To Code Workshop - Part 2

Randomized Control Design Methodology: Applications and Best Practices

April 11, 2017

View from 30,000 feet

1. What are we trying to achieve with the To Code Pilots? A refresher on program evaluations.
2. What are the key requirements for a successful To Code program evaluation?
3. What are the key requirements for a successful randomized control evaluation design?
4. Lessons learned and path forward

Motivating and framing questions

Hypothesis: Much of the existing inefficient stock of equipment can be replaced cost-effectively with more efficient equipment.

1. Is there stranded savings potential in the small and medium-sized commercial sectors?

Yes?

2. Can these stranded savings be cost-effectively realized with To Code incentives?

Yes?

3. How are To Code incentives most effectively deployed?

Part 1

REFRESHER ON PROGRAM EVALUATIONS

We need a credible comparison group!

Goal	Construct **an** estimate of the energy consumption we **would** have observed among participating firms **had the program not been offered**.
Challenge - Part 1	We cannot observe these true "counterfactual' outcomes → We estimate them using data from customers who are not participating.
Challenge - Part 2	For this counterfactual to be credible → Participants and non-participants should be as similar as possible- along observable and non-observable dimensions.
Take-away	Constructing this control group is **difficult** when customers can choose whether to participate or not in the program being evaluated.

With a “clone” business we can identify the impact of the To Code incentives

Beneficiary

“Clone”

Before

Incentive

After

Understanding the To Code challenges:

Goal	What would have happened to owners of boilers **had the program not been offered to them**?
Challenge - Part 1	A given customer either received the offer or did not! → Find a group of customers who did not participate but who are **comparable** to the ones who did.
Challenge - Part 2	Are non-eligible businesses or customers who chose not to participate a credible comparison group?
Take-away	We need to think carefully about how to build the comparison group so it does not over- or under- estimate the impact of the program!

In California, utilities use ratepayer money to incentivize energy efficiency. Important to demonstrate that programs generate savings.

1. **Ex-post surveys** → Recall Bias
2. **Engineering calculations** → Rebound, over-optimism
3. **Regression analysis** → Unobservable confounders

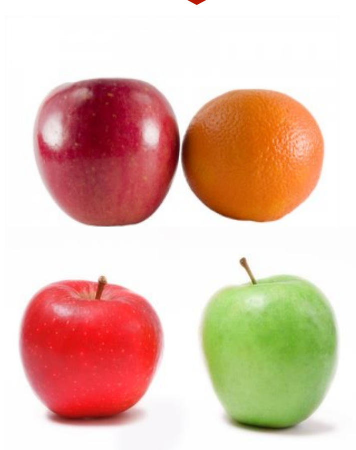
These methods do not allow us to accurately quantify realized savings.

The challenge of non-experimental evaluations

What if we compare…	The problem is…	Which means…
Business X’ s energy usage **before** TC with X’ s energy usage **after** TC?	California’ s economy is growing	X will use more energy to satisfy demand
The change in X’ s energy usage after TC with the change in Y (Y **did not enroll)?**	X is a “green business” and Y is not	X might take many more actions to preserve energy even absent the program.
The change in X’s average energy usage after participating in TC with the change in Y’s (TC **not offered** in Y’s area)	X and Y experience different market shocks	X and Y might have different energy use patterns even absent the program.

How do you choose a comparison group?

Challenge	Approach
How to construct the correct comparison group?	



