

MIT OpenCourseWare
<http://ocw.mit.edu>

Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Analyzing data

Abdul Latif Jameel
Poverty Action Lab

This Lecture

- Dealing with “problems” of non-compliance
 - When people don’t do what you assign them to
- Choosing what effects to report in your study
- External validity
 - When can I be confident my study tells us something about the
- Cost effectiveness
 - How can we compare various interventions?

Outline

- I. ITT and TOT
- II. Choice of Outcomes and Covariates
- III. External Validity
- IV. Cost Effectiveness

The problem

- Individuals allocated to treatment group do not receive treatment
 - *De-worming program*: some students assigned to treatment in treatment schools did not receive medical treatment
- Or individuals assigned to comparison group attempt to move into treatment group
 - *De-worming program*: parents could attempt to move their children from comparison school to treatment school
- *What do you do?*

Sample selection bias

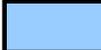
- Use the original assignment
 - i.e. if a child ended up in a treatment school but was from the control, she should be assigned to control when calculating the effect.
- This gives the Intention to Treat estimate (ITT).

Outline

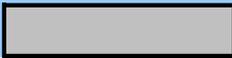
- I. ITT and TOT
- II. Choice of Outcomes and Covariates
- III. External Validity
- IV. Cost Effectiveness

ITT estimator

School 1	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	yes	yes	4
Pupil 2	yes	yes	4
Pupil 3	yes	yes	4
Pupil 4	yes	no	0
Pupil 5	yes	yes	4
Pupil 6	yes	no	2
Pupil 7	yes	no	0
Pupil 8	yes	yes	6
Pupil 9	yes	yes	6
Pupil 10	yes	no	0
Avg. Change			

Avg. Change School 1 
Avg. Change School 2 

==>ITT Effect is 

School 2	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	no	no	2
Pupil 2	no	no	1
Pupil 3	no	yes	3
Pupil 4	no	no	0
Pupil 5	no	no	0
Pupil 6	no	yes	3
Pupil 7	no	no	0
Pupil 8	no	no	0
Pupil 9	no	no	0
Pupil 10	no	no	0
Avg. Change			

Intention to treat (ITT)

- What does “intention to treat” measure?
- Intention to treat: “What happened to the average child who is in a treated school in this population?”
- Is this the right number to look for?
- Remember: In the deworming case many children in treatment schools were not treated and some children in comparison schools were.

When is ITT useful?

- May relate more to actual programs
- For example, we may not be interested in the medical effect of deworming treatment, but what would happen under an actual deworming program.
- In an actual program, some students would miss school (and not get deworming medicine), and others would have taken medicine anyway without the program
- The intention to treat estimate may actually be most relevant to the “program’s impact”

**What if you wanted to know the effect of the
medicine on the children who took the
medicine?**

Partial compliance

Some students in treatment schools not treated

- 1998: 78% of pupils assigned to receive treatment received at least some treatment
- 1999: around 72%
- Absence from school the major cause of non-compliance

Some students in comparison schools treated

- 5% received treatment outside of program

-
- You CANNOT compare the children who took the medicine with those who did not take it
 - Why?
 - But you can re-scale your effect to account for partial compliance.
 - This is called the “Treatment on the Treated” estimator, or ToT

What NOT to do!

School 1	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	yes	yes	4
Pupil 2	yes	yes	4
Pupil 3	yes	yes	4
Pupil 4	yes	no	0
Pupil 5	yes	yes	4
Pupil 6	yes	no	2
Pupil 7	yes	no	0
Pupil 8	yes	yes	6
Pupil 9	yes	yes	6
Pupil 10	yes	no	0
Avg. Change among Treated A=			<input type="text"/>

School 2	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	no	no	2
Pupil 2	no	no	1
Pupil 3	no	yes	3
Pupil 4	no	no	0
Pupil 5	no	no	0
Pupil 6	no	yes	3
Pupil 7	no	no	0
Pupil 8	no	no	0
Pupil 9	no	no	0
Pupil 10	no	no	0
Avg. Change among Not-Treated B=			<input type="text"/>

School 1:

Avg. Change among Treated (A)

