Running Rigorous Evaluations:
Applications to Solar

Catherine Wolfram

Faculty Director – The E2e Project & Energy Institute at Haas

Cora Jane Flood Professor, UC Berkeley

May 2014

Outline

Motivation — The framework — Design considerations — Concluding remarks

Motivation

The case for randomized experiments

Motivation
Why RCTs?

1 **Because they offer a gold standard to measure the impact of a program**
- Measure the real impact of a program in a simple and clean way
- Enable you to identify what really works
- Reveal what's most cost-effective

2 **The use of experiments is becoming more and more widespread**
- Academia
- Governments
- Private Sector

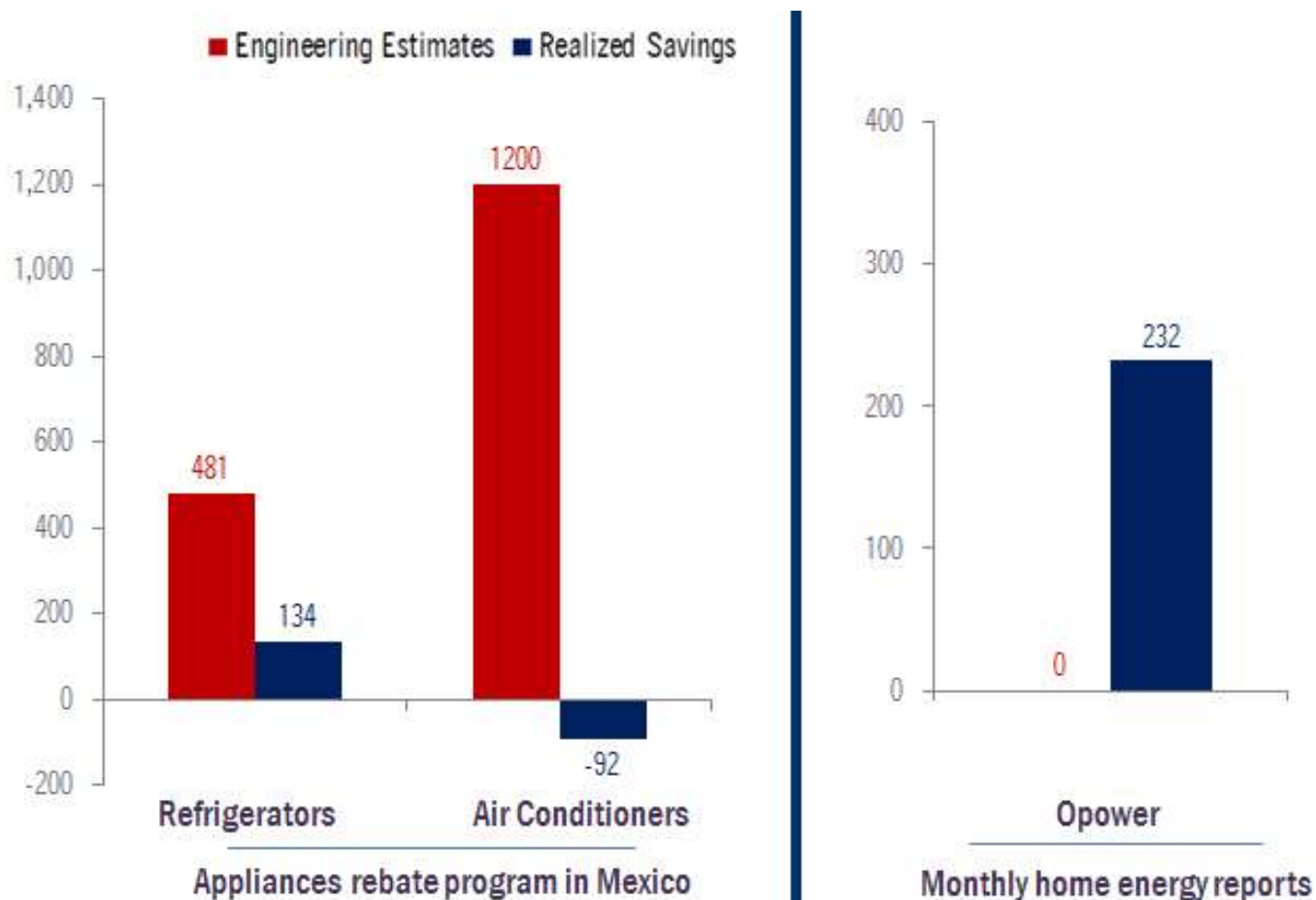**3** **However, RCTs should be done carefully**
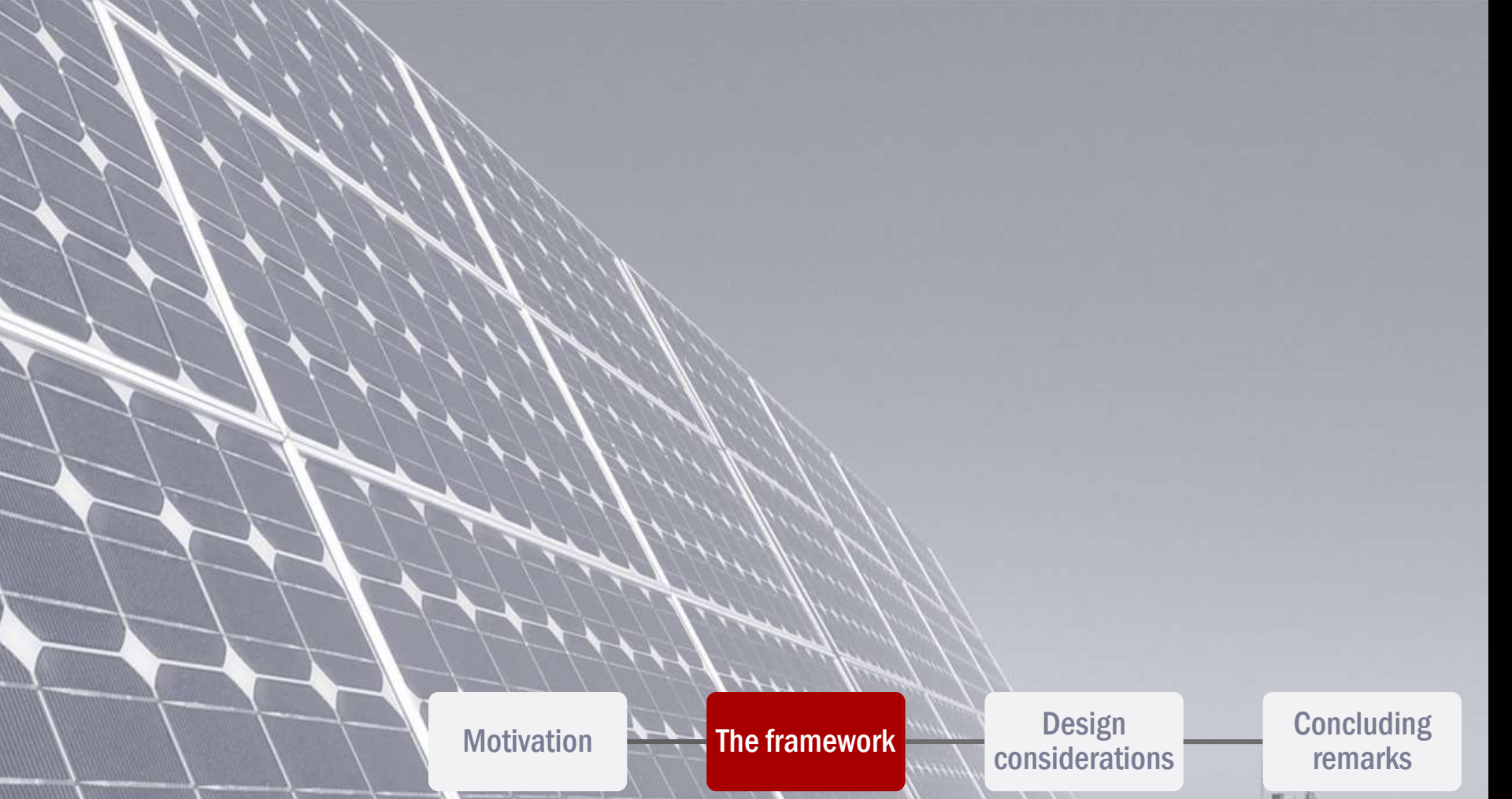- Mistakes will lead to misleading results

Careful measurement can lead to surprising results!

