# Why Evaluate
Using Evidence to Inform Policy

Course Overview

1. Why Evaluate
2. Theory of Change & Measurement
3. Why & When to Randomize
4. How to Randomize
5. Sample Size & Power
6. Ethical Considerations for Randomized Evaluations
7. Threats & Analysis
8. Randomized Evaluation from Start to Finish
9. Applying & Using Evidence
10. The Generalizability Framework

Lecture overview

- **Motivation for impact evaluation**

- Prerequisites for a good impact evaluation

- Case study: Failure to appear in court, New York City, USA

- The future of evaluations and evidence-informed policymaking

47%
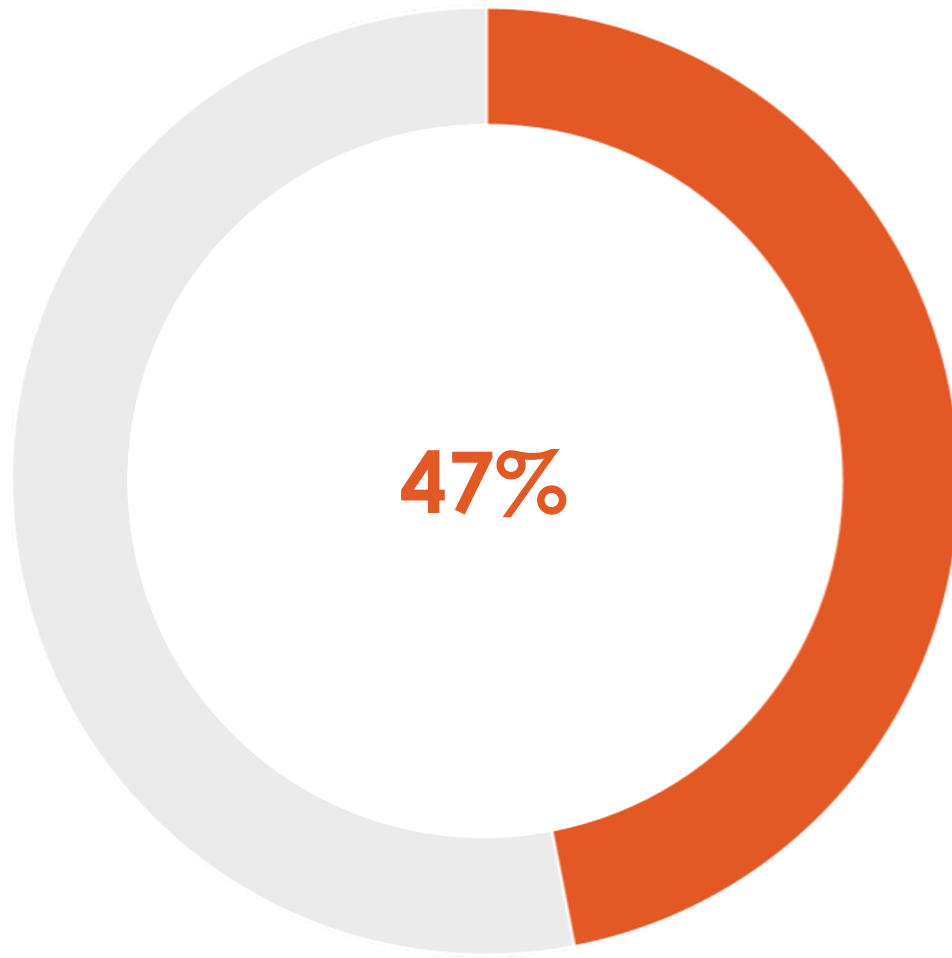
47% of the world's population, over **3 billion people**, live on less than **US $6.85** per day.

The number of people experiencing extreme poverty increased by more than **70 million** in 2020 as a result of the Covid-19 pandemic.

(World Bank 2018, 2022)

The decision-maker's dilemma: Each challenge has many *potentially* good solutions, but time and funding are limited

The decision-maker's dilemma: Each challenge has many *potentially* good solutions, but time and funding are limited

How do you choose between seemingly good program options?

Pair knowledge of local conditions with rigorous evidence from around the world to design your program

The decision-maker's dilemma: Each challenge has many *potentially* good solutions, but time and funding are limited

How do you choose between seemingly good program options?

Pair knowledge of local conditions with rigorous evidence from around the world to design your program

Having chosen one program, how do you know whether it really works or not?
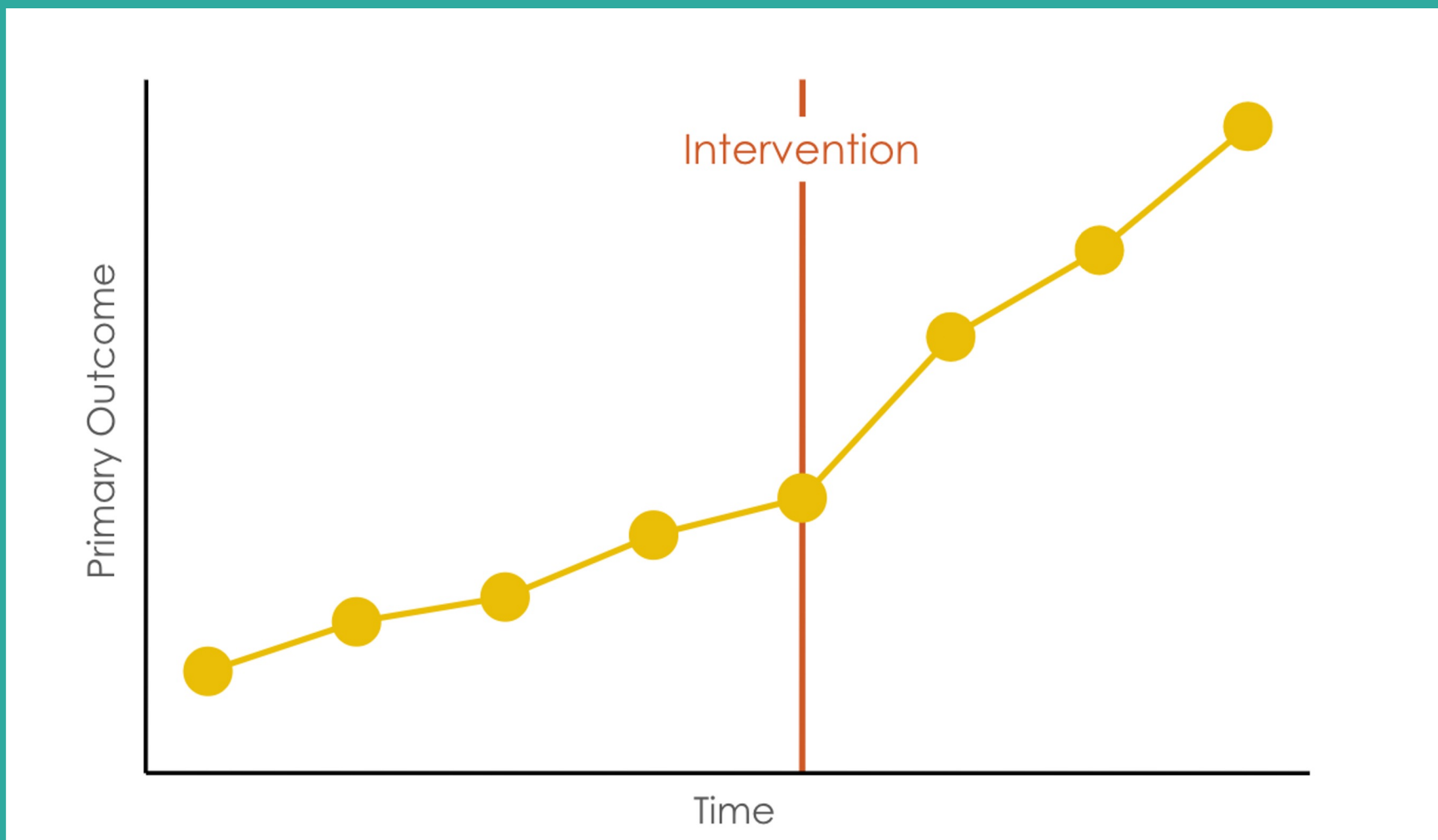
Use data and **impact evaluations** to:

- Test different solutions and see whether they have the intended effect

- Compare different solutions in terms of their cost and magnitude of impact

- Scale most cost-effective solutions

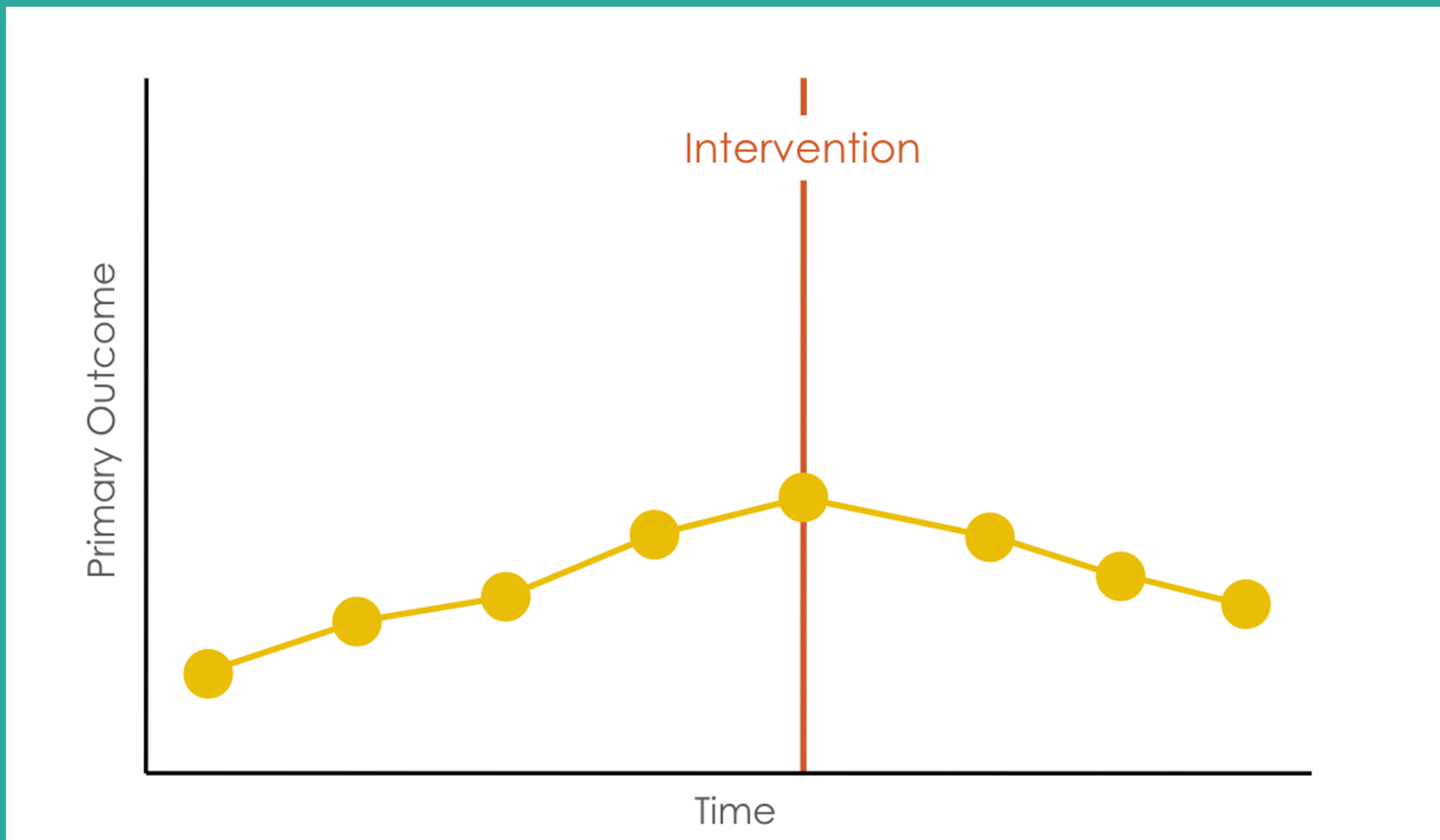# How do you define impact? How do you know if a program is effective?