School 2:

Avg. Change among not-treated (B)

A-B

From ITT to effect of treatment on the treated (TOT)

- The point is that if there is leakage across the groups, the treatment-control difference in impact is smaller...
- but the difference in the probability of getting treated is also smaller.
- Formally this is done by “instrumenting” the probability of treatment by the original assignment.
 - Original assignment is correlated with actual treatment, but uncorrelated with other factors affecting outcomes

Treatment on the treated (TOT)

- The effect of the treatment on those who got the treatment:
 - Suppose children who got the treatment had a weight gain of A , irrespective of whether they were in a treatment or a control school
 - Suppose children who got no treatment had a weight gain of B , again in both kinds of schools
 - We want to know $A-B$, the difference between treated and non-treated students

Treatment on the treated (TOT)

Then...

- $Y(T) = A * \text{Prob}[\text{treated}|T] + B(1 - \text{Prob}[\text{treated}|T])$
- $Y(C) = A * \text{Prob}[\text{treated}|C] + B(1 - \text{Prob}[\text{treated}|C])$
- $A - B = \frac{(Y(T) - Y(C))}{(\text{Prob}[\text{treated}|T] - \text{Prob}[\text{treated}|C])}$

= The “treatment on the treated” effect.

TOT estimator

School 1	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	yes	yes	4
Pupil 2	yes	yes	4
Pupil 3	yes	yes	4
Pupil 4	yes	no	0
Pupil 5	yes	yes	4
Pupil 6	yes	no	2
Pupil 7	yes	no	0
Pupil 8	yes	yes	6
Pupil 9	yes	yes	6
Pupil 10	yes	no	0
Avg. Change Y(T)=			<input type="text"/>

School 2	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	no	no	2
Pupil 2	no	no	1
Pupil 3	no	yes	3
Pupil 4	no	no	0
Pupil 5	no	no	0
Pupil 6	no	yes	3
Pupil 7	no	no	0
Pupil 8	no	no	0
Pupil 9	no	no	0
Pupil 10	no	no	0
Avg. Change Y(C) =			<input type="text"/>

A = Gain if Treated
B = Gain if not Treated

ToT Estimator: A-B

$$A-B = \frac{Y(T)-Y(C)}{\text{Prob(Treated|T)}-\text{Prob(Treated|C)}}$$

Y(T)

Y(C)

Prob(Treated|T)

Prob(Treated|C)

Y(T)-Y(C)

Prob(Treated|T)-Prob(Treated|C)

A-B

TOT

- General technique for when compliance is not perfect, for example, in an encouragement design.
- But: to calculate it, we need to observe the number of people getting treated in both groups to compute:

$$\text{Prob}[\text{treated}|T] - \text{Prob}[\text{treated}|C].$$

- This may not always be easy. Why?
- Possible solution?

TOT: Cautions

- When you can NOT use TOT:
 1. When your encouragement affects the outcome directly, and not just through getting people to take up your treatment. Why?
 2. When you have significant spillovers or externalities. Why?

