J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

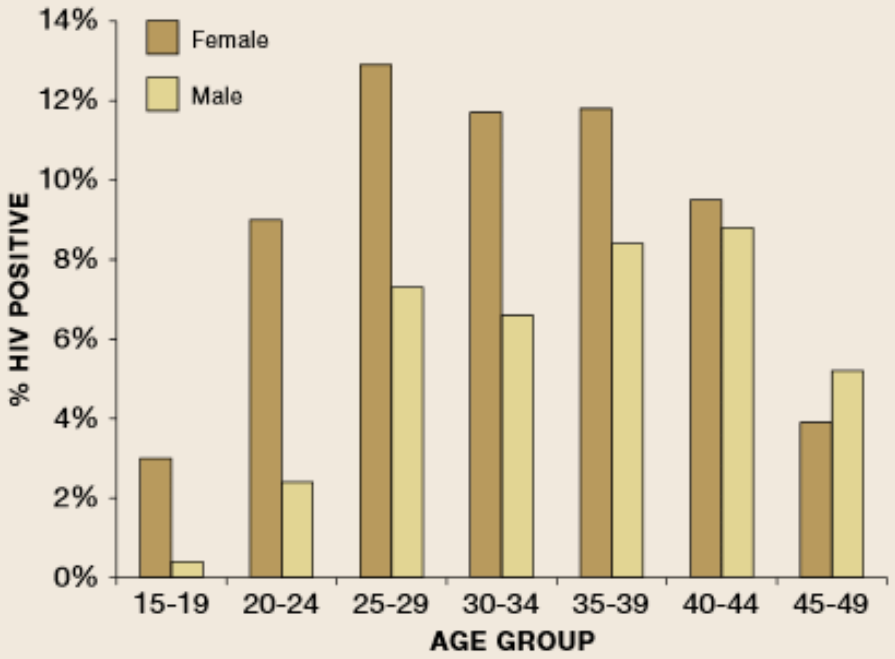
The Generalizability Puzzle

Mary Ann Bates
Deputy 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)

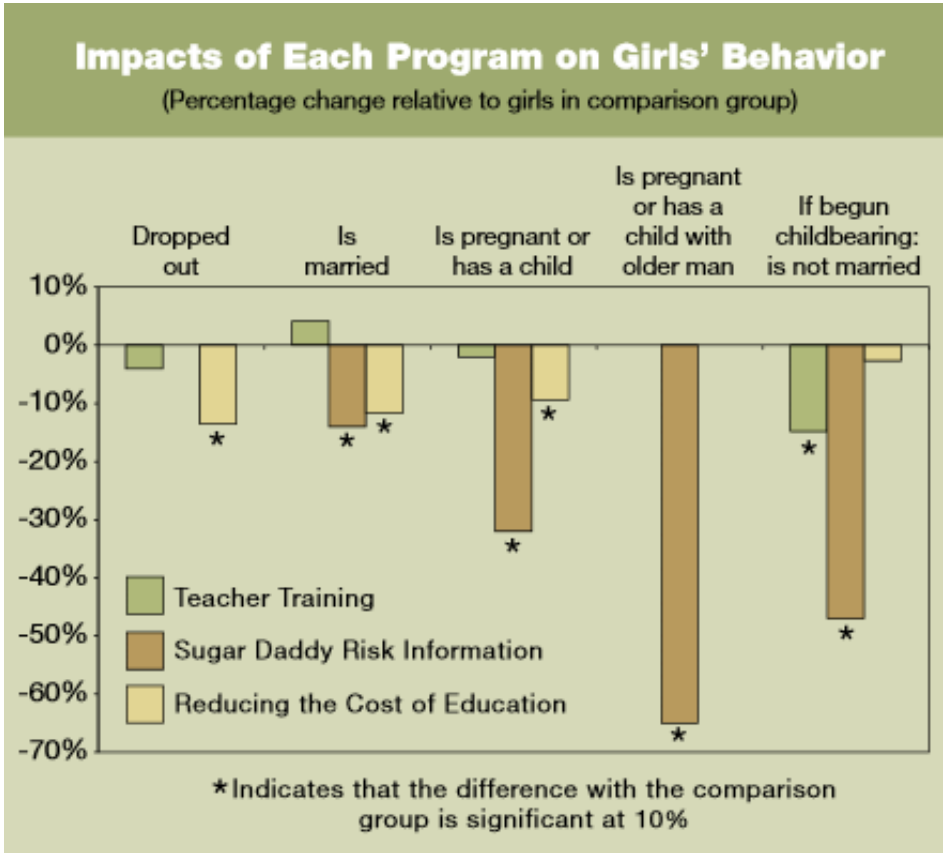
- Location: rural western Kenya
- 71 schools randomly selected from 328 schools
- Trained project staff visited the 8th grade classrooms
  - 10-minute video
  - Detailed stats on the rates of HIV by age and sex from nearby Kisumu
  - 30-minute discussion of cross-generational sex

