14.74

Lecture 5: The long run impact of a childhood intervention

A randomized evaluation

Esther Duflo

Fall 2013

Last lecture, we saw that one impact of nutrition which may be a source of a nutrition-based poverty trap is the long-run effect of proper nutrition in childhood (or even *in utero*).

For example, Anne Case and Christina Paxson find that, in the UK, tall people earn more than shorter people, but that the correlation disappears when you control for IQ. They hypothesize that height is a marker for good nutrition in early childhood.

How would we demonstrate directly that nutrition in early childhood affects long-run income?

We have two problems:

- Data collection–How can we solve that?

- Endogeneity or omitted variable bias.

Today we will examine one solution to the problem of endogeneity and omitted variable bias: Randomized Control Trials. We will focus on one specifc example: deworming in early childhood. Sarah Baird, Joan Hamory

Hicks, Michael Kremer and Edward Miguel: "Worms at work".

This is not really a nutrition intervention. Nevertheless, why is deworming a good window into nutrition in early childhood?

Randomized evaluation: The idea

To test the effect of a policy, we can use *randomized evaluation*, where a randomly selected *treatment group* receives a treatment, while the other group does not (this is the comparison group). We will collect data on both the treatment and the comparison group, and compare the result. Because the treatment and the comparison groups have been randomly selected, any difference between them after the intervention should be due to the intervention itself.

The evaluation problem

Now, a more formal introduction to this problem, with deworming as an example:

Let us call Y_i^T the wage level of an individual i who receives deworming as a child and Y_i^{NT} the wage level of the same individual i if he did not get dewormed. Can we observe Y_i^T and Y_i^{NT} at the same time?

Y_i^T and Y_i^{NT} are called *potential outcomes*.

We are interested in the difference:

$$Y_i^T - Y_i^{NT}$$

That is, the effect of receiving deworming on wages.

The problem: we don't observe individual i both with and without the deworming at the same time. What can we do? We will never know the effect of deworming on a particular individual. We may hope to learn the *average* effect of deworming.

$$E[Y_i^T - Y_i^{NT}]$$

Imagine we have access to data on lots of individuals in the region. Some individuals receive deworming and others do not. Take averages:

$$E[Y_i^T|\text{DW}]-[Y_i^{NT}|\text{no DW}] = E[Y_i^T|T]-E[Y_i^{NT}|NT]$$

Subtract and add $E[Y_i^{NT}|T]$

$$E[Y_i^T|T]-E[Y_i^{NT}|T]-E[Y_i^{NT}|NT]+E[Y_i^{NT}|T] =$$

$$E[Y_i^T - Y_i^{NT}|T] + E[Y_i^{NT}|T] - E[Y_i^{NT}|NT]$$

- The first term $E[Y_i^T - Y_i^{NT}|T]$ is the *treatment effect* that we are trying to isolate: on average, among all the people I give deworming to, what will be the effect of deworming on their wages?

- What is:

  -$E[Y_i^{NT}|T]$?

  -$E[Y_i^{NT}|NT]$?

  - The difference $E[Y_i^{NT}|T] - E[Y_i^{NT}|NT]$?

- Which is likely to be bigger? Why?

The difference is the *selection bias*. It tells us that besides the effect of the deworming, there may be systematic differences between those who receive deworming and those who do not.

What happens when we randomly allocate the treatment?

Suppose that we select the individuals to whom we give deworming randomly within a population of individuals. We observe the wage level both in a *treatment group* (those to whom we gave DW) and for those we have not given DW to, which will form our *control (or comparison) group*.

On average, what do we expect to find if we compare treated individuals and untreated

individuals before the intervention? Or if we compare other characteristics of these individuals that are not likely to be affected by deworming (even after the deworming)?

Compare $E[Y^{NT}|NT]$ and $E[Y^{NT}|T]$
$\rightarrow$ What is $E[Y^{T}|T] - E[Y^{NT}|NT]$ equal to?

Randomized evaluation in practice –with all its trickiness! The deworming example

The program was started in 1999. A first study was undertaken in the short run (before the control group was treated) to look at the effect of the program on school participation, anemia, etc.

A second study was undertaken in the longer run: 7,530 of children surveyed in 2003-2005 and 2007-2009.

In the original program, the decision was taken to randomize at the *school level* (see map). Schools were assigned into three groups:

- Group 1 Deworming program received in 1998-2003
- Group 2 Deworming program received in 1999-2003
- Group 3 Deworming program recevied in 2001-2003

The rest is noise

While, in expectation, treatment and control will have the same potential outcome, in practice, in any sample, they won't: we need to randomized over a sample large enough to have enough *power* in our experiment, to distinguish the effect from zero.

In practice we replace the expectation by the estimated expectation: the sample average. Or we can get the treatment effect directly by running a regression:

$$Y = \beta T + \epsilon$$

Exercise (will be done in recitation as well): Show that $\hat{\beta}$ is equal to $\widehat{E[Y^T]} - \widehat{E[Y^{NT}]}$

One way to see the role of the noise: if we had randomized in the sample, but with a different seed, we would have obtained slightly different results.

This is what the *standard errors* of the estimated treatment effect reflect.

When we randomize at the school level, rather than at the individual level, we need to take it into account when figuring what is the noise of the estimate. Why?

Clustering: we adjust the standard errors (using the "cluster" command in stata, e.g. cluster(schoolid)).

Figuring out the right sample size

To figure out the right sample size, we try to make sure we have enough information to test some specific hypothesis. For example: is the effect zero? Test for this hypothesis: t statistics (coefficient divided by standard error).

- Size of a test: probability of a type I error: we reject H_0 when in fact is is true

- Power of a test, one minus probability of a type II error: we fail to reject H_0 when in fact it is wrong (i.e. we say there is no effect when in fact there is one).

We calculate the power for a given desired size (e.g., a 5% size). The more observations, the more power. We usually try to have a sample size large enough for a power of 80%.

Externalities

Why not randomize *within* school instead? (treat some kids and not some other)

Worms are contagious: if some children are treated, they are less likely to be sick, so even their friends are also less likely to be sick. What does it do to our treatment effect if we randomize within school?

What would be the best way to *test* whether there are such externalities within schools in practice?

What is an approximate way, if deworming was offered on a specific day in class?

How can we test if there is externalities across schools?

There is evidence of externality in the first study both within and across schools (see tables)

Collecting our data

The first evaluation focused on kid school attendance: we can collect this data by visit school a number of times, and take presence in school.

However, for the long run evaluation (wages, and other outcomes) they needed to find people, many years after the fact.

What is the big problem:

If attrition is random (some people just cannot be found because it rains the ray you are looking for them, for example), that just affects our sample size.

But attrition could be non-random, and in particular it could be related to the treatment.

Possible scenarios:

In which direction does each of this scenario bias our results? Why?

Yet, trying to find everyone could be really expensive! To manage attrition in this case they proceeded in two steps:

1. Step one: Try to find everybody, 64% were found.

2. Step two: Focus on a selected subsample (one quarter of not found), for intensive tracking
   Effective tracking rate is:

$$ETR = RTR + (1 - RTR) * ITR$$

The results and what to do with them

Results

- Short-run: Large effects on school participation
- Long-run: See in figure 3, table 3, table 4, table 6, table 7.
- What are the highlights in these results?

Cost effectiveness analysis

- Benefit:

- Costs:

Bottom line: benefits/cost ratio: 23!

Conclusion

School based deworming is extremely cheap and easy to do. In places where worms are a problem (which can be measured with a survey) it is recommended policy (by WHO). This shows that a "small" program can have very large impact (there are very few things we know if that can increase wages by 20% every year!!!)

However one surprising thing is that parents are not doing it themselves. In fact, when cost-sharing was introduced in the schools in treatment 1, take up fell to almost zero.

Possible reasons?