Approach

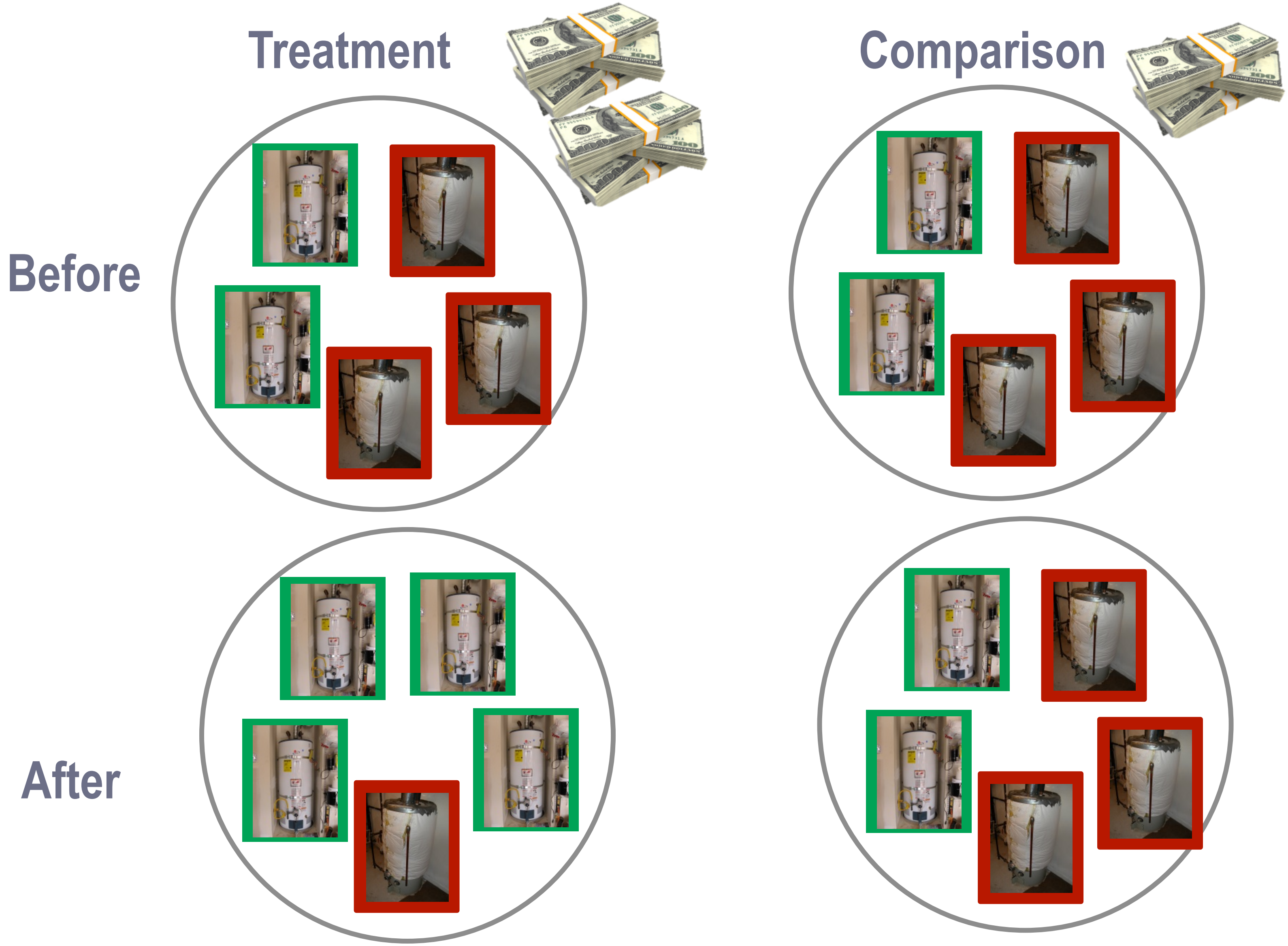
1 Experiments

- RCT
- RED
- Recruit & Delay
- Recruit & Deny

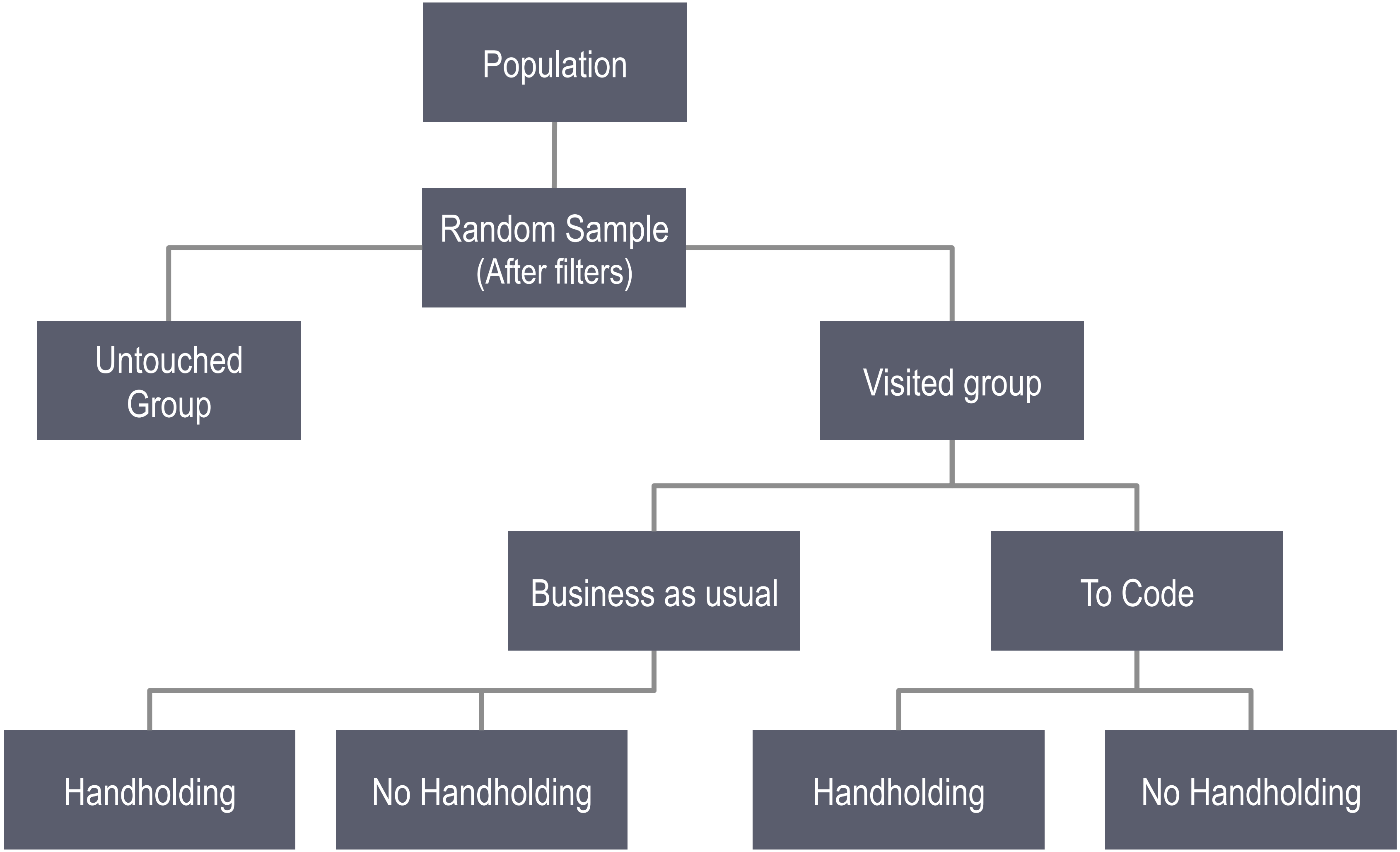
2 Quasi-Experiments

- Differences-in-Differences
- Machine Learning
- Regression Discontinuity
- Matching

Choosing a comparison group



Basic structure of the To-Code pilots



Goals of the pilot

1. Impacts of below code incentives compared to business as usual (BAU) on take-up rates
2. Impacts of "handholding" on take-up rates
3. Energy savings/depth of upgrade
4. Cost-effectiveness

Goals of the pilot

RCT

1. Impacts of below code incentives compared to business as usual (BAU) on take-up rates
2. Impacts of “handholding” on take-up rates
3. Energy savings/depth of upgrade
4. Cost-effectiveness

Are RCTs the only solution?

RCTs	Other Evaluations
Need to plan evaluation strategy and incorporate randomization into program design.	No need to think about evaluation design as program is designed.
↓	↓
Highly credible control group to construct highly credible estimates of program impacts.	Unobservable differences between participants and non-participants will confound your estimates of program impacts.
Really easy to implement	

Are RCTs the only solution?

RCTs	Other Evaluations
Need to plan evaluation strategy and incorporate randomization into program design.	No need to think about evaluation design as program is designed.
Highly credible control group to construct highly credible estimates of program impacts.	Unobservable differences between participants and non-participants will confound your estimates of program impacts.
Really easy to implement	

RCTs require more upfront effort investment.
Pay-off comes after with highly credible results.

