



# Randomized Controlled Trials, Development Economics and Policy Making in Developing Countries

Esther Duflo

Department of Economics, MIT

Co-Director J-PAL

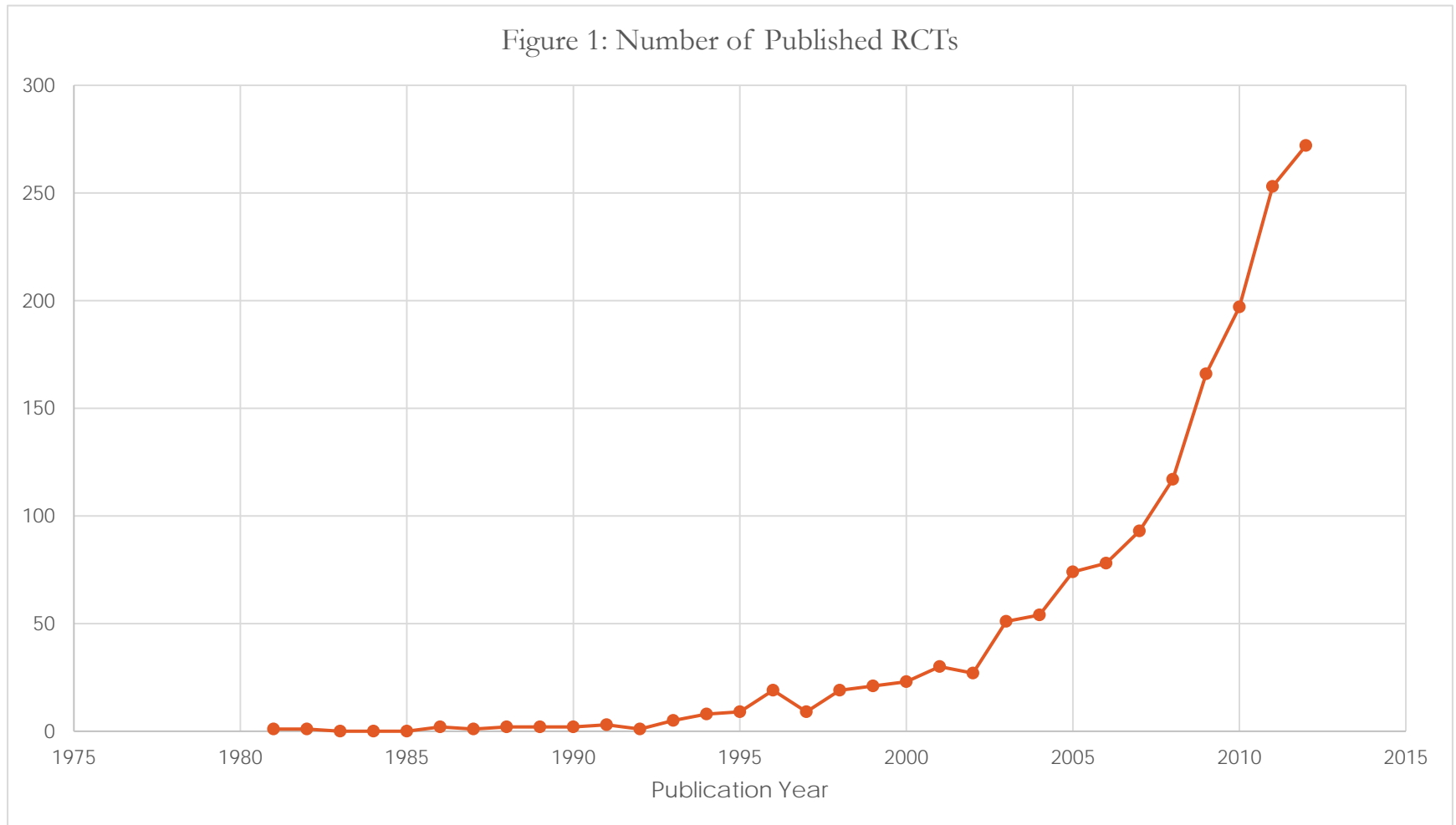
[Joint work with Abhijit Banerjee and Michael  
Kremer]



