Evaluation Design & Method: Case Study of Program Keluarga Harapan

MONITORING & EVALUATION WORKING GROUP - TNP2K
Jakarta, 27 Agustus 2014
Evaluation

- **Operational evaluation** examines how effectively programs were implemented and whether there are gaps between planned and realized outcomes.
- **Impact evaluation** studies whether the changes in well-being are indeed due to the program intervention and not to other factors.
  - Specifically, impact evaluation tries to determine whether it is possible to identify the program effect and to what extent the measured effect can be attributed to the program and not to some other causes.

Impact Evaluation

SOME BACKGROUND & ILLUSTRATION
Duflo, et.al (2006)

“At a given point in time, an individual is either exposed to the program or not. Comparing the same individual over time will not, in most cases, give a reliable estimate of the program's impact since other factors that affect outcomes may have changed since the program was introduced.”
“We cannot, therefore, obtain an estimate of the impact of the program on a given individual. We can, however, obtain the average impact of a program, policy, or variable (we will refer to this as a treatment, below) on a group of individuals by comparing them to a similar group of individuals who were not exposed to the program.”
In reality, use statistics

Treatment

Average of outcomes = 10 units

Comparison

Average of outcomes = 3 units

IMPACT = 10 - 3 = 7 units
Estimating impact of $P$ on $Y$

$$\alpha = (Y \mid P=1) - (Y \mid P=0)$$

**OBSERVE**  ($Y \mid P=1$)  
Outcome with treatment

**ESTIMATE**  ($Y \mid P=0$)  
The Counterfactual

**IMPACT** =  Outcome with treatment - counterfactual

Use *comparison* or *control* group
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.
Counterfactual Criteria

- Treated & Counterfactual
  1. Have identical characteristics,
  2. Except for benefiting from the intervention.

- No other reason for differences in outcomes of treated and counterfactual: Only reason for the difference in outcomes is due to the intervention
Evaluation Design

• Evaluation designs are determined by the choice of methods used to identify a comparison/control group, or in other words, a group of non-participants in a program or a project.

• This comparison/control group should be as similar to the target group as possible, but for the fact that its members do not participate in a program or receive the intervention.
Evaluation Design

• Evaluation designs can be broadly classified into three categories: experimental, quasi-experimental and non-experimental.
  – The term control group is used when the evaluation employs an experimental design and the term comparison group is associated with a quasi-experimental design.
  – In non-experimental design, program participants are compared to non-participants by controlling statistically for differences between participants and non-participants.

• These three evaluation designs vary in feasibility, cost, the degree of clarity and validity of results, and the degree of selection bias.
Selection Bias – unobserved characteristics

Outcome changes observed among these nonrandom groups of individuals would indicate the program impact on motivated participants, but may not reflect how the program on average would affect the target population.
What's wrong?

1. **Selection bias**: People choose to participate for specific reasons

2. Many times **reasons** are related to the **outcome of interest**
   - Job Training: ability and earning
   - Health Insurance: health status and medical expenditures

3. Cannot separately identify impact of the program from these other factors/reasons
Possible Solutions

- Need to guarantee **comparability** of treatment and control groups.

- **ONLY** remaining difference is intervention.
Experimental (randomized)

- Randomized Evaluations go by many names
  - Randomized Controlled Trials
  - Social Experiments
  - Random Assignment Studies
  - Randomized Field Trials
  - Randomized Controlled Experiments

- RCT can solve selection bias
Randomized treatments and comparisons

1. Population

2. Evaluation sample

3. Randomize treatment

As a rule of thumb, randomize at the smallest viable unit of implementation.

External Validity

Internal Validity

= Ineligible

= Eligible

Comparison

Treatment
RCT and Selection Bias

• The problem of selection bias arises because of missing data on the common factors affecting both participation and outcomes.
  – In theory, randomized or experimental evaluation is free from the bias problem whereas the problem is practically unavoidable when non-experimental data are employed.
Keep in Mind

Randomized Assignment

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

We have identified the perfect clone.

Randomized beneficiary  Randomized comparison

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

Most pilots and new programs fall into this category.
Case Study

PROGRAM KELUARGA HARAPAN (PKH)
IMPACT EVALUATION DESIGN AND ESTIMATION METHOD
Program Keluarga Harapan

• First household-based conditional cash transfer program, started in 2007
  – Community based: PNPM Generasi
  – Pilot project for 3 years

• The program is intended to improve the welfare of extremely poor households by providing them with quarterly cash transfers
Program Keluarga Harapan

• At the same time, the program is designed to break the transmission of poverty to next generations by encouraging families to increase their use of public services to, over time, improve the health and education outcomes of their children
  – providing the transfer only to households with pregnant women and/or children, provided that they fulfill specific health and education-related obligations.
## Assistance (WB, 2011)

<table>
<thead>
<tr>
<th>Category</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed cash transfer</td>
<td>200,000</td>
</tr>
<tr>
<td>Cash transfer for per household with</td>
<td></td>
</tr>
<tr>
<td>a. Child aged less than 6 years</td>
<td>800,000</td>
</tr>
<tr>
<td>b. Pregnant or lactating mother</td>
<td>800,000</td>
</tr>
<tr>
<td>c. Children of primary-school age</td>
<td>400,000</td>
</tr>
<tr>
<td>d. Children of secondary-school age</td>
<td>800,000</td>
</tr>
<tr>
<td>Minimum transfer per household</td>
<td>600,000</td>
</tr>
<tr>
<td>Maximum transfer per household</td>
<td>2,200,000</td>
</tr>
</tbody>
</table>
## ASSISTANCE (KEMENSOS, 2014)

<table>
<thead>
<tr>
<th>NO</th>
<th>NOMINAL BANTUAN/TAHUN</th>
<th>BANTUAN TETAP</th>
<th>KOMPONEN PKH</th>
<th>KETERANGAN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>BUMIL/NIFAS/BALITA</td>
<td>USIA SD</td>
</tr>
<tr>
<td>1</td>
<td>800.000,-</td>
<td>300.000,-</td>
<td>-</td>
<td>500.000,-</td>
</tr>
<tr>
<td>2</td>
<td>1.300.000,-</td>
<td>300.000,-</td>
<td>1.000.000,-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>1.000.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bila ada 2 anak SD</td>
</tr>
<tr>
<td>3</td>
<td>1.800.000,-</td>
<td>300.000,-</td>
<td>1.000.000,-</td>
<td>500.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>500.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.500.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bila ada 3 anak SD</td>
</tr>
<tr>
<td>4</td>
<td>2.300.000,-</td>
<td>300.000,-</td>
<td>1.000.000,-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.000.000,-</td>
<td>1.000.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>1.000.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bila ada 2 anak SMP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bila ada 2 anak SD</td>
</tr>
<tr>
<td>5</td>
<td>2.800.000,-</td>
<td>300.000,-</td>
<td>1.000.000,-</td>
<td>500.000,-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sumber: kemensos (2014)
Results Chain of PKH

**Inputs**
- Cash transfer; Pay staff; Staff to deliver service; Check compliance

**Activities**
- Collect eligibility data; set up services; check compliance
- Provide fund for cash transfer; system to check compliance
- Increase use of education and health services; increase spending on goods
- Reduction in current poverty & future poverty, increase in health status

**Outputs**

- [Budgeting](#)
- [Staffing](#)
- [Data collection](#)
- [Train staff](#)
- [Explain](#)
- [Cash transfer delivered](#)
- [Health and educ. Services](#)
- [Data collection](#)
- [Higher school enrolment](#)
- [Higher use of health services](#)
- [Higher years of education](#)
- [Better health](#)
- [Lower poverty](#)

**Results (Supply + Behavior)**

Activities of implementing agencies (SUPPLY SIDE)
PKH and PNPM Generasi

Figure 2.1  
Map of Districts Piloting the Household Conditional Cash Transfer Program

North Sulawesi (5 districts)
- 37 treatment sub-districts
- 31 control sub-districts

Gorontalo (2 districts)
- 4 treatment sub-districts
- 5 control sub-districts

DKI Jakarta (1 district)
- 1 treatment sub-district
- 3 control sub-districts

East Java (21 districts)
- 114 treatment sub-districts
- 192 control sub-districts

NTT (7 districts)
- 38 treatment sub-districts
- 42 control sub-districts

West Java (13 districts)
- 64 treatment sub-districts
- 56 control sub-districts

Note: World Bank, based on data from the Ministry of Social Welfare (Kemensos)

Sumber: World Bank, 2011a
PKH Recently

Sumber: kemensos
PKH Impact Evaluation Flow

**Baseline**
Measurement of outcome prior to intervention → *to improve program design*

**Mid-line**
Measurement of outcome after 2 years of intervention

**End-line**
Measurement of outcome after 6 years of intervention

- **BASELINE SURVEY**
- **FOLLOW-UP SURVEY (1)**
- **FOLLOW-UP SURVEY (2)**

- **Time**
  - 2007
  - 2009
  - 2013

- **INTERVENTION**

Ex-ante

Ex-post
Area Selection and Sub-District Randomization

Figure 2.2 Area selection and sub-district randomization

- **Provincial**
  - i) Provincial government willingness
  - ii) Representation of diverse characteristics

- **District**
  - 80% poorest districts (based on transition rate, malnutrition rate, poverty rate)
  - 20% richest districts (based on transition rate, malnutrition rate, poverty rate)
  - KDP districts in 5 PNPM provinces
  - i) Non-KDP eligible districts in 5 PNPM provinces
  - ii) Districts in DKI Jakarta and West Sumatra
  - Districts selected for PNPM-Generasi: all in Gorontalo and N. Sulawesi, random in other
  - Districts not selected for PNPM-Generasi

- **Sub-district**
  - Sub-district not supply ready
  - Sub-district supply ready
  - PKH treatment group: Randomly select 329 sub-districts for PKH (until provincial quota is filled)
  - PKH control group: Randomly select 259 sub-districts

Based on: Sparrow et al., World Bank 2008.
The PKH *kecamatan* and household-selection processes:

**Step 1:** Determine universe of eligible *kecamatan*

**Step 2:** Choose Evaluation sample

**Step 3:** Randomize program over eligible *kecamatan*

**Step 4:** Select direct beneficiaries (not random)

*Comparison*

- **External Validity**
- **Internal Validity**

*Ineligible* according to demographics and/or selection process

*Eligible* according to demographics and selection process
EVALUATION DESIGN: PKH

• RCT, 360 kecamatan
  – Eligibility of Kecamatan was based on, for example: malnutrition prevalence characteristics, poverty rate, drop-out rate, supply side readiness (education, health)
  – Treatment and control areas were selected randomly from the list of eligible kecamatans
  – Household sample was selected from the list of eligible households in the selected kecamatans
EVALUATION DESIGN: PKH

• However:
  – There were conversion (from control-kecamatans to treatment kecamatans) in 2009 & 2013
    • Non-random conversion in control kecamatans, introducing selection as well as endogeneity bias
    • Statistical power may deteriorate
Baseline and Endline Samples

**Baseline Survey Sampling (2007)**
- 6 provinces, 360 Kecamatan, 14,400 HH
- 180 treatment Kecamatan (PKH)
- 180 Kecamatan was allocated as control

**End-line Survey Sampling (2013)**
- 6 provinces, 450 Kecamatan, 18,000 HH
- 227 treatment Kecamatan, – 179 of which are baseline panel
- 110 + 76 Kecamatan were used as control
Sample Status (2013)

PKH treatment areas randomly selected
769 sub-districts

Treatment areas sample randomly selected (stratified by urban/rural)
180 sub-districts

PKH control areas randomly selected
316 sub-districts

Control areas sample randomly selected (stratified by urban/rural)
180 sub-districts

Treatment Area 1
PKH not implemented (as of 2013): 1 sub-districts

Treatment Area 2
PKH implementation started in 2007/2009: 179 sub-districts

Control Area 1
PKH was not implemented: 110 sub-districts

Control Area 2
Converted into PKH treatment area: 70 sub-districts

39% terkonversi menjadi Kec. PKH
Estimation Strategy

• Instrumental Variable (IV) regression

\[ y_{it} = \beta_0 + \delta_0 t_{it} + \beta_1 PKH_{it}^K + \delta_1 t_{it} \times PKH_{it}^K + X_{it}' \gamma + \varepsilon_{it} \]

• IV is used to overcome the potential endogeneity as the result of non-random conversion

• Instruments used is initial treatment status (lottery) of the PKH kecamatan.
Strategi Estimasi: *Placement Effect*

**PKH treatment areas randomly selected**
- 769 sub-districts

**PKH control areas randomly selected**
- 316 sub-districts

**Treatment areas sample randomly selected (stratified by urban/rural)**
- 180 sub-districts

**Treatment Area 2**
- PKH implementation started in 2007/2009: 179 sub-districts

**Control Area 1**
- PKH was not implemented: 110 sub-districts

**Control Area 2**
- Converted into PKH treatment area: 70 sub-districts

**Control areas sample randomly selected (stratified by urban/rural)**
- 180 sub-districts

**Group A**
- PKH beneficiaries in treatment areas:
  - 3,175 HHs

**Group B**
- Non-beneficiaries in treatment areas:
  - 4,670 HHs

**Treatment Area 1**
- PKH not implemented (as of 2013): 1 sub-district with 50 HHs

4,744 HHs 842 + 2,225 HHs
Strategi Estimasi: Participation Effect

PKH treatment areas randomly selected 769 sub-districts

PKH control areas randomly selected 316 sub-districts

Treatment areas sample randomly selected (stratified by urban/rural) 180 sub-districts

Control areas sample randomly selected (stratified by urban/rural) 180 sub-districts

Treatment Area 1
PKH not implemented (as of 2013): 1 sub-districts

Treatment Area 2
PKH implementation started in 2007/2009: 179 sub-districts

Control Area 1
PKH was not implemented: 110 sub-districts

Control Area 2
Converted into PKH treatment area: 70 sub-districts

50 HHs

4,744 HHs

842 + 2,225 HHs

Group A
PKH beneficiaries in treatment areas:
3,175 HHs

Group B
Non-beneficiaries in treatment areas:
4,670 HHs
Spill-Over Effect

PKH treatment areas randomly selected
769 sub-districts

PKH control areas randomly selected
316 sub-districts

Treatment areas sample randomly selected (stratified by urban/rural)
180 sub-districts

Control areas sample randomly selected (stratified by urban/rural)
180 sub-districts

Treatment Area 1
PKH not implemented (as of 2013): 1 sub-districts
179 sub-districts

Treatment Area 2
PKH implementation started in 2007/2009:
Control Area 1
PKH was not implemented: 110 sub-districts

Control Area 2
Converted into PKH treatment area: 70 sub-districts

Group A
PKH beneficiaries in treatment areas:
3,175 HHs

Group B
Non-beneficiaries in treatment areas:
4,670 HHs

50 HHs

4,744 HHs

842 + 2,225 HHs
Note

• PKH has some significant impacts on some outcomes
• Nevertheless, it seems that the impact are still lower than some international results
THANK YOU
References

• Gertler, et.al (2011), Impact Evaluation in Practice (including supplementary material for presentation), World Bank
• Pokja Monev TNP2K (2014), Evaluasi Dampak Program Keluarga Harapan, materi presentasi internal
• World Bank (2011), Program Keluarga Harapan: Main Findings from the Impact Evaluation of Indonesia’s pilot Household Conditional Cash Transfer Program
• World Bank website
Qualitative Study:
Sample & Data Collection

• Selected area are PKH endline survey area
• Sample: 22 villages in 6 districts/city of PKH pilot (kohort 2007) from 6 pilot PKH province.
• Control: 2 villages were selected from 2 Kecamatanans, from 2 districts.
• HH interviewed:
  • Poor HH, stay poor
  • Poor HH, welfare improved (not Poor HH anymore)
  • Poor HH, welfare deteriorate (from near poor to poor or very poor)
• Most Significance Change (MSC) & In-depth interview methods were used to collect data