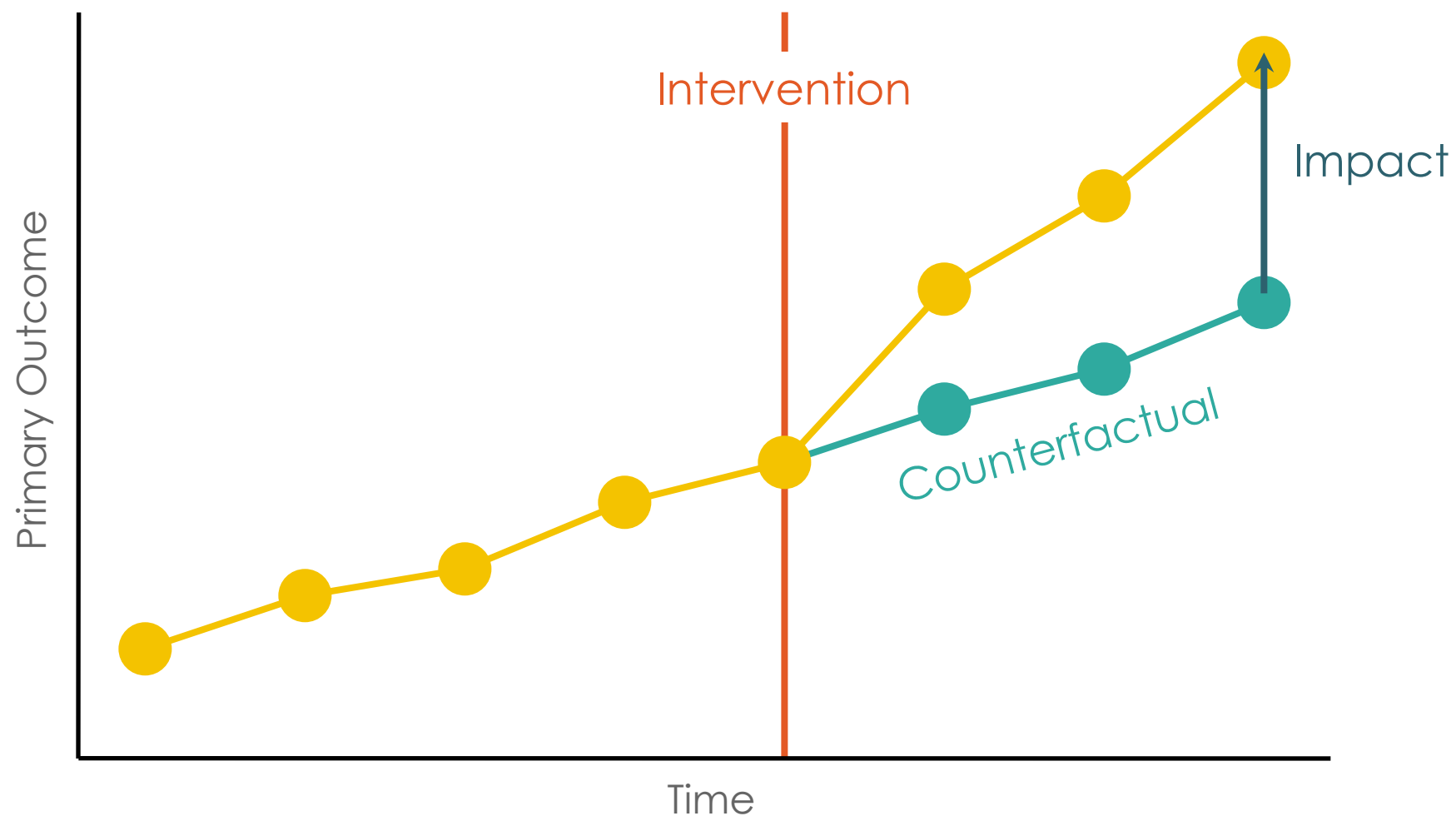
Impact: What is it?
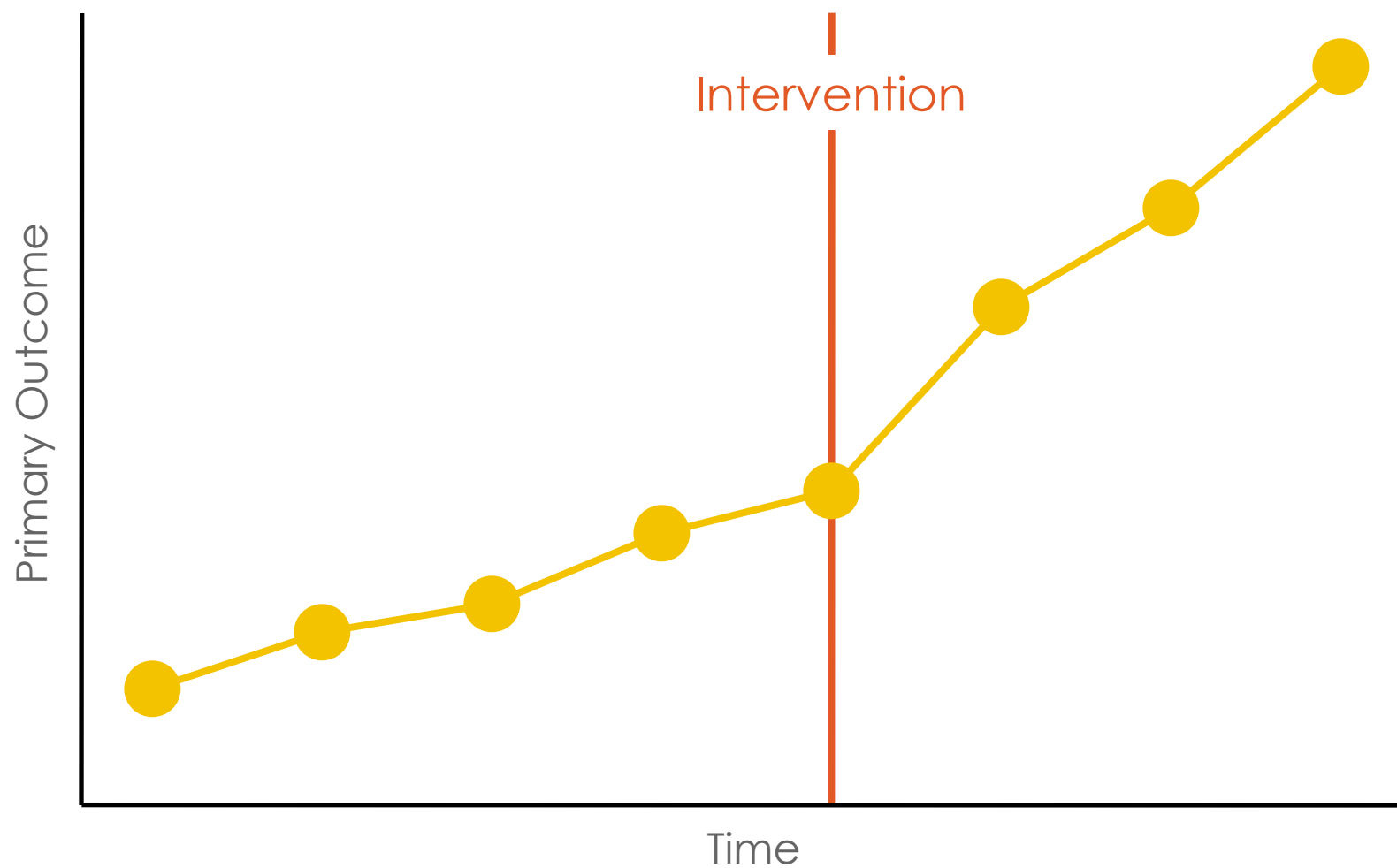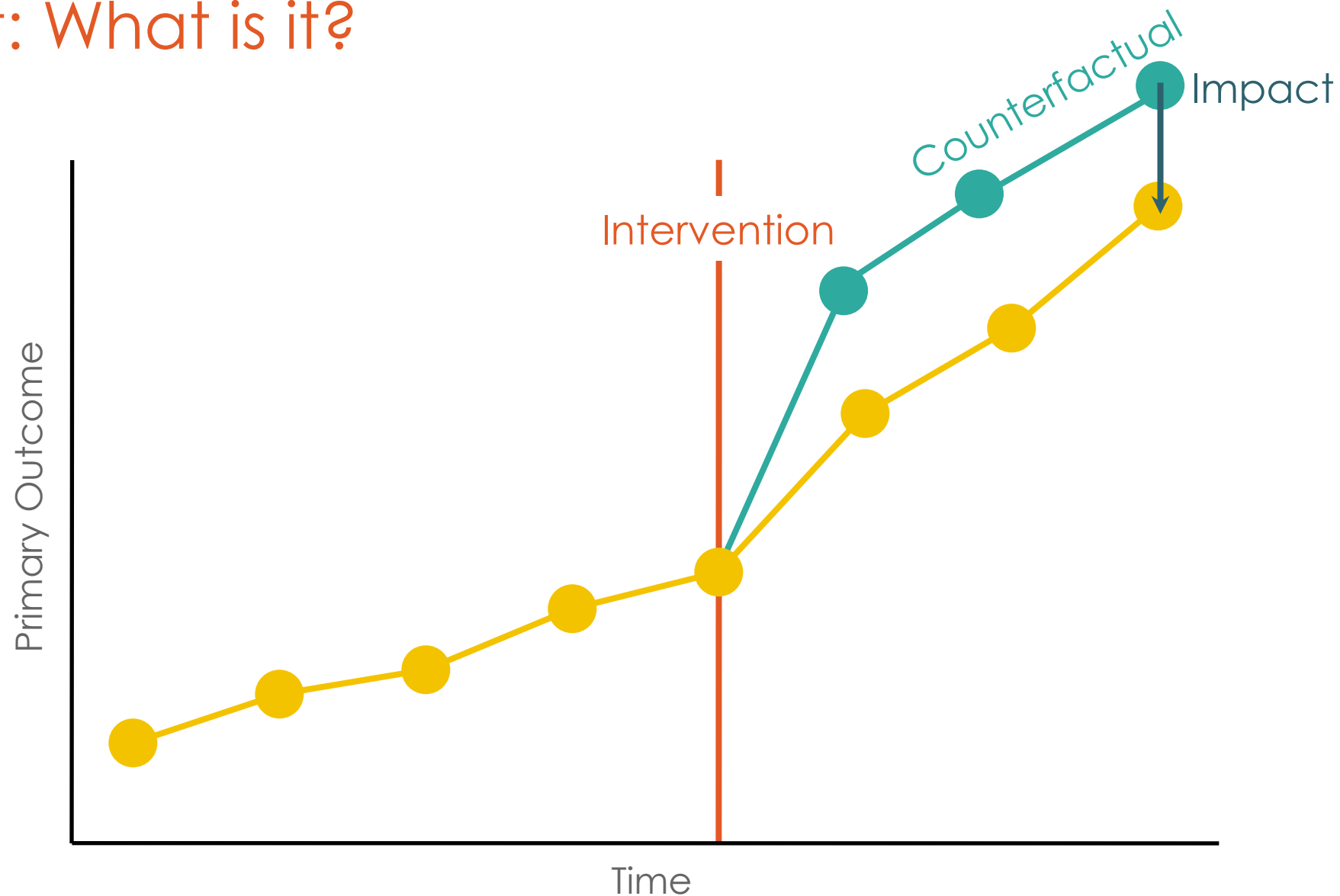
What is the impact of this program?
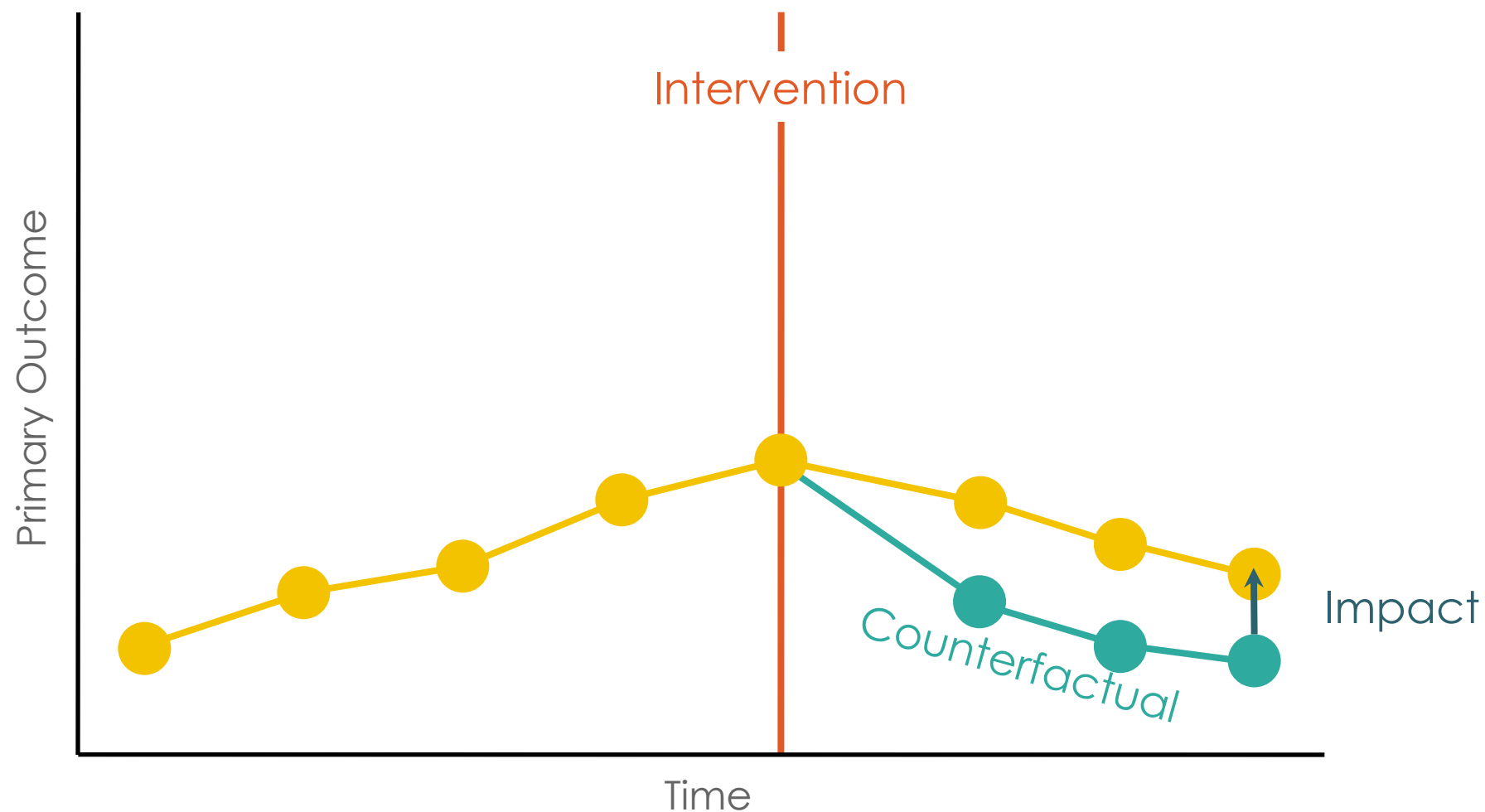
What is the impact of this program?