To Code Workshop - Part 4

Randomized Control Design Methodology: Applications and Best Practices

April 11, 2017

Part 2

DATA INPUTS TO PROGRAM DESIGN

A successful to-code incentive program (No matter how you evaluate it…)

Targeted equipment should satisfy the following criteria:

Market size	1 High saturation of old, inefficient, replaceable equipment.	
Efficiency properties	2 More efficient alternatives/ replacement exists.	3 Easy to measure savings.
Costs and Benefits	4 Returns on replacement investment exceed the costs (from a social perspective).	5 To code rebate could conceivably accelerate replacement decision given costs and benefits.

Measure type, costs, efficiency properties

	Measures	Costs	Savings
Ideal	High saturation of inefficient and replaceable equipment.	Savings could justify costs of replacement.	Easy to calibrate the rebates.
Reality	Difficult to identify good candidates with available utility data.	In some cases can't offer a rebate for a free good, in some cases prohibitive costs.	For lighting, it is especially difficult to measure savings.
Possible Solution	Collection of richer data on equipment saturation, costs, benefits.	HVAC and lighting: impossible to overcome this	Not the main challenge in our context

Identifying equipment to target is challenging!

	Electric utilities	Gas Company
Challenges	Competitive measures for free, title 24, prohibitive costs, small sample.	Identifying presumptively replaceable boilers (3 “mini-pilots”). • Everyone has an old boiler • There are ~400 boilers”
Steps Taken	Failure to identify presumptively eligible equipment.	Ultimately identified atmospheric burners and watertubes, using CIS, as more likely to be below code. Confirmed by mini-pilot.
Status	With no program design, we have nothing to evaluate.	We can get to work once field staff is ready!

Lessons learned about to-code program design

Collaboration	Identifying measures that are good candidates for the **CPUC requested to-code rebate** is a data-intensive (and labor-intensive!) process.	
Independent Technical Expertise	The devil is in the technical details!	Independent technical consultants make a great difference!
IOU Managers	Implementation challenges inevitably come up.	Resourceful IOU managers can help navigate these challenges.
CPUC collaboration	A dynamic and complex policy environment can throw some curve balls	CPUC collaboration helps anticipate program interactions and constraints.

Part 3

DATA INPUTS TO EXPERIMENTAL RESEARCH DESIGN

Low adoption rate is a concern with program and evaluation design

Power calculations: precise estimation of effects

1 **Can we use sample = 30?**

No

An unbiased estimate of program impacts is **not very useful if it is very imprecise**.

2 Precision depends critically on the **expected take-up rates**.

3 **Good data** on baseline levels of eligible equipment saturation is **absolutely essential** for good calibration.

Power calculations (Appendix)

Part art, part science

Educated guesses better than nothing!

Mini-pilots can be used to adjust/refine assumptions.

Low adoption rate is a concern with program and evaluation design

Power calculations: precise estimation of effects	Part art, part science

1 **Can we use sample = 30?**

No

An unbiased estimate of program impacts is **not very useful if it is very imprecise**.

2 Precision depends critically on the **expected take-up rates**.

3 **Good data** on baseline levels of eligible equipment saturation is **absolutely essential** for good calibration.

Power calculations (Appendix)

Educated guesses better than nothing!

Mini-pilots can be used to adjust/refine assumptions.

We cannot evaluate if there is so little interest.

We also probably do not want to run a program like this.

Inputs to a successful randomized evaluation design

Any good program design AND good program evaluation design requires:

1. Estimates of upgrade costs
2. Estimates of upgrade energy savings
3. High customer acceptance rates of baseline rebates
4. Likely acceptance of incentives

Additional requirement for randomized program evaluation:

5. **An opportunity to randomly vary program offers or program recruitment.**

Once we have a viable program design... cue randomization!

Randomization should only vary elements the IOU would use in the real world

Randomize the customer offer

1. Vary the rebate level across participating customers.
2. Vary the form of intervention (e.g. rebate structure, handholding)

Randomize recruitment/encouragement

1. Vary the **level** of effort in engaging/recruiting customers.
2. Vary the **type** of effort in engaging/recruiting: in-house workforce vs. third-party implementers

Gas pilot varies 1-3.

Once we have a viable program design... cue randomization!

