J-PAL Executive Education Course in Evaluating Social Programmes

Course Material

University of Cape Town
23 – 27 January 2012
# Table of Contents

Programme .......................................................................................................................... 3  
Maps and Directions ............................................................................................................. 5  
Course Objectives .................................................................................................................. 7  
J-PAL Lecturers ..................................................................................................................... 9  
List of Participants ................................................................................................................ 11  
  Group Assignment .............................................................................................................. 12  
Case Studies ......................................................................................................................... 13  
  Case Study 2: Learn to Read Evaluations ........................................................................ 18  
  Case Study 3: Extra Teacher Program ............................................................................ 27  
  Case Study 4: Deworming in Kenya ................................................................................. 31  
Exercises ............................................................................................................................... 39  
  Exercise A: Understanding random sampling / the law of large numbers .................... 39  
  Exercise B: Sample size calculations .............................................................................. 40  
  Exercise C: The mechanics of random assignment using MS Excel .............................. 45  
Statistics Review ................................................................................................................... 53  
Practical Tips .......................................................................................................................... 62
# Programme

**J-PAL Executive Education Course in Evaluating Social Programmes, 23 – 27 January 2012**

**University of Cape Town**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>8:30 – 10:00</strong></td>
<td><strong>Welcoming Remarks</strong></td>
<td><strong>Group work on case study</strong></td>
<td><strong>Group work on case study</strong></td>
<td><strong>Lecture 7: Project from Start to Finish</strong></td>
</tr>
<tr>
<td><strong>Lecture 1: What is Evaluation</strong></td>
<td><strong>2: Learn to Read</strong></td>
<td><strong>3: Extra Teacher Programme</strong></td>
<td><strong>- Evaluation Design (ctd.)</strong></td>
<td><strong>by Marianne Bertrand, University of Chicago</strong></td>
</tr>
<tr>
<td><strong>by Marc Shotland, J-PAL Global</strong></td>
<td></td>
<td></td>
<td><strong>Group presentation on mechanisms of randomisation (30min)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>10:30 – 12:30</strong></td>
<td><strong>Group Introduction</strong></td>
<td><strong>Lecture 3: Why Randomise by Michael Rosholm, Aarhus University</strong></td>
<td><strong>Group exercise A on random sampling (45min)</strong></td>
<td><strong>Round table discussion: Challenges of Implementing an RCT</strong></td>
</tr>
<tr>
<td><strong>Group work on case study 1: Women as Policymakers</strong></td>
<td><strong>Decision on group project</strong></td>
<td><strong>Primer on Power Calculation (45min)</strong></td>
<td><strong>Group work on case study 4: Deworming in Kenya (90min)</strong></td>
<td><strong>Group work to finalise presentations</strong></td>
</tr>
<tr>
<td><strong>Lunch</strong></td>
<td><strong>Lunch</strong></td>
<td><strong>Lunch</strong></td>
<td><strong>Lunch</strong></td>
<td><strong>Lunch</strong></td>
</tr>
<tr>
<td><strong>13:30 – 15:00</strong></td>
<td><strong>Lecture 2: Measuring Impacts by Isaac Mbiti, MIT</strong></td>
<td><strong>Lecture 4: How to Randomise by Martin Abel, J-PAL Africa</strong></td>
<td><strong>Lecture 5: Sampling and Sample Size by Rebecca Thornton, University of Michigan</strong></td>
<td><strong>Group presentations (each group: 15 min presentation, 15 min discussion)</strong></td>
</tr>
<tr>
<td><strong>15:30 – 17:00</strong></td>
<td><strong>Group presentation - Theory of change, indicators-</strong></td>
<td><strong>Group presentation - Indicators / Evaluation Design-</strong></td>
<td><strong>Group exercise B on sample size estimation (60min)</strong></td>
<td><strong>Group work on presentat. - Manage Threats, Finalize Presentation-</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Group presentation - Power Calculation-</strong></td>
<td><strong>Group work on presentation (30min)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>5.45 Group Picture</strong></td>
<td><strong>6.00 Braai (BBQ) at UCT (both on upper campus)</strong></td>
<td></td>
</tr>
</tbody>
</table>

*Note: All times are in South Africa Standard Time (SAST)*
Maps and Directions

Cape Town – University of Cape Town

Directions to UCT Middle Campus from the airport
To reach the university from the airport, proceed on the N2 towards Cape Town and take the Muizenberg (M3) off-ramp. Continue until you reach and turn off at the Woolsack Drive / University of Cape Town off ramp. Go straight at the traffic lights on Woolsack Drive and enter middle campus. Follow Cross Campus Road until you come to a stop sign. Take a left and after 100m you see the parking lot for the All Africa House and New Economics Building on the left side (J2/3 on map on next page).

Directions to UCT Middle Campus from Cape Town
UCT’s Middle Campus (Groote Schuur Campus) is situated on the slopes of Devil’s Peak in the suburb of Rondebosch. To reach the middle campus from the city, drive along De Waal Drive or Eastern Boulevards, passing Groote Schuur Hospital on the way. Just past the hospital the road forks. Take the right-hand fork (M3 to Muizenberg). Just beyond Mostert’s Mill (windmill) on your left, take the Woolsack Drive / University of Cape Town turn-off. Go straight at the traffic lights on Woolsack Drive and enter middle campus. Follow the road until you come to a stop sign. Take a left and after 100m you see the parking lot for the All Africa House and New Economics Building on the left side (J2/3 on map on next page).
UCT Lower and Middle Campus: New Economics Building – Course Venue
Course Objectives

Our executive training programme is designed for people from a variety of backgrounds: managers and researchers from international development organisations, foundations, governments and non-governmental organisations from around the world, as well as trained economists looking to retool.

The course is a full-time course. It is important for participants to attend all lectures and group work in order to successfully complete the course and receive the certificate of completion.

Course Coverage

The following key questions and concepts will be covered:

- Why and when is a rigorous evaluation of social impact needed?
- The common pitfalls of evaluations, and how randomisation can help.
- The key components of a good randomised evaluation design.
- Alternative techniques for incorporating randomisation into project design.
- How do you determine the appropriate sample size, identify outcome measures, and manage data?
- Guarding against threats that may undermine the integrity of the results.
- Techniques for the analysis and interpretation of results.
- How to maximise policy impact and test external validity.

The programme will achieve these goals through a diverse set of integrated teaching methods. Expert researchers will provide both theoretical and example-based classes complemented by workshops where participants can apply key concepts to real world examples. By examining both successful and problematic evaluations, participants will better understand the significance of various specific details of randomised evaluations. Furthermore, the programme will offer extensive opportunities to apply these ideas ensuring that participants will leave with the knowledge, experience, and confidence necessary to conduct their own randomized evaluations.
J-PAL Lecturers

Martin Abel
Research Manager, J-PAL Africa

Martin Abel holds an MPA in International Development from the Harvard Kennedy School. He is currently working on labour market and financial literacy projects in South Africa and is heading the capacity building program of J-PAL Africa.

Marianne Bertrand
Professor of Economics
University of Chicago Graduate School of Business

Marianne Bertrand is a Professor of Economics at the University of Chicago Graduate School of Business. Her research focuses on racial bias in the US, affirmative action for disadvantaged castes in India, and corruption in India.

Bruno Crépon
Associate Professor
ENSAE and École Polytechnique

Bruno Crépon is a researcher at Centre de Recherche en Économie et Statistique (CREST) and an Associate Professor at ENSAE and École Polytechnique. The focus of his research is on policy evaluation with special attention to labour market policies.

Isaac Mbiti
Assistant Professor
Southern Methodist University

Isaac Mbiti is an assistant professor in the Department of Economics of Southern Methodist University. His research interests are in economic development, labour economics and demography.
Michael Rosholm  
Professor 
Aarhus University

Michael Rosholm received a Ph.D. in economics from the University of Aarhus, Denmark, for his thesis "Transitions in the Labour Market." Since 2006 he has been a professor at the Business and Social Sciences School at Aarhus University, and the Research Director of the Centre for Research in Integration, Education, Qualifications and Marginalization. He is chairman of the Danish Economic Council and researches the effects of active labour market policies on individuals and firms, immigrants in the labour market, and health and employment.

Marc Shotland  
Senior Research Manager  
J-PAL Global

Marc Shotland holds a Masters in Public Administration in International Development (MPA/ID) degree from Harvard University’s Kennedy School of Government and a Bachelors degree in Economics from Williams College. He first joined Professors Duflo and Banerjee in the summer of 2002 to run randomized evaluations of education interventions as a field research associate in India. In 2004 he joined the Poverty Action Lab’s Cambridge office as a research manager. He left in 2006 to earn his Masters at Harvard before rejoining J-PAL in 2008 in his current position.

Rebecca Thornton  
Assistant Professor of Economics 
University of Michigan

Rebecca Thornton began her appointment as an assistant Professor at the University of Michigan economics department in 2008. Her research focuses on education and health as well as how individuals respond to financial incentives in these areas. She has worked on a randomized evaluation of a merit-based scholarship in Kenya. She is also working on randomized evaluations examining HIV testing and prevention and menstruation and education in Nepal.
## List of Participants

