



J-PAL

ABDUL LATIF JAMEEL POVERTY ACTION LAB

# The Generalizability Puzzle

Mary Ann Bates

Executive Director, J-PAL North America

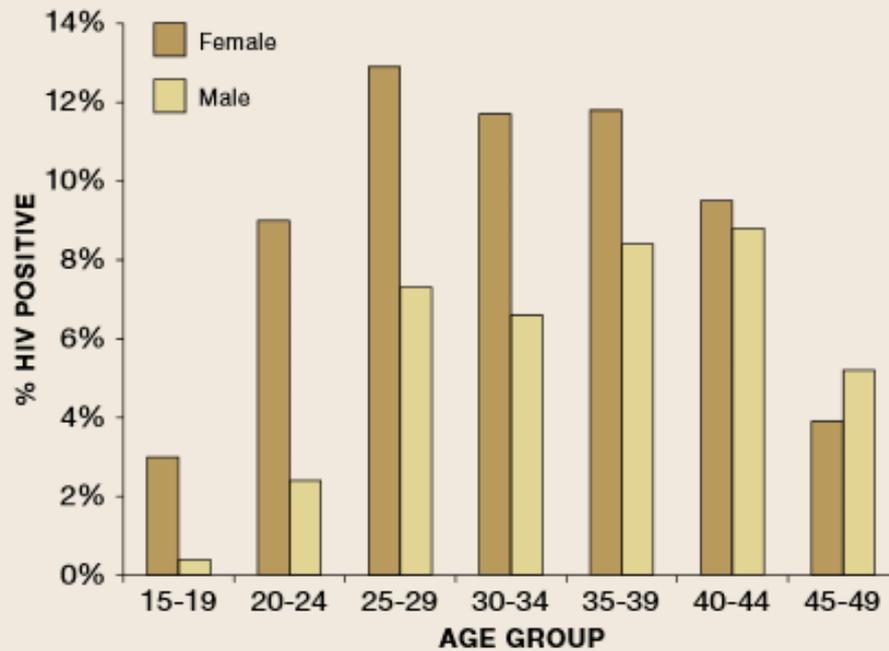
MIT



# Course Overview

1. What is Evaluation?
2. Measurement & Indicators
3. Why Randomize?
4. How to Randomize?
5. Sampling and Sample Size
6. Threats and Analysis
7. Start to Finish
8. Generalizability

**Figure 1: HIV Rates Are Very Different by Age**



Source: The Kenya Demographic and Health Survey (Central Bureau of Statistics, Kenya, 2004)

# Randomized evaluation: Relative risk intervention

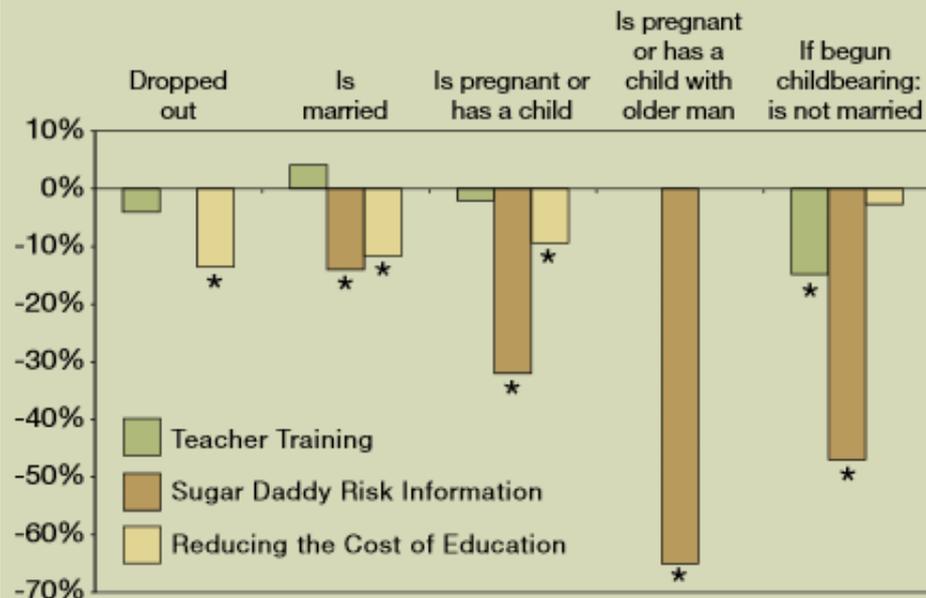
- Study by Pascaline Dupas (Stanford)
- Location: rural western Kenya
- 71 schools randomly selected from 328 schools
- Trained project staff visited the 8<sup>th</sup> grade classrooms
  - 10-minute video
  - Detailed stats on the rates of HIV by age and sex from nearby Kisumu
  - 30-minute discussion of cross-generational sex

# Results

- Childbearing with older men fell by more than 60%
- No offsetting increase in childbearing with same-age peers
- Impact measured by a randomized controlled trial (RCT)
- Much more effective (and cost-effective) than alternative programs

## Impacts of Each Program on Girls' Behavior

(Percentage change relative to girls in comparison group)



\* Indicates that the difference with the comparison group is significant at 10%

Should Rwanda replicate the program?

A. Yes

B. No

# The challenge

- Dramatic rise in the number of rigorous impact evaluations in developing and developed countries in last 20 years
- Unlikely to be rigorous evaluation of the program policy makers wants to introduce in exactly same location

# The generalizability puzzle

- Can a study inform policy only in the location in which it was undertaken?
- Should we use only whatever evidence we have from our specific location?
- Should a new local randomized evaluation always precede scale up?
- Must an identical program or policy be replicated a specific number of times before it is scaled up?
  
- What counts as a “similar enough” new setting?

