



WORLD BANK GROUP
Social Protection & Labor

Introduction to Impact Evaluation

Patrick Premand, Laura Rawlings
SSN Course, May 2016

This material builds on Gertler, P. J.; Martinez, S., Premand, P., Rawlings, L. B. and Christel M. J. Vermeersch, 2010, Impact Evaluation in Practice: Ancillary Material, The World Bank, Washington DC (www.worldbank.org/ieinpractice). The content of this presentation reflects the views of the authors and not necessarily those of the World Bank.

Outline

Monitoring and Evaluation

Evaluation in SSN Projects

Results Chains

The Main Concepts of Impact Evaluation

Choosing the best IE design for your project

Ethical Considerations

Randomization as an operational tool

Evidence Frontiers on Safety Nets?



WORLD BANK GROUP
Social Protection & Labor

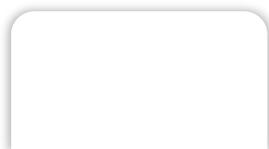
Monitoring and Evaluation

Monitoring



A continuous process of collecting and analyzing information,

- To **compare** how well a project, program or policy is performing against expected results, and
 - To **inform** implementation and program management.
-

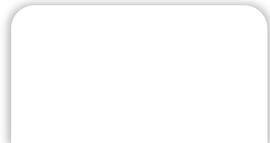


Evaluation



A systematic, objective assessment of an on-going or completed project, program, or policy, its design, implementation and/or results, asking

- **Descriptive Questions** to seek to determine what is taking place and describe aspect of a process.
- **Normative Questions** to compare what is taking place to what should be taking place.
- **Cause-and-Effect Questions** to examine outcomes and assess what difference the intervention makes in outcomes



Complementary Roles of Monitoring and Evaluation

Monitoring		Evaluation	
√	Clarifies program objectives	√	Analyzes why intended results were or were not achieved
√	Links activities and their resources to objectives	√	Assesses specific causal contributions of activities to results
√	Translates objectives into performance indicators and set targets	√	Examines implementation process
√	Routinely collects data on these indicators, compares actual results with targets	√	Explores unintended results
√	Reports progress to managers and alerts them to problems	√	Provides lessons, highlights significant accomplishment or program potential, and offers recommendations for improvement

Evaluations in SSN Projects

Descriptive Process Evaluations -- Assess whether a program is being implemented as planned

- Tailored to program's institutional arrangements and components
- Often include quantitative and qualitative approaches
- Particularly useful at early stages of program implementation

Normative Targeting/Incidence Analysis – Determine whether the program is reaching its intended beneficiaries

- Can be applied at the geographical and household levels
- Includes errors of inclusion and exclusion
- Needs a reference from national measures of poverty (usually direct or proxy measures of income or consumption) against which to benchmark program performance
- Can use national surveys with ID of program beneficiaries, and oversampling if needed and/or regular program registration process

Evaluations in SSN Projects

Causal Impact Evaluations – An assessment of the **causal** effect of a project , program or policy on beneficiaries

- Uses a counterfactual obtained from a control or comparison group to estimate the state of the beneficiaries in the absence of the program
- Relies on baseline and follow-up data on treatment and comparison groups

Useful for:

- Determining intermediate or final outcomes attributable to the intervention
 - Often used to examine questions with less clear answers such as changes in behavior or outcomes with a range of drivers
- Testing program design options
 - For example, different outreach strategies or the relative effectiveness of different benefit packages

Structuring Evaluations

Evaluations are derived from the question posed and should be tailored accordingly

Evaluations can benefit from...

- Combining quantitative and qualitative data
 - Cost – benefit analysis
 - Ensuring timeliness of measuring results, producing information to inform key decisions
 - Early planning!
- Keep an eye on costs and take advantage of available data, national surveys



WORLD BANK GROUP
Social Protection & Labor

Results Chains

Using a Results Chain

A Results Chain maps out a **theory of change** for any project:

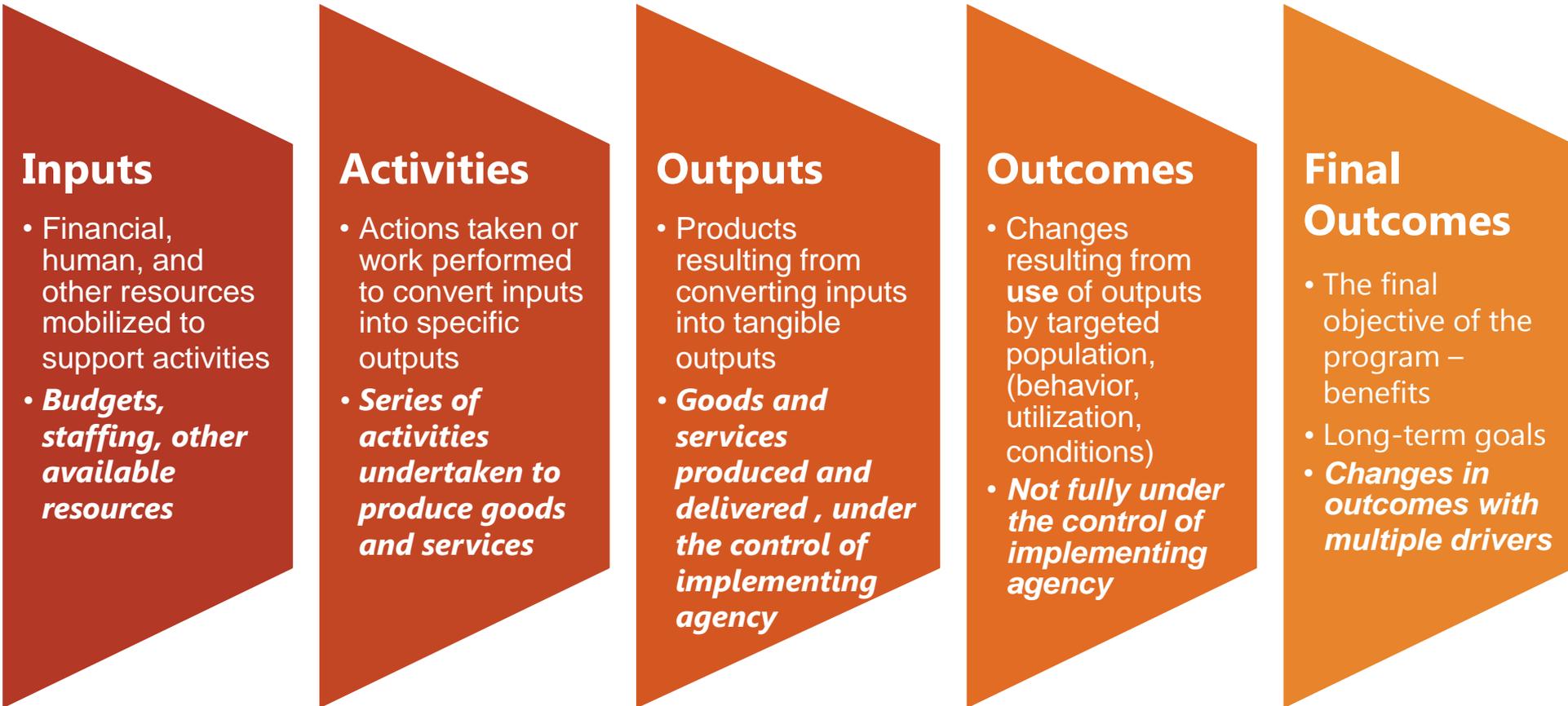
What are the intended results of the program?

How will the intended results be achieved? Are there any critical assumptions? Sufficient resources?

How will we know we have achieved the intended results?



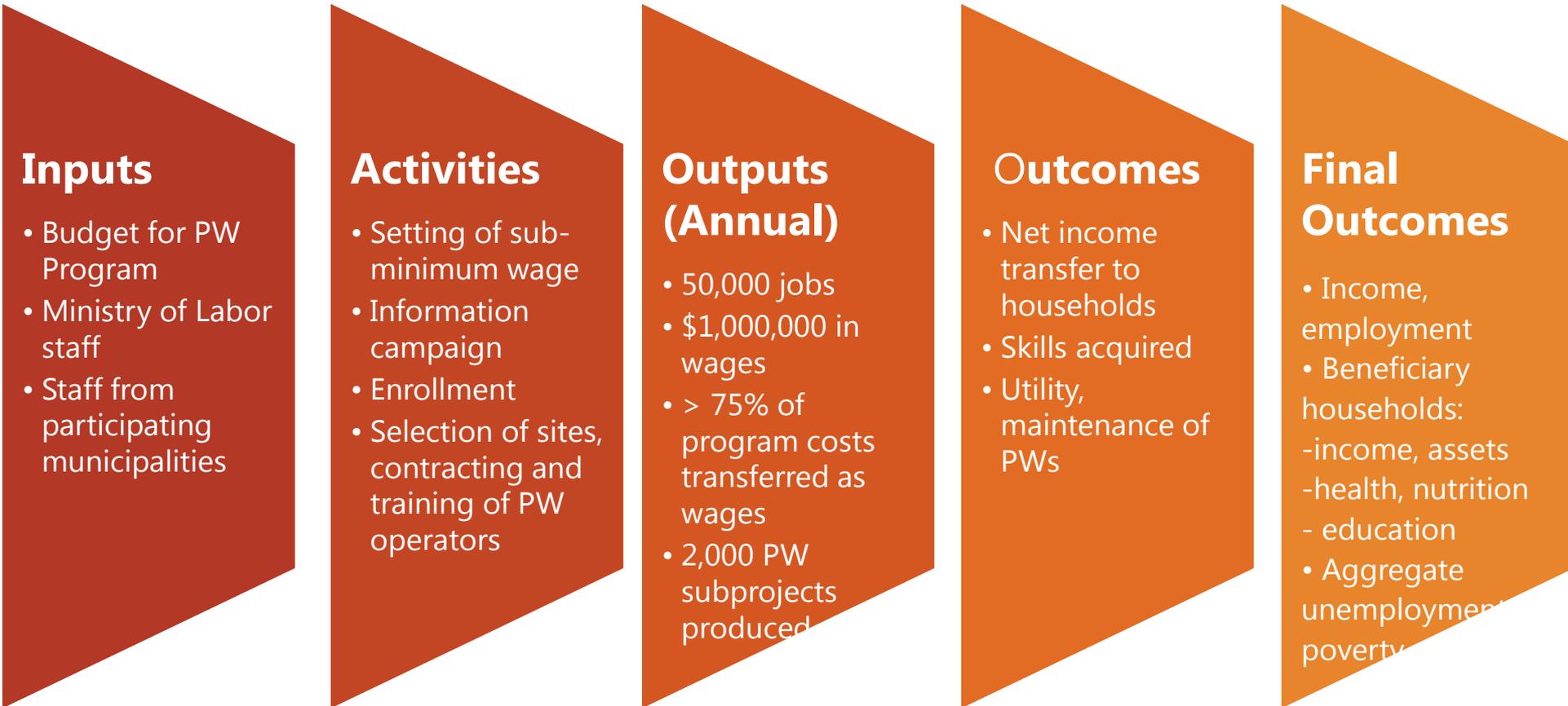
Typical Results Chain



Implementation (SUPPLY SIDE)

Results (DEMAND + SUPPLY)

Public Works Program Results Chain Example



Implementation (SUPPLY SIDE)

Results (DEMAND + SUPPLY)



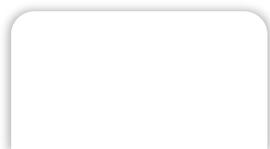
WORLD BANK GROUP
Social Protection & Labor

The Main Concepts of Impact Evaluation

Impact Evaluation needs to be distinguished from other “evaluations”



The objective of impact evaluation is to estimate the **causal** effect or **impact** of a program on outcomes of interest.



The Objective



Estimate the causal effect (impact) of intervention (P) on outcome (Y).

(P) = Program or Treatment

(Y) = Outcome Indicator, Measure of Success

Example: What is the effect of a cash transfer program (P) on Household Consumption (Y)?



Solution

Estimate what **would** have happened to outcomes (Y) in the absence of the program (P).

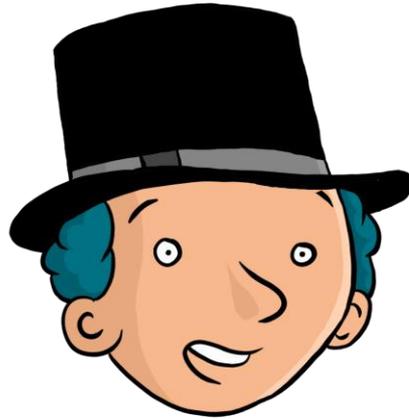
We call this the **Counterfactual**.

The key to a good impact evaluation is a valid estimate of the **counterfactual!**

Counterfactual is estimated based on **comparison groups**

Communicating complex concepts in 3 slides... Example: What is the Impact of...

giving Ali



additional money



(P)

on Ali's consumption

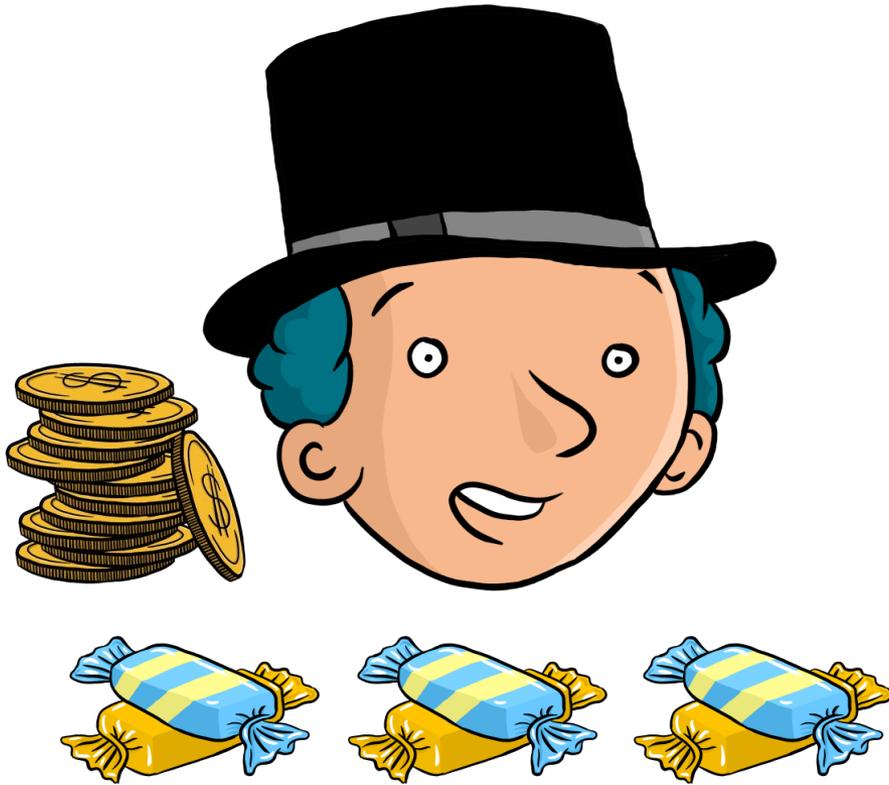


(Y)?



The Perfect Clone

Ali



6 candies

Ali's Clone

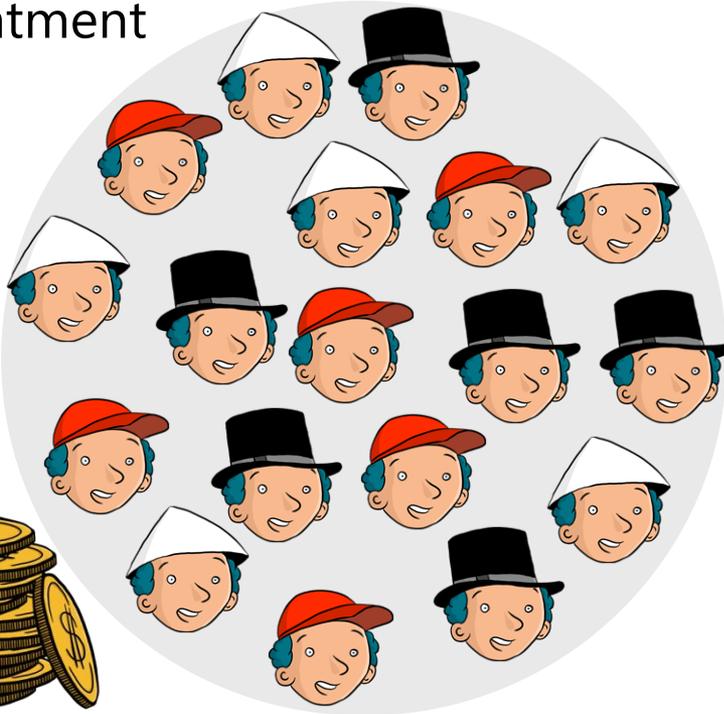


4 candies

$\text{IMPACT} = 6 - 4 = 2$ Candies

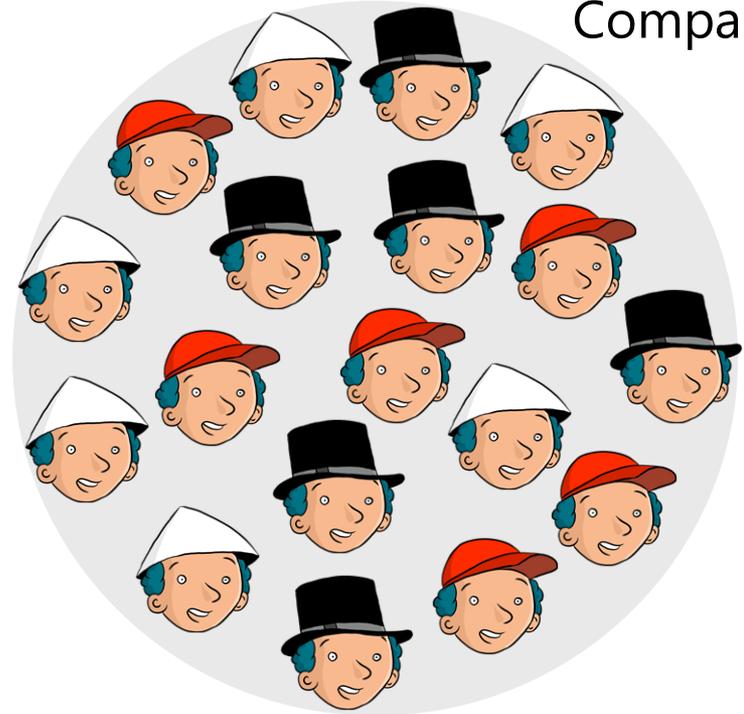
In reality, use statistics

Treatment



Average $Y=6$ candies

Comparison



Average $Y=4$ Candies

$IMPACT=6-4=2$ Candies



WORLD BANK GROUP
Social Protection & Labor

Choosing the best IE design for your project

Finding good comparison groups

We want to find **clones** for the Alis in our programs.

The treatment and comparison groups should

- have identical characteristics

benefiting from the intervention.



With a good comparison group, the **only reason** for different outcomes between treatments and controls is the **intervention (P)**

program eligibility & assignment
to get valid estimates of the
counterfactuals

Two false counterfactuals to avoid



Before
vs After

Compare: Same individuals
Before and After they
receive **P**.

Problem: Other things may
have happened over time.

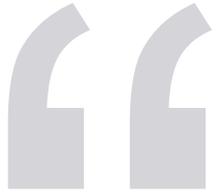
Enrolled
vs Not Enrolled

Compare: Group of
individuals **Enrolled** in a
program with group that
chooses not to enroll.

Problem: Selection Bias.
We don't know why they
are not enrolled.

**Both counterfactuals lead
to biased estimates of the
counterfactual and the
impact.**

The conversation needs to start early



Retrospective Evaluation is necessary when we have to work with a program that has already been roll-out and existing data.

Rarely feasible: baseline data? Information on targeting?



In **Prospective Evaluation**, the evaluation is designed in parallel with the program (and targeting decisions).

The way to go: ensure baseline data is collected, and comparison group exists.



Where do good Comparison Groups come from?

- The rules of program operation determine the evaluation strategy.
- We can almost always find a valid comparison group if:
 - the operational rules for selecting beneficiaries are equitable, transparent and accountable;
 - the evaluation is designed prospectively.
- Evaluation design and program design go hand-in-hand.

5 methods in IE Toolbox

- 5 methods in IE toolbox take different approaches to generate comparison groups and estimate the counterfactual:

1 Randomized Assignment

2 Randomized Promotion

3 Regression Discontinuity Design

RDD

4 Difference-in-Differences

DD

5 Matching

Choosing an IE design for your program

- Design IE prospectively to generate good comparison groups and collect baseline data
- 3 operational questions to determine which method is appropriate for a given program

Money: Does the program have sufficient resources to achieve scale and reach full coverage of all eligible beneficiaries?

Targeting Rules: Who is eligible for program benefits? Is the program targeted based on an eligibility cut-off or is it available to everyone?

Timing: How are potential beneficiaries enrolled in the program – all at once or in phases over time?

Choosing your IE method(s)

Money →	<i>Excess demand</i>		<i>No Excess demand</i>	
Targeting →	<i>Targeted</i>	<i>Universal</i>	<i>Targeted</i>	<i>Universal</i>
Timing ↓				
<i>Phased Roll-out</i>	1 Randomized assignment 4 RDD	1 Randomized assignment 2 Randomized promotion 3 DD with 5 Matching	1 Randomized Assignment 4 RDD	1 Randomized assignment to phases 2 Randomized Promotion to early take-up 3 DD with 5 matching
<i>Immediate Roll-out</i>	1 Randomized Assignment 4 RDD	1 Randomized Assignment 2 Randomized Promotion 3 DD with 5 Matching	4 RDD	If less than full Take-up: 2 Randomized Promotion 3 DD with 5 Matching

Choosing the IE method in Niger

Money	<u>Excess demand</u>		No Excess demand	
Targeting	<u>Targeted</u>	Universal	Targeted	Universal
Timing				
<u>Phased Roll-out</u>	1 Randomized assignment 4 RDD	1 Randomized assignment 2 Randomized promotion 3 DD with 5 Matching	1 Randomized Assignment 4 RDD	1 Randomized assignment to phases 2 Randomized Promotion to early take-up 3 DD with 5 matching
<u>Immediate Roll-out</u>	1 Randomized Assignment 4 RDD	1 Randomized Assignment 2 Randomized Promotion 3 DD with 5 Matching	4 RDD	If less than full Take-up: 2 Randomized Promotion 3 DD with 5 Matching



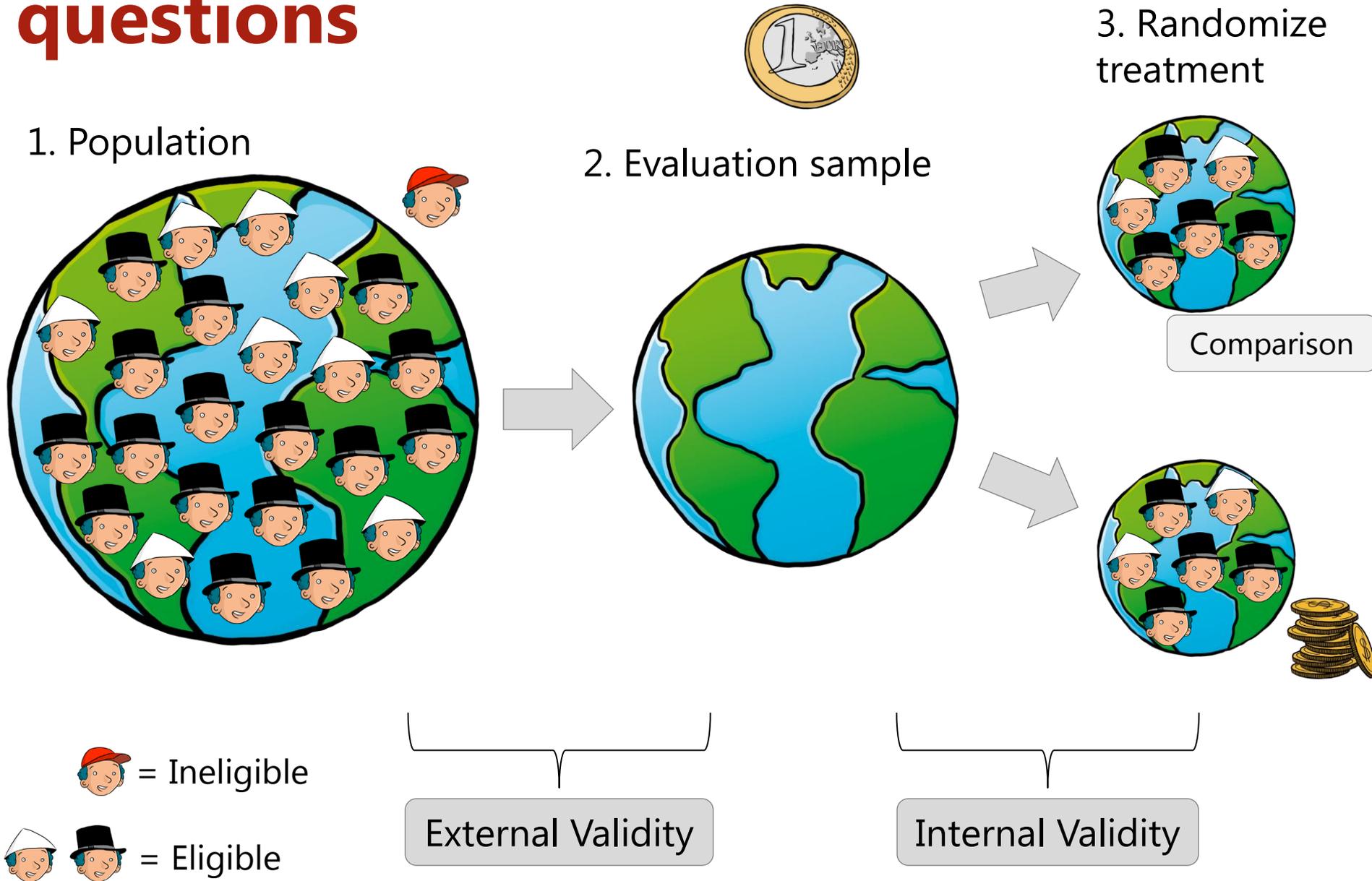
WORLD BANK GROUP
Social Protection & Labor

Randomization as an operational tool

Many projects use randomization as a fair, transparent assignment rule

- Example 1: Public Works in Cote d'Ivoire
 - Originally introduced in post-conflict setting, including to target ex-combattants
 - Public Lottery was considered the most transparent and acceptable selection process
- Example 2: Niger Safety Nets project
 - Geographical targeting only works with higher administrative units (Regions, departments, communes...)
 - How to chose between long list of criteria to select villages within units?
 - Public lottery was deemed the most transparent, and least controversial approach.
 - Project decided to keep using the randomization approach for purely operational reasons, including when not needed for the IE.

Randomization to answer basic IE questions

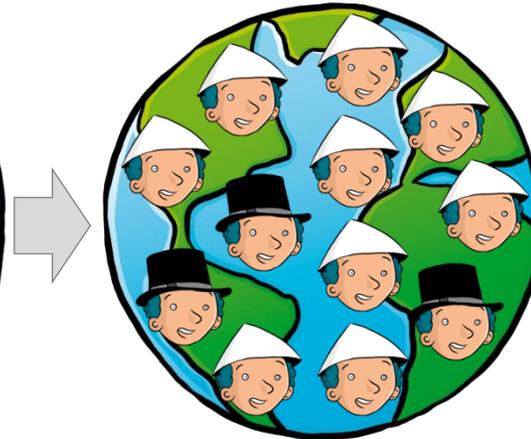


Randomized to answer IE “design” questions

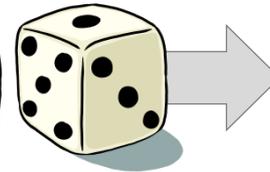
1. Population



2. Evaluation Sample



3. Randomize treatment.



Comparison



 = Not eligible

  = Eligible

Randomized Assignment



In **Randomized Assignment**, large enough samples, produces 2 statistically equivalent groups.

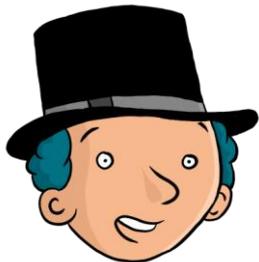
We have identified the perfect **clone**.

Feasible for prospective evaluations with over-subscription/excess demand.

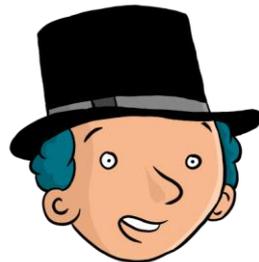
Most pilots and new programs fall into this category.

Consider evaluating relative effectiveness of alternative program design options.

Randomized beneficiary



Randomized comparison





WORLD BANK GROUP
Social Protection & Labor

Ethical Principles

Ethical Principles

- **Basic Considerations for Ethics of Impact Evaluation:**
 - Evaluations should not dictate how programs are assigned; instead, evaluations should be fitted program assignment rules to the extent that those are clear and fair.
 - Groups should not be excluded from an intervention that is known to be beneficial solely for the purpose of an evaluation.
 - There can also be an ethical concern about *not* pursuing an evaluation when programs invest substantial public resources in interventions whose effectiveness is unknown
- **More Specific Considerations when conducting impact evaluation**
 - Human Subjects Research
 - Open Science

Human Subjects Research

Protecting Human Subjects

The World Health Organization (WHO) recommends the following basic criteria for assessing the research projects involving human subjects:

- The rights and welfare of the subjects involved should be adequately protected.
- The researchers should obtain freely given, informed consent from the participants.
- The balance between risk and potential benefits involved should be assessed and deemed acceptable by a panel of independent experts.
- Any special national requirements should be met.

An IRB is a committee that has been formally designated to review, approve, and monitor research involving humans

Open Science

Problem

Publication bias

Data Mining

Lack of Replication

Consequence

No information on what *doesn't* work

Only significant impacts are reported

Mistakes undetected; results between studies not comparable

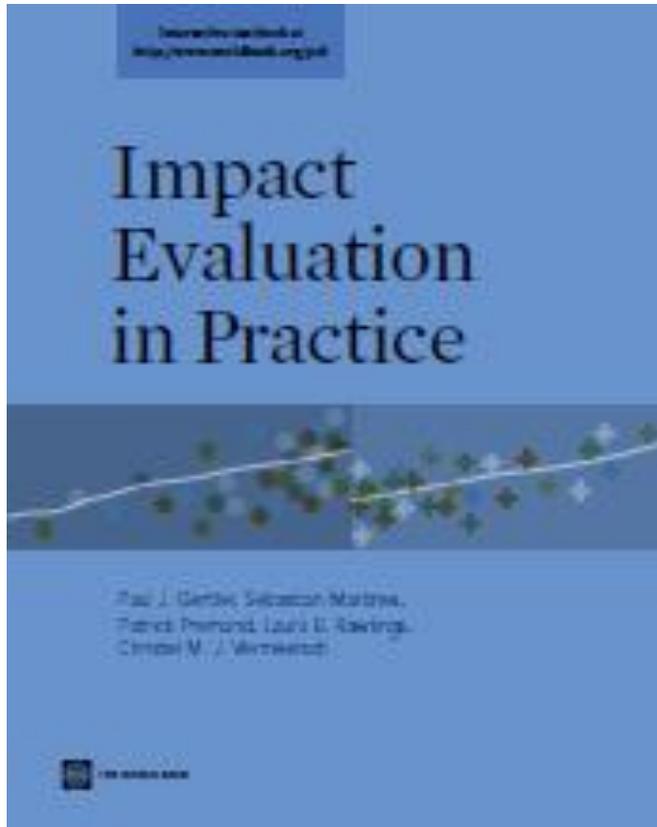
Solution

Trial registries

Pre-analysis plans

Data documentation, changes in journal policies

Thank you!



Reference also:
available in **Spanish**
French and
Portuguese

www.worldbank.org/ieinpractice



Evidence frontiers in SSN?

Spotlights on:

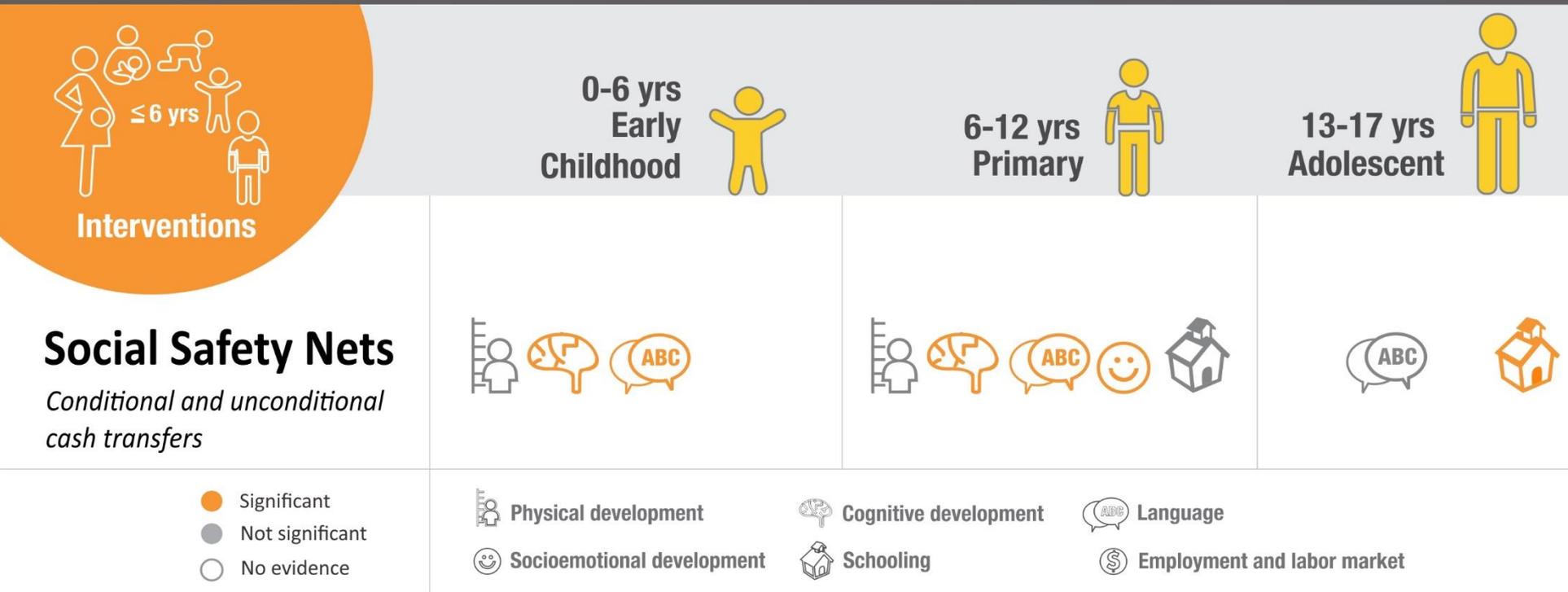
- **Links between SSNs and ECD/Nutrition**
- **Links between SSNs and productive interventions**



WORLD BANK GROUP
Social Protection & Labor

Spotlight on links between SSNs and ECD/Nutrition

Impacts of Early Childhood Interventions on Development Outcomes



SSNs in Early Childhood *can* Improve *some* Development Outcomes

- Tend to improve cognitive and socio-emotional development outcomes
- Tends not to affect physical development outcomes

Some advantages change over time

- Language advantage may diminish with age
- School enrolment benefits seem to increase with age

Sources, IEG: Outcomes age 0-6: <http://wbsupport2ecd>

Outcomes age 6+: <https://ieg.worldbankgroup.org/evaluations/later-impacts-early-childhood-interventions>

Madagascar's social safety net project: testing behavioral interventions for parenting



42

ideas



Why focus on behavioral interventions for parenting?

- 1) Traditional approaches of “training” and “information” have not always been successful in changing behaviors.
- 2) Extreme poor tend to have a reduced “bandwidth”, focusing on elemental needs, but not making the best (long term) decisions (e.g. “buy more rice”, pulling kids out of school).
- 3) Behavioral interventions can be cost-effective: light touch reminders rather than heavy-handed “conditionality”.
- 4) In Africa, social services are rare and thus alternative ways to reach the poor need to be explored.
- 5) Increasing recognition of the role of parenting in ECD and of behavioral change as a complement to income support in SP

Mother Leader Model

The Mother Leaders will be trained in five topics:

- Leadership and Self-confidence
- Citizenship
- Essential Family Practices (best practices for families in hygiene and health)
- Early childhood development (cognitive stimulation)
- Reproductive health

→ Fun, interactive and reflective, drawing on cultural traditions

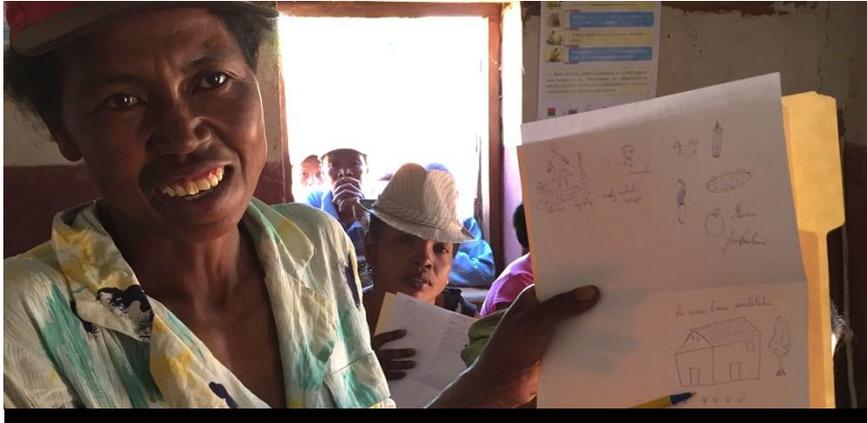
Role of Nudges in the Human Development Cash Transfer

→ Can nudges be used to support mothers' behavioral changes to invest more in their children's health, nutrition and cognitive stimulation?

Hypotheses about the nudges:

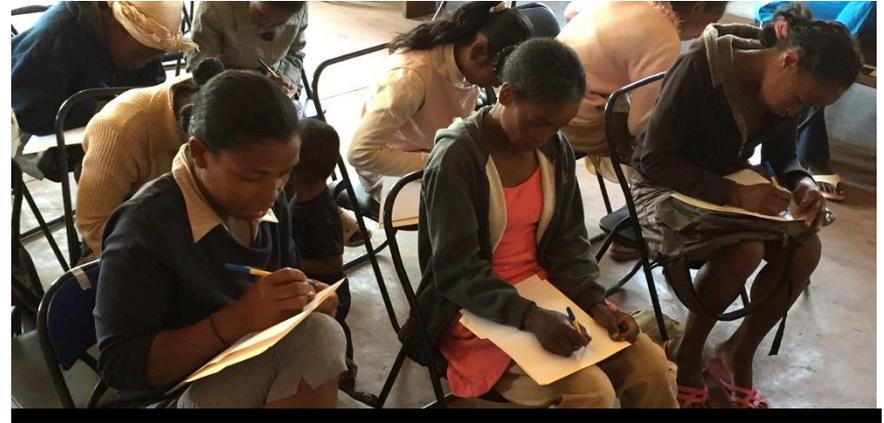
- will build intrinsic motivation
- serve as complements to Mother Leader sessions
- help orient use of cash

Human Development Cash Transfer “Nudges”



Self-Affirmation

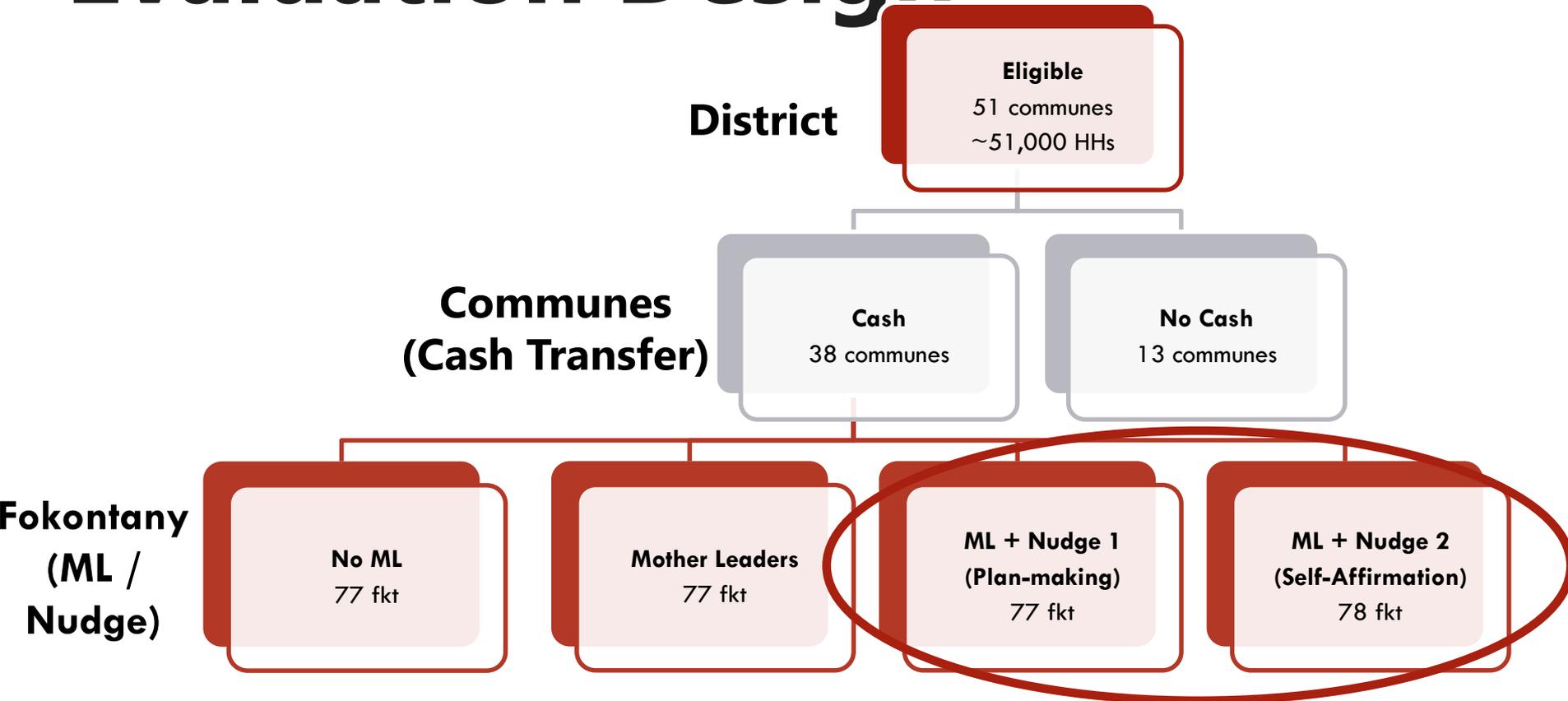
An internal attitude and belief that one has inherent value. The ability to act in response to one's needs and environment.



Plan-Making

The ability to set goals and determine the actions necessary to realize those goals.

HDCT Randomized Control Evaluation Design





WORLD BANK GROUP
Social Protection & Labor

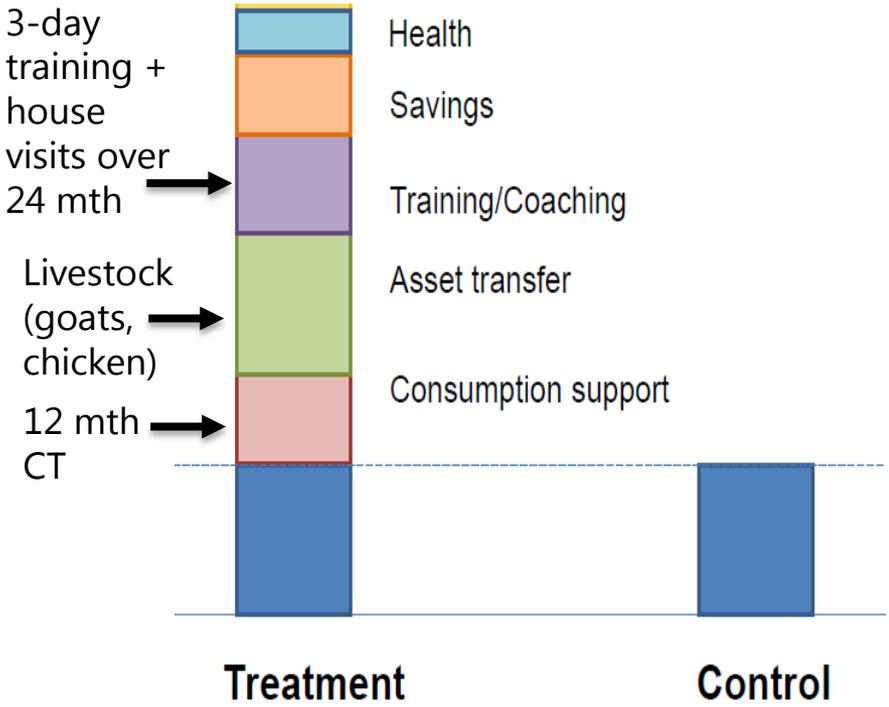
Spotlight on linkages between SSNs and productive interventions

2) International evidence: the BRAC model and CGAP evaluation

The BRAC graduation model ("Targeting the Ultra Poor")

Started for households too poor to participate in microcredit
Multifaceted strategy to unlock multiple constraints to productivity
Run in Bangladesh since 2002

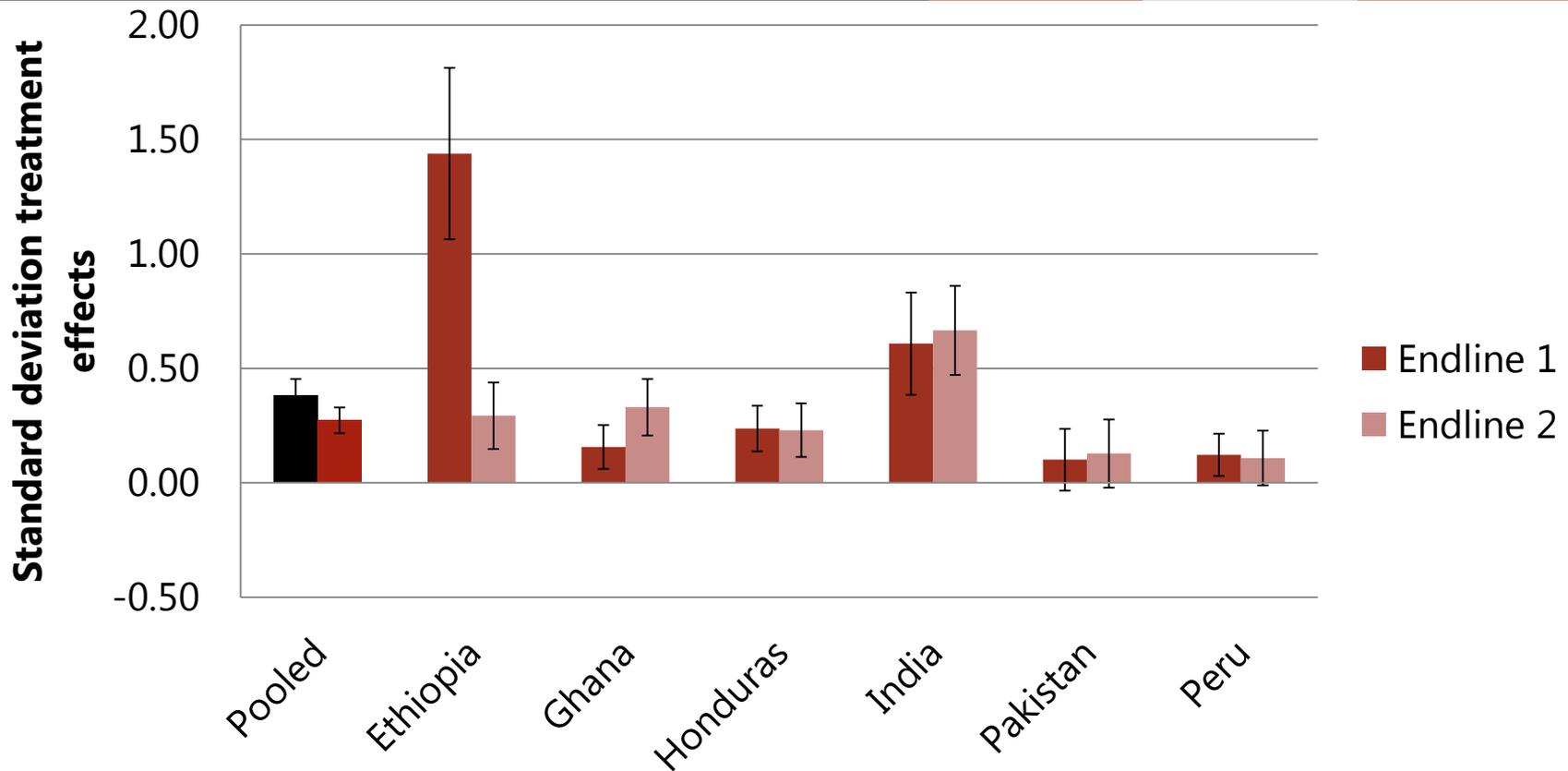
The CGAP multi-country experiment (Science, 2015)