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)
Dropped out
Is married
Is pregnant or has a child
Is pregnant or has a child with older man
If begun childbearing: is not married
10%
0%
-10%
-20%
-30%
-40%
-50%
-60%
-70%
Teacher Training
Sugar Daddy Risk Information
Reducing the Cost of Education
*Indicates that the difference with the comparison group is significant at 10%

Should Rwanda replicate the program?

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?

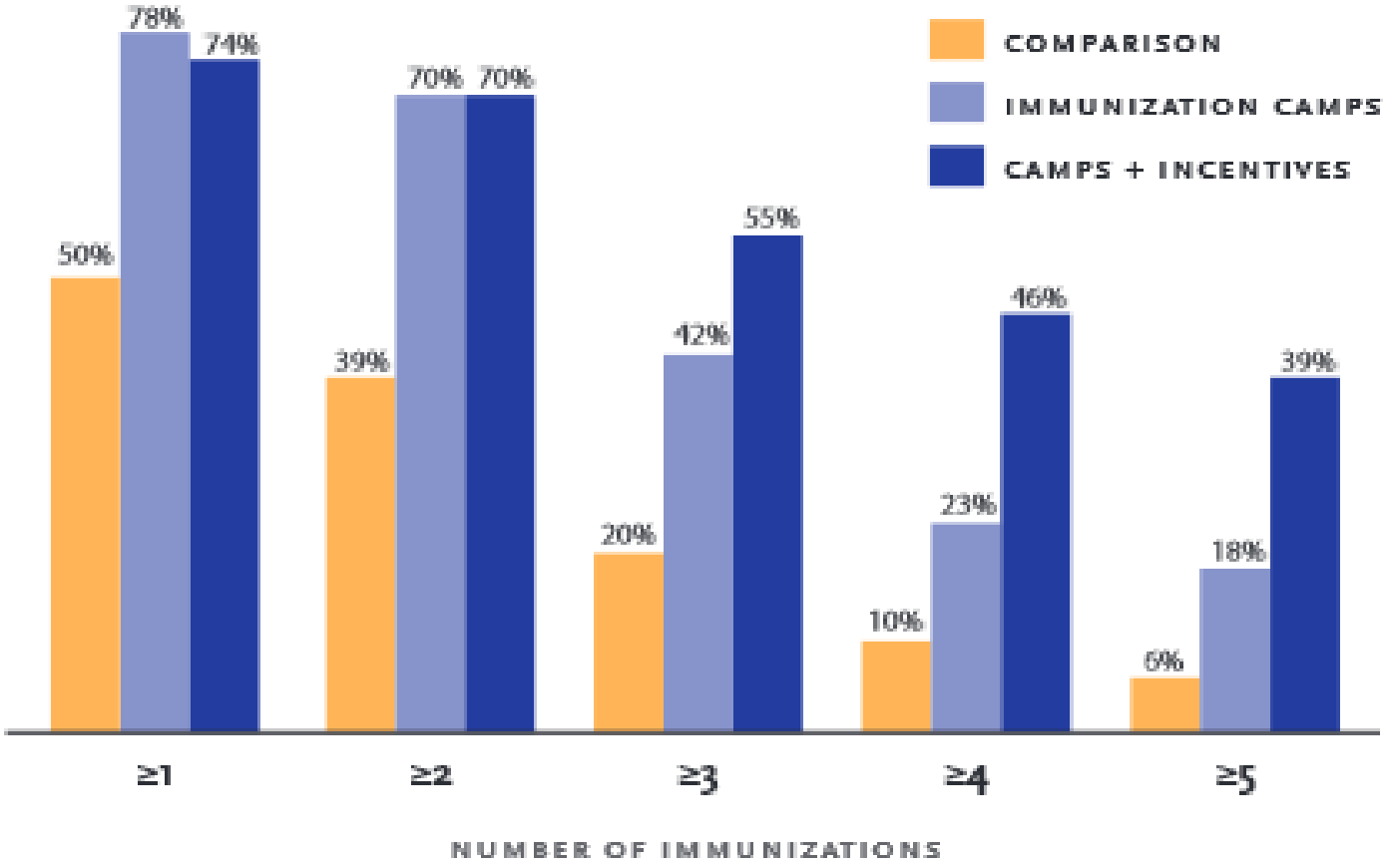
Structured Approach to Evidence in Policy