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>First Name</th>
<th>Organization</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anyadi</td>
<td>Grace</td>
<td>Ghana Health Service</td>
<td>Ghana</td>
</tr>
<tr>
<td>2</td>
<td>Attah</td>
<td>Emmanuel</td>
<td>National Population Commission</td>
<td>Nigeria</td>
</tr>
<tr>
<td>3</td>
<td>Attandroh</td>
<td>Pius</td>
<td>Ghana Health Service</td>
<td>Ghana</td>
</tr>
<tr>
<td>4</td>
<td>Bianco</td>
<td>James</td>
<td>Department for International Development</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>5</td>
<td>Boumas Ngabina</td>
<td>Guennolet</td>
<td>The Earth Institute</td>
<td>Senegal</td>
</tr>
<tr>
<td>6</td>
<td>Brooks</td>
<td>Rachel</td>
<td>mSwali</td>
<td>Kenya</td>
</tr>
<tr>
<td>7</td>
<td>Campey</td>
<td>Mine</td>
<td>Shonaquip</td>
<td>South Africa</td>
</tr>
<tr>
<td>8</td>
<td>Da Maia</td>
<td>Rui</td>
<td>Technical University of Mozambique</td>
<td>Mozambique</td>
</tr>
<tr>
<td>9</td>
<td>da Maia</td>
<td>Carlos</td>
<td>Stellenbosch University</td>
<td>South Africa</td>
</tr>
<tr>
<td>10</td>
<td>Dimova</td>
<td>Sashka</td>
<td>Aarhusen University</td>
<td>Denmark</td>
</tr>
<tr>
<td>11</td>
<td>Dyrberg</td>
<td>Maria</td>
<td>UNICEF</td>
<td>Ghana</td>
</tr>
<tr>
<td>12</td>
<td>Edwards</td>
<td>Ryan</td>
<td>The Foundation for Development Cooperation</td>
<td>Australia</td>
</tr>
<tr>
<td>13</td>
<td>Fakier</td>
<td>Khayaat</td>
<td>Society Work and Development Institute</td>
<td>South Africa</td>
</tr>
<tr>
<td>14</td>
<td>Firestone</td>
<td>Rebecca</td>
<td>Population Services International</td>
<td>US</td>
</tr>
<tr>
<td>15</td>
<td>Herman-Roloff</td>
<td>Amy</td>
<td>Population Services International</td>
<td>Kenya</td>
</tr>
<tr>
<td>16</td>
<td>James</td>
<td>Ambrose</td>
<td>Search for Common Ground</td>
<td>Sierra Leone</td>
</tr>
<tr>
<td>17</td>
<td>Kettle</td>
<td>Stewart</td>
<td>University of Bristol</td>
<td>UK</td>
</tr>
<tr>
<td>18</td>
<td>Linegar</td>
<td>Margaret</td>
<td>Shonaquip</td>
<td>South Africa</td>
</tr>
<tr>
<td>19</td>
<td>Mashalaba</td>
<td>Noni</td>
<td>DTI South Africa</td>
<td>South Africa</td>
</tr>
<tr>
<td>20</td>
<td>Maughan-Brown</td>
<td>Brendan</td>
<td>UCT</td>
<td>South Africa</td>
</tr>
<tr>
<td>21</td>
<td>Morsy</td>
<td>Awny</td>
<td></td>
<td>Egypt</td>
</tr>
<tr>
<td>22</td>
<td>Mtaki</td>
<td>Sunday</td>
<td>TechnoServe</td>
<td>Tanzania</td>
</tr>
<tr>
<td>23</td>
<td>Mukoya</td>
<td>Josiah</td>
<td>Mercy Corps</td>
<td>Kenya</td>
</tr>
<tr>
<td>24</td>
<td>Nelson</td>
<td>Helen</td>
<td>Evaluation Advisor</td>
<td>UK</td>
</tr>
<tr>
<td>25</td>
<td>Nemaware</td>
<td>Vicky</td>
<td>Cell-Life</td>
<td>South Africa</td>
</tr>
<tr>
<td>26</td>
<td>Ngara-Muraya</td>
<td>Rose</td>
<td>KIPPRA</td>
<td>Kenya</td>
</tr>
<tr>
<td>27</td>
<td>Oliya</td>
<td>Francesca</td>
<td>AVSI</td>
<td>Uganda / Italy</td>
</tr>
<tr>
<td>28</td>
<td>Remoortere</td>
<td>Vanessa</td>
<td>Shonaquip</td>
<td>South Africa</td>
</tr>
<tr>
<td>29</td>
<td>Rinaldi</td>
<td>Anna</td>
<td>University of Bari</td>
<td>Italy</td>
</tr>
<tr>
<td>30</td>
<td>Rugema</td>
<td>Faith</td>
<td>Office of President of Rwanda</td>
<td>Rwanda</td>
</tr>
<tr>
<td>31</td>
<td>Scheffler</td>
<td>Elsje</td>
<td>DARE Consult / Shonaquip</td>
<td>South Africa</td>
</tr>
<tr>
<td>32</td>
<td>Sekandi Kalibbala</td>
<td>Julie</td>
<td>Office of the Prime Minister Uganda</td>
<td>Uganda</td>
</tr>
<tr>
<td>33</td>
<td>Stiglic</td>
<td>Ana</td>
<td>GOAL</td>
<td>Uganda</td>
</tr>
<tr>
<td>34</td>
<td>Taruberekera</td>
<td>Noah</td>
<td>Population Services International</td>
<td>South Africa</td>
</tr>
<tr>
<td>35</td>
<td>Umanah</td>
<td>Eyaekop</td>
<td>Ministry of Finance Nigeria</td>
<td>Nigeria</td>
</tr>
<tr>
<td>36</td>
<td>Wormsbaecher</td>
<td>Ingrid</td>
<td>Shonaquip</td>
<td>South Africa</td>
</tr>
</tbody>
</table>
Group Assignment

**Group 1**
**TA: Willa Brown** (J-PAL Africa)
- Mine Campey (Shonaquip)
- Margaret Linegar (Shonaquip)
- Vanessa Remoortere (Shonaquip)
- Elsje Scheffler (Shonaquip)
- Ingrid Wormsbaecher (Shonaquip)

**Group 2**
**TA: Jessica Kiessel** (IPA Ghana)
- Pius Attandroh (Ghana Health Service)
- Grace Anyandi (Ghana Health Service)
- Emmanuel Attah (Nat. Planning Commission)
- James Bianco (DFID)
- Maria Dyrberg (UNICEF)
- Helen Nelseon (DFID)
- Anna Rinaldi (University of Bari)

**Group 3**
**TA: Rebecca Metz** (J-PAL Africa)
- Rebecca Firestone (Population Service Int.)
- Amy Hermann-Roloff (Population Service Int.)
- Stewart Kettle (University of Bristol)
- Vicky Newbaware (Cell-Life)
- Brendan Maughan-Brown (UCT)
- Noah Taruberekera (Population Service Int.)

**Group 4**
**TA: Clare Hofmeyr** (J-PAL Africa)
- Sashka Dimova (Aarhusen University)
- Khayaat Fakier (Society Work and Dev Institute)
- Noni Mashalaba (DTI South Africa)
- Sunday Mtaki (TechnoServe)
- Francesca Oliva (AVSI)
- Ana Stiglic (GOAL)

**Group 5**
**TA: Shawn Powers** (J-PAL Global)
- Guennolet Boumas (The Earth Institute)
- Rui da Maya (Technical Univ. of Mozambique)
- Ambrose James (Search for Common Ground)
- Awny Morsby
- Rose Ngara-Muraya (KIPPRA)
- Faith Rugema (Office of Presidency, Rwanda)

**Group 6**
**TA: Brian Swartz** (IPA Uganda)
- Rachel Brooks (mSwali)
- Carlos da Maia (Stellenbosch University)
- Ryan Edwards (Foundation for Development C.)
- Josiah Mukoya (Mercy Corps)
- Juliet Sekandi (Office Prime Minister, Uganda)
- Eyaekop Umanah (Ministry of Finance, Nigeria)
Case Studies
Case Study 1: Women as Policymakers

This case study is based on “Women as Policy Makers: Evidence from a Randomized Policy Experiment in India,” by Raghabendra Chattopadhyay and Esther Duflo (2004a), *Econometrica* 72(5), 1409-1443.

J-PAL thanks the authors for allowing us to use their paper.
India amended its federal constitution in 1992, devolving power over local development programs from the states to rural councils, known locally as Panchayats (Village Councils). The Village Councils now choose what development programs to undertake and how much of the budget to invest in them. The states are also required to reserve a third of Village Council seats and Village Council chairperson positions for women. In most states, the schedule on which different villages must reserve seats and positions is determined randomly. This creates the opportunity to rigorously assess the impact of quotas on politics and government: Do the policies differ when there are more women in government? Do the policies chosen by women in power reflect the policy priorities of women? Since randomization was part of the Indian government program itself, the evaluation planning centered on collecting the data needed to measure impact. The researchers’ questions were what data to collect and what data collection instruments to use.

### Empowering the Panchayati Raj

Village Councils, known locally as Panchayats, have a long tradition in rural India. Originally, panchayats were assemblies (yat) of five (panch) elders, chosen by the community, convened to mediate disputes between people or villages. In modern times, Village Councils have been formalized into institutions of local self-government.

This formalization came about through the constitution. In 1992, India enacted the 73rd amendment, which directed the states to establish a three-tier Panchayati Raj system. The Village Council is the grassroots unit of this system, each council consisting of councilors elected every five years. The councilors elect from among themselves a council chairperson called a Pradhan. Decisions are made by a majority vote and the chairperson has no veto power. But as the only councilor with a full-time appointment, the chairperson wields effective power.

---

1. **Hypothesis**: a proposed explanation of and for the effects of a given intervention. Hypotheses are intended to be made ex-ante, or prior to the implementation of the intervention.
2. **Indicators**: metrics used to quantify and measure specific short-term and long-term effects of a program.
3. **Logical Framework**: a management tool used to facilitate the design, execution, and evaluation of an intervention. It involves identifying strategic elements (inputs, outputs, outcomes and impact) and their causal relationships, indicators, and the assumptions and risks that may influence success and failure.
4. **Theory of Change**: describes a strategy or blueprint for achieving a given long-term goal. It identifies the preconditions, pathways and interventions necessary for an initiative’s success.
The 73rd amendment aimed to decentralize the delivery of public goods and services essential for development in rural areas. The states were directed to delegate the power to plan and implement local development programs to the Village Councils. Funds still come from the central government but are no longer earmarked for specific uses. Instead, the Village Council decides which programs to implement and how much to invest in them. As of 2005, Village Councils can choose programs from 29 specified areas, including welfare services (for example, public assistance for widows, care for the elderly, maternity care, antenatal care, and child health) and public works (for example, drinking water, roads, housing, community buildings, electricity, irrigation, and education).

Empowering women in the Panchayati Raj

The Village Councils are large and diverse. In West Bengal, for example, each council represents up to 12 villages and up to 10,000 people, who can vary by religion, ethnicity, caste, and, of course, gender. Political voice varies by group identities drawn along these lines. If policy preferences vary by group identity and if the councillors’ identities influence policy choices, then groups underrepresented in politics and government could be shut out as Village Councils could ignore those groups’ policy priorities. There were fears that the newly empowered Village Councils would undermine the development priorities of traditionally marginalized groups, such as women. To remedy this, the 73rd amendment included two mandates to ensure that investments reflected the needs of everyone in the Village Council.

The first mandate secures community input. If Village Council investments are to reflect a community’s priorities, the councillors must first know what those priorities are. Accordingly, Village Councils are required to hold a general assembly every six months or every year to report on activities in the preceding period and to submit the proposed budget to the community for ratification. In addition, the Chairpersons are required to set up regular office hours to allow constituents to formally request services and lodge complaints. Both requirements allow constituents to articulate their policy preferences.

The second mandate secures representation in the council for women. States are required to reserve at least a third of all council seats and Chairperson positions for women. Furthermore, states must ensure that the seats reserved for women are “allotted by rotation to different constituencies in a Panchayat [Village Council]” and that the chairperson positions reserved for women are “allotted by rotation to different Panchayats [Village Councils].” In other words, they have to ensure that reserved seats and chairperson positions rotate evenly within and across the Village Councils.

Randomized quotas in India: What can it teach us?

Your evaluation team has been entrusted with the responsibility to estimate the impact of quotas for women in the Village Councils. Your evaluation should address all dimensions in which quotas for women are changing local communities in India. What could these dimensions be? What data will you collect? What instruments will you use?

As a first step you want to understand all you can about the quota policy. What needs did it address? What are the pros and cons of the policy? What can we learn from it?
Discussion Topic 1: Gender quotas in the Village Councils

1. What were the main goals of the Village Councils?

2. Women are underrepresented in politics and government. Only 10 percent of India’s national assembly members are women, compared to 17 percent worldwide.

   Does it matter that women are underrepresented? Why and why not?

3. What were the framers of the 73rd amendment trying to achieve when they introduced quotas for women?

Gender quotas have usually been followed by dramatic increases in the political representation of women. Rwanda, for example, jumped from 24th place in the “women in parliament” rankings to first place (49 percent) after the introduction of quotas in 1996. Similar changes have been seen in Argentina, Burundi, Costa Rica, Iraq, Mozambique, and South Africa. Indeed, as of 2005, 17 of the top 20 countries in the rankings have quotas.