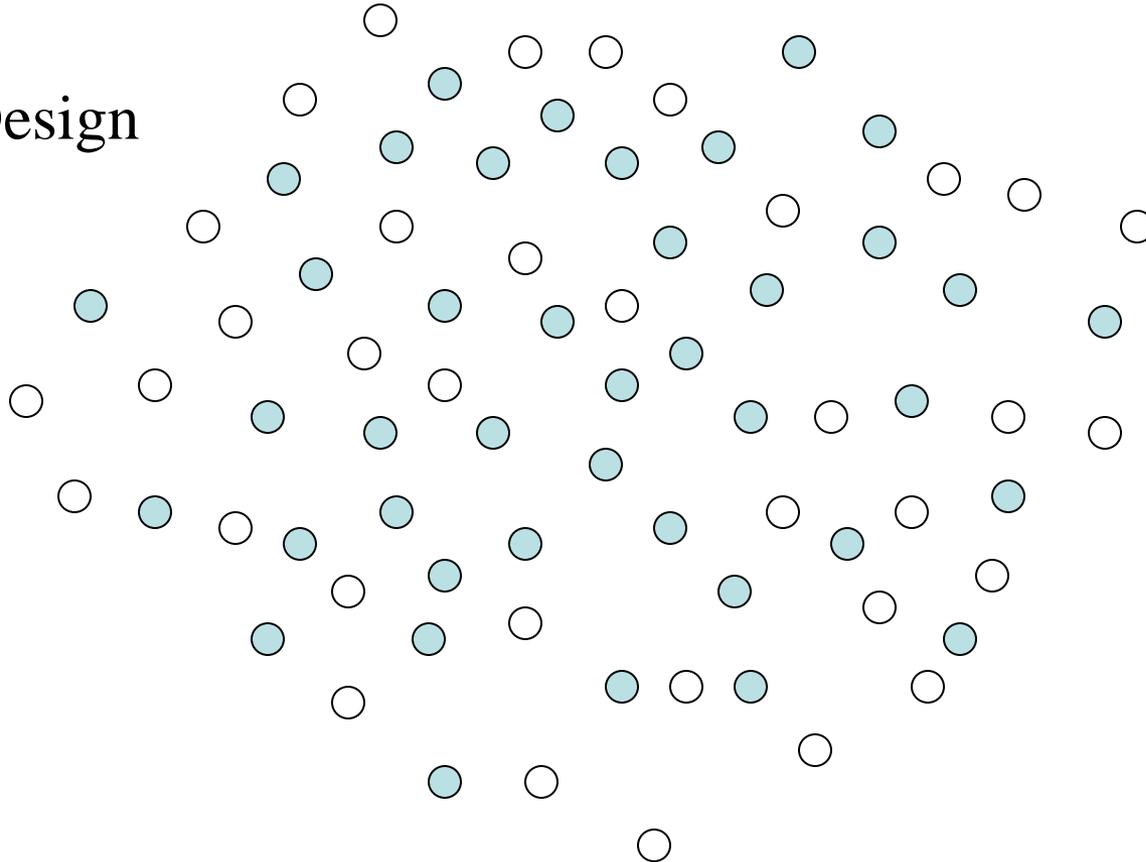
If you have either of these two situations, you should only use ITT

ITT: Another example

- You are starting a malaria prevention program
- You sample 40 villages for the pilot study
- You randomly assign 20 villages to receive the treatment in the first year
 - the remaining 20 villages will be the “comparison” during the pilot (they will receive the treatment later if it works)
- You give the list of 20 treatment villages to your program manager
- Some of the villages that are “comparison” are unhappy. Their leaders talk to your program manager and repeatedly ask him to treat their village.
- The program manager cannot resist the pressure; in the end he cannot fully respect the initial design:
 - He implements the program in only 15 of the 20 villages you selected
 - And also in 2 villages that were in the “comparison”, and in 3 villages out of your sample
- What do you do to measure the impact of your program?

