

IMPACT EVALUATION AND ITS DISCONTENTS

Rizal Adi Prima, Evaluation Coordinator

National Team For Accelerating Poverty Reduction (TNP2K)

National Conference On Monitoring And Evaluation

Bappenas, JAKARTA, 28 – 29 August 2014

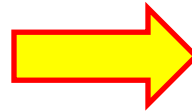
Outline

- Why evaluation
 - The importance of evaluation
 - Counterfactuals
 - Identification
- What kind of policy should be using impact evaluation
- Step by Step
- Evaluation in practice
 - The need of baseline-endline
 - Power calculation

Policy Question

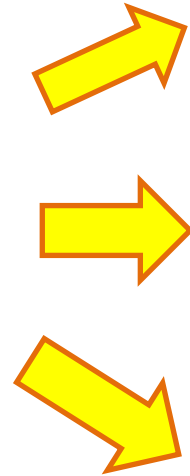
We want to move him from **here...**

... to **here ...**



Policy Question

...with available resources...



...how do we choose the most appropriate solution...

...and how do we **know** we have met our goals...?

How do we **show**?

About a program: what do we want to know?

- Does a program **bring the expected change**, which won't happen without it?
- For every \$1 spent on the program, **how big** is the change?
- Are there other, **cheaper options** with similar results?
- Does a program **bring unintended consequences**?

Why Evaluate?

- Need **EVIDENCE** on what works
 - Rigorous evaluation set up provide you with hard evidence on the effectiveness of a program
- Improve Program/Policy for better **RESULTS**
 - Impact evaluation can also be used to evaluate the current design (ex : eligibility, benefits, timing)
 - Impact evaluation c
- The key for the sustainability of a program is the **INFORMATION** regarding the quality of implementation

When to use Impact Evaluation?

Evaluate impact when project/Program is:

- Innovative
- Replicable/scalable
- Strategically relevant for reducing poverty
- Evaluation will fill knowledge gap
- Substantial policy impact

Definition

Evaluation An *objective and systematic appraisal* about the success of a program or policy, in its relation to its implementation and results (output and impact).

The aim is to decide the relevance and achievement of the programs, its efficiency, credible and useful information for policy making.

- ✓ What is the effect of the program on outcomes?
 - ✓ How much better off are the beneficiaries because of the program/policy?
 - ✓ How would outcomes change if we change the program design?
 - ✓ Is the program Cost effective?
-
- The **Big question** is how to design a good impact evaluation?

What is the Impact of...

giving Mr X



medicines



on Mr X's temperature?

“With-Without”?

Mr X



Temperature: 38 C

Mr X can only choose to take
the medicine or not but
NOT BOTH

Mr X



Temperature:
Unobserved

“With-Without”?

Mr X



Temperature:
Unobserved

IMPACT=Unobserved
Temperature may go down
due to his own immune
system

Mr X



Temperature: 38

The Perfect *Clone*!

Mr X



Temperature: 38 C

Mr X's Clone



Temperature: 41 C

$$\text{IMPACT} = 38 - 41 = -3 \text{ C}$$

But we don't have clones...(yet?)

Why **NOT** “Before-After”?

- Can we assume that the medicine was **the only** factor influencing the temperature of Mr X **over time**?
 - He might have eaten different food or nutrition over time; or different sleep pattern; **or other unobserved things**



Control / comparison “group”

- There might be someone who has **similar characteristics** with Mr X (gender, age, eating habit, weight, height, blood pressure, medical history, etc) and **also having similar medical problem: temperature = 41 C** ---- say: Mr Y
 - **The only difference is that Mr X takes the medicine, but NOT Mr Y**
- Now we could assess the impact of the medicine on lowering the body temperature



