



Experimental impact evaluations of CGIAR research: Inception and capacity building workshop



Organized by the CGIAR Independent Science and Partnership Council's
Standing Panel on Impact Assessment (SPIA)

with

the **Agriculture Technology Adoption Initiative (ATAI)**
Building E-62 MIT Sloan, 50 Memorial Drive, Cambridge MA, 02142

8-9 February 2015

AGENDA

DAY 1, FEB 8th SUNDAY, BUILDING E-62 MIT SLOAN, CLASSROOM 250

- 09.00 Welcome and Introductory Remarks (Doug Gollin, Oxford/ Karen Macours, PSE)
- 9.15 - 10.00 Drought resistance and water saving in rice production in Bangladesh, Alain de Janvry (UC Berkeley)**
- Presentation (30 mins)
 - Feedback and discussions (15 mins)
- 10.00 - 10.15 *Coffee*
- 10.15 - 11.00 Can personalized input recommendations and better practices improve yields? An experimental approach, Xavier Giné (World Bank)**
- Presentation (30 mins)
 - Feedback and discussions (15 mins)
- 11.1 - 12.30 Open session**
- Common challenges and strengths, key indicators for CGIAR – Implications for study design, and donor perspectives, Doug Gollin (Professor, Oxford & SPIA Chair)
- 12.30 - 13.30 *Lunch*
- 13.30 - 14.20 Does modern storage technology solve African farmers' post-harvest problems? A randomized controlled trial of Tanzanian maize farmers, Jacob Ricker-Gilbert (Purdue)**
- Presentation (20 mins)
 - Feedback and discussions (30 mins)
- 14.20 - 15.10 Impact of hermetically sealed grain bags on grain losses, household food security, and aflatoxin contamination among maize smallholders in Kenya, Nilupa Gunaratna (Harvard)**
- Presentation (20 mins)
 - Feedback and discussions (30 mins)
- 15.10 - 15.25 *Coffee*
- 15.25 - 16.15 Household-level impacts of East Coast Fever control by the Infection and Treatment method (ITM), James Rao (ILRI)**
- Presentation (20 mins)
 - Feedback and discussions (30 mins)
- 16.15 - 17.05 Multi-purpose pigeonpea in Southern Malawi – A double-blind impact evaluation, Kai Mausch (ICRISAT)**
- Presentation (20 mins)
 - Feedback and discussions (30 mins)
- 17:05 – 17:10 *Break, stretch your legs*
- 17.10 - 18.00 Key challenges, way ahead and closing remarks, Tavneet Suri (Professor, MIT & ATAI Co-chair)
- 19.30 Group dinner in Cambridge

AGENDA

DAY 2, FEB 9th MONDAY, BUILDING E-62 MIT SLOAN, CLASSROOM 687

09.00 Introduction and objectives (Karen Macours)

09.30 - 12.00 **Group work** (*with a coffee break in between, 10.15-10.30*)

12.00 - 13.00 Lunch

13.15 - 15.30 **Team wrap-up and discussion of next steps/long-term engagement plans**

15.30 - 15.45 Coffee

15.45 - 16.45 **Presentations by teams (10 mins each) on proposal development and way ahead, discussions**

16.45 - 17.00 Closing (Karen Macours)



Experimental impact evaluations of CGIAR research: Inception and capacity building workshop



Organized by the CGIAR Independent Science and Partnership Council's
Standing Panel on Impact Assessment (SPIA)
with
the **Agriculture Technology Adoption Initiative (ATAI)**
Building E-62 MIT Sloan, 50 Memorial Drive, Cambridge MA, 02142
8-9 February 2015

