DECISION-FOCUSED EVALUATIONS:
A PRACTICAL GUIDE

Harlan Downs-Tepper • Global Innovation Week • Sept 28, 2017
1. About IDinsight
2. Why Reliable Data Matters
3. Choosing the Right Tool
4. Parameters of High-Quality Evaluation
5. Zambia Ministry of Health Case
6. USAID d.light Case
7. Looking Forward: Key Lessons
IDinsight helps clients generate and use rigorous evidence to improve social impact.

Photo: Students learning in Andhra Pradesh school.
Our clients

National governments

Social enterprises

Multilateral

Foundations

The Zambian Ministry of Health

Government of India

d.light

STiR education

unicef

The World Bank

Bill & Melinda Gates Foundation

THE WILLIAM AND FLORA HEWLETT FOUNDATION
Key lessons you should walk away with from this session

1. How to think about selecting a **counterfactual** in your context.

2. Know how to **decide on measurement and evaluation tools** for a given question.

3. **Practical ways** to make rigorous evaluations decision-focused, faster, cheaper and less operationally cumbersome.
**Decision-focused evaluation characteristics**

- **Context-specific**
  - Rigorous techniques, including field experiments, at the particular time and location of interest

- **Demand-driven**
  - Research informs a specific decision under consideration

- **Timely**
  - Actionable information delivered in time for decision-making deadlines

- **Cost-effective**
  - Every dollar spent is justified by expected impact
1. About IDinsight

2. Why Reliable Data Matters

3. Choosing the Right Tool

4. Parameters of High-Quality Evaluation

5. Zambia Ministry of Health Case

6. USAID d.light Case

7. Looking Forward: Key Lessons
Should you do an evaluation?

1. Do you have a question about the services you provide?
   - No
   - Yes

2. Will answering your question help you make a decision?
   - No
   - Yes

3. Has anyone already answered your question?
   - Yes
   - No

4. Can you do a “good” evaluation within your organization’s constraints?
   - No
   - Yes

Don’t Evaluate! Evaluate!
Bad evaluation can lead to poor decision-making

% of children immunized, Immunization Project, Rajasthan, India

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad evaluation can lead to poor decision-making</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Rest of study region
- Program villages in study region
Bad evaluation can lead to poor decision-making

% of children immunized, Immunization Project, Rajasthan, India

Program villages in study region

Rest of study region

<table>
<thead>
<tr>
<th>Year</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>4%</td>
<td>4%</td>
<td>14%</td>
<td>40%</td>
<td>40%</td>
<td>40%</td>
</tr>
</tbody>
</table>
What makes data worth listening to?

The **counterfactual**: what would have happened in the absence of your program?

The **randomized evaluation** is the gold standard of measuring the counterfactual, and hence the true impact of a program.
1. About IDinsight
2. Why Reliable Data Matters
3. Choosing the Right Tool
4. Parameters of High-Quality Evaluation
5. Zambia Ministry of Health Case
6. USAID d.light Case
7. Looking Forward: Key Lessons
First, map out your intervention.

- A **theory of change** traces your program’s impact pathways.
- Think hard about what you know and what you don’t.
- Connect *activities* to *outputs* and *outcomes*.

**Example TOC**

<table>
<thead>
<tr>
<th>Input</th>
<th>Activity</th>
<th>Output 1</th>
<th>Output 2</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purchase ITN</td>
<td>Distribute ITN at community point</td>
<td>Recipients hang ITNs</td>
<td>Recipients sleep under bednets</td>
<td>Decrease in malaria morbity and mortality</td>
</tr>
</tbody>
</table>
Then, determine which tool to use for your program’s current stage.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Early stage</th>
<th>Learning/Pilot stage</th>
<th>Scale-up stage or at-scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tool options</td>
<td>Design</td>
<td>Process Improvement</td>
<td>Impact evaluation</td>
</tr>
<tr>
<td></td>
<td>Identify challenges and generate ideas</td>
<td>Make programs run efficiently</td>
<td>Experiment with different ways of running program</td>
</tr>
<tr>
<td></td>
<td>Monitoring Operations</td>
<td>Monitoring Operations</td>
<td>Make sure program continues to run and deliver at scale</td>
</tr>
<tr>
<td></td>
<td>At-scale Impact</td>
<td>At-scale Impact</td>
<td>Test whether program is having intended impact</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tool options</th>
<th>Monitoring data</th>
<th>Impact Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic literature review</td>
<td>Theory of Change review</td>
<td>Process Evaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cost-effectiveness Analysis</td>
</tr>
</tbody>
</table>
Agenda

1. About IDinsight
2. Why Reliable Data Matters
3. Choosing the Right Tool
4. Parameters of High-Quality Evaluation
5. Zambia Ministry of Health Case
6. USAID d.light Case
7. Looking Forward: Key Lessons
How rigorous does your evaluation need to be?

Rigor

- Accuracy
- Precision
- Generalizability

Cost

- Time
- Money
- Expertise
How precise does your evaluation need to be?

- Consider the following scenario:

- Suppose that you want to know the average income per capita in your city. The "true" income per capita is $50,000 and most people (95%) have income between $40,000 and $60,000. Suppose it costs $20 to survey one person about their income.