Imagine that your group is the national parliament of a country deciding whether to adopt quotas for women in the national parliament. Randomly divide your group into two parties, one against and one for quotas.

What data to collect

First, you need to be very clear about the likely impact of the program. It is on those dimensions that you believe will be affected that you will try to collect data. What are the main areas in which the quota policy should be evaluated? In which areas do you expect to see a difference as a result of quotas?

What are all the possible effects of quotas?

Discussion Topic 2: Using a logical framework to delineate your intermediate and final outcomes of interest

1. Brainstorm the possible effects of quotas, both positive and negative.

2. What evidence would you collect to strengthen the case of those who are for or against quotas? For each potential effect on your list, list also the indicator(s) you would use for that effect. For example, if you say that quotas will affect political participation of women, the indicator could be “number of women attending the General Assembly.”

Multiple outcomes are difficult to interpret, so define a hypothesis

Quotas for women could produce a large number of outcomes in different directions. For example, it may improve the supply of drinking water and worsen the supply of irrigation. Without an ex-ante hypothesis on the direction in which these different variables should be affected by the quota policy, it will be very difficult to make sense of any result we find. Think of the following: if you take 500 villages and randomly assign them in your computer to a “treatment” group and a “control” group, and then run regressions to see whether the villages look different along 100 outcomes, would you expect to see some differences among them? Would it make sense to rationalize those results ex-post?
The same applies to this case: if you just present your report in front of the commission who mandated you to evaluate this policy, explaining that the quota for women changed some variables and did not change others, what are they supposed to make of it? How will they know that these differences are not due to pure chance rather than the policy? You need to present them with a clear hypothesis of how quotas are supposed to change policymaking, which will lead you to make predictions about which outcomes are affected.

**Discussion Topic 2 continued...:**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.</td>
<td>What might be some examples of key hypotheses you would test? Pick one.</td>
</tr>
<tr>
<td>4.</td>
<td>Which indicators or combinations of indicators would you use to test your key hypothesis?</td>
</tr>
</tbody>
</table>

**Use a logical framework to delineate intermediate and final outcomes**

A good way of figuring out the important outcomes is to lay out your theory of change; that is, to draw a logical framework linking the intervention, step by step, to the key final outcomes.

**Discussion Topic 2 continued...:**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>What are the steps or conditions that link quotas (the intervention) to the final outcomes?</td>
</tr>
<tr>
<td>6.</td>
<td>Which indicators should you try to measure at each step in your logical framework?</td>
</tr>
<tr>
<td>7.</td>
<td>Using the outcomes and conditions, draw a possible logical framework, linking the intervention and the final outcomes.</td>
</tr>
</tbody>
</table>
This case study is based on “Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in India,” by Abhijit Banerjee (MIT), Rukmini Banerjee (Pratham), Esther Duflo (MIT), Rachel Glennerster (J-PAL), and Stuti Khemani (The World Bank)

J-PAL thanks the authors for allowing us to use their paper
Key Vocabulary

1. **Counterfactual**: what would have happened to the participants in a program had they not received the intervention. The counterfactual cannot be observed from the treatment group; can only be inferred from the comparison group.

2. **Comparison Group**: in an experimental design, a randomly assigned group from the same population that does not receive the intervention that is the subject of evaluation. Participants in the comparison group are used as a standard for comparison against the treated subjects in order to validate the results of the intervention.

3. **Program Impact**: estimated by measuring the difference in outcomes between comparison and treatment groups. The true impact of the program is the difference in outcomes between the treatment group and its counterfactual.

4. **Baseline**: data describing the characteristics of participants measured across both treatment and comparison groups prior to implementation of intervention.

5. **Endline**: data describing the characteristics of participants measured across both treatment and comparison groups after implementation of intervention.

6. **Selection Bias**: statistical bias between comparison and treatment groups in which individuals in one group are systematically different from those in the other. These can occur when the treatment and comparison groups are chosen in a non-random fashion so that they differ from each other by one or more factors that may affect the outcome of the study.

7. **Omitted Variable Bias**: statistical bias that occurs when certain variables/characteristics (often unobservable), which affect the measured outcome, are omitted from a regression analysis. Because they are not included as controls in the regression, one incorrectly attributes the measured impact solely to the program.

Why Learn to Read (L2R)?

In a large-scale survey conducted in 2004, Pratham discovered that only 39% of children (aged 7-14) in rural Uttar Pradesh could read and understand a simple story, and nearly 15% could not recognize even a letter.

During this period, Pratham was developing the “Learn-to-Read” (L2R) module of its Read India campaign. L2R included a unique pedagogy teaching basic literacy skills, combined with a grassroots organizing effort to recruit volunteers willing to teach.

This program allowed the community to get involved in children’s education more directly through village meetings where Pratham staff shared information on the status of literacy in the village and the rights of children to education. In these meetings, Pratham identified community members who were willing to teach. Volunteers attended a training session on the pedagogy, after which they could hold after-school reading classes for children, using materials designed and provided by Pratham. Pratham staff paid occasional visits to these camps to ensure that the classes were being held and to provide additional training as necessary.

Did the Learn to Read project work?

Did Pratham’s “Learn to Read” program work? What is required in order for us to measure whether a program worked or, in other words, whether it had impact?
In general, to ask if a program works is to ask if the program achieves its goal of changing certain outcomes for its participants, and ensure that those changes are not caused by some other factors or events happening at the same time. To show that the program causes the observed changes, we need to simultaneously show that if the program had not been implemented, the observed changes would not have occurred (or would be different). But how do we know what would have happened? If the program happened, it happened. Measuring what would have happened requires entering an imaginary world in which the program was never given to these participants. The outcomes of the same participants in this imaginary world are referred to as the counterfactual. Since we cannot observe the true counterfactual, the best we can do is to estimate it by mimicking it.

The key challenge of program impact evaluation is constructing or mimicking the counterfactual. We typically do this by selecting a group of people that resemble the participants as much as possible but who did not participate in the program. This group is called the comparison group. Because we want to be able to say that it was the program and not some other factor that caused the changes in outcomes, it is important that the only difference between the comparison group and the participants is that the comparison group did not participate in the program. We then estimate “impact” as the difference observed at the end of the program between the outcomes of the comparison group and the outcomes of the program participants.

The impact estimate is only as accurate as the comparison group is successful at mimicking the counterfactual. If the comparison group poorly represents the counterfactual, the impact is (in most circumstances) poorly estimated. Therefore the method used to select the comparison group is a key decision in the design of any impact evaluation.

That brings us back to our questions: Did the Learn to Read project work? What was its impact on children’s reading levels?

In this case, the intention of the program is to “improve children’s reading levels” and the reading level is the outcome measure. So, when we ask if the Learn to Read project worked, we are asking if it improved children’s reading levels. The impact is the difference between reading levels after the children have taken the reading classes and what their reading level would have been if the reading classes had never existed.

For reference, Reading Level is an indicator variable that takes value 0 if the child can read nothing, 1 if he knows the alphabet, 2 if he can recognize words, 3 if he can read a paragraph, and 4 if he can read a full story.

What comparison groups can we use? The following experts illustrate different methods of evaluating impact. (Refer to the table on the last page of the case for a list of different evaluation methods).

Estimating the impact of the Learn to Read project

Method 1:

News Release: Read India helps children Learn to Read.
Pratham celebrates the success of its “Learn to Read” program—part of the Read India Initiative. It has made significant progress in its goal of improving children’s literacy rates through better learning materials, pedagogical methods, and most importantly, committed volunteers. The achievement of the “Learn to Read” (L2R) program demonstrates that a revised curriculum, galvanized by community mobilization, can produce significant gains. Massive government expenditures in mid-day meals and school construction have failed to achieve similar results. In less than a year, the reading levels of children who enrolled in the L2R camps improved considerably.
Just before the program started, half these children could not recognize Hindi words—many nothing at all. But after spending just a few months in Pratham reading classes, more than half
improved by at least one reading level, with a significant number capable of recognizing words and several able to read full paragraphs and stories! On average, the literacy measure of these students improved by nearly one full reading level during this period.

**Discussion Topic 1:**

1. What type of evaluation does this news release imply?
2. What represents the counterfactual?
3. What are the problems with this type of evaluation?

**Method 2:**

**Opinion: The “Read India” project not up to the mark**

Pratham has raised millions of dollars, expanding rapidly to cover all of India with its so-called “Learn-to-Read” program, but do its students actually learn to read? Recent evidence suggests otherwise. A team of evaluators from Education for All found that children who took the reading classes ended up with literacy levels significantly below those of their village counterparts. After one year of Pratham reading classes, Pratham students could only recognize words whereas those who steered clear of Pratham programs were able to read full paragraphs.

*Comparison of reading levels of children who took reading classes Vs. reading levels of children who did not take them*

Notes: Reading Level is an indicator variable that takes value 0 if the child can read nothing, 1 if he knows the alphabet, 2 if he can recognize words, 3 if he can read a paragraph and 4 if he can read a full story.

If you have a dime to spare, and want to contribute to the education of India’s illiterate children, you may think twice before throwing it into the fountain of Pratham’s promises.
Discussion Topic 2:

1. **What type of evaluation is this opinion piece employing?**
2. **What represents the counterfactual?**
3. **What are the problems with this type of evaluation?**

**Method 3:**

**Letter to the Editor: EFA should consider Evaluating Fairly and Accurately**

There have been several unfair reports in the press concerning programs implemented by the NGO Pratham. A recent article by a former Education for All bureaucrat claims that Pratham is actually hurting the children it recruits into its ‘Learn-to-Read’ camps. However, the EFA analysis uses the wrong metric to measure impact. It compares the reading *levels* of Pratham students with other children in the village—not taking into account the fact that Pratham targets those whose literacy levels are particularly poor at the beginning. If Pratham simply recruited the most literate children into their programs, and compared them to their poorer counterparts, they could claim success without conducting a single class. But Pratham does not do this. And realistically, Pratham does not expect its illiterate children to overtake the stronger students in the village. It simply tries to initiate improvement over the current state. Therefore the metric should be *improvement* in reading levels—not the final level. When we repeated EFA's analysis using the more-appropriate outcome measure, the Pratham kids improved at twice the rate of the non-Pratham kids (0.6 reading level increase compared to 0.3). This difference is statistically very significant.

Had the EFA evaluators thought to look at the more appropriate outcome, they would recognize the incredible success of Read India. Perhaps they should enroll in some Pratham classes themselves.

Discussion Topic 3:

1. **What type of evaluation is this letter using?**
2. **What represents the counterfactual?**
3. **What are the problems with this type of evaluation?**

**Method 4:**

**The numbers don’t lie, unless your statisticians are asleep**

Pratham celebrates victory, opponents cry foul. A closer look shows that, as usual, the truth is somewhere in between.

There has been a war in the press between Pratham’s supporters and detractors. Pratham and its advocates assert that the Read India campaign has resulted in large increases in child literacy. Several detractors claim that Pratham programs, by pulling attention away from the schools, are in fact causing significant harm to the students. Unfortunately, this battle is being waged using instruments of analysis that are seriously flawed. The ultimate victim is the public who is looking for an answer to the question: is Pratham helping its intended beneficiaries?

