

Measuring development effectiveness through impact evaluation

Urs Nagel

Regional Evaluation Adviser, UNICEF

Islamabad, 16 April 2015

Contents

- Towards measuring development effectiveness
- Evaluation
- Impact evaluation
- Overview of evaluation approaches
- The international debate

Towards measuring development effectiveness

Busan Partnership for Effective Development Co-operation (2011): Common Principles

- Ownership of development priorities by developing countries
 - Increased government demand for evaluation
- A focus on results: Having a sustainable impact
 - Increased focus on performance and how to measure it
- Partnerships for development: Development depends on the participation of all actors
 - Increased civil society demand for evaluation
- Transparency and shared responsibility: Accountability to all citizens
 - Increased role of evaluation in demonstrating results

2015 International Year of Evaluation

- Ban Ki-Moon said “ I welcome the designation of 2015 as the International Year of Evaluation -- the same year in which the United Nations marks its 70th anniversary and will take momentous decisions about our future prosperity, safety and well-being.
- Evaluation everywhere, and at every level, will play a key role in implementing the new development agenda.
- Evaluation is not easy. Nor is it popular. But it is essential. The current constrained budgetary climate makes it more important than ever.
- All of us share a responsibility to strengthen this important function. I look forward to working with all of you to strengthen evaluation capacity so that it can play its rightful role in building lives of dignity for all.”



The Busan commitment to development effectiveness...

- The establishment of indicators and goals at the country level defined according to the country's own priorities. These will be used to evaluate the progress made by different development actors on their commitments. The results will be made publicly available.

...implies an increasing recognition of the
role of evaluation...

- ...as a tool for accountability
- ...as a tool for learning
- ...as an integral element of managing for development results
- ...as a input for evidence-based policy-making

Evaluation

Where does evidence come from?

- Research
- Audit
- Investigation
- Monitoring
- Evaluation

Definition of the United Nations Evaluation Group (www.unevaluation.org)

“An evaluation is an assessment, as systematic and impartial as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area, institutional performance, etc.

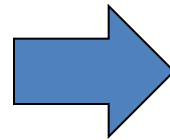
It focuses on expected and achieved accomplishments examining the results chain, processes, contextual factors and causality, in order to understand achievements or the lack thereof.

It aims at determining the relevance, impact, effectiveness, efficiency and sustainability of the interventions and contributions of the Organizations of the UN System.

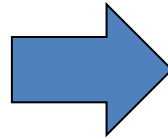
An evaluation should provide evidence-based information that is credible, reliable, and useful, enabling the timely incorporation of findings, recommendations and lessons into the decision-making processes of the Organizations of the UN System and its Members.”

Evaluation asks:

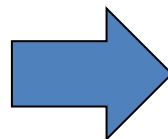
Are policies and programmes making a difference in the lives of people?



Are we doing the right thing?



Are we doing things right?



Why do the things we do work or not work?

