DR Impact Evaluation – Which Design and Analysis Method Is Right for What?

Craig Williamson &
Kelly Marrin

2015 IEPEC Conference — Long Beach, California

What to do, how to estimate DR impacts?

Randomized Control Trials

Fixed Effect Models

WITHIN-SUBJECTS DESIGN

Difference in Differences

Matched Control Groups

Applied Energy Group

What to do, how to estimate DR impacts?

Fixed
Effect

Rando
Contr

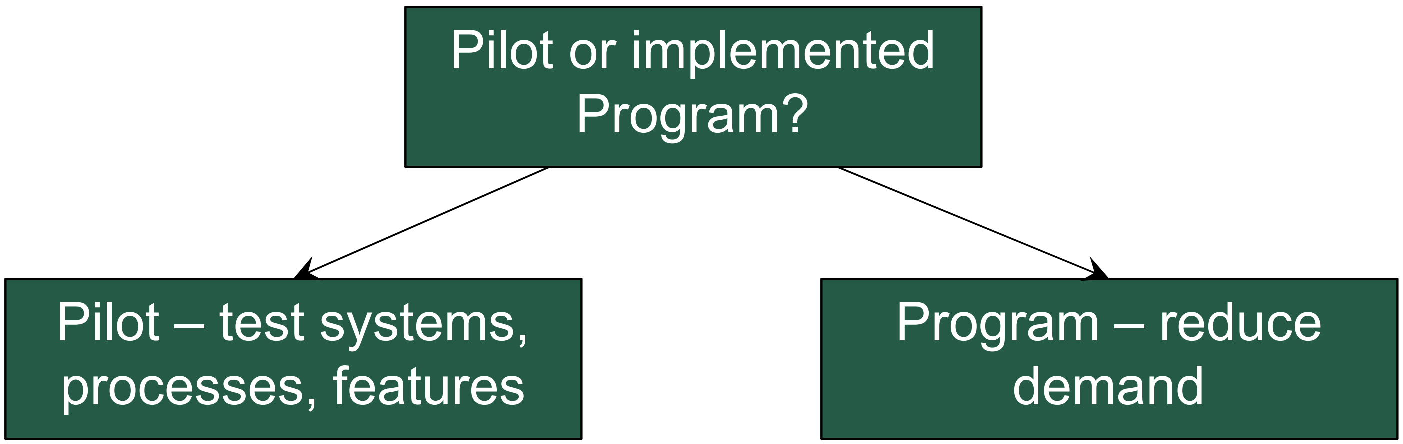
WITHIN-
SU
JECTS
GN

Planning
and Design

e in
ifferences

Matched C
Groups

Start with Type of DR program



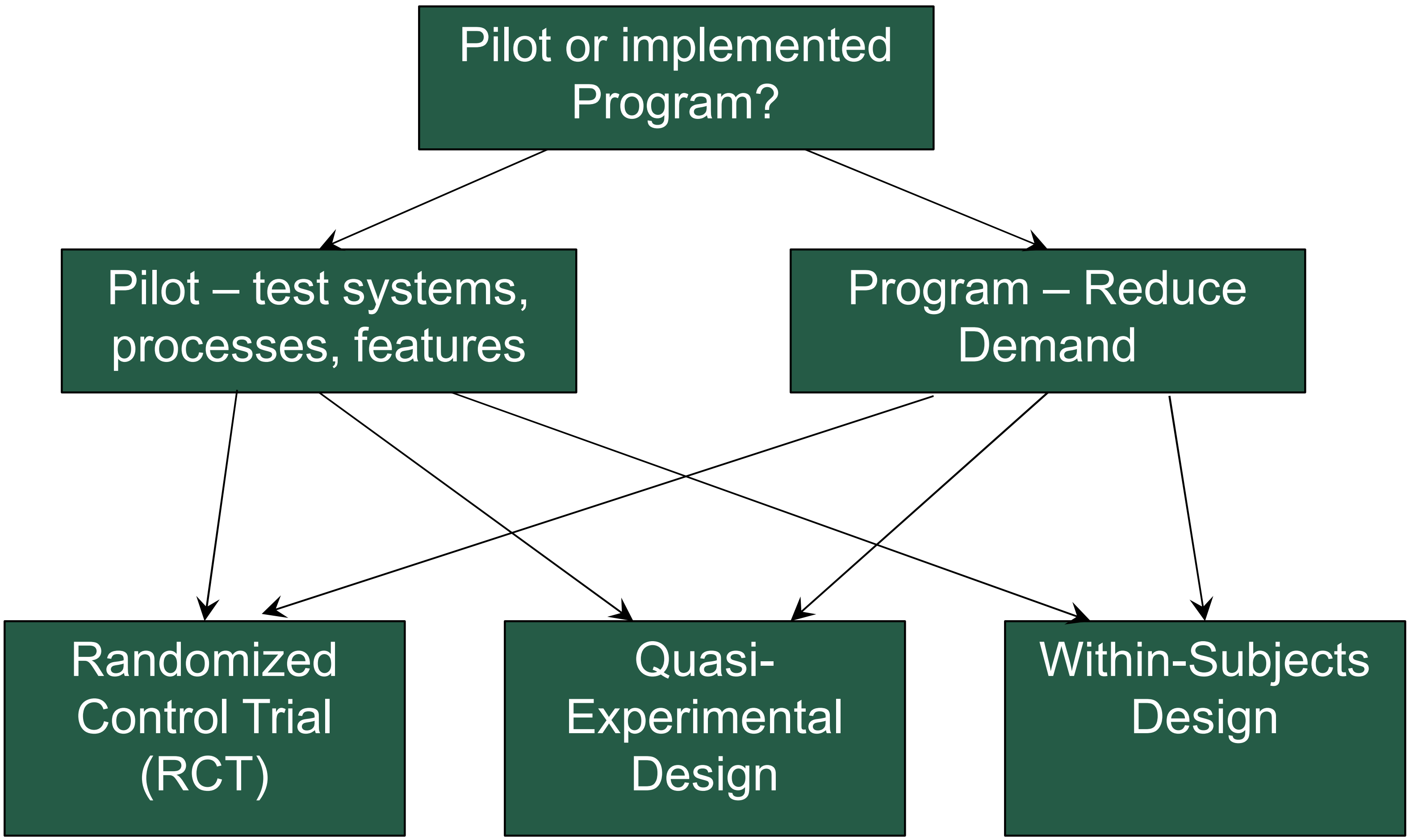
Pilot

- Designed as test or investigation
- Compare or test prices, technologies, processes
- Goal is to answer questions

Program

- Process efficiency is crucial
- Not testing anything – decisions already made
- Goal is to reliably reduce demand

Determine an Experimental Design



Randomized Control Trial

- “Gold standard” for evaluation
- Assignment to treatment and control must be random
  - Must not involve anything related to participation
  - Generally should not be done by program implementer
- Target population must be clearly and explicitly defined
  - Mandatory – entire population, more straightforward
  - Voluntary – treatment and control must both be selected from volunteers for the program, using either recruit-and-deny or recruit-and-delay
  - Targeting certain customers, say higher use customers or customers with central AC, is fine, as long as the randomization is done within the target population

AEG Applied Energy Group

Steps for a Randomized Control Trial

- **Step 1: Design first,** before implementation. Assignment must not consider customer characteristics or customer choices
- **Step 2: Check the randomization.** Compare treatment and control groups based on any measurable characteristics
- **Step 3: Implement with Care.** Avoid excluding customers to the extent possible
- **Step 4: Put treatment and control customers through the same process.** If an unanticipated characteristic makes customers inappropriate to participate, all customers with that characteristic can be removed from both groups, as long as it does not relate to the program
- **Step 5: Account for other program participation.** Avoid double counting of savings from other programs
- **Always stay true to the randomization throughout.**

Randomized Encouragement Design

- Is really just an RCT, with the treatment applied to only a subset of the treatment group
  - There may also be some control group customers with the treatment (i.e. advanced thermostat) – less of an issue with DR
- Customers in the treatment group are encouraged to accept the treatment, but not all will do so
- Analysis must still be done on the entire treatment and control group
  - As a result, impacts are inherently discounted by acceptance rate
- For more on this, see paper and presentation by Lucy Morris & Brian Smith of PG&E from Tuesday morning's 11:00-12:30 session

