SIMULATION BASED POWER ANALYSIS: PRACTICAL LESSONS AND APPLICATIONS

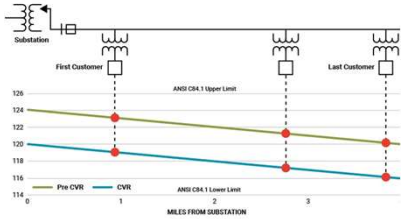
Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

AEIC-WLRA 2021 FALL CONFERENCE, OCTOBER 28-29

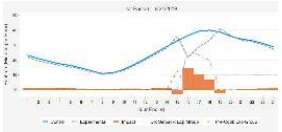
SCIENCE REQUIRES UPFRONT PLANNING AND ANALYSIS
10 STEPS TO GOOD SCIENCE

1. Specify the intervention analyzed in advance
2. Document the hypothesis in advance
3. Identify the data that will be collected and analyzed (e.g., pre-post data)
4. Identify the outcomes that will be analyzed
5. Assign who does or does not receive the intervention, if possible
6. Decide when and how much (dosage) of the intervention is administered
7. Document in advance the techniques and models used to analyze the results
8. Assess the ability to detect a meaningful effect
9. Establish adequate sample sizes
10. Avoid after-the-fact analysis and decisions where there is a temptation to modify models to find the desired results

EXAMPLES OF WHEN POWER ANALYSIS IS USEFUL

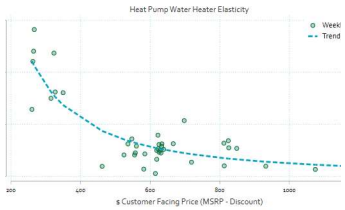


Conservation Voltage Reduction

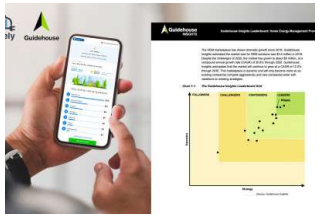


Time Varying Pricing

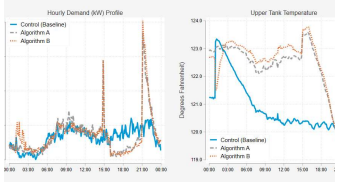
Battery and EV pilots



Enrollment and Incentive Tests



Behavioral Interventions



Other Emerging Tech

- When new technology is costly
- When testing different products (A/B tests)
- When the percent change is subtle or small
- When pilots are constrained by budgets
- When large decisions affect lots of customers (default TOU)

Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

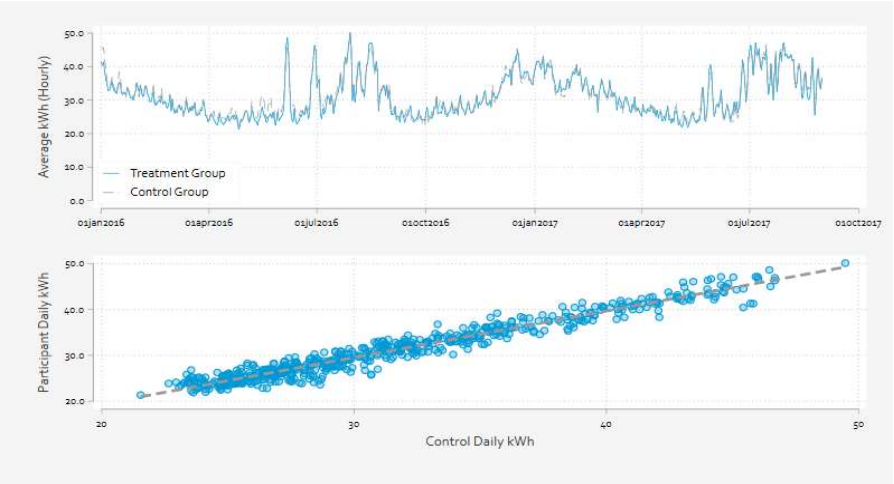
THE ABILITY TO DETECT IMPACTS IS A FUNCTION OF FOUR MAIN FACTORS

1 **The effect or signal size.** Bigger percent changes are easier to detect than small ones

