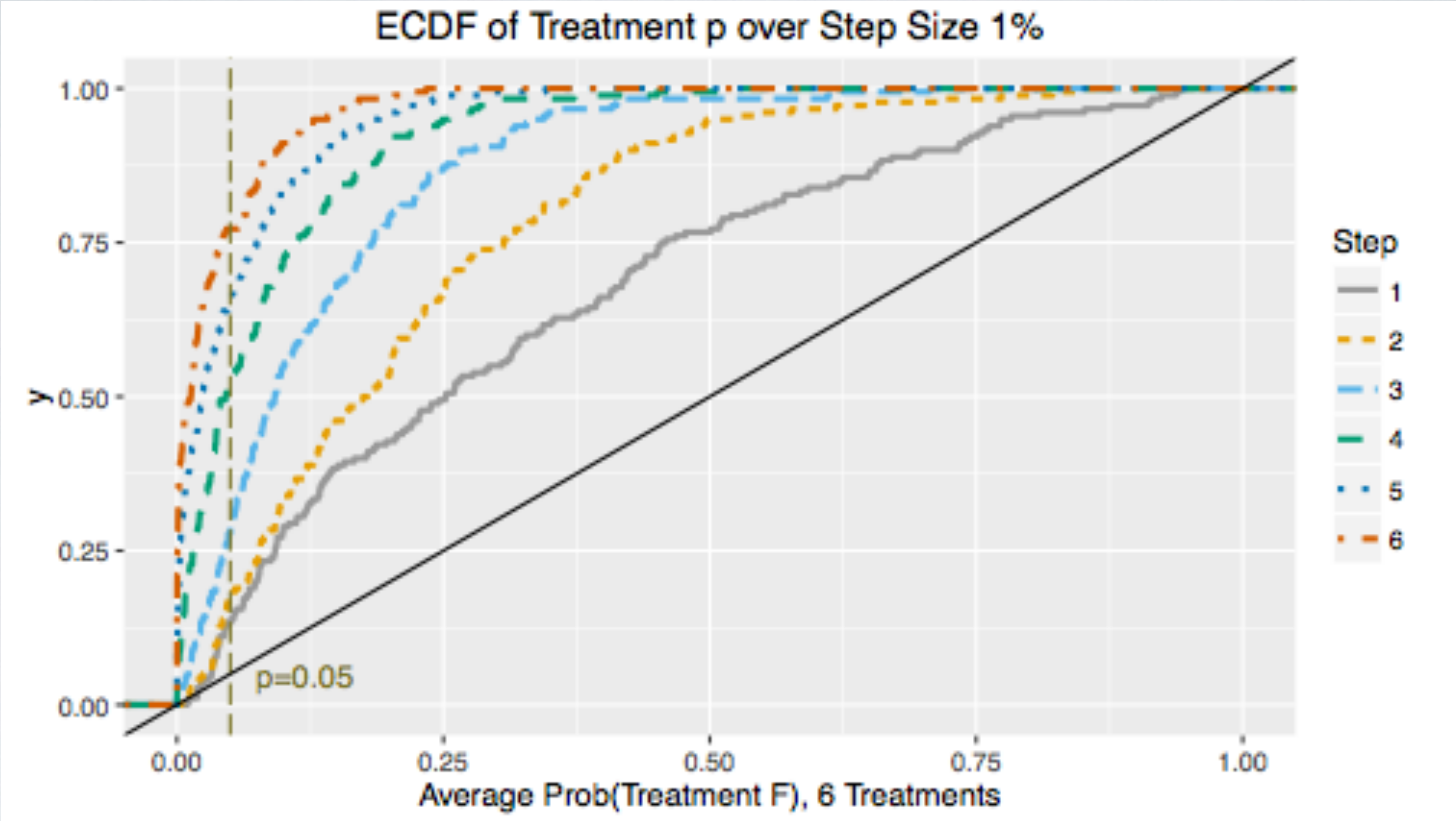
- Different layouts may produce achieved Type I error rates that are much lower than nominal error.
- However, planning experiments requires a compromise between Type I rates (detecting significance where none in present) and Type II error rates (overlooking a true treatment difference).
- We can also attempt to simulate Type II error and compare classes of restricted randomization.

Simulating Type II Error Rates

- Simulating Type I Error rates using uniformity data is straight forward. Since there is no treatment effect, any trial detecting significance can be counted as an error.
- To determine Type II error rates, we need to add a "true" effect, and count the number of trials that fail to detect significance.
- But what is a true effect?