Randomization should only vary elements the IOU would use in the real world

Randomize the customer offer

1. Vary the rebate level across participating customers.
2. Vary the form of intervention (e.g. rebate structure, handholding)

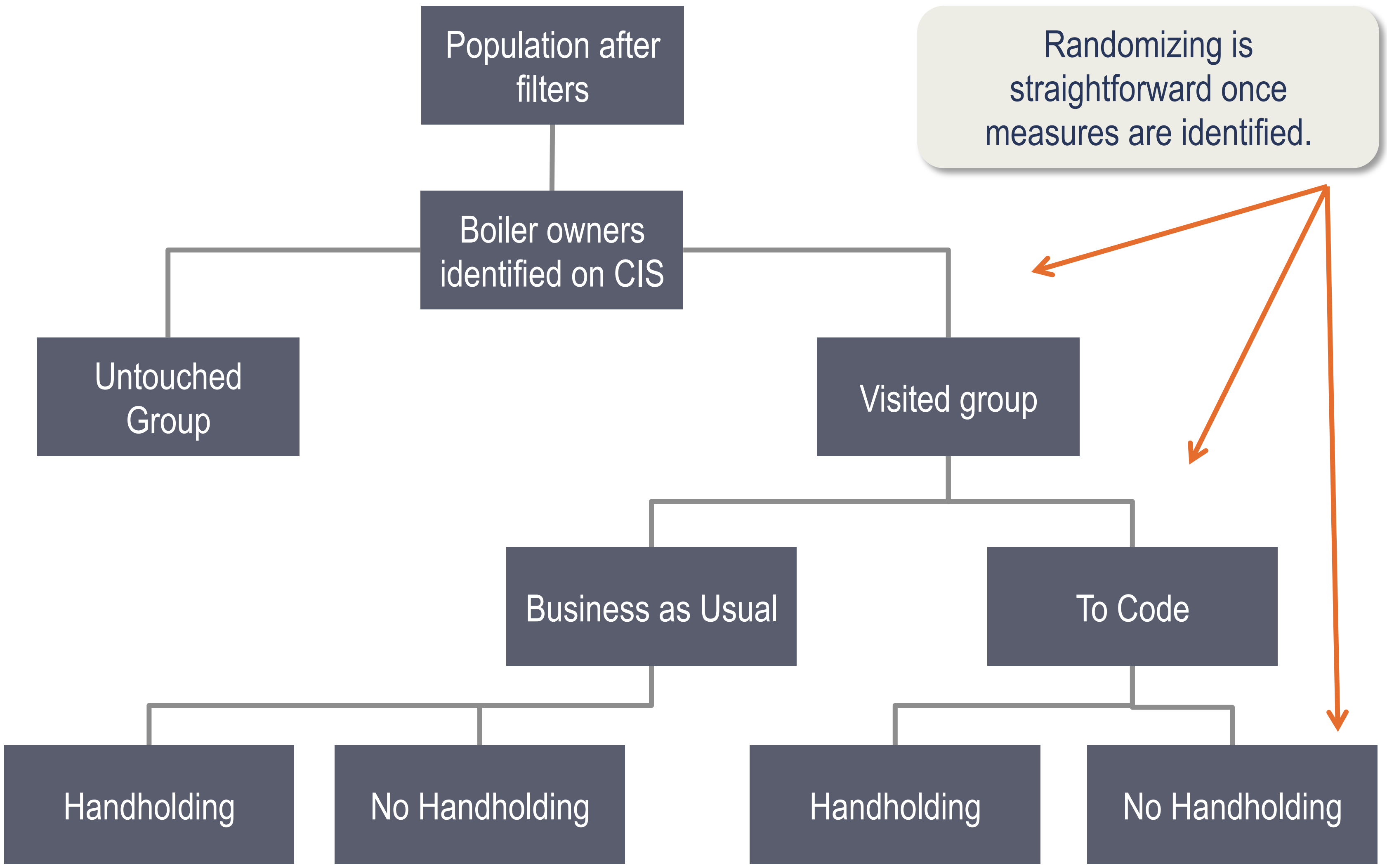
Randomize recruitment/encouragement

1. Vary the level of effort in engaging/recruiting customers.
2. Vary the **type** of effort in engaging/ recruiting: in-house workforce vs. third-party implementers

Adding treatments requires a larger sample size and might cost more.
The final decision on whether to add these treatment arms depended on cost of recruiting, cost of implementing, etc.