# Structured Approach to Evidence in Policy

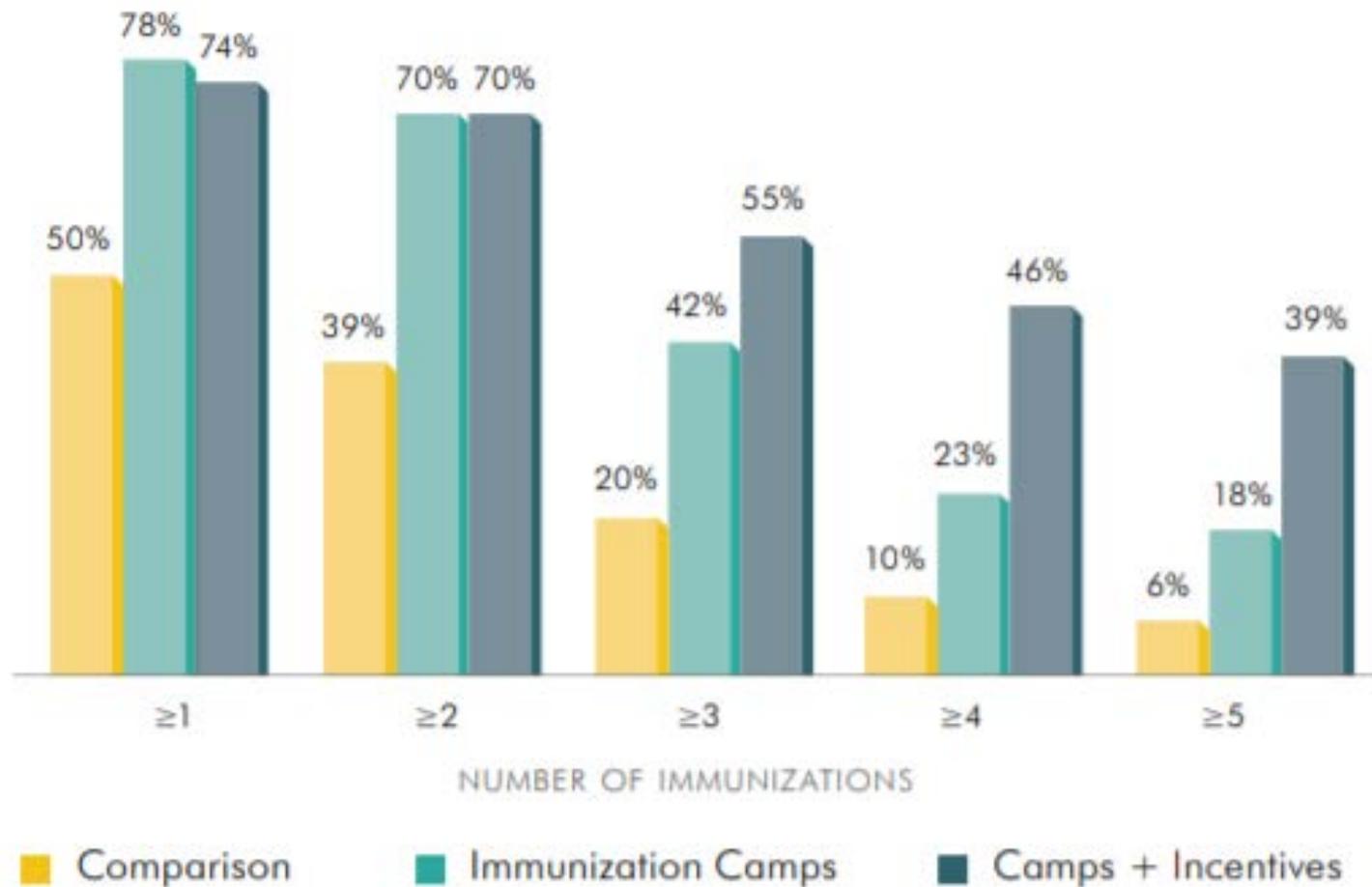
- Evidence from single study just one part of the puzzle
  - We weigh the evidence based on quality and adjust priors
- Combine, theory, descriptive evidence, and results of rigorous impact evaluations to answer:
  - Whether results from one country likely to replicate in another
  - When we need more evaluation and when we don't
- Draw on a theory based review of 70+ RCTs on health econ in dev countries (Kremer and Glennerster, 2012)

# Scaling immunization incentives

- Seva Mandir program to increase immunization rates in rural Rajasthan, tested with RCT
  - Banerjee, Duflo, Glennerster, Kothari, 2010
- Fixing **supply**: regular monthly immunization camps with nurse present without fail
- Building **demand**: 1kg lentils for every vaccination, set of plates on completed immunization schedule