ITT: Another example

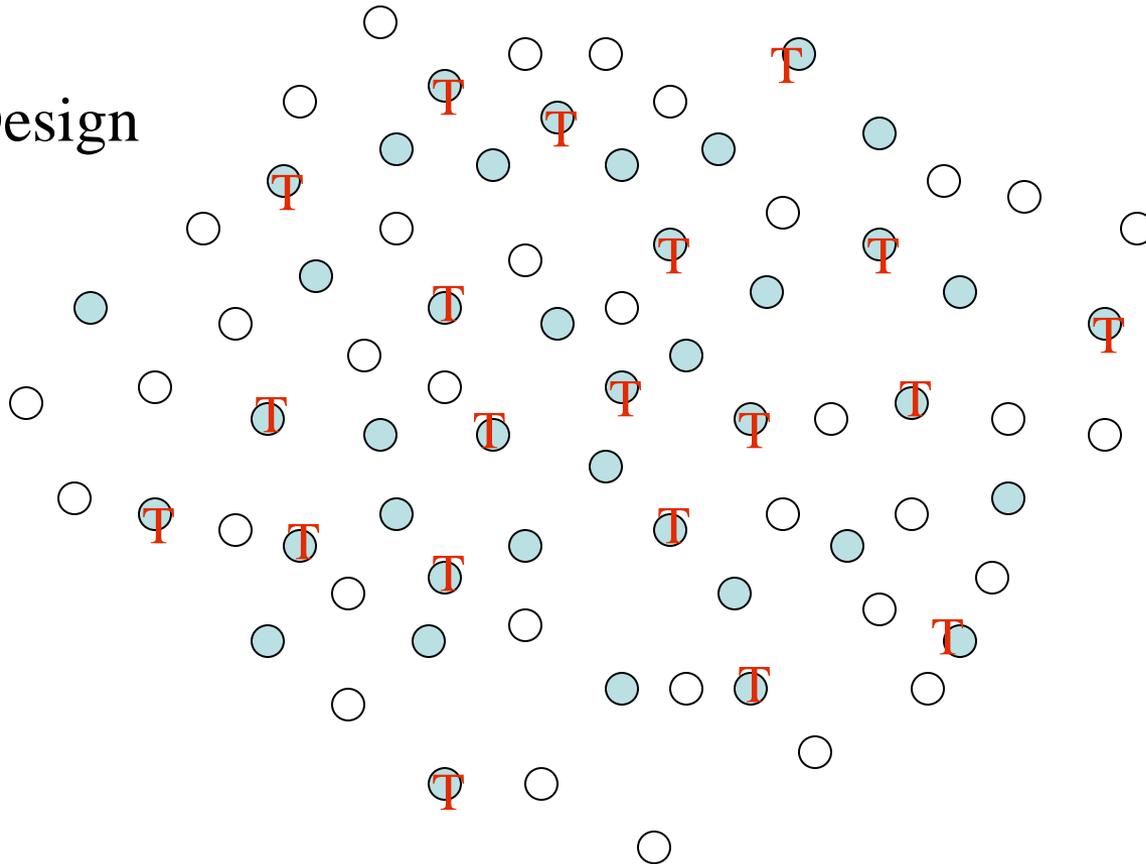
Initial Design



- Your sample
- Other villages

ITT: Another example

Initial Design



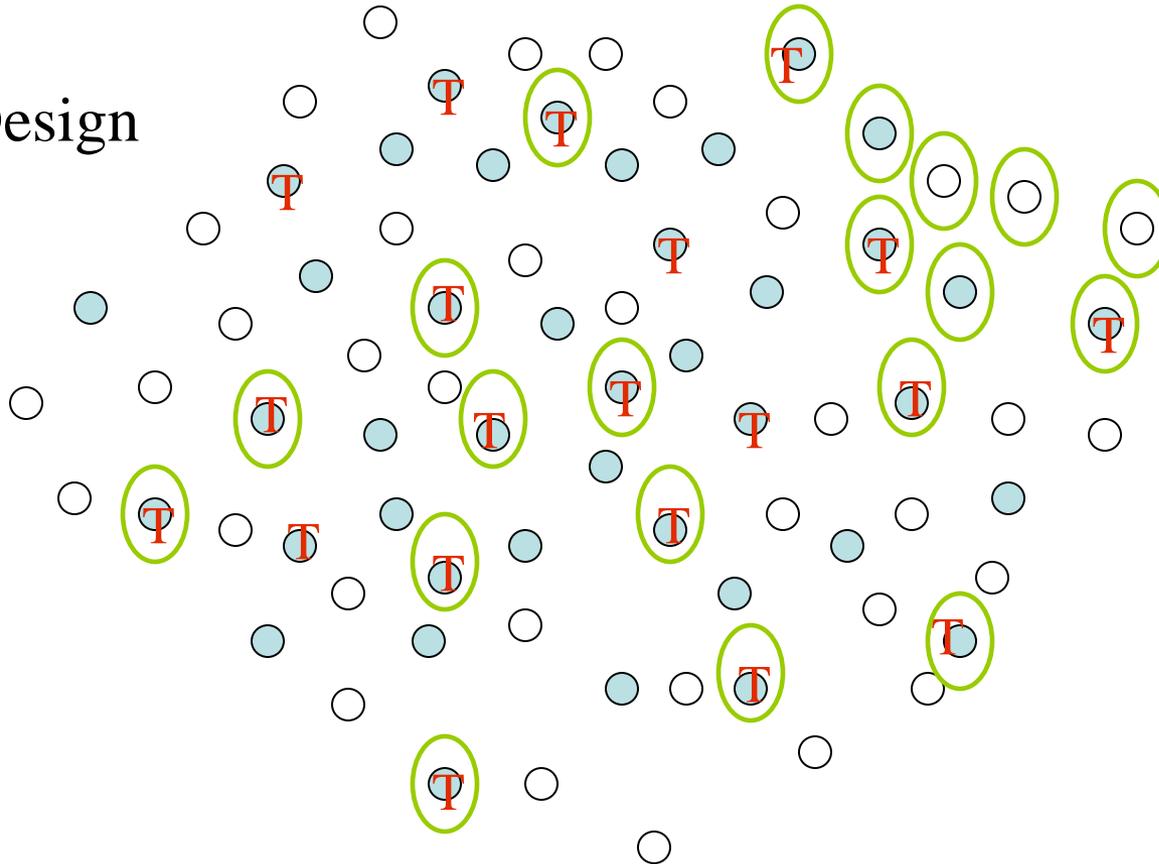
T Your treatment group

● Your sample

○ Other villages

ITT: Another example

Initial Design



T Your treatment group
○ Actual treatment

● Your sample
○ Other villages

How do you measure your impact?

- You cannot compare the actually treated  villages with the sample non-treated villages. 
- Why?
 - Because the villages that should not have been treated but were treated are not randomly selected
 - They are villages with particularly motivated / vocal leaders. They are likely to have better outcomes in any case

How do you measure your impact?

- You can't compare the treated sample villages with the  non-treated sample villages. 
- Why?
 - Because the villages that should have been treated but were NOT treated are not randomly selected
 - They are villages with particularly uncaring leaders (they didn't do anything when they were dropped from the treatment list). They are likely to have worse outcomes in any case

How do you measure your impact?

- Respect the initial assignment!
- You should compare the initial 20 treatment villages with the initial 20 comparison villages
 - Even if some of the treatment villages were not treated
 - Even if some of the comparison villages were treated
 - Ignore the villages outside of your initial sample
- That is, compare [$\textcircled{T} + \textcircled{C}$] with [$\textcircled{C} + \textcircled{C}$]
- This is the ITT estimator: “Intention to Treat”

Compliance Summary

- Intention to Treat
 - Often appropriate for program evaluations
 - Simple to calculate, easy to explain
- Treatment on Treated
 - Measures the treatment effect for those who take the treatment because they are assigned to it
- Think through these issues during the design phase of the experiment

Outline

- I. ITT and TOT
- II. Choice of Outcomes and Covariates**
- III. External Validity
- IV. Cost Effectiveness

Multiple outcomes

- *The more outcomes you look at, the higher the chance you find at least one significantly affected by the program*
- Report results on all measured outcomes, even null results

Heterogeneous treatment effects

- *Data mining vs. policy implication*

→ Report results on all groups

Grouped data

- *People in the same group may be subject to common shocks (other than the treatment)*
 - Need to adjust standard errors for common shocks (clustering)
 - Not very good if number of clusters is small
 - Method when small number of clusters: conduct falsification tests
 - Generate placebo random assignments and see if effect of similar magnitude can be found with likelihood of more than 5% or 10%
 - Look at another outcome that should not have been affected by your treatment –if you find an effect, means there's something else going on...
- Random assignment by group (rather than individual) can have dramatic effects on required sample size

Covariates

- *What to control for?*
 - If stratified randomization: add strata fixed effects
 - Other covariates:
 - Can help increase precision
 - Careful to exclude any variable that could have been affected by the treatment
- Rule: Report both “raw” differences and regression-adjusted results

Outline

- I. ITT and TOT
- II. Choice of Outcomes and Covariates
- III. **External Validity**
“Threats to External Validity”
- IV. Cost Effectiveness

External validity

- Internal validity is a necessary condition for the results of a randomized experiment to be generalizable...
- But it's not sufficient

Threat to external validity: Behavioral responses to evaluations

- One limitation of randomized evaluations is that the evaluation *itself* may cause the treatment or comparison group to change its behavior
 - Treatment group behavior changes: Hawthorne effect
 - Example?
 - Comparison group behavior changes: John Henry effect
 - Example?

Behavioral responses to evaluations

- In addition: a program may generate behavioral responses that would not occur if the program were generalized
 - *Example:* Children may transfer into a school receiving additional inputs, which may effect the original pupils by increasing class size (if class size affects the outcome of interest)

Generalizability of results

Depend on three factors:

- *Program Implementation*: can it be replicated at a large (national) scale?
- *Study Sample*: is it representative?
- *Sensitivity of results*: would a similar, but slightly different program, have same impact?

Trade-off between internal and external validity

Internal validity

- Careful, closely monitored implementation to make sure we know what we are measuring
- Logistical constraints
 - More statistical power if homogenous sample
 - Non-random study location
 - within-site randomization of treatments, but study site not chosen randomly

External validity

- Need to know the impact of a program that's feasible at a large scale with local resources
- Need representative sample

Outline

I. ITT and TOT

II. Choice of Outcomes and Covariates

III. External Validity

IV. Cost Effectiveness

“How do you compare the impact of your program to that of others?”

Comparability tool: Cost-effectiveness

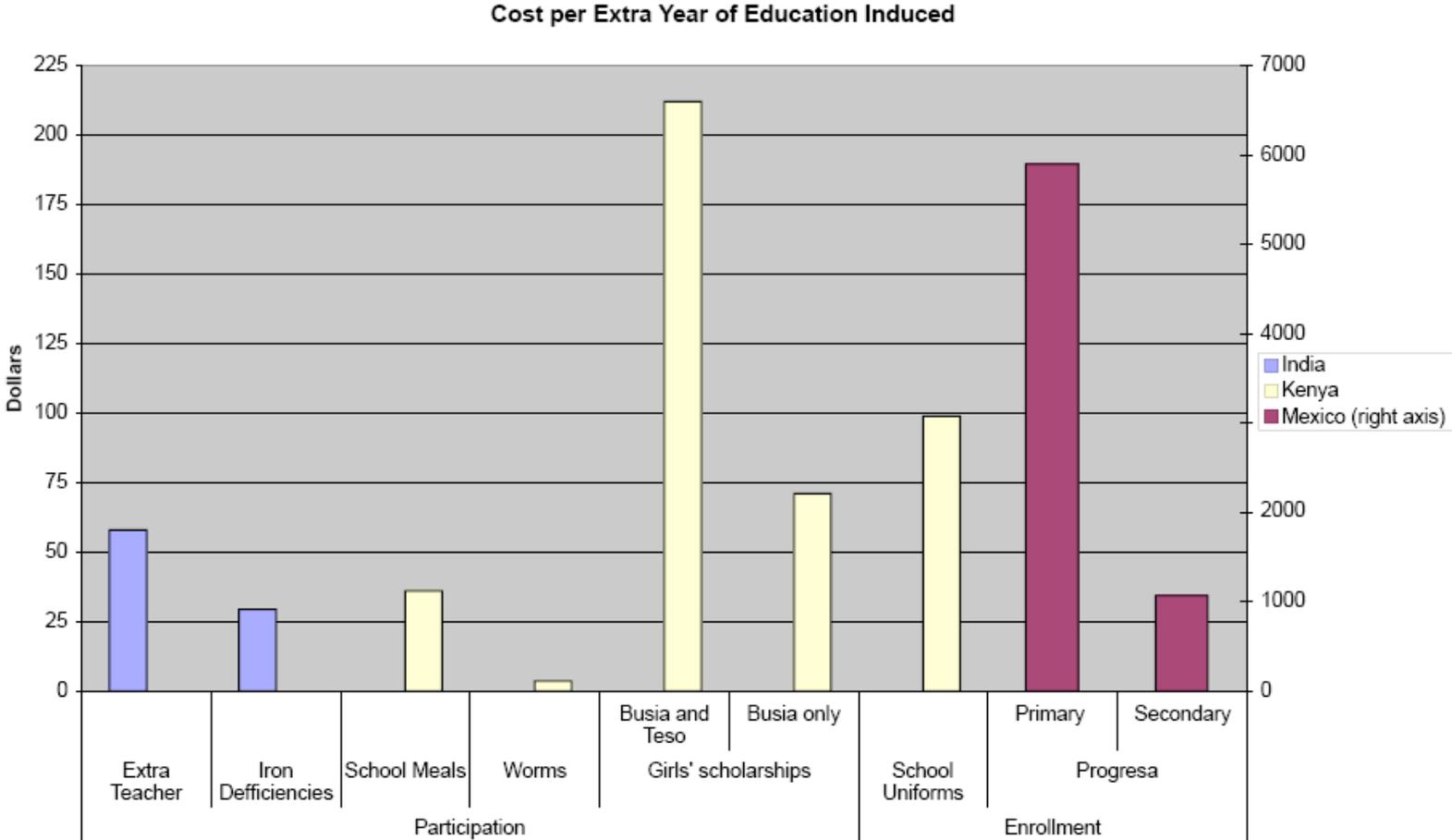
- Could you get the same impact, but at a cheaper price?
- It's important that results of impact studies are expressed in terms of cost-effectiveness so that programs can be compared
- Education programs
 - Cost per additional year of schooling
- Health programs
 - Cost per DALY (Disability-Adjusted Life Year)
 - Cost per death averted

