MONITORING & EVALUATING

Social Safety Net Programs

Laura B. Rawlings
The World Bank
Social Safety Nets Core Course - 2014

Objectives of this session

1. Why Focus on Results
2. Monitoring vs. Evaluation
3. Impact Evaluation
4. Evaluation in SSN Projects
5. Using a Results Chain
6. Putting M&E Into Practice

Annexes: Tips for using, presenting data; Impact evaluation methods

Objectives of this session

1. Why Focus on Results
2. Monitoring vs. Evaluation
3. Impact Evaluation
4. Evaluation in SSN Projects
5. Using a Results Chain
6. Putting M&E Into Practice

Results Based Management is a global trend

What is new about results?

- Managers are judged by their programs' performance, not their control of inputs:
- A shift in focus from inputs to outcomes.
- Establishing links between monitoring and evaluation, policy formulation, and budgets (value for money)
- Critical to effective public sector management

Monitoring vs. Evaluation

	Monitoring	Evaluation
Frequency	Regular, Continuous	Periodic
Coverage	All programs	Selected program, aspects
Data	Universal	Sample based
Depth of Information	Tracks implementation, looks at WHAT	Tailored, often to performance and impact/ WHY
Cost	Cost spread out	Can be high
Utility	Continuous program improvement, management	Major program decisions

Objectives of this session

1. Why Focus on Results
2. Monitoring vs. Evaluation
3. Impact Evaluation
4. Evaluation in SSN Projects
5. Using a Results Chain
6. Putting M&E Into Practice

HUMAN DEVELOPMENT NETWORK

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

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

Objectives of this session

1. Why Focus on Results
2. Monitoring vs. Evaluation
3. Impact Evaluation
4. Evaluation in SSN Projects
5. Using a Results Chain
6. Putting M&E Into Practice

HUMAN DEVELOPMENT NETWORK

Impact Evaluation

An assessment of the causal effect of a project , program or policy on beneficiaries. *Uses a counterfactual…*

- to estimate what the state of the beneficiaries would have been in the absence of the program *(the control or comparison group),* compared to the observed state of beneficiaries *(the treatment group),* and
- to determine intermediate or final outcomes attributable to the intervention .

Impact Evaluation Methods

All impact evaluations estimate the *counterfactual*, using control or comparison groups: *What would the treatment group be like in the absence of the program?*

1. **Experimental/Randomized Assignment**
   - uses randomized assignment to determine who gets program treatment(s) and who is control among eligible beneficiaries
   - can be used ethically in cases where program cannot reach all potential beneficiaries at once; or to test program alternatives
   - random assignment creates statistically equivalent groups (treatment and control) which allows a valid estimate of the counterfactual
2. **Quasi-Experimental**
   - mimics experimental designs
   - methods to create comparison groups include:
     - Regression Discontinuity
     - Differences in Differences
     - Instrumental Variables
     - Statistical Matching

→ **Choice of method depends on context. Rules of program operation are key because they determine eligibility for the program! Use them to ID comparison group.**

When to use Impact Evaluation?

- Evaluate impact selectively, when project is:
  - Innovative
  - Replicable/scalable
  - Strategically relevant
  - Evaluation will fill knowledge gap
  - Substantial policy impact
- Use evaluation within a program to test alternatives, improve programs
  - → Beyond 'does my program work'?
  - → To 'which design is most cost effective?'
  - → *Impact Evaluation 2.0*

HUMAN DEVELOPMENT NETWORK

Objectives of this session

1. Why Focus on Results
2. Monitoring vs. Evaluation
3. Impact Evaluation
4. Evaluation in SSN Projects
5. Using a Results Chain
6. Putting M&E Into Practice

HUMAN DEVELOPMENT NETWORK

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

Objectives of this session

1. Why Focus on Results
2. Monitoring vs. Evaluation
3. Impact Evaluation
4. Evaluation in SSN Projects
5. Using a Results Chain
6. Putting M&E Into Practice

HUMAN DEVELOPMENT NETWORK

Using a Results Chain

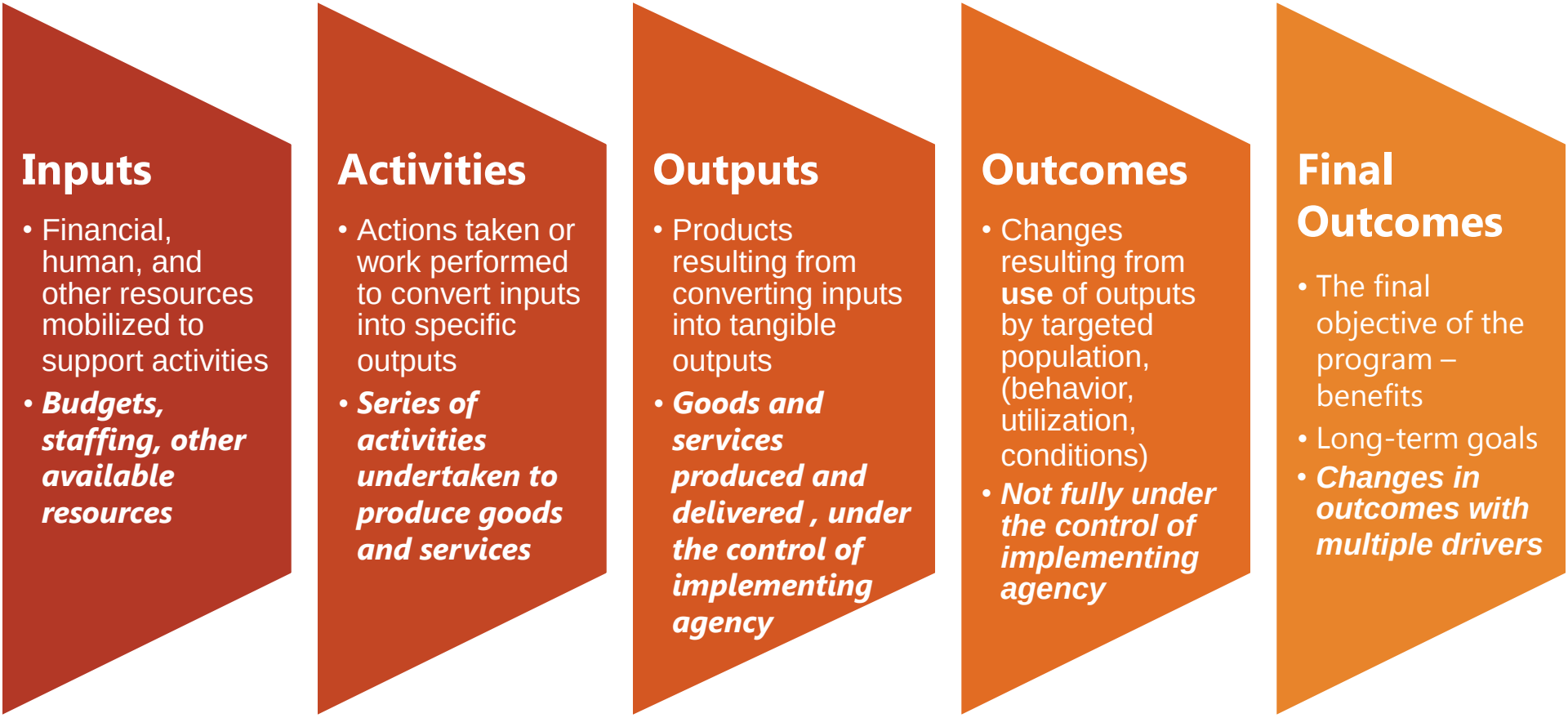
A Results Chain maps out a theory of change for any project:

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?

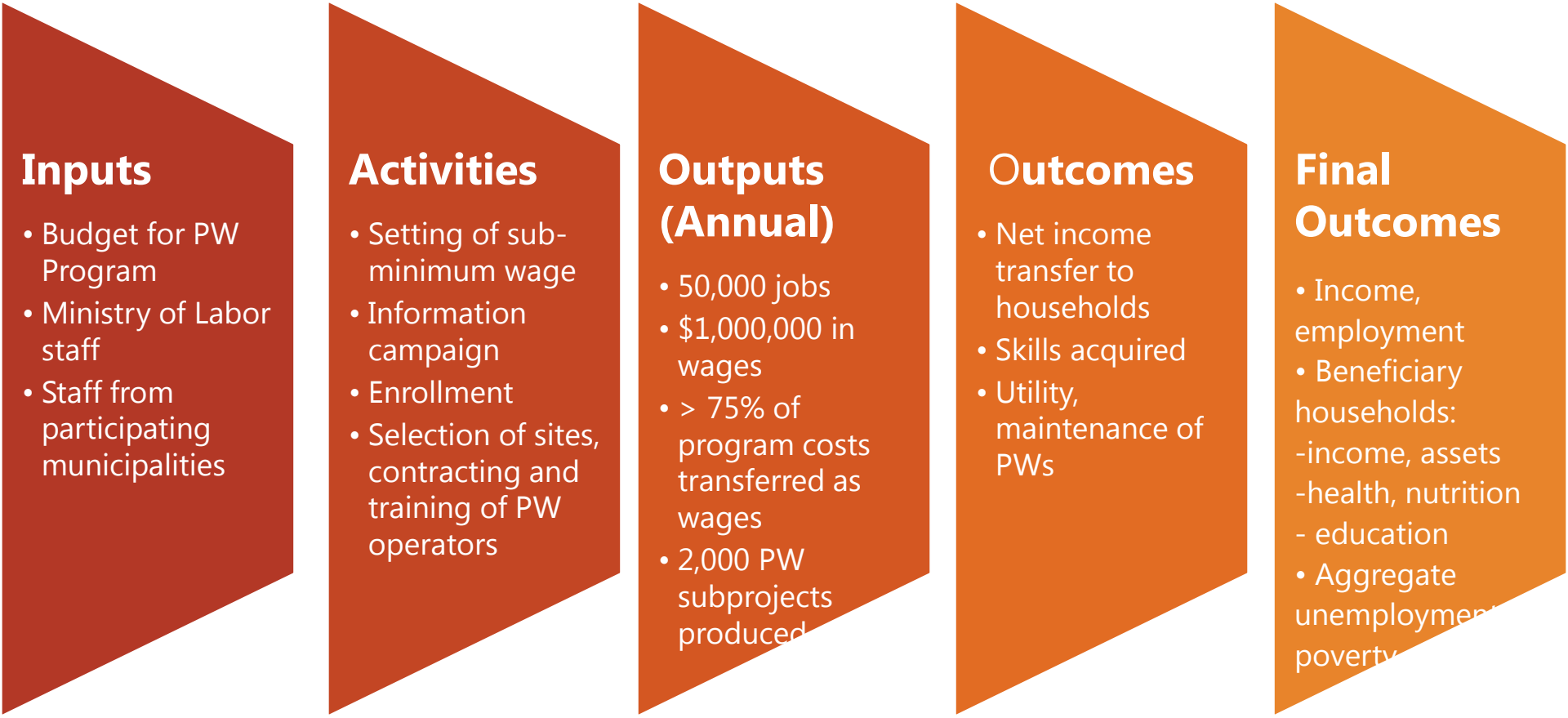
Typical Results Chain



Implementation (SUPPLY SIDE)

Results (DEMAND + SUPPLY)

Public Works Program Results Chain Example



Implementation (SUPPLY SIDE)

Results (DEMAND + SUPPLY)

Exercise: Results Chain

Identify the sequence of inputs, activities, outputs and outcomes:

1. Information about the importance of pre-school made available to program parents
2. Children perform better in primary, secondary
3. Improved rates of age-appropriate enrollment into primary school
4. More children enrolled in pre-school
5. Funds available to promote pre-school enrollment for children of cash transfer program beneficiaries
6. Design information campaigns on the importance of pre-school

Exercise: Results Chain

Identify the sequence of inputs, activities, outputs and outcomes:

5	Funds available to promote pre-school enrollment for children of cash transfer program beneficiaries	Input
6	Design information campaigns on the importance of pre-school for program parents	Activity
1	Information about the importance of pre-school made available to program parents	Output
4	More children enrolled in pre-school	Outcome
3	Improved rates of age-appropriate enrollment into primary school	Outcome
2	Children perform better in primary, secondary	Outcome

Objectives of this session

1. Why Focus on Results
2. Monitoring vs. Evaluation
3. Impact Evaluation
4. Evaluation in SSN Projects
5. Using a Results Chain
6. Putting M&E Into Practice

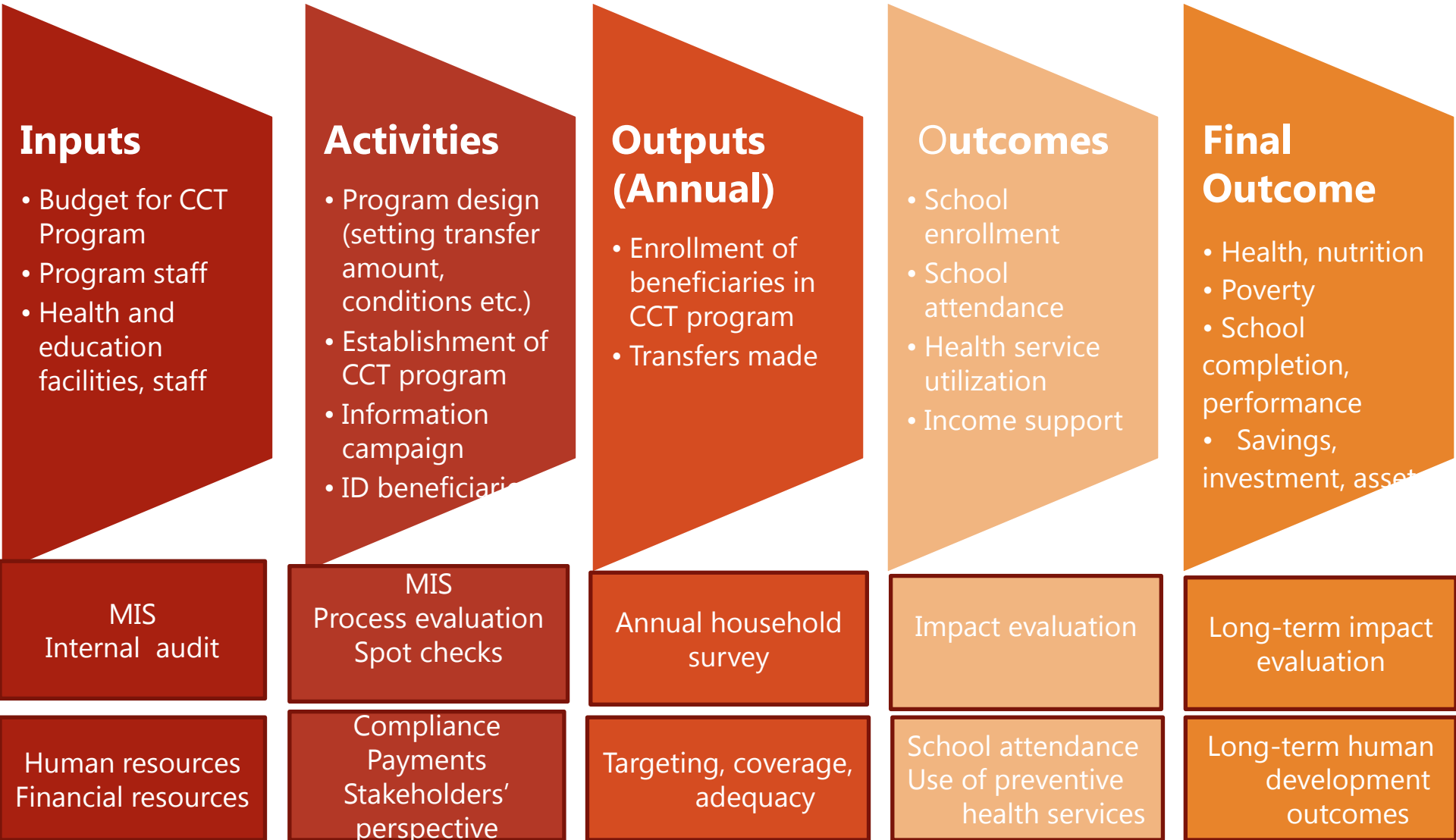
HUMAN DEVELOPMENT NETWORK

Implementing the Results Chain

Jamaica PATH CCT Program

- Example of how a well-structured program level M&E helped shape program design and inform policy decisions
- Program of Advancement Through Health and Education (PATH)
- Conditional cash transfer (CCT) program aimed at linking social assistance with human capital accumulation
- Primarily child grants to poor children <19 conditional on school, health care usage

Jamaica CCT Results Chain, Instruments, Indicators



Use of PATH M&E Results

Instruments	Key Indicator	
	Results	**Use**
Inputs: • **Management Info System (MIS)** • **Internal Audit**	○ Some lag in payments ○ Good compliance with conditions ○ Slower take up rate of program	○ Adjustments to payment system ○ Intensified outreach
Activities: • **Process evaluation** • **Spot checks**	○ Application process seen as burdensome ○ Stakeholders not clear on program rules ○ Weak system for verifying eligibility of new beneficiaries ○ Delays in appeals processing ○ Strong demand for jobs/ training	○ Revamping of MIS ○ Revised operations manual ○ Social workers used as focal points to access a variety of social services ○ "Steps to Work", new program created with focus on employment, labor markets skills development

Use of PATH M&E Results

Instruments	Key Indicator	
	Results	**Use**
Outputs: • **Special targeting assessment (using annual household survey)**	○ PATH better at reaching the poor than other Jamaican safety net programs ○ Not as good as other internationally	○ Improved the beneficiary identification system ○ Expanded training for social workers to help verify eligibility ○ More frequent recertification
Outcomes: • **Impact evaluation**	○ Education: School attendance improved slightly (by about half a day in a 20 day period). No impact on enrollment ○ Health: 30% increase in use of preventive health services	○ Focused main education objective on school completion ○ Introduced differentiated benefit levels to provide incentives for completion (gender, age) ○ Introduced a bonus for completing high school

Lessons Learned

- A well articulated approach to M&E is critical to good program management and to informing policy
- Impact evaluations are powerful for informing key program and policy decisions
- Good monitoring systems
  - Allow for results-based planning and management
  - Facilitate project preparation, supervision and reform and program implementation
  - Monitoring and IE play important complementary functions

Lessons Learned

What does it take to get there?