- 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



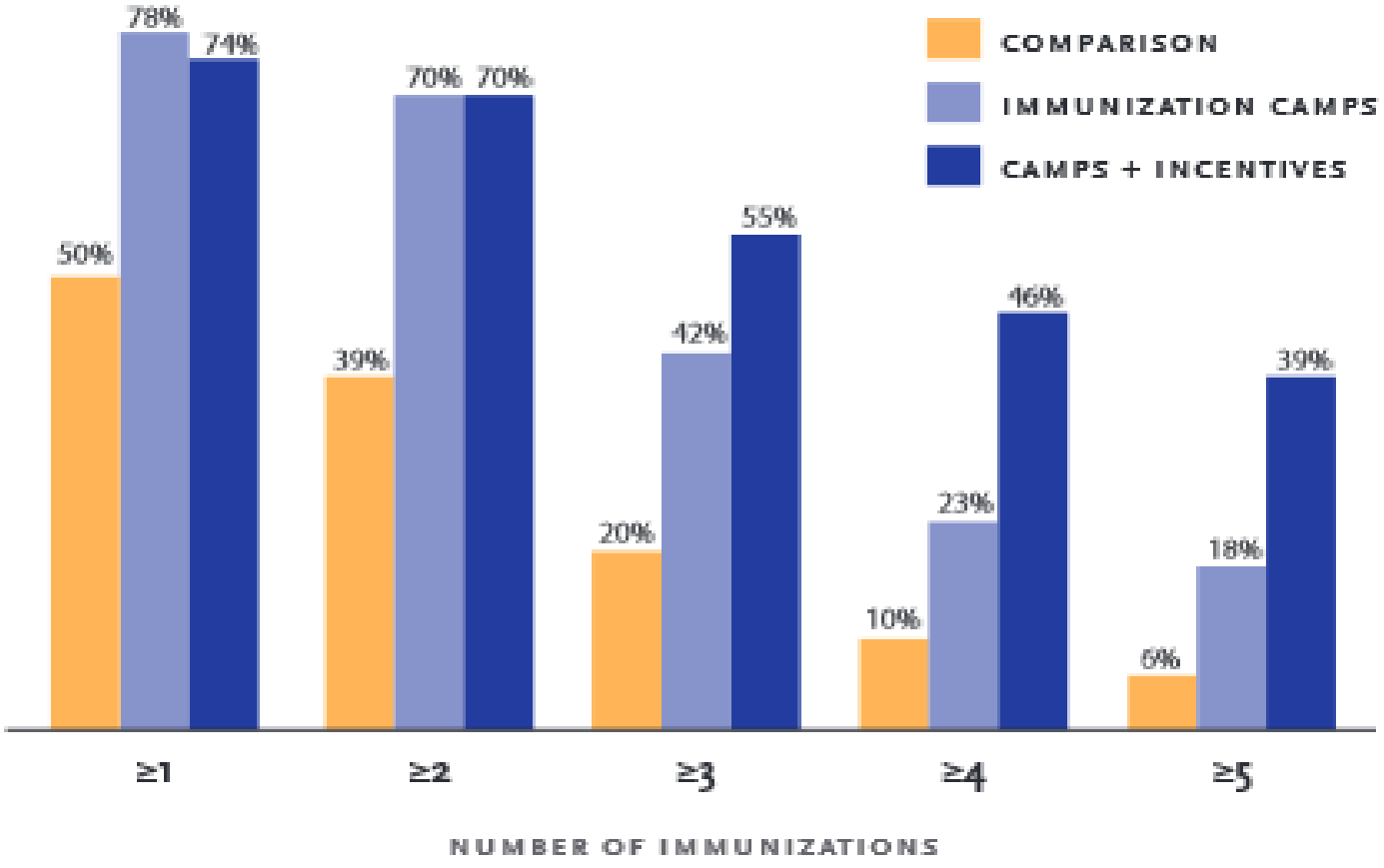
Viewing evidence in isolation

- 