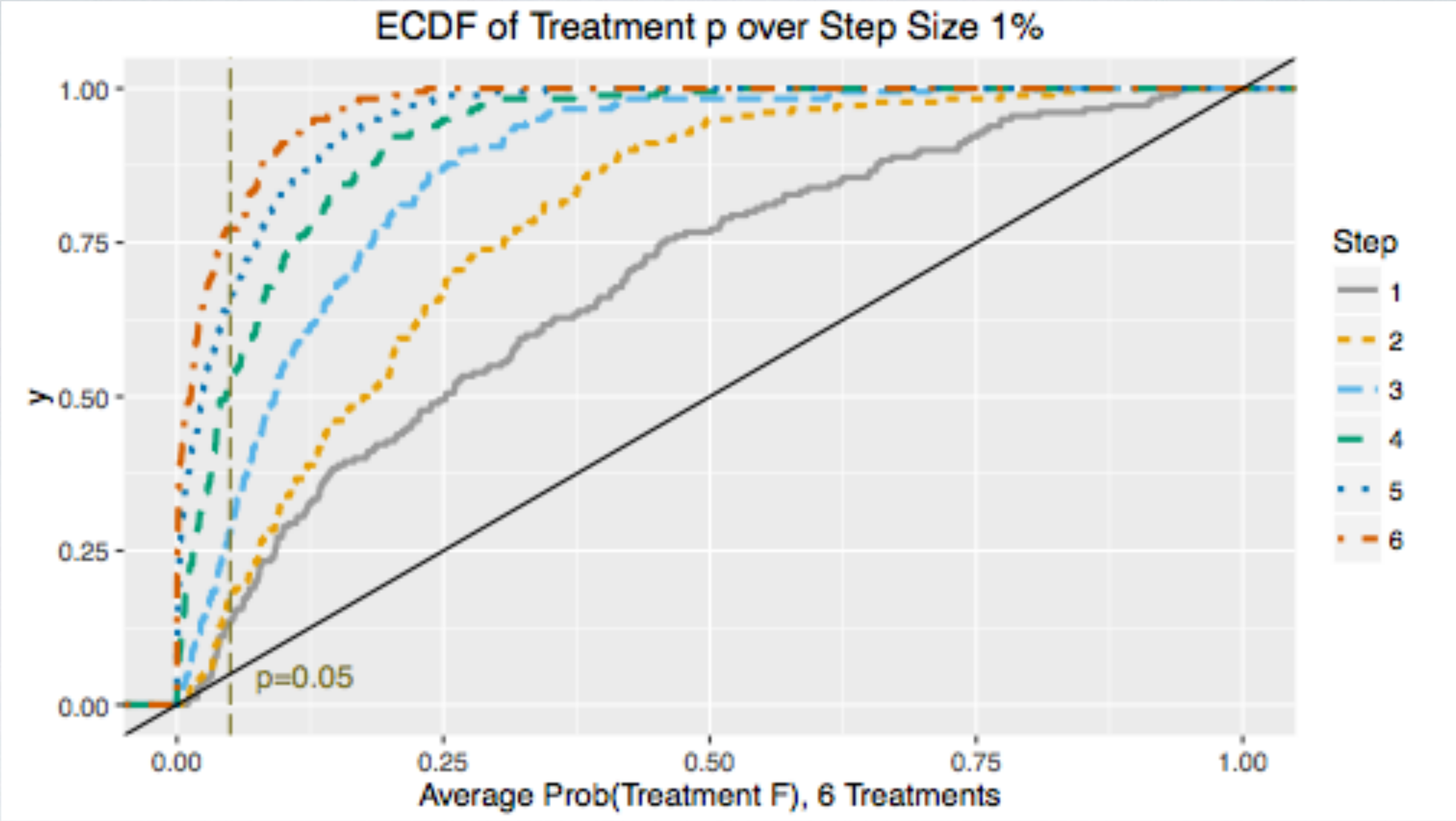
Simulating Type II Error Rates

- To determine a true effect, we start with a small value, ~1% of the grand mean, add this value to a single treatment, perform AOV and check treatment p.
- Do this for each treatment (of 6), and each location (of 180) in our field.
- If we haven't detected significance in at least 864 trials (80% of 180 x 6), increment our effect value and repeat.
- This gives an estimate of effect size required to achieve 80% power.



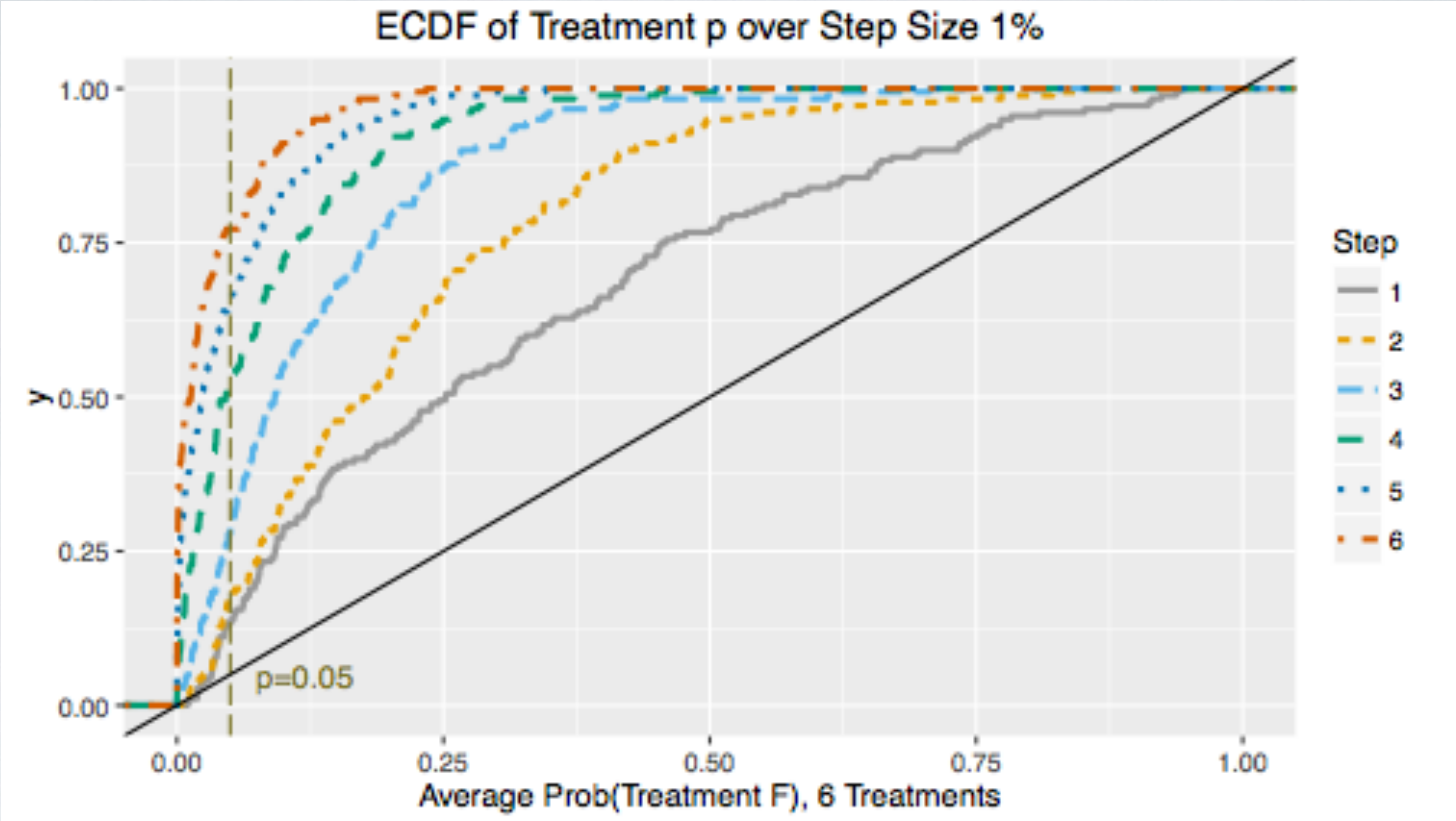
Incrementing Effect Size

As we increment effect size for a single RCB trial, we see the Treatment p distribution shifted toward the left.