Quasi-Experimental Design

- Not as rigorous as RCT
- Sometimes the best realistic option
  - Available population, customer satisfaction considerations, etc.
- Select similar customers from non-participant population to be in the control group
  - Match on as many observable characteristics as possible
  - But you can never know how good the match on unobservable characteristics really is
  - Importantly, unobservable characteristics may have driven the decision to participate, and those same characteristics may also drive other energy use decisions

Sample Control Group Matching Algorithm

Allocate control group candidates into enrollment groups and geographic and central AC ownership groups

Make 1st comparison using monthly billing data
(3-to-1 control-to-participant match)

Process and categorize interval data into 4 day types consisting of on- and off-peak hourly data

Calculate average day types for each customer, excluding outlying days

Make 2nd comparison using hourly interval data and average day types, ensure no duplicate matches
(1-to-1 control-to-participant match)

Applied Energy Group

Within-Subjects Design – Two Approaches

- Customer's own load is used as control
- Two basic approaches
  - Prior year comparison
  - Non-event day comparison

Prior year similar days

- Estimates event day impacts
- Captures long term behavioral changes
- Economics, weather, standards can have influence
- Allows use of all extreme days as events

Similar non-event days

- Estimates event day impacts
- Better captures short term effects
- Does not capture long term behavioral changes
- Requires not calling events on some extreme days

Design

- Randomized Control Trial
- Quasi-experimental Design
- Within-Subjects Design

≠

Analysis

- Difference in Differences
- Regression
- Elasticity Models

Three Analytical Methods in General Use

- Difference in Differences
  - Difference calculated between treatment and control both before and after start of treatment
  - Pre-treatment difference between groups is removed from impact
  - Can be calculated directly or using a DID regression
- Regression (usually Fixed Effects)
  - Regression model can account for other factors
  - Weather normalization possible; Often more precise by accounting for more variation in energy use
- Elasticity Modeling
  - Elasticity is modeled using nonlinear regression
  - Models are applied to prices to estimate impacts
  - Can provide scenario modeling

Much more detail and fancy equations with lots of Greek letters in the paper

Difference in Differences

- ***Strengths of DID*****:** Flexibility; simple and direct calculation and estimation; results in unbiased estimates; able to adjust for pre-treatment differences; very few assumptions needed.
- ***Weaknesses of DID*****:** Does not allow for scenarios with weather or other factors; correctly accounting for covariance can be challenging; cannot account for other factors (without a DID regression).
- ***DID is appropriate for*****:** Pilots or programs; RCT; matched control group; within subjects design; ex post impact estimates.

Regression

- ***Regression strengths*: Relatively flexible; able to account for pre-treatment differences using a fixed effect or indicator variable; easily able to incorporate weather; useful for scenario analysis.
- ***Regression weaknesses*: Imposes some assumptions on the data; requires a more expertise and experience to execute well.
- ***Regression is appropriate for*: Pilots or programs; RCT; matched control groups; within subjects design; ex-post, ex-ante, and weather normalized impacts.

Elasticity Modeling

- ***Elasticity model strengths*: Ability to predict rate impacts under different pricing and weather scenarios (the only method of the three that can do this); results include elasticity estimates.**
- ***Elasticity model weaknesses*: Much more complex execution and estimation; many assumptions made about structure and the nature of the relationship of price to energy use; interpretation of results is more complex; not practical for general evaluation purposes.**
- ***Elasticity model appropriate for*: Studies with the goal of estimating price response; pilots only, with RCT, matched control group, or within subjects design.**

Thank you

Craig Williamson
Practice Area Lead
Program Evaluation & Load Analysis
Applied Energy Group
cwilliamson@appliedenergygroup.com
(720) 233-1500

Kelly Marrin
Senior Project Manager
Program Evaluation & Load Analysis
Applied Energy Group
kmarrin@appliedenergygroup.com
(909) 730-7425