BUILDING EVIDENCE TO INFORM PRACTICE AND POLICY: DESIGN CHOICES AND CONSIDERATIONS FOR IMPACT STUDIES

NANDITA VERMA

USING PILOTS TO MEASURE IMPACTS
HOSTED BY EDUCATION REVIEW OFFICE
NEW ZEALAND

JULY 12-13, 2018
OVERVIEW

- Evidence-building in the US context – brief background
  - Where MDRC fits in
- Launching a pilot to assess impacts – choices and considerations
  - What to know in advance
  - Key questions, supporting design, sample, methods
  - Operational issues – getting it off the ground
  - Monitoring, assessment, tracking
  - Dissemination
- Beyond the pilot – scale-up, sustainability
Long journey! A few turning points in the tale

The Dept. of Health and Human Services (HHS) played a significant role crucial (strong champions) in this journey

Today, bipartisan consensus around the need to build evidence around effectiveness – billions spent on human services programs and policies

Demand at federal level for testing program effectiveness

- Some agencies adopting a tiered-evidence approach to grantmaking

Private foundations also play a critical role in shaping, supporting innovation and evidence-building.
MDRC – WHO WE ARE

- Mission-driven, nonpartisan, nonprofit, independent social policy organization
- Founded in 1974 by the Ford Foundation and some federal agencies
- Dedicated to learning what works to improve programs and policies for low-income families
  - Families with children (focus on early developmental outcomes)
  - K-12 education (helping students succeed in education)
  - Postsecondary education
  - Work, housing, income security (raising the prospects of low-wage workers, the unemployed, and communities where poverty is concentrated)
  -Disconnected youth (foster care, juvenile justice)
  - Criminal justice (supports for incarcerated adults)
- Develop and evaluate solutions to problems (demonstrations vs. evaluations).
MDRC – DISTINGUISHING FEATURES

- Commitment to evidence
  - Research rigor, comprehensive studies
- Focus on initiatives where evidence matters and potential to scale
- Engage with policymakers, build partnerships with the programs we study
- Proactive dissemination
  - To inform public policy and improve practice in field
- Combination of research, operational know how, and analytical skills.
**PURPOSE OF EVALUATIONS – RECAP**

- Provides continuous learning opportunities for programs and organizations
  - Build on strengths
  - Remedy shortcomings
- Identify gaps in evidence where further innovations and testing are needed
  - Program gaps
  - Target group gaps
- Facilitates evidence-based policy making
  - Invest in what works
  - OMB places a premium on evidence
  - Innovation funds are based on evidence

BEFORE YOU BEGIN AN IMPACT STUDY....

- Know what you want – outcomes are not the same as impacts
- Tailor evaluation design to match study goals and questions
- Think nested designs and multiple perspectives – evaluations addresses more than 1 question; different lenses provide for more nuanced analysis and interpretation
- Develop comprehensive measurement strategies – think big, build out data collection as resources permit (in RCT: informed consent defines data collection)
- Plan for multiple data points, (baseline, follow-up), data and resources permitting; full sample and subgroups
- Plan to closely monitor implementation and treatment contrast!
- Elicit stakeholder engagement at different stages (early, interim, final, beyond).
The goal is to build reliable evidence about the effectiveness of programs and policies designed to improve the lives of low-income people, so:

- Do they work, how, and for whom?
- Are they cost-effective?
- If they do not work, why not?
- How can they be improved?
- What more can be done?
- Should they be replicated?
BE AWARE OF THE CHALLENGES – WHAT TO OVERCOME IN IMPACT STUDIES?

- Gaining the initial and ongoing cooperation of relevant stakeholders (administrators, organizations, target population…)
- **Raising funds** for research, and the new program, esp. in special demos
- Obtaining research subject cooperation
- Acquiring necessary data
- Meeting ethical and legal standards
- Assuring the program has a fair test – steady state operations
- Keeping evaluation going to observe potential effects.
OUTCOMES VS. IMPACTS: IMPORTANT DISTINCTION