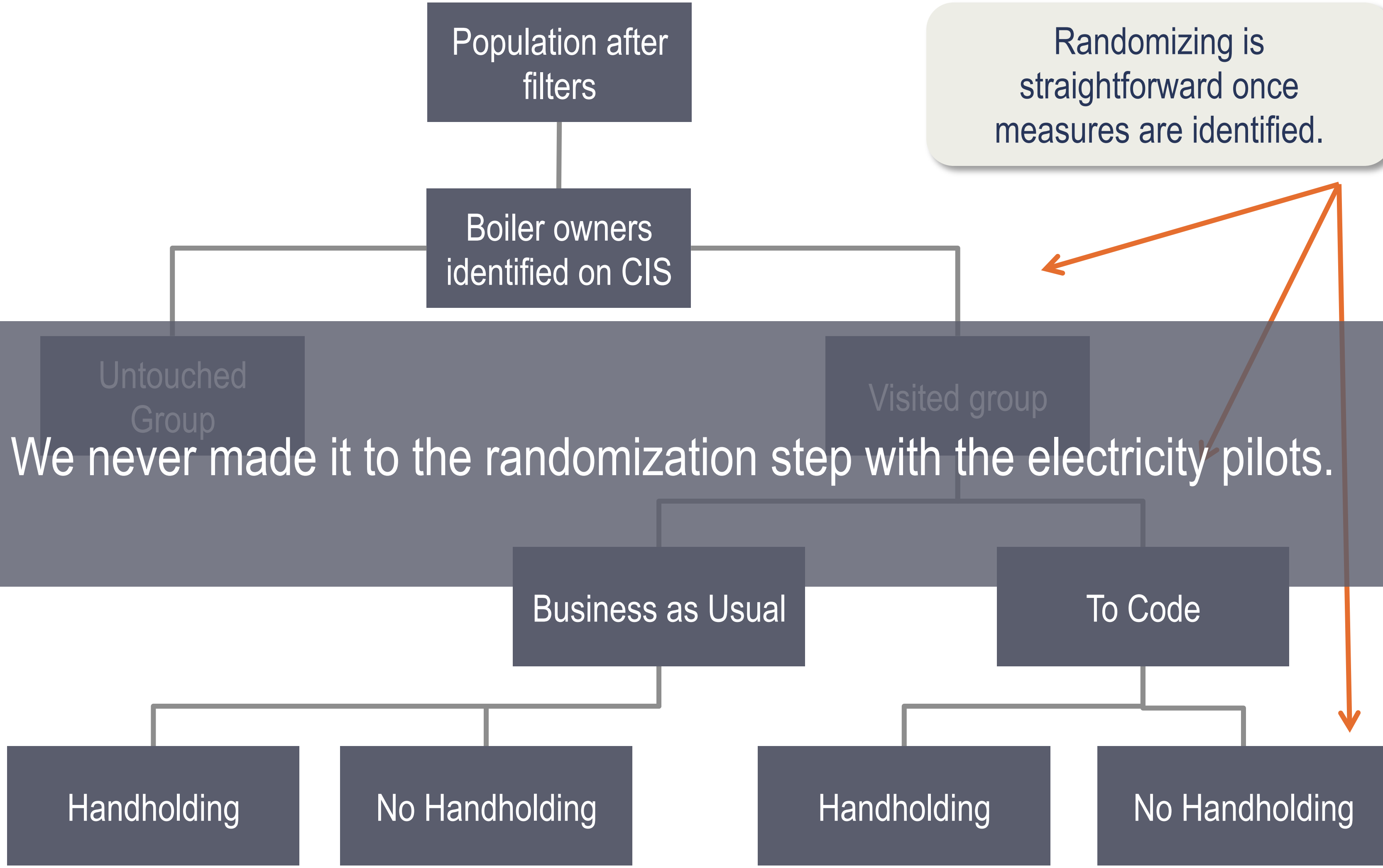
Gas pilot varies 1-3.

Recap of the Gas Pilot



Recap of the Gas Pilot

Randomizing is straightforward once measures are identified.



We never made it to the randomization step with the electricity pilots.

Punchline

To Code was a really tough nut to crack.

Hard to demonstrate the randomized evaluation concept with no viable electricity measures to evaluate!

Part 3

GOING FORWARD

Lessons learned

Why did the **gas** pilot get off the ground?