- Clients willing to learn, take risks, experiment, and collaborate → change in perspective "*from threats to tools*"
- Strong support of M&E by senior government champions and demand for transparency by civil society → champions, demands
- Donor and government desire to focus on M&E processes and goals → change in culture
- Cross-sectoral collaboration in the government (especially Ministry of Finance) & donors → collaboration

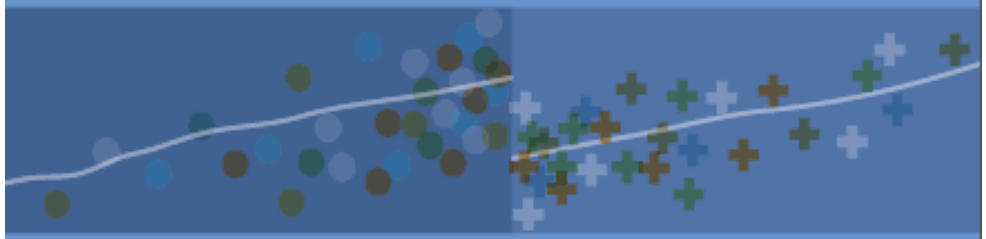
Conclusions

- Monitoring and evaluation are separate, complementary functions, but both are key to results-based management
- Good M&E is crucial not only to effective program management but can be a driver for reform
- Have a good M&E plan before you roll out/scale-up and use it to inform the journey!
- Design the timing and content of M&E results to further evidence-based dialogue
- Good monitoring is essential to good impact evaluation

THANK YOU

Interactive textbook at
http://www.worldbank.org/pdt

Impact Evaluation in Practice



Paul J. Gertler, Sebastian Martinez,
Patrick Premand, Laura B. Rawlings,
Christel M. J. Vermeersch

THE WORLD BANK

ANNEX 1:
Tips for Using, Presenting Data

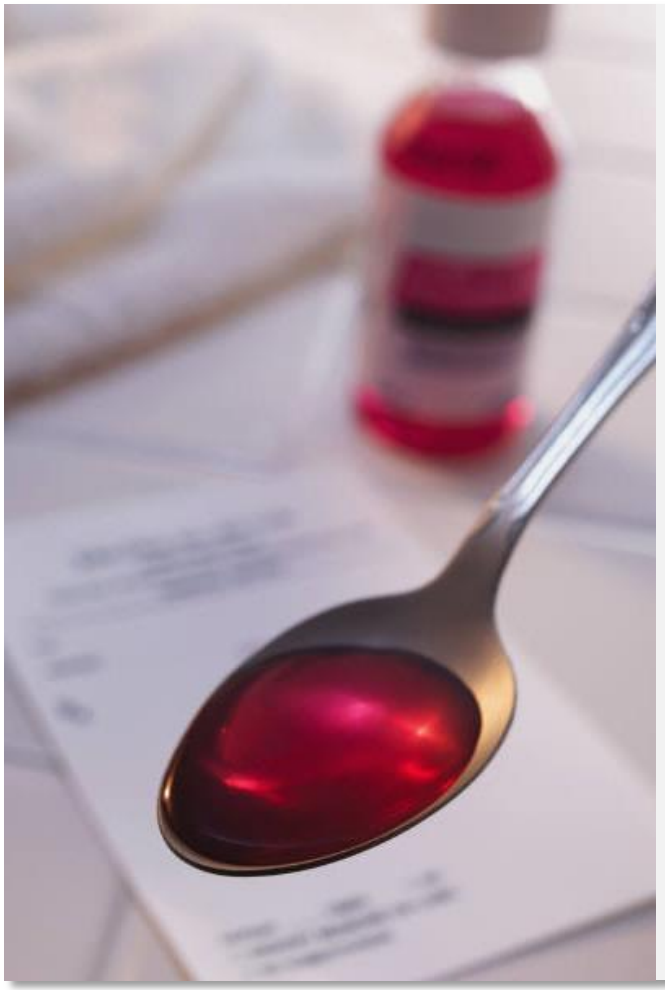
Develop a Data Collection Plan

- Identify what specific data are needed
- Identify which data are available
- Identify how the data will be collected
- Identify who will be responsible for collecting and reporting the data
- Identify when the data will be collected and reported, including how frequently
- Identify costs and sources of financing
- Identify who will use the data

Example: Data Collection and Reporting Plan

Indicator	Data Source	Data Collection Method	Who will collect data	Frequency of Collection	Cost of collection	Difficulty to collect	Who will analyze & report data
1							
2							
3							

SMART: Identifying good indicators



Specific

Measurable

Attributable

Realistic

Targeted

Quick Tips

on making performance monitoring really useful…

1) Provide frequent, timely information to program staff.

2) Set targets for each performance indicator.

3) Provide sub-group data. Disaggregate data by customer and service characteristics (*region*).

4) Do regular, basic, analysis of the data, especially comparisons.

5) Require explanations for unexpected findings.

6) Report findings in a user-friendly way.

7) Hold "How Are We Doing?" sessions after each performance report.

8) Use "Red-Yellow-Green Lights" to identify programs/projects needing attention.

9) Link outcome information to program costs.

10) Use/create a unique beneficiary ID to link data across programs and expand the analysis.