This report uses sophisticated statistical methods to measure the true impact of Pratham programs. We were concerned about other variables confounding previous results. We therefore conducted a survey in these villages to collect information on child age, grade-level, and parents’ education level, and used those to predict child test scores.
Looking at Table 1, we find some positive results, some negative results and some “no-results”, depending on which variables we control for. The results from column (1) suggest that Pratham’s program hurt the children. There is a negative correlation between receiving Pratham classes and final reading outcomes (-0.68). Column (3), which evaluates improvement, suggests impressive results (0.24). But looking at child outcomes (either level or improvement) controlling for initial reading levels, age, gender, standard and parent’s education level – all determinants of child reading levels – we found no impact of Pratham programs.

Therefore, controlling for the right variables, we have discovered that on one hand, Pratham has not caused the harm claimed by certain opponents, but on the other hand, it has not helped children learn. Pratham has therefore failed in its effort to convince us that it can spend donor money effectively.

Discussion Topic 4:
1. What type of evaluation is this report utilizing?
2. What represents the counterfactual?
3. What are the problems with this type of evaluation?

NOTE: Data used in this case are real. “Articles” on the debate were artificially produced for the purpose of the case. Education for All (EFA) never made any of the claims described herein
<table>
<thead>
<tr>
<th>Methodology</th>
<th>Description</th>
<th>Who is in the comparison group?</th>
<th>Required Assumptions</th>
<th>Required Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quasi-Experimental Methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pre-Post</strong></td>
<td>Measure how program participants improved (or changed) over time.</td>
<td>Program participants themselves—before participating in the program.</td>
<td>The program was the only factor influencing any changes in the measured outcome over time.</td>
<td>Before and after data for program participants.</td>
</tr>
<tr>
<td><strong>Simple Difference</strong></td>
<td>Measure difference between program participants and non-participants after the program is completed.</td>
<td>Individuals who didn’t participate in the program (for any reason), but for whom data were collected after the program.</td>
<td>Non-participants are identical to participants except for program participation, and were equally likely to enter program before it started.</td>
<td>After data for program participants and non-participants.</td>
</tr>
<tr>
<td><strong>Differences in Differences</strong></td>
<td>Measure improvement (change) over time of program participants relative to the improvement (change) of non-participants.</td>
<td>Individuals who didn’t participate in the program (for any reason), but for whom data were collected both before and after the program.</td>
<td>If the program didn’t exist, the two groups would have had identical trajectories over this period.</td>
<td>Before and after data for both participants and non-participants.</td>
</tr>
<tr>
<td><strong>Multivariate Regression</strong></td>
<td>Individuals who received treatment are compared with those who did not, and other factors that might explain differences in the outcomes are “controlled” for.</td>
<td>Individuals who didn’t participate in the program (for any reason), but for whom data were collected both before and after the program. In this case data is not comprised of just indicators of outcomes, but other “explanatory” variables as well.</td>
<td>The factors that were excluded (because they are unobservable and/or have been not been measured) do not bias results because they are either uncorrelated with the outcome or do not differ between participants and non-participants.</td>
<td>Outcomes as well as “control variables” for both participants and non-participants.</td>
</tr>
<tr>
<td><strong>Statistical Matching</strong></td>
<td>Individuals in control group are compared to similar individuals in experimental group.</td>
<td><strong>Exact matching:</strong> For each participant, at least one non-participant who is identical on selected characteristics. <strong>Propensity score matching:</strong> non-participants who have a mix of characteristics which predict that they would be as likely to participate as participants.</td>
<td>The factors that were excluded (because they are unobservable and/or have been not been measured) do not bias results because they are either uncorrelated with the outcome or do not differ between participants and non-participants.</td>
<td>Outcomes as well as “variables for matching” for both participants and non-participants.</td>
</tr>
<tr>
<td><strong>Regression Discontinuity Design</strong></td>
<td>Individuals are ranked based on specific, measurable criteria. There is some cutoff that determines whether an individual is eligible to participate. Participants are then compared to non-participants and the eligibility criterion is controlled for.</td>
<td>Individuals who are close to the cutoff, but fall on the “wrong” side of that cutoff, and therefore do not get the program.</td>
<td>After controlling for the criteria (and other measures of choice), the remaining differences between individuals directly below and directly above the cut-off score are not statistically significant and will not bias the results. A necessary but sufficient requirement for this to hold is that the cut-off criteria are strictly adhered to.</td>
<td>Outcomes as well as measures on criteria (and any other controls).</td>
</tr>
<tr>
<td>Methodology</td>
<td>Description</td>
<td>Who is in the comparison group?</td>
<td>Required Assumptions</td>
<td>Required Data</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Instrumental Variables</strong></td>
<td>Participation can be predicted by an incidental (almost random) factor, or “instrumental” variable, that is uncorrelated with the outcome, other than the fact that it predicts participation (and participation affects the outcome).</td>
<td>Individuals who, because of this close to random factor, are predicted not to participate and (possibly as a result) did not participate.</td>
<td>If it weren’t for the instrumental variable’s ability to predict participation, this “instrument” would otherwise have no effect on or be uncorrelated with the outcome.</td>
<td>Outcomes, the “instrument,” and other control variables.</td>
</tr>
<tr>
<td><strong>Experimental Method</strong></td>
<td><strong>Randomized Evaluation</strong> Experimental method for measuring a causal relationship between two variables.</td>
<td>Participants are randomly assigned to the control groups.</td>
<td>Randomization “worked.” That is, the two groups are statistically identical (on observed and unobserved factors).</td>
<td>Outcome data for control and experimental groups. Control variables can help absorb variance and improve “power”.</td>
</tr>
</tbody>
</table>
Case Study 3: Extra Teacher Program

This case study is based on the paper “Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” by Esther Duflo (MIT), Pascaline Dupas (UCLA), and Michael Kremer (Harvard)

J-PAL thanks the authors for allowing us to use their paper
Key Vocabulary

1. **Level of Randomization**: the level of observation (ex. individual, household, school, village) at which treatment and comparison groups are randomly assigned.

Confronted with overcrowded schools and a shortage of teachers, in 2005 the NGO International Child Support Africa (ICS) offered to help the school system of Western Kenya by introducing contract teachers in 120 primary schools. Under its two year program, ICS provided funds to these schools to hire one extra teacher per school. In contrast to the civil servants hired by the Ministry of Education, contract teachers are hired locally by school committees. ICS expected this program to improve student learning by, among other things, decreasing class size and using teachers who are more directly accountable to the communities they serve. However, contract teachers tend to have less training and receive a lower monthly salary than their civil servant counterparts. So there was concern about whether these teachers were sufficiently motivated, given their compensation, or qualified given their credentials.

What experimental designs could test the impact of this intervention on educational achievement? Which of these changes in the school landscape is primarily responsible for improved student performance?

**Over-crowded schools**

Like many other developing countries, Kenya has recently made rapid progress toward the Millennium Development Goal of universal primary education. Largely due to the elimination of school fees in 2003, primary school enrollment rose nearly 30 percent, from 5.9 million to 7.6 million between 2002 and 2005.\(^2\)

Without accompanying government funding, however, this progress has created its own set of new challenges in Kenya:

1) **Large class size**: Due to budget constraints, the rise in primary school enrollment has not been matched by proportional increases in the number of teachers. (Teacher salaries already account for the largest component of educational spending.) The result has been very large class sizes, particularly in lower grades. In a sample of schools in Western Kenya, for example, the average first grade class in 2005 was 83 students. This is concerning because it is believed that small classes are most important for the youngest students, who are still acclimating to the school environment. The Kenyan National Union of Teachers estimates that the country needs an additional 60,000 primary school teachers in addition to the existing 175,000 in order to reach all primary students and decrease class sizes.

2) **Teacher absenteeism**: Further exacerbating the problem of pupil-teacher ratios, teacher absenteeism remains high, reaching nearly 20% in some areas of Kenya.

There are typically no substitutes for absent teachers, so students simply mill around, go home or join another class, often of a different grade. Small schools, which are prevalent in rural areas of developing countries, may be closed entirely as a result of teacher absence. Families have to consider whether school will even be open when deciding

---

whether or not to send their children to school. An obvious result is low student attendance—even on days when the school is open.

3) **Heterogeneous classes:** Classes in Kenya are also very heterogeneous with students varying widely in terms of school preparedness and support from home.

Grouping students into classes sorted by ability (*tracking, or streaming*) is controversial among academics and policymakers. On one hand, if teachers find it easier to teach a homogeneous group of students, tracking could improve school effectiveness and test scores. Many argue, on the other hand, that if students learn in part from their peers, tracking could disadvantage low achieving students while benefiting high achieving students, thereby exacerbating inequality.

4) **Scarce school materials:** Because of the high costs of educational inputs and the rising number of students, educational resources other than the teacher are stretched, and in some cases up to four students must share one textbook. And an already over-burdened infrastructure deteriorates faster when forced to serve more children.

5) **Low completion rates:** As a result of these factors, completion rates are very low in Kenya with only 45.1% of boys and 43.3% of girls completing the first grade.

All in all, these issues pose new challenges to communities: how to ensure minimum quality of education given Kenya’s budget constraints.

**What are Contract Teachers?**

Governments in several developing countries have responded to similar challenges by staffing unfilled teaching positions with locally-hired contract teachers who are not civil service employees. The four main characteristics of contract teachers are that they are: (1) appointed on annual renewable contracts, with no guarantee of renewed employment (unlike regular civil service teachers); (2) often less qualified than regular teachers and much less likely to have a formal teacher training certificate or degree; (3) paid lower salaries than those of regular teachers (typically less than a fifth of the salaries paid to regular teachers); and (4) more likely to be from the local area where the school is located.

**Are Contract Teachers Effective?**

The increasing use of contract teachers has been one of the most significant policy innovations in providing primary education in developing countries, but it has also been highly controversial. Supporters say that using contract teachers is an efficient way of expanding education access and quality to a large number of first-generation learners. Knowing that the school committee’s decision of whether or not to rehire them the following year may hinge on performance, contract teachers are motivated to try harder than their tenured government counterparts. Contract teachers are also often more similar to their students, geographically, culturally, and socioeconomically.

Opponents argue that using under-qualified and untrained teachers may staff classrooms, but will not produce learning outcomes. Furthermore the use of contract teachers de-professionalizes teaching, reduces the prestige of the entire profession, and reduces motivation of all teachers. Even if it helps in the short term, it may hurt efforts to recruit highly qualified teachers in the future.

While the use of contract teachers has generated much controversy, there is very little rigorous evidence regarding the effectiveness of contract teachers in improving student learning outcomes.
The Extra Teacher Program Randomized Evaluation

In January 2005, International Child Support Africa initiated a two year program to examine the effect of contract teachers on education in Kenya. Under the program, ICS gave funds to 120 local school committees to hire one extra contract teacher to teach an additional first grade class. The purpose of this intervention was to address the first three challenges: class size, teacher accountability, and heterogeneity of ability. The evaluation was designed to measure the impact of class-size reductions, the relative effectiveness of contract teachers, and how tracking by ability would impact both low and high-achieving students.

Addressing Multiple Research Questions through Experimental Design

Different randomization strategies may be used to answer different questions. What strategies could be used to evaluate the following questions? How would you design the study? Who would be in the treatment and control groups, and how would they be randomly assigned to these groups?

