Impact Evaluation: the Basics

Counterfactuals and randomization

William Jack
Georgetown University
Having an impact

• What does it mean?
  – Having things turn out differently (better) than they otherwise would have

• On whom?
  – Kids, women, rural residents, minorities, HIV/AIDS sufferers, mothers, unemployed youth,.....

• On what?
  – Well being, happiness, freedom, dignity
  – Income, security, opportunity
  – Health, education, employment, etc.
  – CD4 counts, test scores, assisted deliveries, etc.
Measuring impact

Policy change
Innovation
Intervention

What would have happened, the **counterfactual**

What actually happens

**Impact**

**Uh-oh!**
Elements of having an impact, and knowing you did

The art

Innovation

The science

Measure what happens

Predict what would have happened

Compare
Candidate counterfactuals I: “Before”

What happened before

What would have happened

What actually happens

Impact

Impact
Candidate counterfactuals II: the untreated

Treated

Untreated

Outcome for the treated

Outcome for the untreated

Impact
When is before-after a good comparison?

**Answer:** When nothing would have changed in the absence of the innovation.
When is treated-untreated a good comparison?

**Answer:** When the two groups are the same in all relevant respects.

- **Observed outcome for treated group**
- **Treated-untreated comparison**
- **Evolution of untreated**
  
  = Assumed counterfactual
But often the treated and untreated are different.

Time

Outcome

Before | After

Evolution of untreated = Assumed counterfactual

Observed outcome for treated group

Treated-untreated comparison overstates the impact
Finding the perfect match

Treated sample

Untreated sample
Comparing perfect matches

Difference in outcome

Impact on yellow types
Impact on purple types
Impact on green types
Impact on olive types
Impact on red types
Impact on blue types
Impact on orange types

Negative impact

Untreated sample  Treated sample
What if there are no perfect matches?

• Each individual has a projected outcome absent the innovation
Density functions

Number of people

\( f(y | T = 0) \)

\[ \alpha + \varepsilon_i \rightarrow \bar{y}_C = \alpha \rightarrow \alpha + \varepsilon_j \]

Outcome, \( y \), in the absence of the innovation
Distribution of outcomes when treated

- Each individual has a projected outcome in the presence of the innovation

\[ f(y | T = 1) \]

Outcome, \( y \), in the presence of the innovation

\[ \alpha \bar{y}_T = \alpha + \beta \]
Comparing distributions

- But we still need to decide who is in each group

Distribution of outcomes without treatment

\[ f(y \mid T = 0) \]

Distribution of outcomes with treatment

\[ f(y \mid T = 1) \]
Averages

• In terms of averages:

\[ \bar{y}_C = \alpha \quad \bar{y}_T = \alpha + \beta \]

• \( \beta \) is the difference between the averages of the treated and the untreated groups
Regressions

• At the individual level:

\[ y_i = \alpha + \beta T_i + \varepsilon_i \]

– where \( T_i = 1 \) if individual \( i \) is treated, and 0 if no

– and \( \varepsilon_i \) is the amount by which \( i \)'s outcome differs from the average for her/his group
Geometry of regression

Outcome, $y_i$

Regression line
slope = $\beta$

Observed average outcome for treated group

Observed average outcome for untreated group

Treatment status, $T_i$
But who’s in each group?

Number of people

\[ f(y \mid T = 0) \]
Random sampling

\[ f(y \mid T = 0) \]

\[ f(y \mid T = 1) \]
Random sampling

\[ f(y \mid T = 0) \]

\[ f(y \mid T = 1) \]
Random sampling

$$f(y | T = 0)$$

$$f(y | T = 1)$$
Choosing a random sample

Population

Random sample
Randomized assignment

Population
Random sample

Randomized assignment
Treatment

Control
Eligibility

- We only need a random sample if we want to say something about the impact on the population as a whole.

- Randomization gives us “internal validity”.

Treatment

Control
Voluntary participation

- We can’t say what would have been the effect on these guys
- We lack “external validity”
What does random mean?