From evidence to evidence-based policy advocacy and programming: Some issues

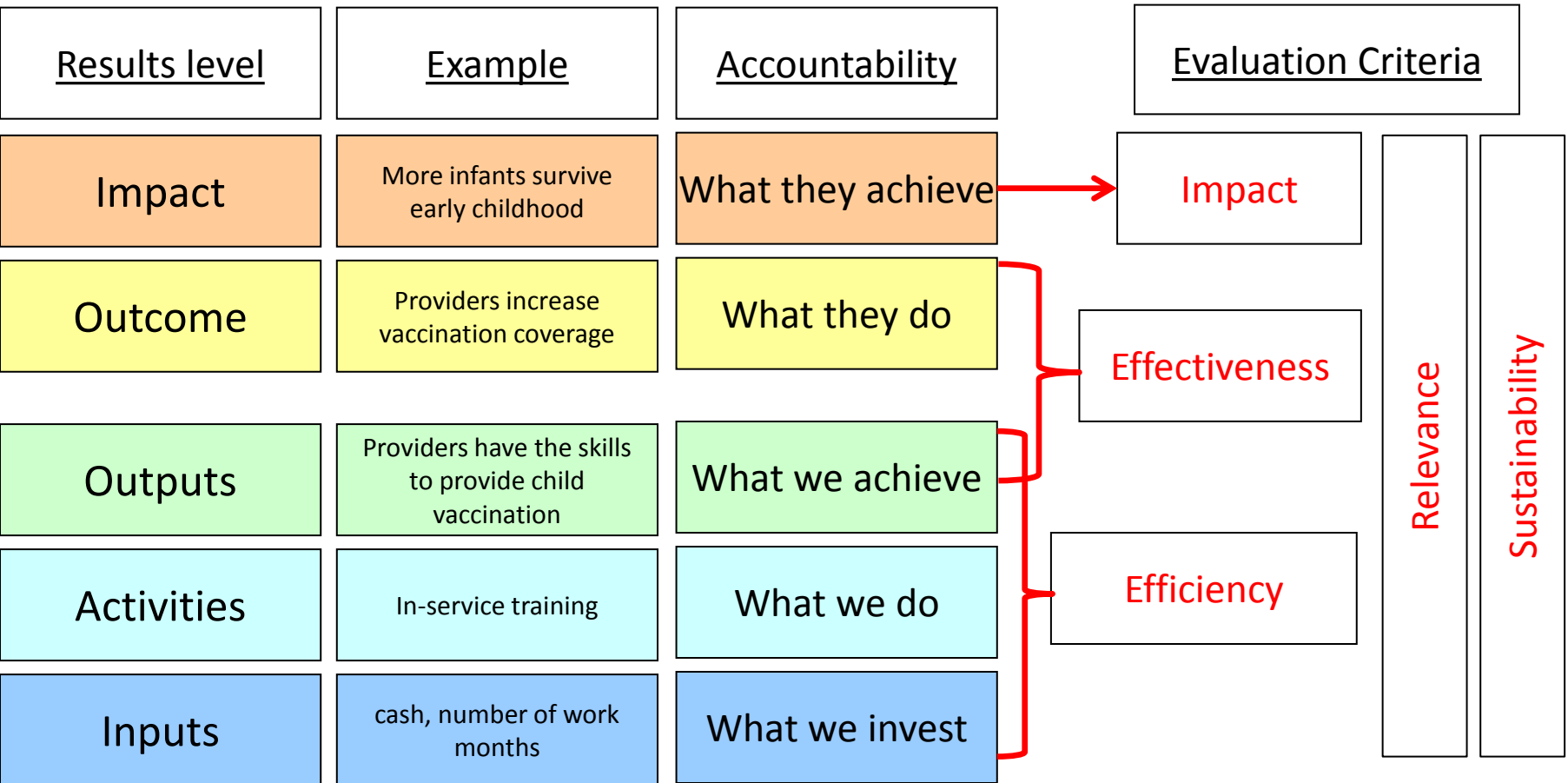
- Is it possible to 'measure' social development?
- Technical process within a political environment
- Requires a political strategy (including an understanding of political processes)

Impact

OECD DAC definition

- The positive and negative changes produced by an intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the initiative on the local social, economic, environmental and other development indicators.

Relationship between results and levels and evaluation criteria



An emerging understanding of impact evaluation

- Impact = the extent to which effects can be directly attributed to causes

<u>Results level</u>	<u>Example</u>	<u>Accountability</u>
Impact	More infants survive early childhood	What they achieve
Outcome	Providers increase vaccination coverage	What they do
Outputs	Providers have the skills to provide child vaccination	What we achieve
Activities	In-service training	What we do
Inputs	cash, number of work months	What we invest

impact evaluation
=
determining attribution

Impact evaluations seek to assess

- Positive or negative effects
- Intended and unintended effects
- The causes of the observed effects

Designing an impact evaluation

- Is an impact evaluation the best type of evaluation? How will the results be used?
- What is the theory of change?
 - Causal chain
 - Underlying assumptions
- What are the main evaluation questions?
- How evaluable is the initiative? What real-world constraints are there?
- What approaches and methods will help answer the questions in the best way?