Discussion Topic 1: Testing the effectiveness of contract teachers

1. What is the relative effectiveness of contract teachers versus regular government teachers?

Discussion Topic 2: Looking at more general approaches of improving education

1. What is the effect of grouping students by ability on student performance?
2. What is the effect of smaller class sizes on student performance?

Discussion Topic 3: Addressing all questions with a single evaluation

1. Could a single evaluation explore all of these issues at once?
2. What randomization strategy could do so?
Case Study 4: Deworming in Kenya

Managing threats to experimental integrity

This case study is based on Edward Miguel and Michael Kremer, “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,” Econometrica 72(1): 159-217, 2004

J-PAL thanks the authors for allowing us to use their paper
Key Vocabulary

1. **Phase-in Design**: a study design in which groups are individually phased into treatment over a period of time; groups which are scheduled to receive treatment later act as the comparison groups in earlier rounds.

2. **Equivalence**: groups are identical on all baseline characteristics, both observable and unobservable. Ensured by randomization.

3. **Attrition**: the process of individuals joining in or dropping out of either the treatment or comparison group over the course of the study.

4. **Attrition Bias**: statistical bias which occurs when individuals systematically join in or drop out of either the treatment or the comparison group for reasons related to the treatment.

5. **Partial Compliance**: individuals do not comply with their assignment (to treatment or comparison). Also termed "diffusion" or "contamination."

6. **Intention to Treat**: the measured impact of a program that includes all data from participants in the groups to which they were randomized, regardless of whether they actually received the treatment. Intention-to-treat analysis prevents bias caused by the loss of participants, which may disrupt the baseline equivalence established by randomization and which may reflect non-adherence to the protocol.

7. **Treatment on the Treated**: the measured impact of a program that includes only the data for participants who actually received the treatment.

8. **Externality**: an indirect cost or benefit incurred by individuals who did not directly receive the treatment. Also termed "spillover."

Between 1998 and 2001, the NGO International Child Support Africa implemented a school-based mass deworming program in 75 primary schools in western Kenya. The program treated the 30,000 pupils enrolled at these schools for worms—hookworm, roundworm, whipworm, and schistosomiasis. Schools were phased-in randomly.

Randomization ensures that the treatment and comparison groups are comparable **at the beginning**, but it cannot ensure that they remain comparable **until the end of the program**. Nor can it ensure that people comply with the treatment they were assigned. Life also goes on after the randomization: other events besides the program happen between initial randomization and the end-line. These events can reintroduce selection bias; they diminish the validity of the impact estimates and are threats to the integrity of the experiment.

How can common threats to experimental integrity be managed?

**Worms—a common problem with a cheap solution**

Worm infections account for over 40 percent of the global tropical disease burden. Infections are common in areas with poor sanitation. More than 2 billion people are affected. Children, still learning good sanitary habits, are particularly vulnerable: 400 million school-age children are chronically infected with intestinal worms.

Worms affect more than the health of children. Symptoms include listlessness, diarrhea, abdominal pain, and anemia. Beyond their effects on health and nutrition, heavy worm infections can impair children’s physical and mental development and reduce their attendance and performance in school.
Poor sanitation and personal hygiene habits facilitate transmission. Infected people excrete worm eggs in their feces and urine. In areas with poor sanitation, the eggs contaminate the soil or water. Other people are infected when they ingest contaminated food or soil (hookworm, whipworm, and roundworm), or when hatched worm larvae penetrate their skin upon contact with contaminated soil (hookworm) or fresh water (schistosome). School-age children are more likely to spread worms because they have riskier hygiene practices (more likely to swim in contaminated water, more likely to not use the latrine, less likely to wash hands before eating). So treating a child not only reduces her own worm load; it may also reduce disease transmission—and so benefit the community at large.

Treatment kills worms in the body, but does not prevent re-infection. Oral medication that can kill 99 percent of worms in the body is available: albendazole or mebendazole for treating hookworm, roundworm, and whipworm infections; and praziquantel for treating schistosomiasis. These drugs are cheap and safe. A dose of albendazole or mebendazole costs less than 3 US cents while one dose of praziquantel costs less than 20 US cents. The drugs have very few and minor side effects.

Worms colonize the intestines and the urinary tract, but they do not reproduce in the body; their numbers build up only through repeated contact with contaminated soil or water. The WHO recommends presumptive school-based mass deworming in areas with high prevalence. Schools with hookworm, whipworm, and roundworm prevalence over 50 percent should be mass treated with albendazole every 6 months, and schools with schistosomiasis prevalence over 30 percent should be mass treated with praziquantel once a year.

**Primary School Deworming Program**

International Child Support Africa (ICS) implemented the Primary School Deworming Program (PSDP) in the Busia District in western Kenya, a densely-settled region with high worm prevalence. Treatment followed WHO guidelines. The medicine was administered by public health nurses from the Ministry of Health in the presence of health officers from ICS.

The PSDP was expected to affect health, nutrition, and education. To measure impact, ICS collected data on a series of outcomes: prevalence of worm infection, worm loads (severity of worm infection); self-reported illness; and school participation rates and test scores.

**Evaluation design — the experiment as planned**

Because of administrative and financial constraints the PSDP could not be implemented in all schools immediately. Instead, the 75 schools were randomly divided into 3 groups of 25 schools and phased-in over 3 years. Group 1 schools were treated starting in both 1998 and 1999, Group 2 schools in 1999, and Group 3 starting in 2001. Group 1 schools were the treatment group in 1998, while schools Group 2 and Group 3 were the comparison. In 1999 Group 1 and Group 2 schools were the treatment and Group 3 schools the comparison.

**Figure 1:** The planned experiment: the PSDP treatment timeline showing experimental groups in 1998 and 1999

<table>
<thead>
<tr>
<th>Group</th>
<th>1998</th>
<th>1999</th>
<th>2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Treatment</td>
<td>Treatment</td>
<td>Treatment</td>
</tr>
<tr>
<td>Group 2</td>
<td>Comparison</td>
<td>Treatment</td>
<td>Treatment</td>
</tr>
<tr>
<td>Group 3</td>
<td>Comparison</td>
<td>Comparison</td>
<td>Treatment</td>
</tr>
</tbody>
</table>
Threats to integrity of the planned experiment

Discussion Topic 1: Threats to experimental integrity

Randomization ensures that the groups are equivalent, and therefore comparable, at the beginning of the program. The impact is then estimated as the difference in the average outcome of the treatment group and the average outcome of the comparison group, both at the end of the program. To be able to say that the program caused the impact, you need to be able to say that the program was the only difference between the treatment and comparison groups over the course of the evaluation.

1. What does it mean to say that the groups are equivalent at the start of the program?

2. Can you check if the groups are equivalent at the beginning of the program? How?

3. Other than the program’s direct and indirect impacts, what can happen over the course of the evaluation (after conducting the random assignment) to make the groups non-equivalent?

4. How does non-equivalence at the end threaten the integrity of the experiment?
Managing attrition—when the groups do not remain equivalent

Attrition is when people join or drop out of the sample—both treatment and comparison groups—over the course of the experiment. One common example in clinical trials is when people die; so common indeed that attrition is sometimes called experimental mortality.

Discussion Topic 2: Managing Attrition

You are looking at the health effects of deworming. In particular you are looking at the worm load (severity of worm infection). Worm loads are scaled as follows:

- Heavy worm infections = score of 3
- Medium worm infections = score of 2
- Light infections = score of 1

There are 30,000 children: 15,000 in treatment schools and 15,000 in comparison schools. After you randomize, the treatment and comparison groups are equivalent, meaning children from each of the three categories are equally represented in both groups.

Suppose protocol compliance is 100 percent: all children who are in the treatment get treated and none of the children in the comparison are treated. Children that were dewormed at the beginning of the school year (that is, children in the treatment group) end up with a worm load of 1 at the end of the year because of re-infection. Children who have a worm load of 3 only attend half the time and drop out of school if they are not treated. The number of children in each worm-load category is shown for both the pretest and posttest.

### Worm Load

<table>
<thead>
<tr>
<th>Worm Load</th>
<th>Pretest Treatment</th>
<th>Pretest Comparison</th>
<th>Posttest Treatment</th>
<th>Posttest Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5,000</td>
<td>5,000</td>
<td>0</td>
<td>Dropped out 5,000</td>
</tr>
<tr>
<td>2</td>
<td>5,000</td>
<td>5,000</td>
<td>0</td>
<td>5,000</td>
</tr>
<tr>
<td>1</td>
<td>5,000</td>
<td>5,000</td>
<td>15,000</td>
<td>5,000</td>
</tr>
<tr>
<td>Total children tested at school</td>
<td>15,000</td>
<td>15,000</td>
<td>15,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>

1. a. At posttest, what is the average worm load for the treatment group?  
   b. At posttest, what is the average worm load for the comparison group?  
   c. What is the difference?  
   d. Is this outcome difference an accurate estimate of the impact of the program? Why or why not?  
   e. If it is not accurate, does it overestimate or underestimate the impact?  
   f. How can we get a better estimate of the program's impact?

2. Besides worm load, the PSDP also looked at outcome measures such as school attendance rates and test scores.  
   a. Would differential attrition (i.e. differences in drop-outs between treatment and comparison groups) bias either of these outcomes? How?  
   b. Would the impacts on these final outcome measures be underestimated or overestimated?

3. In Case 1, you learned about other methods to estimate program impact, such as pre-post, simple difference, differences in differences, and multivariate regression.  
   a. Does the threat of attrition only present itself in randomized evaluations?
Managing partial compliance—when the treatment group does not actually get treated or the comparison group does get treated

Some people assigned to the treatment may in the end not actually get treated. In an after-school tutoring program, for example, some children assigned to receive tutoring may simply not show up for tutoring. And the others assigned to the comparison may obtain access to the treatment, either from the program or from another provider. Or comparison group children may get extra help from the teachers or acquire program materials and methods from their classmates. In any of these scenarios, people are not complying with their assignment in the planned experiment. This is called “partial compliance” or “diffusion” or, less benignly, “contamination.” In contrast to carefully-controlled lab experiments, diffusion is ubiquitous in social programs. After all, life goes on, people will be people, and you have no control over what they decide to do over the course of the experiment. All you can do is plan your experiment and offer them treatments. How, then, can you deal with the complications that arise from partial compliance?

Discussion Topic 3: Managing partial compliance

Suppose none of the children from the poorest families have shoes and so they have worm loads of 3. Though their parents had not paid the school fees, the children were allowed to stay in school during the year. Parental consent was required for treatment, and to give consent, the parents had to come to the school and sign a consent form in the headmaster’s office. However, because they had not paid school fees, the poorest parents were reluctant to come to the school. Consequently, none of the children with worm loads of 3 were actually dewormed. Their worm load scores remained 3 at the end of the year. No one assigned to comparison was treated. All the children in the sample at the beginning of the year were followed up, if not at school then at home.

<table>
<thead>
<tr>
<th>Worm Load</th>
<th>Pretest Treatment</th>
<th>Pretest Comparison</th>
<th>Posttest Treatment</th>
<th>Posttest Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5,000</td>
<td>5,000</td>
<td>5,000</td>
<td>5,000</td>
</tr>
<tr>
<td>2</td>
<td>5,000</td>
<td>5,000</td>
<td>0</td>
<td>5,000</td>
</tr>
<tr>
<td>1</td>
<td>5,000</td>
<td>5,000</td>
<td>10,000</td>
<td>5,000</td>
</tr>
<tr>
<td>Total children tested at school</td>
<td>15,000</td>
<td>15,000</td>
<td>15,000</td>
<td>15,000</td>
</tr>
</tbody>
</table>

1. Calculate the impact estimate based on the original group assignments. 
   a. This is an unbiased measure of the effect of the program, but in what ways is it useful and in what ways is it not as useful?

You are interested in learning the effect of treatment on those actually treated (“treatment on the treated” (TOT) estimate). 

2. Five of your colleagues are passing by your desk; they all agree that you should calculate the effect of the treatment using only the 10,000 children who were treated.
   a. Is this advice sound? Why or why not?

3. Another colleague says that it’s not a good idea to drop the untreated entirely; you should use them but consider them as part of the comparison.
   a. Is this advice sound? Why or why not?

4. Another colleague suggests that you use the compliance rates, the proportion of people in each group that did or did not comply with their treatment assignment. You should divide the “intention to treat” estimate by the difference in the treatment ratios (i.e. proportions of each experimental group that received the treatment).
   a. Is this advice sound? Why or why not?
Managing spillovers—when the comparison, itself untreated, benefits from the treatment group being treated

People assigned to the control group may benefit indirectly from those receiving treatment. For example, a program that distributes insecticide-treated nets may reduce malaria transmission in the community, indirectly benefiting those who themselves do not sleep under a net. Such effects are called externalities or spillovers.

**Discussion Topic 4: Managing spillovers**

In the deworming program, randomization was at the school level. However, while all boys at a given treatment school were treated, only girls younger than thirteen received the deworming pill. This was due to the fact that the World Health Organization (WHO) had not tested (and thus not yet approved) the deworming pill for pregnant women. Because it was difficult to determine which girls were at risk of getting pregnant, the program decided to not administer the medication to any girl thirteen or older. (Postscript: since the deworming evaluation was implemented, the WHO has approved the deworming medication for pregnant women).

Thus at a given treatment school, there was a distinct group of students that was never treated but lived in very close proximity to a group that was treated.

Suppose protocol compliance is 100 percent: all boys and girls under thirteen in treatment schools get treated and all girls thirteen and over in treatment schools as well as all children in comparison schools do not get treated.

You can assume that due to proper randomization, the distribution of worm load across the three groups of students is equivalent between treatment and control schools prior to the intervention.

<table>
<thead>
<tr>
<th>Worm Load</th>
<th>Treatment</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All boys</td>
<td>Girls &lt;13 yrs</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>10000</td>
<td>5000</td>
</tr>
<tr>
<td>Total children tested at school</td>
<td>20000</td>
<td>20000</td>
</tr>
</tbody>
</table>

1. **a.** If there are any spillovers, where would you expect them to show up?  
**b.** Is it possible for you to capture these potential spillover effects? How?

2. **a.** What is the treatment effect for boys in treatment v. comparison schools?  
**b.** What is the treatment effect for girls under thirteen in treatment v. comparison schools?  
**c.** What is the direct treatment effect among those who were treated?  
**d.** What is the treatment effect for girls thirteen and older in treatment v. comparison schools?  
**e.** What is the indirect treatment effect due to spillovers?  
**f.** What is the total program effect?
References:
Exercises

Exercise A: Understanding random sampling / the law of large numbers

In this exercise, we will visually explore random samples of different sizes from a given population. In particular, we will try to demonstrate that larger sample sizes tend to be more reflective of the underlying population.

Your Group Leader has the data for this exercise.

1) Open the file “ExerciseA_SamplingDistributions_NEW.xlsm”.
2) If prompted, select “Enable Macros”.
3) Navigate to the “Randomize” worksheet, which allows you to choose a random sample of size “Sample Size” from the data contained in the “control” worksheet.
4) Enter “10” for “Sample Size” and click the “Randomize” button. Observe the distribution of the various characteristics between Treatment, Control and Expected. With a sample size this small, the percentage difference from the expected average is quite high for reading scores. Click “Randomize” multiple times and observe how the distribution changes.
5) Now, try “50” for the sample size. What happens to the distributions? Randomize a few times and observe the percentage difference for the reading scores.
6) Increase the sample size to “500”, “2000” and “10000”, and repeat the observations from step 5. What can we say about larger sample sizes? How do they affect our Treatment and Control samples? Should the percentage difference between Treatment, Control and Expected always go down as we increase sample size?
Exercise B: Sample size calculations

Key Vocabulary:

1. **Power**: the likelihood that, when the program has an effect, one will be able to distinguish the effect from zero given the sample size.
2. **Significance**: the likelihood that the measured effect did not occur by chance. Statistical tests are performed to determine whether one group (e.g. the experimental group) is different from another group (e.g. comparison group) on the measurable outcome variables used in the evaluation.
3. **Standard Deviation**: a standardized measure of the variation of a sample population from its mean on a given characteristic/outcome. Mathematically, the square root of the variance.
4. **Standardized Effect Size**: a standardized measure of the [expected] magnitude of the effect of a program.
5. **Cluster**: the level of observation at which a sample size is measured. Generally, observations which are highly correlated with each other should be clustered and the sample size should be measured at this clustered level.
6. **Intra-cluster Correlation Coefficient**: a measure of the correlation between observations within a cluster; i.e. the level of correlation in drinking water source for individuals in a household.

The Extra Teacher Program (ETP) case study discussed the concept of cluster randomized trials. The Balsakhi example used in the prior lecture introduced the concept of power calculations. In the latter, we were interested in measuring the effect of a treatment (balsakhis in classrooms) on outcomes measured at the individual level—child test scores. However, the randomization of balsakhis was done at the classroom level. It could be that our outcome of interest is correlated for students in the same classroom, for reasons that have nothing to do with the balsakhi. For example, all the students in a classroom will be affected by their original teacher, by whether their classroom is unusually dark, or if they have a chalkboard; these factors mean that when one student in the class does particularly well for this reason, all the students in that classroom probably also do better—which might have nothing to do with a balsakhi.

Therefore, if we sample 100 kids from 10 randomly selected schools, that sample is less representative of the population of schools in the city than if we selected 100 random kids from the whole population of schools, and therefore absorbs less variance. In effect, we have a smaller sample size than we think. This will lead to more noise in our sample, and hence a larger standard error than in the usual case of independent sampling. When planning both the sample size and the best way to sample classrooms, we need to take this into account.

This exercise will help you understand how to do that. Should you sample every student in just a few schools? Should you sample a few students from many schools? How do you decide?

We will work through these questions by determining the sample size that allows us to detect a specific effect with at least 80% power. Remember that power is the likelihood that when the treatment has an effect you will be able to distinguish it from zero in your sample.

In this example, “clusters” refer to “clusters of children”—in other words, “classrooms” or “schools”. This exercise shows you how the power of your sample changes with the number of clusters, the size of the clusters, the size of the treatment effect and the Intraclass Correlation Coefficient. We will use a software program developed by Steve Raudebush with funding from the William T. Grant Foundation. You can find additional resources on clustered designs on their web site.
Section 1: Using the OD Software

First download the OD software from the website (a software manual is also available):

http://sitemaker.umich.edu/group-based/optimal_design_software

When you open it, you will see a screen which looks like the one below. Select the menu option “Design” to see the primary menu. Select the option “Cluster Randomized Trials with person-level outcomes,” “Cluster Randomized Trials,” and then “Treatment at level 2.” You’ll see several options to generate graphs; choose “Power vs. Total number of clusters (J).”

A new window will appear:

Select $\alpha$ (alpha). You’ll see it is already set to 0.050 for a 95% significance level.

First let’s assume we want to test only 40 students per school. How many schools do you need to go to in order to have a statistically significant answer?

Click on $n$, which represents the number of students per school. Since we are testing only 40 students per school, so fill in $n(1)$ with 40 and click OK.

Now we have to determine $\delta$ (delta), the standard effect size (the effect size divided by the standard deviation of the variable of interest). Assume we are interested in detecting whether there is an increase of 10% in test scores. (Or more accurately, are uninterested in a decrease of less than 10%.) Our baseline survey indicated that the average test score is 26, with a standard deviation of 20. We want to detect an effect size of 10% of 26, which is 2.6. We divide 2.6 by the standard deviation to get $\delta$ equal to 2.6/20, or 0.13.

Select $\delta$ from the menu. In the dialogue box that appears there is a prefilled value of 0.20 for delta(1). Change the value to 0.13, and change the value of delta (2) to empty. Select OK.
Finally we need to choose $\rho$ (rho), which is the intra-cluster correlation. $\rho$ tells us how strongly the outcomes are correlated for units within the same cluster. If students from the same school were clones (no variation) and all scored the same on the test, then $\rho$ would equal 1. If, on the other hand, students from the same schools are in fact independent—and there were no differences between schools - then $\rho$ would equal 0.

You have determined in your pilot study that $\rho$ is 0.17. Fill in rho(1) to 0.17, and set rho(2) to be empty.

You should see a graph similar to the one below.

You’ll notice that your x axis isn’t long enough to allow you to see what number of clusters would give you 80% power. Click on the button to set your x axis maximum to 400. Then you can click on the graph with your mouse to see the exact power and number of clusters for a particular point.
Exercise 3.1:
How many schools are needed to achieve 80% power? 90% power?

Now you have seen how many clusters you need for 80% power, sampling 40 students per school. Suppose instead that you only have the ability to go to 124 schools (this is the actual number that was sampled in the Balsakhi program).

Exercise 3.2:
How many children per school are needed to achieve 80% power? 90% power? Choose different values for n to see how your graph changes.

Finally, let’s see how the Intraclass Correlation Coefficient (ρ) changes power of a given sample. Leave rho(1) to be 0.17 but for comparison change rho(2) to 0.0.

You should see a graph like the one below. The solid blue curve is the one with the parameters you’ve set - based on your pretesting estimates of the effect of reservations for women on drinking water. The blue dashed curve is there for comparison – to see how much power you would get from your sample if ρ were zero. Look carefully at the graph.

Exercise 3.3:
How does the power of the sample change with the Intraclass Correlation Coefficient (ρ)?
Exercise 3.4: 
Try generating graphs for how power changes with cluster size (n), intra-class correlation (rho) and effect size (delta). 
You will have to re-enter your pre-test parameters each time you open a new graph.
Exercise C: The mechanics of random assignment using MS Excel

Part 1: Simple Randomization

Like most spreadsheet programs, MS Excel has a random number generator function. Say we had a list of schools and wanted to assign half to treatment and half to control.

(1) We have our list of all schools.
(2) Assign a random number to each school:

The function RAND() is Excel’s random number generator. To use it, in Column C, type in the following = RAND() in each cell adjacent to every name. Or you can type this function in the top row (row 2) and simply copy and paste to the entire column, or click and drag.

Typing = RAND() puts a 15-digit random number between 0 and 1 in the cell.

(3) Copy the cells in Column C, then paste the values over the same cells
The function, =RAND() will re-randomize each time you make any changes to any other part of the spreadsheet. Excel does this because it recalculates all values with any change to any cell. (You can also induce recalculation, and hence re-randomization, by pressing the key F9.)

This can be confusing, however. Once we’ve generated our column of random numbers, we do not need to re-randomize. We already have a clean column of random values. To stop excel from recalculating, you can replace the “functions” in this column with the “values”.