Example:

Cost-effectiveness in education

- Many different approaches have been tried to increase school enrolment: which is the most cost effective?
 - Reducing the cost of school is clearly effective in increasing attendance:
 - Progresa program in Mexico: social welfare payments tied to school attendance
 - Providing free uniforms to school children
 - School meals for preschoolers
 - Incentives to learn increase learning
 - Scholarships for girls
 - Health can have an important education impact
 - Deworming

Cost-effectiveness of education



The case of deworming:

Cost effectiveness and welfare analysis

- Educational cost effectiveness
- Health cost effectiveness
- Human capital investment
- Externalities

Deworming: Educational cost- effectiveness

What is the cost of an additional year of school participation?

- School participation increased by .15 years per treated child
 - Cost of treatment is \$.49 per child in a large-scale program (includes cost of pills, staff, transport)
 - Costs only \$3.27 per additional year of school participation.
 - Much cheaper than other ways of increasing schooling
- The most effective intervention is not always the most obvious!

Health cost effectiveness of school-based deworming

- One of many worthy health interventions competing for scarce resources—annual government per capita expenditures on health is \$5 in Kenya.
- Effectiveness measure: cost per DALY
- Cost per pupil treated is \$0.49

Health cost effectiveness

- Treatment of schistosomiasis is very cost effective, but only meets the strictest standard when take into account the externalities present.
 - \$12 per DALY saved
 - The externality benefits account for 61.1% of the DALY reduction
- Treatment of geohelminths would not meet standard cost
 - Effectiveness criteria in the poorest countries based on health impact alone: \$155 per DALY saved.

Human capital investment

What is the rate of return to deworming in future earnings?

- Knight and Sabot (1987) estimate the avg. labor market return to an additional year of primary schooling in Kenya is 17%
 - Years of schooling alone account for approximately 40% of the total return to education controlling for scores
 - If deworming increases participation but has minimal cognitive gains, the direct return to an additional year of primary education would be 7%
- Deworming increases net present value of wages by more than \$20 per treated child at a cost of only \$0.49
- Excludes health and nutritional gains, thus understating benefits
- Labor market benefits alone seem to far outweigh cost

Conclusion:

Rigorous evaluation impacts policy

- Results of evaluation directly affect the actions of the partner
- Deworming adopted nation-wide in Uganda, in progress in Kenya
- PROGRESA adopted all over Latin America and elsewhere
- Experiments that showed that the threat of audits reduces corruption in Indonesia have led the World Bank and the GOI to increase the probability of audits nationwide

Additional Resources

- Using Randomization in Development Economics Research: A Toolkit
 - Duflo, Glennerster, Kremer
 - JPAL Working Paper
- “Mostly Harmless Econometrics”
 - Angrist and Pischke