Annual actual vs. estimated savings
kilowatt hours saved per household

The Framework

What are RCTs all about?

How can we measure the benefit of a prescription drug on health?

Meet Meredith

She's an Economics professor at Berkeley

We want to understand the impact of a **cholesterol medicine** on her **health**

What are we trying to measure?

The difference between Meredith's health when she receives the treatment and when she doesn't reveals the impact of the medicine

Problem: We don't observe both of these...

... unless we could clone Meredith and treat one version!

Treat the real Meredith...

... and observe her clone without the medicine

Compare pill-popping Meredith to Catherine, who does not take the medicine?

Observables:

- Women
- Married, 2 kids each
- Similar in age
- Both Econ professors
- Both at UC Berkeley

Is Catherine a good counterfactual?

Problem: Catherine has different genes and lifestyle (worse). So this comparison will overstate the benefit of the pill for Meredith.

But, you CAN "clone" groups

1   **Randomization provides you with *groups* that are, on average, equivalent**

**2**   **On observable and non-observable characteristics**
- **Treatment**: some people with high cholesterol, some people with normal cholesterol
- **Control**: some people with high cholesterol, some people with normal cholesterol

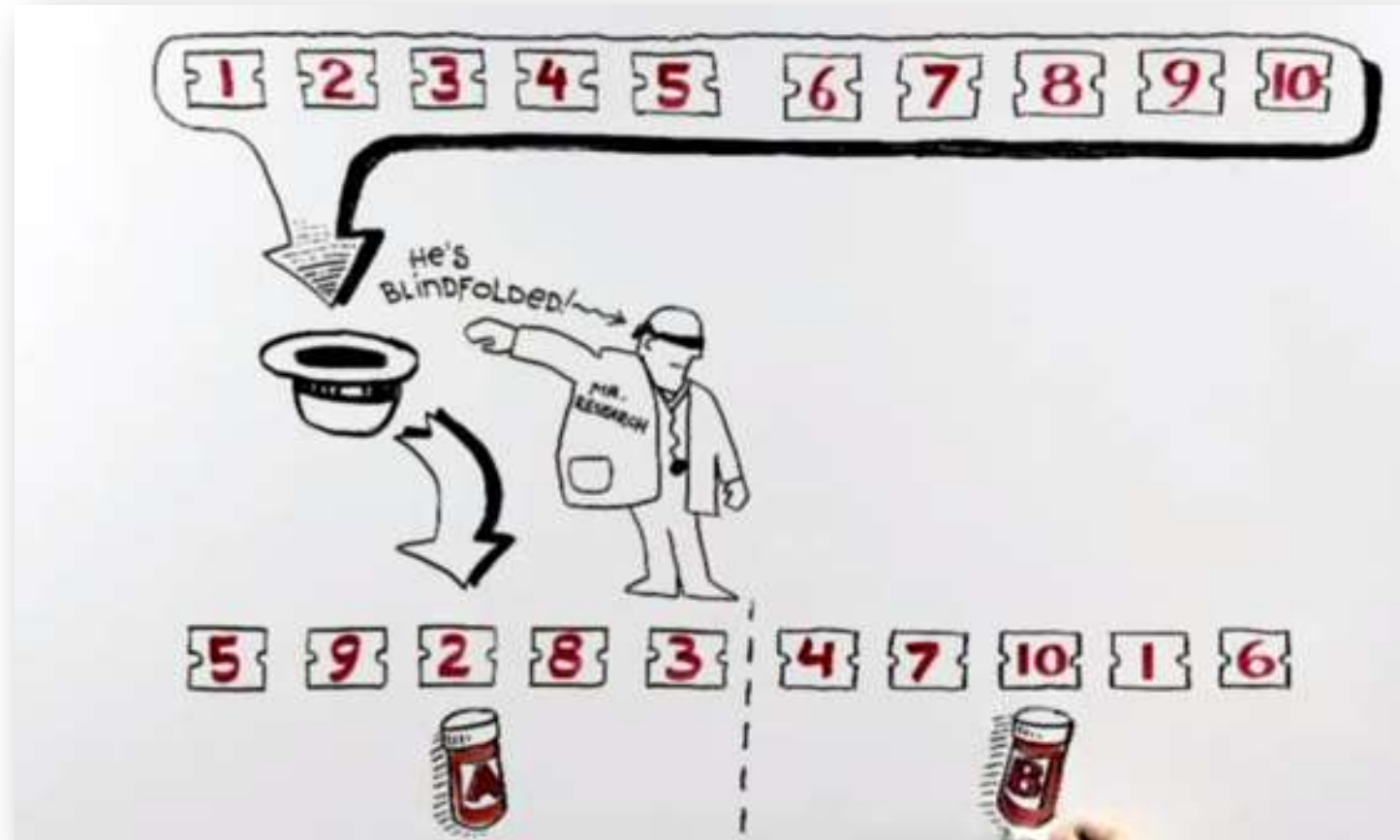**3**   **Comparing unlike groups (e.g., Catherine & Meredith) leads to "selection bias"**
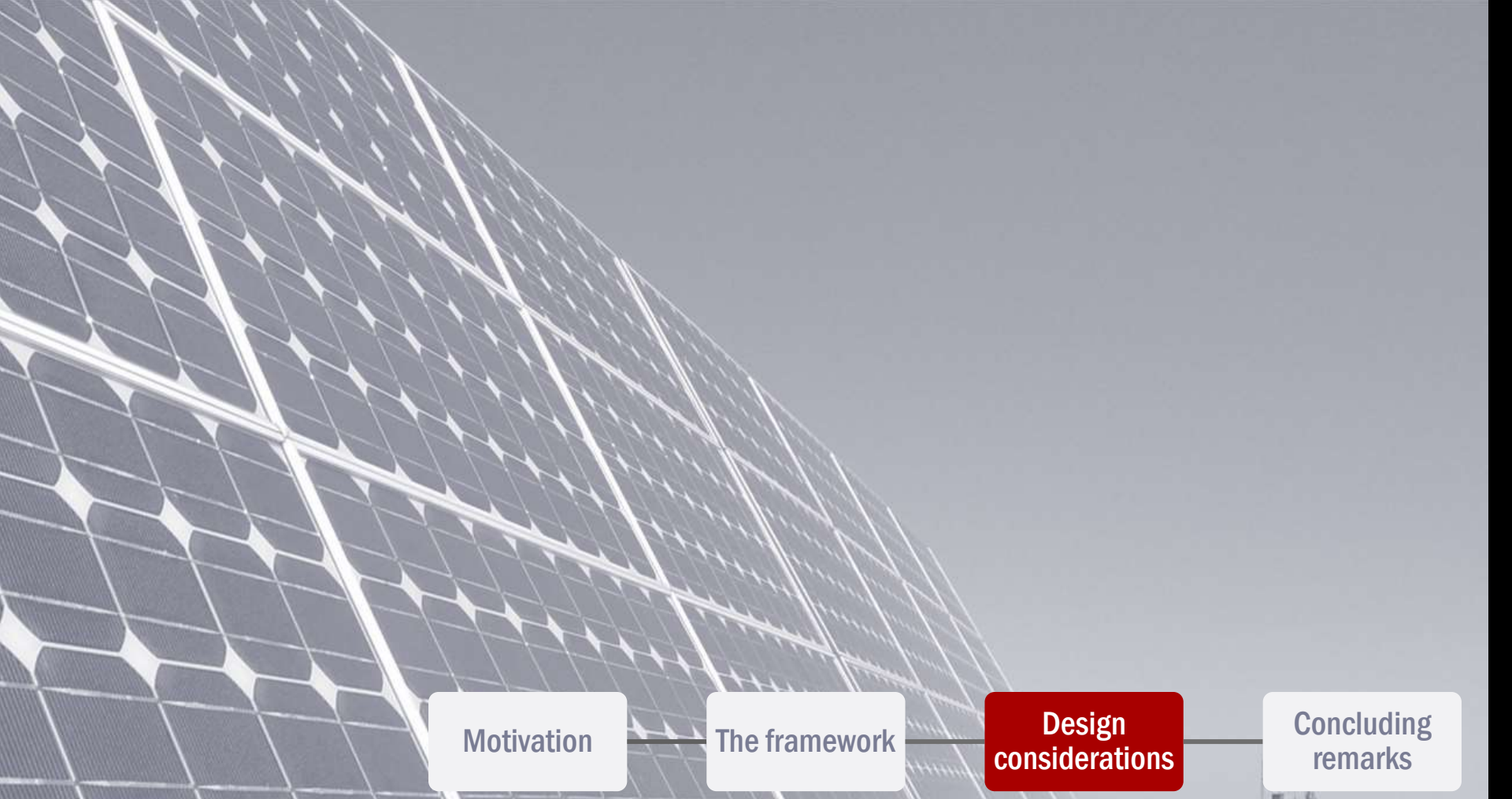
POPULATION