2 **The inherent volatility and background noise in the data.** Impacts are usually easier to detect in end use data than in whole building data

3 **The ability to filter out noise or explain variability.** Models, control groups, and quasi-experimental designs are simply tools to filter out noise and allow the effect to be more easily detected

4 **Sample/population size.** The more sites that are included the less noisy the results



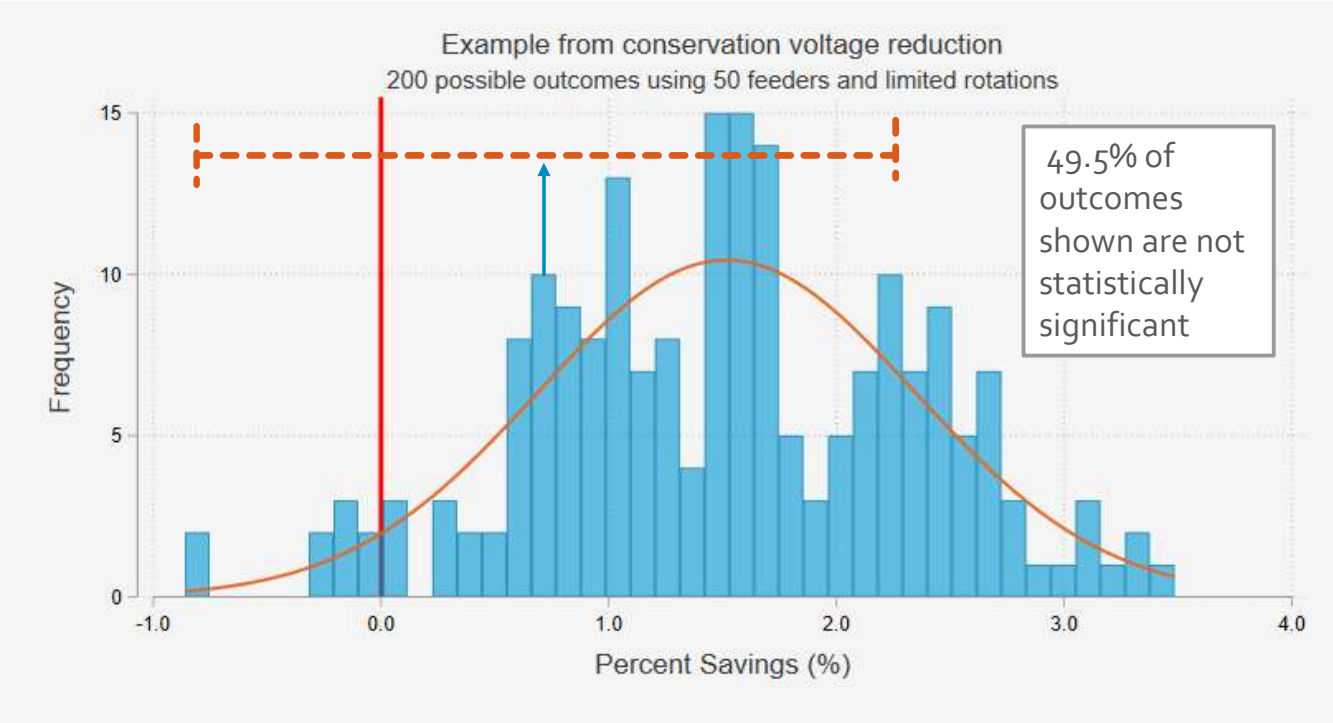
Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

WHEN WE RUN AN EXPERIMENT OR PILOT, WE ATTEMPT TO AVOID TWO TYPES OF ERRORS

Source: https://medium.com/@neeraj.kumar.iitg/statistical-performance-measures-12bad66694b7

- The focus is often on avoiding false positives:
  - By default, we set the probably of false positives at 5% (a 95% interval)
- But failure to identify an effect if it exist, can be just as costly and lead to bad decisions.
  - Insufficient sample sizes
  - Improper experimental design
  - Use of poor models
- We often can't repeat an experiment thousands of times