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

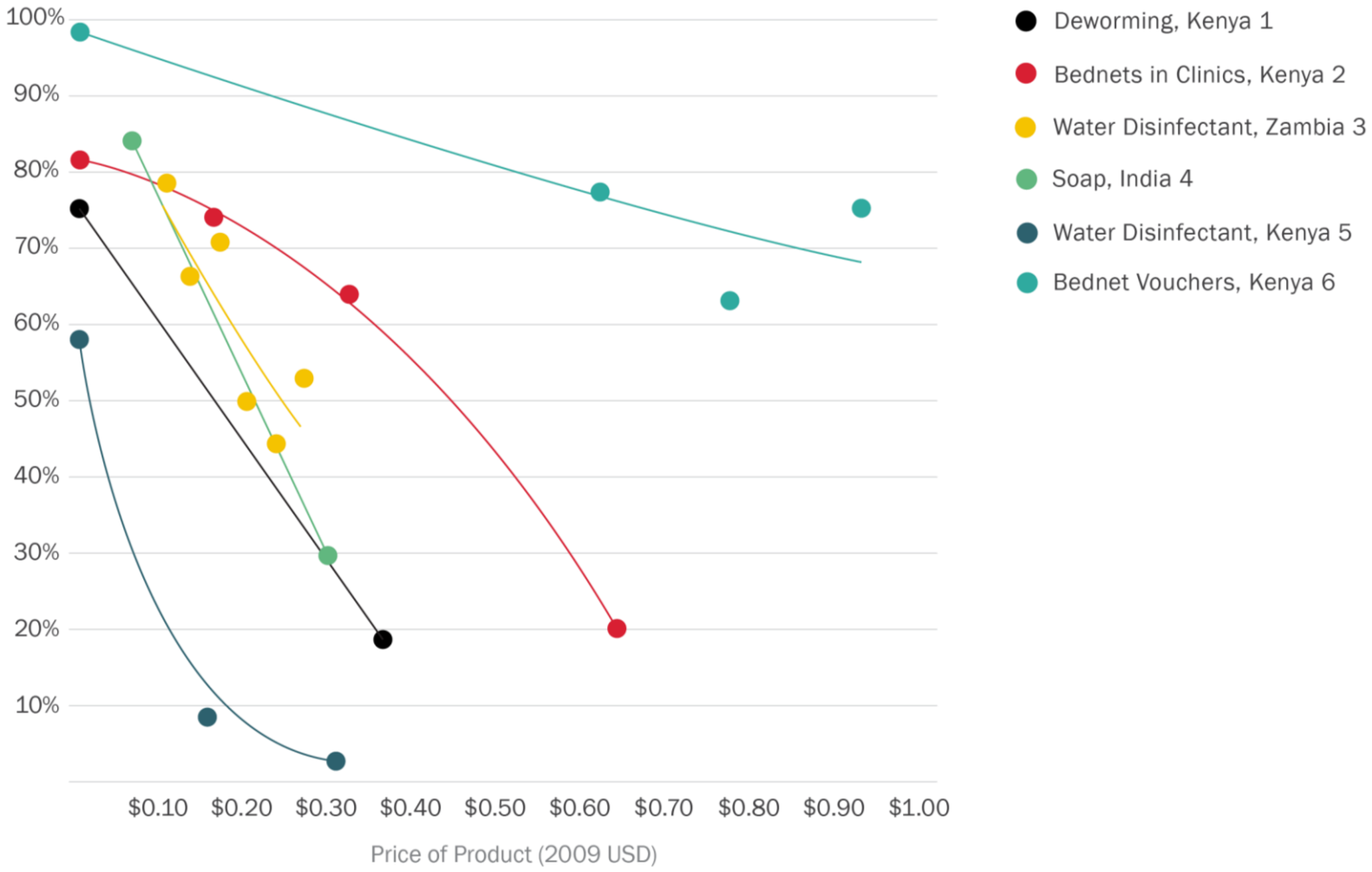


Generalizability Framework

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)