SAMPLE

TREATMENT

CONTROL

random

Compare control vs. treatment
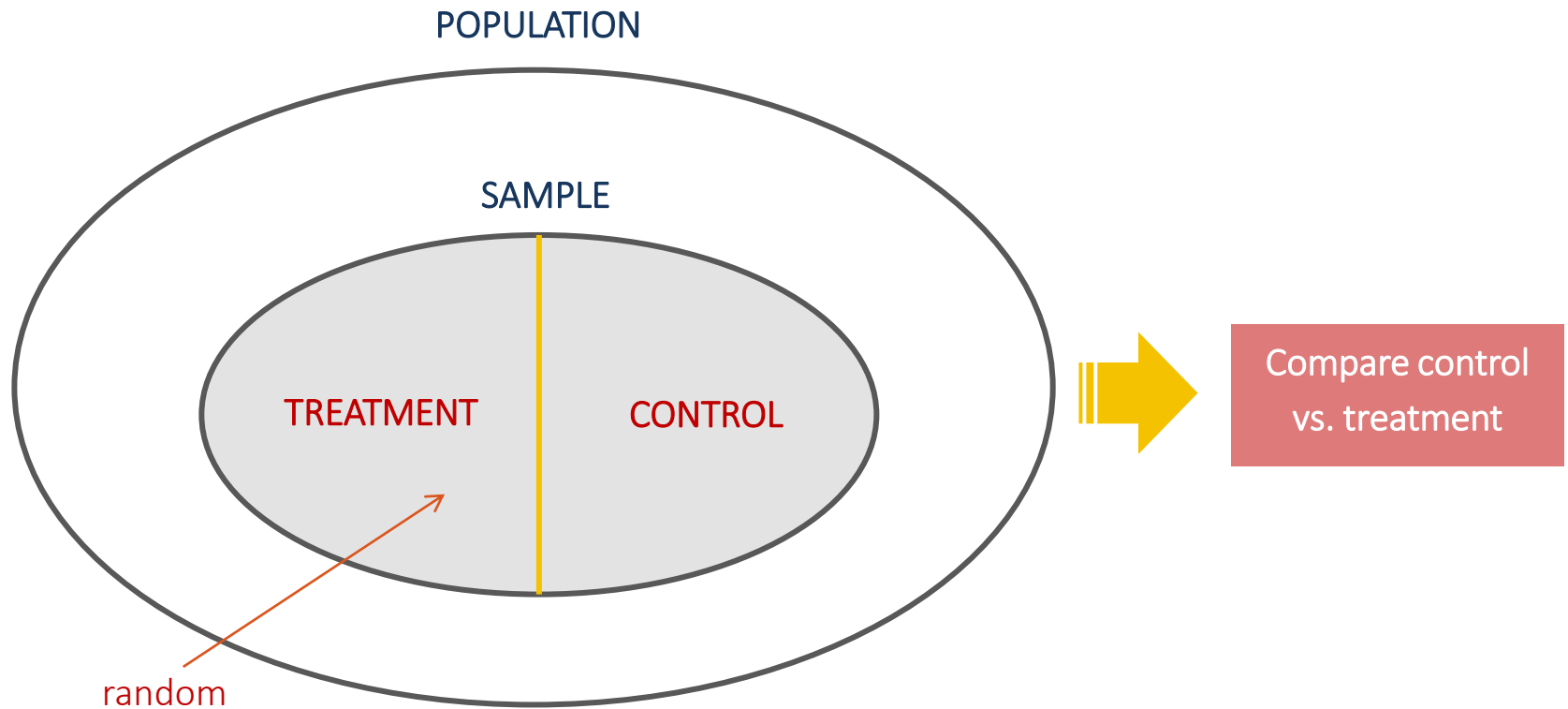
How does randomization work in practice?
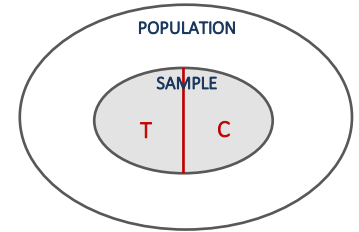