Possible evaluation approaches

Evaluation design: Overview

- Some broader methodological considerations:
 - Experimental or quasi-experimental designs: do we have control or comparison groups?
 - Or non-experimental design?
 - Quantitative or qualitative approaches?
 - Or mixed methods?

Comparison or control group: identifying the counterfactual

A group with characteristics that are similar/(statistically) identical to the target/treatment group of your programme

Programme beneficiaries

Young men ages 15-19 living in an urban community in country X; *participate in workshops*

Comparison/control group

Young men ages 15-19 living in a different urban community in country X (preferably in a different area); *do not participate in workshops*

Why? Helps to determine if the change you measure among beneficiaries is due to your programme activities

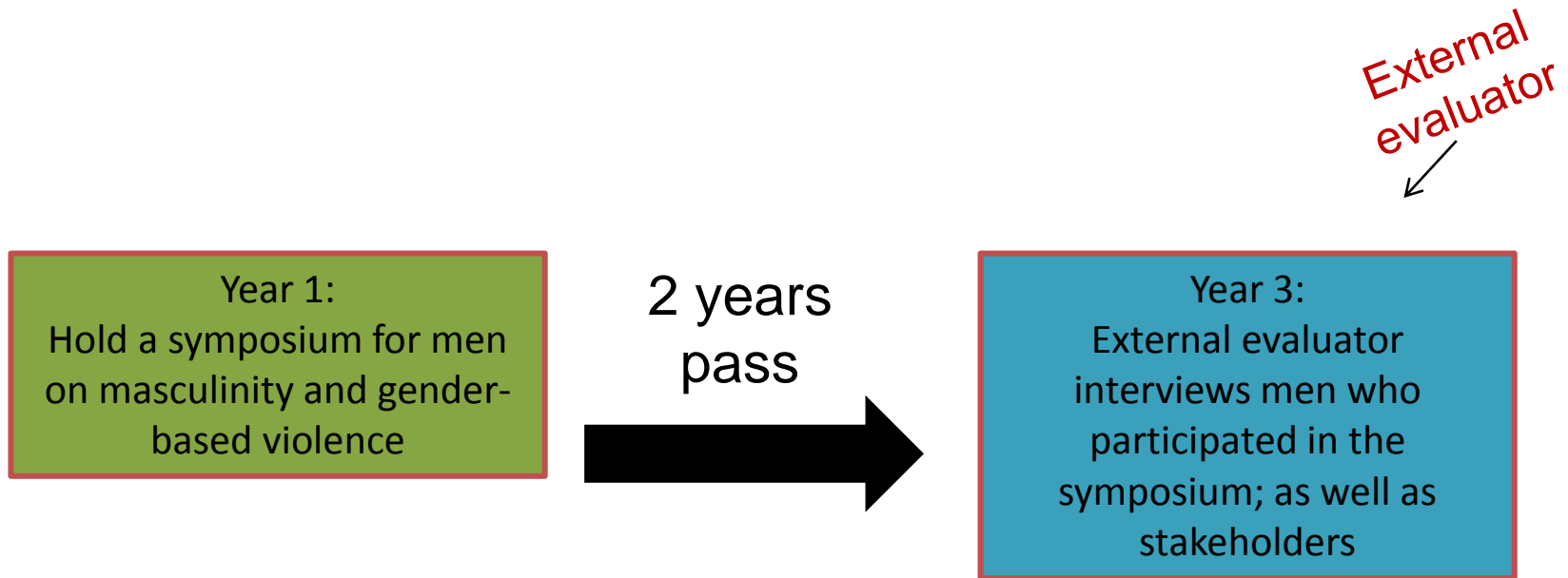
Types of evaluation designs

- i. Post-test only, no comparison group
 - ii. Pre-test, post-test, no comparison group
 - iii. Pre-test, post-test with comparison group
 - iv. Randomized controlled trial with control group
- Non-experimental designs*
- Quasi-experimental designs*
- Experimental designs*