To do this, highlight all values in Column C. Then right-click anywhere in the highlighted column, and choose Copy.

Then right click anywhere in that column and chose Paste Special. The “Paste Special window will appear. Click on “Values”.
(4) Sort the columns in either descending or ascending order of column C:

Highlight columns A, B, and C. In the data tab, and press the Sort button:

A Sort box will pop up.

In the Sort by column, select “random #”. Click OK. Doing this sorts the list by the random number in ascending or descending order, whichever you chose.
There! You have a randomly sorted list.

(5) Sort the columns in either descending or ascending order of column C:
Because your list is randomly sorted, it is completely random whether schools are in the top half of the list, or the bottom half. Therefore, if you assign the top half to the treatment group and the bottom half to the control group, your schools have been “randomly assigned”.

In column D, type “T” for the first half of the rows (rows 2-61). For the second half of the rows (rows 62-123), type “C”
Re-sort your list back in order of school id. You’ll see that your schools have been randomly assigned to treatment and control groups.

<table>
<thead>
<tr>
<th>SchoolID</th>
<th>SchoolName</th>
<th>Random#</th>
<th>T.C</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Babajpur G M M Kumar Shala No. 1</td>
<td>0.80541713</td>
<td>C</td>
</tr>
<tr>
<td>107</td>
<td>Babajpur Kanya Shala No. 2</td>
<td>0.65079324</td>
<td>C</td>
</tr>
<tr>
<td>107</td>
<td>Babajpur Kanya Shala No. 7</td>
<td>0.92443604</td>
<td>C</td>
</tr>
<tr>
<td>108</td>
<td>Babajpur Mishra Shala No. 8</td>
<td>0.81342516</td>
<td>C</td>
</tr>
<tr>
<td>112</td>
<td>Babajpur Marathi Mishra Shala No.12</td>
<td>0.68559637</td>
<td>C</td>
</tr>
<tr>
<td>113</td>
<td>Babajpur Kanya Shala No. 13</td>
<td>0.68559637</td>
<td>C</td>
</tr>
<tr>
<td>114</td>
<td>Babajpur Mishra Shala No. 14</td>
<td>0.94501796</td>
<td>C</td>
</tr>
<tr>
<td>117</td>
<td>Babajpur Kanya Shala No. 17</td>
<td>0.45205278</td>
<td>T</td>
</tr>
<tr>
<td>118</td>
<td>Babajpur Mishra Shala No. 18</td>
<td>0.18534539</td>
<td>C</td>
</tr>
<tr>
<td>119</td>
<td>Babajpur Kanya Shala No. 15</td>
<td>0.60772095</td>
<td>C</td>
</tr>
<tr>
<td>120</td>
<td>Babajpur Mishra Shala No. 20</td>
<td>0.83092642</td>
<td>C</td>
</tr>
<tr>
<td>121</td>
<td>Babajpur Mishra Shala No. 21</td>
<td>0.85013499</td>
<td>C</td>
</tr>
<tr>
<td>125</td>
<td>Babajpur Kanya Shala No. 25</td>
<td>0.30572157</td>
<td>T</td>
</tr>
<tr>
<td>126</td>
<td>Babajpur Kanya Shala No. 26</td>
<td>0.53308033</td>
<td>C</td>
</tr>
<tr>
<td>127</td>
<td>Babajpur Mishra Shala No. 27</td>
<td>0.45035747</td>
<td>T</td>
</tr>
<tr>
<td>128</td>
<td>Babajpur Mishra Shala No. 28</td>
<td>0.27464689</td>
<td>C</td>
</tr>
<tr>
<td>130</td>
<td>Babajpur Hindi Mishra Shala No. 30</td>
<td>0.62037959</td>
<td>T</td>
</tr>
<tr>
<td>131</td>
<td>Babajpur Mishra Shala No. 31</td>
<td>0.77709494</td>
<td>C</td>
</tr>
<tr>
<td>132</td>
<td>Babajpur Kanya Shala No. 32</td>
<td>0.23251221</td>
<td>T</td>
</tr>
<tr>
<td>201</td>
<td>Fatehpura Kanya Shala No. 1</td>
<td>0.91652716</td>
<td>C</td>
</tr>
<tr>
<td>202</td>
<td>Fatehpura Kanya Shala No. 2</td>
<td>0.66693643</td>
<td>C</td>
</tr>
<tr>
<td>205</td>
<td>Fatehpura Kanya Shala No. 8</td>
<td>0.49008271</td>
<td>T</td>
</tr>
<tr>
<td>209</td>
<td>Fatehpura Kanya Shala No. 9</td>
<td>0.62054343</td>
<td>C</td>
</tr>
<tr>
<td>210</td>
<td>Fatehpura Kanya Shala No. 10</td>
<td>0.17075854</td>
<td>T</td>
</tr>
<tr>
<td>211</td>
<td>Fatehpura Kanya Shala No. 11</td>
<td>0.17075854</td>
<td>T</td>
</tr>
<tr>
<td>212</td>
<td>Fatehpura Kanya Shala No. 12</td>
<td>0.36694815</td>
<td>T</td>
</tr>
<tr>
<td>216</td>
<td>Fatehpura Kanya Shala No. 16</td>
<td>0.56569267</td>
<td>T</td>
</tr>
<tr>
<td>219</td>
<td>Fatehpura Kanya Shala No. 19</td>
<td>0.45511079</td>
<td>T</td>
</tr>
<tr>
<td>220</td>
<td>Fatehpura Kanya Shala No. 10</td>
<td>0.77059484</td>
<td>C</td>
</tr>
<tr>
<td>301</td>
<td>N. Sayajigunj Mishra Shala No. 1 (center)</td>
<td>0.93891849</td>
<td>C</td>
</tr>
</tbody>
</table>
Part 2: Stratified Randomization

Stratification is the process of dividing a sample into groups, and then randomly assigning individuals within each group to the treatment and control. The reasons for doing this are rather technical. One reason for stratifying is that it ensures subgroups are balanced, making it easier to perform certain subgroup analyses. For example, if you want to test the effectiveness of a new education program separately for schools where children are taught in Hindi versus schools where children are taught in Gujarati, you can stratify by “language of instruction” and ensure that there are an equal number schools of each language type in the treatment and control groups.

(1) We have our list of schools and potential “strata”.

Mechanically, the only difference in random sorting is that instead of simply sorting by the random number, you would first sort by language, and then the random number. Obviously, the first step is to ensure you have the variables by which you hope to stratify.

(2) Sort by strata and then by random number.

Assuming you have all the variables you need: in the data tab, click “Sort”. The Sort window will pop up. Sort by “Language”. Press the button, “Add Level”. Then select, “Random #”. 

![Image of sorting process](image.png)
**Assign Treatment – Control Status for each group.**

Within each group of languages, type “T” for the first half of the rows, and “C” for the second half.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>132</td>
<td>Babajpora Mishra Shala No. 32</td>
<td>Gujarati</td>
<td>Mishra</td>
<td>0.8931976</td>
<td>C</td>
</tr>
<tr>
<td>101</td>
<td>615</td>
<td>Wadi Mishra Shala No. 15</td>
<td>Gujarati</td>
<td>Mishra</td>
<td>0.9142383</td>
<td>C</td>
</tr>
<tr>
<td>102</td>
<td>618</td>
<td>Wadi Kumar Shala No. 18</td>
<td>Gujarati</td>
<td>Kumar</td>
<td>0.9229356</td>
<td>C</td>
</tr>
<tr>
<td>103</td>
<td>408</td>
<td>Raopura Kanya Shala No. 8</td>
<td>Gujarati</td>
<td>Kanya</td>
<td>0.9286077</td>
<td>C</td>
</tr>
<tr>
<td>104</td>
<td>502</td>
<td>Sheher Vibhag Mishra Shala No. 2</td>
<td>Gujarati</td>
<td>Mishra</td>
<td>0.9649163</td>
<td>C</td>
</tr>
<tr>
<td>105</td>
<td>311</td>
<td>Sayajiganj Mishra Shala No. 11</td>
<td>Gujarati</td>
<td>Mishra</td>
<td>0.9595266</td>
<td>C</td>
</tr>
<tr>
<td>106</td>
<td>344</td>
<td>Sayajiganj Mishra Shala No. 44</td>
<td>Gujarati</td>
<td>Mishra</td>
<td>0.9688884</td>
<td>C</td>
</tr>
<tr>
<td>107</td>
<td>347</td>
<td>Sayajiganj Hindi Mishra Shala No. 47</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.0163449</td>
<td>T</td>
</tr>
<tr>
<td>108</td>
<td>332</td>
<td>Sayajiganj Hindi Mishra Shala No. 32</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.1528766</td>
<td>T</td>
</tr>
<tr>
<td>109</td>
<td>342</td>
<td>Sayajiganj Hindi Mishra Shala No. 42</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.2646791</td>
<td>T</td>
</tr>
<tr>
<td>110</td>
<td>215</td>
<td>Fatehpura Hindi Mishra Shala No. 15</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.3142377</td>
<td>C</td>
</tr>
<tr>
<td>111</td>
<td>325</td>
<td>Sayajiganj Hindi Mishra Shala No. 25</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.4291589</td>
<td>T</td>
</tr>
<tr>
<td>112</td>
<td>638</td>
<td>Wadi Hindi Mishra Shala No. 38</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.6772441</td>
<td>C</td>
</tr>
<tr>
<td>113</td>
<td>130</td>
<td>Babajpora Hindi Mishra Shala No. 30</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.7053783</td>
<td>C</td>
</tr>
<tr>
<td>114</td>
<td>315</td>
<td>Sayajiganj Hindi Mishra Shala No. 15</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.7955643</td>
<td>C</td>
</tr>
<tr>
<td>115</td>
<td>626</td>
<td>Wadi Hindi Mishra Shala No. 26</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.8910818</td>
<td>C</td>
</tr>
<tr>
<td>116</td>
<td>345</td>
<td>Sayajiganj Hindi Mishra Shala No. 46</td>
<td>Hindi</td>
<td>Mishra</td>
<td>0.9051467</td>
<td>C</td>
</tr>
<tr>
<td>117</td>
<td>303</td>
<td>N. Sayajiganj Marathi Mishra Shala No. 3</td>
<td>Marathi</td>
<td>Mishra</td>
<td>0.0364843</td>
<td>T</td>
</tr>
<tr>
<td>118</td>
<td>523</td>
<td>Sheher Vibhag Marathi Mishra Shala No. 23</td>
<td>Marathi</td>
<td>Mishra</td>
<td>0.1834626</td>
<td>T</td>
</tr>
<tr>
<td>119</td>
<td>409</td>
<td>Raopura Marathi Mishra Shala No. 9</td>
<td>Marathi</td>
<td>Mishra</td>
<td>0.7676874</td>
<td>T</td>
</tr>
<tr>
<td>120</td>
<td>611</td>
<td>Wadi Marathi Shala No. 11</td>
<td>Marathi</td>
<td>Mishra</td>
<td>0.8847497</td>
<td>T</td>
</tr>
<tr>
<td>121</td>
<td>329</td>
<td>Sayajiganj Marathi Mishra Shala No. 29</td>
<td>Marathi</td>
<td>Mishra</td>
<td>0.8992805</td>
<td>C</td>
</tr>
<tr>
<td>122</td>
<td>112</td>
<td>Babajpora Marathi Mishra Shala No. 12</td>
<td>Marathi</td>
<td>Mishra</td>
<td>0.9430188</td>
<td>C</td>
</tr>
<tr>
<td>123</td>
<td>327</td>
<td>Sayajiganj Marathi Mishra Shala No. 27</td>
<td>Marathi</td>
<td>Mishra</td>
<td>0.9516261</td>
<td>C</td>
</tr>
<tr>
<td>124</td>
<td>617</td>
<td>Wadi Marathi Mishra Shala No. 17</td>
<td>Marathi</td>
<td>Mishra</td>
<td>0.9648496</td>
<td>C</td>
</tr>
</tbody>
</table>
Statistics Review
Preparation for Power Calculation Lecture: Sample Means and the normal distribution

For statistics, we want to work with...