1. IOU partners (Thank you, Juan!)
2. Reasonable data on equipment saturation
3. Successful identification of measures that satisfy program criteria!

Being implemented

Why are **electricity** pilots still grounded?

1. To code rebates can't compete with free measures
2. Some measures: saturation rates too low
   - Pool pumps
   - Schools and thermostats
3. Other measures: replacement costs prohibitive
   - Cranes
   - Shutting down streets

Strong IOU partnerships notwithstanding, we could not identify measures that fit to code criteria.

What is the counterfactual evaluation?

There are trade offs between RCTs and traditional approaches

1. High upfront engagement
2. Highly credible after-the-fact estimates

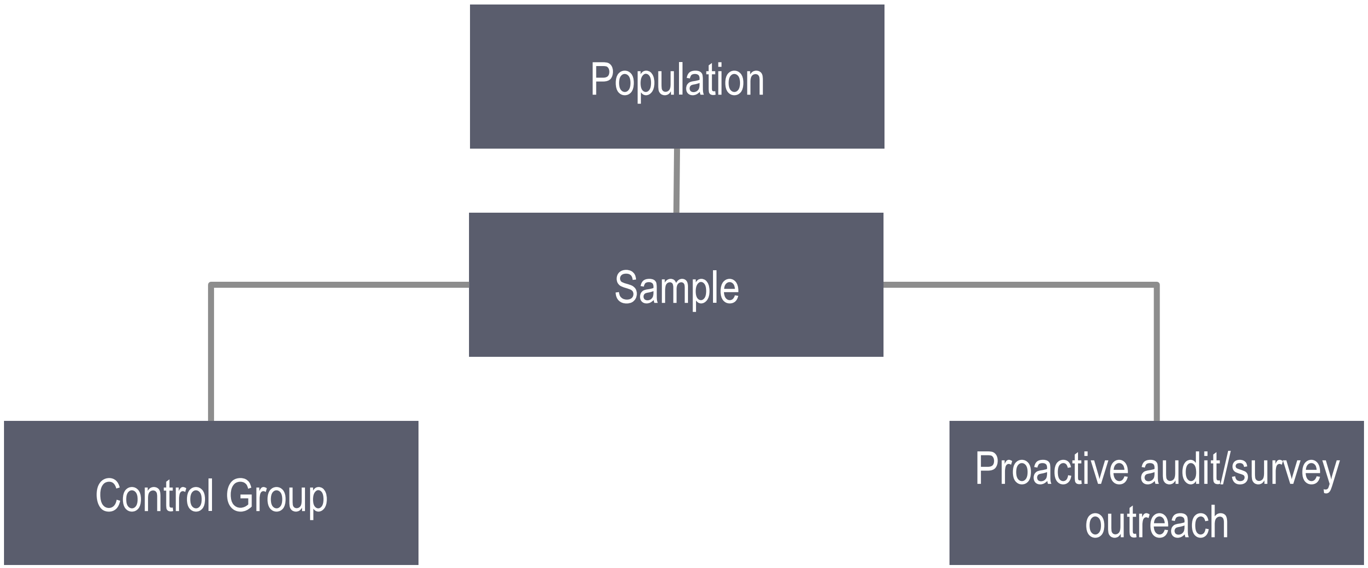
What is the counterfactual outcome of these To Code pilots?

3. Failure to launch electricity pilots due to a failure to find presumptively eligible measures/equipment.

Going forward

Collaboration	Repeated interactions between researchers and utility implementers	Thank you Juan, Shawn, Loan, Naila, Mary, Megan, David, and Edmond!
Data	RCTs require up-front investment in data collection to inform program design	Increased, systematic data collection efforts could support future program design
Other evaluation opportunities	To Code was a hard nut to crack!	We hope there are more opportunities to work together

