IEPEC International Energy Program Evaluation Conference

WHEN YOU CAN'T GO FOR THE GOLD

Approaches to Evaluating a Smart Thermostat Demand Response Program

August 9, 2017

Opinion **Dynamics**

Program Overview

- Evaluated a smart thermostat pilot to assess:
  - **Demand Response**
  - Energy Efficiency
  - TOU and Bill Impacts
  - Free Ridership and Willingness to Pay
  - Customer Journey, Engagement and Satisfaction
- Designed as a BYOT program that offered three distinct vendor devices (over 10 devices)
- Enrolled just under 1,500 residential participants
- Called for system reliability rather purposes

Opinion **Dynamics**

Research Approach

Evaluation Approach

- Employed a Randomized Controlled Trial (RCT) approach for demand response (DR) events by randomly assigning participants to treatment and control status for each event
- An RCT eliminates bias from self-selection by participants and from different comparison days
- Distinct from typical approach that uses non-event weather days to serve as reference load

Random Assignment – Group Assignment per Event

- Treatment and control groups change status on different event days

	Group			
Event	A	B	C	D
1	Treatment	Control	Treatment	Control
2	Treatment	Control	Control	Treatment
3	Control	Treatment	Control	Treatment
4	Control	Treatment	Treatment	Control
5	Treatment	Control	Treatment	Control
6	Treatment	Control	Control	Treatment
7	Control	Treatment	Control	Treatment
8	Control	Treatment	Treatment	Control
9	Treatment	Control	Treatment	Control

Research Overview

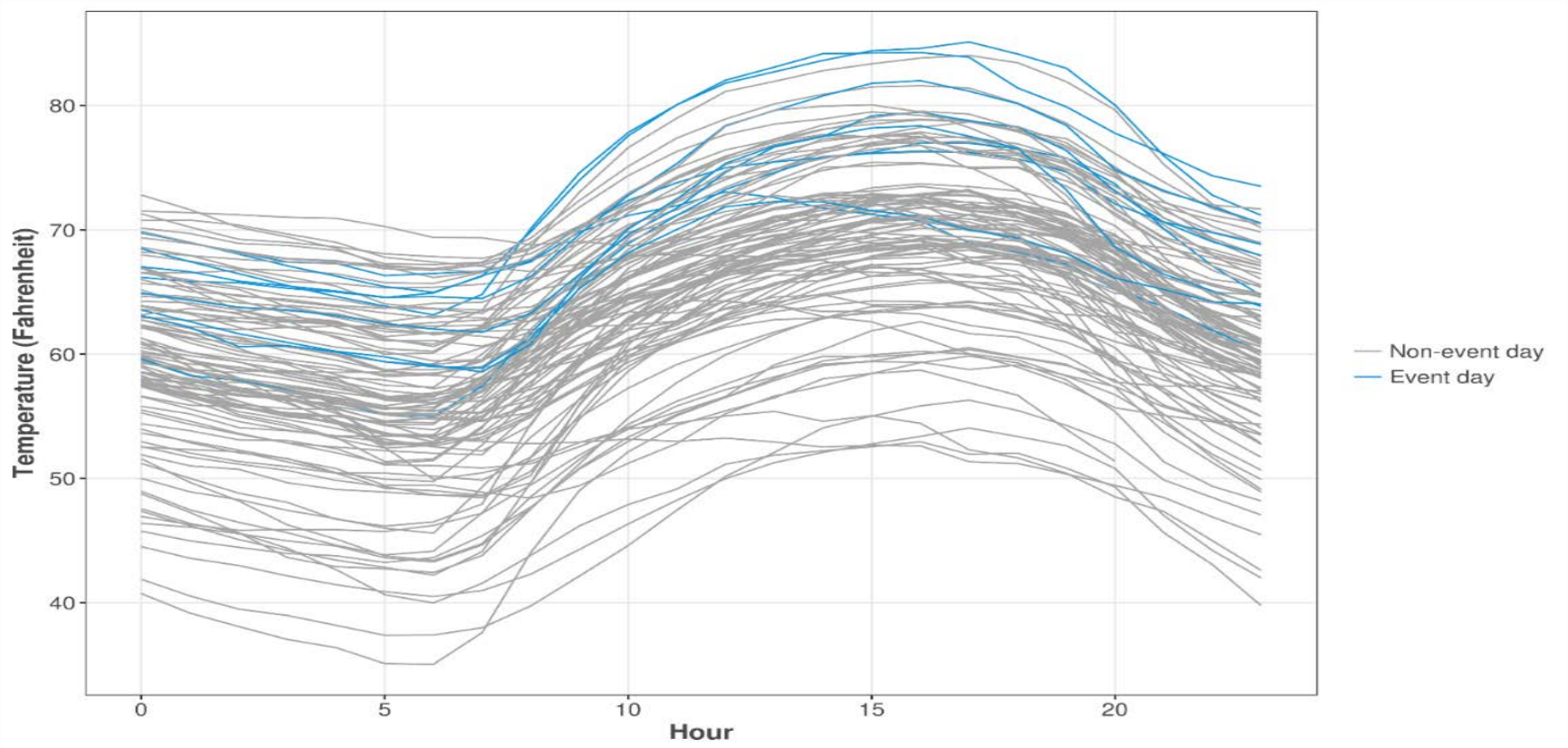
- Focus on DR impacts using two approaches:

Research Design	Matching Approach	Modeling Approach	Comparison
Experimental	Random Assignment	Difference	Event Day
Quasi-Experimental	Mahalanobis Distance Day Matching	Linear Fixed Effects Regression	Similar Day

- Compare results by approach to test accuracy and bias of assessing impacts

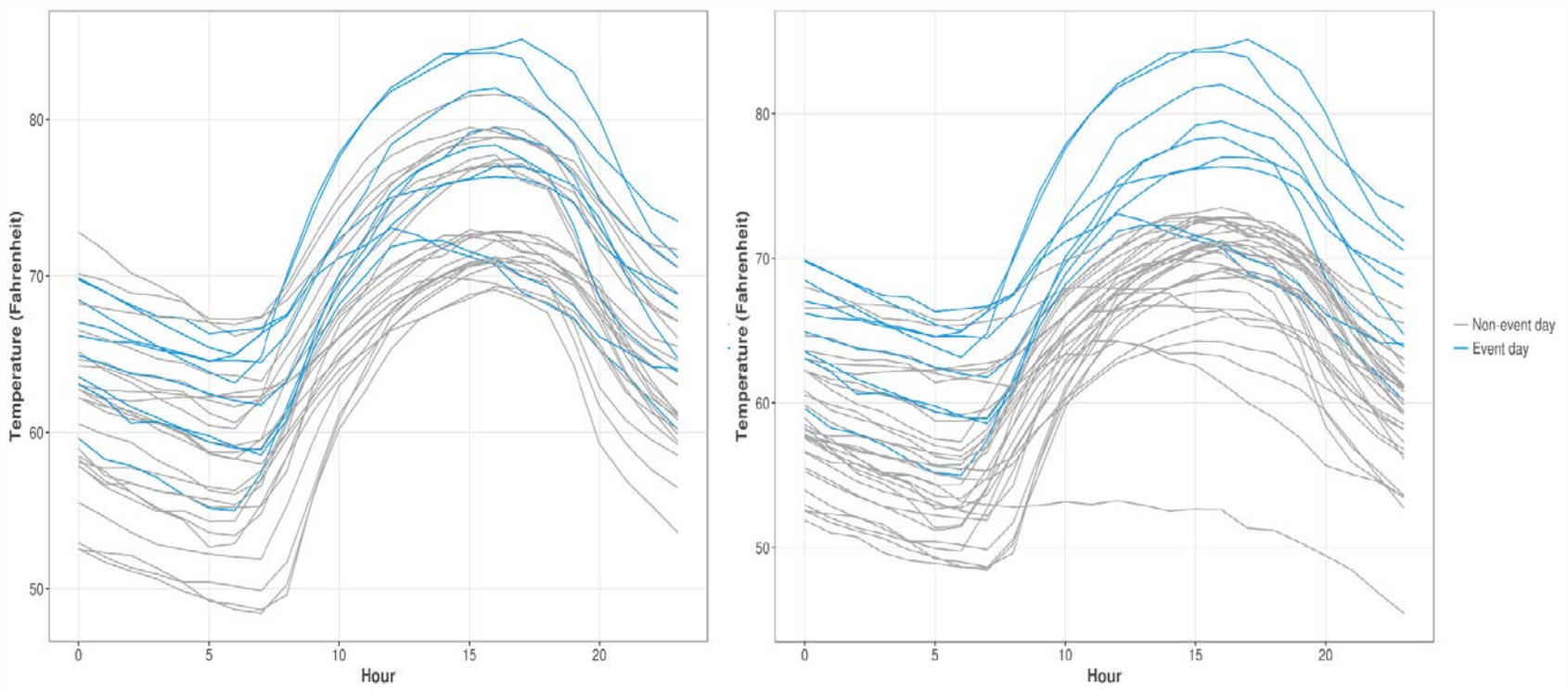
Weather Matching – All Days

Event Day and Non-Event Day Temperatures before Matching



Weather Matching – Matched Days

Event Day and Non-Event Day Temperatures after Mahalanobis Matching
Well-Matched (L), Poorly Matched (R)