PARTICIPANT LIST

1	Jenny Aker	Tufts
2	Richard Caldwell	Bill and Melinda Gates Foundation
3	Kyle Emerick	Tufts
4	Xavier Gine	World Bank
5	Doug Gollin	Oxford and SPIA Chair
6	Nilupa Gunaratna	Harvard
7	Ben Jacques	ATAI
8	Alain de Janvry	UC Berkeley
9	Lakshmi Krishnan	ISPC Secretariat and SPIA
10	Karen Macours	Paris School of Economics and SPIA
11	Kai Mausch	ICRISAT
12	James Rao	ILRI
13	Jacob Ricker-Gilbert	Purdue
14	Tavneet Suri	MIT and ATAI Co-Chair
15	John Firth	PhD candidate, MIT
16	Benjamin Marx	PhD candidate, MIT
17	Ben Roth	PhD candidate, MIT
18	Mahvish Shaukat	PhD candidate, MIT
19	Ashish Shenoy	PhD candidate, MIT

Workshop proceedings

Experimental impact evaluations of CGIAR research

Inception and capacity building workshop report, 8-9 February 2015

Doug Gollin introduced SPIA and the rationale for the joint workshop with ATAI. The objectives of the workshop were to (1) provide feedback to the funded projects to help them proceed in a direction that reflects SPIA priorities and generate credible evidence of impact for important CGIAR interventions or further develop theories (on adoption and impact); (2) identify opportunities to identify common grounds between projects and potentially harmonize a few key impact indicators; (3) provide feedback to the unfunded projects to help them develop proposals and make it more fundable.

Day 1, morning session: three commissioned studies

Alain de Janvry presented the project assessing the impacts of drought tolerant rice (Swarna-Sub1) in Bangladesh (UC Berkeley, Tufts, IRRI). Since the field work (piloting phase) was delayed by national strikes in Bangladesh, the presentation and discussion focused on the larger set of projects on rice technologies. He drew on results from a flood tolerant rice experiment in India (both rice varieties were developed under the IRRI STRASA project), and spoke to implications for the Bangladesh experiment. Primary research questions of interest in all cases are: the impact(s) of yield protection from shocks; behavioral responses of farmers to risk reduction, including changes in their risk management techniques and agricultural investments (that may in turn affect expected yields); labor market spillovers; and water market spillovers. In terms of technological efficacy, risk resistance in crops typically comes with trade-offs, but field results for Sub1 shows no yield penalty under non-stress conditions, and experimental plots (not field verification) for the flood-tolerant variety (Sahbhagi Dhan-SD) show similar results.

Results from Swarna Sub-1 experiment in Odisha, India show a 45% yield advantage after 10 days of submergence, and an avoided yield loss of 682 kg/ha (for particular floods of that year). More importantly, there is indication of differential benefits as plots cultivated by Scheduled Castes tend to have 21% more floor days (SCs as a community have historically been subject to social exclusion and are some of the poorest). Hence, there are efficiency and equity gains. There is also evidence that farmers do adjust their behaviors to downside risks, and this can multiply the production gain from technology: e.g. an increase in fertilizer use and rice area planted, and an increased use of more labor-intensive/yield-increasing transplanting techniques. But, it is difficult to separate the pure risk management effect from expected yield effect.

Richard Caldwell noted that the replacement rate for Swarna is low (was introduced almost 36 years ago), and asked if there was evidence that farmers are expanding cultivation. Karen Macours raised a question about moving beyond yield as a primary outcome indicator, idiosyncratic as it is, and the possibility of introducing a little more detail in welfare measures (beyond the metrics proposed). However, the researchers are hesitant to introduce detailed consumption section given its increases survey time and hence response fatigue. There was some discussion on the possibility of collecting this data outside of the main survey schedule (say, through mobile phones).

Xavier Gine presented the project on personalized (agricultural) input recommendations and information on better farming practices in Mexico. In the first phase of the project, farmers were individually assigned to receive either soil analyses results, soil analyses with best practices recommendation to obtain 5 t/ha maize yields, or soil analyses + best practices with targeted foliar fertilizer offers (all farmers choose a plot for testing, yields are measured at this plot level). First round results revealed, among other things, that there was an absence of appropriate fertilizer packages in the markets (limiting farmers' ability to customize inputs) and liquidity constraints in purchasing these inputs. These are issues that will be explored in the SPIA funded second wave.