Generalizability Framework

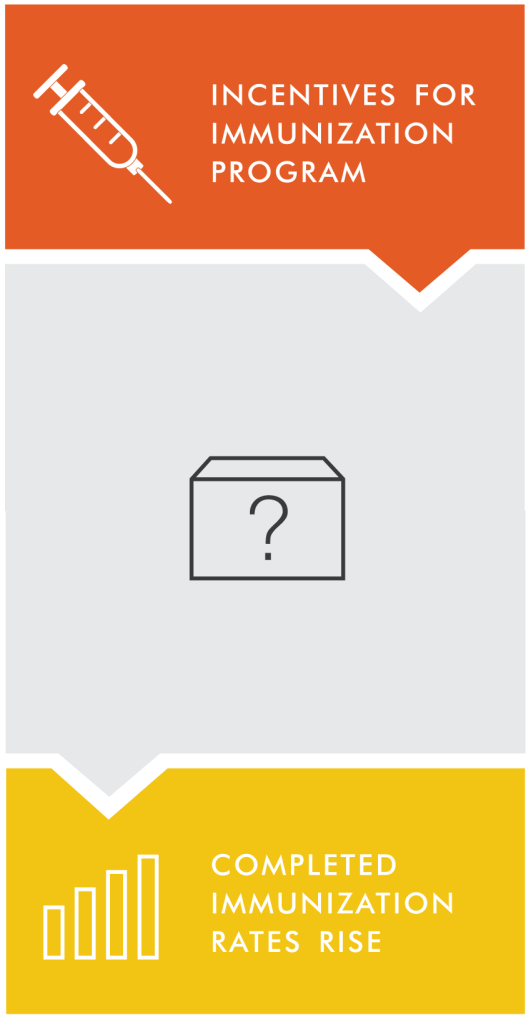
Is either country a good potential scale up location?

Immunization rates by antigen		
	Country 1	**Country 2**
DPT1	84	47
DPT3	74	41
Measles	67	41
Fully immunized	49	38

What local implementation issues would you consider?

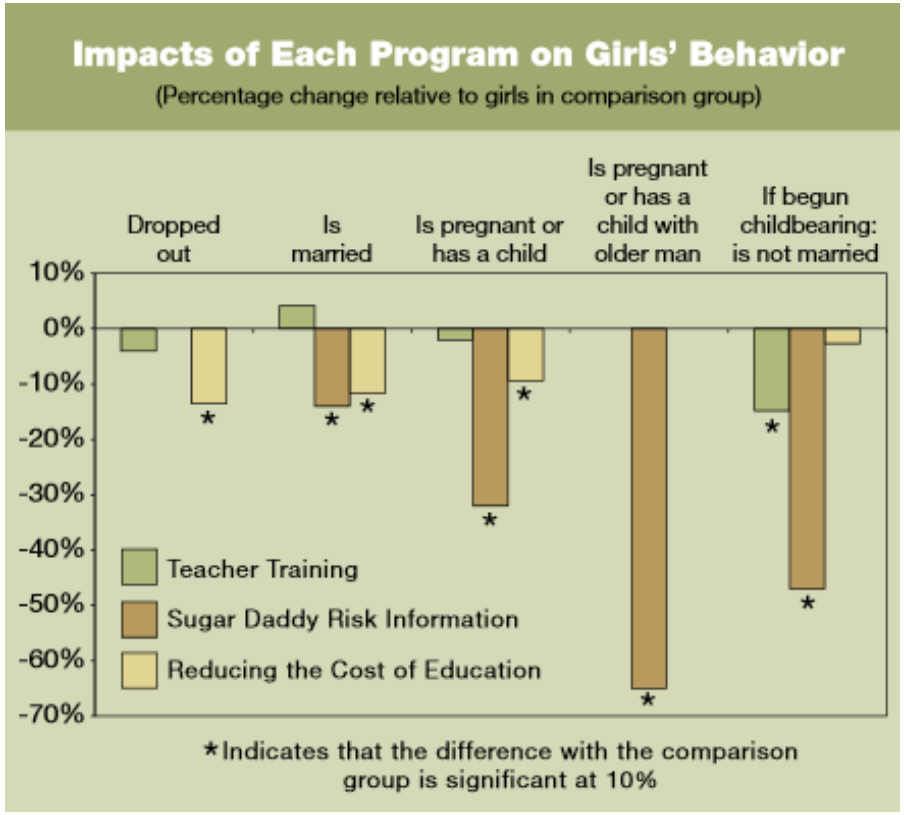
Local Evidence on Implementation

- 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


INCENTIVES FOR IMMUNIZATION PROGRAM
?
COMPLETED IMMUNIZATION RATES RISE


INCENTIVES FOR IMMUNIZATION PROGRAM
LOCAL CONDITIONS
1. Parents want to vaccinate
2. Parents can access clinic
3. Provider presence sufficient
4. Full immunization schedule is salient
GENERALIZED LESSONS ON BEHAVIOR
1. Minimal risk from overvaccination
2. Parents procrastinate or fail to persist
3. Parents are highly sensitive to price of preventative health
LOCAL IMPLEMENTATION
1. Incentives delivered to clinics
2. Incentives delivered to parents
COMPLETED IMMUNIZATION RATES RISE

