The Generalizability Puzzle

J-PAL

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

Example:
HIV Relative Risk Information Campaign in Kenya

A "Relative Risk Information Campaign" led to a significant reduction in unwanted teenage childbearing with older partners.

Dupas 2011

Photo: Aude Guerricci, for evaluation "HIV/AIDS Prevention Through Relative Risk Information for Teenage Girls in Kenya"

Randomized evaluation: Relative Risk Information

- 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

Men's HIV Rates by Age
in Kisumu, Kenya, 2001

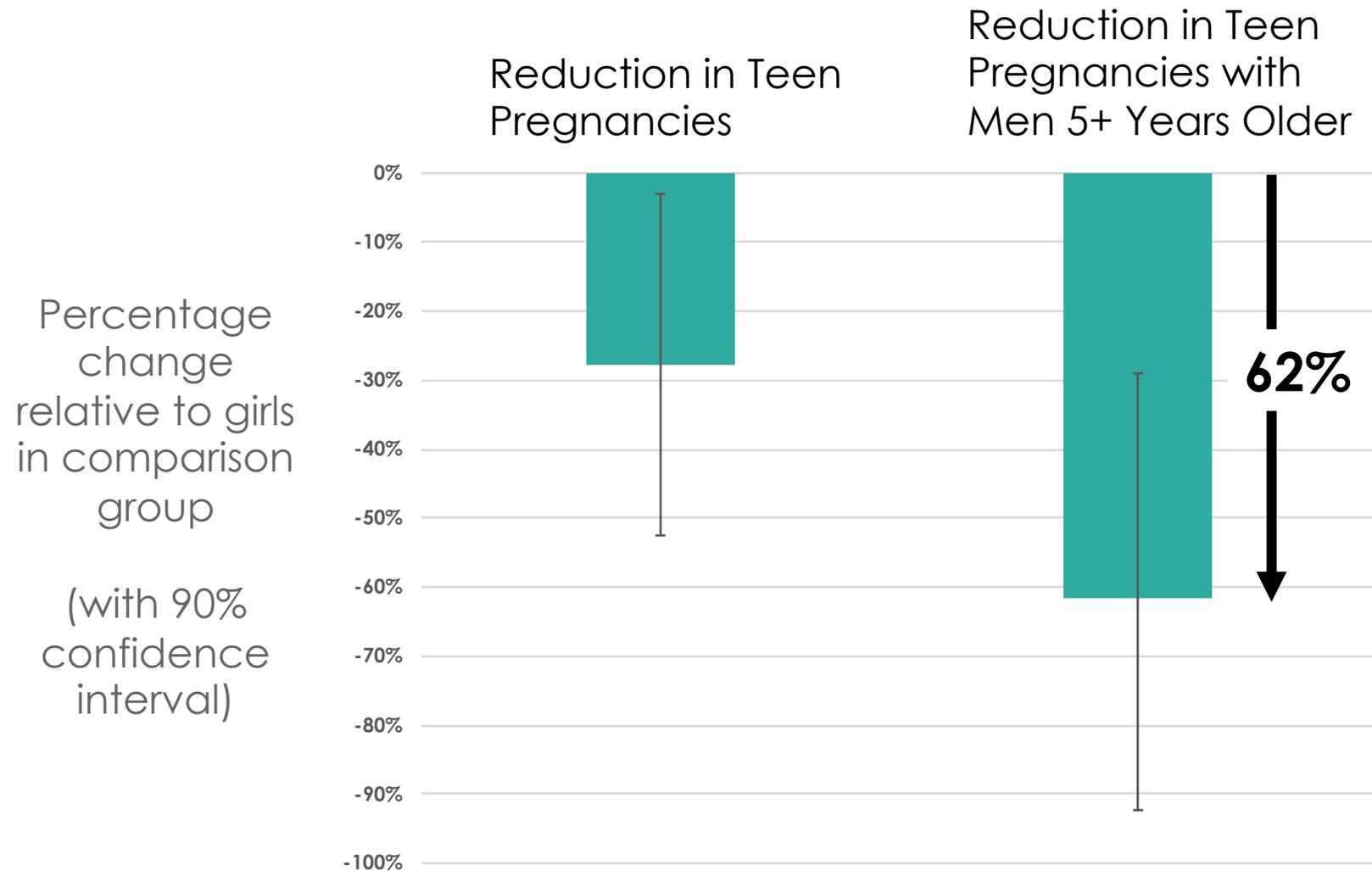| Age	HIV prevalence rate
Age 15-19	4%
Age 20-24	13%
Age 25-29	28%
Age 30-39	32%

Republic of Kenya Ministrry of Health, "AIDS in Kenya," 2001.

Example:
HIV Relative Risk Information Campaign in Kenya

Photo of a relative risk education session in Botswana. young1ove.org

HIV Relative Risk Information Campaign Reduced Teen Pregnancies in Kenya

# Should Rwanda replicate the program?

A. Yes

B. No

Imagine that you are the head of a school district

Students are falling behind grade level in reading and math.

How can you know whether a tutoring intervention that worked somewhere else (for example, Saga Innovations) will also work for you?

Imagine that you are the head of a school district

- Students are falling behind grade level in reading and math

- How can you know whether something that worked somewhere else will also work for you?

- Tempting to focus on geography. (Was it tested nearby? How near is near enough?)

