Public Goods for more impact from impact evaluations and next steps for 3ie

David McKenzie, *World Bank*

What public goods would contribute to improving the use of evidence in policy decisions?

1) Improvements in Measurement Techniques and Technologies

- Example: programs to grow small-scale enterprises (e.g. business training, microfinance, mentoring, grants, value-chain insertion, etc.)

  - How do we know these programs have worked?

    $\Rightarrow$ Want to see that the firms are **more profitable, more productive, and are using better techniques.**

    $\Rightarrow$ But measuring each of these outcomes difficult.

Consequences

- Underpowered studies with lots of noise, and concerns about interventions changing the outcomes of interest.

Table 5: Power of Studies to Detect Increases in Profits or Sales

Study	Group or Individual Randomization?	Sample Sizes in Treatment (T) and Control (C) Groups	C.V. Profits	C.V. Revenues	Attendance Rate	Power to Detect Increase of:			
						25% in Profits	50% in Profits	25% in Revenues	50% in Revenues
Berge et al. (2012)	Group	119 (T), 116 (C) groups (a)	0.80	1.47	83%	0.631-0.842	0.996-1.000	0.239-0.365	0.705-0.897
Bruhn and Zia (2012)	Individual	297 (T), 148 (C)	2.69	n.a.	39%	0.070	0.132	n.a.	n.a.
Calderon et al. (2012)	Two-stage	164 (T), 711 (C) (c)	1.51	1.53	65%	0.263 (b)	0.754 (b)	0.257 (b)	0.743 (b)
De Mel et al. (2012)	Individual	200 (T1), 200 (T2), 228 (C)	0.49	0.91	70%	0.990	1.000	0.632	0.994
Drexler et al. (2012)	Individual	402 (T1), 404 (T2), 387 (C)	n.a.	1.63	49%	n.a.	n.a.	0.231	0.686
Giné and Mansuri (2011)	Group	373 (T), 374 (C) groups	n.a.	n.a.	50%	n.a.	n.a.	n.a.	n.a.
Glaub et al. (2012)	Individual	56 (T), 53 (C)	n.a.	n.a.	84%	n.a.	n.a.	n.a.	n.a.
Karlan and Valdivia (2011)	Group	138 (T), 101 (C) groups	-24.96	2.30	80%	0.057 (b)	0.078 (b)	0.120-0.757	0.335-1.000
Klinger and Schündeln (2011)	Individual RD	377 (T), 278 (C)	n.a.	2.51	n.a.	n.a.	n.a.	0.259 (d)	0.746 (d)
Mano et al. (2012)	Individual	47 (T), 66 (C) (b)	1.23	1.20	87%	0.188	0.571	0.195	0.592
Sonobe et al. (2011)									
Tanzania	Individual	53 (T), 59 (C)	1.99	1.61	92%	0.109	0.292	0.141	0.414
Ethiopia	Individual	56 (T), 47 (C)	1.94	2.49	75%	0.087	0.204	0.072	0.142
Vietnam - Steel	Individual	110 (T), 70 (C)	1.59	0.93	39%	0.075	0.153	0.124	0.353
Vietnam - Knitwear	Individual	91 (T), 70 (C)	-8.15	2.34	59%	0.052	0.058	0.074	0.150
Valdivia (2012)	Individual	709 (T1), 709 (T2), 565 (C)	n.a.	2.29	51%	n.a.	n.a.	0.207	0.626

2) *Research/knowledge of how to improve program take-up*

- Low take-up is bad from both research and policy perspective
  - Research – kills our statistical power
  - Policy – not targeting our programs efficiently or not reaching the desired population.

- Solution: cross-project program of "little IE" experimentation with ways to improve take-up.

Example

- (Opt-in) "Place a check in the box if you will get a Flu shot this Fall **– 42% ticked the box**

- (Active choice) "Place a check in one box: I will get a flu shot this Fall or, I will not get a flu shot this Fall": **62% ticked the box for a flu shot**

- (Enhanced Active Choice) Place a check in one box: I will get a Flu Shot this Fall to reduce my risk of getting the flu and I want to save $50 or, I will not get a Flu Shot this Fall even if it means I may increase my risk of getting the flu and I don't want to save $50: **75% ticked the box for a flu shot.**

3) Funding proof-of-concept pilot interventions

- Current model is one where researchers take programs governments and NGOs are planning on doing as given, try and tweak design around the margins to allow evaluation in some cases, and then evaluate.

- As Howard noted – a lot of 3ie early evaluations were researchers persuading small NGO to do something – what he calls "formative research" is proof-of-concept stage though.

- This means knowledge and insights from researchers of what might be better policies often never gets tested
  - 3ie explicitly disallows funding the intervention itself
  - NSF very reluctant to fund interventions
  - $\Rightarrow$ Consequence – limited policy innovation – governments/NGOs often reluctant to try untested ideas of academics…

  USAID DIV process of funding proof of concept as a stage I, followed by impact evaluation as stage II is nice exception here.

4) Long-term evidence

- Point raised by Scott, Nancy and John this morning.
- Incentives to follow-up long-term not so strong for researchers or operational staff.
- Should be good lower-risk for funders – a lot of issue is that intervention gets delayed or falls through – here you just need to fund a survey…

- Need to think about what "long-term" means
– want more 5-7 year impacts
  - also then 10 year + impacts

Issues of design – what to do when control group now treated?

Which actions should 3ie consider as part of its strategic plan for generating high quality evidence that contributes to effective policies for the poor?

Recommended actions

1) Fund studies aimed at improving measurement of key outcomes

2) Fund and help researchers coordinate across projects on "little IE" - cross-cutting evaluations that would do well to be included in say 5 big IEs.

3) Fund proof of concept interventions.

4) Try to make case to donors/global community of value of doing this type of work in Middle Income countries – avoid excessive focus on low-income countries.

5) Keep up the good things – e.g. flexible timelines that recognize realities of process

6) Reduce the bureaucratic process – whole policy implementation plans and updates process needs refinement.