Using *Counterfactual*

Mr X



Temperature: 38 C

Mr Y



the state of "Mr X" would have been in the absence of medicine

Temperature: 41 C

$$\text{IMPACT} = 38 - 41 = -3 \text{ C}$$

Evaluation design

- Experimental
 - Randomly assign participants into **control** and **treatment**
- Quasi-experimental
 - Exploits existing events/information to divide observations into **treatment** and **comparison**
- Non-experimental
 - Do not assign and compare groups, looks at identifying characteristics, frequency, and associations
- *Not to be confused with: quantitative vs qualitative – which are **methods** or **techniques***

Choosing your IE method(s)

Choose the **best possible design** given the operational context:

Best Design



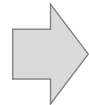
- Best comparison group you can find

Have we controlled for everything?



- Internal validity
- Good comparison group

Is the result valid for *everyone*?

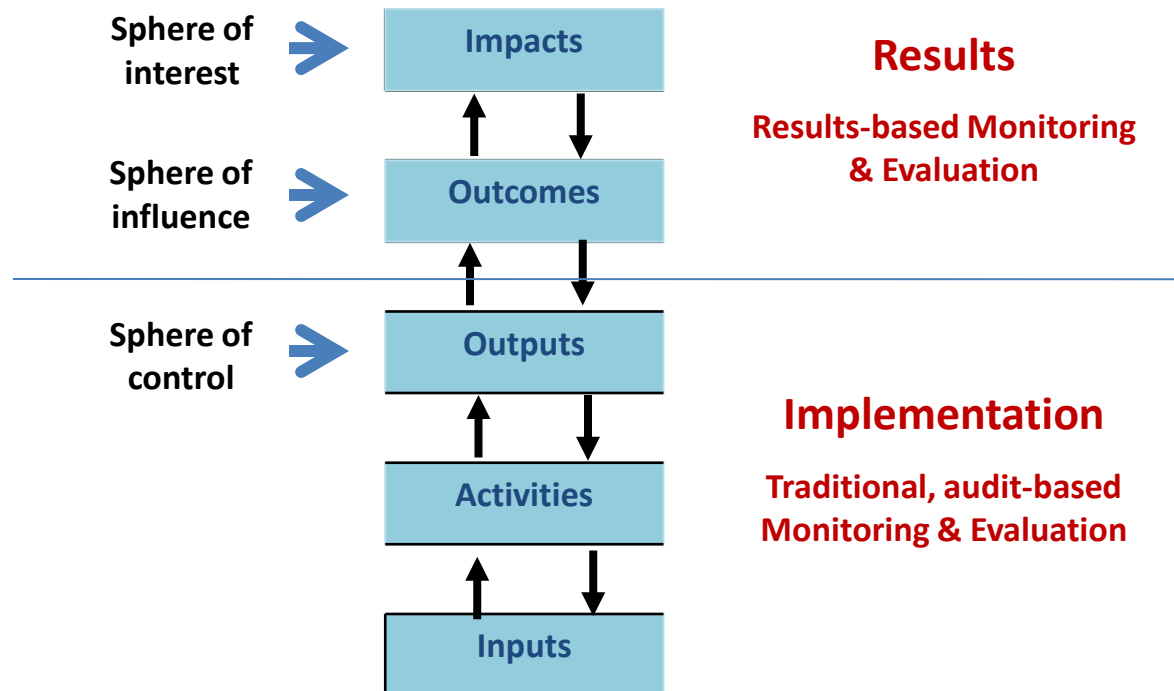


- External validity
- Local versus global treatment effect
- Evaluation results apply to population we're interested in