Study Results

Results – It's all about the counterfactual

- Matching biases impact estimates downward
- Well-matched weather day results are closer to RCT results than poorly matched days

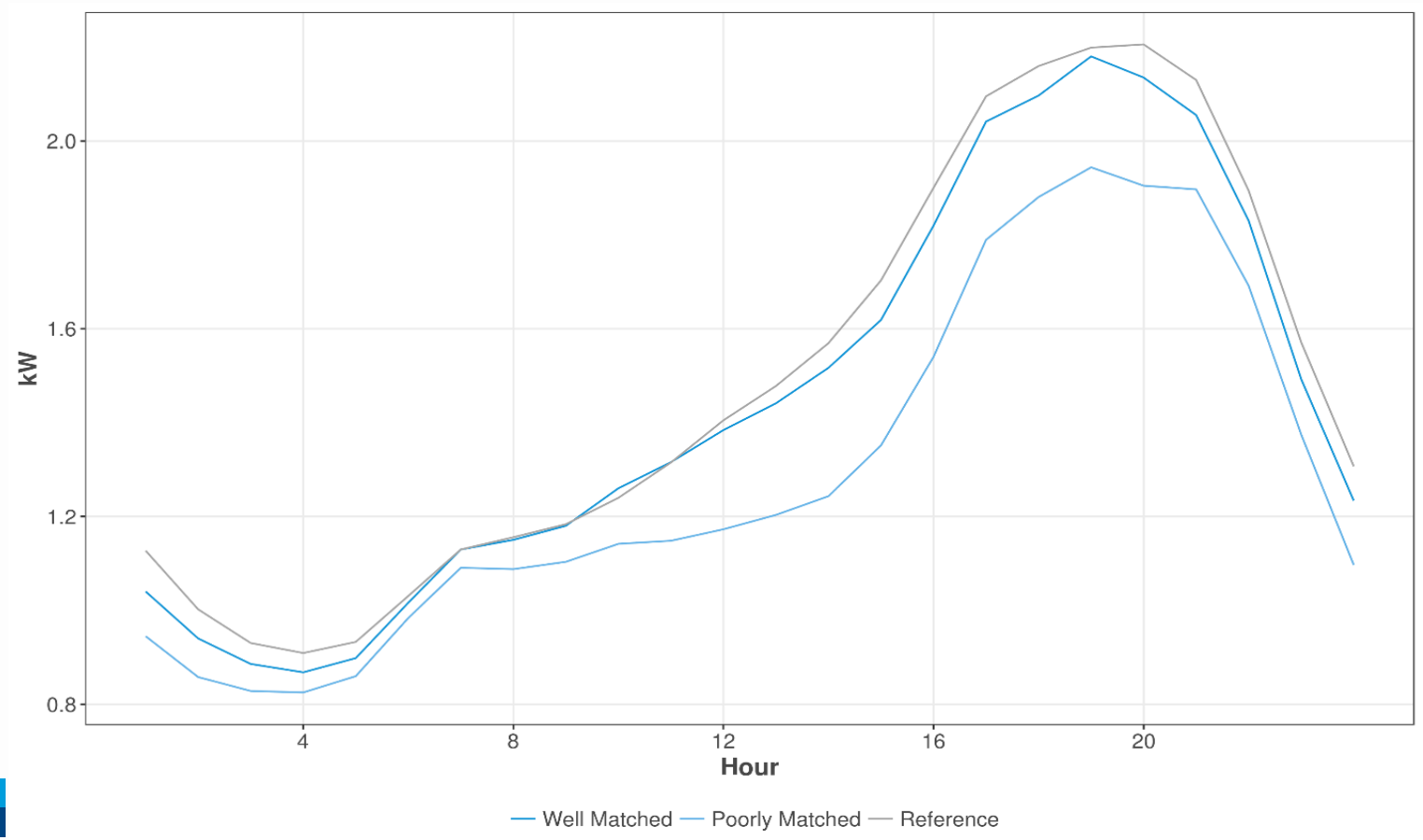
Research Design	Matching Approach	Reference kW	Per T-Stat kW Demand Reduction	Standard Error
Experimental	Random Assignment	1.88	0.45	0.01
Quasi-Experimental	Well-Matched Day Matching	1.75	0.37	0.01
	Poorly-Matched Day Matching	1.48	0.10	0.01

Conclusion 1: Design Matters!