# Randomized controlled trials have greatly expanded in the last two decades

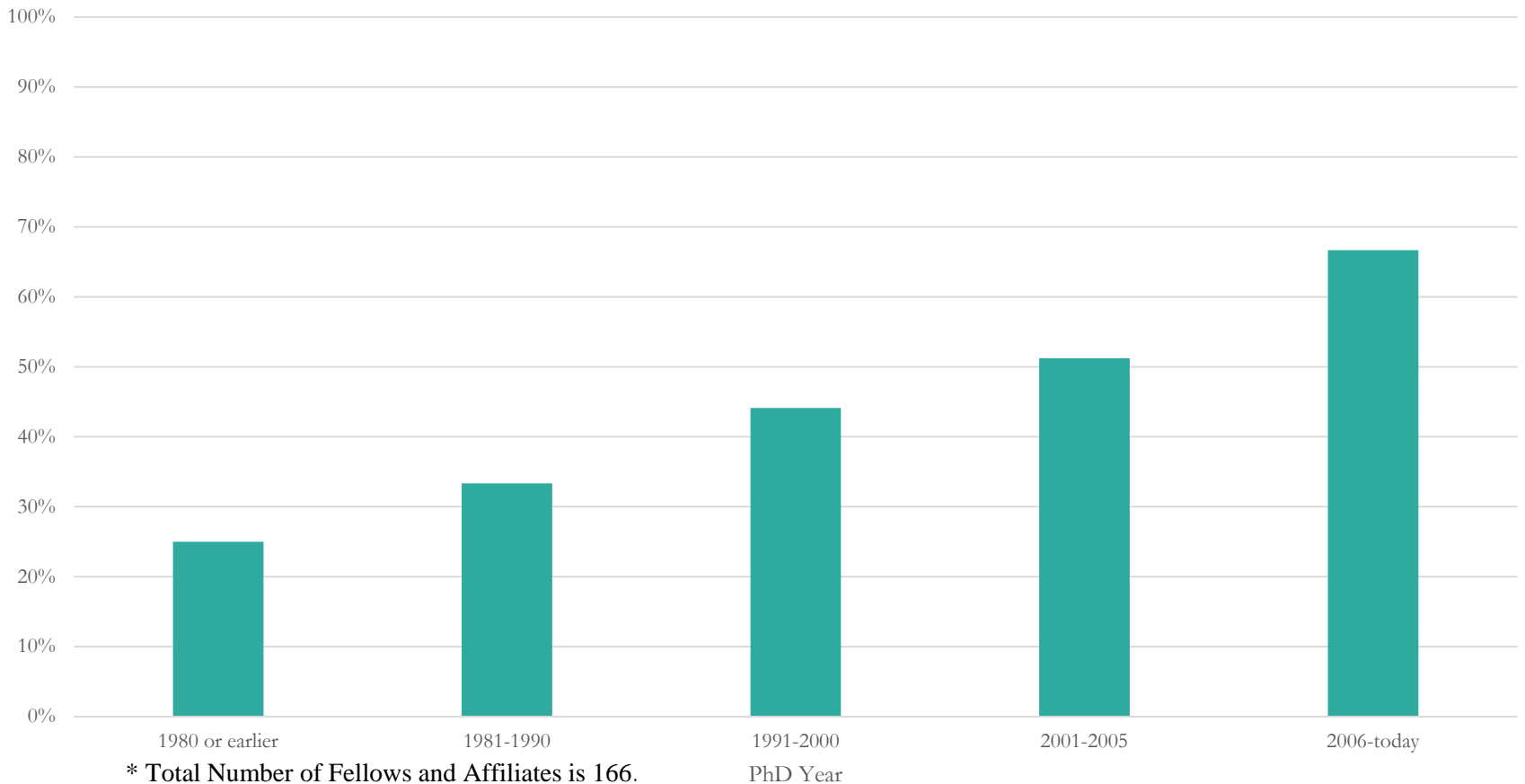
- Randomized controlled Trials were progressively accepted as a tool for policy evaluation in the US through many battles from the 1970s to the 1990s.
- In development, the rapid growth starts after the mid 1990s
  - Kremer et al, studies on Kenya (1994)
  - PROGRESA experiment (1997)
- Since 2000, the growth have been very rapid.