During the discussion, a number of questions were raised in relation to the implementation components and how this influences research design (adoption and compliance). Currently, a farmer can transfer the in-kind grant (a voucher for input purchase) to a neighbor, and the agro-dealer has no way of identifying treated households from the control/ non-experimental ones, with implications for adoption rates, contamination, and spillover. Researchers are considering having someone verify farmer identity at the agro-dealer. Since randomization is at the individual level, there is also the possibility of farmers sharing information from soil testing of their plots with neighbors (control or non-experimental), and it was suggested that researchers track such information sharing (by asking if neighbors, of each treated household, had received this information and updated land use practices).

Additionally, at the moment, there is no enforcement of recommendations (from the 'shopping list' developed with soil testing results) i.e. farmers can use the voucher to buy whatever combination of inputs they want, and this is another aspect that researchers may monitor (and enforce). Overall, since each farmer chooses a plot for testing based on which recommendations are presented, the farmer choice of plots for testing at baseline and their actions in each season (if they choose to dilute recommended inputs across all plots, if they switch plots – have one tested, but cultivate another) has implications for analysis. While AEWS will be closely working with farmers and monitoring the activity, and such behavior can be closely monitored, there exists a certain level of rigidity to the experimental design.

A specific recommendation to the researchers was on input recommendation tables they present to farmers (based on a non-linear model): currently, they present 3 options to the farmers, each stating what the target output per hectare is and what that option will cost. This implies that the farmer cannot do any other type of trade-off analysis (say, not between package of inputs, but between quantities of various inputs): is this something that can be built-in? Researchers were also cautioned on the difficulty of measuring yields for inter-cropped plots – besides the fact that farmers inter-crop already and yields may be under-reported. There was also a question on scalability – what results from this experiment would imply for soil analysis. Would every plot in Mexico need to be sampled? This is unrealistic. Researchers could then start to explore other levels of aggregation that may be appropriate as the experiment progresses.

Mushfiq Mobarak presented on leveraging social networks to improve agricultural extension services (focused on inter-cropping) in hilly regions of Nepal. He also drew on his on-going work in Malawi along similar lines. The Malawi experiment focuses on pit planting technique in maize that helps retain soil moisture and minimizes soil disturbance (arid areas), with the residue serving as useful material for manure and mulch. Both experiments look at low adoption rates or non-adoption from an information failure perspective: while the traditional response to this has been a focus on agricultural extension system, recent literature that suggests that social learning might be more effective. In addition, in some contexts, shortage of extension workers is a significant constraint.

The Malawi experiment explores alternative dissemination mechanisms (traditional extension officer, extension officer with lead farmer, extension officer with peer farmer) with a second layer of intervention testing the influence of incentives. Results show that peer farmers (without incentives) don't learn about technology themselves, and do not put a lot of effort in communicating with others (and hence they don't adopt). In fact, this arm is indistinguishable from pure control. However, with a small incentive, peer farmers become the best performers and there is increased adoption of practices (11.5 pp increase in pit planting, 26.8 pp in composting). This raises questions such as who optimal peer farmers are, how monitoring could be made more cost-effective etc. – some of which are explored in the Nepal experiment.

In Nepal, besides testing a similar dissemination strategy to Malawi, the incentive layer will become more nuanced (no incentive, performance based incentive, and flat payment – i.e. without monitoring). The technology that is being tested is maize inter-cropping (with tomatoes, ginger or French beans). Outcomes measured include knowledge of technology, adoption, and yield effects through household and on-farm

monitoring surveys. Baseline data suggests that between 74-89% of households had never heard of the technology, and between 88-97% had never used it. There is also some indication that farmers rely more on social networks than formal extension systems for information.