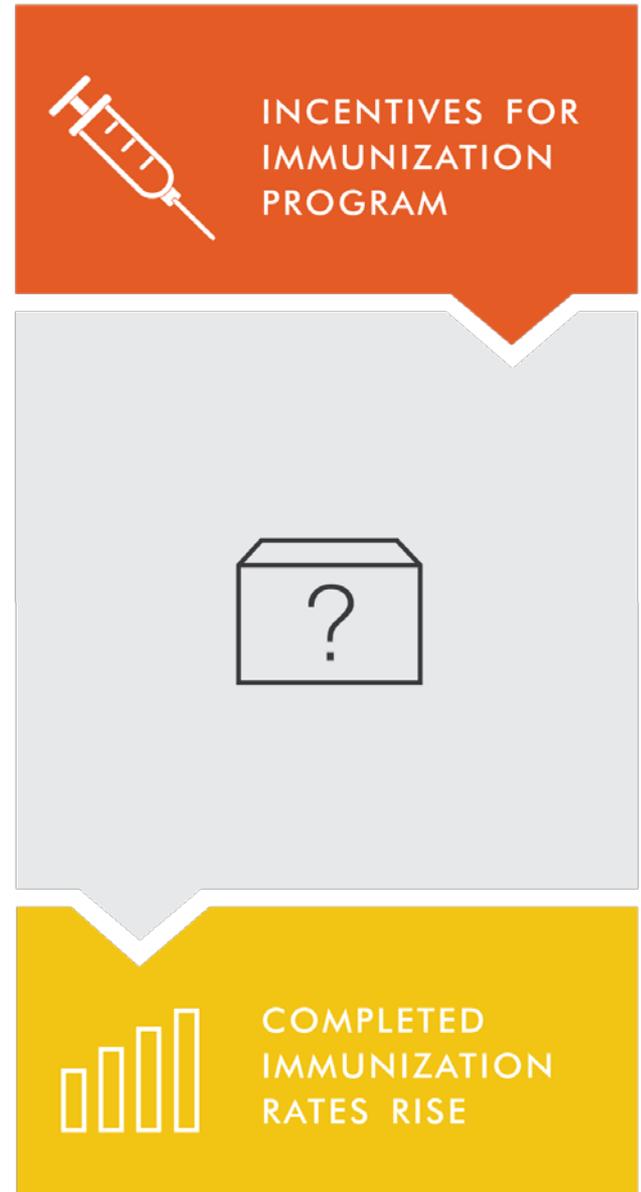


**FIGURE 1: NUMBER OF IMMUNIZATIONS RECEIVED BY CHILDREN AGED 1-3 YEARS**



# Viewing evidence in isolation

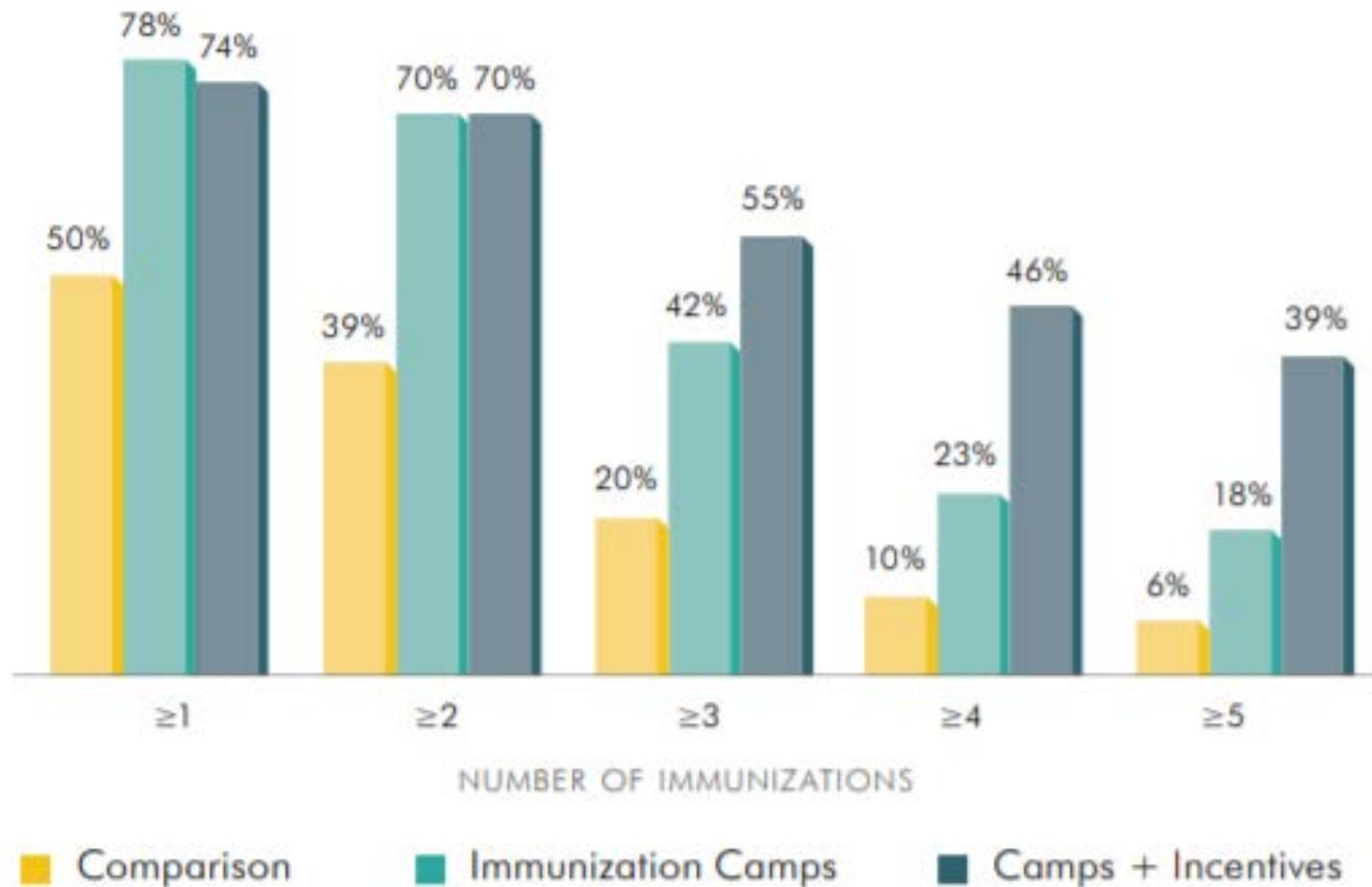
- If a government in West Africa wanted to improved immunization rate, should they consider noncash incentives?
- Only one RCT in South Asia not Africa
- Program conducted by NGO not government
- Lentils not core part of local diet



# Generalizability Framework



**FIGURE 1: NUMBER OF IMMUNIZATIONS RECEIVED BY CHILDREN AGED 1-3 YEARS**



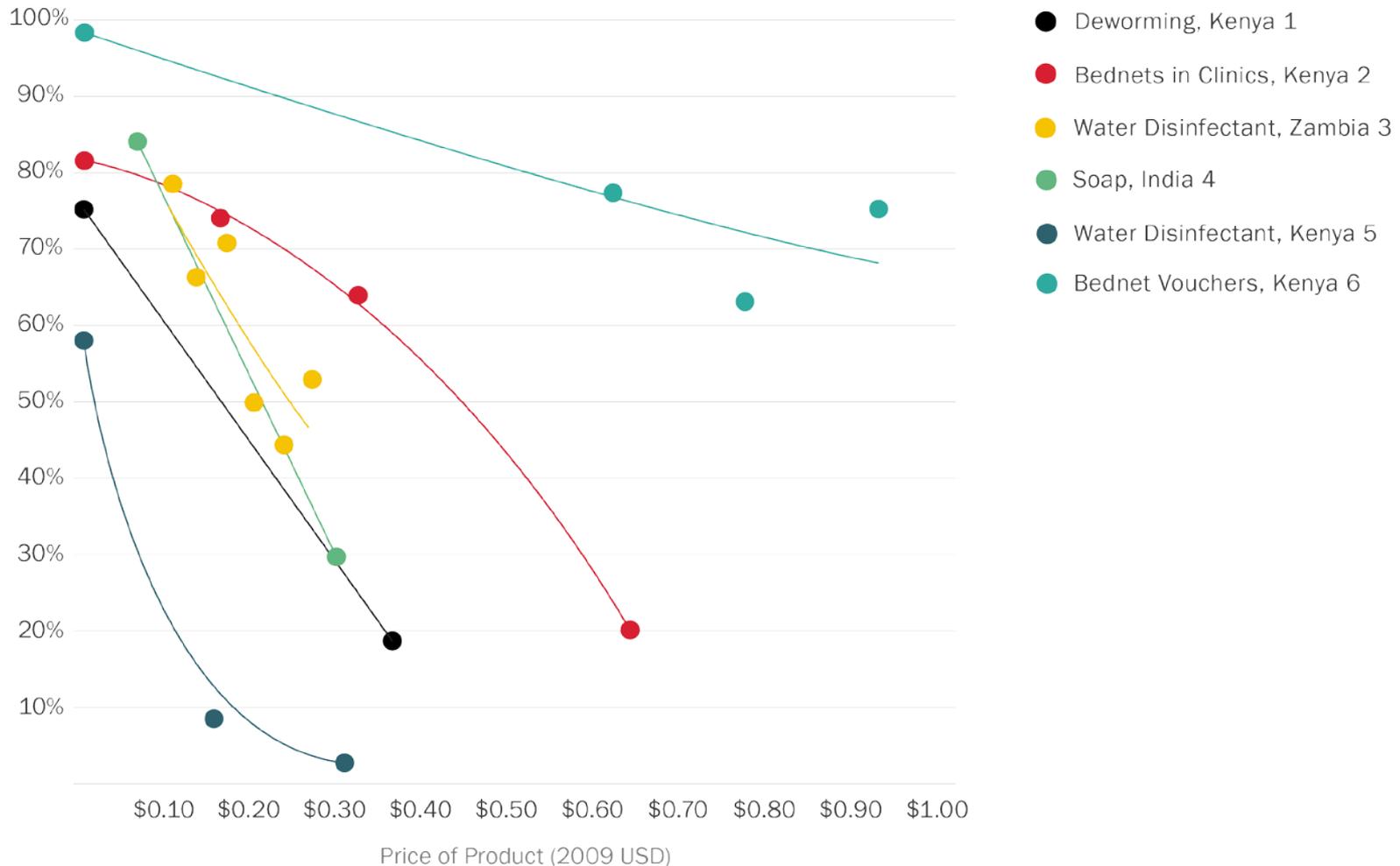
# Generalizability Framework



# Evidence on present bias

- People procrastinate and find hard to stick with behavior they believe is good for them and their children
  - Good theoretical work showing how small changes to a standard discounting model produces series of testable conclusions and can explain many stylized facts (e.g. Laibson, 1997)
  - Small changes in price of preventative products sharply reduces take up (9+ RCTs)
  - People are willing to pay to tie their own hands with commitment savings products: difficult to explain unless people know they are present biased (e.g. Gine et al. 2010)