- Experimental design produces the least biased, most accurate impact results
- Helps to assess impacts for smaller population groups (across vendors, devices, etc.)
- Both well-matched and poorly-matched quasi-experimental impacts are biased low compared to RCT
- RCT does a better job at identifying the best control (e.g., the actual weather, humidity, and day, on the day of the event)

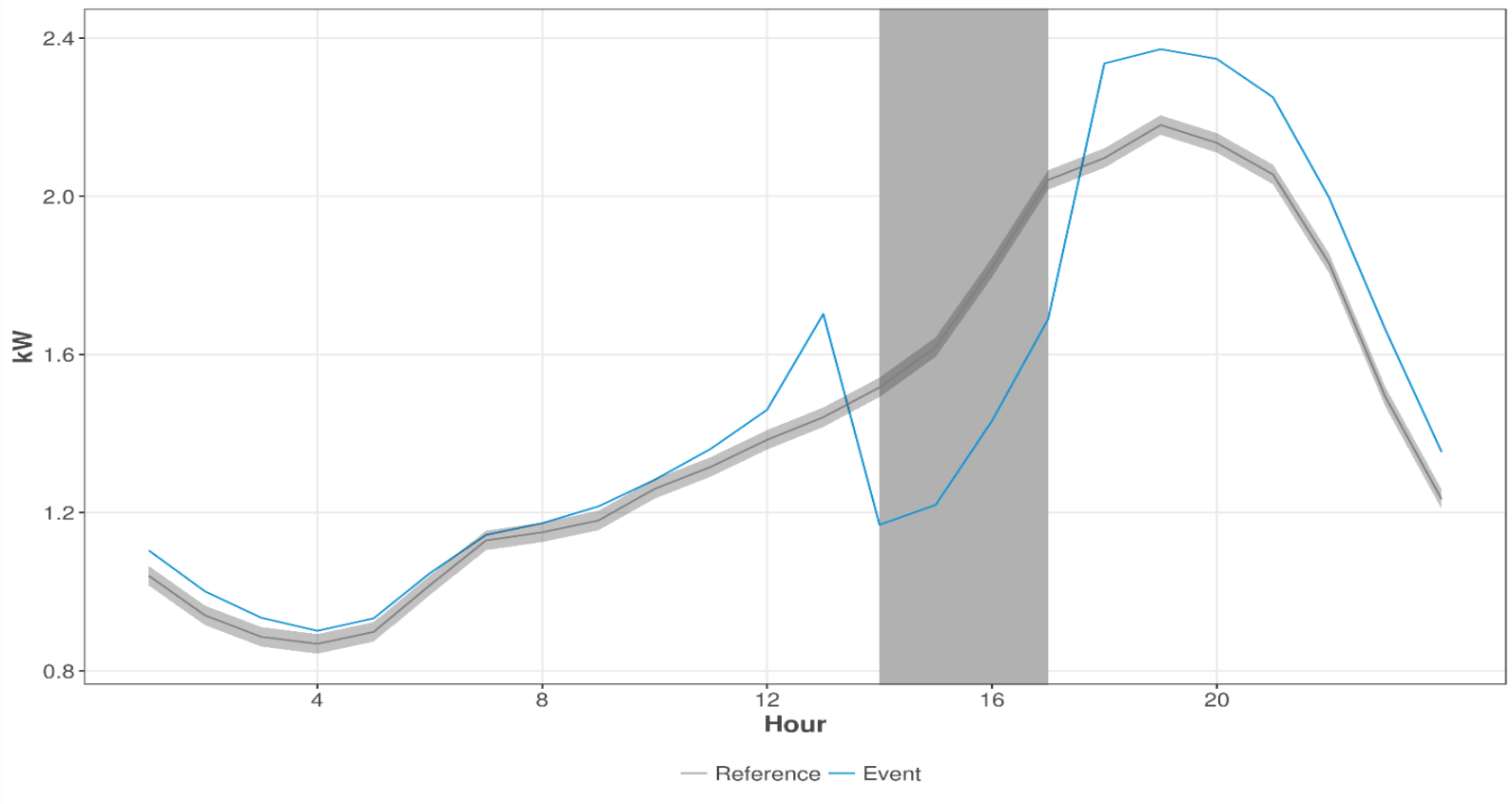
Conclusion 1: A Closer Look at Reference Load

Average RCT Reference vs Modeled Reference Event Day Usage



Conclusion 2: Weather Matters!