Source: Harry Hatry, Urban Institute

ANNEX 2:
Impact Evaluation Methods

Randomized Assignment

Randomized Offering/Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

Impact Evaluation Toolbox

Randomized Assignment

Randomized Offering/Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

Impact Evaluation Toolbox

SIEF

Randomized Assignment of Treatments & Comparison

Eligibles > Number of Benefits

- Randomize!
- Lottery for who is offered benefits
- Fair, transparent and ethical way to assign benefits to equally deserving populations.

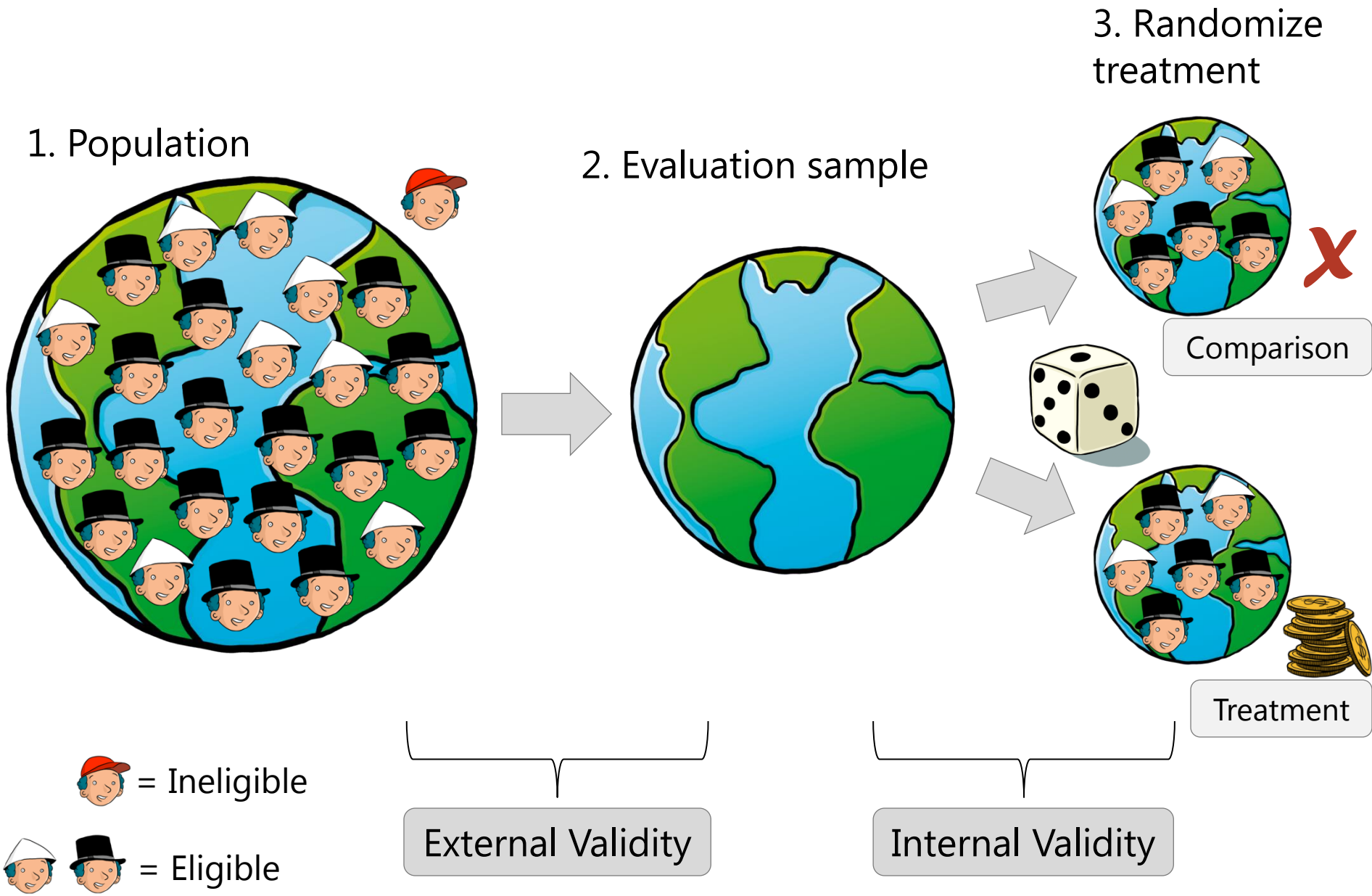
Oversubscription

- Give each eligible unit the same chance of receiving treatment
- Compare those offered treatment with those not offered treatment *(comparisons)*.

Randomized Phase In

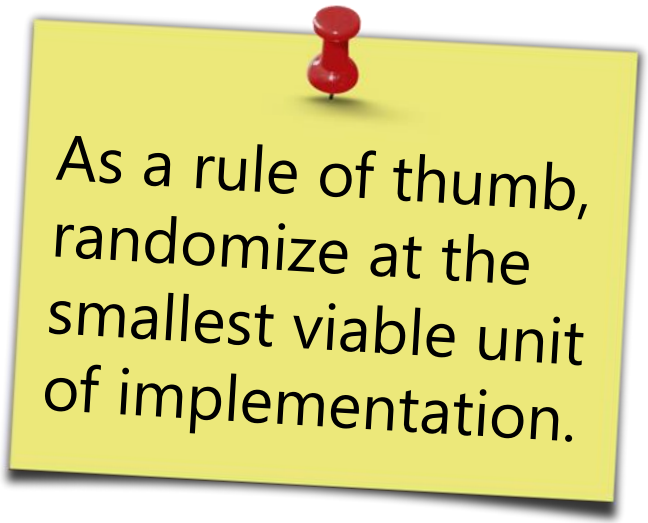
- Give each eligible unit the same chance of receiving treatment first, second, third…
- Compare those offered treatment first, with those offered later *(comparisons)*.

Randomized treatments and comparisons



Unit of Randomization

- Choose according to type of program
  - Individual/Household
  - School/Health Clinic/catchment area
  - Block/Village/Community
  - Ward/District/Region



- Keep in mind
  - Need "sufficiently large" number of units to detect minimum desired impact: Power.
  - Spillovers/contamination
  - Operational and survey costs

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

IE Methods Toolbox

What if we can't *choose*?