# Price Sensitivity of Preventative Health

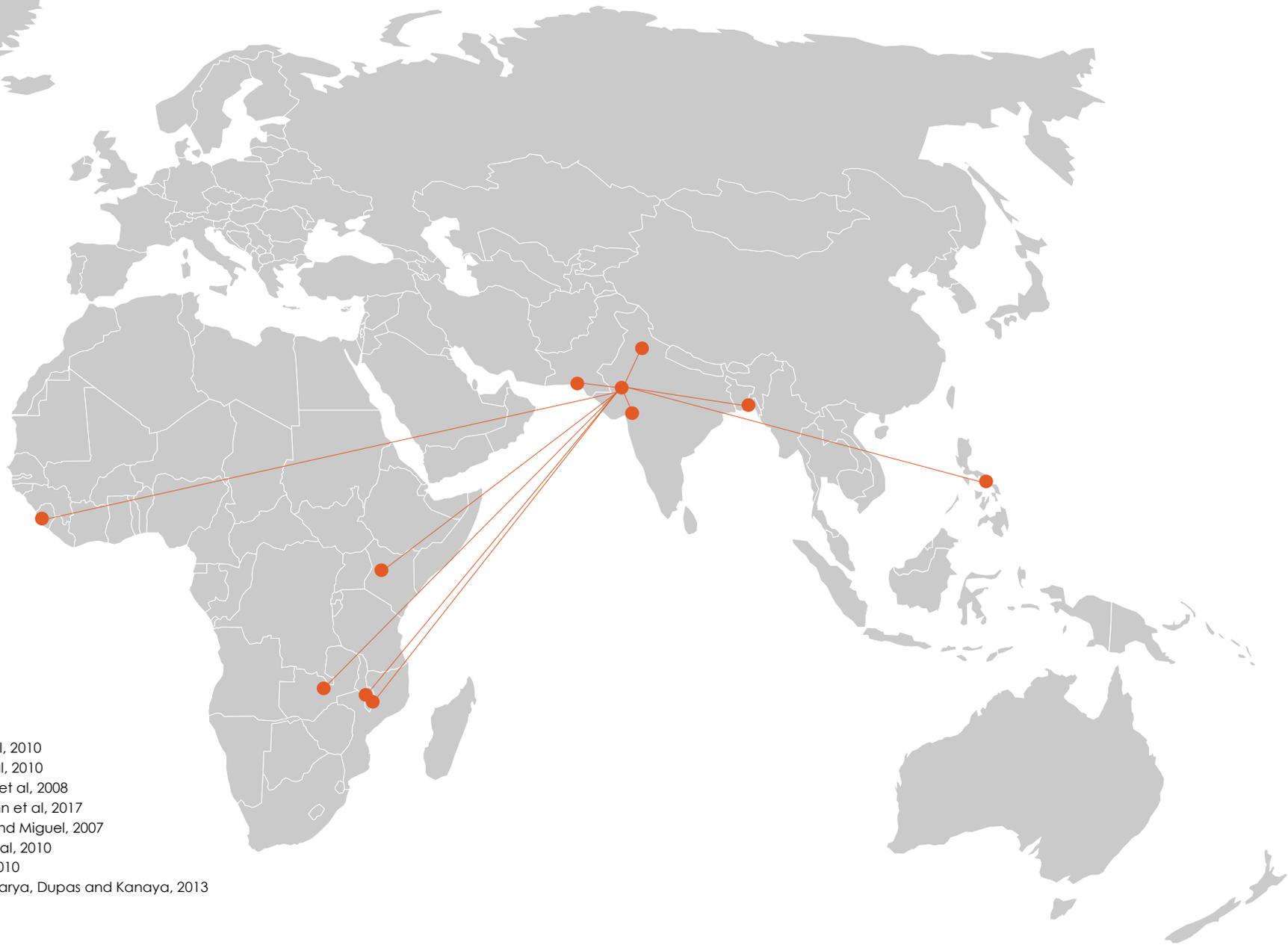


Source: Kremer and Miguel 2007, Ashraf et al 2010, Spears 2010, Dupas et al in process, & Dupas 2013. All as summarized in J-PAL Policy Bulletin. 2011..



## Small incentives can have big impacts on behavior

- 30+ RCTs of CCTs but usually much bigger incentives (Fiszbein and Schady, 2009)
- Malawi: smaller CCT same impact as bigger CCT (Baird et al 2010)
- Small incentives for HIV testing (Thornton 2008 Malawi), age of marriage (Field et al, in progress Bangladesh)



Sources:

Gine et al, 2010

Baird et al, 2010

Thornton et al, 2008

Buchmann et al, 2017

Kremer and Miguel, 2007

Ashraf et al, 2010

Spears, 2010

Bhattacharya, Dupas and Kanaya, 2013

# Generalizability Framework



Is either country a good potential scale up location?

<b>Immunization rates by antigen</b>		
	<b>Country 1</b>	<b>Country 2</b>
DPT1	84	47
DPT3	74	41
Measles	67	41
Fully immunized	49	38

Which country is a good potential scale up location?

A. Country 1

B. Country 2

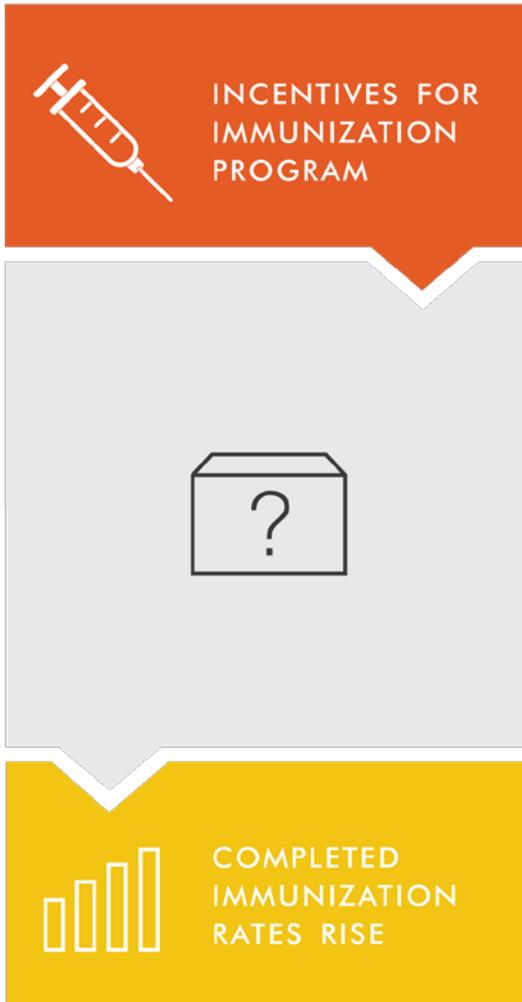
C. Neither

D. Both

What local implementation issues would you consider?