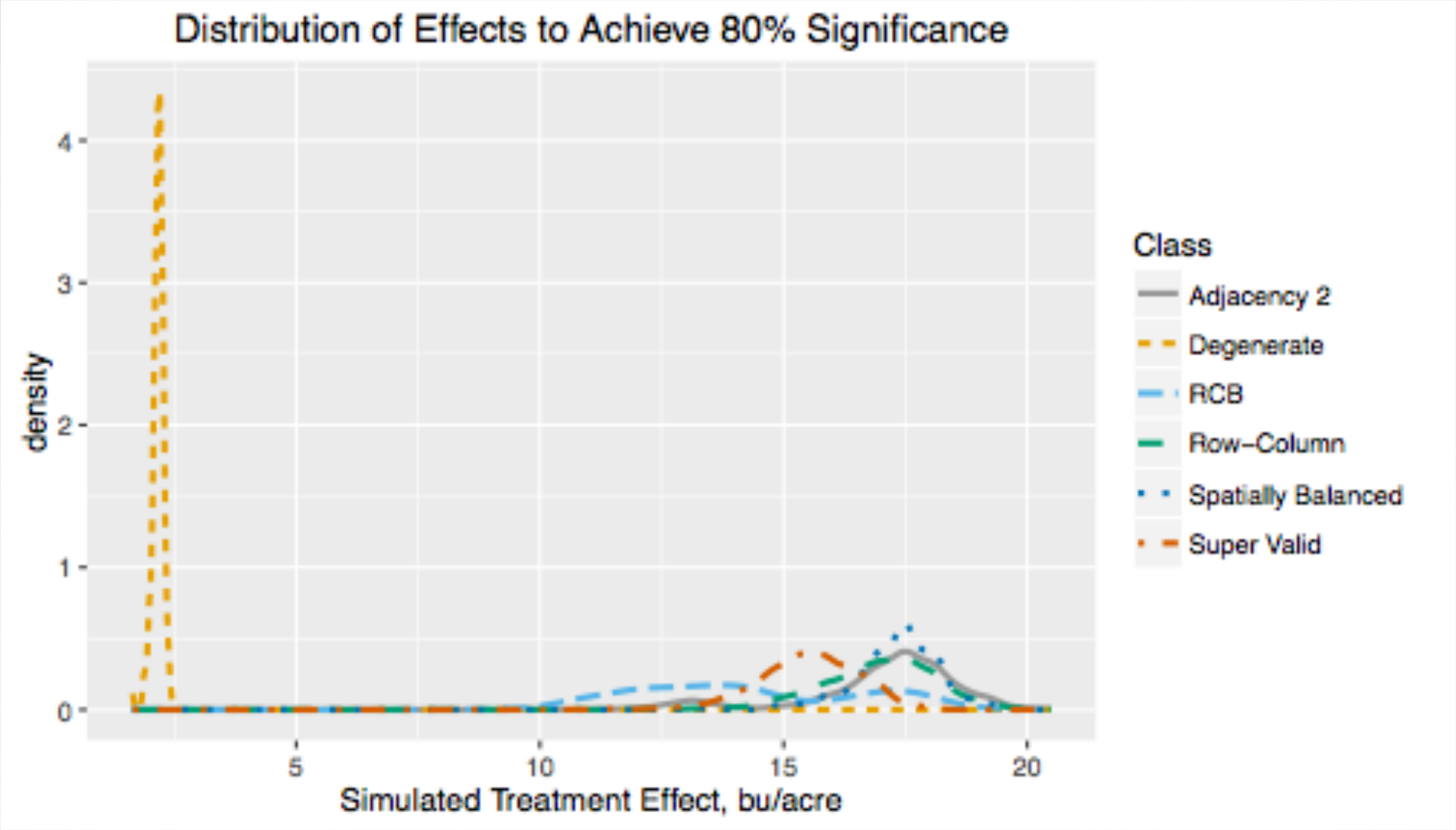
Incrementing Effect Size

True effects shift treatment p distribution to the left.



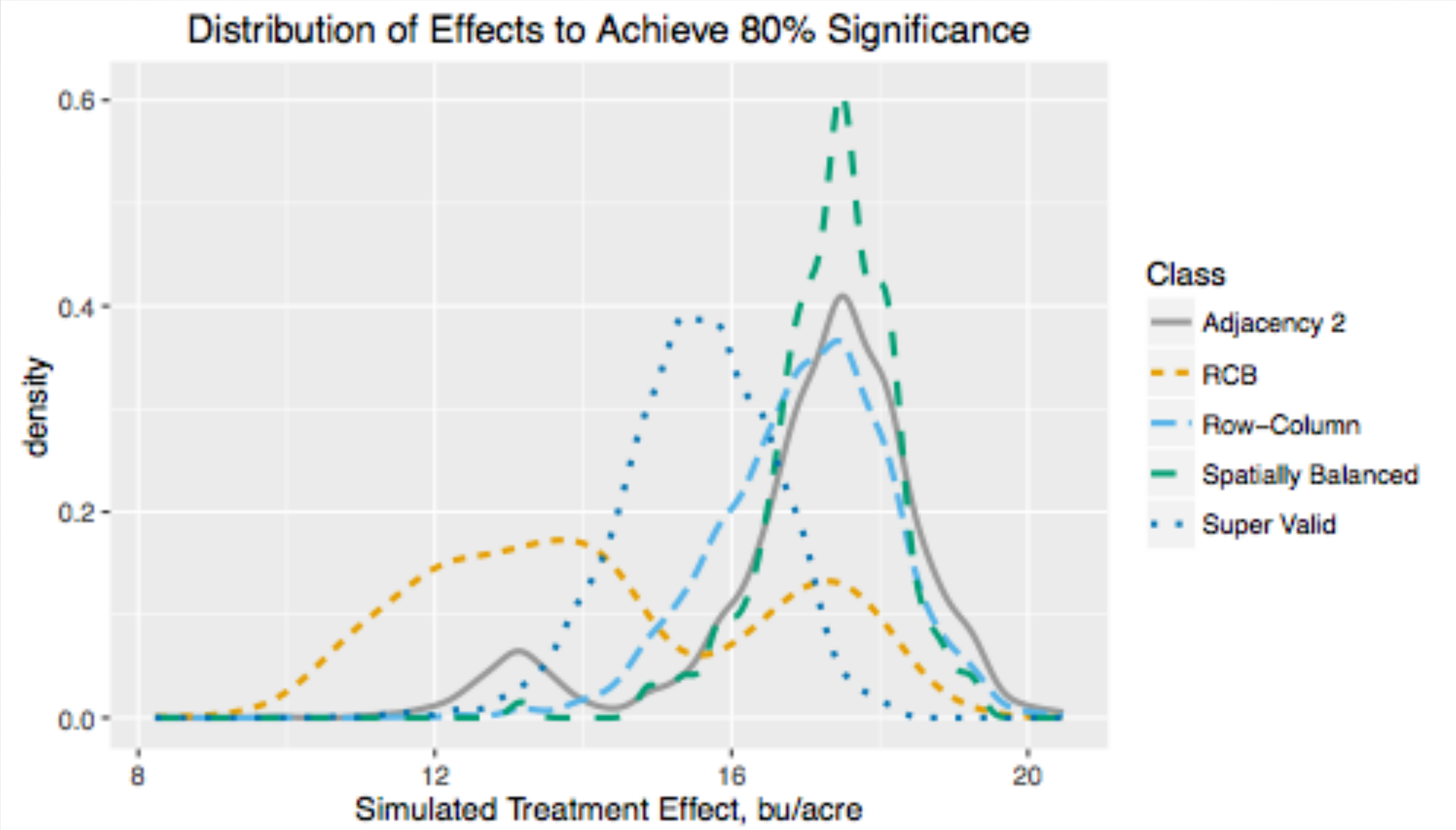
Incrementing Effect Size

Remember, null treatment effects confounded with spatial effects also shifted treatment p distributions to the left.