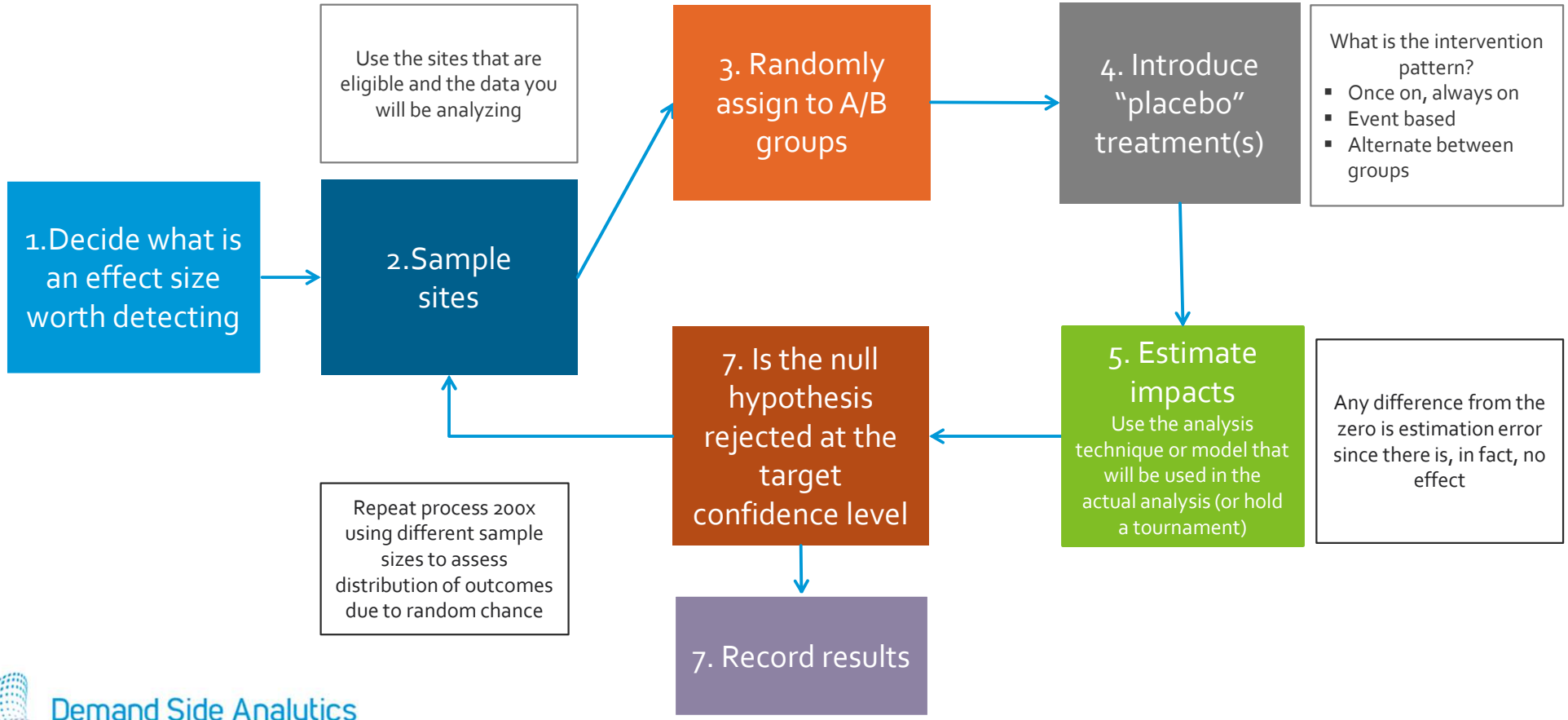
WHAT EXACTLY IS POWER ANALYSIS?



- **Statistical power.** This reflect the ability to detect a meaningful effect size and conclude it is statistically significant
- **Effect size.** The study is designed to detect a change in energy (or participation) of 1.5%. This often based on some preliminary cost-effectiveness screening
- **Sample size.** In this case, we have a limited number of circuit feeders
- **Margin of error.** Also know as the confidence interval. A metric for precision.