Impact: What is it?

Impact: What is it?

Impact: What is it?

Impact: How can we measure it?

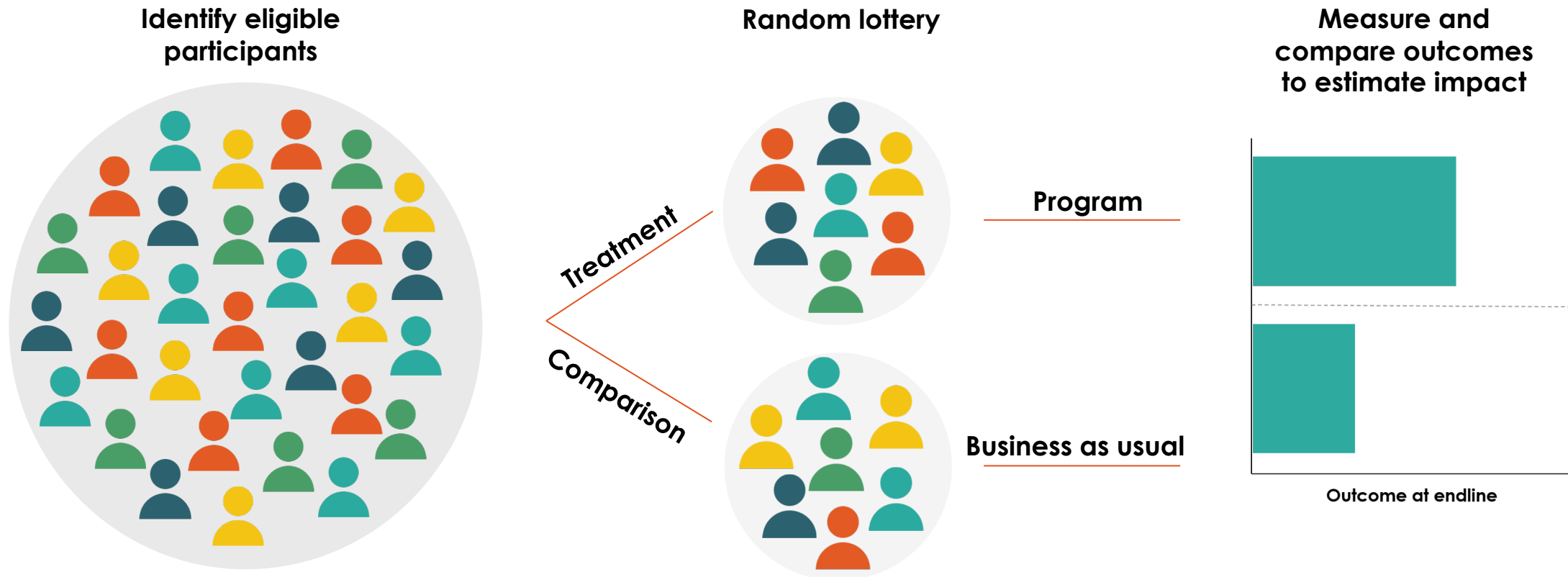**The impact of a program is defined as a comparison between:**

- **What actually happens** after the program has been introduced

- **What would have happened** had the program not been introduced (i.e., the "counterfactual")

In order to assess the impact of a program, we need to understand the counterfactual, i.e., the state of the world that program participants would have experienced in the absence of the program

- **Problem:** The counterfactual never happened so it cannot be observed

- **Solution:** We need to "mimic" or construct the counterfactual

This can be done in different ways, but in this course we will primarily focus on Randomized Controlled Trials (RCTs)

Randomized evaluations use random assignment to mimic the counterfactual and estimate a program's impact

Identify eligible participants

Random lottery

Measure and compare outcomes to estimate impact

Treatment

Comparison

Program

Business as usual

Outcome at endline

Key advantage of randomized evaluations: Due to random assignment, members of the treatment and comparison groups do not differ systematically at the outset of the evaluation. Thus, any difference that subsequently arises between them can be attributed to the program, rather than to other factors.

Lecture overview

- Motivation for impact evaluation

- **Prerequisites for a good impact evaluation**

- Case study: Failure to appear in court, New York City, USA

- The future of evaluations and evidence-informed policymaking

A good impact evaluation builds on good program design and implementation

Conduct a needs assessment

NEEDS ASSESSMENT

What is the problem?

- What is the extent of the problem?

- Who is most in need?

What are its contributing factors?

- Build a hypothesis about what causes the problem and propose possible solutions

A good impact evaluation builds on good program design and implementation

Conduct a needs assessment

↓

Design the program and build a theory of change

💡

THEORY OF CHANGE

Decide on a program to address the identified needs

- What are the inputs or activities?

What steps are needed for the program to achieve the desired change in outcomes?

- What assumptions need to hold?

| Input | Output | Intermediate Outcomes | Final Outcomes |

A good impact evaluation builds on good program design and implementation

Conduct a needs assessment

↓

Design program and
build a theory of change

↓

**Implement program and conduct
a process evaluation**

**IMPLEMENTATION AND
PROCESS EVALUATION**

- Was the program carried out as planned?