Non-experimental designs

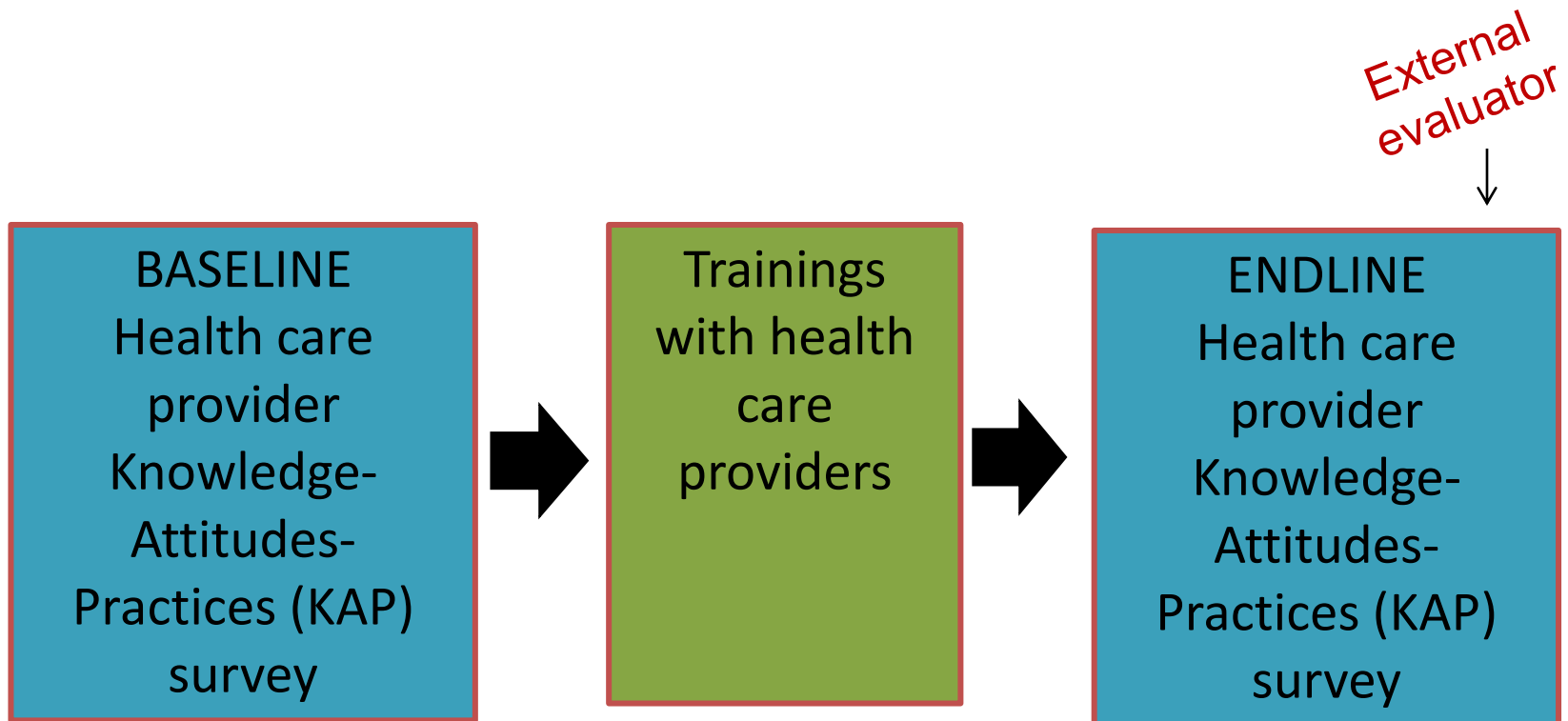
(i) Post-test only, no comparison group

Example (i): Retrospective Qualitative Study



(ii) Pre-test, post-test, no comparison group

Example (ii): IPPF/WHR



Importance of baseline data

- Provides a comparison point for changes your project aims to make
- Offers important information about the context in which the work will take place
- Provides the external evaluator with important information needed to be able to document *change over time*