Impact measured at 12 and 24 months



Income and revenues

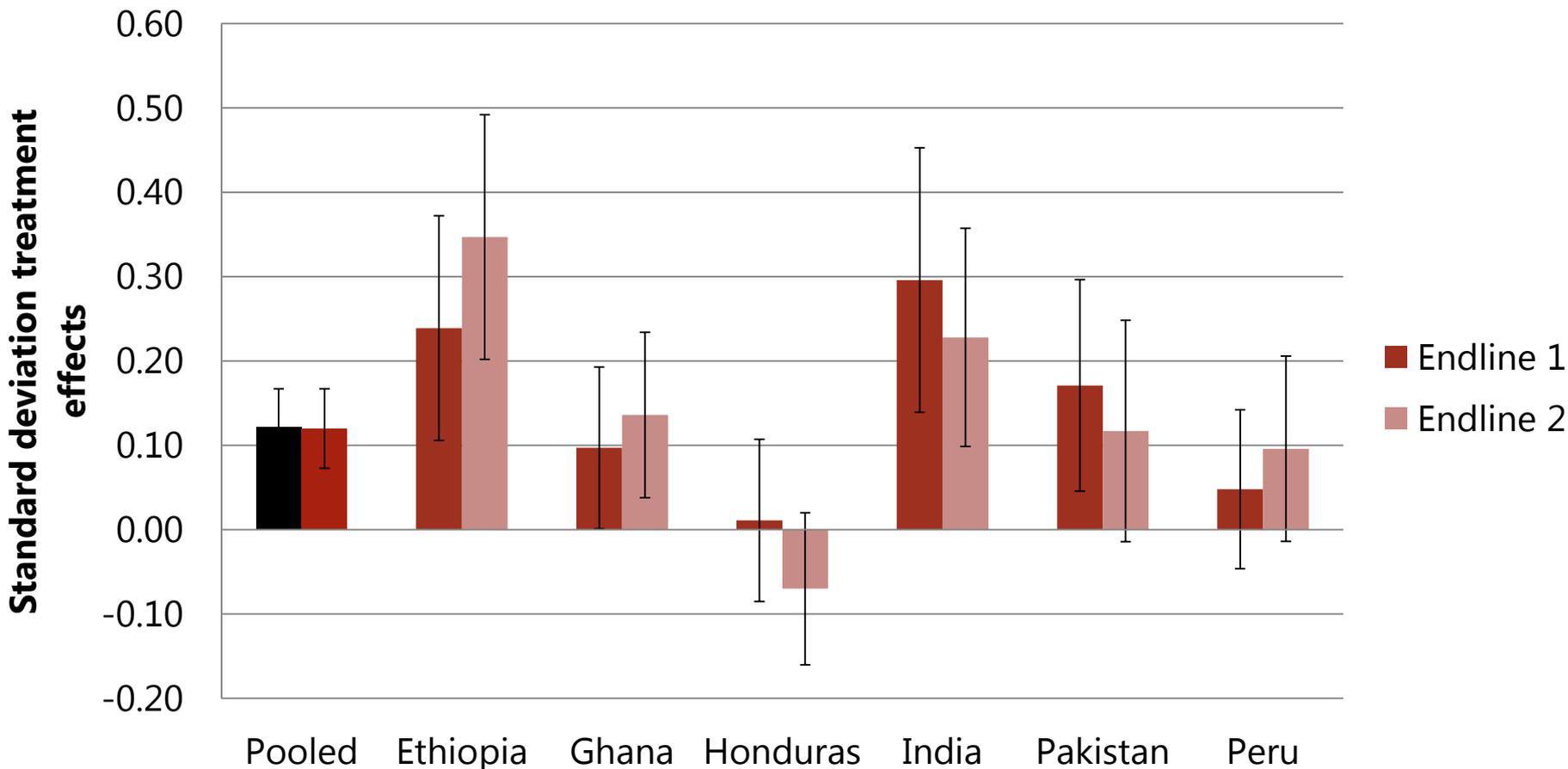


Monthly livestock revenues

Pooled endline 1: \$85.81 (control mean \$73.52): **+117%**

Pooled endline 2: \$55.50 (control mean \$80.62): **+33%**

Per capita consumption



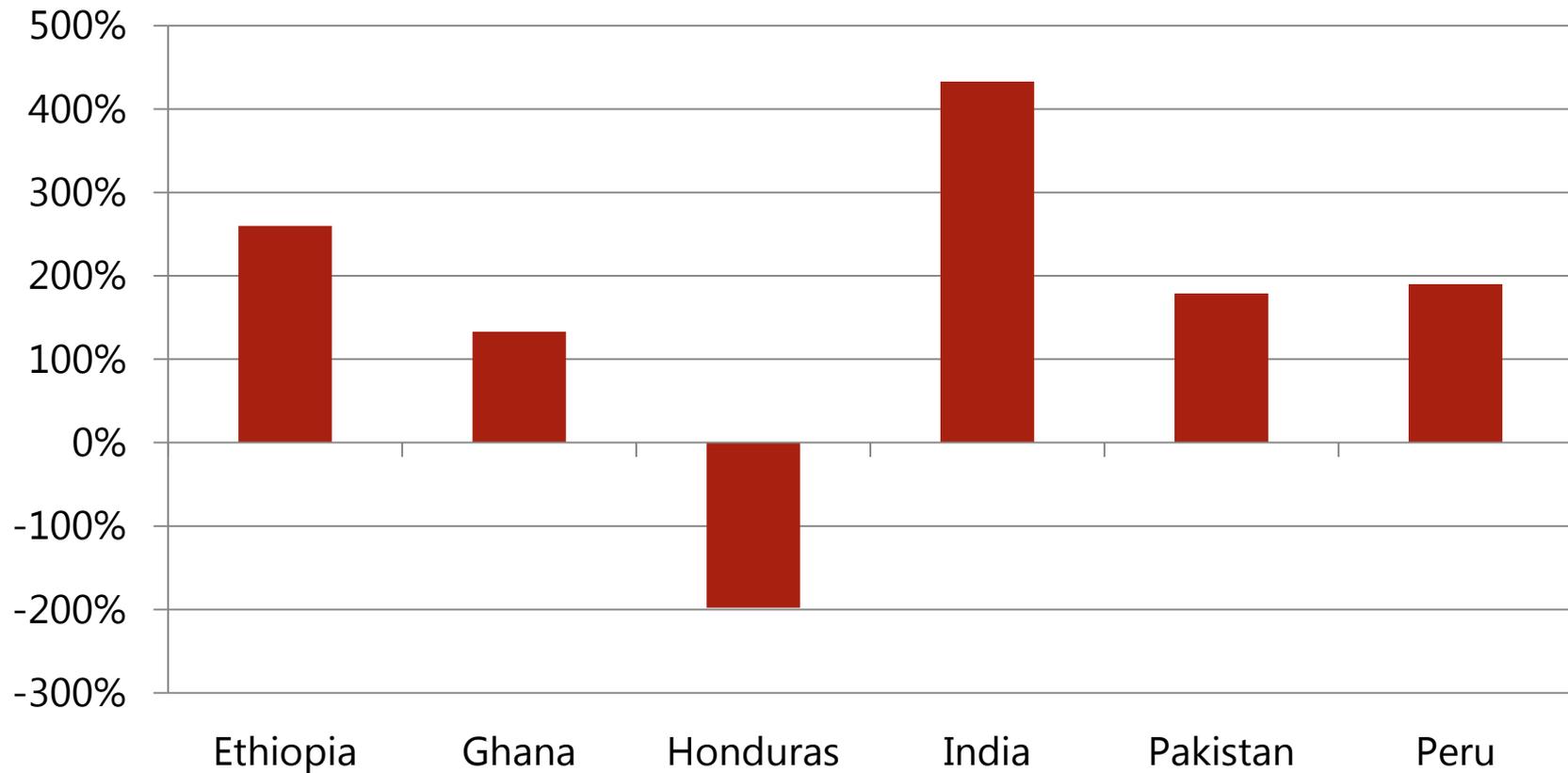
Per capita consumption, month

Pooled endline 1: \$4.55 (control mean \$78.80): **+6%**

Pooled endline 2: \$3.36 (control mean \$68.80): **+5%**

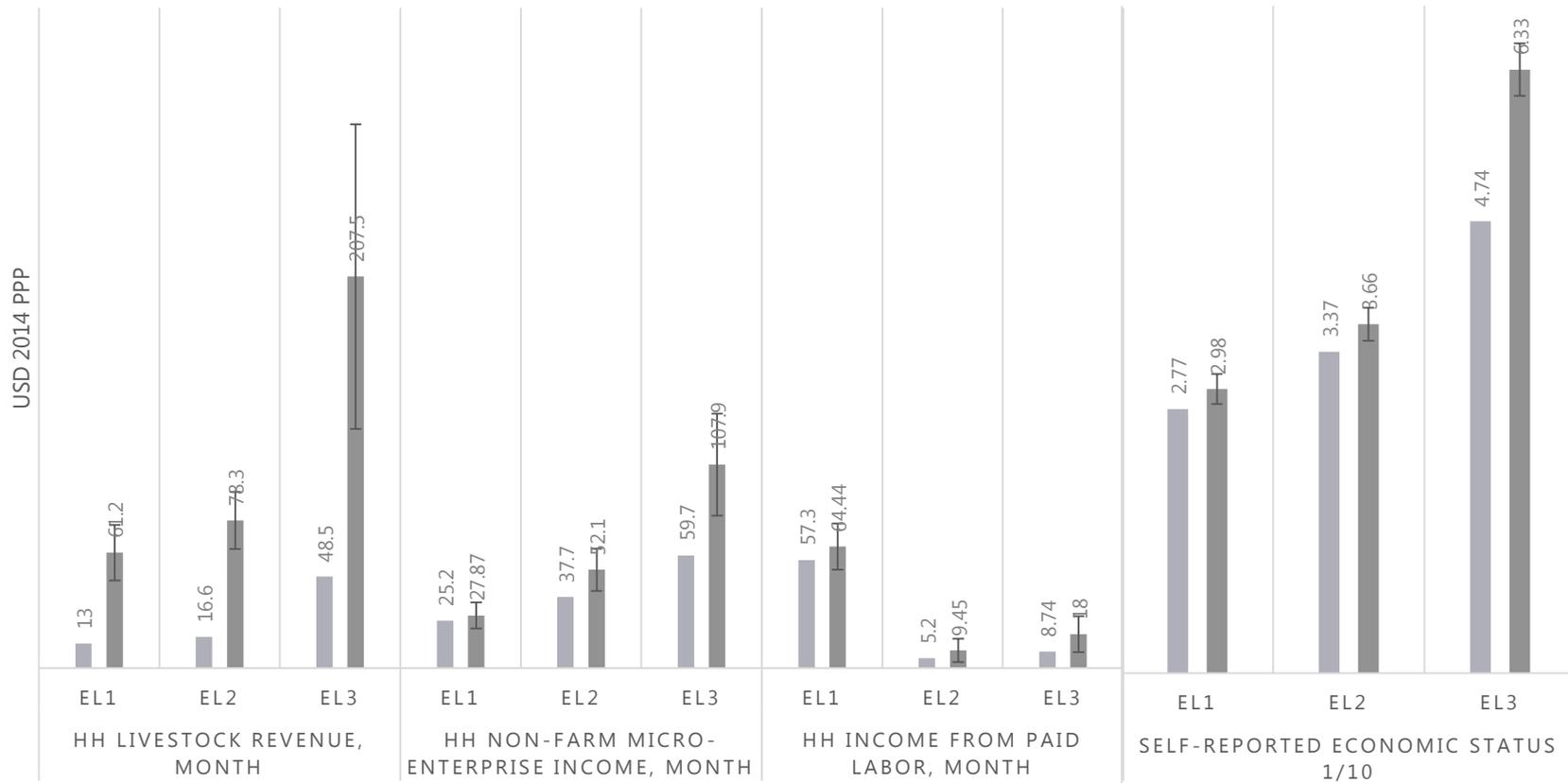
Cost-effectiveness

Total Benefits / Total Costs by Country



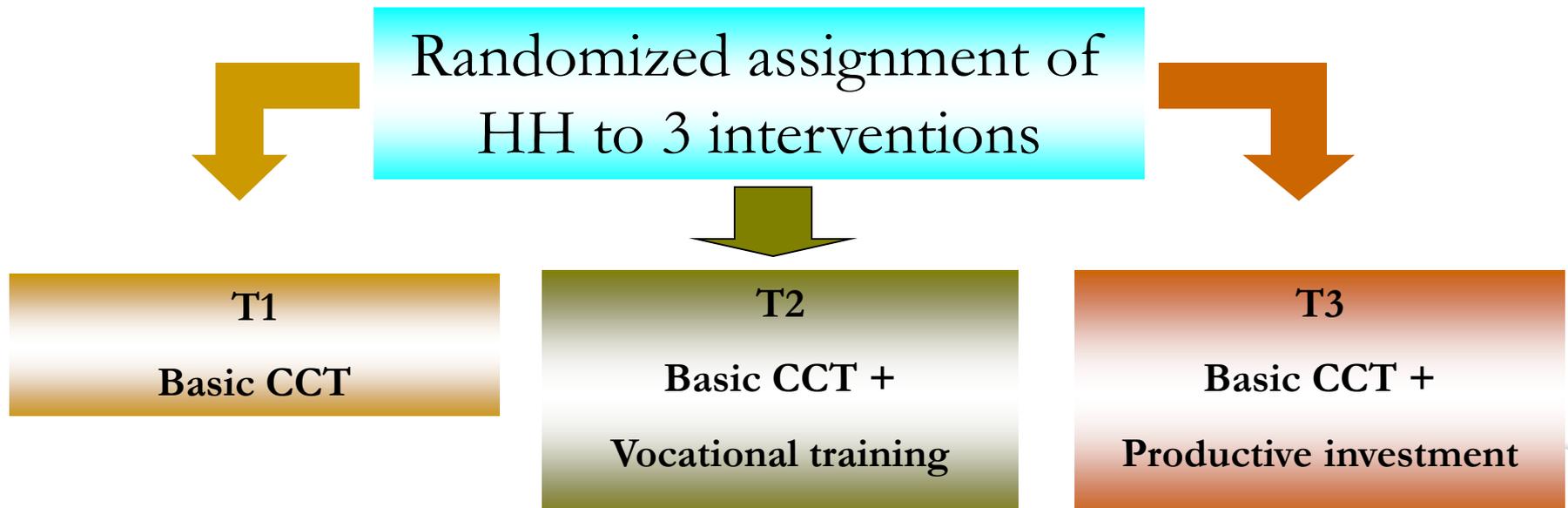
Long-term income and revenues (West Bengal)

■ Control ■ TUP



How to Build Household Resilience?

- Can Safety Nets Fulfill twin objectives
 - Offer short-term protection
 - Improve risk-management against climatic risks in the medium/long term
- Nicaragua “Atencion a Crisis pilots”



2 years after the end of the intervention, are beneficiaries still better protected against shocks?

○ CCT + Productive investment grant

- Leads to diversification out of agriculture (non-ag self-employment)
- **Fosters Resilience**: both income and consumption are protected against the negative welfare effects of drought shocks
- **Offers promotion**: there is also an average impacts on consumption and income

○ CCT + Training

- **Fosters Resilience**: both income and consumption are protected against the negative welfare effects of drought shocks
- Driven by diversification at times of shocks (non-ag self-employment)
- no average impacts on consumption and income

○ Traditional CCT

- Short-term impacts for consumption/income disappear for those with CCT
- After the end of the program, no protection against shock



WORLD BANK GROUP
Social Protection & Labor

Where do you see the frontier?

THANKS!