Following the presentation, there was some discussion on identification of peer farmers. Heterogeneity in effects was obvious for peer farmers in Malawi: was this being driven by identity of peer farmers or the number of peer farmers per village? Data suggests that identity matters a lot – similarity between peer farmers and adopters is driving success, and also those who are most likely to convince are those who report having access to inputs and markets. Some concern was expressed about relying on yields as primary outcome measure, particularly as yield reporting is known to be inaccurate for inter-cropping (and is difficult to verify). One could explore the possibility of high frequency data on yields (of tomatoes, ginger, French beans) that are regularly harvested through mobile phones. It was also noted that instead of focusing on yields, one could look impacts from a profit maximization perspective (even as this involves other issues – differential prices and wages). Another option is look at dietary impacts, but an indicator like nutritional diversity would have to be at the individual household member level.

In discussing adoption constraints, it was noted that there was dis-adoption of pit-planting in Malawi – indicative of other things that are going on, and similarly in the Nepali context, there is a similar question on the role of information and delivery vis-à-vis other constraints. However, it is important to recognize that dis-adoption can occur for positive reasons – if farmers are learning, you would expect adoption/dis-adoption cycles.

Summary of general discussions

Both during the open discussion session after the first three presentations, and the late afternoon wrap-up session, participants spoke to the usefulness/appropriateness of RCTs; efficacy of technology being assessed; the choice of topics and research questions – in particular, some of the big puzzles in agricultural technology adoption and impact research; and measurement issues.

Doug Gollin noted that while it may seem ideal to run a million experimental trials to understand what technology works, in which context, this is far from cost-effective or feasible. At the same time, a few studies that have a with/without comparison do not necessarily provide the kind of information one looks for from such trials. One could then attempt to find studies that are a part of a broader learning agenda for the CGIAR. In that context, even if a technology fails along the way, there is value in learning about intermediate outcomes and pathways. Currently, there is healthy skepticism about what the scientists are telling us about technologies and we recognize that the biggest constraint to adoption is ineffective technology (a first order question). To take the example of what de Janvry and team note in their STRASA trials, is there a systematic process that leads to overestimates of yields by scientists who develop these varieties. Then, what can the CGIAR do to improve these estimates?

On measurement related issues, Gollin noted that measurement of inputs and outputs, of shadow prices is difficult and non-trivial. And, there is plenty of room to advance what we do, given the recognition of heterogeneity. We should no longer continue to use a common set of prices (e.g. there is significant seasonal and by activity variation in wages). Another measurement issue is the focus on crop-specific yields - it is important to think about household/farm level.

Continuing on these lines, given the measurement issues for labor, wages, prices etc. Karen Macours encouraged researchers to think about other welfare effects at the household level, and pre-emptively think about what we could learn if there are 'zero results', in some senses. For example, in case of Gine's work even if there is no difference in tailored versus untailored inputs, one could still understand through knowledge related questions in the survey if farmers changed their priors about land quality and respond to that. Richard

Caldwell also noted that BMGF were developing harmonized indicators and measurement guidelines for the projects they are funding.

In responding to the what/when to RCT question, Tavneet Suri began by noting that one could evaluate technologies themselves or evaluate aspects that are further away from technologies. In either case, it is clear that one needs better experimental station trials to start with – something that SPIA currently doesn't focus on. Alain de Janvry suggested a similar approach – thinking of a hierarchy of research questions and experiments: technical questions on efficacy of technology is best answered in labs; questions on adoption and usage – observe real-world behavior and understand dimensions that could influence outcomes – dimensions that can then be tested through experiments; behavioral responses to the technology and this inducing a whole set of new behaviors – ensure you observe these in your experimental trials etc.