# Local Evidence on Implementation

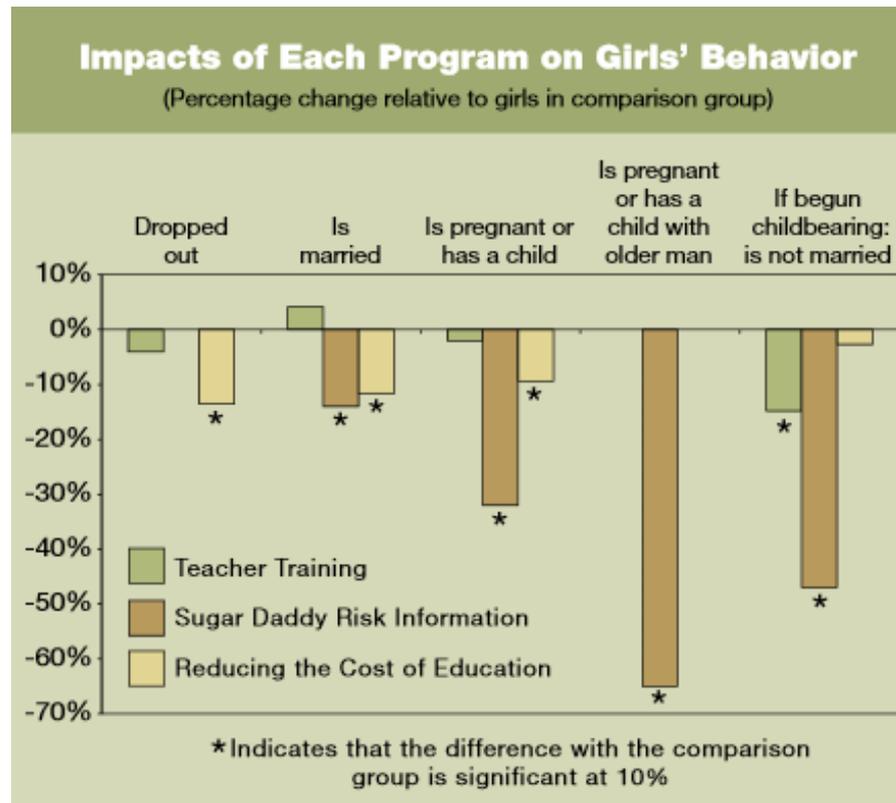
- This is where the switch from reliable NGO to government delivery will be critical
- Result with a government might be different than with NGO, should we do an RCT?
- What other information, evidence might be useful?
- Would be good to have more evidence on how to improve incentives for effective delivery within government



- 1. Minimal risk from overvaccination
- 2. Parents procrastinate or fail to persist
- 3. Parents are highly sensitive to price of preventative health



# Would the “Sugar Daddies” program work in Rwanda?



# Generalizability Framework: HIV Relative Risk Program

- Girls trade off the costs and benefits of sex
  - Older men give more gifts and can support you if you get pregnant
  - Girls know that unprotected sex can lead to HIV
  - Girls don't know older men riskier than younger men
- Impact of information on behavior depends on how it changes people's priors
- Key question for scaling is **prior beliefs in new populations**

What local information would be relevant?

What conditions would need to be similar?



## Local descriptive data (collected in a few weeks)

- In Rwanda, men ages 25-29 have an HIV rate of **1.7 percent** compared with 28 percent in the district in Kenya where the original evaluation was carried out.
- 42 percent of students estimated that more than 20 percent of men in their 20s would have HIV
- **Less than 2 percent** of surveyed students correctly identified the HIV prevalence rate for men in their 20s as being less than 2 percent.
- **In which direction would a risk awareness program change the Rwandan students' prior beliefs?**

Should Rwanda replicate the program?

A. Yes

B. No



## INFORMATION ON RELATIVE RISK OF HIV BY AGE

1. Relationships between older men and adolescent girls are common
2. Older men offer more financial protection against pregnancy
3. Older men have higher rates of HIV than younger men
4. Girls do not know that older men have higher HIV than younger men
5. Girls trade off costs and benefits of sex with different partners



## LOCAL CONDITIONS

1. Increasing perceived relative risk of HIV with one group leads to reduction in sexual activity with that group



## GENERALIZED LESSONS ON BEHAVIOR



## LOCAL IMPLEMENTATION

1. Relative risk information can be conveyed effectively to girls



## RISKY SEX WITH OLDER MEN REDUCES, LESS RISK OF HIV

# Teaching at the right level



If  $3x - 10 = 24$ , then  $x = ?$

For all  $a$  and  $b$ ,  
 $6a^2b^3 - 3a^2b$  is equivalent to  
which of the expressions?