- Nowhere is identical to "here."
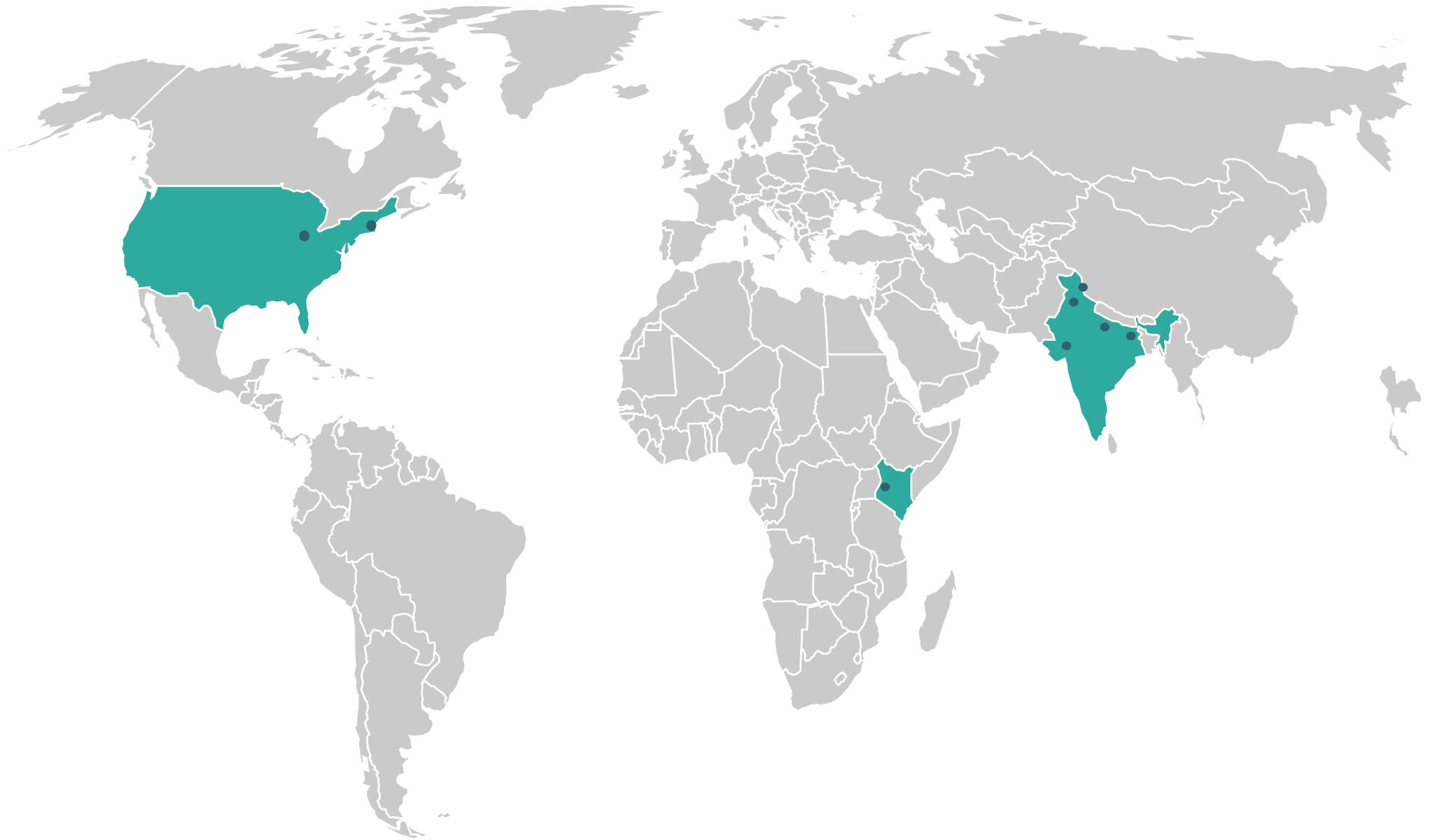
The challenge

- 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

Four misguided questions

- 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?

The generalizability puzzle framework

- Instead of focusing on place, focus on people
    - Key conditions and general lessons about behavior

- 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)

Applying the
# Generalizability Puzzle Framework

Three examples

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

A parent receives a kilogram of lentils at a vaccination clinic in Rajasthan, India.

NUMBER OF IMMUNIZATIONS RECEIVED BY CHILDREN AGED 1-3 YEARS

NUMBER OF IMMUNIZATIONS

Comparison — Immunization Camps — Camps + Incentives
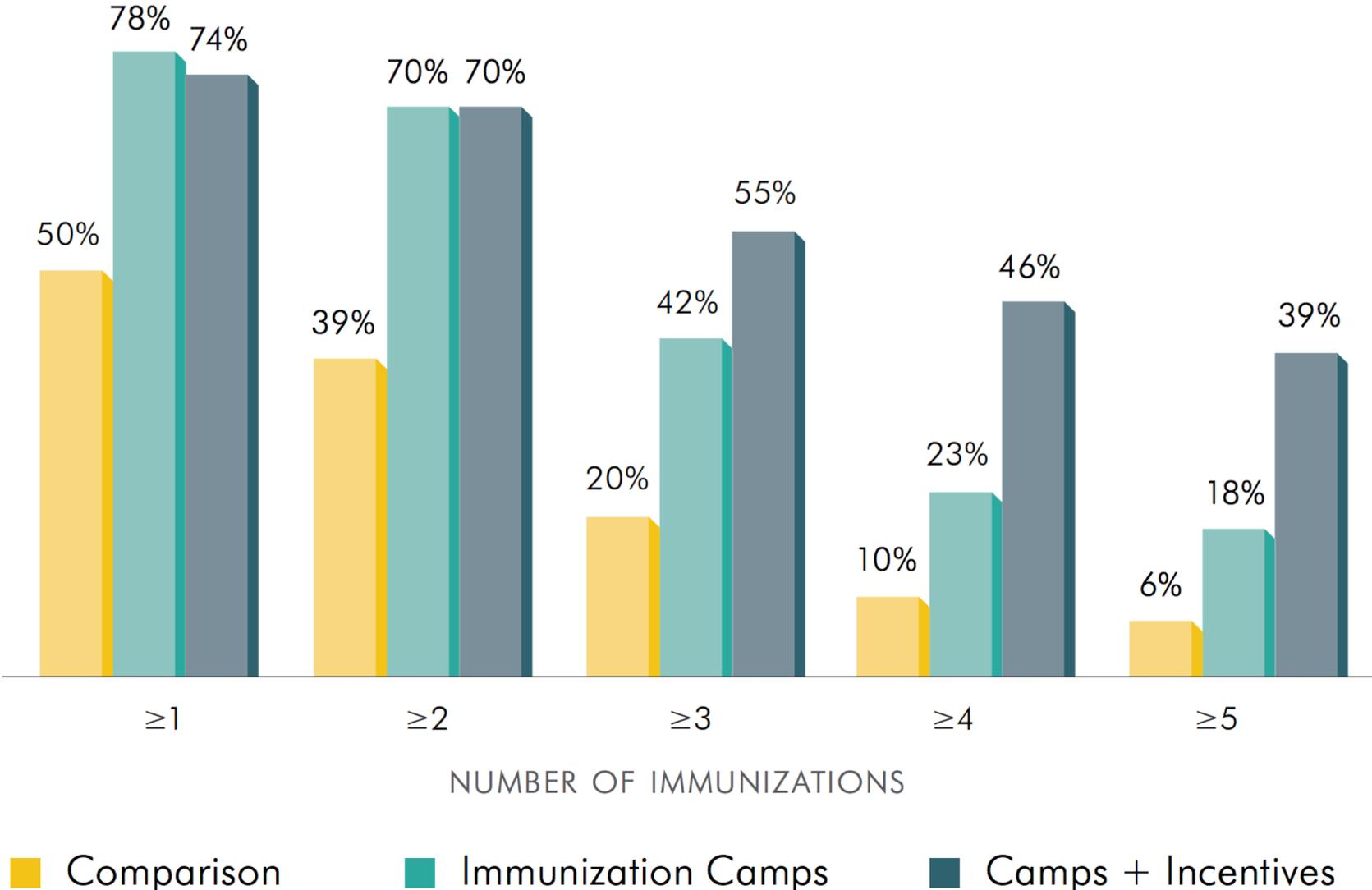
Viewing evidence in isolation