- A Normal Distribution...
- (a “Bell Curve”)

How do we get from here...
The normal distribution is

- A sampling distribution
- A distribution of “means” if we sampled the population an infinite number of times

Take 5 and 10 draws of 10...
With 50 and 100 samples...

Frequency of Means With 50 Samples

Frequency of Means with 100 Samples

With 500 and 1000 samples

Frequency of Means With 500 Samples

Frequency of Means With 1000 Samples
Size of Sample

- In our last example we drew samples of 10
  - (Sample size was 10)
- We observed the mean and plotted it on our graph
- After we repeated that 1000 times and looked at the “distribution of means,” we saw it approached a bell curve
- What happens if we increase our sample size?
50 and 100 samples of 10 and 50

With 500 and 1000 samples
Basic lessons:

- If we take enough samples from our data, the distribution of their means will eventually approach a normal distribution (Central Limit Theorem)
- More individuals per sample => quicker convergence to normal and mean
Group Presentation

Participants will form 4-6 person groups which will work through the design process for a randomised evaluation of a development project. Groups will be aided in this project by both the faculty and teaching assistants, with the work culminating in presentations at the end of the week.

The goal of the group presentation is to consolidate and apply the knowledge of the lectures and thereby ensure that participants leave with the knowledge, experience, and confidence necessary to conduct their own randomised evaluations. We encourage groups to work on projects that are relevant to participants’ organisations.

All groups will present on Friday. The 15-minute presentation is followed by a 15-minute discussion led by J-PAL affiliates and staff. We provide groups with template slides for their presentation (see next page). While the groups do not need to follow this exactly, the presentation should have no more than 9 slides (including title slide, excluding appendix) and should include the following topics:

- Brief project background
- Theory of change
- Evaluation question
- Outcomes
- Evaluation design
- Data and sample size
- Potential validity threats and how to manage them
- Dissemination strategy of results

Please time yourself and do not exceed the allotted time. We have only a limited amount of time for these presentations and follow a strict timeline to be fair to all groups.
Group Presentation Template

Background

- Talk briefly about general context, needs assessment, problem you want to solve.

Theory of Change

- Describe the specific intervention that you are evaluating.
- Talk about how it will solve part of the problem you described in the background.
- You may want to mention other causes of a problem that your intervention will not solve.
- (You can use the TOC template in the appendix.)

Evaluation Questions and Outcomes

- These should be directly linked to the TOC described above.
- What outcomes do you need to measure to test your research hypothesis?

Evaluation Design

- Unit of randomization, type of randomization (why did you choose these?)
- The actual randomization design- i.e. specific treatment group(s)

Data and Sample Size

- Outcomes
- Tell us where you will get the data – survey? Administrative?
- Power calc:
  - Justify where you got effect size and rho from, don’t make it up.
  - You may need to do separate power calc's for separate outcomes.
Potential challenges

- Talk about threats (attrition, spillover, etc.) and how you want to manage them.
- You may need to revise your power calcs.

Results

- Why (and for whom) they would be useful.
- How would you disseminate them.

Appendix

Theory of change

Need Assessment

Intervention

Final Outcome
Practical Tips

Transport from Airport

Participants are in charge of organizing their own transport from the airport. There are metered taxis available. Standard rates are between 10 and 12 Rand per km. A trip to UCT / Rondebosch should not cost more than R200. Mention that you need a receipt before entering a cab.
One of cheapest cab services is Excite Cab: 021 418 4444 (9 Rand per km)

Shopping

On campus: There are cafeterias in Middle Campus selling a range of food products, from sandwiches to vegetarian Indian food.

Off campus: Southern Suburbs (if your hotel is in Rondebosch or Claremont)
Cavendish Square is located just off Claremont Main Road and is bound to satisfy all your shopping desires. If one is looking to do a little grocery shopping, the Woolworths food is located on the bottom floor. Alternatively, if you wish to avoid the rush of the mall, you can find a Pick n’ Pay supermarket a little further down Main Road towards Rondebosch. A smaller Pick n’ Pay and Woolworths food are also situated closer to the university on Rondebosch Main Road and are a 15 minute walk from UCT.

<table>
<thead>
<tr>
<th>Cavendish Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open times: 09:00-19:00 (Mon-Sat), 10:00-17:00 (Sundays)</td>
</tr>
<tr>
<td>Address: Dreyer Street, Claremont</td>
</tr>
<tr>
<td>Website: <a href="http://www.cavendish.co.za">www.cavendish.co.za</a></td>
</tr>
<tr>
<td>Call: 021 657 5620</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pick n’ Pay (Claremont)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open times: 08:00 – 19:00 (Mon-Sat), 08:00- 17:00 (Sun)</td>
</tr>
<tr>
<td>Address: Corner Campground &amp; Main Road, Claremont</td>
</tr>
<tr>
<td>Call: 021 674 5908</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pick n’ Pay (Rondebosch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open times: 08:00-21:00 (Mon-Sun)</td>
</tr>
<tr>
<td>Address: Shop No 1, Village Centre, Main Road, Rondebosch</td>
</tr>
<tr>
<td>Call: 021 685 4001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Woolworths (Rondebosch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open times: 09:00-21:00 (Mon-Fri), 09:00-20:00 (Sat-Sun)</td>
</tr>
<tr>
<td>Call: 021 685 4416</td>
</tr>
</tbody>
</table>

Other large malls in Cape Town:
The Waterfront (for when you are closer to town or visiting Robben Island)
Website: [www.waterfront.co.za](http://www.waterfront.co.za)
Alternative places to buy groceries and snacks:
Many of the petrol stations around Cape Town have little food stores.

Restaurants
Cape Town is known for its diverse array of dining and cuisine. Here is but a small list of well-known restaurants that you may wish to try.

**Budget**
*(Main meal under R50)*

1) **Eastern Food Bazaar**
   Cuisine: Indian, Chinese
   Location: City Bowl
   Contact: 021 461 2458

2) **Mzolis**
   Cuisine: African, BBQ
   Location: Gugulethu
   Contact: 021 638 1355
   **You will need a guide or someone from Cape Town who knows the area!**

3) **Food Lovers Market**
   Cuisine: Deli, Buffet – Basically everything
   Location: Claremont
   Contact: 021 674 7836

**Medium price range**
*(Main meal between R50 and R100)*

1) **Col Cacchio Pizzeria**
   Cuisine: Pizza
   Location: Claremont (Cavendish), Camps Bay
   Contact: 021 674 6387/ 021 438 2171

2) **Kirstenbosch Tea Room**
   Cuisine: Coffee Shop
   Location: Kirstenbosch Gardens, Newlands (Not for dinner)
   Contact: 021 797 4083

3) **Rhodes Memorial Restaurant**
   Cuisine: Bistro, Coffee Shop
   Location: Rhodes Memorial Restaurant (Not for dinner)
   Contact: 021 687 0000

4) **Fadela Williams**
   Cuisine: Cape Malay
   Location: Claremont
   Contact: 021 671 0037

5) **Hussar Grill**
   Cuisine: Grills
   Location: Rondebosch
   Contact: 021 689 9516

6) **Addis in Cape**
   Cuisine: Ethiopian
   Location: City Bowl
   Contact: 021 424 5722
Higher End
(Main meal - R100 and above)

1) *Die Wijnhuis
   Cuisine: Mediterranean, Italian
   Location: Newlands
   Contact: 021 671 9705

2) *Barristers Grill
   Cuisine: Grill and Seafood
   Location: Newlands
   Contact: 021 671 7907

3) Panama Jack's Taverna
   Cuisine: Seafood
   Location: Table Bay harbour
   Lunch rates are lower. For example they offer a half-kilo of prawns for only R60 during the week
   Contact: 021 448 1080

4) Olympia Cafe
   Cuisine: Deli, Bakery, Coffee Shop
   Location: Kalk Bay
   Contact: 021 788 6396

5) *Bihari
   Cuisine: Indian
   Location: Newlands
   Contact: 021 674 7186

6) Jonkershuis Constantia Eatery
   Cuisine: Bistro
   Location: Constantia
   Contact: 021 794 4813

7) Moyo
   Cuisine: African
   Location: Stellenbosch
   Contact: 021 809 1133
Internet Access

Most hotels will have access otherwise ask for directions to your nearest internet café.

Electricity

Voltage: 220/230 V

Adapter: You will need an adaptor for Plug M and sometimes plug C. Plug C is the two-pin plug commonly used in Europe.

Money

Withdrawing: We suggest that you use the campus ATM machines. They are situated on Middle Campus (next to the cafeteria), and Upper Campus (ground floor of the Leslie Social Science building and next to the library).

Credit Cards: When paying by credit card, we suggest that you ask vendors to swipe the card in your presence.

Exchange Rate: The current exchange rate is 8.15 South Africa Rand to the US-Dollar.
1. Vineyard Hotel
2. Southern Sun
3. Graca Machel
4. UCT Economics Building
5. Cavendish Square (Woolworths food inside)
6. Pick n’ Pay (Claremont)
7. Pick n’ Pay (Rondebosch)
8. Woolworths (Rondebosch)
9. Food Lovers Market (See No. 5)
10. Col Cacchio Pizzeria (See No. 5)
11. Rhodes Memorial Restaurant
12. Fadela Williams
13. Hussar Grill
14. Die Wijnhuis
15. Barristers Grill
16. Bihari
Health and Emergencies

On campus:
1) Campus Protection Services: 021 650 2222/3
2) UCT Emergency Controller: 021 650 2175/6

Off Campus
1) Kingsbury Hospital (Wilderness Road, Claremont): 021 670 4000
2) Constantiaberg Medi-Clinic Hospital (Burnham Road, Plumsted): 021 799 2911 / 021 799 2196 (Emergency number)
3) Kenilworth Medicross (67 Rosmead Avenue, Kenilworth): 021 670 7640 – for doctor’s visits

State Emergency Number (Police and Ambulance Services): 10111
Private Ambulance Services: Netcare911: 082 911

J-PAL Africa Staff Contact Details

Kamilla Gumede (Executive Director): 082 312 3635
Martin Abel (Research Manager): 073 325 4438
Raissa Fabregas (Research Manager): 0727291483
Willa Brown (Research Analyst): 071 747 5635
Megan Blair (Research Analyst): 079 516 8360
Rebecca Metz (Research Analyst): 073 236 3147
Clare Hofmeyr (Policy Analyst): 072 246 8747
Letitia Sullivan (Admin Officer): 021 650 5981