Non-experimental designs: no comparison groups, but baseline information

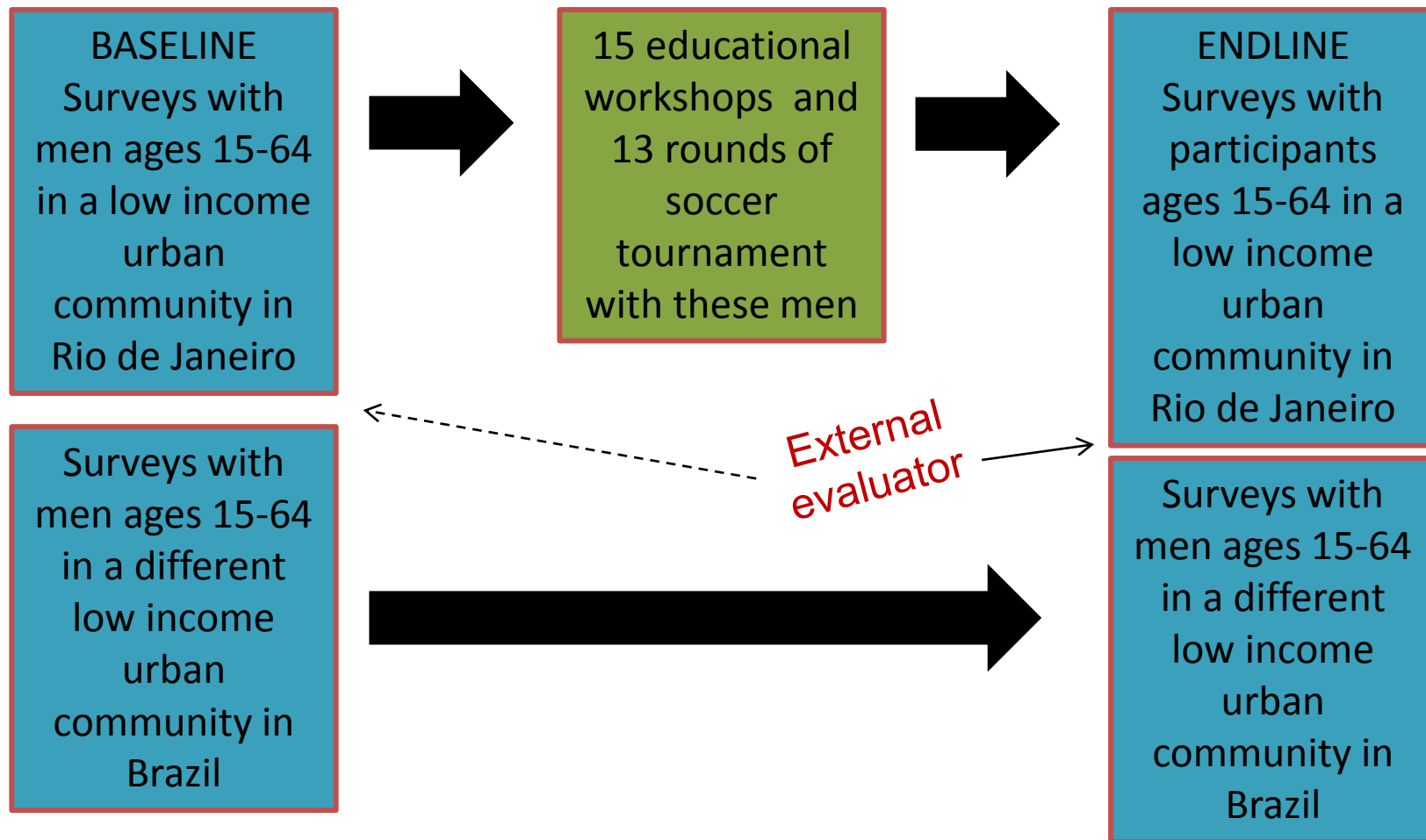
- Single and multiple case studies
- Longitudinal study
- Interrupted time series