- Outcomes alone are not sufficient for judging effectiveness of social investments
  - Outcomes can be driven by other factors (strong economy, natural progress that highly motivated participants would have made over time anyway)
  - As a practical matter, program operators must use outcomes to monitor program performance and motivate staff. (They're not likely to have access to control group data)
  - It's not that impacts matter and gross outcomes do not; both are important, but tell us different things that must be reconciled in the ongoing effort to build program effectiveness evidence under real-world conditions
  - Policy, funding, program decisions should avoid directing resources toward programs that may have high outcomes but little, if any, value added.
- Both outcomes and impacts can guide policy and practice.
OUTCOMES VS. IMPACTS: SIMPLE ILLUSTRATION

Program Outcome: Employment Rate

- Program A: 80 percent
- Program B: 60 percent
OUTCOMES VS. IMPACTS: SIMPLE ILLUSTRATION

Program Outcome: Employment Rate

Program A

Program B

Program Group
Comparison Group
Outcomes vs. Impacts:
Impacts on quarterly earnings, Years 1 - 3

Source: MDRC calculations from NDNH data (FSS evaluation)
SO COMPARISONS MATTERS FOR ESTIMATING IMPACTS: HOW TO GET VALID COMPARISON?

What would have happened if *everything* was the same, except that the “treatment” or particular change *wasn’t implemented*? What’s the value-added?

- Same group, previous time period (pre/post)
- Similar group, same time period (comparison group, using different matching methods)
- Extensions and combinations of these approaches
- Randomly assign some to “business as usual” some to the “treatment” or intervention of interest (randomized controlled trials, or RCTs)
WHY NOT JUST COMPARE PARTICIPANTS TO NON-PARTICIPANTS?

Differences in outcomes between these two groups could be the result of the types of individuals who enroll in the program.
Differences in outcomes between program and control groups are thus the result of the program and not the types of individuals who enroll in the program.
THREE REFERENCE POINTS TO SUPPORT ILLUSTRATIONS

- **Creating Moves to Opportunity (CMTO).** New demonstration to increase housing choice for housing voucher holders. Test underway in 1 site; site reconnaissance underway for 3-4 site expansion effort. *(Choices and considerations in launch effort.)*

- The federal **Family Self-Sufficiency Program:** Main program to support the economic mobility of housing voucher recipients. Operated since 1990s, through $70 million in annual grants to 700+ housing agencies. Model includes goal-setting, case management, and escrow / financial incentive. *(Successive testing and model iteration.)*

- The **Jobs Plus Program:** Place-based employment and self-sufficiency model scaled up by HUD. Model includes employment services, work-based rent incentive, and community support for work. *(Scaling up evidence-based models and repeated tests).*
Designing A New Impact Study: Choices and Considerations
THE CREATING MOVES TO OPPORTUNITY (CMTO) DEMONSTRATION – CASE STUDY 1

- **Initiated by** researchers Raj Chetty, Nathan Hendren, Larry Katz (Harvard) and Stefanie DeLuca (Johns Hopkins); MDRC study partner, focused on expansion effort

- **Policy context**: Most voucher holders are concentrated in high or moderate-level poverty areas (often segregated). Encouraging (strong) evidence that *moving to higher opportunity areas* may be good for younger children. Limited (strong) evidence of effective mobility programs

- **Learning agenda**: Produce strong evidence on *effectiveness* of mobility interventions. Do they help families move to higher opportunity areas? Do the children have better school outcomes and life experiences? Do the parents experience better economic mobility and well-being outcomes? Cost-benefit results – for families, PHAs, taxpayers.
## WHERE TO CONDUCT THE STUDY? SITE SELECTION APPROACH

<table>
<thead>
<tr>
<th>Considerations</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHA</strong></td>
<td>Operating a strong mobility program, or having an advanced design</td>
</tr>
<tr>
<td></td>
<td>Strong capacity to operate the program</td>
</tr>
<tr>
<td></td>
<td>Strong PHA commitment to the housing mobility interventions</td>
</tr>
<tr>
<td></td>
<td>Openness to participating in a research demonstration, including RCT</td>
</tr>
<tr>
<td></td>
<td>Program size to support sample goals</td>
</tr>
<tr>
<td></td>
<td>Mix of agencies</td>
</tr>
<tr>
<td><strong>Funder prospects</strong></td>
<td>Strong prospects for local funding</td>
</tr>
<tr>
<td><strong>Program model</strong></td>
<td>Approaches that would add to the diversity of mobility interventions</td>
</tr>
<tr>
<td><strong>Local settings</strong></td>
<td>Would add geographic, demographic, housing market diversity</td>
</tr>
</tbody>
</table>
EARLY DECISIONS IN A RANDOM ASSIGNMENT FRAMEWORK (2-ARM TEST)

Participants – WHO, WHAT, WHERE?