$$8 + 14 - 7$$

$$7 \times 4$$



Photo: Aude Guerruci



Photo: Pratham

$$\frac{\pi}{\pi} 1500 = \frac{\pi}{\pi} (5x)^2 \cdot 6x$$

$$1500 = 25x^2 \cdot 6x$$

$$\frac{1500}{1.5} = \frac{1.5x^3}{1.5}$$

$$1000 = x^3$$

$$\sqrt[3]{1000}$$

$$x = 10$$

O.R

$$r = 5m$$

$$h = 60m$$

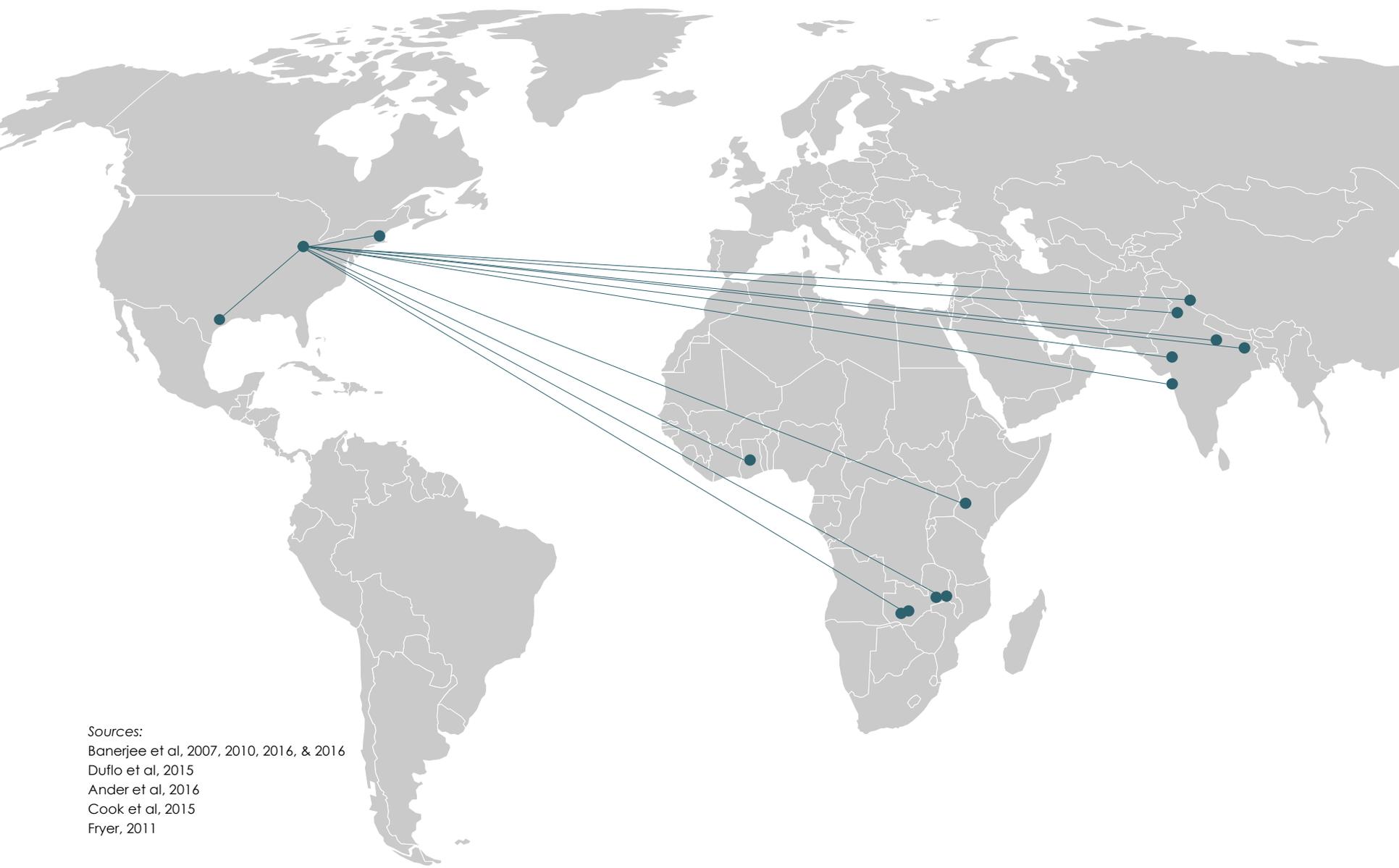
$$3a + 2a^2 - b + b^2 + a^2$$

$$3A + 3A^2 - b$$

$$2(4x+6) + 3(2x+1)$$

$$8x+12 + 6x+3$$

$$14x+15$$



Sources:  
Banerjee et al, 2007, 2010, 2016, & 2016  
Duflo et al, 2015  
Ander et al, 2016  
Cook et al, 2015  
Fryer, 2011

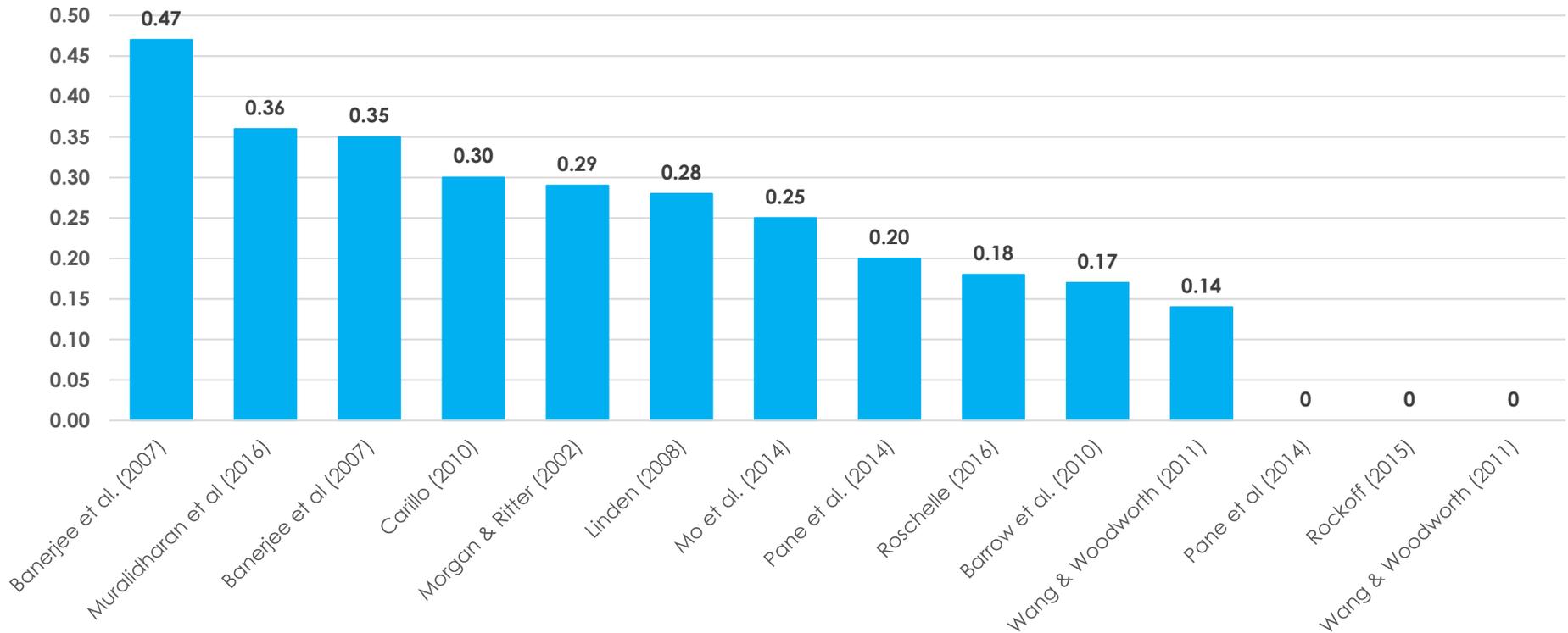
# Targeted Instruction Increases Learning

Series of studies shows targeted instruction can work in a variety of contexts:

1. Extra Teacher Programme in Kenya (Duflo et al 2011)
2. Balsakhi Assistant Programme in India (Duflo et al 2007)
3. Read India Programme (Banerjee et al 2007)
4. India Reading Camps (Banerjee et al 2010)
5. Haryana Learning Enhancement Programme (Berry et al 2013)
6. TCAI Programme in Ghana (Duflo and Kiessel in progress)
7. Computer Assisted Learning (Duflo et al 2007)