Quasi-experimental designs

(iii) Pre-test, post-test with comparison group

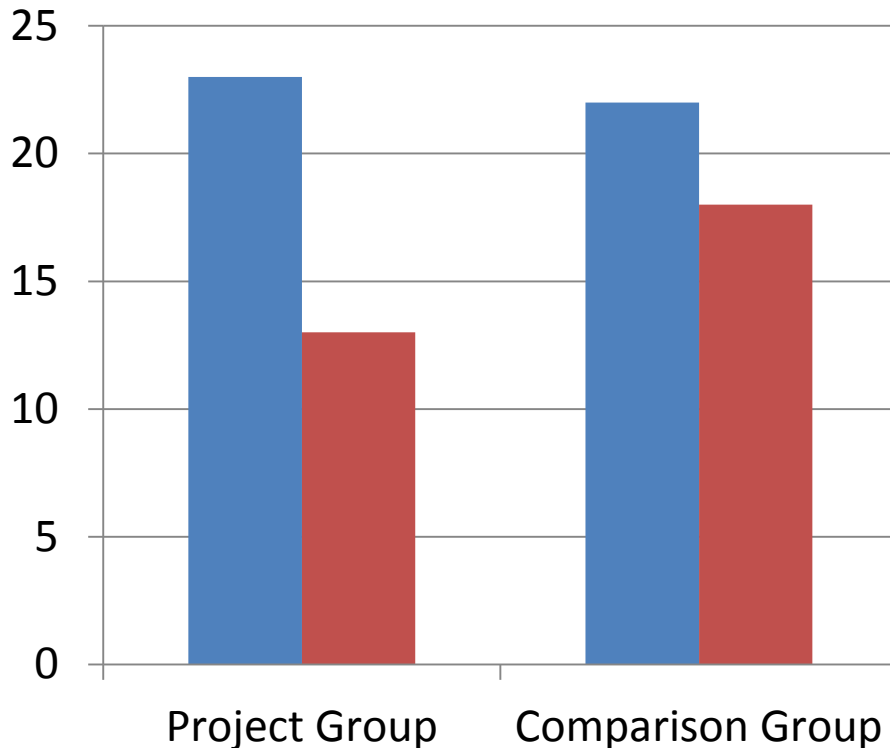
Example (iii): Engaging Men to Prevent Gender-Based Violence

Multi-country intervention and impact evaluation



(iii) Pre-test, post-test with comparison group

Example (iii): Engaging Men to Prevent Gender-Based Violence – *analyzing the difference-in-differences*
Multi-country intervention and impact evaluation



■ Baseline
■ Endline

Percentage of men who agree with the statement "There are times when a woman deserves to be beaten."