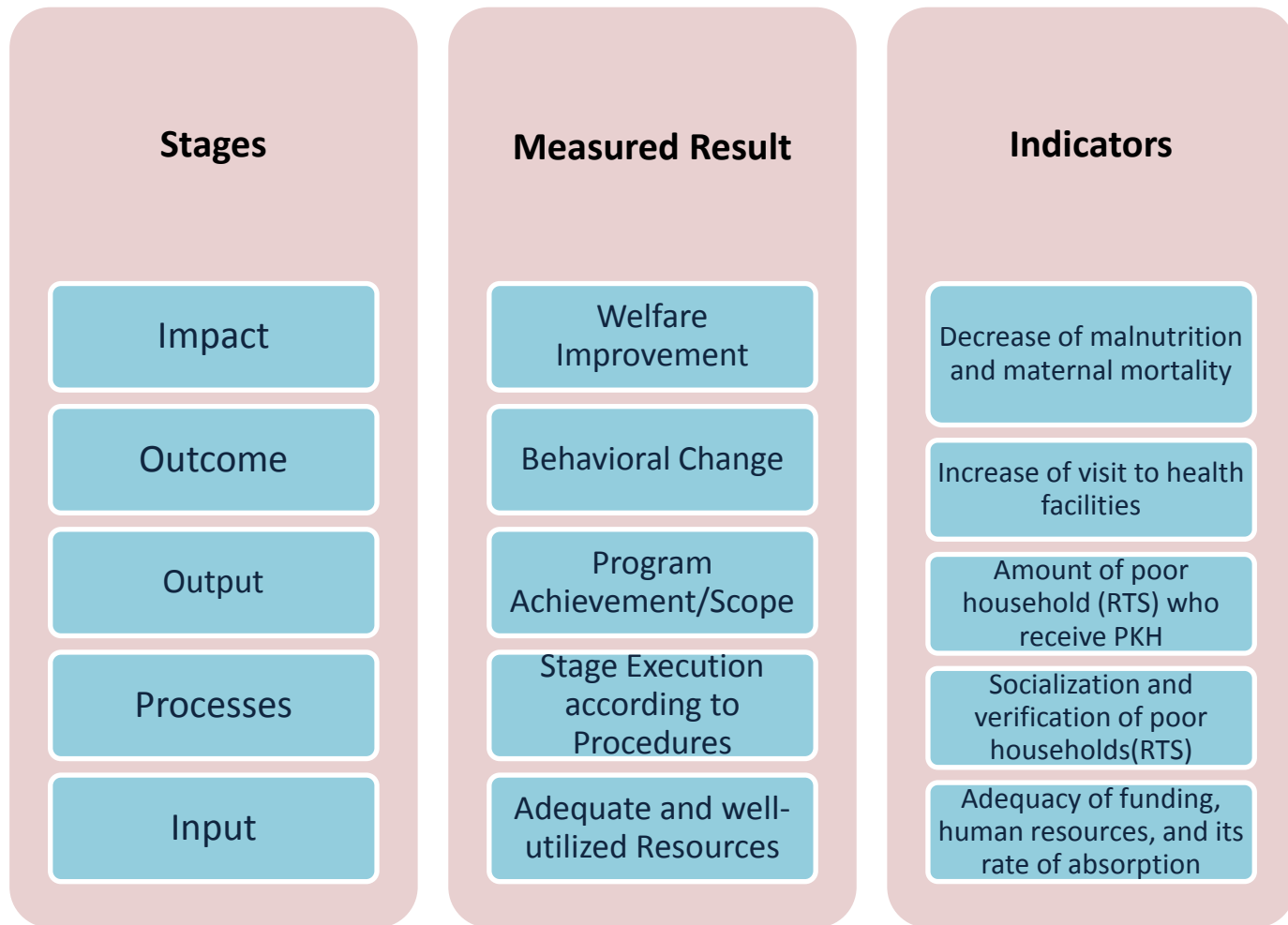
Steps for Impact Evaluation

1. Define the problem/motivation
2. Define the expected **results**
3. Gather **baseline** data/information
4. Gather Endline/Midline data/Information
5. Evaluate the results

Step 2. Different levels of results – “Theory of Change”



Example: Program Keluarga Harapan



Step 3. Select key indicators

- **Indicator:** a specific variable, that when tracked systematically over time, indicates progress (or lack thereof) toward an outcome or impact
- Indicators: translating an abstract concept into a more concrete and measured one
 - “Higher competence of labor force” → “% of high school graduates passing standardized test”
- SMART: **S**pecific **M**easurable **A**chievable
Relevant **T**ime-bound