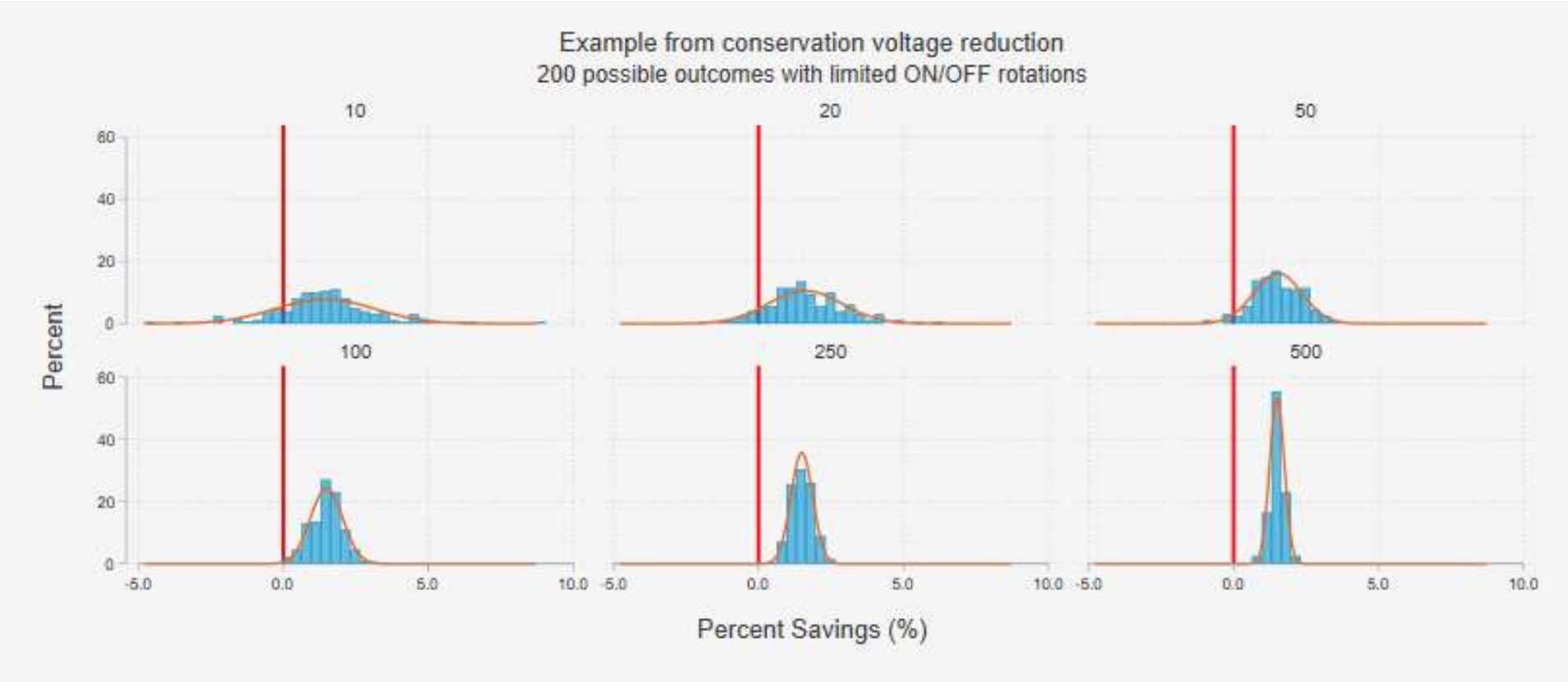
Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

HOW DO YOU IMPLEMENT IT?