- It's not always possible to choose a control group. What about:
  - National programs where everyone is eligible?
  - Programs where participation is voluntary?
  - Programs where you can't exclude anyone?

Can we compare Enrolled & Not Enrolled?

Selection Bias!

Randomly offering or promoting program

- **If you can exclude some units, but can't force anyone:**
  - Offer the program to a random sub-sample (e.g. offer training to teachers in treatment schools)
  - Many will accept, some will not accept

  → Randomized offering

- **If you can't exclude anyone, and can't force anyone**:
  - Making the program available to everyone
  - But provide additional promotion, encouragement or incentives to a random sub-sample:
    - Additional Information.
    - Encouragement.
    - Incentives (small gift or prize).
    - Transport (bus fare).

  → Randomized promotion

Randomly offering or promoting program

Necessary conditions:

1. Offered/promoted and not-offered/ not-promoted groups are comparable:
   - Whether or not you offer or promote is not correlated with population characteristics
   - Guaranteed by randomization.
2. Offered/promoted group has higher enrollment in the program.
3. Offering/promotion of program does not affect outcomes directly.

Other Impact Evaluation Methods

Prospective evaluation:

1. Randomized evaluations
2. Double-difference (DD) methods

Retrospective evaluations:

3. Propensity score matching (PSM)
4. Regression discontinuity (RD) design
5. Instrumental variable (IV) methods

[and combinations of methods]

Quasi-experimental Methods
(as opposed to RCTs)

- Comparison group non-randomly constructed by evaluator

- Challenge: evaluator can never be sure if behaviour of comparison group mimics that of treatment group without program: selection bias

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

IE Methods Toolbox

Discontinuity Design

Many social programs select beneficiaries using an index or score:

Anti-poverty Programs	Targeted to households below a given poverty index/income
Pensions	Targeted to population above a certain age
Education	Scholarships targeted to students with high scores on standardized tests
Agriculture	Fertilizer program targeted to small farms less than given number of hectares)

Effect of secondary education scholarships on enrolment and later life outcomes

Goal

Improve educational and later life outcomes of poor and vulnerable youth

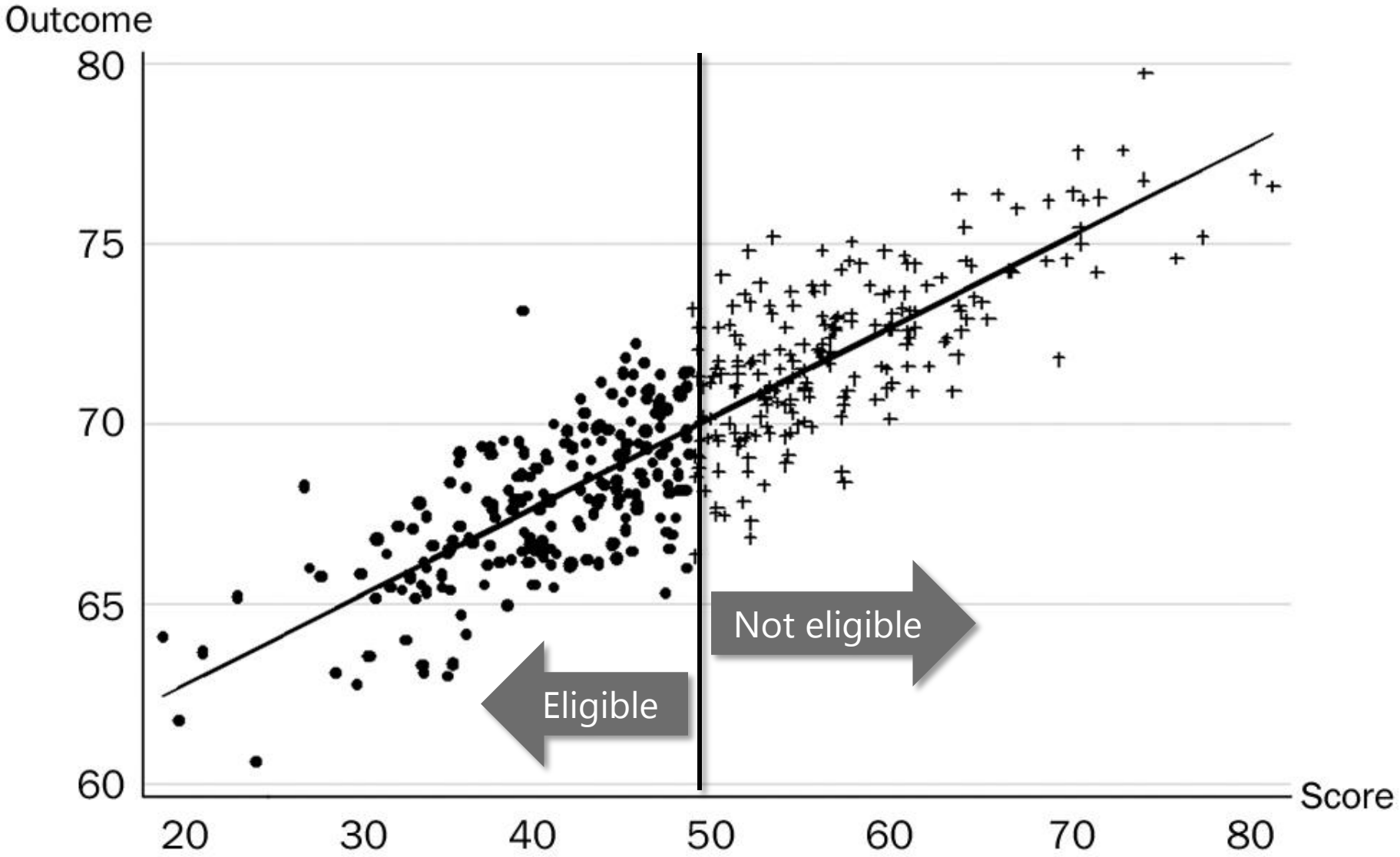
Method

- Students with a vulnerability score ≤50 are poor
- Students with a vulnerability score >50 are not poor