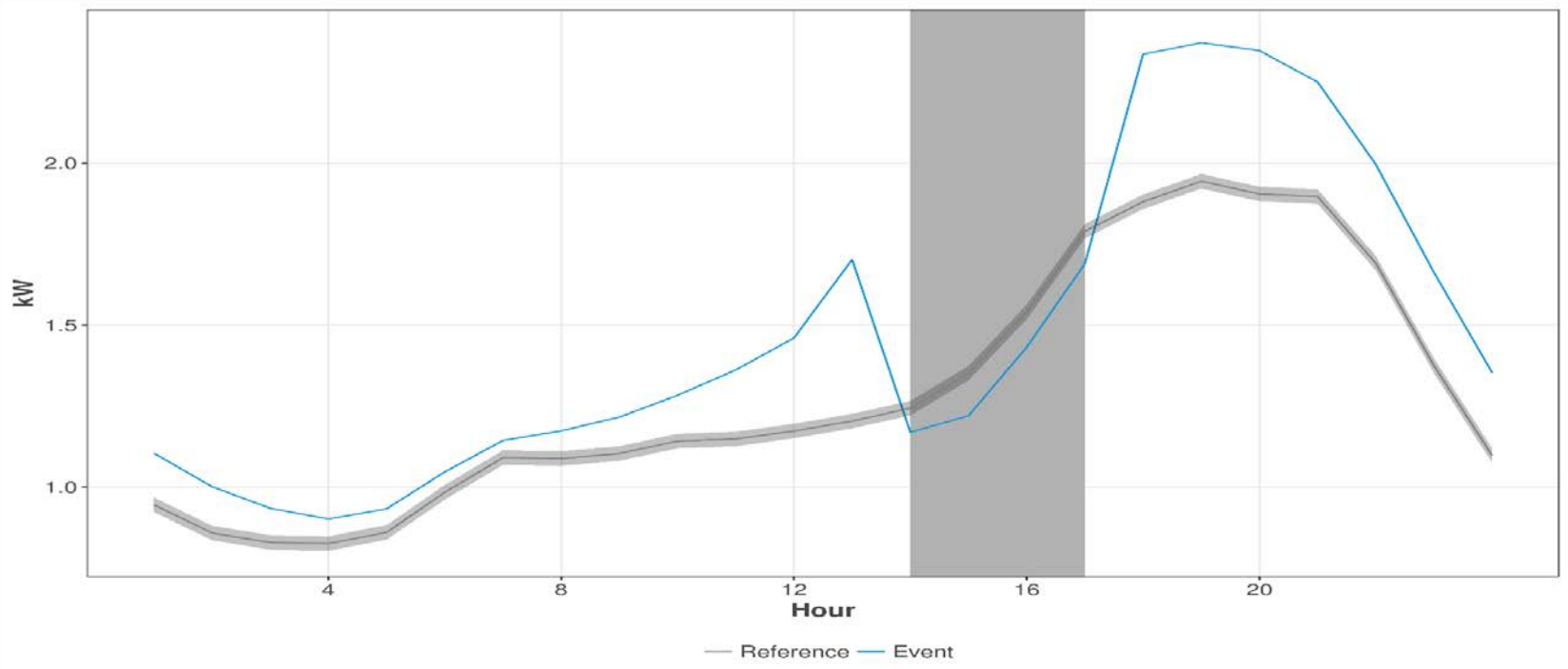
Average 2015 Summer Ex Post Demand Response Event Impacts (Well Matched)



Opinion Dynamics

Conclusion 2: Cool Days Underestimate Reference Load

Average 2015 Summer Ex Post Demand Response Event Impacts (Poorly Matched)



Events are typically called on hottest days, so non-event days will be cooler leaving cooler days with lower demand as comparison days

Study Implications

Implications for Incorporating Experimental Design

- Goal is to design DR event protocols to achieve maximum DR and high degrees of accuracy
  - Requires embedding RCTs in advance of summer event season
- For RCTs:
  - Develop control groups that are appropriately scaled to estimate impacts while maximizing demand reduction
  - Pilots should employ RCT as best practice
  - RCTs can answer other questions related to control strategies, opt-out rates, etc.

Implications for Day Matching

- Carefully select comparison days to accurately predict reference load. During event periods:
  - Embed experimental design for test events or non-emergency events across various weather conditions and use it to help scale results from the day matching approach
  - Report model validation statistics to demonstrate how well the weather days matched event days, and what types of interpretations are appropriate based on the modeled results

Implications for Program Impacts and CE

- Bias matters for program administrators
  - If quasi-experimental results are biased low, evaluators underestimate both impacts and cost-effectiveness
  - Results in potentially cancelling or reducing the size of a program that is performing well or anticipating more load reduction than available
- Reflects increasing risk to administrators as we move from energy conservation priorities to load management efforts

Thank you

Olivia Patterson
opatterson@opiniondynamics.com
510-444-5050 ext 9191

Opinion **Dynamics**