# Cameron et al (2016): RCT in development



# BREAD Affiliates doing RCT

Figure 4. Fraction of BREAD Affiliates & Fellows with 1 or more RCTs

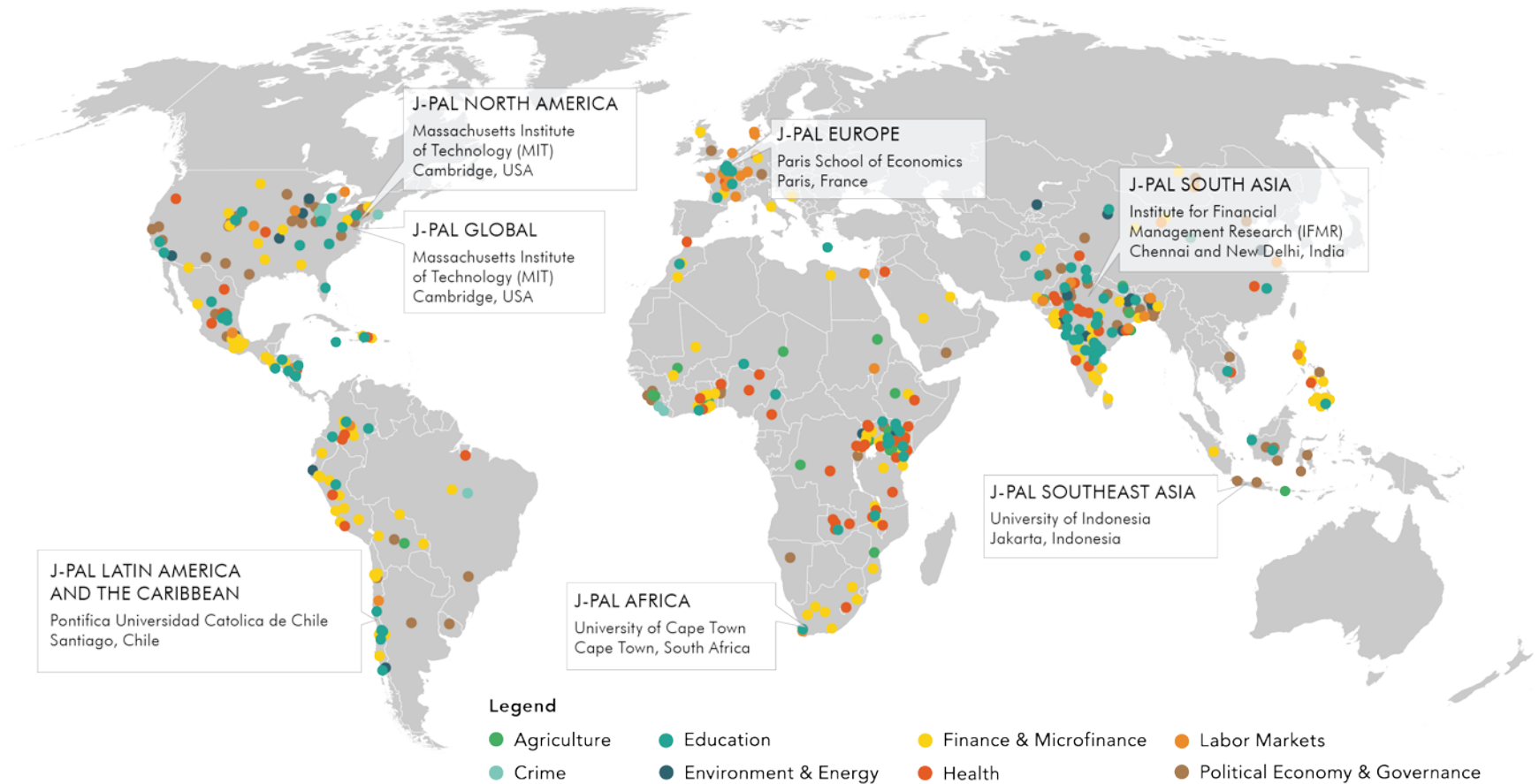


# Top Journals

**Table 2: Papers in Top 5 Journals**

<b>Year</b>	<b>Total # of Papers</b>	<b># of Development Papers</b>	<b># of which are RCTs</b>
2015	271	32	10
2000	215	21	0
1990	278	17	0

# Many sectors, many countries



# Why have RCT had so much impact?

- Focus on identification of causal effects (across the board)
- Assessing External Validity
- Observing Unobservables
- Data collection
- Iterative Experimentation
- Unpack impacts

# Focus on Identification... across the board!

- The key advantage of RCT was perceived to be a clear identification advantage
- With RCT, since those who received a treatment are randomly selected in a relevant sample, any difference between treatment and control must be due to the treatment
- Most criticisms of experiment also focus on limits to identification (imperfect randomization, attrition, etc. ) or things that are not identified even by randomized trials (distribution of treatment effects, effects elsewhere).

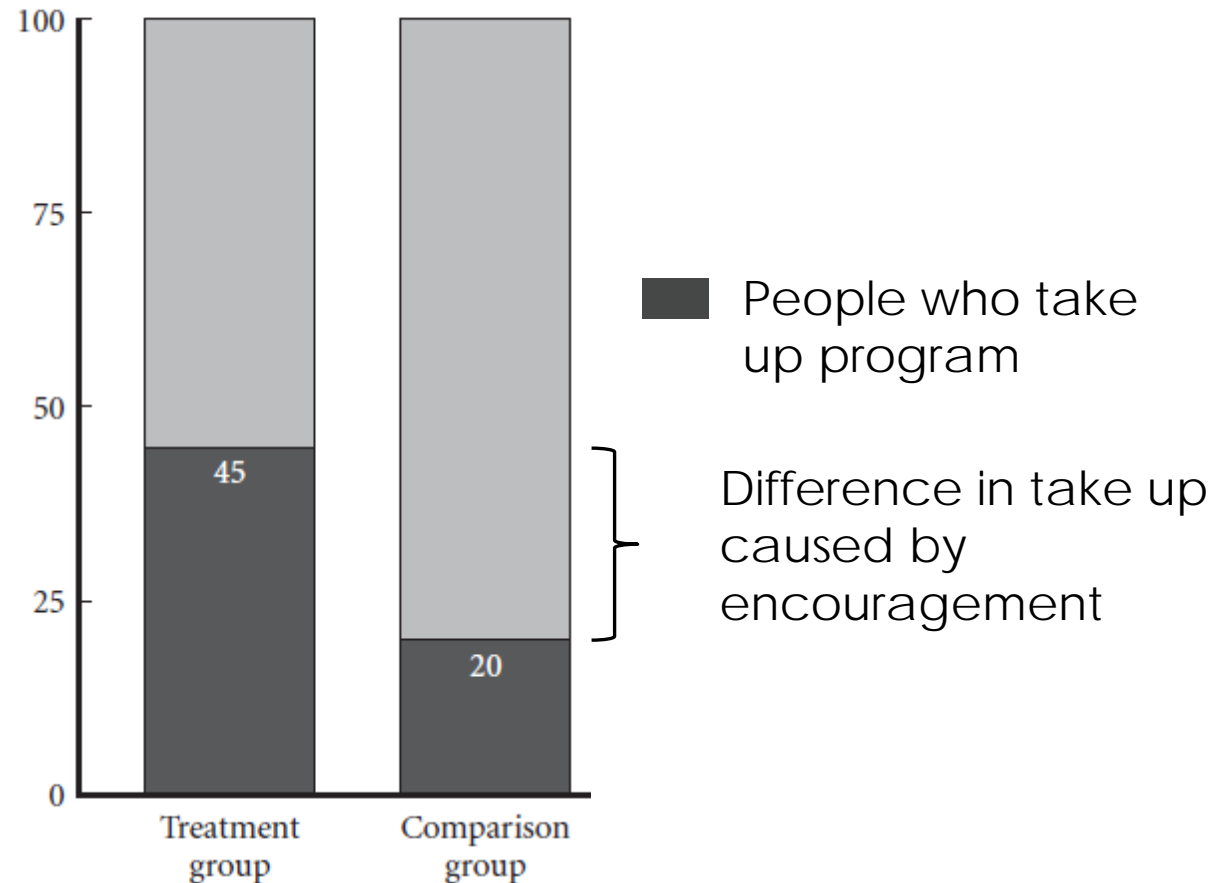


# Focus on Identification... across the board!

- Before the explosion of RCT in development, a literature on RCT in labor and public finance has thought of other ways to identify causal effects
- In development economics, there was a joint development of the two literatures (natural experiment and RCT), which has made both literatures stronger, and perhaps less different than we initially thought they would be:
  - Natural experiments think of RCT as a natural benchmark (not just an hypothetical gold standard).
  - Development of methods to go beyond simple comparison of treatment and control in experiments, and richer designs

# Encouragement design

Percentage of people eligible for the program



# Focus on Identification... across the board!

- Before the explosion of RCT in development, a literature on RCT in labor and public finance has thought of other ways to identify causal effects
- In development economics, there was a joint development of the two literatures (natural experiment and RCT), which has made both literatures stronger, and perhaps less different than we initially thought they would be:
  - Natural experiments think of RCT as a natural benchmark (not just an hypothetical gold standard). Extremely well identified non randomized studies.
  - Development of methods to go beyond simple comparison of treatment and control in experiments, and richer designs
- Ultimately, the advantage of RCT in terms of identification is a matter of degree, rather than a fundamental difference.

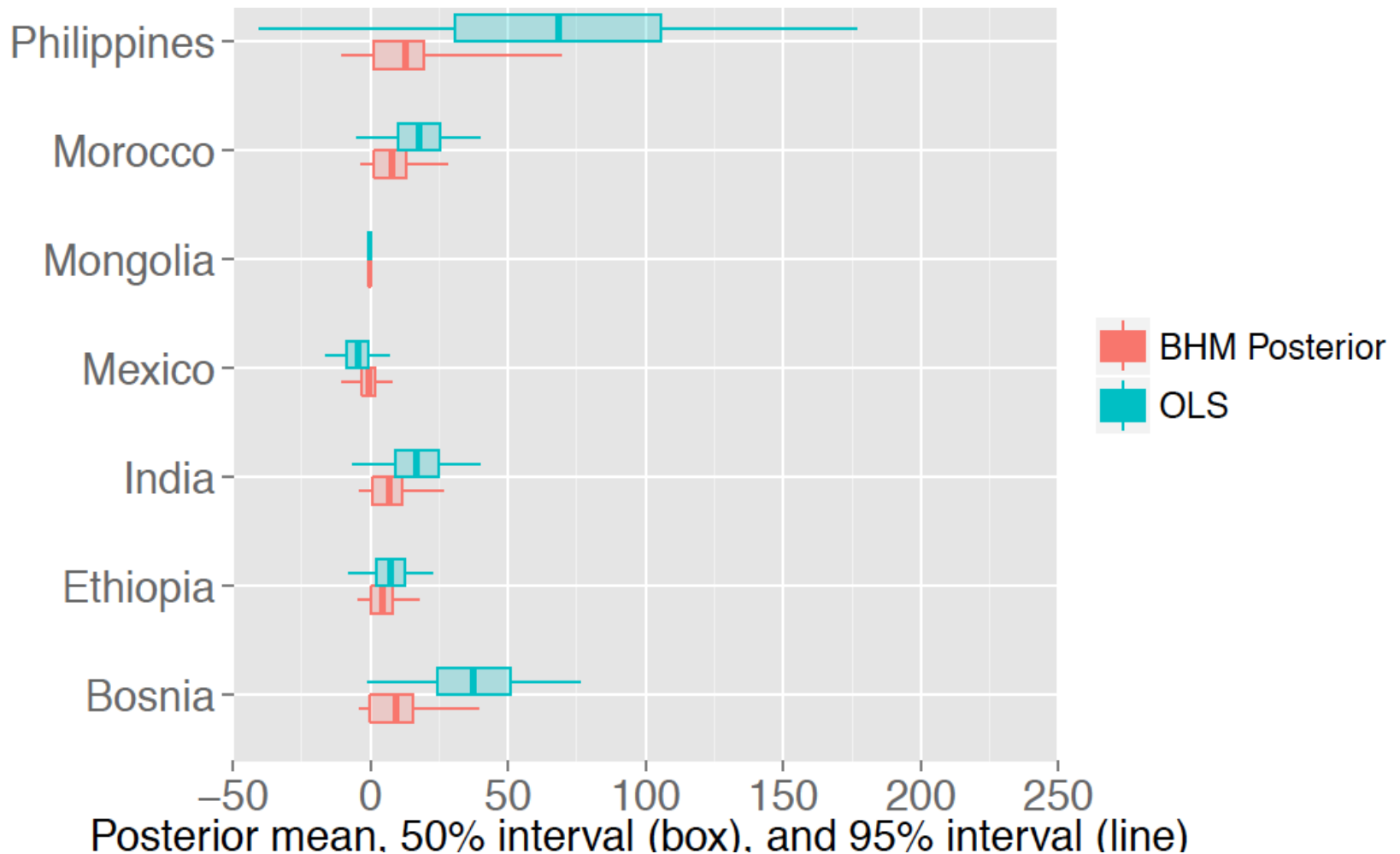
# Why have RCT had so much impact?

- Focus on identification of causal effects (across the board)
- *Assessing External Validity*
- Observing Unobservables
- Data collection
- Iterative Experimentation
- Unpack impacts

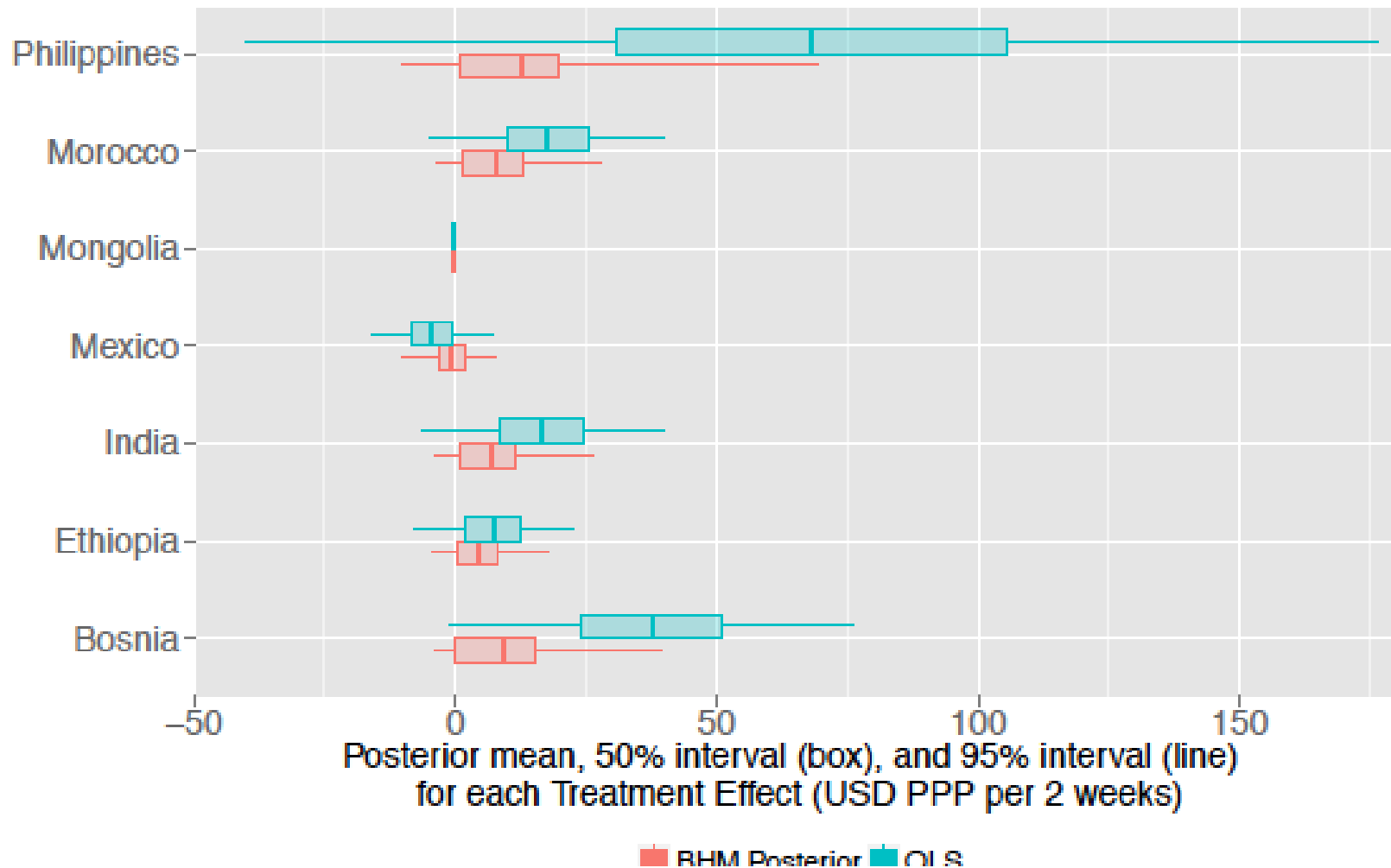
# External Validity

- Will results obtained somewhere generalize elsewhere?
- A frequent criticism of RCT is that they don't guarantee external validity
- Which is quite right, but it is not like they are less externally valid...
- And because they are internally valid, and because you can control where they will take place:
  - compared across contexts.
  - they can be purposefully run in different contexts
  - Prediction can be made of what the effects of related programs could be.

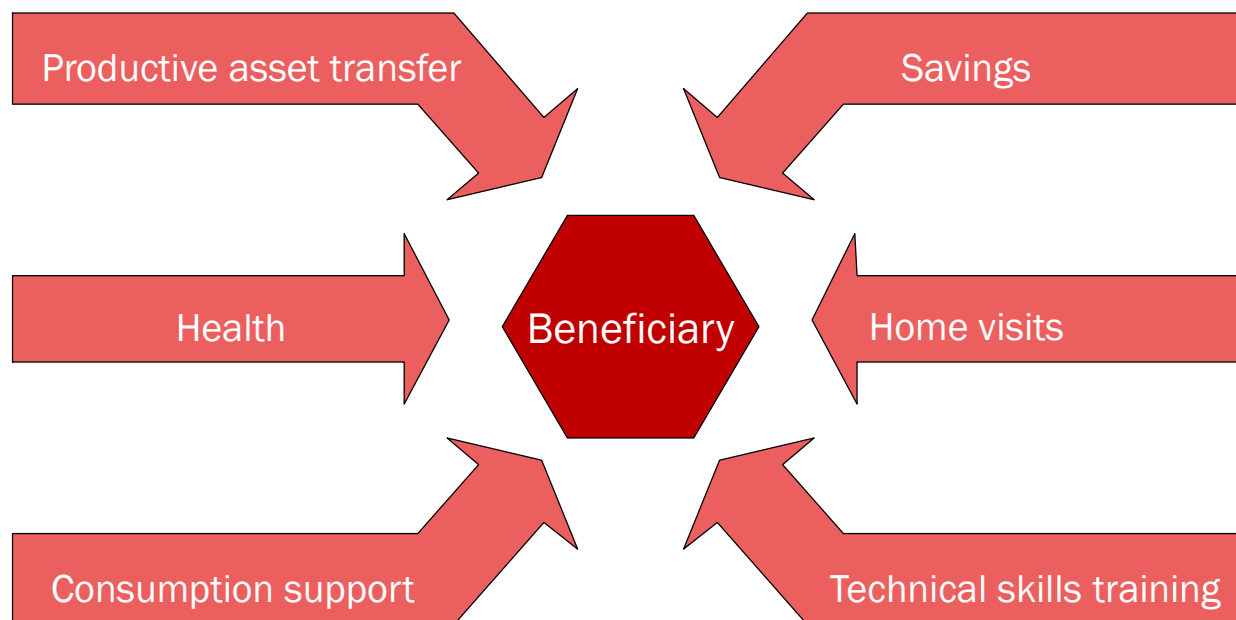
# Bayesian Hierarchical Modelling of all the MF results : Profits Meager (2015)



# Bayesian Hierarchical Modeling-- Meta analysis (consumption)



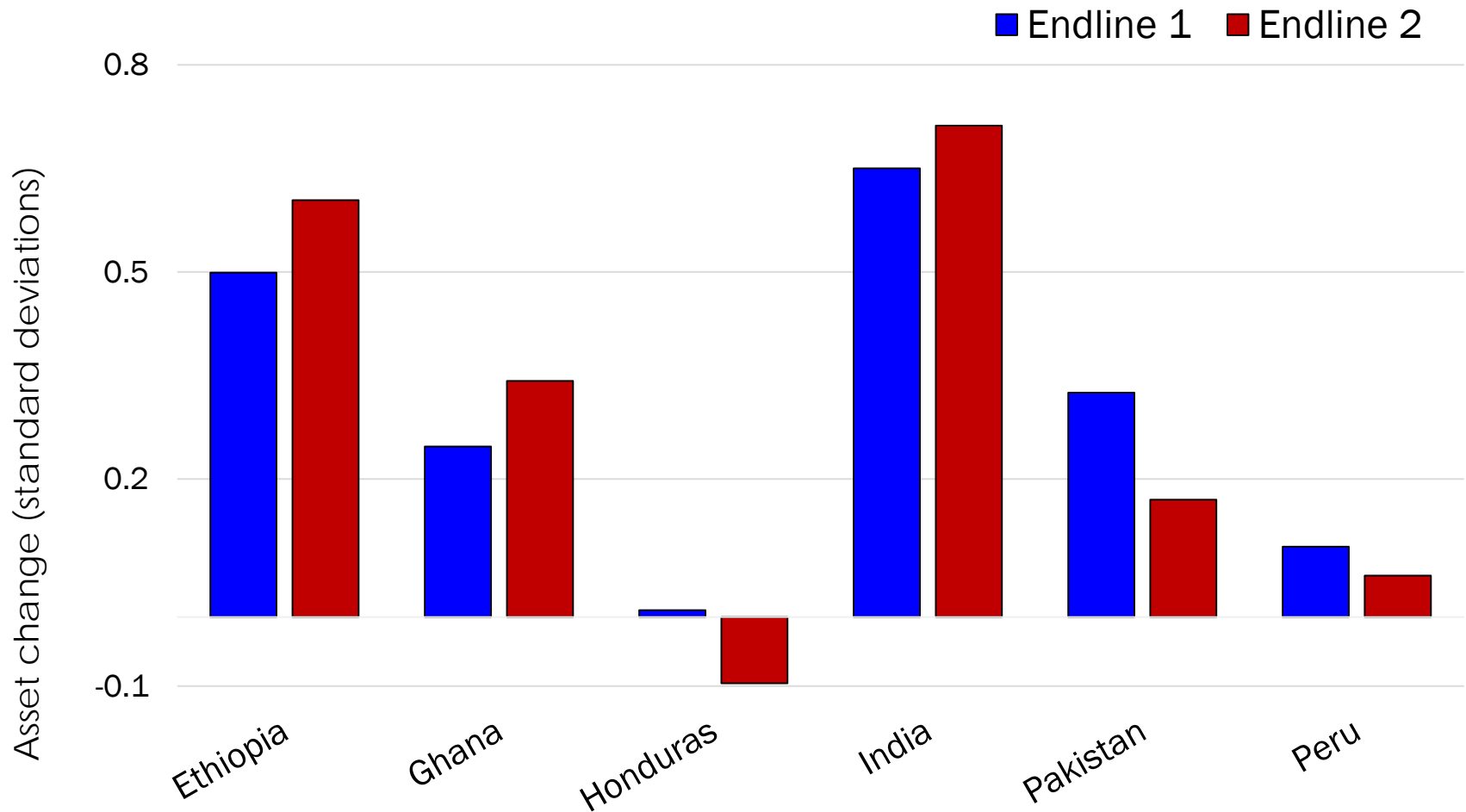
## Example 2: Targeting the Ultra Poor Program: Coordinated evaluation in several countries



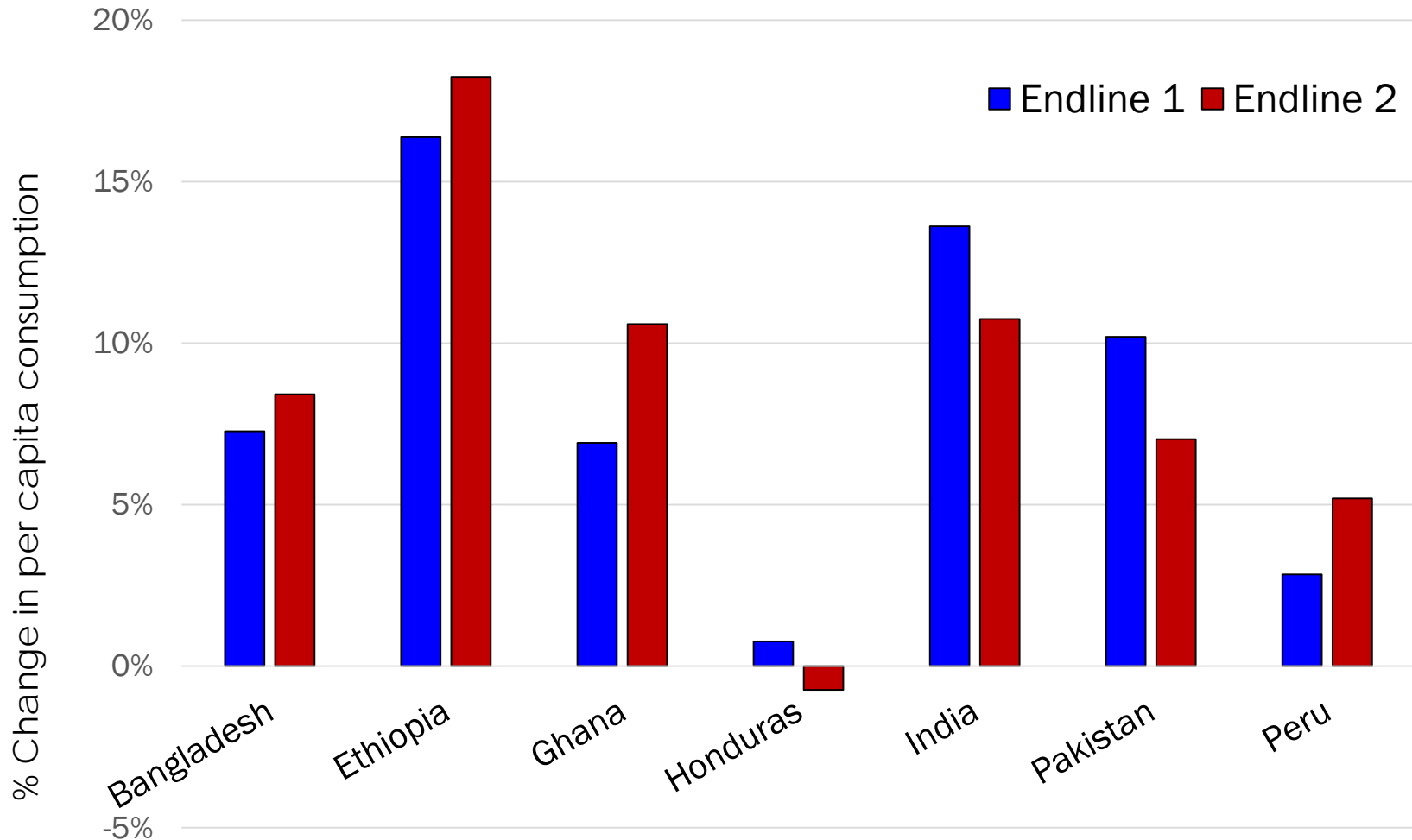
[Banerjee et al, 2015](#)



# Country by country results: Assets



# Country by country results: Consumption



# Structured Speculation

- Ultimately, if the results are similar it is nice, but if they are different the ex-post analysis is speculative.
- Banerjee, Chassang, Snowberg (2016) propose to be explicit about such speculation, and that researchers should predict what the effect may be for other interventions, or in other contexts.
- This can then motivate running such experiments, and guesses can be falsified.
- Example: Dupas (2014)—Effect of short run subsidies on long run adoption depend on the timing of costs and benefits, and how quickly uncertainty about them is resolved: this allows her to classify the goods.

# Why have RCT had so much impact?

- Focus on identification of causal effects (across the board)
- Assessing External Validity
- Observing Unobservables
- Data collection
- Iterative Experimentation
- Unpack impacts

# Observing unobservables

- Some things simply cannot be observed in the wild, with naturally occurring variation
- Negative income tax experiment was designed as an experiment to separate income and substitution effects
- Many experiments in development are designed likewise to capture such effects:
  - Karlan Zinman Observing Unobservables
  - Cohen Dupas and Ashraf Dupas Shapiro: selection and treatment effect of prices.
  - Bertrand et al. Corruption in driving licences in Delhi.

# Why have RCT had so much impact?

- Focus on identification of causal effects (across the board)
- Assessing External Validity
- Observing Unobservables
- Data collection
- Iterative Experimentation
- Unpack impacts

# Innovative data collection

- Innovative data collection does not require an experiment.
- But experiments have two features which have motivated creativity in measurement
  - We know precisely what we are trying to measure: payoff to the person who is designing the questionnaire
  - We know that there will likely be enough power to measure such effects
- As a result, lots of innovation in measurement:
  - Borrowing from other fields : psychology, political science, agriculture, web scraping, wearable technology,
  - Inventing new methods: e.g. Olken 2007

# Why have RCT had so much impact?

- Focus on identification of causal effects (across the board)
- Assessing External Validity
- Observing Unobservables
- Data collection
- Iterative Experimentation
- Unpack impacts



# Iterative experimentations

- Some great natural experiment leave us with some unanswered questions:
  - Why are elite school not working for the marginal child?
  - Why are (some) charter school working so well?
- One other advantage of experiments is that one is never stuck with one particular surprising answer: you can continue to experiment in the same setting till you have some clarity.
- Example: Duflo, Kremer, Robison multi-year work on fertilizer.
  - People don't use fertilizer, even though it is profitable
  - One set of experiment on financing
  - One set on learning and social learning.

# Why have RCT had so much impact?

- Focus on identification of causal effects (across the board)
- Assessing External Validity
- Observing Unobservables
- Data collection
- Iterative Experimentation
- Unpack impacts

# Unpack impacts

- This is a related point, but more narrowly focused on policy design.
- There are many many possible ways to design a particular programs
- Usually, one version is tried out
- But if it works what was essential?
  - Effort to unpack Conditional Cash Transfer
  - Example of doing everything at once: Raskin program, Indonesia
    - Many people do not receive the rice they are eligible for, or over pay
    - Would transparency help?

They distribute 4 version of a cards to eligible villagers in 378 villages, randomly chosen out of 572

A



B



C



D



# Other sources of variation and results

- They also varied:
  - Public (common) knowledge of the program
  - Fraction of people who get the physical card
- Results:
  - Making the card distribution public knowledge makes it more effective
  - The physical card matter: information (in the form of list) alone is not sufficient
  - (Perception of) accountability does not seem to make much of a difference.
- The government decided to scale the version of the card with most info and the list to 65 million beneficiaries!

# What has been the policy impacts of RCTs?

- Is the Raskin Case unique or unusual?
  - A study designed by researchers with several treatments and an underlying economic model, destined to be published in a to academic journal
  - but that still had large policy impact
- Some have argued that the research impact of RCT has potentially come at the expense of real-world impact:
  - Researchers' and policy makers interests may diverge
  - Research slow down the process of iteration
- Evidence
  - Out of 700 projects on going or completed on the J-PAL site, there are only 9 story of scale up or policy impact.
  - However, this is not a census of J-PAL study (or of RCT). Story selected for high impacts : the sum of people reached is about 200 million.