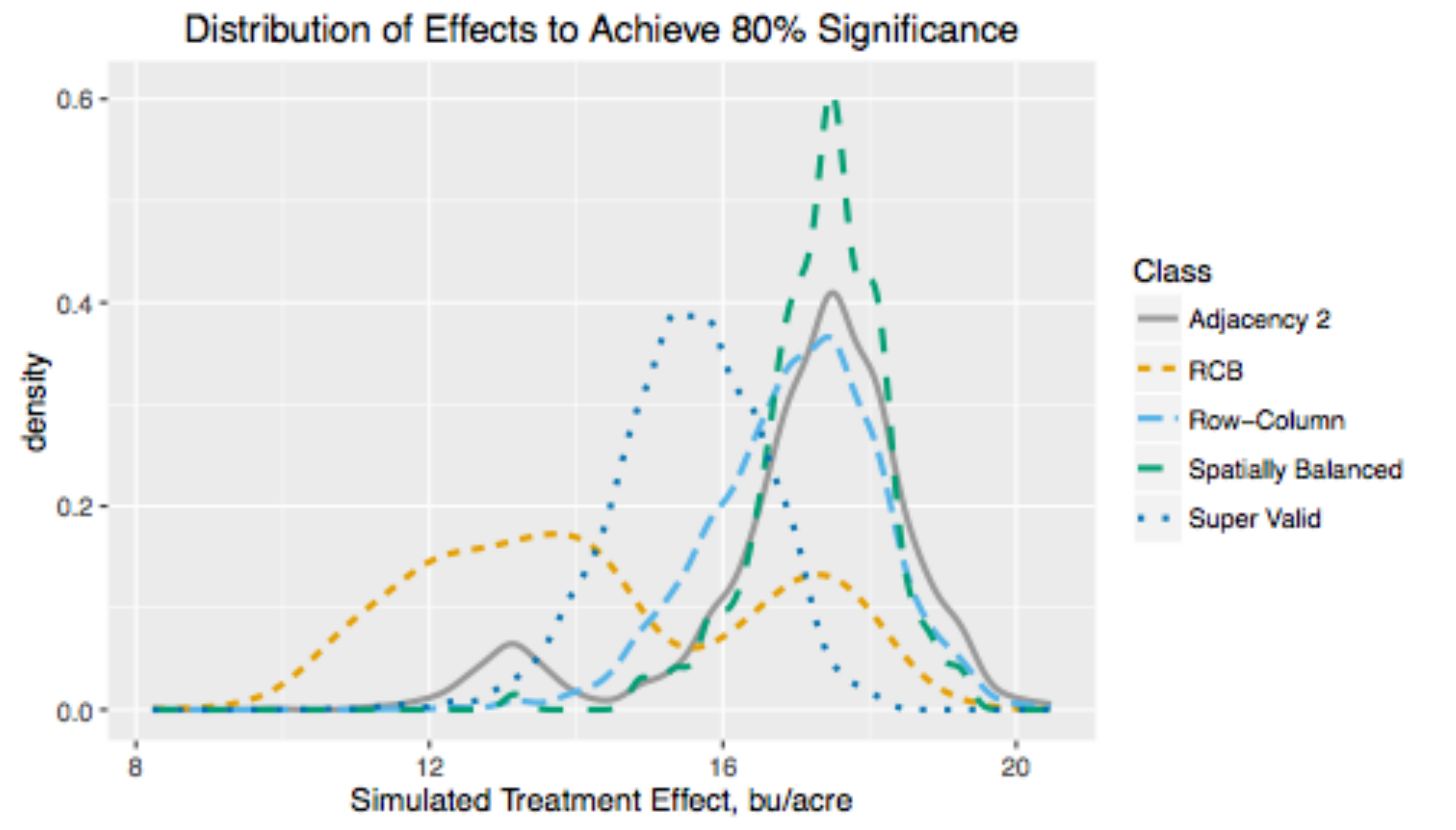
Pooled Power Analysis

An unrandomized layout is almost certain to detect very small treatment effects as significant.



Pooled Power Analysis

Layouts that tend toward lower Type I error rates require larger absolute treatment effects (~17 bu/acre) to achieve a desired power.

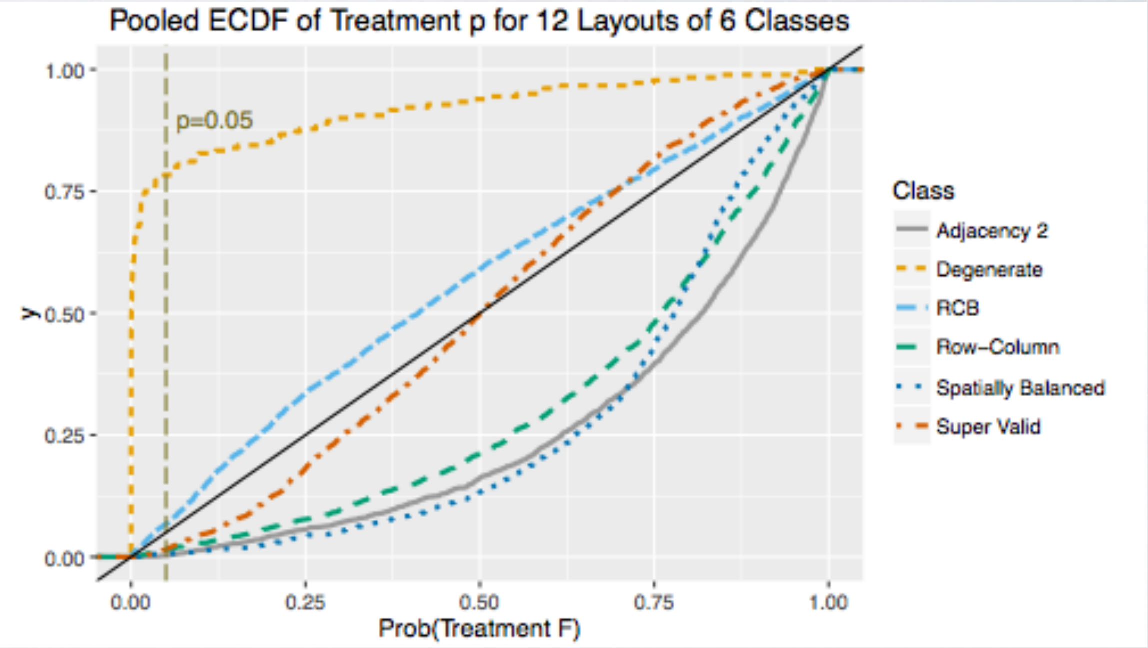


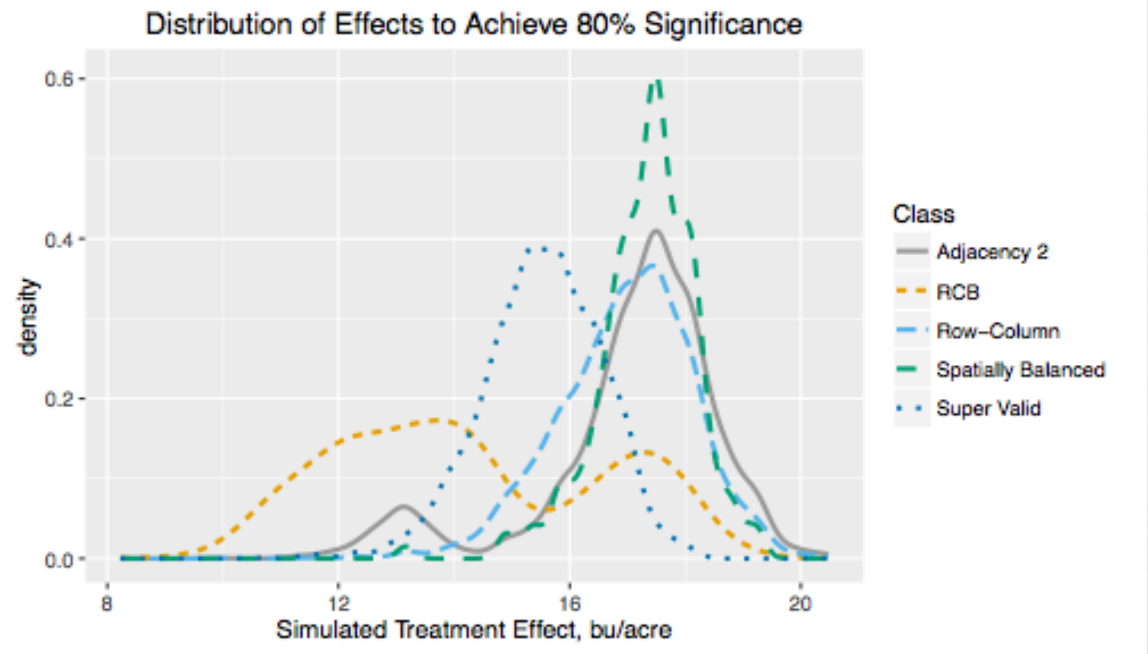
Pooled Power Analysis

Super valid layouts offer a compromise between Type I and Type II error rates.

Recommended Replicates

- If we suspect that a layout has a tendency to low p-values, would we use the same number of replicates?
- If we suspect that a layout has a tendency to require larger treatment effects, would we use the same number of replicates?
- Given an arbitrary layout, can we predict these tendencies?

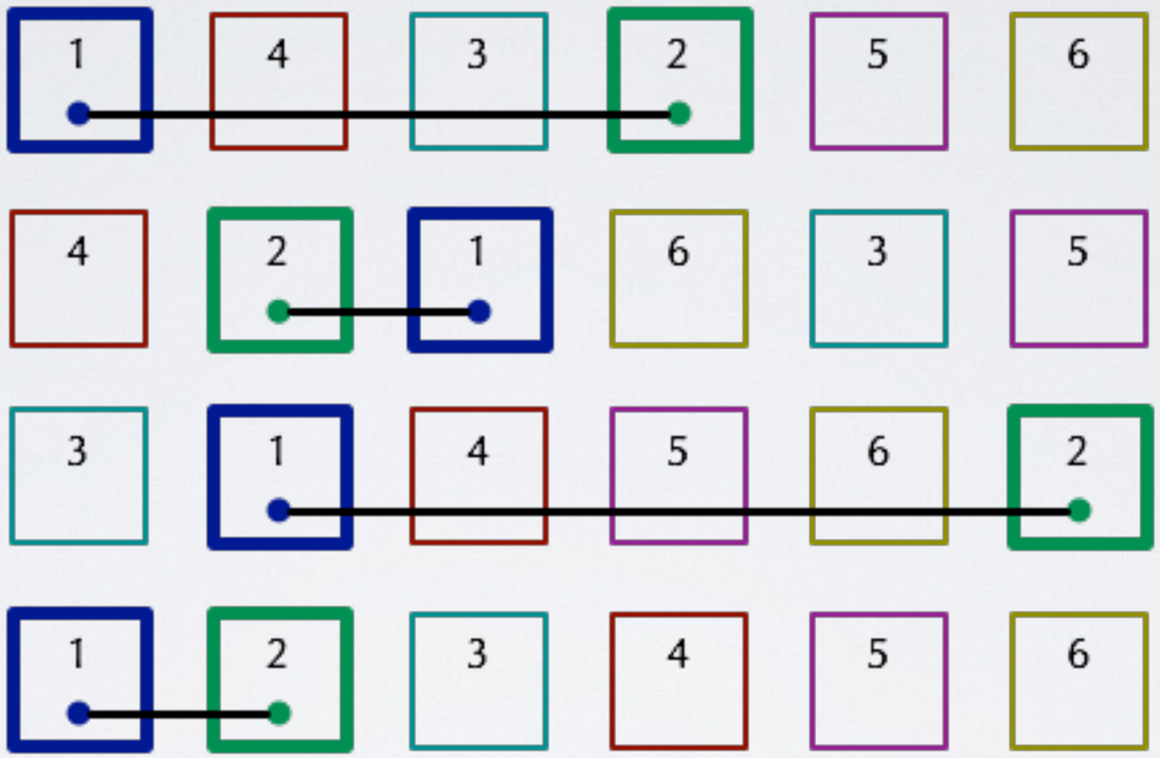




Average Distance of Treatment Comparison

- van Es and van Es, "Spatial Nature of Randomization and Its Effect on the Outcome of Field Experiments", Agron J, 85:420-428 (1993).
- Comparison between treatments 1 and 2 is made from data taken from 4 plots for each treatment.
- Measure the plot-to-plot distance for each plot containing treatment 1 to the paired plot, within replicates, containing treatment 2, for a total of 4 distances.
- ADTC for the treatment pair 1-2 is the average of the 4 distances.

Distances, Treatments 1-2



The average distance for the contrast between treatments 1 and 2 is computed by averaging the linear distances between plot centers, including plot width (4m) and buffer space (0.5m)

Average Dispersion

- To provide an estimate of how well a treatment is dispersed relative all other treatments, compute the average of ADTC for all comparisons including that treatment.

- Contrasts including Treatment 1
  - ADTC 1-2 = 10.125 ADTC 1-3 = 7.875 ADTC 1-4 = 7.875
  - ADTC 1-5 = 14.625 ADTC 1-6 = 15.75