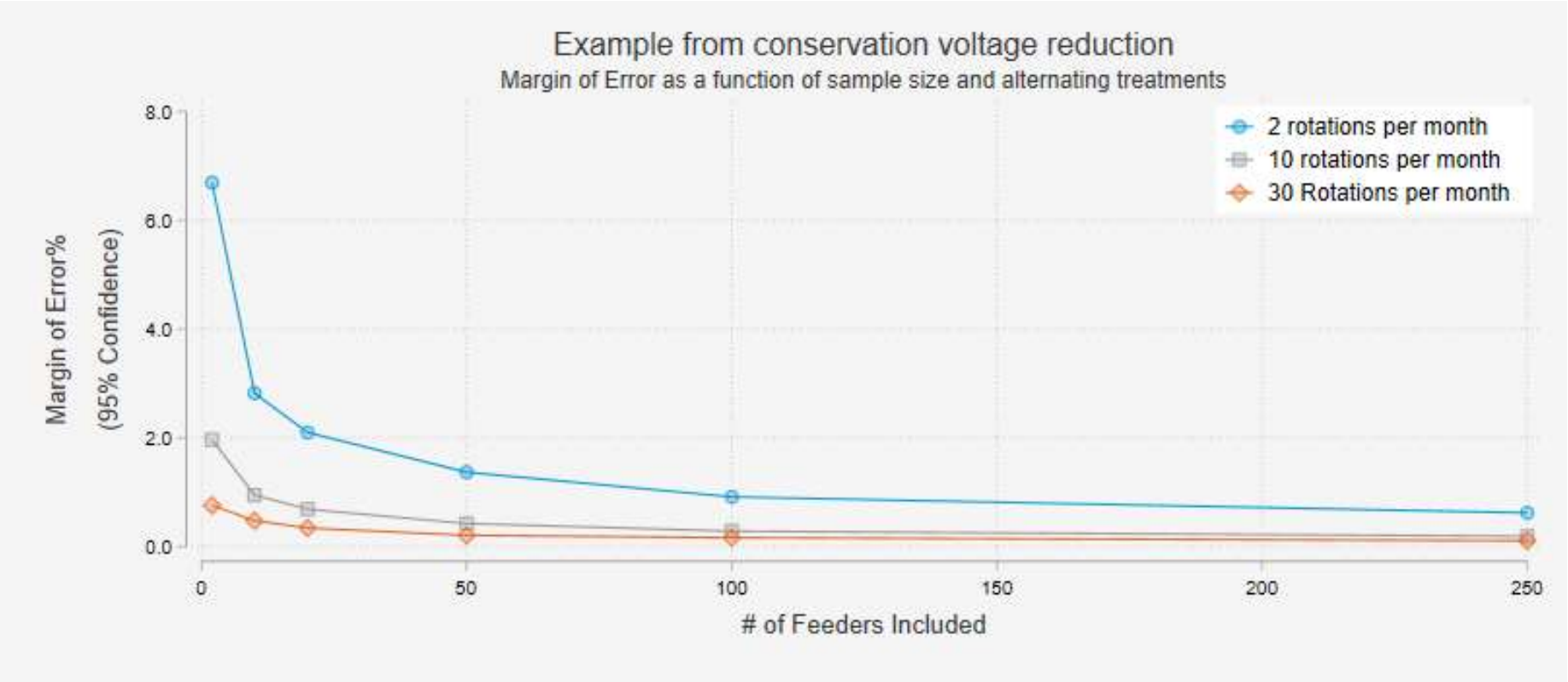
Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