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 peoples 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?**


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

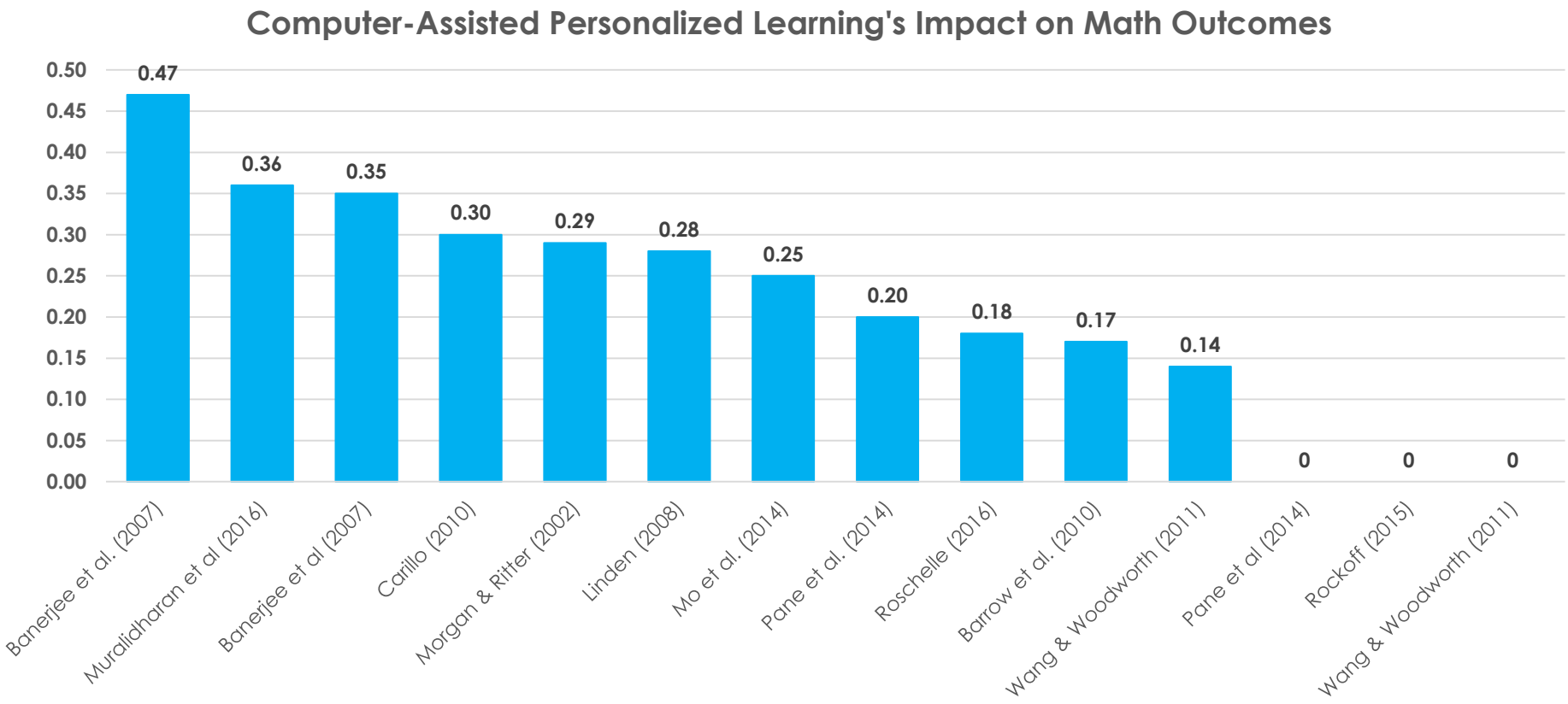
Teaching at the right level

Targeted Instruction Increases Learning

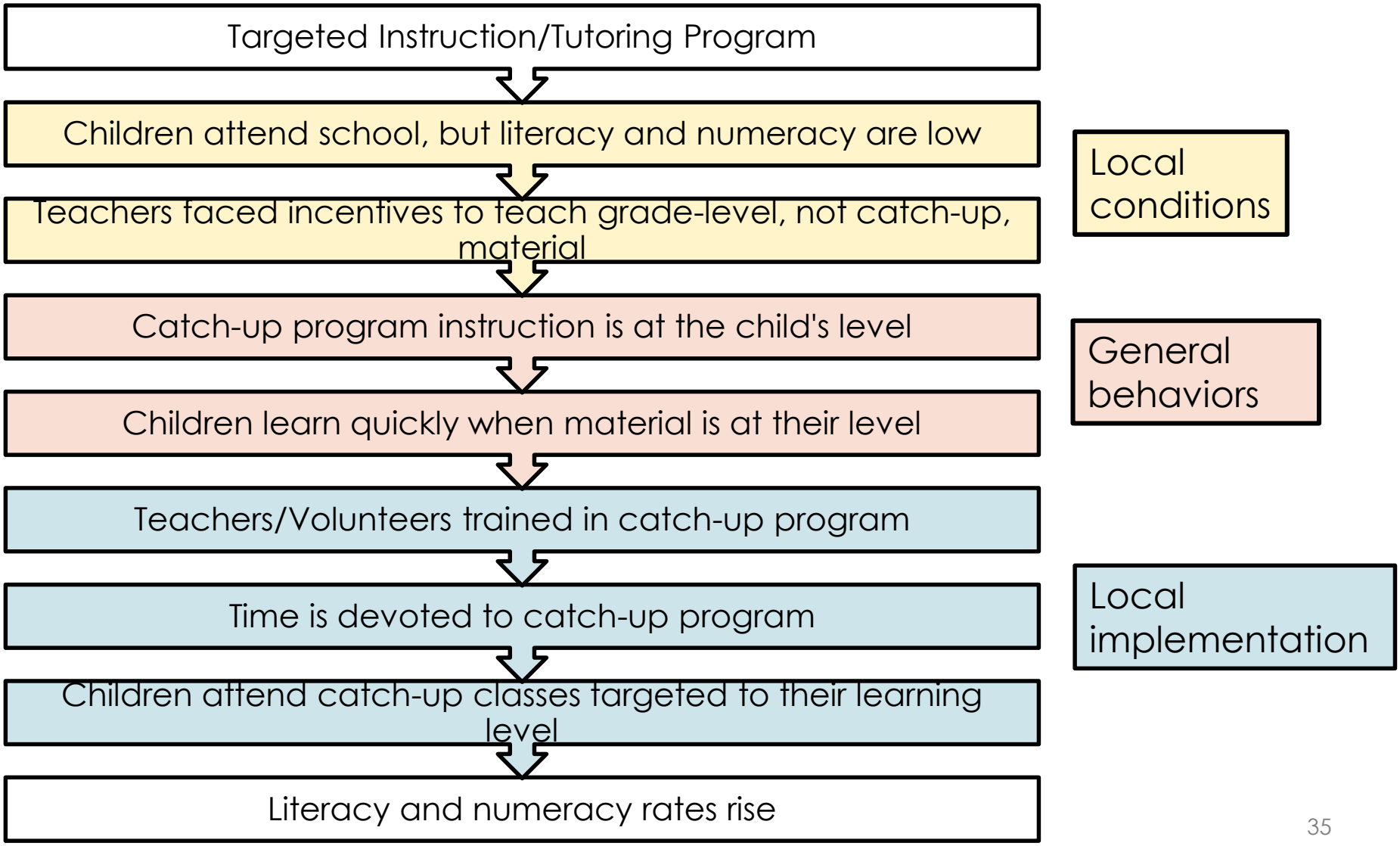
Series of studies shows targeted instruction can work in a variety of contexts:

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



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 failing to grasp basic skills.

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

Literature Reviews vs Meta-Analysis

- Meta analysis common in medicine, literature reviews common in economics
- Benefits of meta analysis:
  - explicit criteria for inclusion reduces risk bias in picking studies
  - Pooling results from many studies gives more power
  - Useful when testing identical programs
- Literature review rely on judgement and theory
  - Cross cutting lessons that are not from testing same program
  - Descriptive data can be used to support argument

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

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