Suri also noted that experiments tend to operate at a small scale for the household – one cannot expect the household to not respond to the (small) scale of intervention, and that technologies have specific traits that farmers prefer (information on which is rarely collected during studies) which may explain their adoption decisions in the future. Hence, one prudent strategy might be to go back and identify places where RCTs or rigorous trials were done before, and keep measuring adoption (at intensive and extensive margins) over time, even if not impact. Briefly, Gollin noted that SPIA is addressing this through one activity (long-term or large-scale call for proposals). Speaking to the adoption issue, Richard Caldwell noted that there is no accountability for adoption at CGIAR (in addition to absence of systematic collection of adoption and diffusion data), and wondered out loud if incentives should be put in place for that, that they should have to make a business case for technologies.

Suri then summarized some of the unanswered questions related to agricultural technology adoption and impact: researchers don't have a sense of seasonal price differences (how often prices change, how big a magnitude of change etc.), and more importantly, a story of why price differentials exist. For instance, even where price differentials are observed, quantities don't change – so, what is the market failure? (infrastructure and transport?). She also cautioned against focusing on profits because aspects like labor and theories on labor markets become important.

Day 1, afternoon session and Day 2: four non-commissioned studies

The focus of the afternoon session on day 1, and (a full) day 2 of the workshop were the 4 non-funded studies. On day 1, **Ricker-Gilbert** presented a proposal evaluating the impact of PICS (hermetically-sealed) storage bags on maize farmer's post-harvest problems in Tanzania. **Nilupa Gunaratna** presented an impact assessment of the hermetically-sealed bags (IITA related) on grain losses, food security, and aflatoxin contamination among maize farmers in Kenya. **James Rao** presented a proposal on East Coast Fever (ECF) control by infection and treatment method in Tanzania. **Kai Mausch** presented a double blind trial on improved pigeonpea in Southern Malawi. All four presenters received feedback from other researchers, and on day 2, graduate students affiliated to ATAI opened the detailed discussion session (1-1.5 hours) on each proposal with a commentary on research hypothesis, and research/intervention design. Since these proposals are being developed further, this summary will not delve into feedback given to each presenter, and will instead focus very briefly on common discussion points.

In case of the hermetically sealed bags, a higher-level question is why people don't hold maize: is it a post-harvest liquidity issue or an insects/mold issue? Currently, one could in fact argue that there is excessive storage at the household level. Specific to the interventions proposed, there are questions on the importance of insect/mold loss – while self-reported loss varies between 5-10%, agricultural ministries and extension workers see to think post-harvest loss is a huge problem. Measurement is complicated by the fact that farmers might sell in order to avoid losses. An indirect indicator of losses could then be usage of insecticide against weevils. Similarly, in case of aflatoxin, measurement is fairly noisy (can vary within one bag). Technological

efficacy (bags per se) seems uncertain – e.g. it is not entirely clear what happens if contaminated grain is stored inside the bags, and it appears that drying of maize before storage is a critical practice. Similarly, in case of East Coast Fever, the ITM does not protect livestock against other tick-borne diseases, and while some of the low adoption levels observed may be explained by logistics (difficulty in transporting/storing vaccine, or unavailability of smaller package sizes of vaccines) or by negative experience (stories of mortality and morbidity spreading, or dissuasion by animal health workers and vets who had more to gain from treating ECF), there is room to use existing data on roll-out – perhaps through a large-scale survey – to shed light on these adoption-related issues.

For the pigeonpea experiment, it could be made clearer what the researchers hope to learn from a double-blind trial. In general, people may have sub-optimal responses to the treatment and if this affects their management response, what will this indicate? While it is true that disentangling the effect of genetics versus efforts will inform ICRISAT research efforts, if researchers think framing matters, this could be tested in other ways – particularly as there isn't a strong hypothesis, currently, of how farmers will react to the new varieties (will they put more effort or less effort?). Also, while pigeonpea makes for a good double blind – seeds cannot be differentiated – it is still possible that differences show up during the growing season or that farmers throw the seeds that were given away. Hence, across all four proposals, there is room to do substantial diagnostics to develop research hypothesis on adoption and farmer response, and frame issues better before designing an RCT or non-experimental IA.