- Average Distance, Treatment 1 Contrasts = 11.25

- Standard Deviation, Treatment 1 Contrasts = 3.73

Summarizing ADTC

Standard Deviation of Average Distances = 0.61

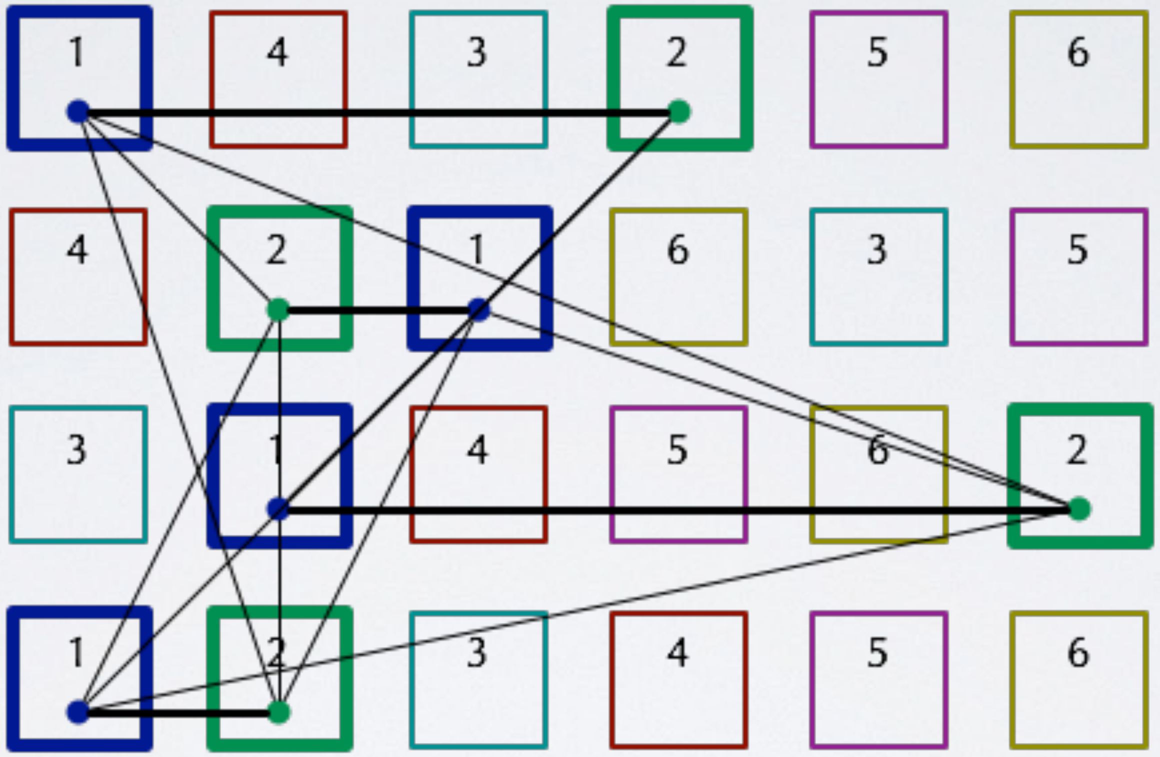
Treatment

Trt	Code	At Edge	Ave Dist.	StDev	Min	Max
1		*2*	**11.3**	**3.73**	7.9	**15.8**
2		**3**	10.4	*0.94*	**9**	*11.3*
3		**3**	*9.9*	1.85	7.9	12.4
4		**3**	*9.9*	1.85	7.9	12.4
5		**3**	10.4	3.32	*5.6*	14.6
6		*2*	**11.3**	**3.73**	*5.6*	**15.8**
Average		2.7	10.5	2.57	7.3	13.7