# Over 200 million people reached through scale-ups of programs evaluated by J-PAL researchers

Program	People Reached (mn)
School-based Deworming	95
Raskin: Subsidized Rice (Indonesia)	66
Teaching at the Right Level (India)	34
Generasi: Conditional Community Block Grants (Indonesia)	6
Chlorine Dispensers for Safe Water (East Africa)	0.5
Free Insecticidal Bednets	Policy influence
Police Skills Training	Policy influence
TOTAL	202 mn



# Getting a sense of overall influence is difficult

- Returns to R&D highly skewed.
  - Most scholarly articles are never cited
  - Most start-ups fail
  - Venture capitalists get most of their revenue from a small number of investments
- Still huge payoffs to R&D
- Ideas take a long time to percolate through the system, and many RCT are fairly recent
- Many RCT find that things DO NOT work as well as hoped (microcredit, smokeless stoves)



# RCTs and real world impact: The case study of DIV

- To solve the “census” problem, we focus on one case study: USAID DIV.
- USAID’s Development Innovation Ventures (DIV) offers an opportunity to compare outcomes in selected sample of award winners:
  - **DIV has open approach:** no top-down restriction on sector, strategy
    - Grantees include social entrepreneurs, NGOs and development researchers
  - **Staged financing** (Pilot  $\leq$  \$150,000; Testing  $\leq$  \$1,500,000; Transition to Scale  $\leq$  \$15,000,000)
  - **Openness on scaling strategy** (Commercial, public-sector or hybrid funding)
    - Emphasis on cost effectiveness;
    - Attention to management team, external commitments, but no rigid litmus test

# Methods

- **Coverage:** 43 DIV awards made from 2010-2012; total value \$17.3m
- **Here just examine reach,** the estimated number of people exposed to the original and adapted versions of the innovation, after the DIV funding.
  - Do not compare measures of the size of impact per beneficiary
  - Do not estimate the likelihood that reach will be sustained or increased in the future
  - Does not assume the credit to further expansion all goes to DIV
  - One (of several) components of social return calculation
- Specifically, we focus on number of awards reaching more than 100k or more than 1 M people.

# DIV-Supported several Innovations Reaching > 100K People

## 5 INNOVATIONS REACHED MORE THAN 1 MILLION PEOPLE

- Voter Report Cards (2 awards)
- Election Monitoring Technology
- Consumer Action and Matatu Safety
- Digital Attendance Monitoring
- Dispensers for Safe Water

## 6 INNOVATIONS REACHED MORE THAN 100K AND LESS THAN ONE MILLION PEOPLE

- Scaling CommCare for Community Health Workers (2 awards)
- d.light Innovative Financing for Solar Systems
- Sustainable Distribution for Improved Cookstoves
- Recruiting and Compensating Community Health Workers
- VisionSpring BoPtical Care
- Renewable Powered Microgrids for Rural Lighting

# High reach of innovations with RCTs/involvement of development economics researchers

- 24 awards had RCT/researcher involvement, of which:
  - 42% (10 awards) reached more than 100,000 people
  - 25% (6 awards) reached more than one million people
- 19 awards did not have RCT, of which:
  - 16% (3 awards) reached more than 100,000 people
  - No awards have yet reached more than one million people
- Overall DIV numbers favorable relative to many impact investors
  - Arbitrage opportunity from openness to multiple types of innovation?
  - Discipline of evidence useful?

While early stage awards have low probability of attaining reach, they have high expected reach per dollar spent

Award Stage	Number of Awards	Award Value	Fraction Reaching More than 100,000 people	Fraction Reaching More than 1,000,000 people	People Reached	Expenditure per Person Reached
Stage 1 (< \$100,000)	24	\$2,353,136	17% (4/24)	8% (2/24)	6,723,733	\$0.35
Stage 2 (<\$1,000,000)	18	\$9,557,926	39% (7/18)	17% (3/18)	16,931,044	\$0.56
Stage 3 (<\$15M)	1	\$5,516,606	100% (1/1)	100% (1/1)	2,754,193	\$2.00

# Pathways to reach

- DIV awards for innovations with RCTs reached > 100,000 people through a **variety of partnerships**:
  - Country governments (e.g. Zambia CHW recruiting, India biometric monitoring)
  - Donors (e.g. cookstoves in Ethiopia and Sudan)
  - NGOs/Social Enterprises (e.g. Dispensers for Safe Water)
  - Private sector firms (e.g. newspapers, banks, insurance companies, Qualcomm, Safaricom)
- Three of five innovations reaching more than one million people had earlier RCTs demonstrating impact, potential for cost effectiveness: researcher/project selection?

# Why might projects involving RCT be more likely to have future reach?

- Convincing force of evidence [most projects that do not involve RCT try to scale through retail sale, which is harder]
- Nothing to do with the RCT per se:
  - Close involvement of researchers help ideas grounded in basic science percolate research (like in biotech).
  - In particular: Influence of behavioral economics/information: focus on low cost interventions , which are more likely to scale
- Selection of good projects [willing to do an RCT]

# Conclusion

- The projects evaluated by RCT that then have reached many people tend to be low-cost, well defined, simple.
- So what has made RCT useful as a research tool (ability to iterate, zero-down to component, test a theory) is exactly what has turned out to make them policy relevant: details matter tremendously, and RCT tend to get the details right.
- An alternative pathway: BRAC, PROGRESA. Complexed interventions replicated in many contexts.
- And a third one: innovation within existing governments and institutions.
  - Tamil Nadu innovation Fund, Nudge Unit
  - Gujarat Pollution Control Board



# Conclusion

- For RCT to move from the large research impacts to large policy impact, we need a range of complementary institutions:
  - Meta-analysis
  - Review article
  - Review panels
  - Registry of Experiments has started and is successful (706 studies as of June 8),
- Appropriate support and experiment to support the learning needed to move from successful pilot to policy at scale
  - Iterations to design scalable (robust) versions and measure their effects
  - Equilibrium effects
  - Political economy /industrial organization of implementation