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

INCENTIVES FOR IMMUNIZATION PROGRAM

?

COMPLETED IMMUNIZATION RATES RISE

Generalizability Framework

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. Parents procrastinate or fail to persist
2. 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

NUMBER OF IMMUNIZATIONS RECEIVED BY CHILDREN AGED 1-3 YEARS

NUMBER OF IMMUNIZATIONS

Comparison ■ Immunization Camps ■ Camps + Incentives

Generalizability Framework

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. Parents procrastinate or fail to persist
2. 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
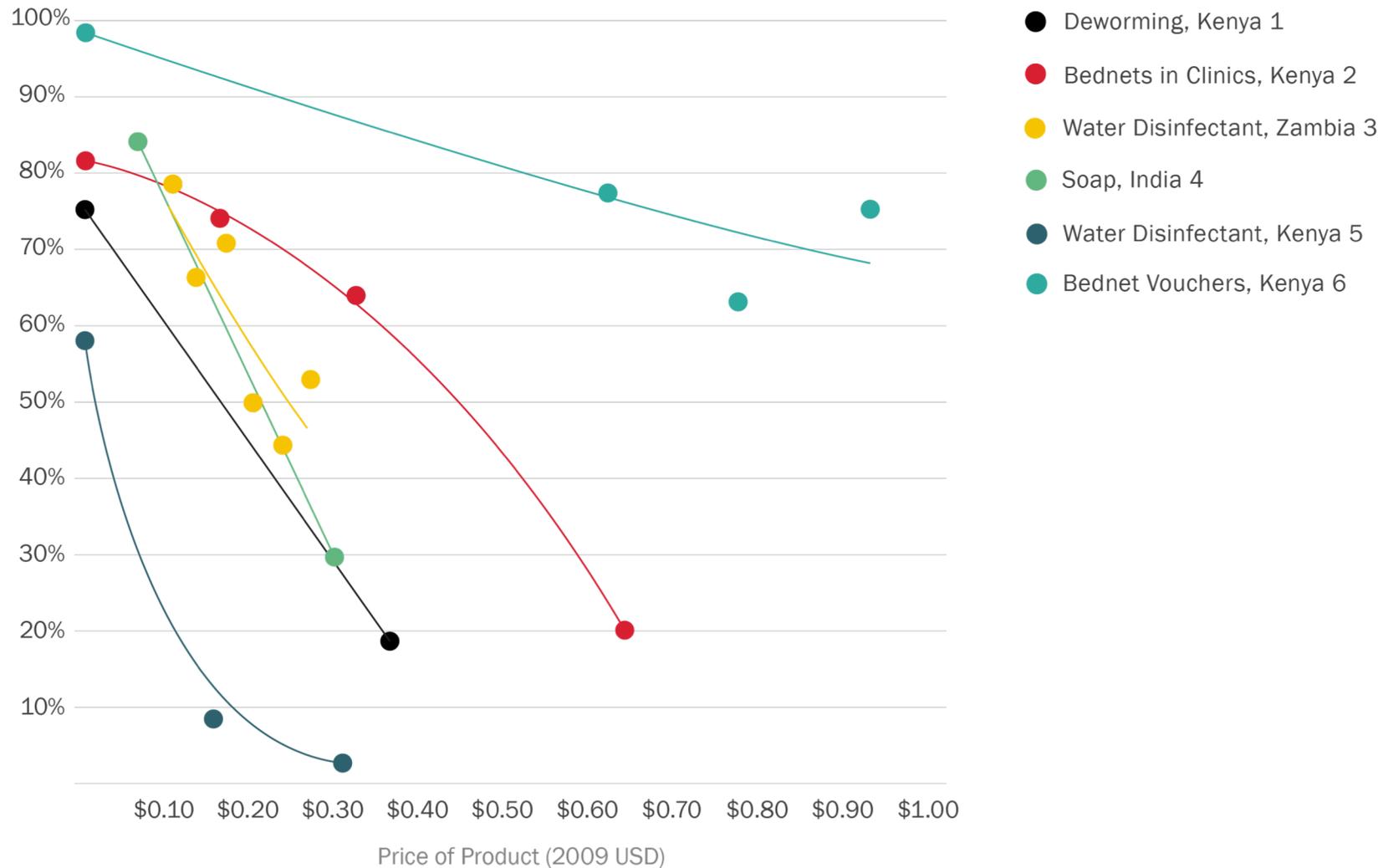
Evidence on present bias

- 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 Preventive Health

- ● Deworming, Kenya 1
- ● Bednets in Clinics, Kenya 2
- ● Water Disinfectant, Zambia 3
- ● Soap, India 4
- ● Water Disinfectant, Kenya 5
- ● Bednet Vouchers, Kenya 6

Price of Product (2009 USD)

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)
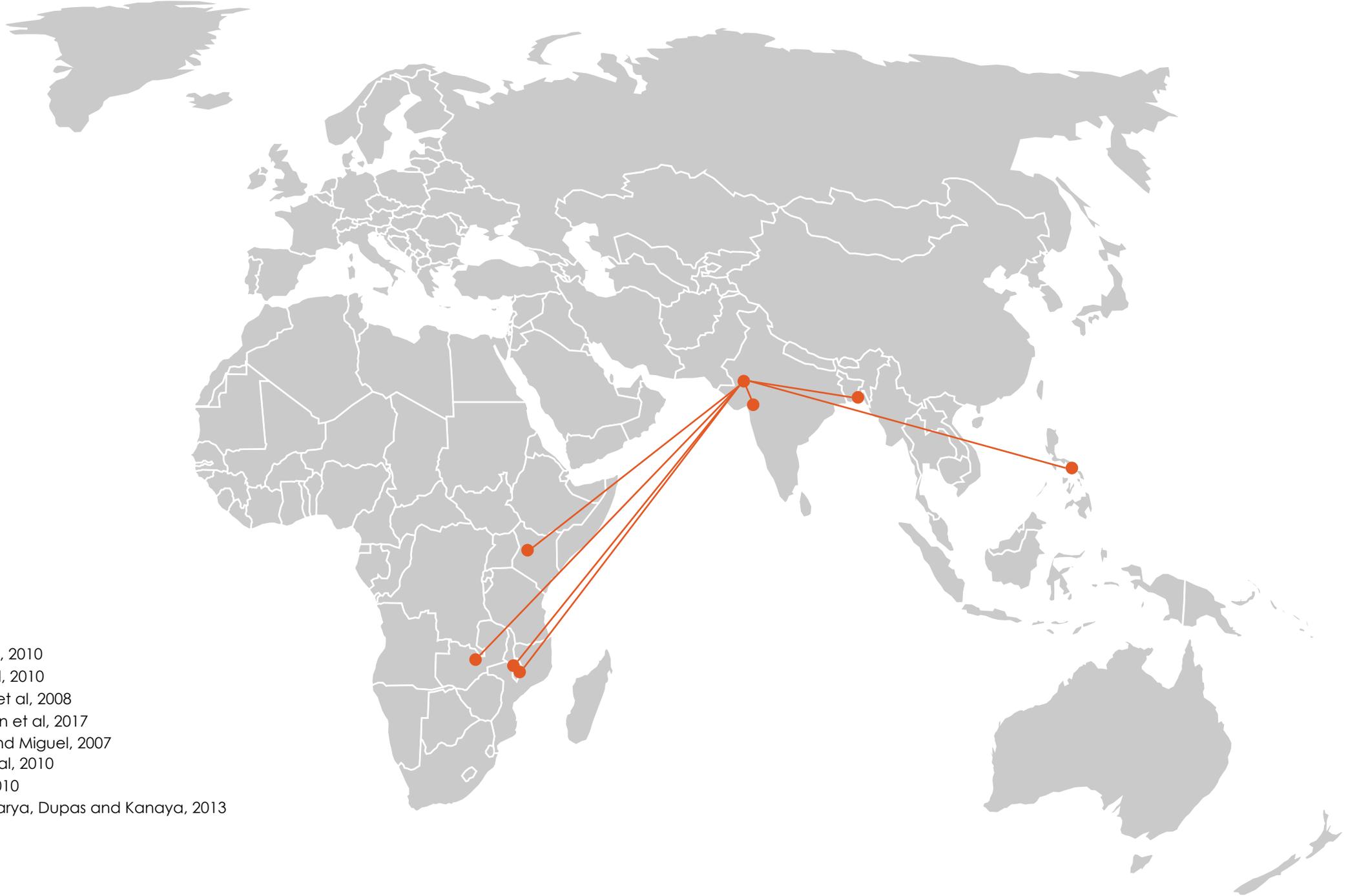
Sources:
Gine et al, 2010
Baird et al, 2010
Thornton et al, 2008
Buchmann et al, 2017
Kremer and Miguel, 2007
Ashraf et al, 2010
Spears, 2010
Bhattacharya, Dupas and Kanaya, 2013

Generalizability Framework

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

1. Parents procrastinate or fail to persist
2. Parents are highly sensitive to price of preventative health

GENERALIZED LESSONS ON BEHAVIOR

LOCAL IMPLEMENTATION

1. Incentives delivered to clinics
2. Incentives delivered to parents

COMPLETED IMMUNIZATION RATES RISE

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

Which country is a good potential scale up location?

A. Country 1

B. Country 2

C. Neither

D. Both

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

Left diagram:

INCENTIVES FOR IMMUNIZATION PROGRAM

?

COMPLETED IMMUNIZATION RATES RISE

Right diagram:

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. Parents procrastinate or fail to persist
2. 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

Applying the Generalizability Puzzle Framework

Three examples

1. Scaling immunization incentives
2. Relative risk education program
3. Teaching at the right level

Would the HIV Relative Risk Information Campaign work in Rwanda?

Results of the study in Kenya:

Percentage change relative to girls in comparison group

(with 90% confidence interval)

Reduction in Teen Pregnancies

Reduction in Teen Pregnancies with Men 5+ Years Older

62%
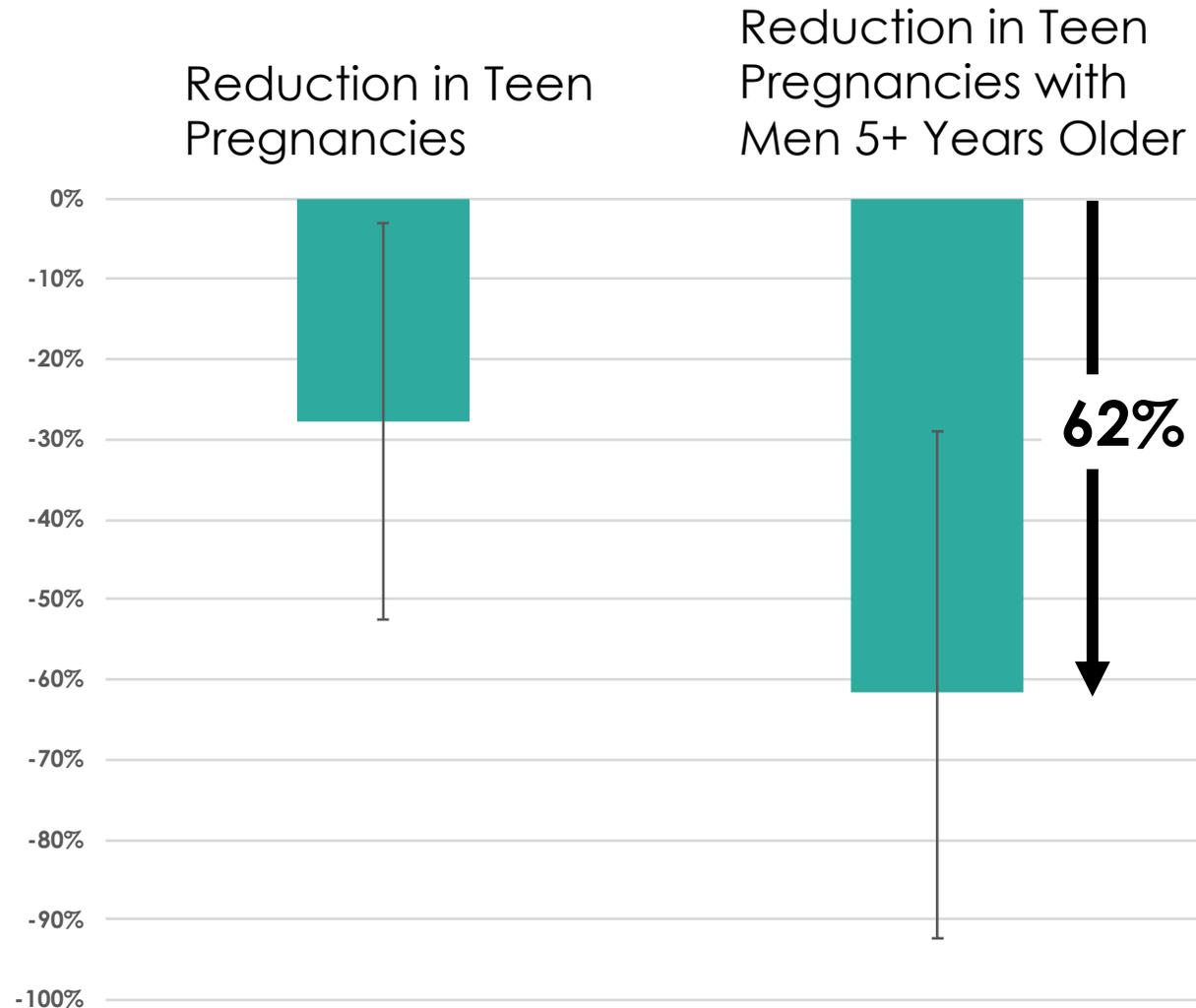
Generalizability Framework: HIV Relative Risk Program

- What informs girls' choices of sexual partners?

  – 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?

INFORMATION ON RELATIVE RISK OF HIV BY AGE

LOCAL CONDITIONS

GENERALIZED LESSONS ON BEHAVIOR

LOCAL IMPLEMENTATION

RISKY SEX WITH OLDER MEN REDUCES, LESS RISK OF HIV

Local descriptive data (collected in a few weeks)

- In Rwanda, men ages 25-29 have an HIV rate of 1.7 percent

- 98% of students overestimated the rate of HIV among men ages 25-29

- **In which direction would a risk awareness program change the Rwandan students' prior beliefs?**

Should Rwanda replicate the program?

A. Yes

B. No

Young 1ove

EVIDENCE-BASED

We comb academic papers for relevance to our mission, model and niche, and sufficiently rigorous evidence. Our team sifts through jargon, equations and other arcane details tucked away in these papers, and then pulls out and codifies the theory of change behind the proven social impact. We then solicit feedback from experts in the field and put pen to paper, creating evidence-based curriculum. The final step in the translation process is personifying our curricula via trained peer facilitators who deliver our evidence-based messages in partnership with government in schools throughout Eastern and Southern Africa, continuing to learn as we scale.

FOR YOUTH BY YOUTH

We believe proven health and education needs to be taught *by youth for youth*. You can't send old officials to teach kids about sex and stigmatized topics. It doesn't work. It's not relatable. It's boring. It doesn't have impact. We make sure the messages we pick up - the ones research has shown work -- also get delivered in a way our target audience deserves: by youth for youth.

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

GENERALIZED LESSONS ON BEHAVIOR

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

LOCAL IMPLEMENTATION

1. Relative risk information can be conveyed effectively to girls

RISKY SEX WITH OLDER MEN REDUCES, LESS RISK OF HIV

Applying the Generalizability Puzzle Framework

Three examples

1. Scaling immunization incentives
2. Relative risk education program
3. Teaching at the right level

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 x 4

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

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

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

$$1000 = x^3$$

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

$$\boxed{x = 10}$$

6x

$$r = 5_m$$
$$h = 60_m$$

O.R

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

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

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

$$8x+12 + 6x+1$$

$$1^3x +$$
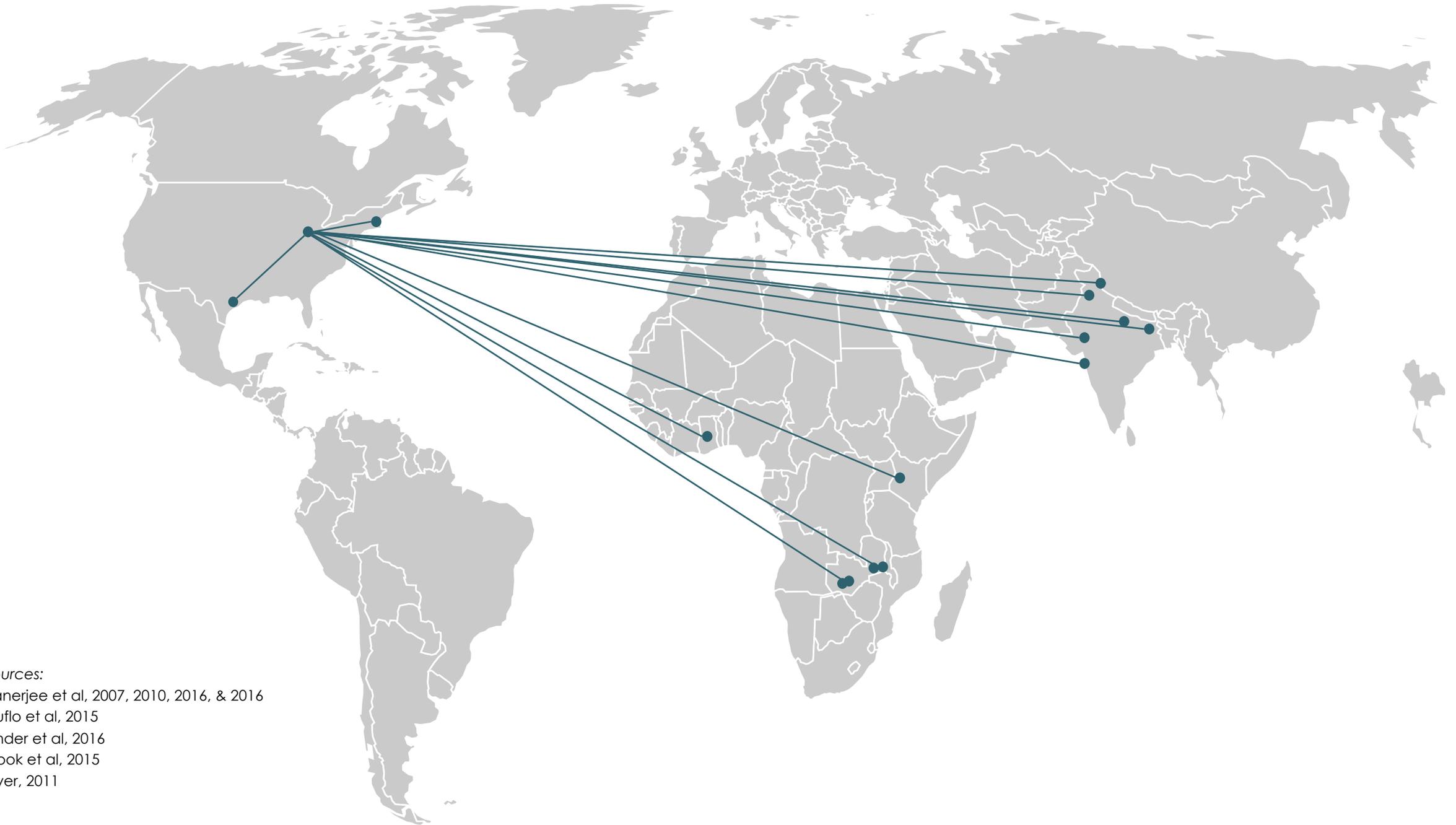
Sources:
Banerjee et al, 2007, 2010, 2016, & 2016
Duflo et al, 2015
Ander et al, 2016
Cook et al, 2015
Fryer, 2011

Targeted instruction increases learning

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

1. Balsakhi Assistant Programme in India (Duflo et al 2007)
2. Read India Programme (Banerjee et al 2007)
3. Computer Assisted Learning (Duflo et al 2007)
4. India Reading Camps (Banerjee et al 2010)
5. Extra Teacher Programme in Kenya (Duflo et al 2011)
6. Haryana Learning Enhancement Programme (Berry et al 2013)
7. TCAI Programme in Ghana (Duflo and Kiessel 2012)
8. Match Education and Youth Guidance in Chicago (Cook et al 2014)
9. Match Education in Boston (Cook et al 2015)
10. Saga Innovations in Chicago (Davis et al 2017)

Personalized learning is highly effective across studies

Computer-Assisted Personalized Learning's Impact on Math Outcomes
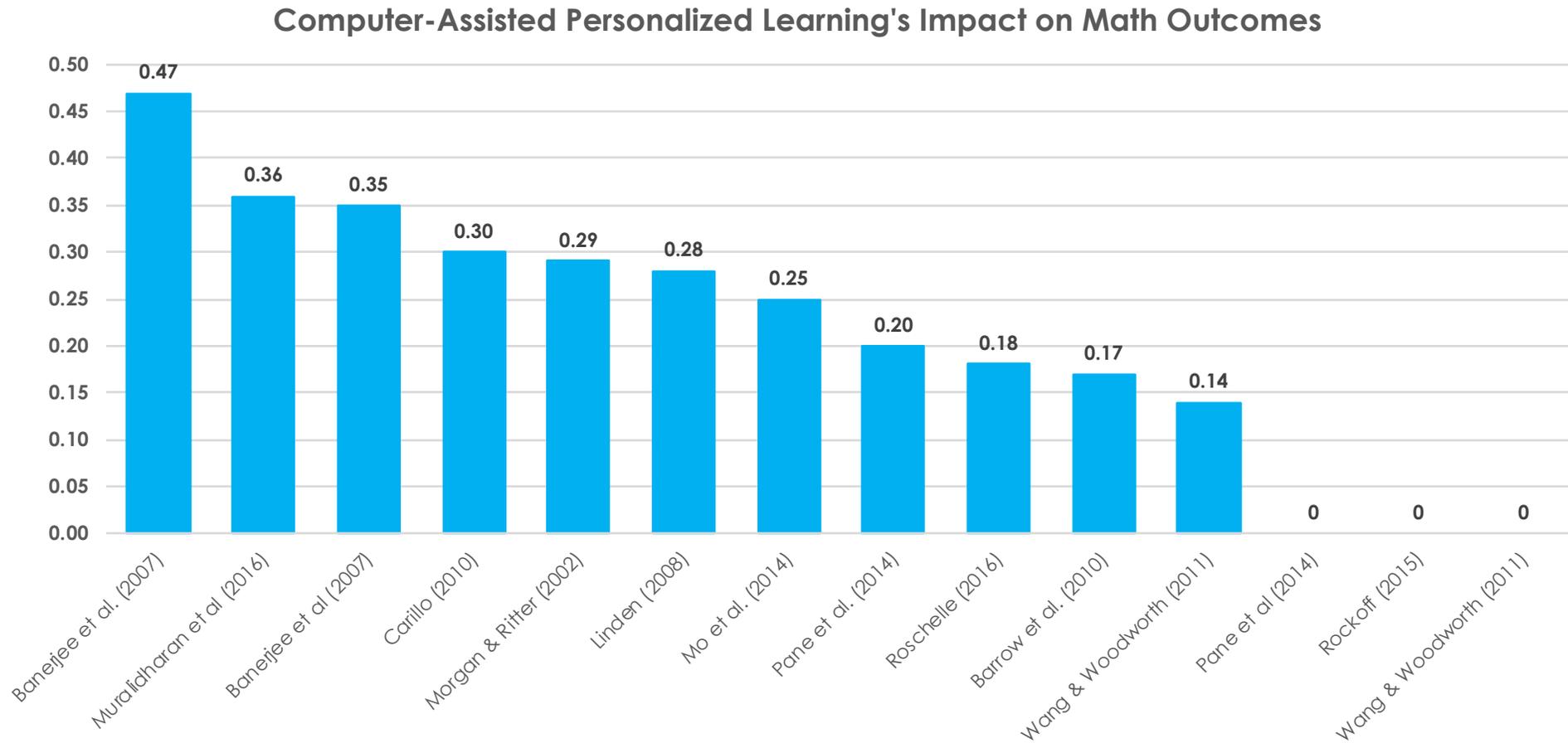
For details see J-PAL North America's review:
Education Technology: An Evidence-Based Review by Escueta et al.

TARGETED INSTRUCTION / TUTORING PROGRAM

LOCAL CONDITIONS

1. Children attend school, but literacy and numeracy rates are low
2. Teachers face incentives to teach grade-level material, not catch-up material

GENERALIZED LESSONS ON BEHAVIOR

1. Catch-up program instruction is at the student's level
2. Students learn when material is at their level

LOCAL IMPLEMENTATION

1. Teachers/volunteers trained in catch-up program
2. Time is devoted to catch-up program
3. Students attend catch-up classes targeted to their learning level

LITERACY AND NUMERACY RATES RISE

Many Implementation Models

Who should lead the program?	Where should the program be held?	When should the program 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		

TEACHING AT THE RIGHT LEVEL

Zambia

This case study shares J-PAL Africa and Pratham's experience of working with the Ministry of General Education in Zambia to build the Catch Up programme, where global evidence, local adaptation, and iterative testing were used to improve learning outcomes.

Scenario

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 Saga's program for teaching at the right level, and want to explore whether it makes sense for you to implement this program in your schools.

- 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?

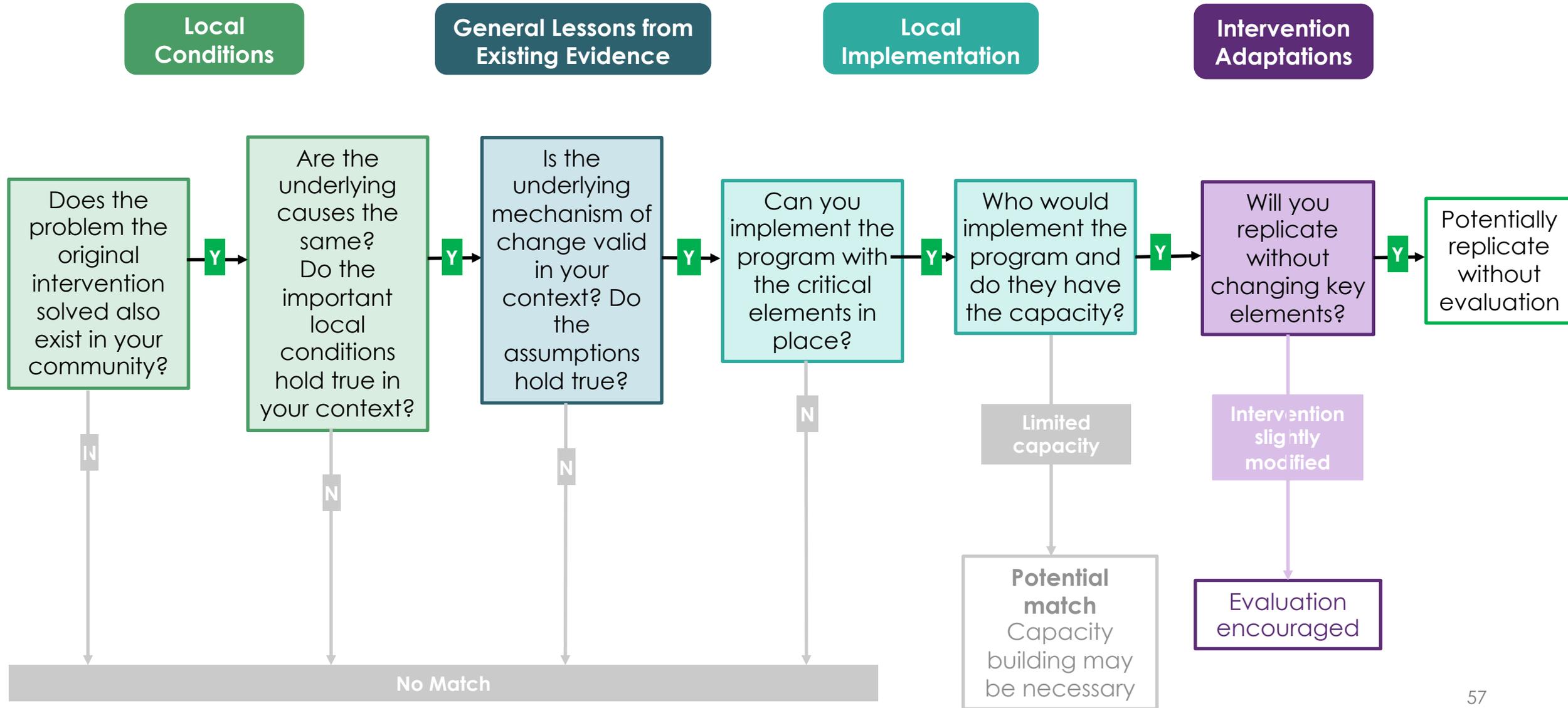# Indicators and Data for Decision-Making

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?

Local Conditions	Indicators	Data Sources
1. Students (at least some) are **performing below grade level**		
2. There are **varying levels of student achievement** in classrooms, with some students performing above, at, and below grade level.		
3. Teachers **teach at one level for all students** in their classroom, 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?

A Decision Tree

Local Conditions

General Lessons from Existing Evidence

Local Implementation

Intervention Adaptations

Does the problem the original intervention solved also exist in your community?

→ **Y** → Are the underlying causes the same? Do the important local conditions hold true in your context?

→ **Y** → Is the underlying mechanism of change valid in your context? Do the assumptions hold true?

→ **Y** → Can you implement the program with the critical elements in place?

→ **Y** → Who would implement the program and do they have the capacity?

→ **Y** → Will you replicate without changing key elements?

→ **Y** → Potentially replicate without evaluation

N (from first box) → No Match

N (from second box) → No Match

N (from third box) → No Match

N (from fourth box) → No Match

Limited capacity → **Potential match** Capacity building may be necessary

Intervention slightly modified → Evaluation encouraged

No Match

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 between evidence in new areas, versus more on improving evidence on refining a program
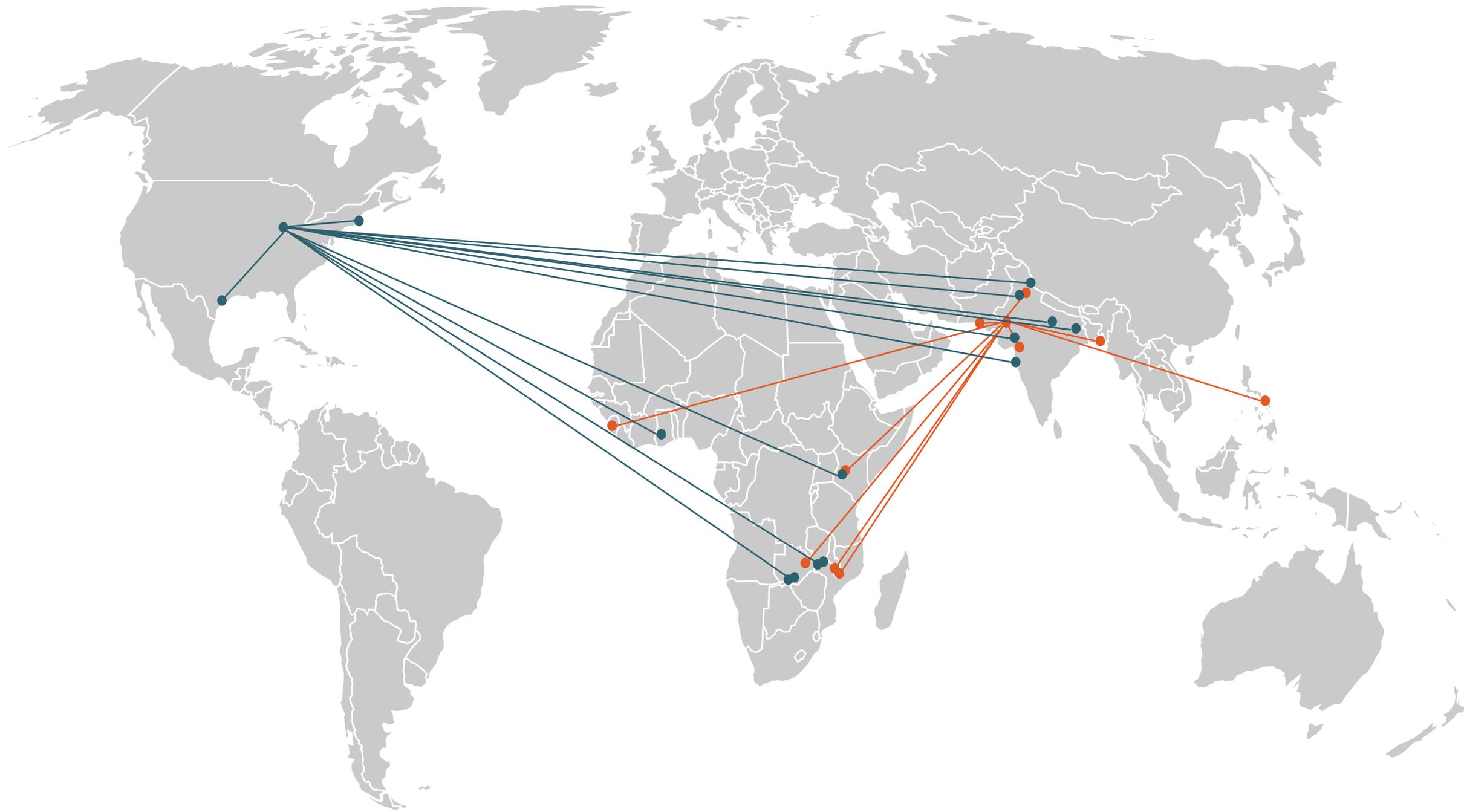
Over 400 million people reached by scale ups of programs found to be effective by J-PAL RCTs

Evidence to Policy

Evidence to Policy

Evidence from randomized evaluations is changing how we understand and address problems related to poverty. Policymakers, practitioners, and funders worldwide are increasingly applying this learning to social policies and programs.

Over 400 million people have been reached by programs that were scaled up after being evaluated by J-PAL affiliated researchers. Many more have benefitted from the several broader ways evidence can inform policy, outlined below.

Continue reading ⌄

Pathways to Policy Change

Below, you will find six pathways through which evidence can have an impact on policy and case studies that illustrate partnerships leading to policy impact.

Shifting global thinking

Knowledge generated by randomized evaluations has fundamentally shaped our understanding of many social policies.

Example case studies:

Free bednets to fight malaria

More...

Institutionalizing evidence use

Many organizations, including governments and large NGOs, have institutionalized processes for rigorously evaluating innovations and incorporating evidence into decision-making.

Example case studies:

A government innovation lab to improve education

More...

Applying research insights

Lessons from randomized evaluations have informed the design of programs.

Adapting and scaling a program

Programs originally evaluated in one context have been adapted and scaled in others.

Pathways to Policy Change

Below, you will find six pathways through which evidence can have an impact on policy and case studies that illustrate partnerships leading to policy impact.

Shifting global thinking

Knowledge generated by randomized evaluations has fundamentally shaped our understanding of many social policies.

Example case studies:

Free bednets to fight malaria

More...

Institutionalizing evidence use

Many organizations, including governments and large NGOs, have institutionalized processes for rigorously evaluating innovations and incorporating evidence into decision-making.

Example case studies:

A government innovation lab to improve education

More...

Applying research insights

Lessons from randomized evaluations have informed the design of programs.

Example case studies:

Fund flow reform for social program delivery

More...

Adapting and scaling a program

Programs originally evaluated in one context have been adapted and scaled in others.

Example case studies:

Teaching at the right level to improve learning

Targeting the ultra-poor to improve livelihoods

More...

Scaling up an evaluated pilot

Innovate, test, scale: Replicating and expanding a successful evaluated pilot to similar contexts.

Example case studies:

Targeted information to improve social assistance

Simplified reminders to increase take-up of tax credits

More...

Scaling back an evaluated program

Innovate, test, reassess: Partners have scaled down, redesigned, or decided to not move forward with programs that were evaluated and found to be ineffective.

Example case studies:

Limits of technological solutions to provider monitoring

Unintended effects of anonymous resumes

Further reading and resources

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

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

- J-PAL Evidence to Policy page: http://www.povertyactionlab.org/evidence-to-policy/

References, Reuse, and Citation