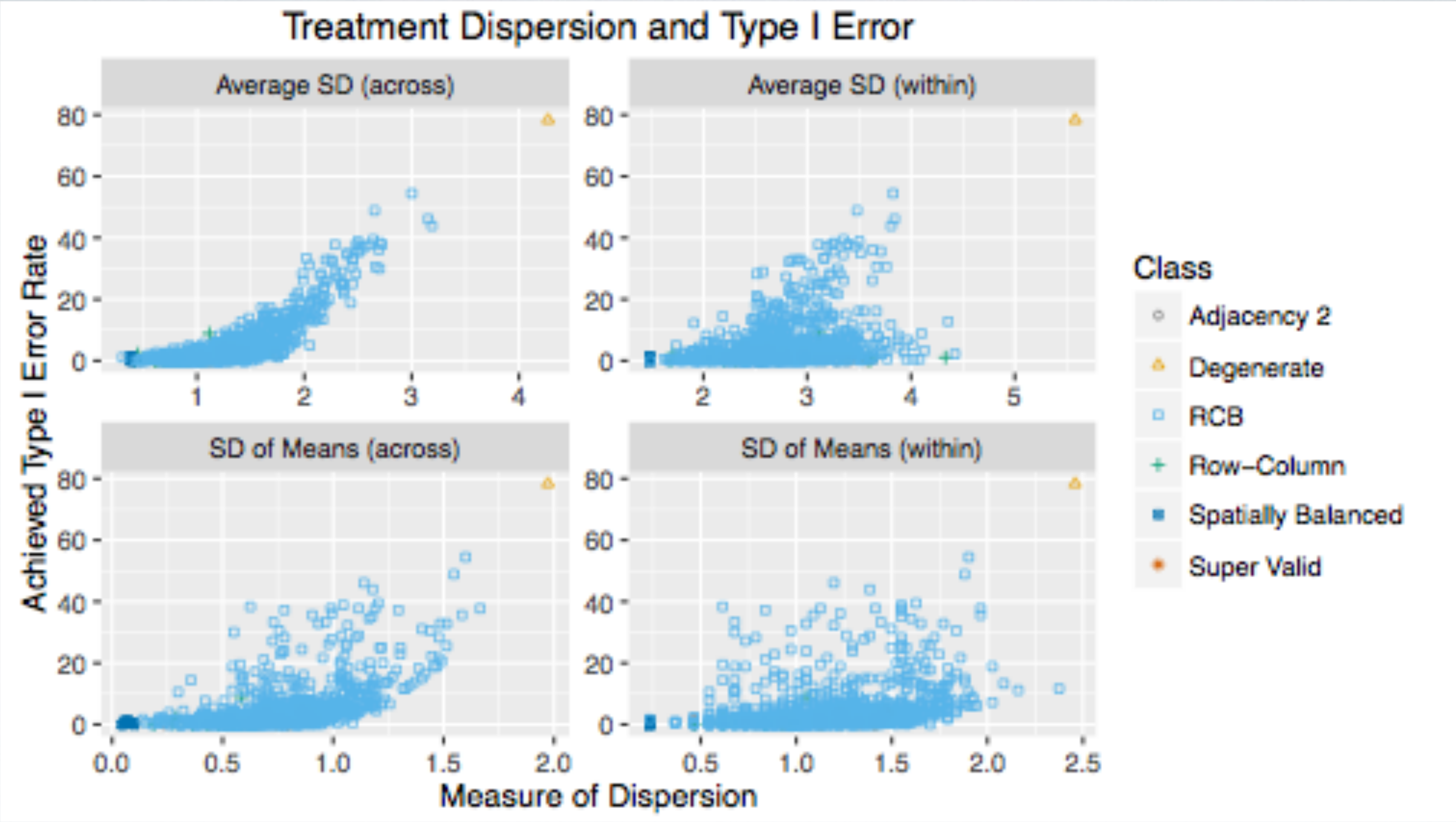
Average of Standard Deviations

Two choices for a single value summarizing ADTC

Distances, Treatments 1-2, Across Replicates

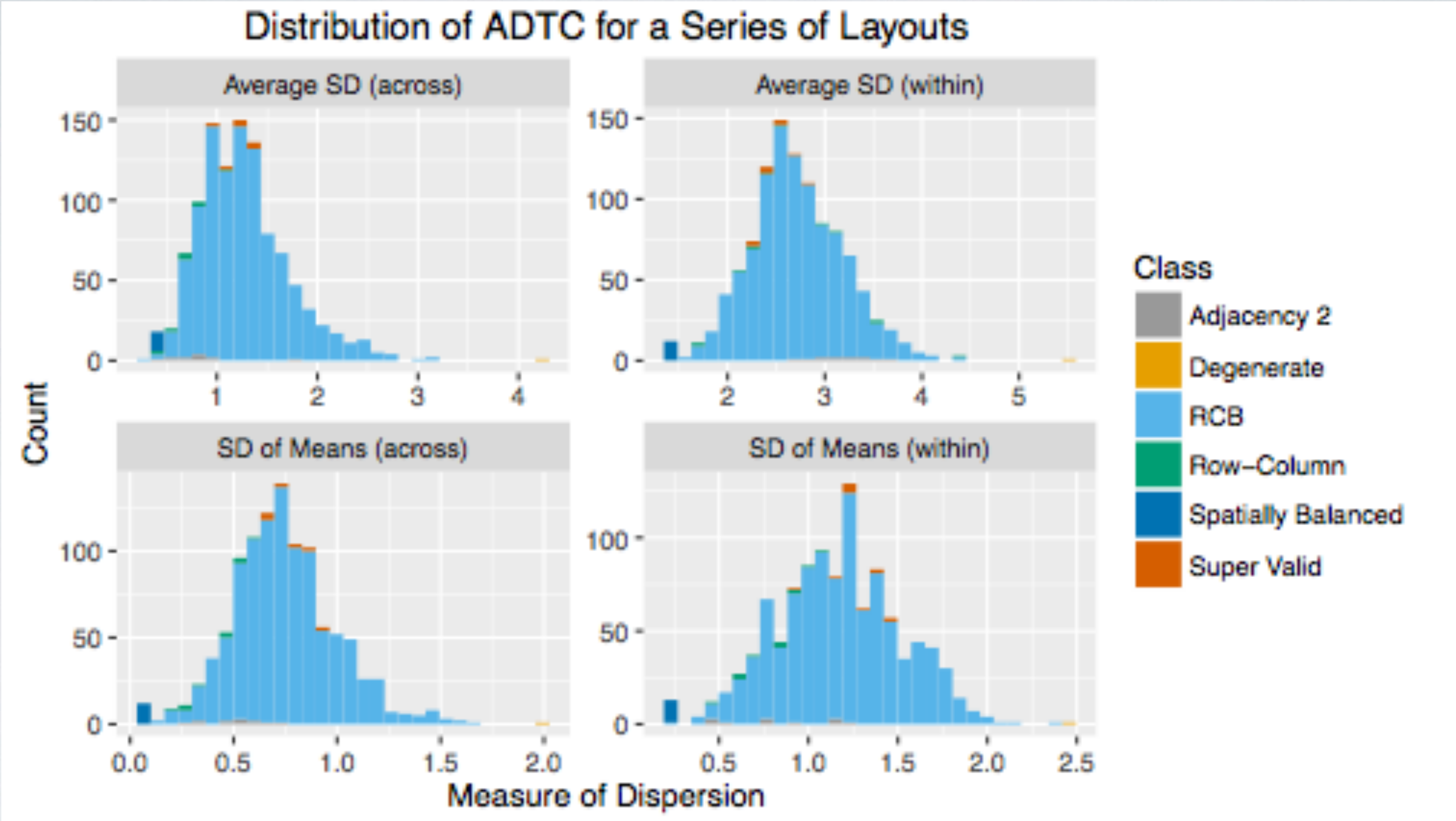


Measure dispersion by computing ADTC with distances across replicates as well as within replicates.



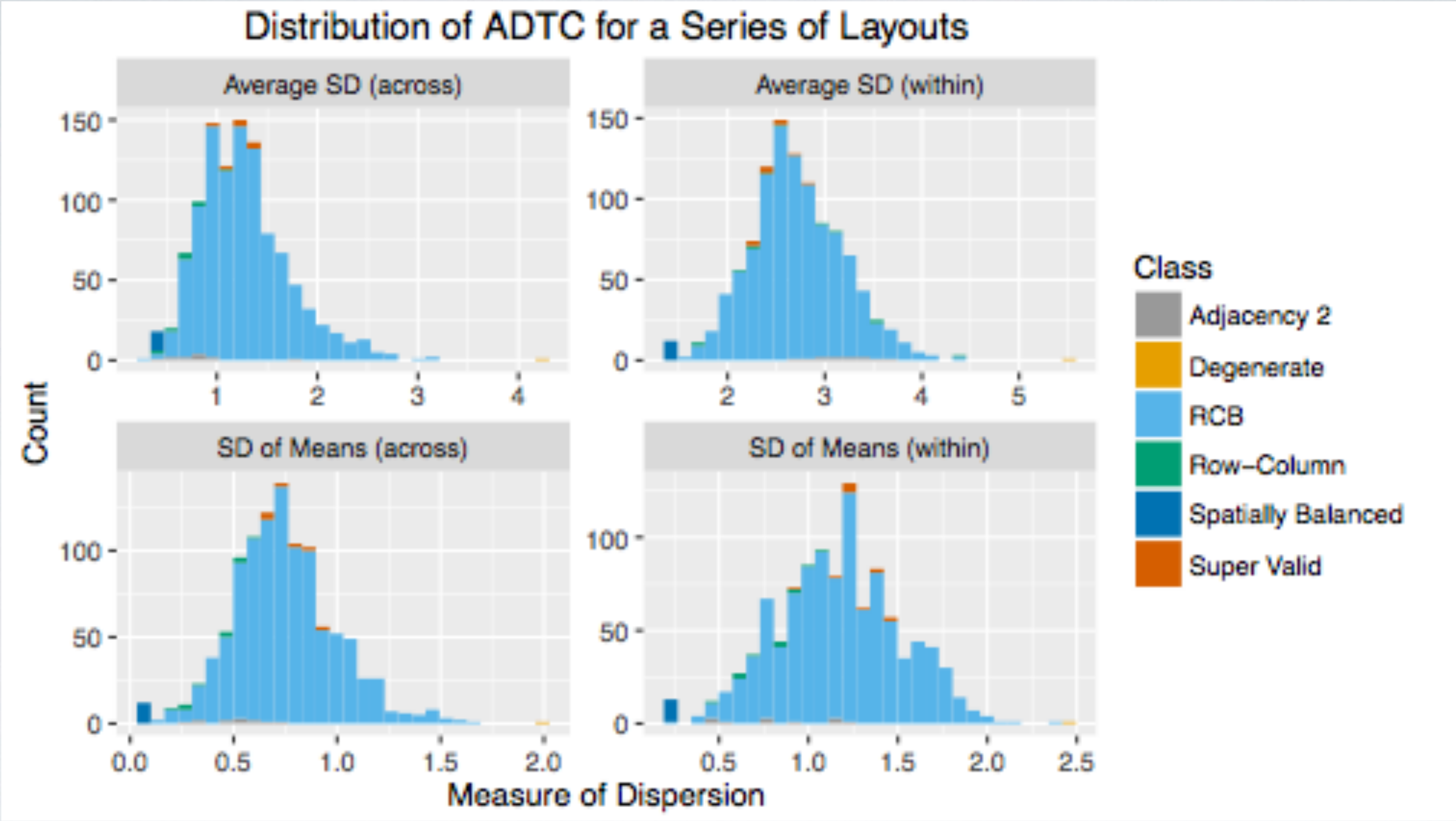
ADTC as a Predictor of Type I Error

12 trials of each class of restricted randomizations, along with 1000 unrestricted RCB layouts



Distribution of ADTC

Adjacency, row-column and spatially balanced randomizations tend toward low ADTC



Distribution of ADTC

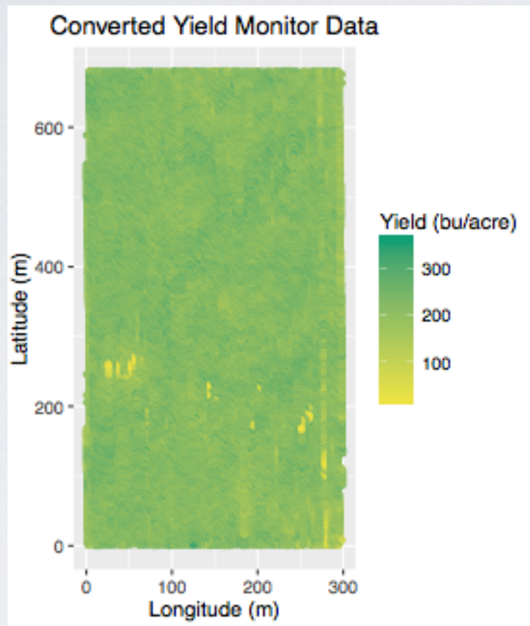
Super valid layouts tend toward mean ADTC