Voluntary or mandatory enrollment – HOW? HOW MANY

Informed Consent and baseline data collected – WHAT?

Random Assignment:
- Batch or individual?
  - Program group: Enroll in program
  - Control group: Receive other services in community

Service contrast?
THINGS TO KEEP IN MIND WITH RCTS

- All program participants will be informed that a random assignment process (or lottery) will be used before they enroll in the study.
- The lottery process is the only route into the program during the intake period.
- TA necessary to train program staff on how to conduct the lottery, explain the study, collect baseline information, and answer questions.
- Track everyone who is randomly assigned into the study – including program group members who don’t participate in the program or drop out early – for the study period, funding permitting.
DEFINE DATA NEEDS. WEIGH CHOICES AND TRADEOFFS

- **Types** – each adds a unique perspective, addresses different questions
  - Pre-baseline data
  - Baseline (at enrollment)
  - Program participation
  - Program outcomes

- **Sources** – what is essential, feasible?
  - Administrative data
  - Case files/program records
  - Follow-up survey(s)
  - Field research, observations, other data options

- **Respondents** – who is being tracked, whose perspective is critical?
  - Study participants, households (?)
  - Staff (administrators, supervisors, frontline)
  - Key stakeholders and partners.
**MONITOR IMPLEMENTATION AND PARTICIPATION**

- What we get from monitoring implementation
  - What services were provided (quantity, quality, content, cost)?
  - How were they provided and by whom?
  - Who engaged – take-up, dosage, exposure?
  - Where did the intervention take place? (context, contrast)
  - Any threats to treatment contrast?
- Inform / enrich interpretation of findings – unpack puzzling or expected results (black box)
- Provide information to help with program improvement in addition to findings
Participant engagement
Types of services used, Years 1-3

Source: MDRC calculations of PHA program data
### Assess treatment contrast – is there a there there? (Service receipt in prior 12 Months)

<table>
<thead>
<tr>
<th>Used services (%)</th>
<th>FSS</th>
<th>Control</th>
<th>(Impact)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any services</td>
<td>79</td>
<td>69</td>
<td>10***</td>
</tr>
<tr>
<td>Financial counseling</td>
<td>41</td>
<td>18</td>
<td>23***</td>
</tr>
<tr>
<td>Job search, post employment</td>
<td>47</td>
<td>32</td>
<td>15***</td>
</tr>
<tr>
<td>Education, training</td>
<td>31</td>
<td>25</td>
<td>6***</td>
</tr>
<tr>
<td>Homeownership prep.</td>
<td>17</td>
<td>6</td>
<td>11***</td>
</tr>
<tr>
<td>Supportive services</td>
<td>20</td>
<td>17</td>
<td>3*</td>
</tr>
<tr>
<td>Social services</td>
<td>19</td>
<td>19</td>
<td>-</td>
</tr>
<tr>
<td>Used services in 3+ domains</td>
<td>56</td>
<td>39</td>
<td>16***</td>
</tr>
</tbody>
</table>

Source: MDRC calculations from FSS 36-Month Survey
## Early impacts on employment and earnings, Years 1 - 3.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>FSS Group</th>
<th>Control Group</th>
<th>Difference (Impact)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ever employed (%)</td>
<td>83</td>
<td>81</td>
<td>1</td>
</tr>
<tr>
<td>Average quarterly employment rate (%)</td>
<td>64</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>Total earnings ($)</td>
<td>41,603</td>
<td>41,421</td>
<td>182</td>
</tr>
<tr>
<td>Annual earnings &gt; $25K (%)</td>
<td>20</td>
<td>20</td>
<td>--</td>
</tr>
<tr>
<td>Sample size (N = 2,548)</td>
<td>1,282</td>
<td>1,266</td>
<td></td>
</tr>
</tbody>
</table>

Source: MDRC calculations from NDNH data
Beyond the full sample.
Promising effects on earnings for those not working at study enrollment (FSS)

- Multi-site study
  - Pooled analysis
  - Site-specific
  - Justifiable?
- Full sample
- Subgroups
  - Confirmatory
  - Exploratory
COST STUDY

- Net cost: how much to operate the program (compared to some alternative)
- Cost-Effectiveness: cost per outcome
- Cost-Benefit: Quantifies the benefits relative to the costs
  - Requires a lot of assumptions
  - Often done from multiple perspectives—the individual, the government, and society
Insights From Other Approaches?
IS AN RCT ALWAYS APPROPRIATE?