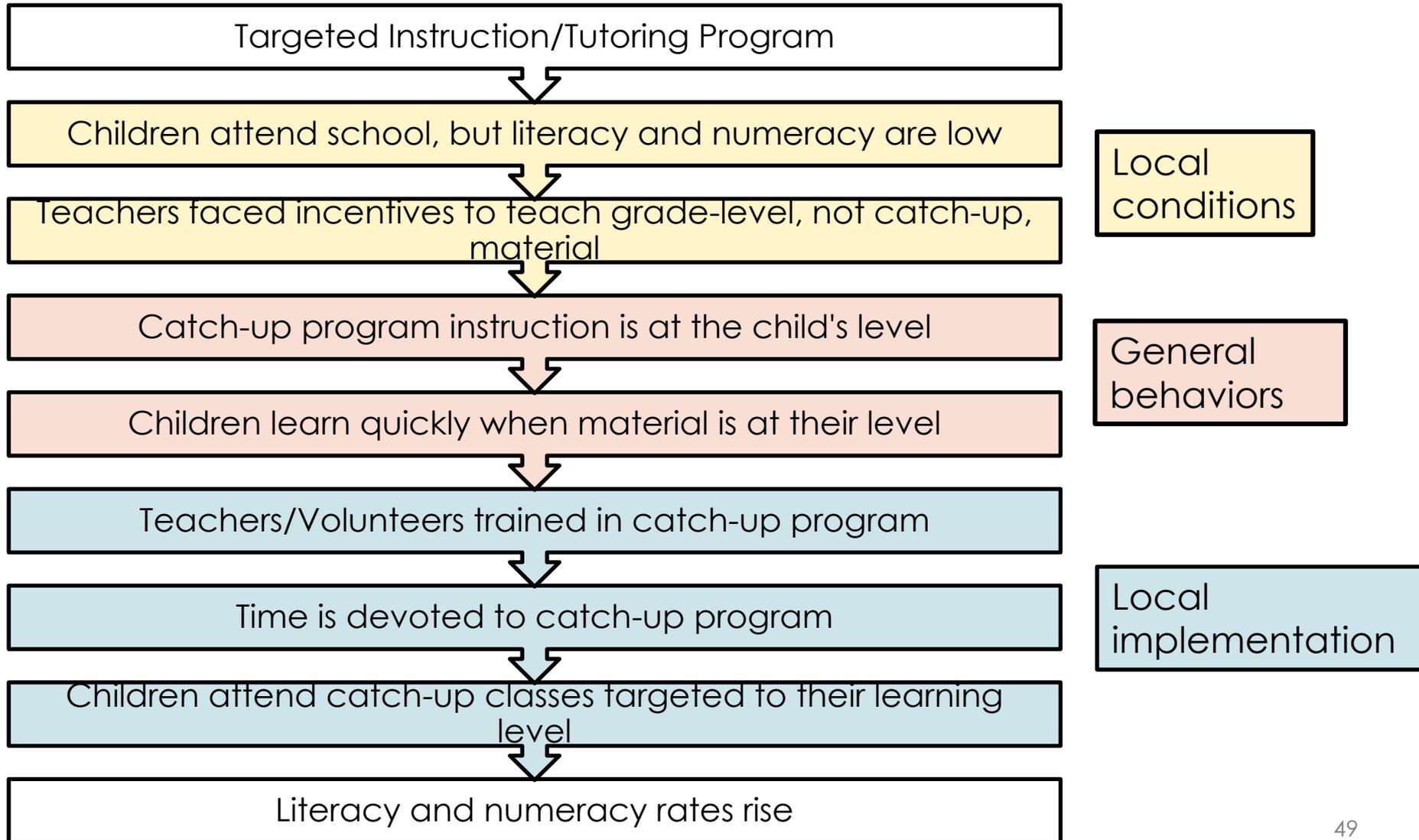
# Personalized learning is highly effective across studies

Computer-Assisted Personalized Learning's Impact on Math Outcomes



For details see J-PAL North America's review:  
Education Technology: An Evidence-Based Review by Escueta et al.

# Necessary Steps for Targeted Instruction



# Many Implementation Models

Who should lead the programme?	Where should the programme be held?	When should the programme be held?
1. Teachers	1. In schools	1. During the school day
2. Low-cost Tutors	2. Outside of schools	2. After school hour
3. Unpaid volunteers		3. On holiday breaks
4. Computer-Assisted		

Main lesson: Create a dedicated time to focus on the learning level of each child, especially those who have fallen behind on basic skills.

Results replicated in tutoring program in Chicago.  
Working with Government of Zambia to scale.

# Activity

You are the leaders of the Los Angeles Unified School District, and are looking for ways to boost student performance in your schools.

You recently heard about teaching at the right level, and want to explore whether it makes sense for you to implement this program in your schools.

## Small Group Discussion

- What metrics and data would you use to assess whether the important local conditions hold in your school district?
- How would you determine what grades and students to target?

*See Handout*

**TEACHING AT THE RIGHT LEVEL: WORKSHEET**  
You are the head of the Los Angeles Unified School District. How would you assess whether or not to implement SAGA Innovations' "tutoring model" in your schools?

Local Conditions	Metrics	Data Sources	How would you determine which grade levels and students to target?
Students (or at least some) are performing below grade level.			
There are varying levels of student achievement in classrooms, with some students performing above, at, and below grade level.			
Teachers teach of one level for all students in their classrooms, for either practical reasons or the school's incentive structure.			

# Considering Implementation

- Is implementing this program without modifying the critical components realistic in your particular context?
- Who in your jurisdiction would implement the program? What is their current capacity and experience implementing similar programs, and how confident are you in their implementation abilities?
- What kind of implementation capacity is needed to implement the program or policy (e.g. new business processes, staffing, funding, etc.)?
- Are there any local hurdles to implementation that need to be overcome?

# Resources In Development

**IMPROVING ADHERANCE TO IMMUNIZATION SCHEDULES: SUMMARY**

PROBLEM	LOCAL CONDITIONS	WHY IT WORKED	IMPLEMENTATION	OUTCOMES
What was the problem the intervention sought to address?  Parents were not taking their children to clinics to receive the full schedule of shots to become immunized, so children were not getting completely immunized.	What were the local conditions relevant to the program's success?  1. Parents wanted their children to get immunized, or at the very least had no reservations about immunizations. 2. Small clinics, such as the local health center, were getting a 40% hit, a strong performance for immunization. 3. Parents' willingness to send their child to a clinic was sensitive to small changes in price.	How were the generalized needs met through the intervention to be achieved?  Small clinics often lacked capacity from spending with predominantly healthy children. Parents tended to procrastinate or fail to enroll.  Offering incentives can be an effective way to improve persistence and overcome these small costs.  There is minimal risk from drug vaccination.	How did the program in practice work, and what implementation elements were critical to its success?  A mobile immunization team conducted weekly vaccination clinics where parents could bring their children to be vaccinated through a sign.  Parents who come to the clinic at their own volition are often more given to small incentives (e.g., free book, small bag of snacks).  <b>Implementation Elements:</b> 1. Flexible access to immunizations was provided in a convenient location. 2. Incentives were given in small amounts each time the parent got their child immunized. 3. The program was easily accessible to parents living around the program.	What was the impact of the intervention?  The intervention increased immunization rates in high-risk areas by 39 percent.

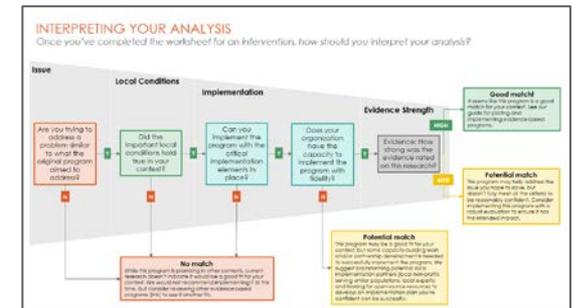
**Strength of evidence:** There is a substantial amount of rigorous evidence indicating that small costs often deter people from enrolling with preventive health behaviors in many different contexts, and small incentives can help overcome procrastination.

