Framework for impact evaluation

NEDA Workshop on Impact Evaluation
Mitzie Irene P. Conchada, Ph.D.
July 2018

Roadmap for today

- 130-300 — Framework for impact evaluation
  Causal inference and counterfactual
- 300-315 — Break
- 315-500 — Preparing for evaluation: theory of change and results chain, evaluation questions, outcome and performance indicators, checklist

Assessing where we are

- What are we currently working on?

- How do we prepare for an evaluation?

Objectives of impact evaluation

- Policy decisions
  - Curtailing inefficient programs
  - Scaling up programs that are effective
  - Selecting among various program alternatives
- Explore different types of policy questions
- Construct a comparison group that is similar to the treatment group (internal validity)

Purpose and scope of impact evaluation

- Assess the causal effect of public policy interventions
  - Job training programs on earnings and employment
  - Infrastructure projects
  - Class size on test scores
  - Minimum wage on employment

Video on building blocks of impact evaluation

Causality with potential outcomes

- Treatment

  Di: indicator of treatment intake for unit i

$$Di = \begin{cases} 1 \; if \; unit \; i \; received \; treatment \\ 0 \; otherwise \end{cases}$$

- Outcome

  Yi = observed outcome variable of interest for unit i

- Potential outcomes

Y_{0i} and Y_{1i}: potential outcomes for I

Y_{1i}= potential outcome for unit i with treatment

Y_{0i} = potential outcome for unit i without treatment

- Treatment effect
  - Also called causal effect of the treatment on the outcome for unit i is the difference between its two potential outcomes: Y_{1i} - Y_{0i}
- Observed outcomes are realized as:

$$Y_i = Y_{1i}D_i + Y_{0i}(1 - D_i) \quad \text{or} \quad Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

- Fundamental problem of causal inference
  - Cannot observe both potential outcomes (Y_{1i} and Y_{0i})

Exercise on causal inference and counterfactuals

- Refer to page 56-58
- Does the before-and-after comparison control for all the factors that affect health expenditures over time?
- Based on these results produced by the before-and-after analysis, should HISP be scaled up nationally?

Evaluating the Impact of HISP: Doing a Before-and-After Comparison of Outcomes

Recall that the Health Insurance Subsidy Program (HISP) is a new program in your country that subsidizes the purchase of health insurance for poor rural households and that this insurance covers expenses related to health care and medicine for those enrolled. The objective of HISP is to reduce what poor households spend on primary care and medicine and ultimately to improve health outcomes. Although many outcome indicators could be considered for the program evaluation, your government is particularly interested in analyzing the effects of HISP on per capita yearly out-of-pocket expenditures (subsequently referred to simply as *health expenditures*).

HISP will represent a hefty proportion of the national budget if scaled up nationally—up to 1.5 percent of gross domestic product (GDP) by some estimates. Furthermore, substantial administrative and logistical complexities are involved in running a program of this nature. For these reasons, a decision has been made at the highest levels of government to introduce HISP first as a pilot program and then, depending on the results of the first phase, to scale it up gradually over time. Based on the results of financial and cost-benefit analyses, the president and her cabinet have announced that for HISP to be viable and to be extended nationally, it must reduce yearly per capita health expenditures of poor rural households by at least US$10 on average, compared to what they would have

financial and cost-benefit analyses, the president and her cabinet have announced that for HISP to be viable and to be extended nationally, it must reduce yearly per capita health expenditures of poor rural households by at least US$10 on average, compared to what they would have spent in the absence of the program, and it must do so within two years.

HISP will be introduced in 100 rural villages during the initial pilot phase. Just before the start of the program, your government hires a survey firm to conduct a baseline survey of all 4,959 households in these villages. The survey collects detailed information on every household, including their demographic composition, assets, access to health services, and health expenditures in the past year. Shortly after the baseline survey is conducted, HISP is introduced in the 100 pilot villages with great fanfare, including community events and other promotional campaigns to encourage households to enroll.

Of the 4,959 households in the baseline sample, a total of 2,907 enroll in HISP, and the program operates successfully over the next two years. All health clinics and pharmacies serving the 100 villages accept patients with the insurance scheme, and surveys show that most enrolled households are satisfied with the program. At the end of the two-year pilot period, a second round of evaluation data is collected on the same sample of 4,959 households.[3]

The president and the minister of health have put you in charge of overseeing the impact evaluation for HISP and recommending whether or not to extend the program nationally. Your impact evaluation question of interest is, what is the impact of HISP on poor households' out-of-pocket health expenditures? Remember that the stakes are high. If HISP is found to reduce health expenditures by US$10 or more, it will be extended nationally. If the program did not reach the US$10 target, you will recommend against scaling it up.

The first "expert" consultant you hire indicates that to estimate the impact of HISP, you must calculate the change in health expenditures over time for the households that enrolled. The consultant argues that because HISP covers all health costs, any decrease in expenditures over time must be attributable to the effect of HISP. Using the subset of enrolled households, you calculate their average health expenditures before the implementation of the program and then again two years later. In other words, you perform a before-and-after comparison. The results are shown in table 3.1. You observe that the treatment group reduced its out-of-pocket health expenditures by US$6.65, from US$14.49 before the introduction of HISP to US$7.84 two years later. As denoted by the value of the t-statistic (*t-stat*), the difference between health expenditures before and after the program is *statistically significant*.[4] This means that you find strong evidence against the claim that the true difference between expenditures before and after the inter-

reduced its out-of-pocket health expenditures by US$6.65, from US$14.49 before the introduction of HISP to US$7.84 two years later. As denoted by the value of the t-statistic (*t-stat*), the difference between health expenditures before and after the program is *statistically significant*.[4] This means that you find strong evidence against the claim that the true difference between expenditures before and after the intervention is zero.

Even though the before-and-after comparison is for the same group of households, you are concerned that other circumstances may have also changed for these households over the past two years, affecting their health expenditures. For example, a number of new drugs have recently become available. You are also concerned that the reduction in health expenditures may have resulted in part from the financial crisis that your country recently experienced. To address some of these concerns, your consultant conducts a more sophisticated *regression analysis* that will try to control for some additional factors.

Table 3.1 Evaluating HISP: Before-and-After Comparison

	After	Before	Difference	*t*-stat
Household health expenditures (US$)	7.84	14.49	–6.65**	–39.76

Note: Significance level: ** = 1 percent.

- Break

Causal inference and counterfactuals

- Causal inference
  - Cause and effect relationship
  - Challenge in impact evaluation studies is establishing causality by empirically establishing to what extent the program contributed in change in outcome

$$\Delta = (Y|P = 1) - (Y|P = 0)$$

Formula: causal impact of a program P on outcome Y

- Counterfactual
  - How do we measure what would have happened if the other circumstance had prevailed?
  - How do we measure $(Y|P = 0)$?
  - Find a “perfect clone”

Figure 3.1 The Perfect Clone

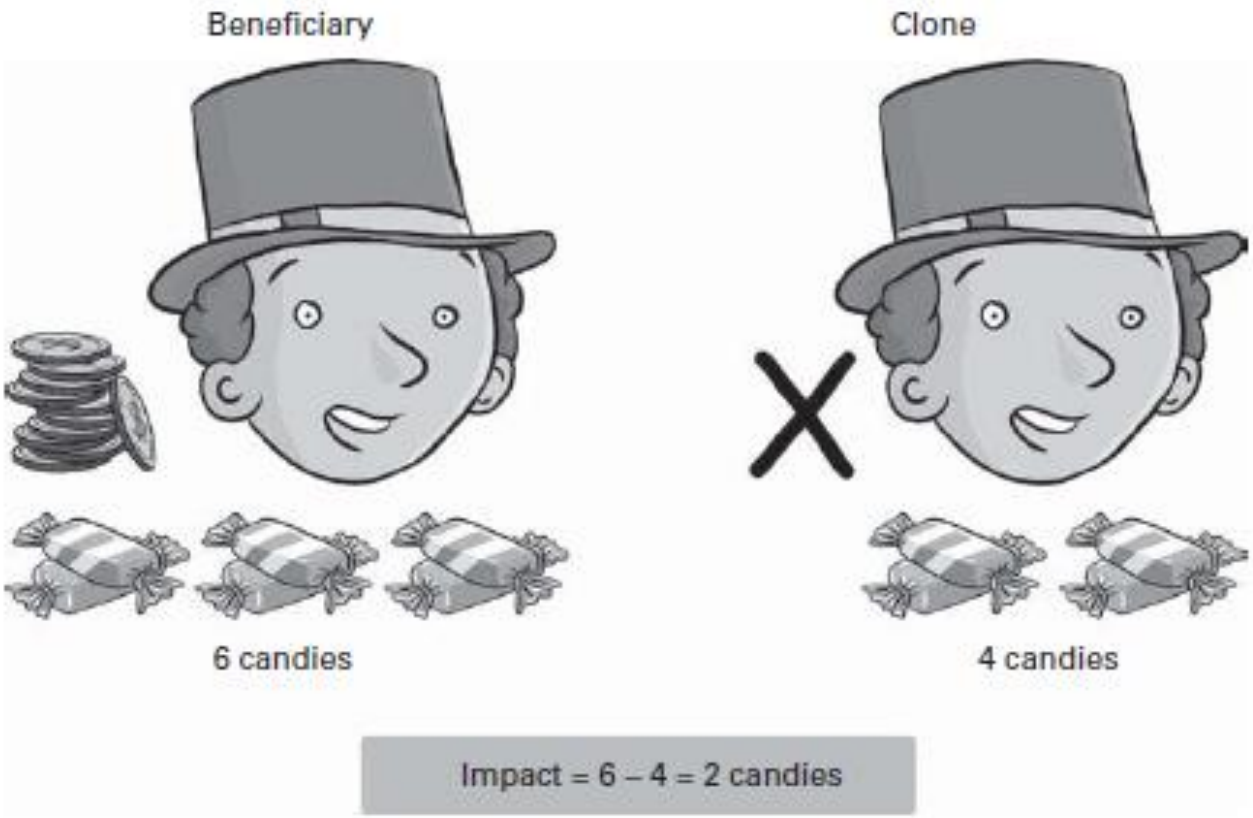
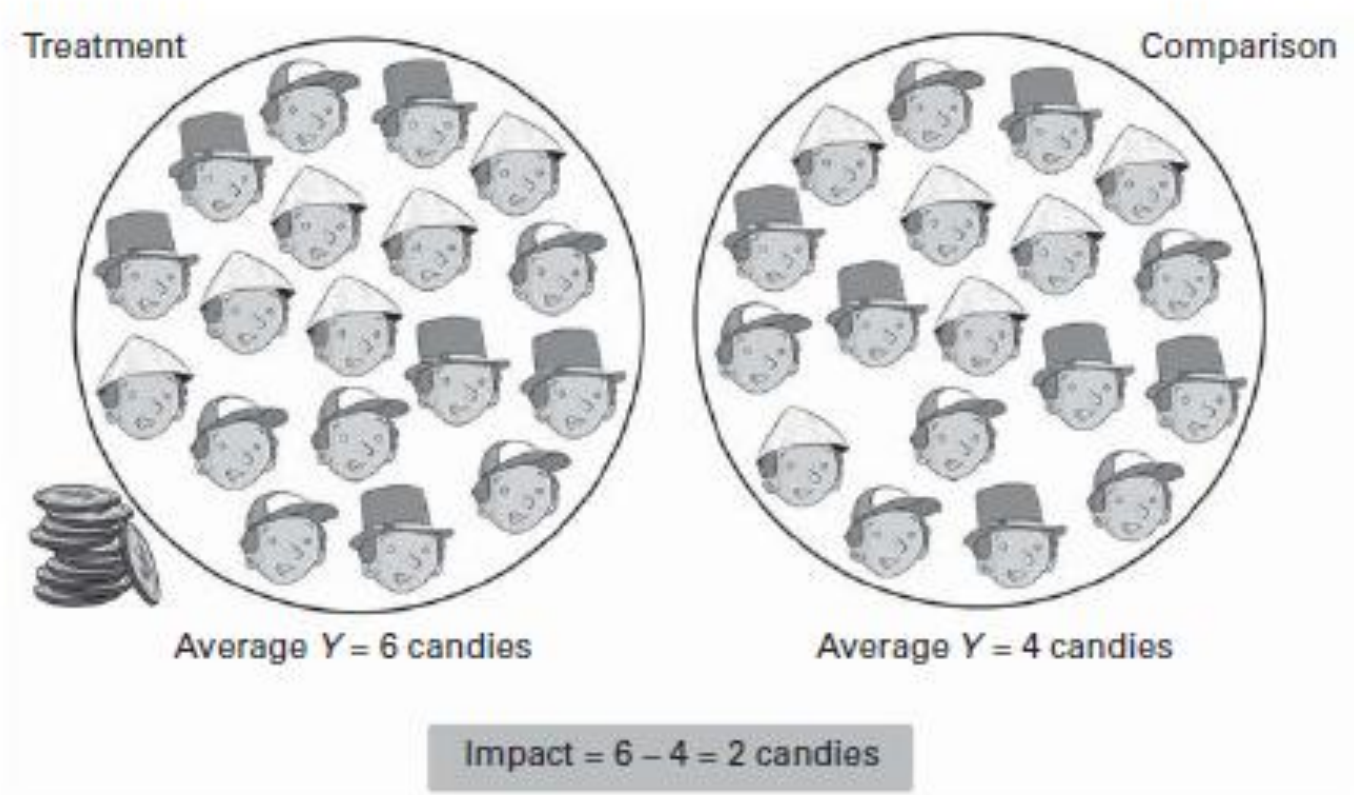


Figure 3.2 A Valid Comparison Group

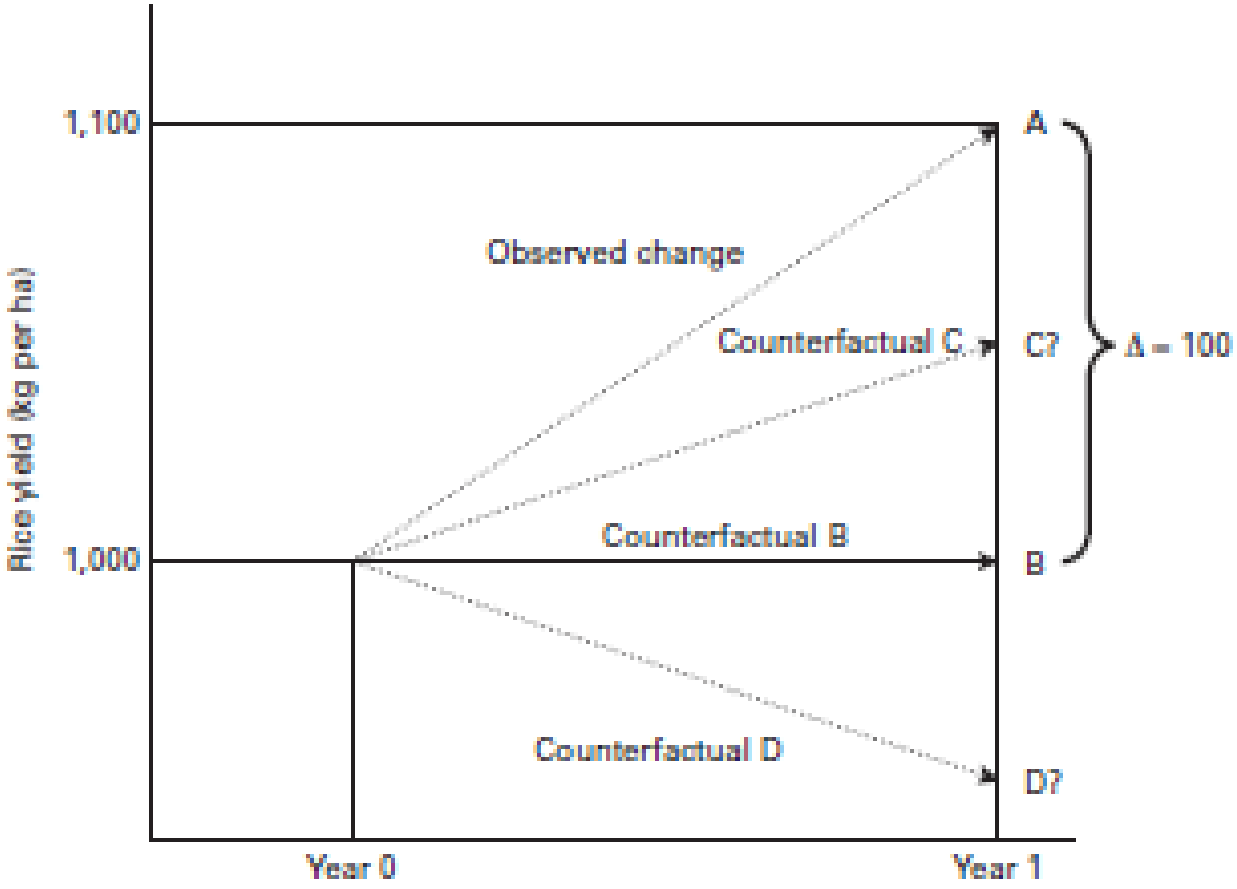


- Estimating counterfactual
  - Individual level ⟹ unit level
  - Challenge: identify treatment and comparison group that are statistically identical
  - 3 ways in which treatment and control group should be same:
    1. Average characteristics should be same even in absence of program i.e. same income levels
    2. Treatment should not affect comparison group – directly or indirectly
    3. Outcomes in control group = outcomes in treatment group

- 2 counterfeit estimates of counterfactual
  - Before-and-after comparisons
    - Problem: estimated counterfactual $(Y|P = 0)$ as outcome for treatment group before intervention started (baseline survey)
    - If baseline survey data is different from actual
    - i.e. microfinance program for poor in rural households – giving out of fertilizers to increase rice production

- Case study of microfinance program for poor, rural farmers

Case study of microfinance for poor, rural farmers: before and after estimates of the program



Note: Δ = Change in rice yield (kg); ha = hectares; kg = kilograms.

- Comparing enrolled and non-enrolled (self-selected groups)
  - Selection based on preferences, decisions or unobserved characteristics of potential participants
  - i.e. vocational training program for unemployed youth
    - Those who chose to participate in program may be more motivated to earn income
    - Unobserved preferences – based on interview (biased)

Theory of change

Theory of change

- Describes how intervention delivers desired results
- Depicts a sequence of events leading to outcomes
- Explores conditions and assumptions
- Shows the causal logic behind program – via map
- Ways to depict theory of change
  - Results chain
  - Theoretical models
  - Logic models
  - Logical frameworks
  - Outcome models

- 3 elements of theory of change
  - What are the RESULTS you are trying to achieve?
  - What are the STEPS or ACTIVITIES that you will take in order to achieve these results?
  - HOW will these steps lead to these results?
    - Usually guided by assumptions

- Results chain
  - Establishes causal logic from start to end of program
  - Elements:
    - Inputs – i.e. budget, staff
    - Activities – actions taken
    - Outputs – tangible goods and services
    - Outcomes – results likely to be achieved
    - Final outcomes – long term goals; objectives of program

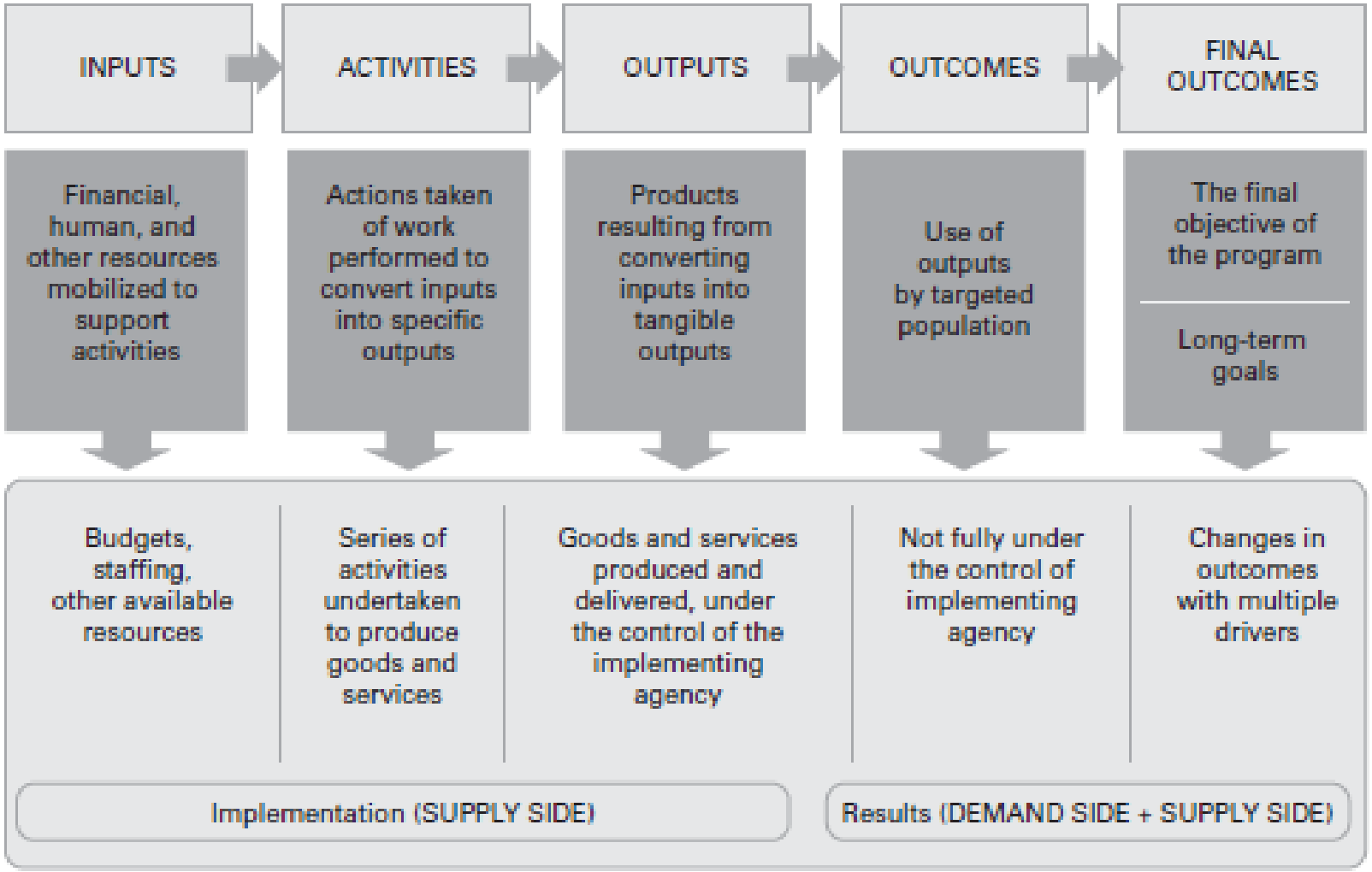
Results should be….

- Specific
- Measurable
- Achievable
- Realistic
- Timebound

Creating a theory of change

1. Results – what are you trying to achieve
2. Assumptions – work backwards and identify assumptions that support results
3. Activities – define activities that can be undertaken

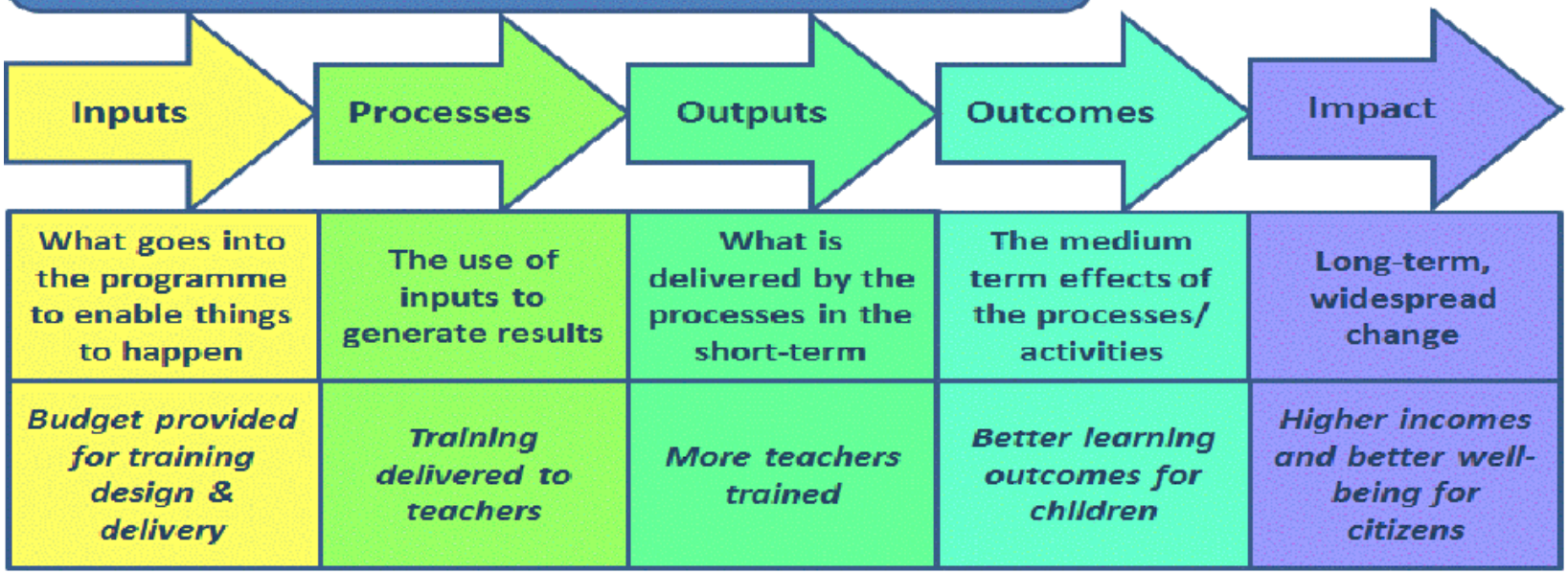
Figure 2.1 The Elements of a Results Chain



Source: Gertler, et al., 2016. Impact evaluation in practice

1. In PbR, payment should be linked to outcomes or outputs such as improved learning or increased enrolment.

2. But it may also be for an intermediate output or process if it can be shown that this is a measurable improvement in performance for the recipient.



3. Payments made on the basis of outcomes are a particularly important and innovative form of PbR, one that DFID is keen to do more of.

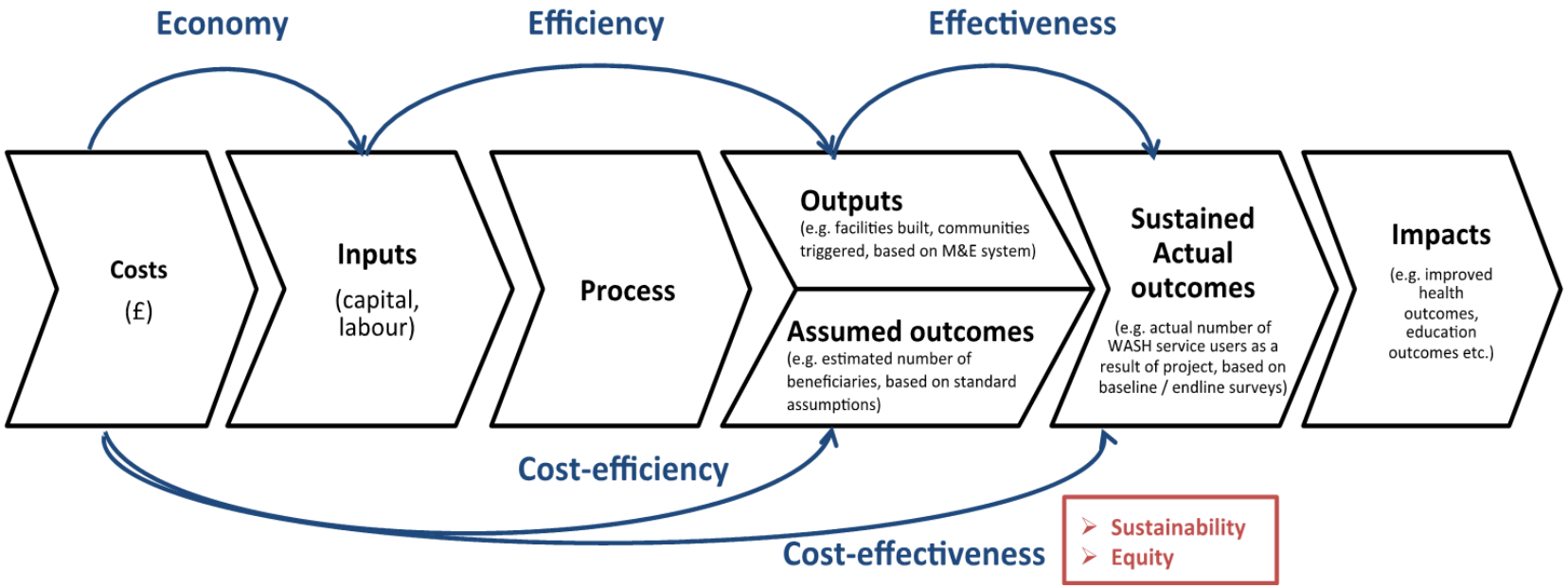
PbR – Payment by results strategy

Source: https://www.gov.uk/government/publications/dfids-strategy-for-payment-by-results-sharpening-incentives-to-perform/payment-by-results-strategy-sharpening-incentives-to-perform

Example 1: Results Chain

	Activities	Outputs	Outcomes	Longer-term Outcomes
Education	○ Teacher training ○ Textbooks developed	○ Teachers trained in new methods ○ Textbooks delivered	○ New methods used ○ Increased completion rates	○ Increased test scores ○ Increased labor productivity
Health	○ Doctors hired ○ Birth attendants trained	○ New doctors practicing ○ Attendants applying methods	○ Increased use of health clinics for deliveries	○ Improved maternal mortality
Social Protection and Labor	○ CCTs delivered ○ Targeting system ○ MIS	○ CCTs delivered to target households in accordance with conditions	○ Increased food consumption ○ Increased child health visits	○ Decreased poverty ○ Lower child mortality

Source: http://slideplayer.com/slide/2442387/



- Source: http://vfm-wash.org/tag/vfm/

Exercise on results chain

Box 1.3: Impact Evaluation Findings on Electricity Infrastructure

Although the number of impact evaluations on energy interventions has grown more slowly than in other sectors, studies conducted to date offer exciting evidence as to the effects of electricity access. The findings from these studies show effects on a range of outcomes from education, to health, income, and gender equality. At the same time, most of these results are from specific situations and interventions, so that additional studies are needed to verify the generalizability of findings.

1. Electricity connection can lead to *changes in time use*, particularly to increased study time for children, longer working hours, and increased time spent on nonagricultural income-generating activities for adults (Barron and Torero 2015, Grimm et al. 2013, Dasso and Fernandez 2015, Arraiz and Calero 2015).
2. Increased study time due to electricity access can lead to *improved educational outcomes* for children (Arraiz and Calero 2015, Khandker et al. 2013). However, it may also lead to increased childhood employment at the expense of education (Squires 2015).
3. Time use changes from electrification can lead to *microbusiness generation* (Dinkelman 2011, Khandker et al. 2013, Rao 2013, Dasso and Fernandez 2015).
4. Increased employment due to electricity can lead to *increased income, consumption, and expenditure* (Dinkelman 2011, Khandker et al. 2013, Rao 2013, Dasso and Fernandez 2015).
5. Electricity access may lead to *improved health* measured as a decline in reported respiratory infections and other smoke-related illnesses. This appears to follow improvement in indoor air quality as households substitute kerosene for electricity (ADB 2010, Barron and Torero 2015).
6. Some results have suggested that electricity access can *enhance family planning*. Increased TV viewing due to electrification has been observed to reduce fertility rates, partly as a result of higher exposure to family planning information that helps increase utilization of contraception (Grimm et al. 2015).
7. Other findings suggest that electrification can lead to *improved gender equality*. Effects on education have been found to be more positive for girls than for boys (van de Walle et al. 2013), as have effects on employment (Barron and Torero 2015,

- Source: White and Raitzer, 2017

Specifying evaluation questions

- Helps in providing focus for the research – what is the impact or changes directly attributable to a program
- Characteristics
  - Well-defined
  - Testable hypothesis
- Example: High school mathematics reform program
  - What is the effect of a new mathematics curriculum on test scores?
  - Identify the elements
  - Identify hypotheses

Selecting outcome and performance indicators

- Clear evaluation question
  - Need to identify outcome measures to assess results
- Clear objectives – program success
- Include stakeholders in evaluation team
- Indicators are good measures if:
  - S – specific: to measure the information required as closely as possible
  - M – measurable: to ensure that information can be readily obtained
  - A – attributable: to ensure that each measure is linked to the project's efforts
  - R – realistic: to ensure that the data can be obtained in a timely fashion, reasonable frequency, and reasonable cost
  - T – targeted: to the objective population

Checklist: getting data for indicators

- Are the indicators clearly specified
- Are the indicators SMART
- What is the source of data for each indicators
- With what frequency will data be collected
- Who is responsible for collecting the data
- Who is responsible for analysis and reporting
- What resources are needed to produce the data
- Is there appropriate documentation
- What are the risks involved

What can you do with your theory of change?

- Create an initial theory of change and get feedback
- Define a plan to monitor and evaluate theory of change
- Use the information you get from your monitoring and evaluation in two ways
  - Build into future work
  - Use it to revise your TOC

Break

Assignment: theory of change

- Select a project you work on
- Define the result, assumptions and activities
- Guide questions:
  - Does your theory make sense?
  - Are there assumptions you are missing?
  - Are there activities you need to add?

THEORIES OF CHANGE FOR INFRASTRUCTURE

- Refer to document on " Impact Evaluation Findings on Electricity Infrastructure" White and Raitzer, 2017 file:///C:/Users/SOE/Downloads/impact-evaluation-development-interventions-guide.pdf

- Identify any theory of change in the programs (infra) in low income countries

Box 1.5: Learning about Project Design from Impact Evaluation: Energy-Efficient Light Bulbs in Pakistan

The Government of Pakistan launched a national program costing $60 million to replace 30 million incandescent light bulbs with compact fluorescent light bulbs (CFLs) in the residential sector. ADB supported this program with a $40 million loan. An impact evaluation undertaken during the preparation of the loan found the following results:

- A significant minority (11%) of households are unaware of CFLs.
- The substantial majority of households have incorrect knowledge of CFLs' greater efficiency. CFLs last at least 10 times as long as incandescent bulbs. However, one-third of respondents replied they did not know the difference, and a quarter said that CFLs last just twice as long. Fewer than 10% replied that CFLs last 10 times as long.
- Benefits are overestimated if based on adoption alone since there is a "rebound effect" as households consume more light when using CFLs rather than incandescent bulbs.

The first two findings show the importance of including a demand (consumer education) component in the program, and the third informs the economic analysis.

Source: Chun and Jiang (2013).

- Source: White and Raitzer, 2017