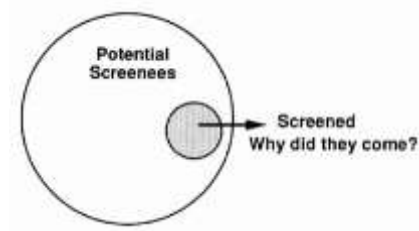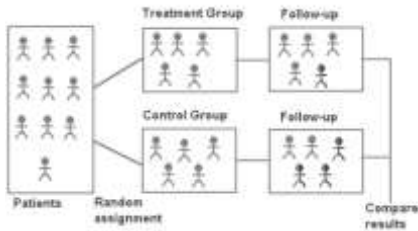
It is exactly like a lottery, or like drawing names from a hat

Design Considerations
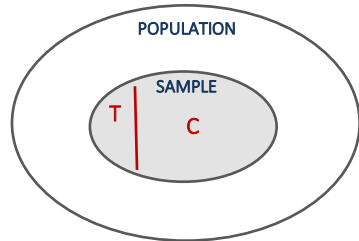
Some important statistical considerations while designing an RCT

Design
Aspects to be considered

Think Upfront

Building randomization into a **budget-limited** pilot can be easy

Avoid Selection Bias

The selection **needs to be random**

Run a Power Calculation

The larger the impact you are expecting, the **smaller the sample you need**

Plan the Groups' Sizes

You don't need to have a 50/50 split

Can estimate upfront the size of the group you need.

Especially if treatment is costly

Decide the Level of Randomization

In many instances, you can't randomize at the individual level (and that's ok)