• Something about being chosen by “chance”

• Perhaps generated by a physical system
  – Coin toss, roll of dice, observation of a quantum mechanical system
Pseudo-random numbers

• These are generated by a computer

• E.g., let \( r \) be drawn from the uniform distribution on \([0,1]\)

• \( \text{prob}(r < r_0) = r_0 \)

• Assign to treatment if \( r < 1/2 \), and control if \( r \geq 1/2 \)
Mutli-treatment experiments

- Sometimes we want to compare two or more different kinds of treatments
  - E.g., different intensities, different means of delivery, etc.
Cross comparisons

• Suppose you want to test two types of intervention – e.g. to improve school attendance

  School feeding programme  Vs  Conditional cash transfer

• Is either better than nothing?
• Which one is better than the other?
• Are there important synergies?
2 x 2 design

Population → Eligible sample

<table>
<thead>
<tr>
<th></th>
<th>Feeding programme</th>
<th>No feeding programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCT</td>
<td>Feeding + CCT</td>
<td>T2</td>
</tr>
<tr>
<td>No CCT</td>
<td>T1</td>
<td>C</td>
</tr>
</tbody>
</table>
What happens if we don’t randomize?

$$f(y \mid T = 0)$$

$$f(y \mid T = 1)$$

True effect

Over-estimate of impact due to non-randomization
Biased estimates of the treatment effect

Outcome, $y_i$

Regression line slope is too large

Observed average outcome for treated group

Observed average outcome for untreated group

0

1

Treatment status, $T_i$
Compliance, ITT, and TOT

• What if some people assigned to the treatment don’t take part?
  – Should we ignore them?
  – Or put them in the control?

• What if some assigned to the control do take part?
  – Should we include them as treated?
An RCT with full compliance

• Suppose we randomly assign individuals to the treatment group, and guarantee full compliance

• That is,
  – *all* individuals assigned to the treatment group actually get treated, and
  – *none* of those assigned to the control group do
Incomplete compliance

• But what if we can’t ensure perfect compliance?

• Here, assignment to treatment increases the chance of being treated, but not from 0 to 100 percent.
Can we compare the treated with the untreated?

What about comparing the treated in the treatment group with the untreated in the control?

Again, these groups look like they might be different.
Intent to treat

• Let’s just compare all those in the treatment group with all those in the control group
• We know (by randomization) that they are statistically the same
• This is the Intent to Treat, or “ITT”, estimate
• It is the average impact of being assigned to the treatment group
• It is not the impact of the treatment itself
Compliers and non-compliers

• Think of there being three types of subjects
  1. Never treated, share $\phi_N$
  1. Compliers, share $\phi_C = (1 - \phi_N - \phi_A)$
  1. Always treated, share $\phi_A$

• Within each type, the treatment and control groups are statistically identical
Which type can we look at?

• We can calculate the effect of treatment *neither* on the “Never treated”
  – there is no-one who gets treated
• .....*nor* on the “Always treated”
  – there is no-one who doesn’t get treated

• But we *can* (indirectly) calculate the effect of treatment on the compliers – those whose treatment status is affected by the assignment
Calculating the treatment effect on compliers

• Label the three types of people by $i = N$ (never treated), $A$ (always treated), and $C$ (compliers)

• Let $Y_i^0$ be the outcome for type $i$ people when they are not treated, and $Y_i^1$ be the outcome when they are treated
Some simple algebra

• The mean outcome for those assigned to the control group is
  
  \[ Y^C = \phi_N Y^0_N + \phi_C Y^0_C + \phi_A Y^1_A \]

• The mean outcome for those assigned to the treatment group is
  
  \[ Y^T = \phi_N Y^0_N + \phi_C Y^1_C + \phi_A Y^1_A \]

• The difference is
  
  \[ \Delta Y = Y^T - Y^C = \phi_C (Y^1_C - Y^0_C) \]
The IV estimate

\[ \Delta Y = Y^T - Y^C = \phi_C (Y^1_C - Y^0_C) \]

• Hence,

\[ Y^1_C - Y^0_C = \left( \frac{Y^T - Y^C}{\phi_C} \right) \]

Wald estimate of the Local average treatment effect, or “Treatment on the treated”

• The TOT “scales” the ITT estimate by one over the fraction of the sample who are affected by the assignment