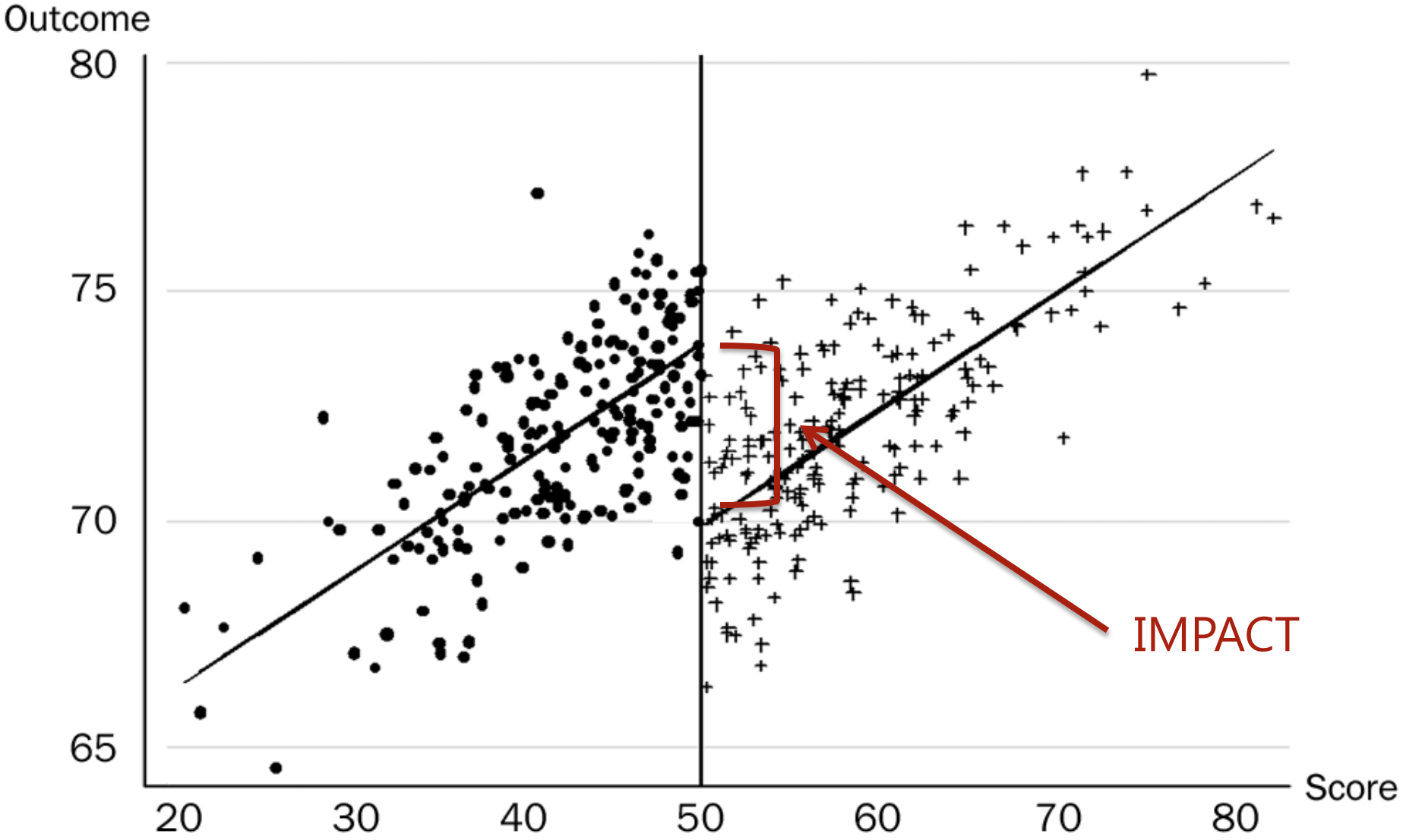
Intervention

Poor students receive scholarships

Regression Discontinuity Design-Baseline



Regression Discontinuity Design-Post Intervention



Discontinuity Design

- We have a continuous eligibility index with a defined cut-off
  - Households with a score ≤ cutoff are eligible
  - Households with a score > cutoff are not eligible
  - Or vice-versa
- Intuitive explanation of the method:
  - Units just above the cut-off point are very similar to units just below it – *good comparison*.
  - Compare outcomes *Y* for units just *above and below* the cut-off point.
- Strategy estimates a **local** average treatment effect.

For a discontinuity design, you need:

1) Continuous eligibility index
2) Clearly defines eligibility cut-off.