Ex: Impact of teachers with masters degrees on student performance – classroom

Choose the Treatment Carefully

You can only design treatment around variables that can be manipulated

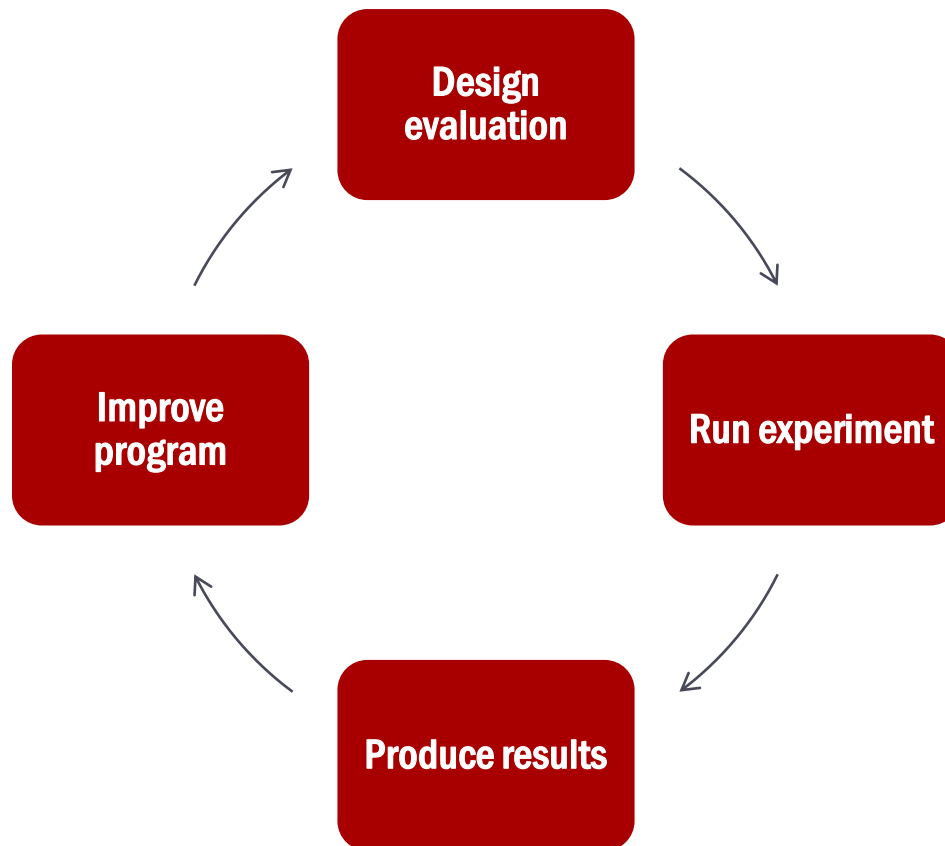Gender, age, height cannot be changed!
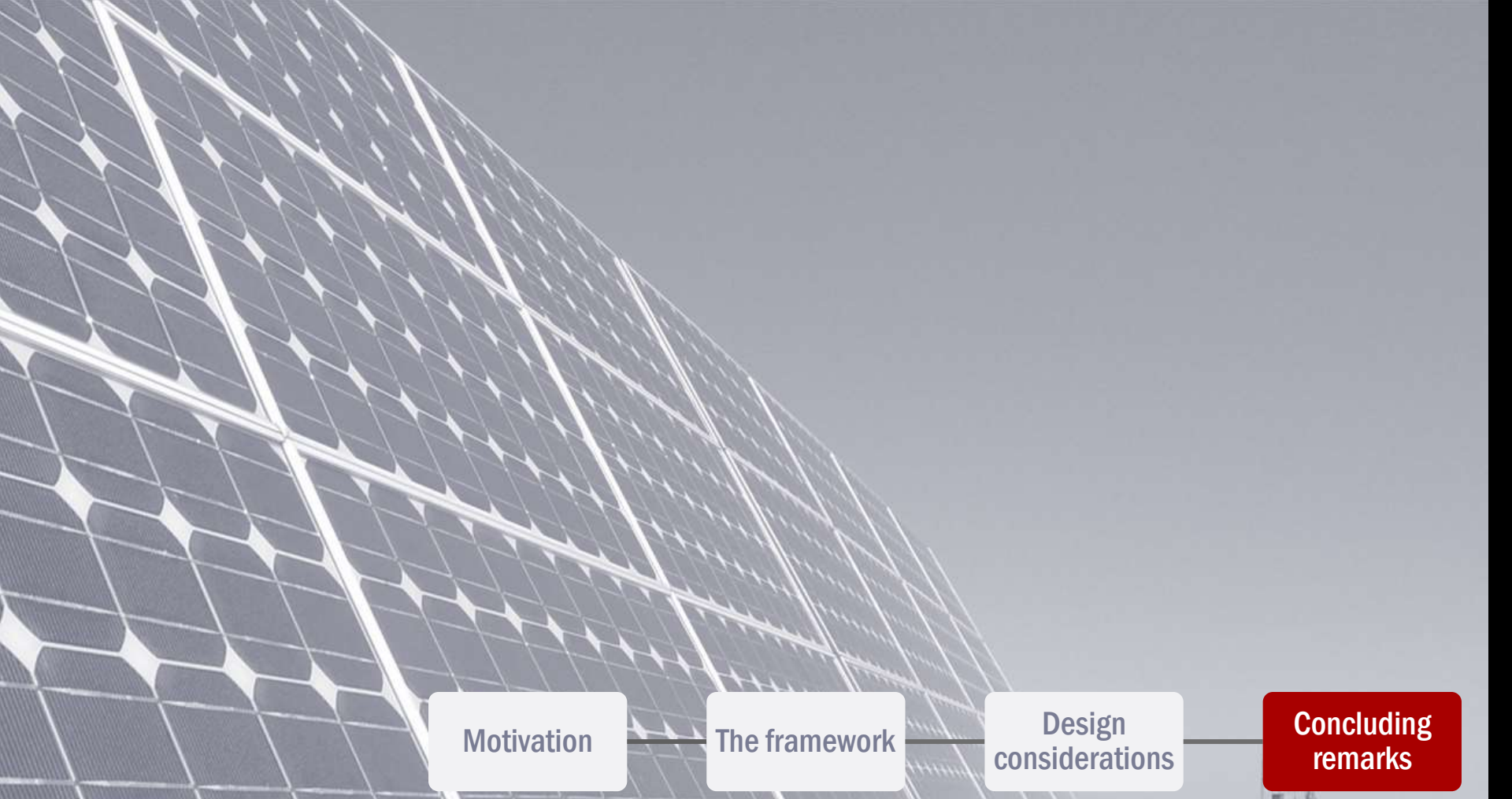
Concluding remarks