- Is the program reaching the target population? Do people use the services?

A good impact evaluation builds on good program design and implementation

Conduct a needs assessment

↓

Design program and build a theory of change

↓

Implement program and conduct a process evaluation

↓
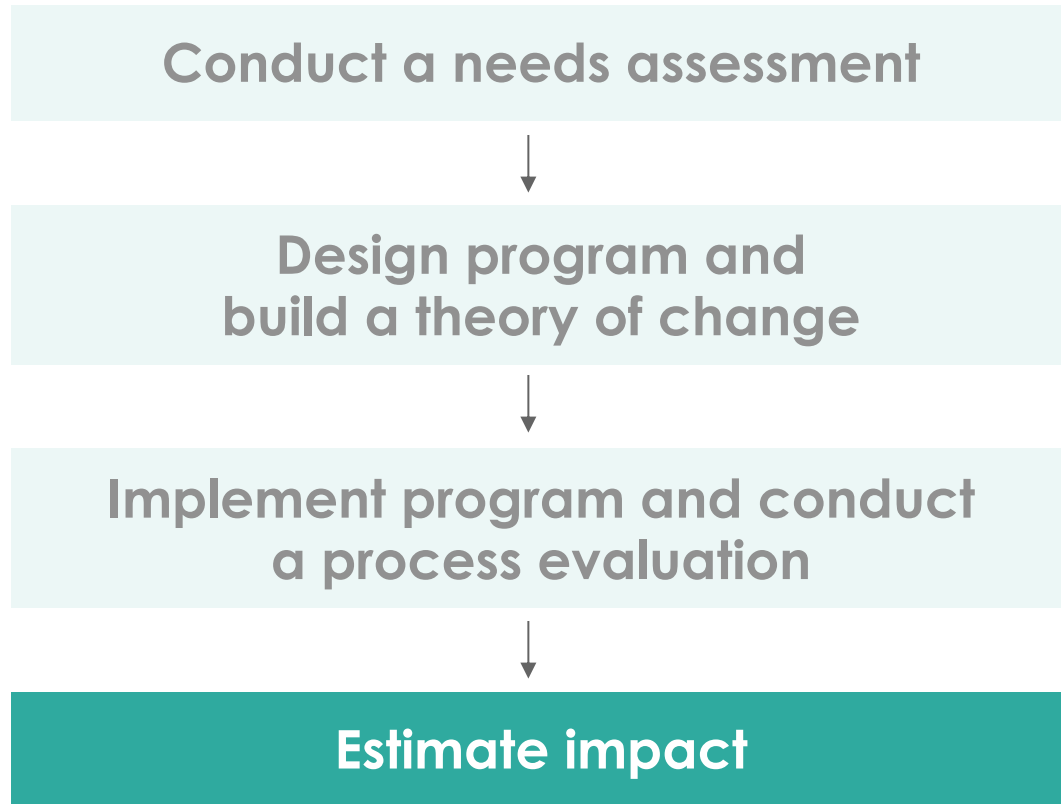
Estimate impact

IMPACT ESTIMATION

- Did the program have the intended effects?

- If so, how big were these effects?

A good impact evaluation builds on good program design and implementation

Conduct a needs assessment

↓

Design program and build a theory of change

↓

Implement program and conduct a process evaluation

↓

Estimate impact

↓

Conduct a cost-effectiveness analysis
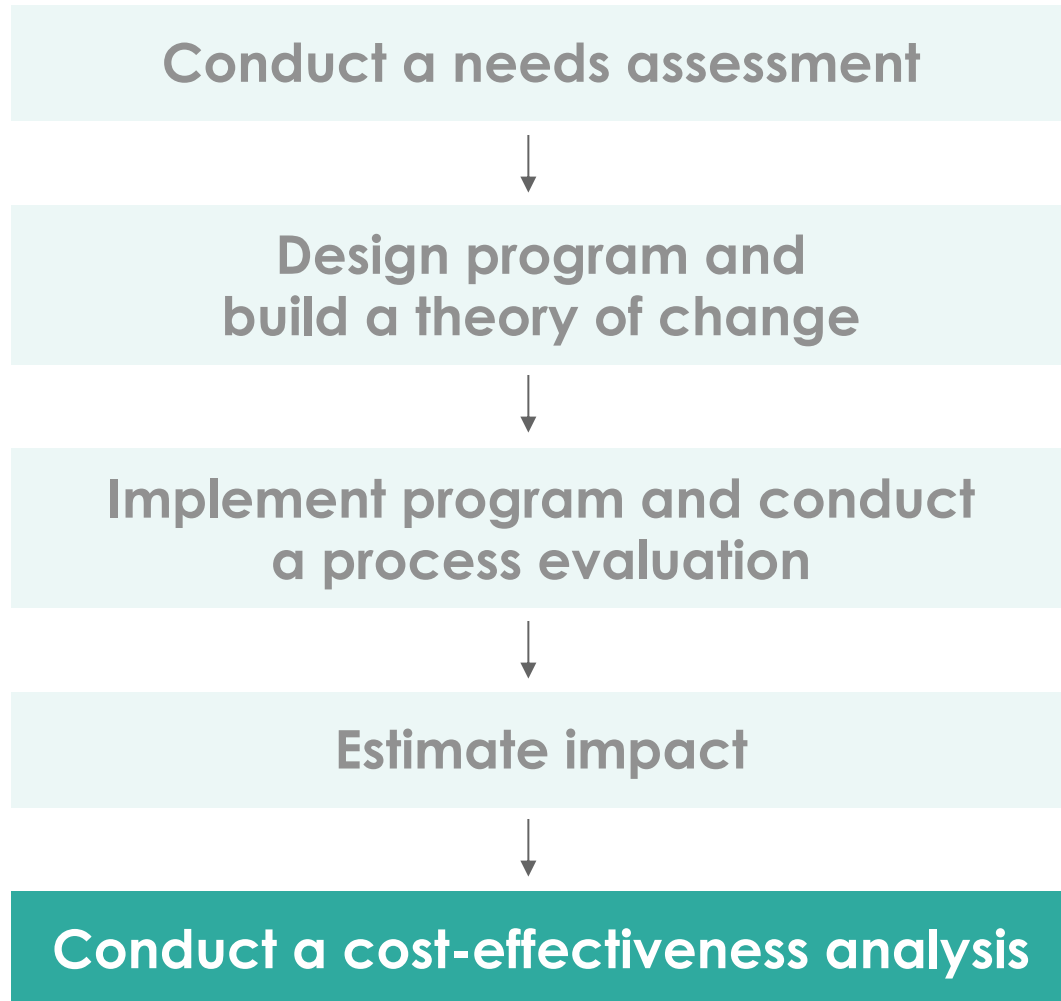
COST-EFFECTIVENESS ANALYSIS

- What is the cost of the program compared to its impact?

- Given magnitude of impact and cost, how does it compare to alternatives?

Lecture overview

- Motivation for impact evaluation

- Prerequisites for a good impact evaluation

- **Case study: Failure to appear in court, New York City, USA**

- The future of evaluations and evidence-informed policymaking

What is the challenge? Failure to appear in court

Each year, millions of people in the United States are required to appear in court for low-level offenses.

However, many people miss their court dates.

Consequences include:
Additional fines and fees, a suspended driver's license, and, possibly, an arrest warrant.

What is the challenge? Failure to appear in court in New York City, NY, USA

Many New Yorkers do not appear for their court dates (known as "failure to appear," or FTA)

- About 40% of people missed their court date for low-level offenses in 2015 in NYC

This leads to arrest warrants, which are costly and burdensome

What is causing high FTA rates, and how can we improve this?

Needs assessment → Design & ToC → Implementation & process evaluation → Estimate impact → CEA

Have you ever missed an appointment? If so, why?

Failure to appear:
What are possible reasons?

Intentional "contempt of court"

Costs of attending court outweigh the risks

Misperceptions about court attendance rates

Insufficient awareness about court date, location, and/or consequences (e.g., arrest)

| Needs assessment | → | Design & ToC | → | Implementation & process evaluation | → | Estimate impact | → | CEA |

What are potential policy solutions?

Failure to appear:
What are possible solutions?