KEY OUTPUTS – DISTRIBUTION OF ERRORS



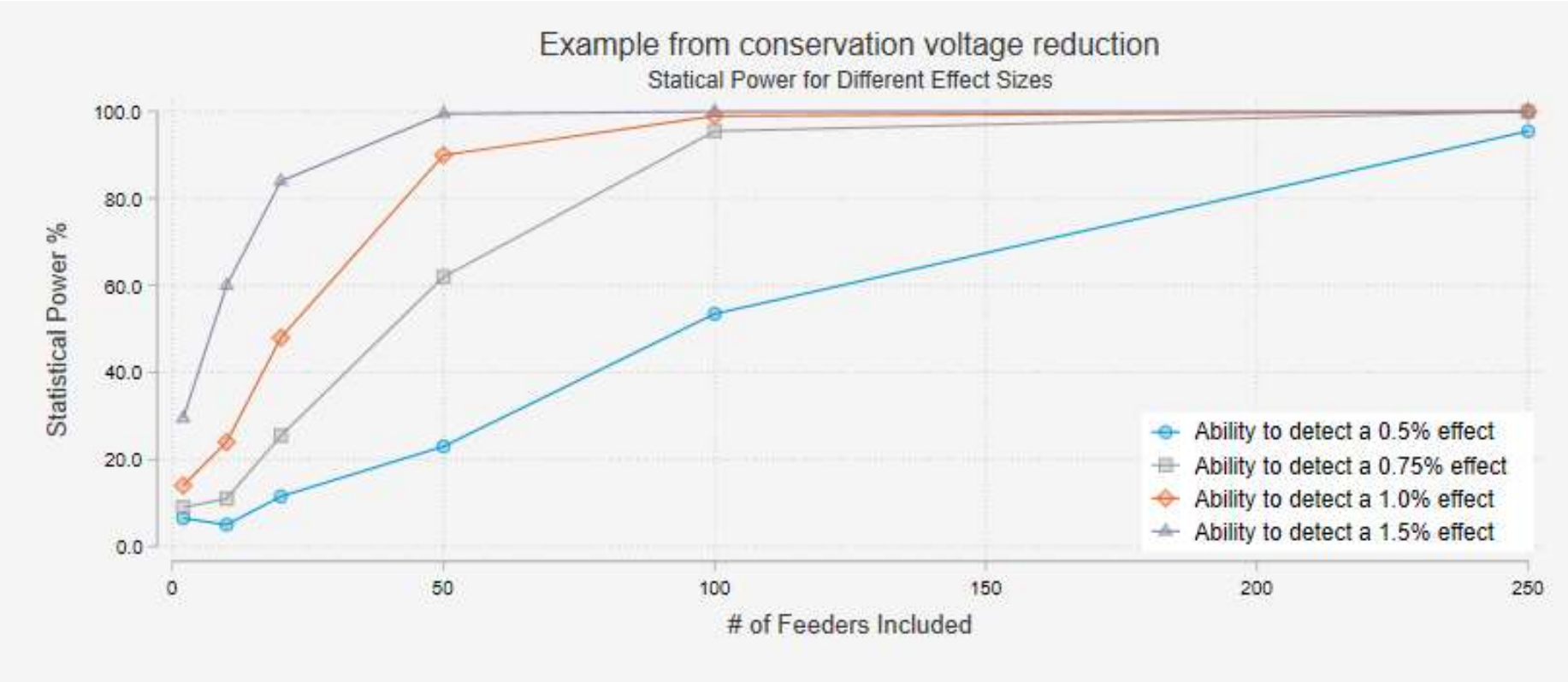
Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