- How many people do you need to survey in order to get a precise measurement of income per capita?
How precise does your evaluation need to be?

Estimates of Average Income

Income (USD)

54,000
52,000
50,000
48,000
46,000

$200
$2,000
$20,000
$200,000
$2 million

Sample Size

10 100 1,000 10,000 100,000
How generalizable does your evaluation need to be?

- Consider the following scenario: Suppose that you need to select a sample for your evaluation. You can choose either:

A. 10 people per district in 100 districts
B. 100 people per district in 10 districts

1. Which sample is more likely to produce biased results?
   - Equally prone to bias
2. Which sample is more likely to produce generalizable results?
   - (A)
3. Which sample is likely to have higher data collection costs?
   - (A)
Agenda

1. About IDinsight
2. Why Reliable Data Matters
3. Choosing the Right Tool
4. Parameters of High-Quality Evaluation
5. Zambia Ministry of Health Case
6. USAID d.light Case
7. Looking Forward: Key Lessons
Motivation
How to decrease maternal and infant mortality in Zambia?
The Decision in Question
Should the Ministry of Health provide “mama kits” at clinics to reduce maternal and infant mortality?
How would you answer MoH’s question?

1. What method would you choose?

2. What outcome?

3. What timeline?
Randomizing intervention delivery allows us to isolate impact.

Note: stylized randomization for illustrative purposes only.
Randomizing intervention delivery allows us to isolate impact.

Note: stylized randomization for illustrative purposes only.
Randomization successfully achieved balance between treatment and control

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mama Kit health centers</th>
<th>Comparison health centers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pregnancy to delivery ratio</td>
<td>0.52</td>
<td>0.53</td>
</tr>
<tr>
<td>Number of previous pregnancies</td>
<td>3.87</td>
<td>3.94</td>
</tr>
<tr>
<td>Maternal age</td>
<td>25.13</td>
<td>25.42</td>
</tr>
<tr>
<td>Number of ANC visits</td>
<td>1.70</td>
<td>1.72</td>
</tr>
</tbody>
</table>
Constraints

1. Limited information on effectiveness of mama kits, and no information in local context.

2. Limited budget.

3. Need to reach a decision rapidly to inform next budgeting cycle.
Mama Kits theory of change

Mama Kit distributed at facilities → Mothers deliver in facilities → Decreased maternal mortality

**Outcome variable**

Link well-established in public health literature

Incentive link uncertain in this context
Evaluation Design

- **Method:** randomized evaluation
- **Sample size:** 2219 eligible women at 30 health centers
- **Outcome variable:** percentage of women delivering in a medical facility
- **Timeline:** 9 months
- **Cost:** $70,000
$4 kits increased institutional delivery by...
$4 kits increased institutional delivery by... 47%

Estimated cost-effectiveness: $5,183 per life saved
DECISION: Mama kits scaled nationally.
Mama Kits: Key takeaways

• A proximate outcome allowed for a shorter study.

• A policy-relevant effect size kept costs down.

• Data collection tailored to specific needs.

• An identification strategy that made sense for this context.
1. About IDinsight
2. Why Reliable Data Matters
3. Choosing the Right Tool
4. Parameters of High-Quality Evaluation
5. Zambia Ministry of Health Case
6. USAID d.light Case
7. Looking Forward: Key Lessons
d.light aims to improve energy access, education, health, and environmental outcomes by selling off-grid solar lanterns in developing countries, to displace hazardous kerosene lanterns.
The Decision in Question

Should USAID allocate over $10M in grant funding to help scale up d.light’s home solar system in eastern Uganda?
Criteria and constraints

• USAID’s scale-up funding criteria:
  o Organization must provide rigorous evidence that it causally improves social outcomes such as student learning, health, or livelihoods

• d.light constraints
  • Partner for scale-up
  • Geography
  • Business model
Designing an evaluation

Compare on:

• Rigor in assessing causal impact
• Cost
• Timeline
• Informing the decision in question

If you were USAID – would you fund it?
Sell d.light products using normal business operations. Provide products free of charge or heavily subsidized.

RCT options and requirements:

1. Randomized controlled trial
   - Would require a very large sample size

2. Freely provide products or heavily subsidized
   - Requires working with a different distribution partner
   - Different geography
   - Poorer households
   - Different business model
Statistical matching

• Allows working with the existing partner, using current business model, in intended geography

• Requires identifying comparison households in nearby villages or neighboring regions

• Requires larger (2x) sample size at baseline to identify enough similar comparison households (“matches”)

• Statistical matching
d.light customers  500
Comparison households  1,500

Comparison household 1
Household size: 5 people
Rooms in house: 4
Education of HH head: Primary
Occupation of HH head: Farmer
Distance to urban center: 30km
Mobile phone: Yes
Floor material: Packed earth

Comparison household 2
Household size: 6 people
Rooms in house: 3
Education of HH head: Primary
Occupation of HH head: Wage labor
Distance to urban center: 23km
Mobile phone: Yes
Floor material: Cement

Comparison household 3
Household size: 6 people
Rooms in house: 4
Education of HH head: None
Occupation of HH head: Wage labor
Distance to urban center: 18km
Mobile phone: No
Floor material: Packed earth

D.light customer household
Household size: 6 people
Rooms in house: 4
Education of HH head: Primary
Occupation of HH head: Wage labor
Distance to urban center: 20km
Mobile phone: Yes
Floor material: Cement
Two evaluation design options – which would you choose?

<table>
<thead>
<tr>
<th></th>
<th>RCT</th>
<th>Matching with diff-in-diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodological rigor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(internal validity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution partner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographical similarity to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scale up region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of interpreting/understanding results</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Two evaluation design options – which would you choose?

<table>
<thead>
<tr>
<th></th>
<th>RCT</th>
<th>Matching with diff-in-diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodological rigor (internal validity)</td>
<td>Better</td>
<td></td>
</tr>
<tr>
<td>Distribution partner</td>
<td></td>
<td>Better</td>
</tr>
<tr>
<td>Geographical similarity to scale up region</td>
<td></td>
<td>Better</td>
</tr>
<tr>
<td>Target population</td>
<td></td>
<td>Better</td>
</tr>
<tr>
<td>Cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of interpreting/understanding results</td>
<td></td>
<td>Better</td>
</tr>
</tbody>
</table>
Shortcomings of non-randomized experiments

• Unobserved characteristics may differ between d.light and non-d.light recipients that could affect use
• Matching design requires a larger sample size than randomized evaluation design
• Analysis can be more complicated than with randomized evaluation
• Matching design tends to be more expensive, all else being equal
• Can be less compelling (rightfully so) to many donors (and yourself!)
Summary: what characterizes a decision-focused evaluation?

- **Context-specific**
  - Rigorous techniques, including field experiments, at the particular time and location of interest

- **Demand-driven**
  - Research informs a specific decision under consideration

- **Timely**
  - Actionable information delivered in time for decision-making deadlines

- **Cost-effective**
  - Every dollar spent is justified by expected impact
1. About IDinsight
2. Why Reliable Data Matters
3. Choosing the Right Tool
4. Parameters of High-Quality Evaluation
5. Zambia Ministry of Health Case
6. USAID d.light Case
7. Looking Forward: Key Lessons
Practical tips to decrease cost and timeframe

1. Smaller, focused questionnaires
2. Proximate outcome variables
3. Skip baseline survey if not needed
4. Larger, policy-relevant effect sizes (smaller sample sizes)
5. Rolling sample sizes, early stopping rules
6. Existing data and data collection systems
7. Electronic surveys/mobile data collection
Spotting opportunities for IE

• Depending on the program and available data, different opportunities may arise for solving the fundamental problem of causal inference.

<table>
<thead>
<tr>
<th>Opportunity</th>
<th>Analysis Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>A pilot is planned before scale-up</td>
<td>• Randomize who gets to participate in the pilot</td>
</tr>
<tr>
<td></td>
<td>• Collect data on those who don’t get the pilot</td>
</tr>
<tr>
<td>A phased roll-out is planned</td>
<td>• Randomize the order of the rollout</td>
</tr>
<tr>
<td></td>
<td>• Collect data on those who don’t get the pilot initially</td>
</tr>
<tr>
<td></td>
<td>• Ensure that those who will receive the program later don’t know they will get it in future</td>
</tr>
<tr>
<td>Presence of natural experiment</td>
<td>• Check to make sure that the groups that have been randomly split are similar on important characteristics</td>
</tr>
</tbody>
</table>

• However, just because you can do an IE doesn’t mean that you should do an IE
Guiding principles of decision-oriented impact evaluation

1. What specific decision will the impact evaluation inform? What actions will you take if you find a negative effect? Zero effect? Positive effect?

2. Can you create a comparison group that is very similar (on observable and unobservables) to the group that receives your program?

3. Can you get results on the timeframe you need before you have to make a decision? Are there proximate outcomes that you can quickly measure that are closely linked to the ultimate outcome you care about?

4. Is the cost of the evaluation justified given the size/importance of the decision it will inform?

5. What financial and human resources do you need to act upon the findings of the impact evaluation?
Thank you!
Appendix
Is an evaluation worth the cost?

1. How many people (& how intensely) will the decision impact?

2. How much $ will the decision influence?

3. What is the quality and applicability of existing evidence and how much uncertainty is there?

4. Are there political/administrative constraints preventing you from making a decision based on the impact evaluation?

5. How much will the evaluation cost?
Is an evaluation worth the cost? A simplified SROI model

Estimate:

1. The expected outcomes after a decision without evaluation evidence
2. The expected outcomes after a decision with evaluation evidence
3. The additional cost incurred due to the experiment

$$SROI = (2-1)/3$$
Maintaining quality with a lean survey

• Start with a well-established tool in your context
  o This will contribute to construct validity, credibility with outside audiences, and the ability to compare your result with other evaluations.

• Focus on your most important outcome of interest

• Trim the survey to fit your needs: see which questions are already highly correlated or inappropriate in your context
  o E.g. do students max out an exam?