Going forward:
Build on this collaboration to fill some data gaps



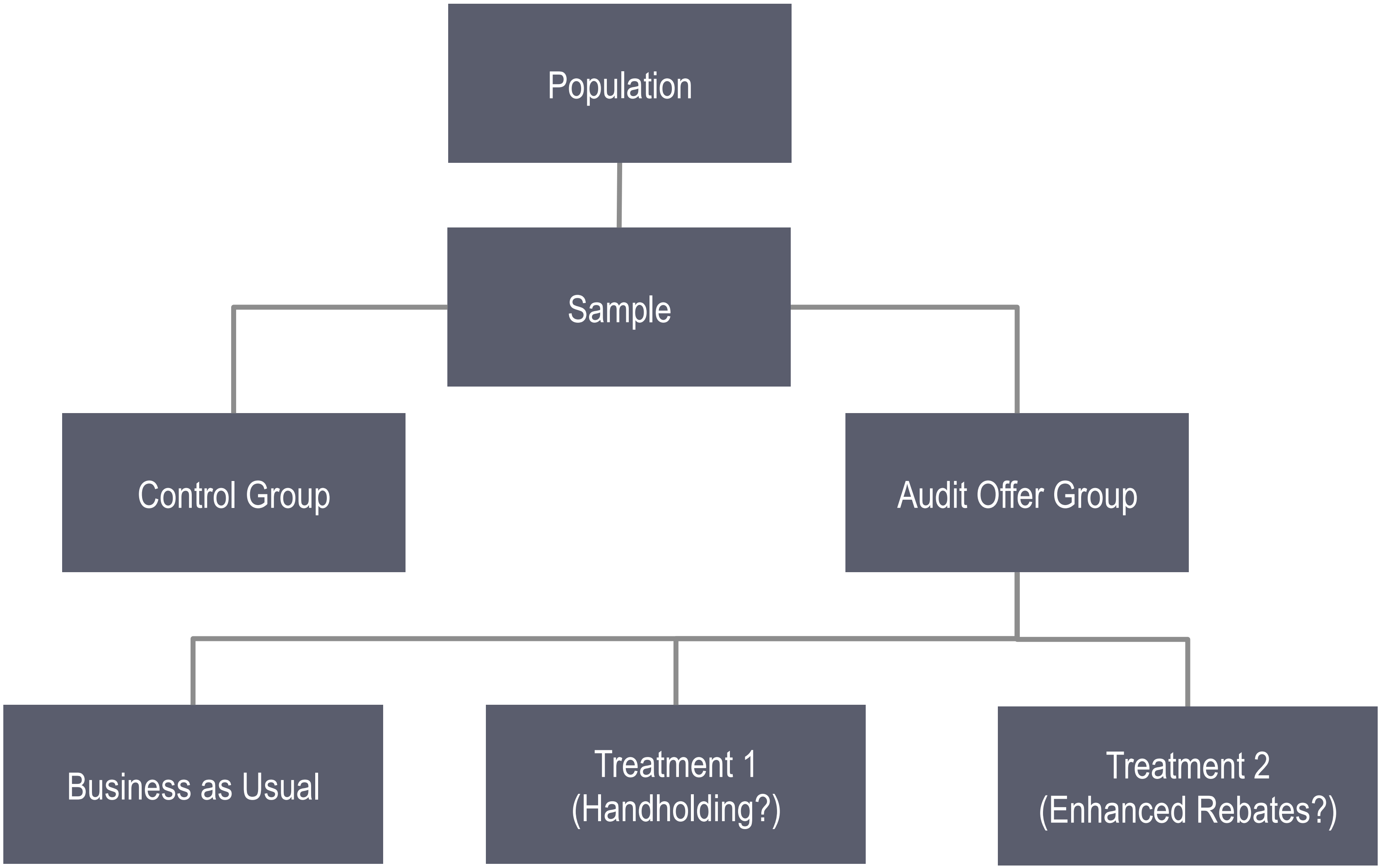
1

Audit/survey intervention for C/I customers to collect information about existing equipment/efficiency potential.

2

Random assignment of audit intervention: to assess whether the delivery of an audit affects energy consumption patterns.

Going forward:
Pursue other promising evaluation opportunities



Thank you!
Please feel free to contact us

UC BERKELEY OFFICE

319, Giannini Hall
Berkeley, California 94720
Email: e2e@berkeley.edu
Tel: 510.642.9590
Fax: 510.643.5180

UNIVERSITY OF CHICAGO OFFICE

1155 E. 60th St.
Chicago IL 60637
Tel 773.702.0627

MIT OFFICE

MIT Building E19-411
400 Main Street, 4th Floor
Cambridge, MA 02142-1017
Tel: 617.253.3551
Fax: 617.253.9845

Appendix
Optimal Design Outcomes as MDEs Change