References and Related Work

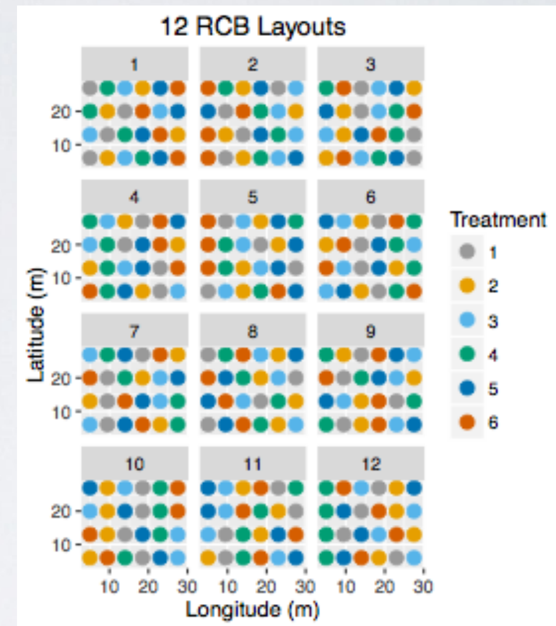
- Bailey, R. A. ''Experiments in rectangular areas: design and randomization." Journal of Agricultural, Biological, and Environmental Statistics 17.2 (2012): 176-191.
- van Es, H. M., and C. L. van Es. "Spatial nature of randomization and its effect on the outcome of field experiments." Agronomy Journal 85.2 (1993): 420-428.
- van Es, H. M., et al. "Spatially-balanced complete block designs for field experiments." Geoderma 140.4 (2007): 346-352.
- Williams, Emlyn R., and Hans-Peter Piepho. "An evaluation of super-valid restricted randomization." Journal of Agricultural, Biological, and Environmental Statistics 19.4 (2014): 470-478.
- Forkman, Johannes. "A comparison of super-valid restricted and row–column Randomization." Journal of Agricultural, Biological, and Environmental Statistics 21.2 (2016): 243-260.
- Piepho, Hans-Peter, Emlyn R. Williams, and Volker Michel. "Beyond Latin squares: A brief tour of row-column Designs." Agronomy Journal 107.6 (2015): 2263-2270.

TL; DR

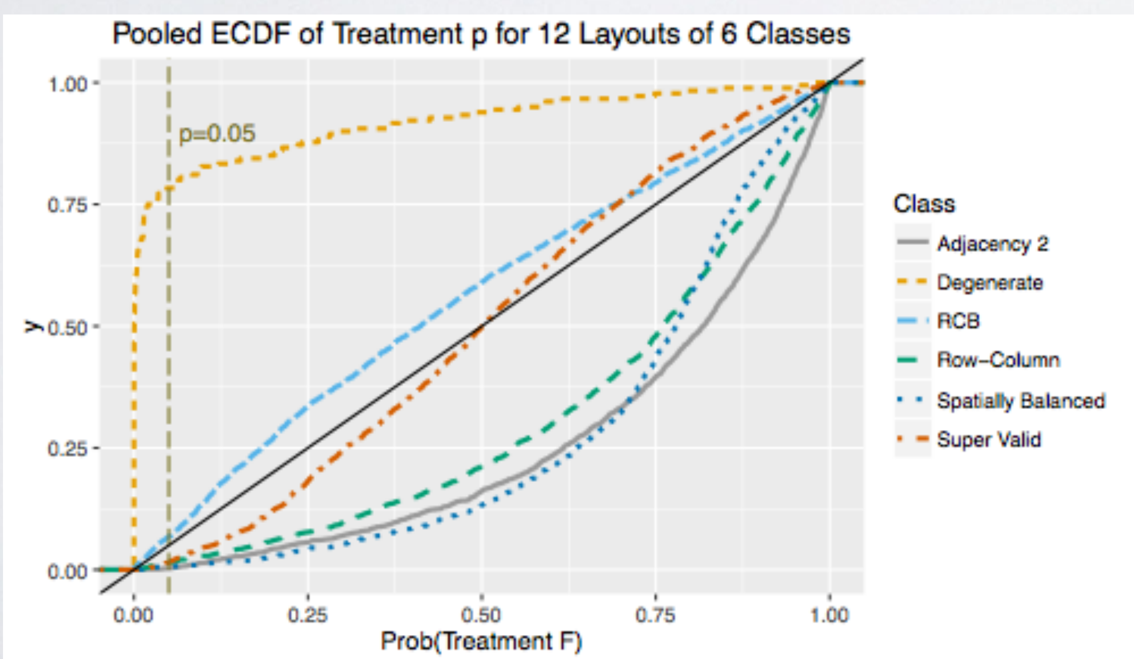
How do

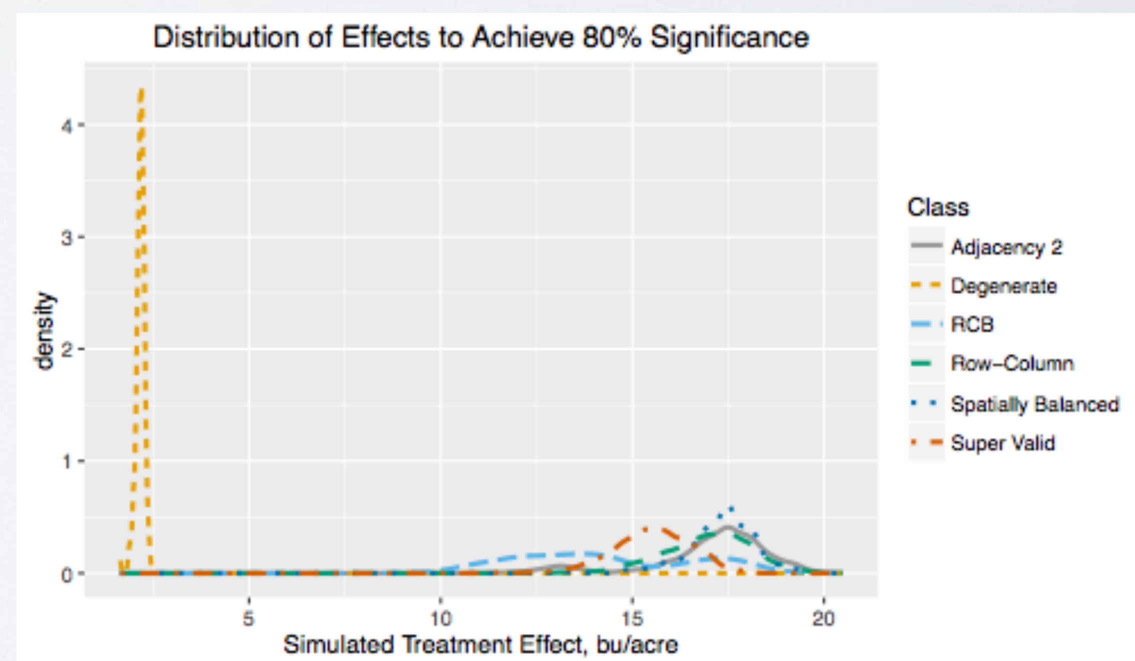


and



combine to produce





?