Intentional "contempt of court"	Increase punishment
Costs of attending court outweigh the risks	Reduce costs
Misperceptions about court attendance rates	Correct misperceptions
Insufficient awareness about court date, location, and/or consequences	Make basic information and consequences more salient, send reminders

Needs assessment → Design & ToC → Implementation & process evaluation → Estimate impact → CEA

Rigorous evaluation enables us to determine which solutions will work

How does the program work?

Need

High FTA rates

Inputs

1. Redesigned form
2. Text messages

Outputs

Court information is clearer, individuals receive reminders

Intermediate Outcomes

Increased awareness of court information

Increased awareness of consequences

Final Outcomes

Increased likelihood that Individuals appear in court

Needs assessment → **Design & ToC** → Implementation & process evaluation → Estimate impact → CEA

How was the program implemented?

- Focused on criminal summonses, typically issued for the lowest level of criminal offenses, in New York City

- Primary sample includes nearly 400,000 summonses issued in New York City between January 2016 and June 2017

- Everyone eventually received redesigned summons form between March and August 2016

- People who received the new summons form could provide their cell phone number; ~11% of those issued a summons provided their phone numbers

- Used administrative data collected by the New York State Office of Court Administration, including: gender, date of birth, and address; information about the violation; and court outcomes

Needs assessment → Design & ToC → **Implementation & process evaluation** → Estimate impact → CEA
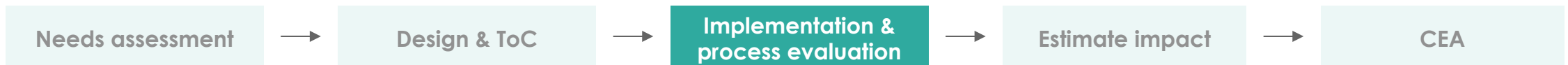
Step 1: Everyone received a redesigned form

① Clear title describes the purpose and required action.

② The date, time, and location of the appearance is moved to the top, where it is more likely to be read.

③ The consequence of missing is clearly articulated and framed to spur loss aversion, the human tendency to feel losses more severely than equivalent gains.

Cooke, et al., 2018

Step 2: Text message reminders (randomized)

- Sample: 23,000 people who provided their phone numbers

- Individuals were **randomly assigned** to receive text messages before court or to the comparison group, which did not get messages

- Message content was varied to identify what info is most effective at reducing FTA: *consequences* versus *plan-making* versus *both*

Cooke, et al., 2018

| Needs assessment | → | Design & ToC | → | **Implementation & process evaluation** | → | Estimate impact | → | CEA |
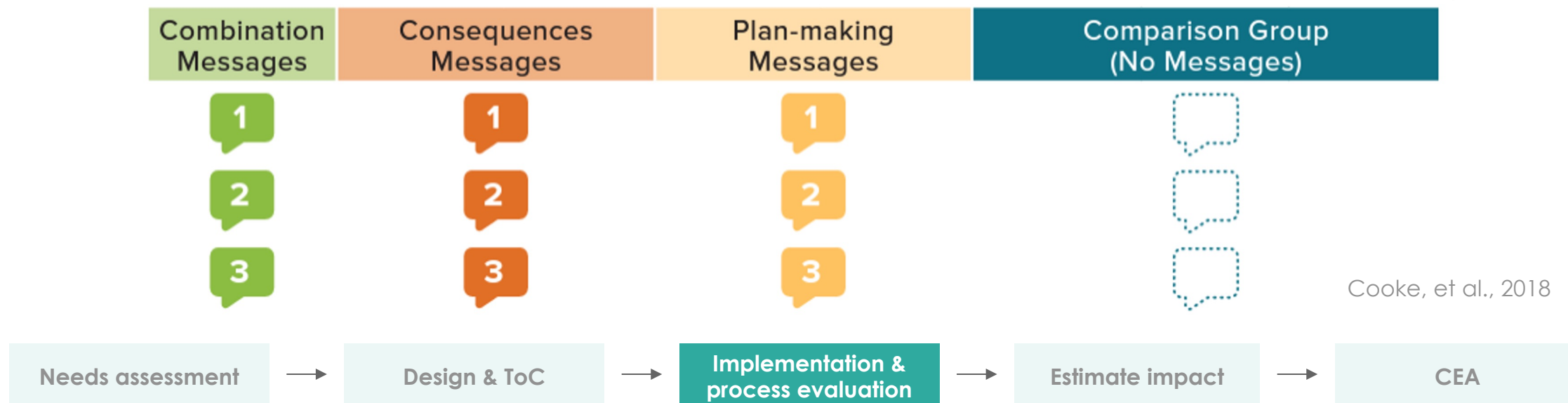
Process evaluation considerations

- Were text messages successfully delivered to the numbers provided?

- Were text messages delivered as planned/scheduled (7 days, 3 days, 1 day prior to the court date)?

Process evaluation helps determine if the intervention was delivered as intended

| Needs assessment | → | Design & ToC | → | Implementation & process evaluation | → | Estimate impact | → | CEA |
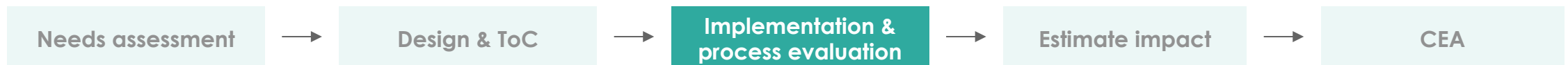
Estimate impact by comparing outcomes between treatment and comparison group

Identify eligible participants

Random lottery

Measure and compare outcomes to estimate impact

Treatment

Comparison

Program

Business as usual

Outcome at endline

| Needs assessment | → | Design & ToC | → | Implementation & process evaluation | → | Estimate impact | → | CEA |
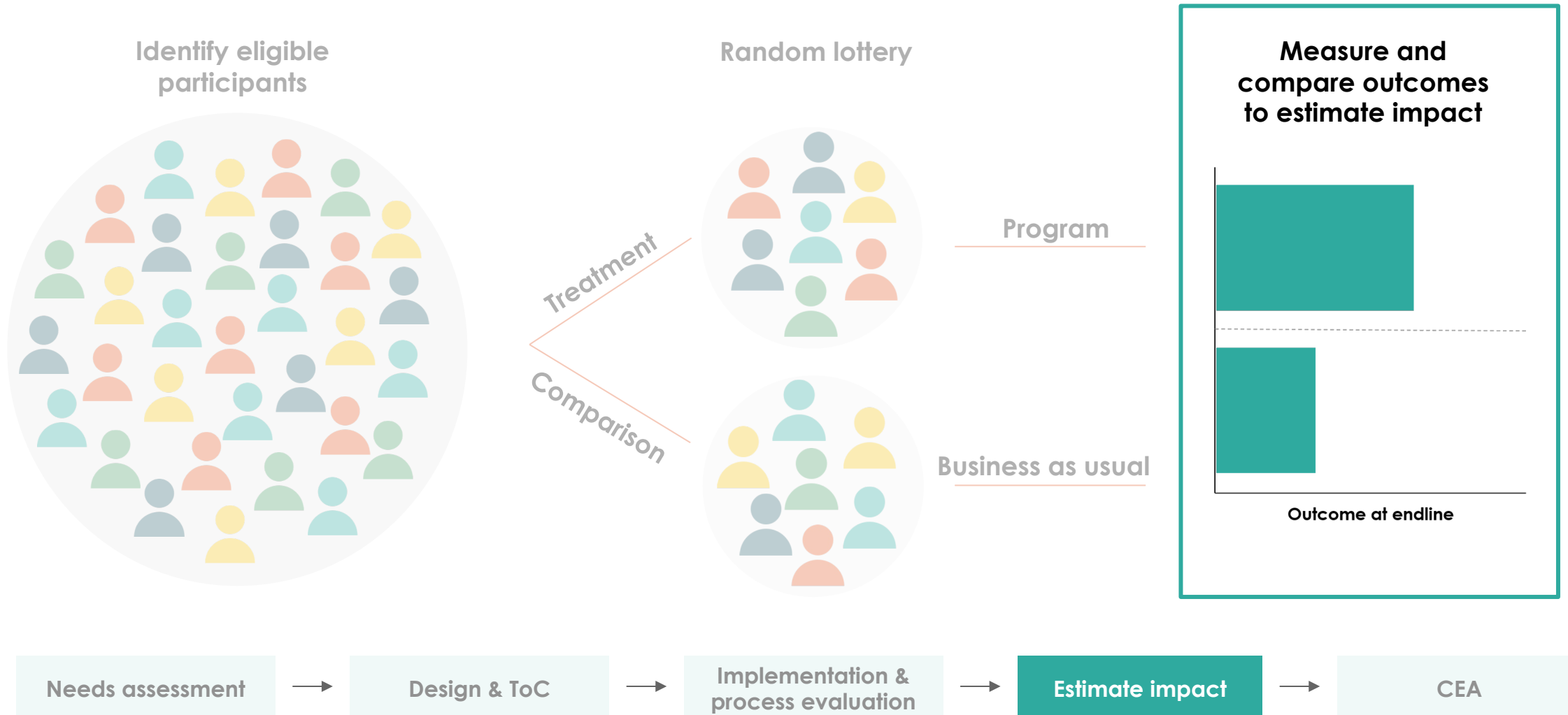
Finding #1: Receiving text messages before court decreased FTA overall