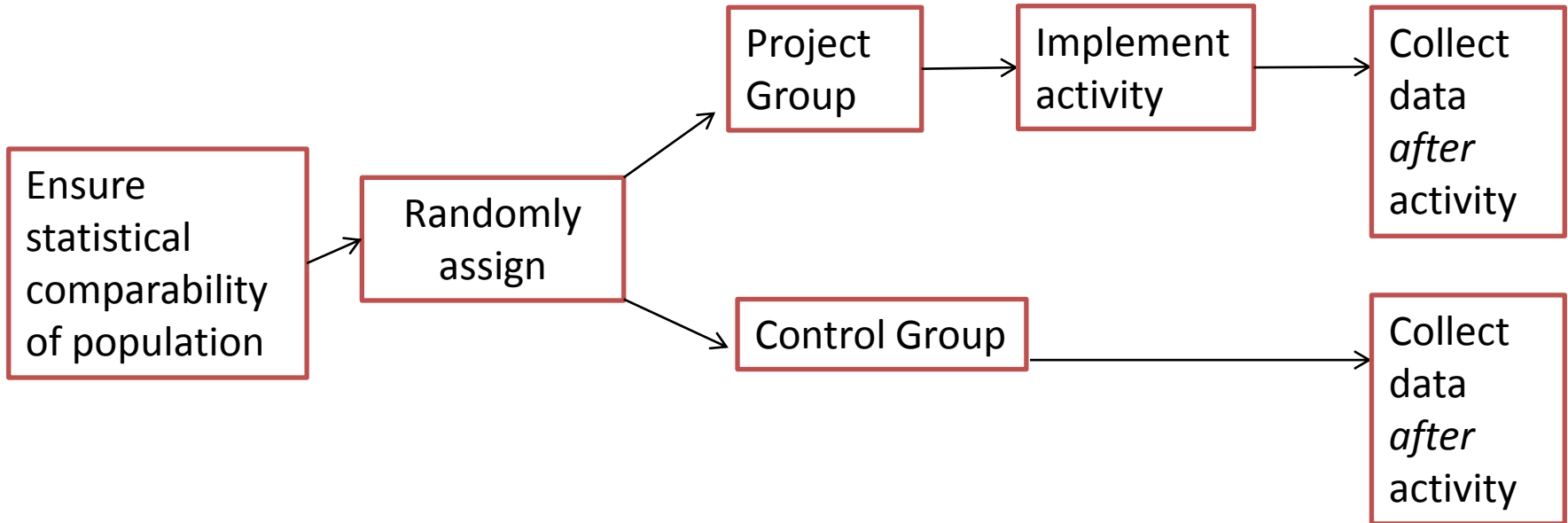
Quasi-experimental designs

- Compensating for absence of randomly-selected comparison groups:
 - Statistical (propensity score) matching
 - Regression discontinuity

Experimental designs

(iv) Randomized controlled trial (RCT)

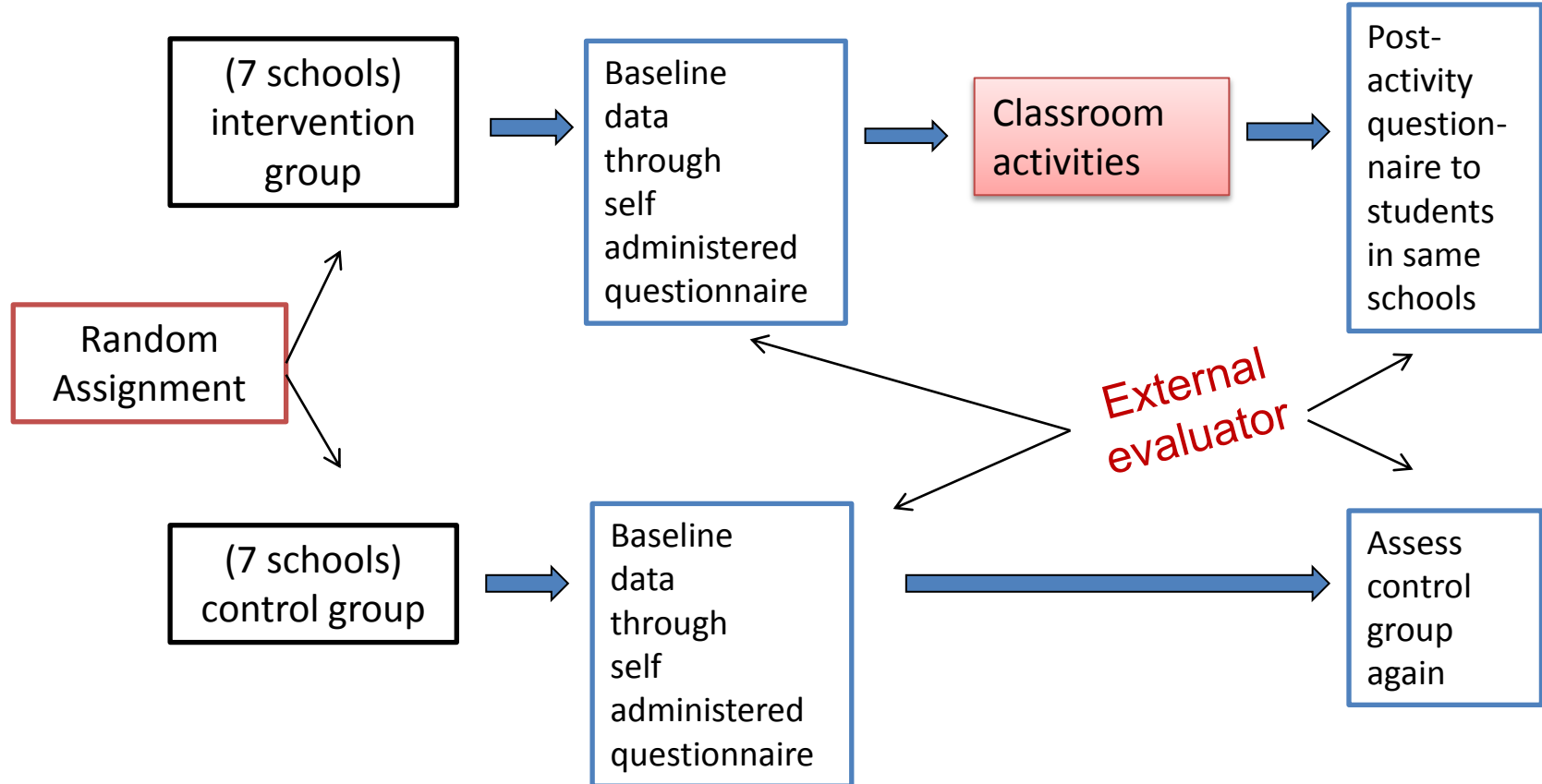
Random assignment to treatment and control group