Summary of key local conditions, general lessons on behavior, and critical implementation elements for program or policy

**IMPROVING ADHERANCE TO IMMUNIZATION SCHEDULES: WORKSHEET**  
Should I implement this in my community?

Local Context Questions	Implementation Questions
<p><b>What is the primary problem you are trying to solve? How exactly will the intervention in your program be implemented to address it?</b></p> <p>1. Are there any other conditions, risk factors, or barriers that could prevent the intervention from being successful? How would you address these?</p> <p>2. How does your community, region, or area compare to the target population? What are the strengths and weaknesses of your community?</p> <p>3. How do you think the program will be implemented in your community?</p> <p>4. How do you think the program will be implemented in your community?</p> <p>5. How do you think the program will be implemented in your community?</p> <p>6. How do you think the program will be implemented in your community?</p> <p>7. How do you think the program will be implemented in your community?</p> <p>8. How do you think the program will be implemented in your community?</p> <p>9. How do you think the program will be implemented in your community?</p> <p>10. How do you think the program will be implemented in your community?</p>	<p><b>Is the "Program Plan" (your intervention) likely to be successful in your community?</b></p> <p>1. How do you think the program will be implemented in your community?</p> <p>2. How do you think the program will be implemented in your community?</p> <p>3. How do you think the program will be implemented in your community?</p> <p>4. How do you think the program will be implemented in your community?</p> <p>5. How do you think the program will be implemented in your community?</p> <p>6. How do you think the program will be implemented in your community?</p> <p>7. How do you think the program will be implemented in your community?</p> <p>8. How do you think the program will be implemented in your community?</p> <p>9. How do you think the program will be implemented in your community?</p> <p>10. How do you think the program will be implemented in your community?</p>

Worksheet to assess potentially bringing this program or policy to your jurisdiction



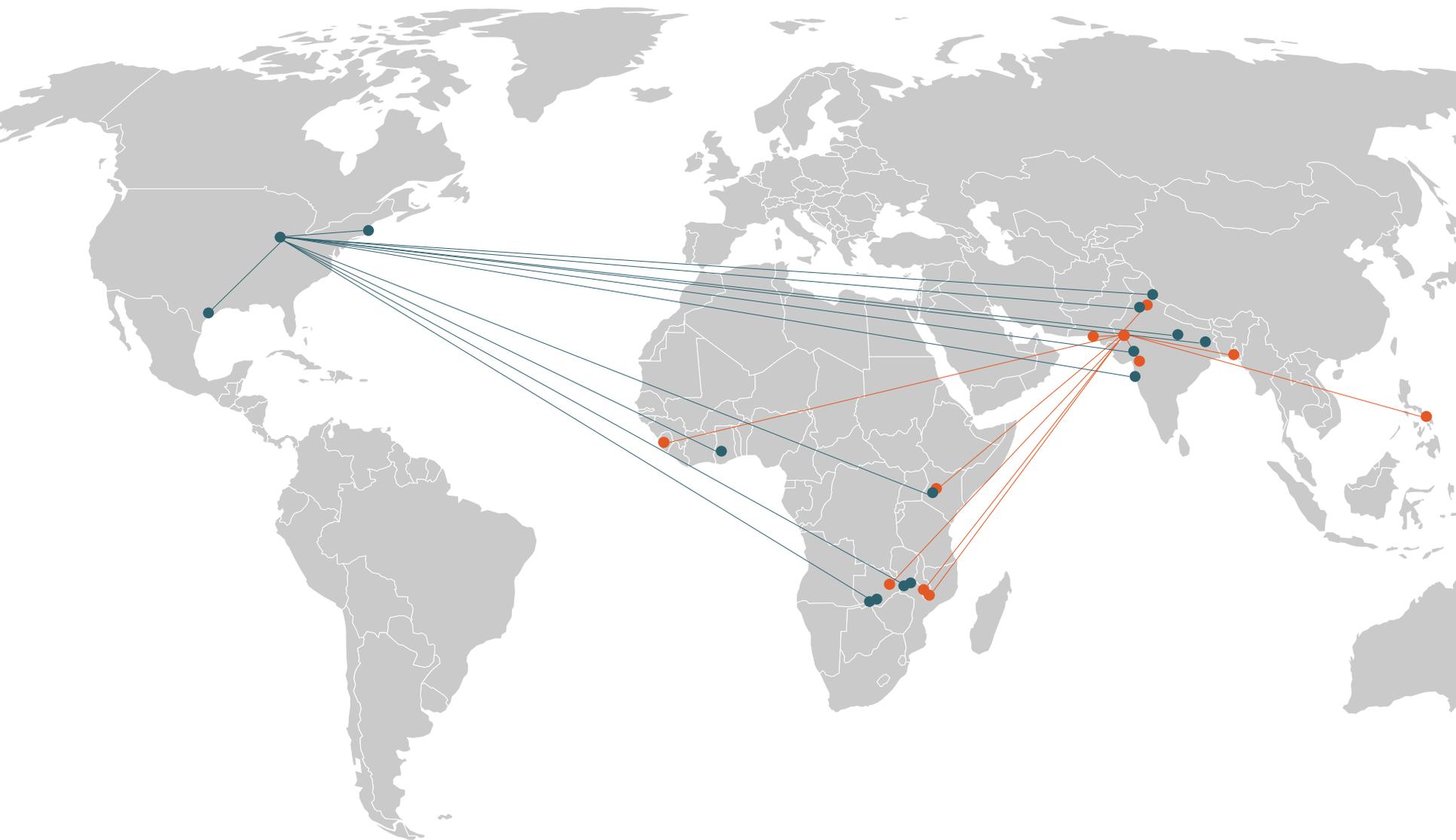
Decision tree on how to proceed based on your assessment

# Conclusion

- Does evidence from RCTs replicate to new context? Too big a question, need to break it down:
  - What is the theory of change behind the RCT?
  - Do the local conditions hold for that theory to apply?
  - How strong is the evidence for the general behavioral change?
  - What is the evidence that the implementation process can be carried out well?

# Conclusion

- If we have enough evidence to act, do we have enough evidence to stop evaluating impact? (always monitor)
  - we often need to act even when evidence is thin
- Often big overlap between when have enough evidence to launch big new initiative and when still worth evaluating
  - Questions may remain about best way to implement
- Trade off of between evidence in new areas, vs more on improving evidence on refining a program



# Over 300 million people reached by scale ups of programs found to be effective by J-PAL RCTs



# Policy Influence in North America

- Cognitive behavioral therapy
  - Becoming a Man Program
- Behavioral nudges
  - Letters to encourage tax filing and EITC take up
- Teaching at the right level
  - SAGA Innovations (formerly Match Tutoring)
- Summer youth employment
  - Chicago (with further testing in NYC, Boston, Philadelphia)



J-PAL

ABDUL LATIF JAMEEL POVERTY ACTION LAB

## For more reading and resources

Kremer and Glennerster, 2012, Chapter in  
Handbook of Health Economics

Bates and Glennerster, 2017, "The Generalizability  
Puzzle" Stanford Social Innovation Review

[www.povertyactionlab.org](http://www.povertyactionlab.org)