Fishbane, et al., 2020

| Needs assessment | → | Design & ToC | → | Implementation & process evaluation | → | **Estimate impact** | → | CEA |
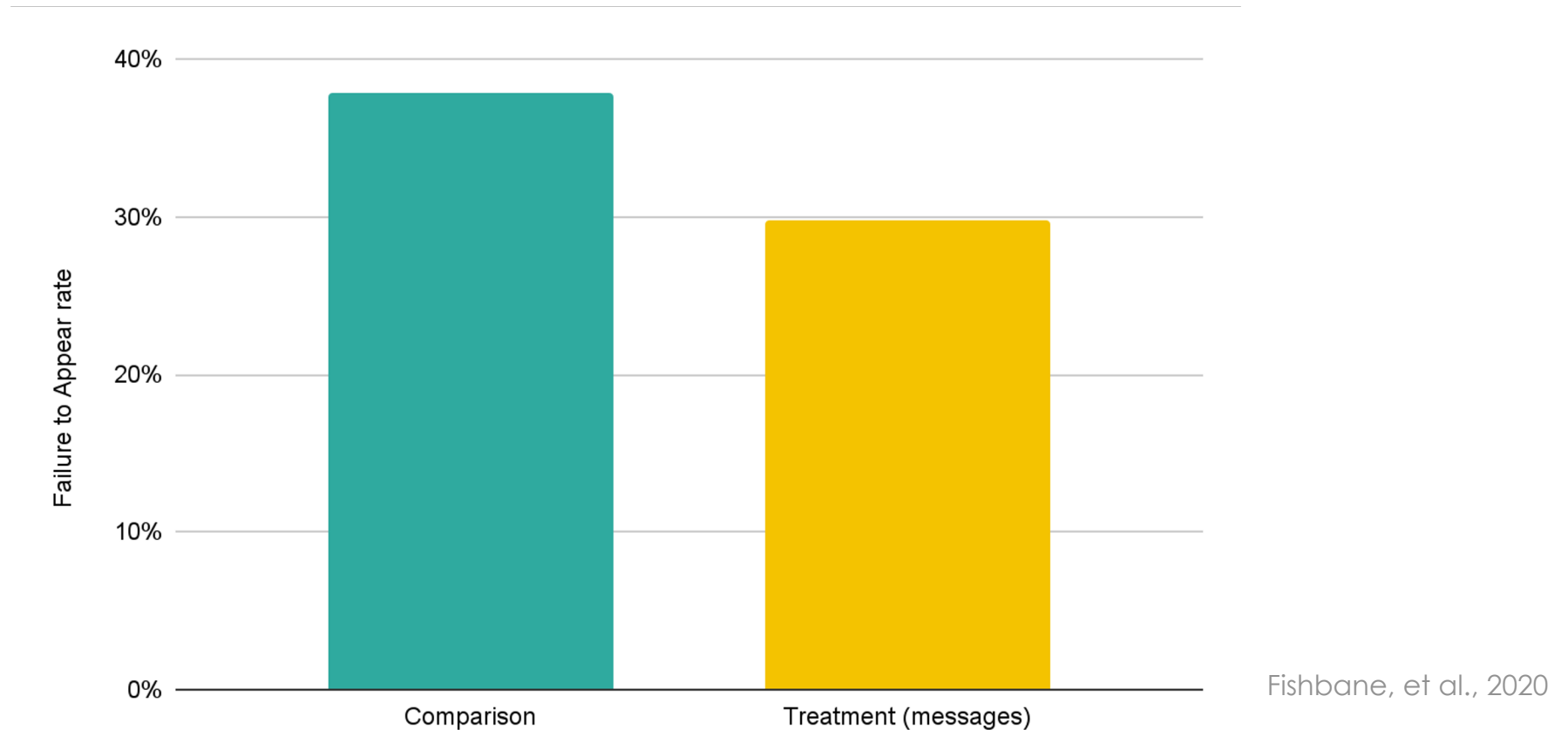
Finding #2: The combination and consequences message sets were the most effective at reducing FTA

FTA Rate by Type of Message

40% — 37.8% (Comparison Group (No Messages))

26% DECREASE — 28.1% (Combination)

24% DECREASE — 28.9% (Consequences)

16% DECREASE — 31.8% (Plan-Making)

Cooke, et al., 2018

| Needs assessment | → | Design & ToC | → | Implementation & process evaluation | → | Estimate impact | → | CEA |
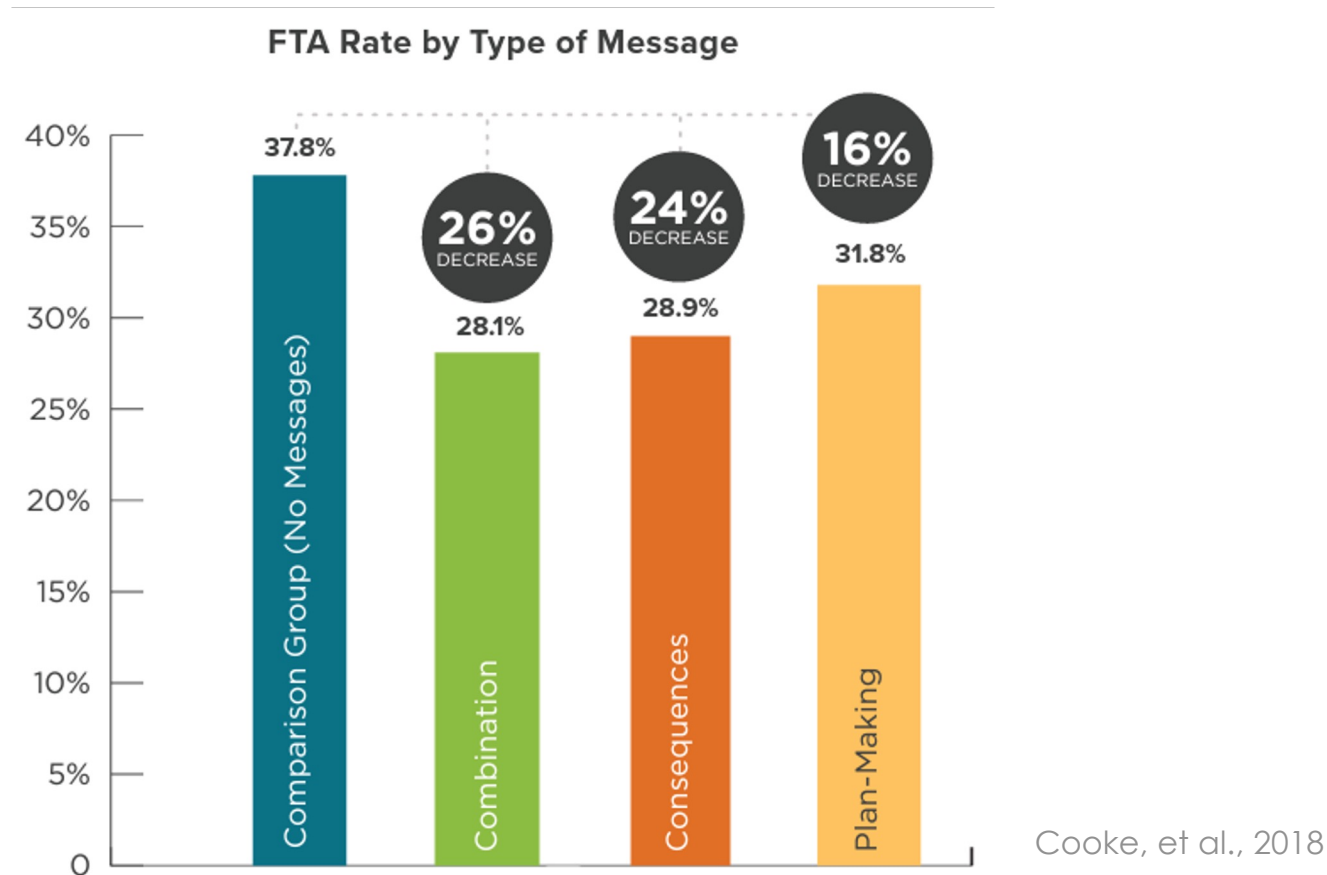
What can we learn from this evaluation?