What makes a successful RCT?

Let's start at the very beginning

Building an RCT into the program from the beginning can foster a **virtuous cycle** of evaluation

Pilot programs are a great opportunity for RCTs

Why randomize your next pilot? Because...

| It estimates the true impact | You want the pilot to be a representative sample of your population in order to discover the true potential impact. |

| It is a fair way to allocate resources | You want to be fair when selecting the participants for your pilot – everyone has an equal chance of getting in. |

| It tests your underlying model | You want to make sure the logistics of implementation work and refine it. |

| It assesses the scaling up process | You want to get a first causal assessment of treatment impacts and understand costs and benefits of scaling up. |

Conclusions
For successful RCTs

(1) **Be willing to partner, especially with independent researchers**
- Engage researchers to participate in program design
- Provide easy access to data for experiment design and evaluation

(2) **Develop a strong test-learn-refine culture**
- Include RCT evaluation in programs at inception

(3) **Develop strong partnerships. Share expertise and insist on rigor from:**
- Partner organizations
- Program implementers
- Other stakeholders

(4) **RCTs allow you to identify the real impact of your program**
- Including behavioral responses that are otherwise impossible to detect
- Companies such as Google run randomized experiments (A/B tests) regularly to learn about their customers

Thank you!
Please feel free to contact us

Berkeley Office	MIT Office
Energy Institute at Haas **E2e Project** **2547 Channing Way** **Berkeley, California 94720-5180** **Email: e2e@berkeley.edu** Tel: 510.642.9590 Fax: 510.643.5180	**Center for Energy and Env. Policy Research** **E2e Project** **MIT Building E19-411** **400 Main Street, 4th Floor** **Cambridge, MA 02142-1017** **Email: e2e@mit.edu** Tel: 617.253.3551 Fax: 617.253.9845

University of California
Berkeley
Haas School of Business

ENERGY INSTITUTE AT HAAS

MIT CEEPR

MITei
MIT Energy Initiative

Coming Soon: Chicago Office

THE UNIVERSITY OF
CHICAGO