- For impact questions, RCT logic always applicable
- For deciding whether to measure impacts, consider
  - The question you’re asking (process map)
  - The readiness of the intervention (developing or developed?)
  - The difference between the intervention and “compared with what” (a little? a lot?)
  - The resources available (time, data, people)
- For continuous improvement, other approaches helpful. Some examples…
PREDICTIVE ANALYTICS: A ROBUST APPROACH TO TARGETING

- Use power of MIS & machine learning to analyze data
- Align risk measures with program interventions & goals
- Develop continuous measure of risk
- Intervene & assess

***Better targeting and use of resources***
THE PROMISE OF BEHAVIORAL SCIENCE

Focus on how people actually make decisions and behave

Actions do not reflect intentions

Framing matters

Small—often low cost—changes can have outsized effects

Understand the context

Apply behavioral strategies
Iterative and Successive Testing
FEDERAL FAMILY SELF-SUFFICIENCY PROGRAM – CASE STUDY 2

- **2007 – 2014:** First RCT conducted in New York City. Three-arm test – FSS only, FSS+incentives, control.
  - No positive effects for the full sample. No effects for the main FSS program. Positive effects for FSS+incentives subgroup not working at study entry.

- **2012 – current:** 18-site national evaluation. Do you see different results in other settings?

- **2017 – current:** Can you do better than FSS? What would a stronger program look like? Will it be effective? New 2-site demonstration to test a more robust model.
Replication, Scale-up Efforts, Sustainability
Can’t assume that positive effects demonstrated in the original evaluation will hold when programs are replicated/scaled up

Not every iteration of an evidence-based program can incorporate an impact evaluation

Scaled-up effort may include model refinements

Sustainability vision should be folded into the scale-up plan
Program outcomes can be judged against by asking the following questions:

- Are the program components that drove the positive results being implemented with the same intensity and quality as in the original impact evaluation?
- Is the program serving a population with the same challenges or has it enrolled an easier-to-serve group?
- Are the same outcome measures being used as in the original evaluation?
- How does the operating context, such as the policy environment and local economy, compare with that of the original evaluation?
THE TESTING AND SCALE UP OF JOBS PLUS – CASE 3

1998-2006
Tested in six cities.
Positive impacts on earnings over 7-years of follow up

2011
Replicated in two more cities.
Implementation study only

2015-2017
Dept. of Housing and Urban Devpt. launches nationwide replication. 4-year grants to 24 sites. Implementation and impact studies

2009, 2011: NYC expanded program to 24 public housing developments
TYING IT ALL TOGETHER

- Innovation in social policy is moving toward better programs, more effective policies, more responsible use of taxpayer dollars
- Need more examples of what works
- We have robust methods to test the effectiveness of programs
  - And respond to the need for rapid-cycle assessments
- A culture of continuous improvement and accountability could direct investments in programs with strongest evidence.
Thank You!
nandita.verma@mdrc.org
EXAMPLES OF POLICY BASED ON EVIDENCE:

- Survival of Job Corps and Head Start from the War on Poverty era
- Welfare reform
- EITC
- $1.5 billion investment in Home Visiting based on NFP results
- $50 million in HUD budget for Jobs Plus replication
- Increased federal contribution for Youth ChalleNGe
FACTORS AFFECTING THE COST OF AN EVALUATION

- Type of study
- Breadth of outcomes
- Data sources
- Number of sites
- Sample enrollment period
- Follow-up period

Important to consider the costs relative to the ROI
WHY BEHAVIORAL SCIENCE?

It’s a continuous improvement process.

Many direct-service organizations struggle with program enrollment and participation. We use an analytic framework that changes the way organizations think, and behavioral diagnosis directly addresses these challenges.

It’s cost-effective.

The interventions are typically lower-cost, implemented with limited resources and adapted to an organization’s operating budget.

It’s scalable.

Solutions are embedded within current procedures; as a result, interventions that work can often be scaled with little disruption to the system.
Monthly Escrow Balance and Credit Accrual, Years 1 to 3

Source: MDRC calculations from PHA FSS escrow data.