Step 4. Gather baseline data

Prior to intervention we collect information from households designated as the beneficiaries of the program

- Primary data Collection
 - For Baseline Survey of PKH evaluation we gathered information almost 14.000 household
 - For BSM baseline Survey we surveyed 5.020 households and 12.000 school aged children
- Administrative data can also be used as a baseline, but we have to make sure the quality of the information
- Alternatively design a **Quasi-Experimental evaluation**

Step 5. Gather Endline Data

- Collect information of both treatment and comparison group. To capture the state of the two after the introduction of the program
- Challenges :
 - Not all the designated beneficiaries eventually comply/exposed to the program
 - Tracking Individuals and/or Household has their own difficulties

Step 6. Evaluating results

- Descriptive questions – “What is”
 - What proportion of women participated in the program?
 - How often does the household visit a health provider?
- Cause-and-Effect questions – looking for causality
 - Did distributing School Vouchers Increase School participation?

Evaluation Can be Costly...

- Evaluation Design **ideally** need a baseline data, to capture the state of treatment and comparison group prior to an intervention
- A baseline-Endline data can be generated from administrative data or a carefully designed Primary data collected for program evaluation.
 - Power of the Test and Effect Size matters
 - Field Survey can be time consuming and expensive
- Alternatively we can use quasi-experimental set up to artificially construct a comparison group
 - Open the possibilities of using a secondary data set

On Going Impact Evaluation

- TNP2K with the support of PRSF has been designing impact evaluation for Government anti poverty programs :
 - PNPM, using a data of SEDAP
 - Micro Credit, a non experimental design evaluation using secondary data set of VIMK
 - PKH, Baseline-Endline with Instrumental Variable adjustment of an RCT set up
 - BSM, using Baseline-Endline exploiting a pipeline method

Checklist

- What do we know about our program?
 - What is the output? Outcome? Impact?
- What do we want to know from our program?
 - The process/how it actually works?
 - The outputs – what it has been deliver?
 - The outcome – what changes it has brought? For whom?
 - The impact – what would have happened without our program?
- What do we have to answer them?
 - Data that we have?
 - Data that we want to gather?
 - Data that we are able to gather? → What resources do we have?

THANK YOU

References

- Duflo, E. et.al (2006), “Using Randomization in Development Economics Research: A Toolkit, *NBER Technical Working Paper Series 333*
- Khandker, et. al (2010), *Handbook on Impact Evaluation: Quantitative Methods and Practices*, World Bank
- Gertler, et.al (2011), *Impact Evaluation in Practice (including supplementary material for presentation)*, World Bank
- Perdana, Ari, (2014)*Monitoring and Evaluation for Social Program. Yayasan Indonesia Mengajar*
- *International Program for Development Evaluation Training, 2012*
- Training Materials, (2012) *Turning Promises to Evidence” The Regional Impact Evaluation Workshop KDI school and the World Bank, Seoul.*
- Pokja Monev TNP2K (2014), *Evaluasi Dampak Program Keluarga Harapan*, materi presentasi internal
- Purnagunawan, W. (2014), *Introduction to Impact Evaluation: Basic Theory and Concept*, Training Materials for CEDS Econ Training Fest, Bandung
- World Bank (2011), *PKH Evidence and Policy Implications: summary of Results from Impact Evaluation, Operation Analysis, and Spot Checks*, Materials for Diskusi Pokja Kebijakan Monev TNP2K
- World Bank (2011), *Program Keluarga Harapan: Main Findings from the Impact Evaluation of Indonesia’s pilot Household Conditional Cash Transfer Program*
- World Bank website
 - <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20188242~menuPK:412148~pagePK:148956~piPK:216618~theSitePK:384329,00.html#quasi>