Offers the greatest possibility of attributing outcomes/change to the intervention; resource intensive

(iv) Randomized controlled trial, pre-test post-test with control group

Example (iv): Safe Dates



Data collection methods

Quantitative and qualitative methods

- Quantitative – usually larger samples
 - Mathematical/statistical analysis
 - Usually based on surveys, aptitude and behavioural tests, etc.
- Qualitative – usually smaller samples
 - Interpretative analysis
 - Usually based on participant observation, key informant interviews, focus groups, etc.

Quantitative and qualitative methods

- Quantitative
 - Can provide strong statistical evidence on developmental change
 - Can support the determination of attribution in impact evaluations
- Qualitative
 - Can help obtain a more nuanced picture of developmental reality
 - Can provide strong evidence about delivery processes, including the implications for disadvantaged groups
 - Can capture unintended effects or results
 - Helps understand the “why” as opposed to just the “what”

The international debate on impact evaluation

Advantages of experimental and quasi-experimental designs

- RCT seen as “gold standard”: greatest statistical accuracy
- Quasi-experimental designs: second-best option – adequate statistical rigour
- Rely on quantitative methods which are often considered scientifically ‘superior’ to qualitative methods
- Provide clear answers on the net effect

Criticism of experimental and quasi-experimental designs

- Can only be used in some circumstances
- Difficult to capture complexity of development, in different circumstances
- Political reality may not permit random assignment
- Ethical issues in random assignment
- High cost of experimental and quasi-experimental methods
- Tells you about the before and after, but not about the 'why' and 'how'

Evolving practice

- Combining quantitative with qualitative methods to not only establish, but also to understand, causality
- Openness to a range of qualitative methods that seek to overcome the absence of counterfactuals and baselines
 - Importance of triangulation
 - Ensuring multiple sources of evidence
 - Establishing credibility through plausibility

Towards strengthening opportunities for impact evaluation

- Ensuring that (impact) evaluation is embedded in programming (RBM)
 - Establishing clear theories of change ex ante
 - Establishing baselines as a norm
 - Putting in place rigorous monitoring systems
- Ensuring ownership: government, civil society, development partners
- Advocating the value of evidence-based policy-making and programming

Concluding observations

- Impact evaluation means different things to different people and organizations – important to clarify intention and context
- Different methods ‘measure’ different things – the complexity of reality requires the use of complimentary methods
- The availability of rigorous evidence does not mean it will be used – need to understand the role of evidence in the political debate

Thank you