This intervention is a low-cost way to reduce FTA

The redesigned summons form and text messages avoided over ~30,000 arrest warrants over three years, and ~20,000 people had their cases fully dismissed instead of having an open warrant

Beyond addressing the negative impact of warrants on people's lives, the intervention reduced costs for the criminal legal system, saving hundreds of thousands of dollars, and was inexpensive to implement at $4,500 USD per year (<1 cent per msg)

As a result of the evaluation, the redesigned summons forms are now used for all court summonses in NYC, and text message reminders are sent to all summons recipients who provide a cell phone number

Needs assessment → Design & ToC → Implementation & process evaluation → Estimate impact → **CEA**

This was just one of the pathways to policy change

Shifting global thinking
Free bednets to fight malaria

Institutionalizing evidence use
US Office of Evaluation Sciences

Applying research insights
Agricultural information delivery

Adapting and scaling a program
Teaching at the Right Level

Scaling up evaluated pilots
Improve access to social assistance

Scaling back an evaluated program
Limits of technological solutions to provider monitoring

Lecture overview

- Motivation for impact evaluation

- Prerequisites for a good impact evaluation

- Case study: Failure to appear in court, New York City, USA

- **The future of evaluations and evidence-informed policymaking**

The use of RCTs is expanding to tackle key policy-research challenges in new areas

Climate change

Advance evidence-based policies in climate change mitigation and adaptation.

Social protection

Improve the effectiveness of programs to reduce poverty, inequality, vulnerability, and risk.

Discrimination and racial equity

Identify effective approaches to counter discriminatory practices and reduce prejudice.

Misinformation

Draw out lessons for combating misinformation and eliciting behavior change.
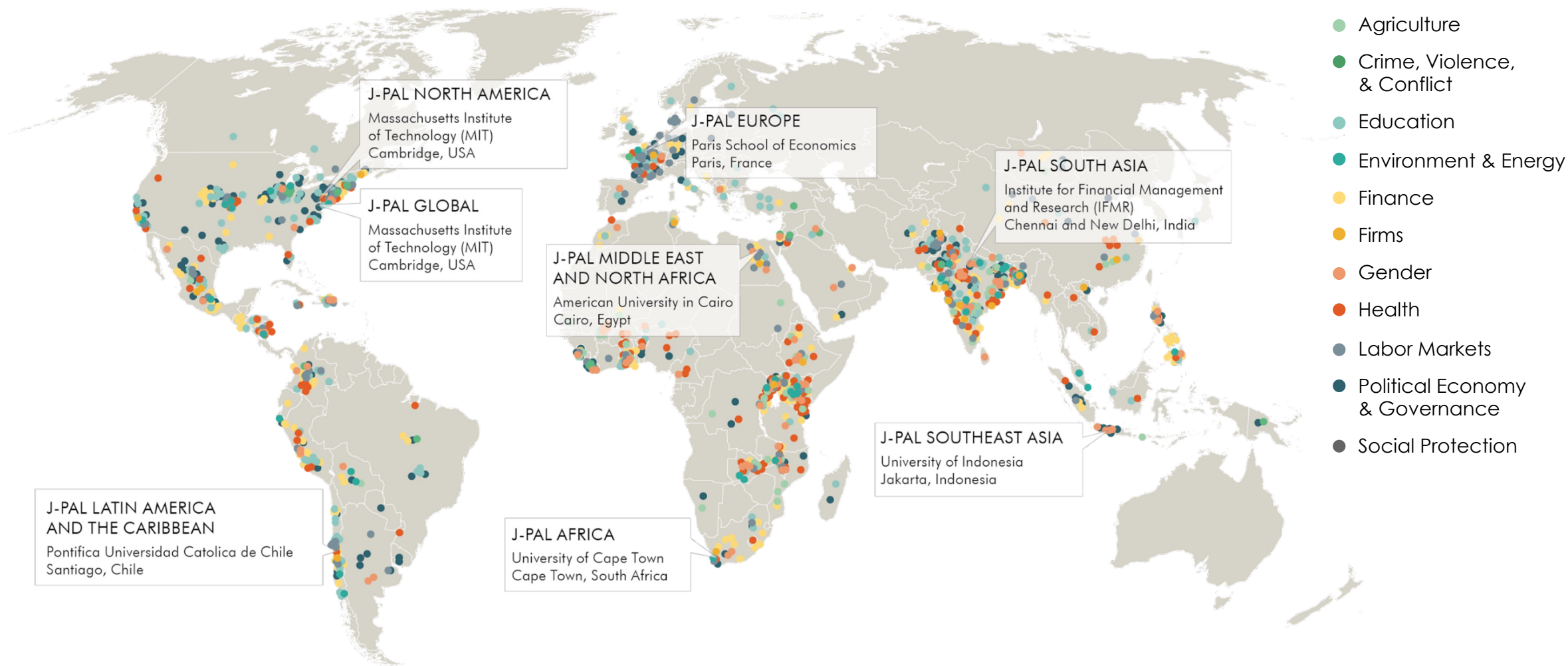
The future of work

Test strategies for job creation and helping workers adapt to changing economies.

Big data

Increase the use of big data for evidence-informed decision-making.

At J-PAL, our researchers have 1,100+ projects in 95 countries in all social sectors, each one with a local partner

J-PAL NORTH AMERICA
Massachusetts Institute of Technology (MIT)
Cambridge, USA

J-PAL GLOBAL
Massachusetts Institute of Technology (MIT)
Cambridge, USA

J-PAL EUROPE
Paris School of Economics
Paris, France

J-PAL SOUTH ASIA
Institute for Financial Management and Research (IFMR)
Chennai and New Delhi, India

J-PAL MIDDLE EAST AND NORTH AFRICA
American University in Cairo
Cairo, Egypt

J-PAL SOUTHEAST ASIA
University of Indonesia
Jakarta, Indonesia

J-PAL LATIN AMERICA AND THE CARIBBEAN
Pontifica Universidad Catolica de Chile
Santiago, Chile

J-PAL AFRICA
University of Cape Town
Cape Town, South Africa

- Agriculture
- Crime, Violence, & Conflict
- Education
- Environment & Energy
- Finance
- Firms
- Gender
- Health
- Labor Markets
- Political Economy & Governance
- Social Protection

Conclusion: Evidence is key to good policy making

A fundamental dilemma for decision-makers is to select the best possible program to address a given challenge facing society

The only way to know whether a program works is to evaluate its impact

A good impact evaluation builds on the careful design and implementation of a program

Impact evaluations can be costly, but not evaluating a program can be even more costly

References and resources

Beyond Bail: Using Behavioral Science to Improve Timely Court Appearance (references)

- J-PAL Evaluation Summary
- Research Paper
- Research Article (*Science*)

Evidence to Policy

- Pathways to Policy Change

Growth is not enough

- J-PAL blog
- Project Syndicate

Reuse and citation

To reference this lecture, please cite as:

J-PAL. "Lecture: Why Evaluate." Abdul Latif Jameel Poverty Action Lab. 2023. Cambridge, MA