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

IE Methods Toolbox

Effect of per capita financing of primary schools on educational outcomes

Goal

Improve equity in education and strengthen quality of teaching

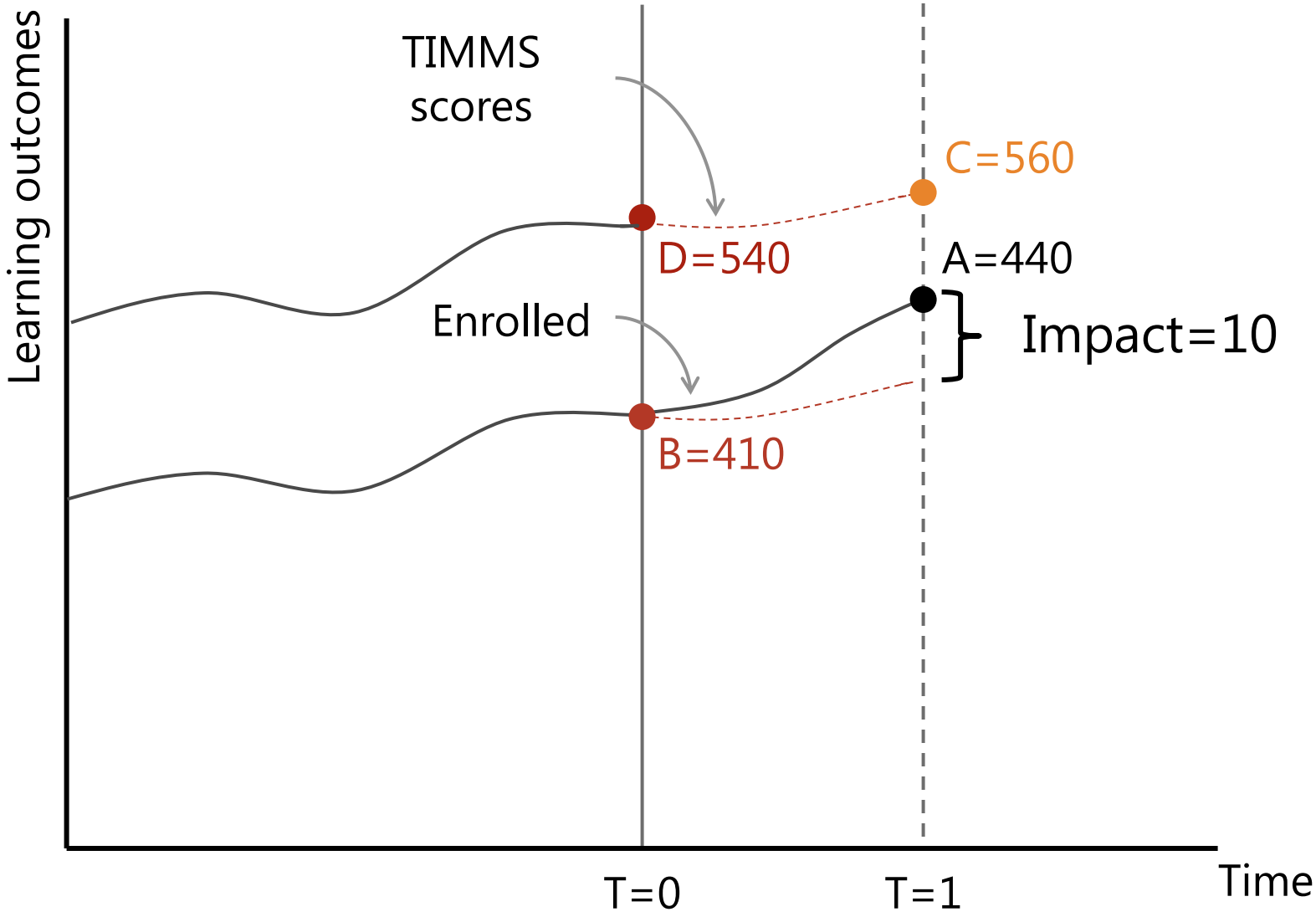
Intervention

Introduce per capita financing and decentralize decision making powers to the school level. Poorest districts selected first

Evaluation Method

Compare the change in learning outcomes of pupils in 30 pilot districts with the change in learning outcomes of pupils in 30 non-pilot districts

Impact =(A-B)-(C-D)=(A-C)-(B-D)



Difference-in-differences *(Diff-in-diff)*

Y=TIMMS test scores
P=School decentralization program

	Pilot districts		Non pilot districts	
After	440	-	560	-
Before	410		540	
Difference	**+30**	**-**	**+20**	**= 10**

Diff-in-Diff: $\text{Impact} = (Y_{t1} - Y_{t0}) - (Y_{c1} - Y_{c0})$

Keep in Mind

Difference-in-Differences

Differences in Differences combines *Enrolled & Not Enrolled* with *Before & After*.

Slope: Generate counterfactual for change in outcome

Trends –slopes- are the same in treatments and comparisons *(Fundamental assumption)*.

To test this, at least **3 observations** in time are needed:

- **2** observations **before**
- **1** observation **after**.

Choosing an IE method

- Choice of method depends on context. Rules of program operations are key because they determine who gets the program.
- The rules of program operations determine which impact evaluation method can be used (not vice versa)
- Do this at the same time as you are designing the program – for a prospective evaluation.
- We can almost always find valid comparison groups (counterfactuals) if the operational rules for selecting beneficiaries are equitable, transparent and accountable.

Who gets the program?

- Eligibility criteria
  - Are benefits targeted?
  - How are they targeted?
  - Can we rank eligible's priority?
  - Are measures good enough for fine rankings?
- Roll out

  Equal chance to go first, second, third?

The Method depends on the rules of operation

		Targeted	Universal
In Stages	Without cut-off	○ Randomization	○ Randomized Rollout
	With cut-off	○ RD/DiD ○ Match/DiD	○ RD/DiD ○ Match/DiD
Immediately	Without cut-off	○ Randomized Promotion	○ Randomized Promotion
	With cut-off	○ RD/DiD ○ Match/DiD	○ Randomized Promotion

Ethical Considerations

Do not delay benefits

Rollout base on budget/administrative constraints

Equity

Equally deserving beneficiaries deserve an equal chance of going first

Transparent & accountable method

- Give everyone eligible an equal chance
- If rank based on some criteria, then criteria should be quantitative and public

Overall Messages

Impact evaluation

Is useful for:

- Validating program design
- Adjusting program structure
- Communicating to finance ministry & civil society

Evaluation design

Derived from clear, transparent rules of program operation

A good one requires estimating the counterfactual:

- What would have happened to beneficiaries if had not received the program
- Need to know all reasons why beneficiaries g[illegible] program & others did not

Remember

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

Remember

“

To estimate impact, we need to estimate the **counterfactual.**

- what would have happened in the absence of the program and
- use comparison or control groups.

Remember

“

Choose the best evaluation method that is feasible in the program’s operational context.

HUMAN DEVELOPMENT NETWORK

Remember

> We have a **toolbox** with **5 methods** to identify good comparison groups.

References (Methods)

- 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).
  Spanish and French versions available.

- Khandker, Shahidur R., Gayatri B. Koolwal, and Samad A. Samad, 2009, *Handbook on Impact Evaluation: Quantitative Methods and Practice. Washington DC: The World Bank, 2009.*