KEY OUTPUTS – MARGIN OF ERROR



Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

KEY OUTPUT – POWER CURVES



Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

KEY OUTPUT – POWER TABLES

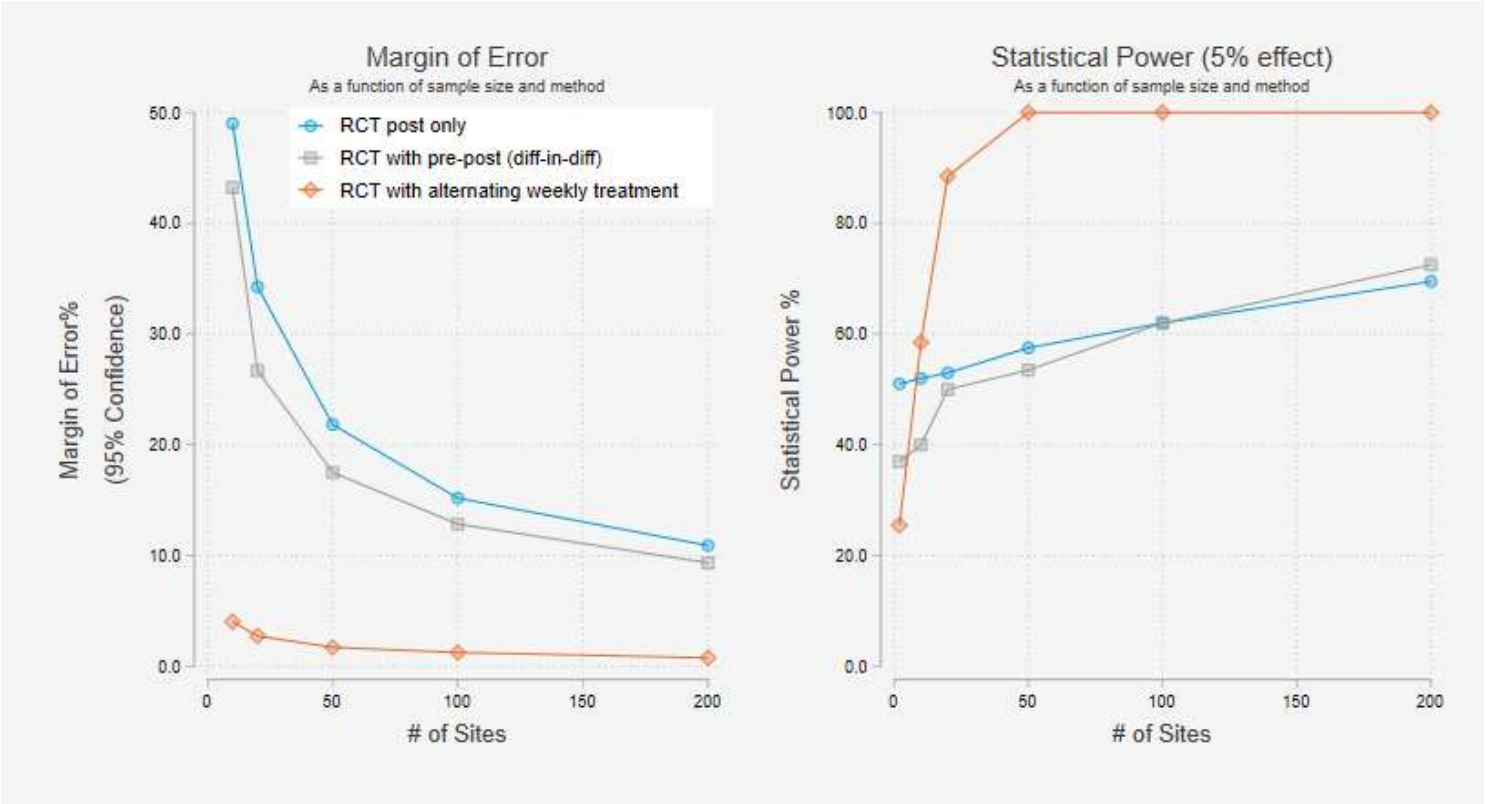
# of Feeders	# of A/B Rotations 2	5	10	30
2	9.0%	19.5%	29.5%	61.0%
10	16.0%	33.5%	60.0%	89.5%
20	23.5%	49.0%	84.0%	99.5%
50	49.5%	87.5%	99.5%	100.0%
100	78.5%	97.0%	100.0%	100.0%
250	98.0%	100.0%	100.0%	100.0%
500	100.0%	100.0%	100.0%	100.0%

- There are multiple acceptable solution
- Use more feeders, or dispatch and rotate the CVR more often
- One of the solution is far cheaper

Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

COMPARING DIFFERENT DESIGNS

BATTERY STORAGE – SAME TIME PERIOD, SAME MODEL, SAME DATA



- Data: AMI data for sites with battery storage for full time frame
- Time Period: May-Sep 2021
- The different research designs were applied to the same sample pulls
- The alternating treatment design is far more powerful

Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

COMPARING DIFFERENT MODEL SPECIFICATIONS BEHAVIORAL PROGRAM EXAMPLE

Treatment and Control Group Sizes	95% Confidence Interval (One-sided)		
	Standard fixed effects diff-in-diff	Lagged Dependent Variable	Pre-treatment seasonal use interacted with month
1,000 each	1.69%	1.75%	1.65%
5,000 each	0.74%	0.74%	0.69%
10,000 each	0.52%	0.51%	0.48%
15,000 each	0.42%	0.41%	0.39%
20,000 each	0.37%	0.36%	0.33%

- The same process can be used to compare different model specifications and analysis techniques
- The table shows the difference of three common models used to detect savings for behavioral programs
- The model specification was a subject of much debate at the time, but difference were minimal

KEY TAKEAWAYS

- The industry as whole is running randomized control trials more often
- The ability to detect an effect if it exists (statistical power) is just as important as avoiding concluding there is an effect when there is none
- Simulation based power analysis is the best way to assess the appropriate sample sizes, research design, and statistical model
- It also requires one to engage in science and define all the key elements up front

QUESTIONS?

Josh Bode
Partner
Demand Side Analytics
jbode@demandsideanalytics.com
415.786.0707

Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS