

Module 2:

Evaluation of labour market policies

Sher Singh Verick

3rd SARNET Training on Labour Economics for Young Scholars

Deputy Director (a.i.)

ILO Decent Work Team for South Asia and Country Office for India

27 October – 7 November 2015

Outline

- Evidence-based policymaking
- Evaluation problem
- Theory of change
- Different evaluation methods
- Impact of active labour market policies

Key studies

- Gerteler P. et al. 2011. Impact Evaluation in Practice. World Bank, Washington, D.C.
- Glennerster, R. and K. Takavarasha 2013. Running Randomized Evaluations: A Practical Guide. Princeton University Press, Princeton.
- Kluve, J. 2006. The effectiveness of European active labour market policy. IZA Discussion Paper, No. 2018.
- J. Kluve et al. 2007. Active Labour Market Policies in Europe: Performance and Perspectives. Berlin and Heidelberg: Springer: 153-203.

Evidence-based policymaking

- What is the impact of a vocational training programme on employment and earnings?
- Does career counselling improve employment chances (and reduce unemployment)?
- Do entrepreneurship schemes result in sustainable businesses?
- Does a minimum wage reduce poverty?
- Do firing rules reduce inflows into employment?

Evidence-based policymaking

- With scarce resources, it is critical to know whether a government policy or programme is working
 - Should the government pursue this approach? Should the interventions be scaled up?
 - What is the most cost-effective way to achieve policy objective?
 - What is the cost-benefit ratio?

Evidence-based policymaking

- Impact evaluation – evaluate outcomes and identify whether interventions made a difference (i.e. are directly attributable to the programme and not some other factor)
 - Causal impact
- Monitoring
 - Tracking progress of the implementation of the programme
 - Mostly involves tracking inputs, activities and outputs

What is evaluation?



The different steps of evaluation

- | | |
|-----------------------|---|
| 1. Needs Assessment | A. What is the problem? |
| 2. Theory of change | B. How, in theory, does the program fix the problem? |
| 3. Process Evaluation | C. Does the program work as planned? |
| 4. Impact Evaluation | D. Were its goals achieved? The magnitude? |
| 5. Cost Effectiveness | E. Given magnitude and cost, how does it compare to alternatives? |

Needs assessment

- What are the key labour market challenges?
- What are the main factors driving poor outcomes (e.g. youth unemployment)?
 - Youth unemployment rate is 25%
 - Employers complain that graduates are unemployable
 - But the economy is not creating jobs.
- But what is really the cause of unemployment? A very careful needs assessment is required!

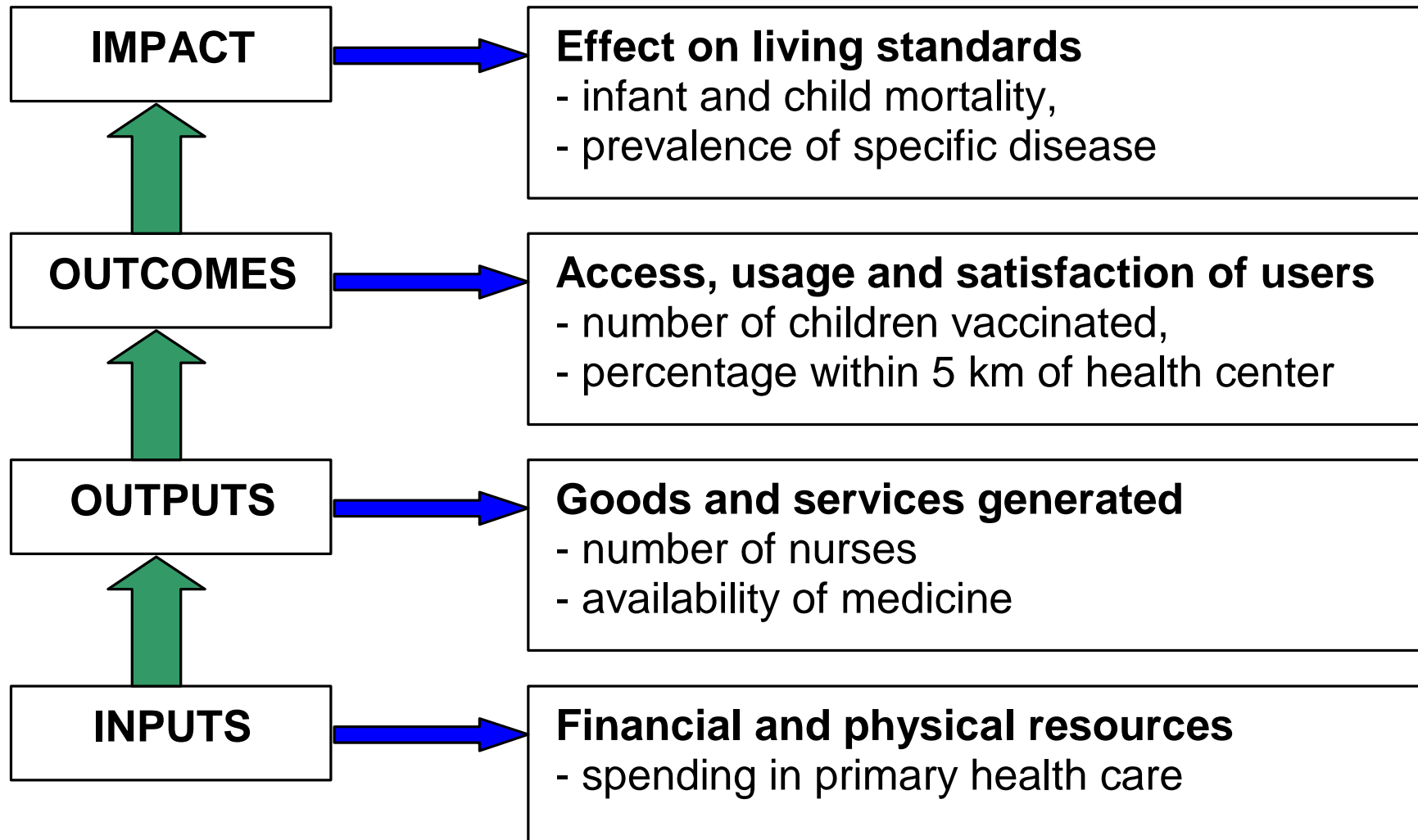
Needs assessment

- How will the programme address the needs in the labour market?
 - What are the prerequisites to meet the needs?
 - How and why are those requirements currently lacking or failing?
 - How does the program intend to target or circumvent shortcomings?
 - What services will be offered?

Theory of change

- “A theory of change is a description of how an intervention is supposed to deliver the desired results. It describes the causal logic of how and why a particular project, program, or policy will reach its intended outcomes.” *Impact Evaluation in Practice*
- Theory of change is necessary for
 - Designing an evaluation/intervention
 - Specifying the research question

Log frame



Evaluation problem

- Interested in the casual effect of a programme D on an outcome of interest Y

$$\alpha = Y(T = 1|D = 1) - Y(T = 1|D = 0)$$

Evaluation problem

- The (evaluation) problem is that we cannot observe the same person person/entity in a world without the intervention
 - Counterfactual – by definition, this cannot be observed
- The challenge of impact evaluation methodologies is to construct an estimate for the counterfactual

Evaluation problem: comparison group

- Need a control/comparison group that will allow us to attribute any change in the “treatment” group to the programme (causality)
 - A valid control group would have the same characteristics of the treatment group except for the participation in the programme

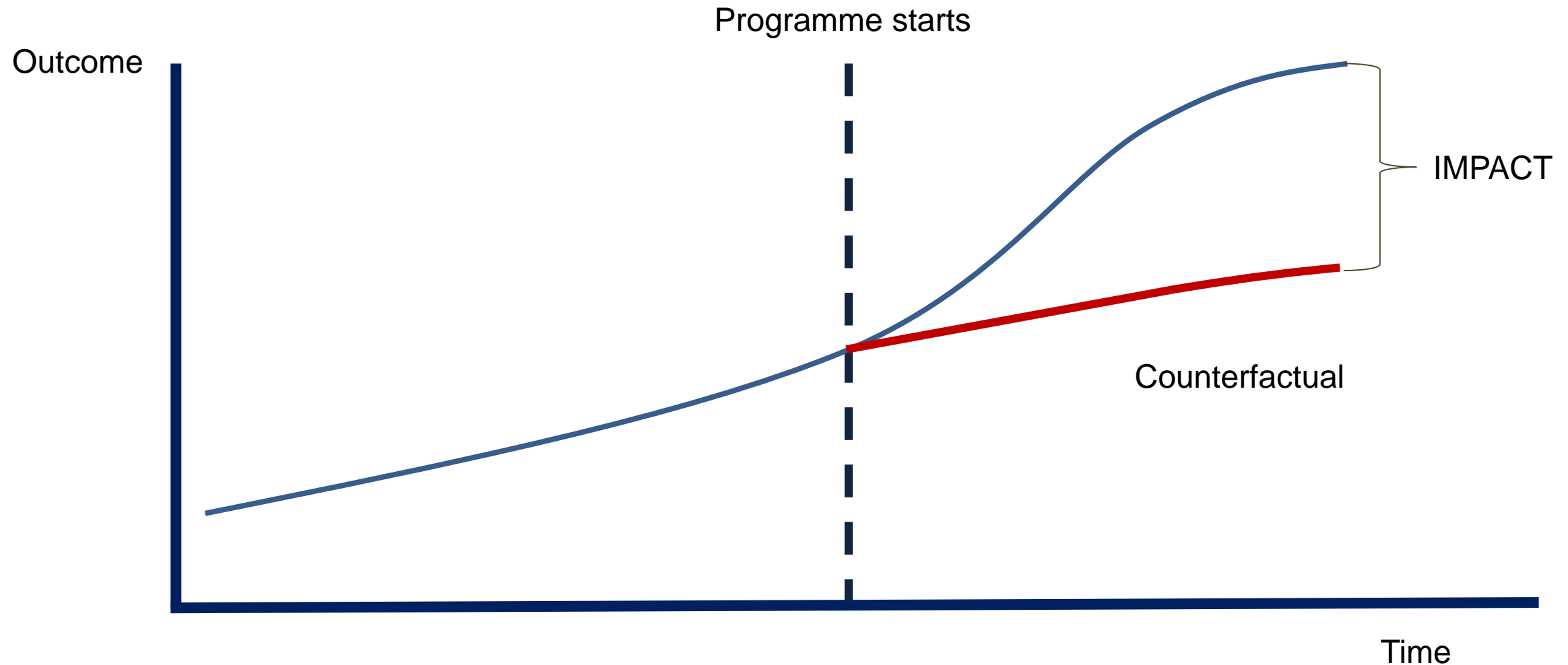
Evaluation problem: comparison group

- Two key problems for the construction of the comparison group
 - Programmes are targeted -> programme areas differ in observable and unobservable
 - Individual participation is (usually) voluntary -> participants will differ from non-participants in observable and unobservable ways
- A comparison of participants and an arbitrary group of non-participants can lead to an invalid estimate of the effect of the programme -> **biased estimates**

Two type of estimates

- Estimated impact α is called the “intention-to-treat” estimate (ITT) if
 - Calculation done using all participants who are offered the programme though some of them may choose not to participate
 - More important if the interest is in the average impact of the programme on the population ***targeted*** by the intervention

What is the impact of a programme?



Two type of estimates

- Estimated impact α is called the “treatment-on-the-treated” estimate (TOT) if
 - Impact evaluation is applied to units to whom the programme had been offered and who have actually enrolled it
- ITT = TOT if there is full compliance

Two misleading comparisons

- Before and after (pre-post) comparison
 - Compares outcomes for participants before and after the implementation of the programme
 - Comparison group – participants themselves before the intervention serve as the counterfactual
 - Required assumptions: treatment group's outcomes would have stayed the same in the absence of the programme –
- But there can be changes in the external environment, individuals, regression to the mean

Example: providing fertilizer to farmers

- The intervention: provide fertilizer to farmers in a poor region of a country (call it region A)
 - Programme targets poor areas
 - Farmers have to enroll at the local extension office to receive the fertilizer
 - Starts in 2002, ends in 2004, we have data on yields for farmers in the poor region and another region (region B) for both years
- Farmers which received fertilizer had a *decrease* in yields from 2002 to 2004
- But further study reveals there was a national drought, and everyone's yields went down (failure of the reflexive comparison)

Two misleading comparisons

- A cross-section comparison of differences between programme participants and nonparticipants
 - Comparison group (counterfactual) – individuals who didn't participate but data is collected after the implementation of the programme
 - Required assumptions: Participants and nonparticipants are identical except for the programme participation AND they were equally likely to enter the programme (i.e. no selection effect)
 - But selection bias occurs when the reason for which an individual participates in a programme is correlated with outcomes – e.g. individual is highly motivated

Impact evaluation methodologies

1. Multivariate regression
2. Statistical matching
3. Difference-in-difference
4. Regression discontinuity design
5. Instrumental variables
6. Randomized evaluations

Multivariate regression

- Compare participants and nonparticipants
- Need data on explanatory variables that explain the differences between the treatment and comparison groups
 - Assume that there are no other characteristics that differ between the two groups and might affect outcome variables
- Better than the simple before and after approach but it is hard to claim there are no unobservable traits like motivation that would differ between groups and affect outcomes

Matching

- A form of multivariate regression
- Programme participants are compared to a group of nonparticipants that is constructed by finding people whose observable characteristics are similar to those in the treatment group
- Each participant is paired with one or more nonparticipant that are similar based on observable characteristics
- Assumes that, conditional on the set of observables, there is no selection bias based on unobserved heterogeneity

Matching

- When the set of variables to match is large, often match on a summary statistics: the probability of participation as a function of the observables (the propensity score)
- Very data intensive and requires a large sample
- Strong identification assumptions

Difference-in-differences

- Observations over time: compare observed changes in the outcomes for a sample of participants and non-participants

Difference-in-differences

- Identification assumption: the selection bias is time-invariant ('parallel trends' in the absence of the program)
- Counter-factual: changes over time for the nonparticipants
- Constraint: Requires at least two cross-sections of data, pre-program and post-program on participants and nonparticipants
 - Need to think about the evaluation ex-ante, before the programme begins

Regression discontinuity

- The regression discontinuity design is an impact evaluation method that can be used for programs that have a continuous eligibility index with a clearly defined cutoff score to determine who is eligible and who is not.
- Sources of discontinuity
 - Social programmes often use an index to decide who is eligible to enrol in the program and who is not.
 - Labour laws only apply above a certain threshold

Regression discontinuity

- Comparison group are individuals/units close to the cutoff but do not receive the programme
- But require a clear discontinuity between those eligible and those who are not
- Moreover, the estimated impact is only valid in the neighbourhood of the cutoff (local average impact)
 - Cannot get an average treatment effect for all participants
 - Raises problems of statistical power (fewer observations)

Instrumental variables

- IV approach can be used if there is a variable that predicts participation but not outcomes (it is uncorrelated)
 - Interest rate for take up micro-credit; compulsory schooling laws
- Comparison group: individuals are less likely to participate in or benefit from programme because of nearly random factor
- The estimated effect is local: IV identifies the effect of the programme only for the sub-population of those induced to take-up the program by the instrument
- Validity of the instrument can be questioned, but cannot be tested.

An example of regression discontinuity

- Effects of a social assistance program in Quebec, Canada on labour market outcomes.
 - Provides help to the unemployed but offered lower payments to individuals under the age of 30 with no children, compared to individuals older than 30—\$185 a month versus \$507
- Lemieux and Milligan (2005) limited the sample to men without children and without a high school diploma and gathered data.
- Men close to the discontinuity (between the ages of 25 and 29) are very similar on observable characteristics.
- Comparing men on both sides of the eligibility threshold, the authors found that access to greater social assistance benefits actually reduced employment by about 4.5 percent for men in this age range without children.

Randomized evaluations

- Randomly assign individuals/units to the treatment or control groups
- Key advantage of experiments: members of the groups (treatment and control) do not differ systematically at the outset of the experiment
 - Any difference that subsequently arises between them can be attributed to the program rather than to other factors.

Randomized evaluations

- If properly designed and conducted, randomized experiments provide the most credible method to estimate the impact of a programme
 - Randomization assures that there is no systematic difference between groups
- RE can be done where the eligible population is greater than the number of places in the programme or needs to be phased in
- But many challenges exist
 - Political constraints, ethical questions
 - Expensive and lengthy
 - Can result in spillovers/crossovers
 - REs are not good at capturing general equilibrium effects

Beyond impact – useful information for policymakers

- Cost-effectiveness analysis
 - Compares the relative performance of two or more programmes or programme alternatives
- Cost-benefit analysis
 - Estimates the total expected benefits of the programme (difficult to measure!) compared to the total expected costs

Progres/Oportunidades

- Began in 1998, provides cash transfers to poor mothers in rural Mexico conditional on their children's enrolment in school, with their attendance confirmed by the teacher
- The grants, for children in grades 3 through 9, amount to about 50 per cent to 75 per cent of the private cost of schooling and are guaranteed for three year

Progresar/Oportunidades

- Need to phase in the large-scale social programme, about two-thirds of the localities (314 out of 495) were randomly selected to receive the program in the first two years, and the remaining 181 served as a control group before entering the program in the third year.
- Schultz (2004) found an average increase in enrolment of 3.4 per cent for all students in grades 1–8, with the largest increase among girls who had completed grade 6, at 14.8 per cent

Impact of ALMPs: findings of evaluations

Intervention	Empirical evidence	Additional constraints for developing countries
Job search assistance	Generally positive impact; inexpensive and cost-effective	Lack of employment services (public and private); more relevant for formal transactions in the labour market; less relevant when structural and cyclical unemployment is high
Training	Stronger positive impact over longer term; on-the-job training more effective; training targeting youth have been generally less successful	Training facilities are inadequate and inappropriate
Wage/employment subsidies	Less effective (mixed evidence); can generate deadweight costs and substitution effects	Expensive and are only relevant to the formal sector
Public works/employment	Public works are an effective social protection scheme but are unlikely to improve employment outcomes; subsidized public sector employment tend to be ineffective	Working conditions tend to be poor in infrastructure-related programmes; poor link to training, entrepreneurship and other complementary interventions
Entrepreneurship incentives	Some positive impact for older/better educated	Low success rate

Evaluation of ALMPs (cont)

- New meta-analysis (Card et al. 2010, Kluve 2010) suggests that
 - Classroom and OTJ training programmes yield more positive impacts in medium-term (2 years)
 - Found also US: JSA dominates training in SR but over longer term, gains to HC development policies are larger

Evaluation of ALMPs (cont)

- Card et al. (2010):

“... few studies include enough information to make a crude assessment of benefits of the programme relative to costs. Indeed, many studies completely ignore the “cost” side of the evaluation problem.”

Some other insights about what works

- Programmes tend to work better when they are
 - Part of comprehensive packages
 - Oriented to labour demand
 - Linked to a real workplace
 - Carefully targeted