WORKSHOP REPORT

New approaches to assessing the impact of agricultural research on poverty and undernutrition:
A brainstorming workshop on ex-post impact assessment
IFPRI, Washington DC, 3rd & 4th December 2010

MAIN OUTCOMES

- Micro- and macro-level approaches to assessing the impact of agricultural technologies on poverty were reviewed. While there was some discussion about indicators and measurement issues, and about complementarities (realised and potential) between macro and micro levels of analysis, the main theme of the discussions came back to causal identification – the ability of a particular evaluation design to generate a rigorous counterfactual.
- There was broad agreement that impact assessment in the CGIAR should aim to be more rigorous.
- A strong case was made for greater use of experimental methods (such as Randomised Control Trials – RCTs) in impact assessment, but this was not an unqualified recommendation. Some participants argued that RCTs are useful but not the only way of generating rigorous counterfactuals. The requirement to design and integrate RCTs *ex ante* of an identifiable 'intervention' to disseminate a technology, the relatively small scale at which they could be applied, and the ability to assess only adopter level (i.e., producer level) short-term impacts were also noted as limitations of RCTs.
- Given the complexity and diversity of agricultural livelihood strategies, and the large number of social structures that condition the relationship between technology and poverty, practitioners need to be very clear about the theories motivating impact evaluation designs. It is essential to be able to demonstrate the pathways through which impact occurs or does not occur for each case.
- A number of participants identified a need for a system-wide approach to filling cross-cutting data gaps (household panel data; adoption surveys; poverty maps) to enable the tracking of adoption and to lay the foundation for rigorous impact assessment. This could best be initiated at the country level.
- For micro-level analyses to better feed into Computable General Equilibrium (CGE) models, data about household income and expenditure should be collected.
- Donors reinforced the need for good impact assessment – with budgets tight they need to be able to justify allocations to agricultural research relative to other possible means for reducing poverty.

ACTION POINTS

- All CGIAR centres are urged to start developing detailed ideas for full proposals of empirical case-studies that could be funded in 2011.
- A steering committee of SPIA, IFPRI and USAID is managing a budget in 2011 of $500,000 on the topic of poverty and undernutrition impacts, to allocate to empirical studies. In addition, SPIA has approximately $100,000 allocated in its budget for these studies. SPIA will put out a call for proposals for its funds in early 2011.
- There are a number of additional resources that could be tapped into to fund good micro-level studies - 3IE[1] and ATAI[2] for example.
- Alain de Janvry has invited centres to send scientists to a "Match-Making" exercise (matching interested economists from the top research universities in the US to field-based organisations with interventions to assessed) at the Brookings Institution, January 21st.
- Per Pinstrup-Andersen has invited centre scientists to submit ideas for how to fill crucial data gaps at the global level, to enable us to better evaluate change in agricultural systems, nutrition and poverty.
- SPIA will develop a typology of methods, based on the discussion on Day 1, for consultation with the group.

---

[1] www.3ieimpact.org International Initiative on Impact Evaluation
[2] http://atai-research.org/ Agricultural Technology Adoption Initiative

ISSUES DISCUSSED

SESSION 2: OVERVIEW OF APPROACHES AND METHODS

Bob Herdt on the literature to date: While there have been a few very good studies that have helped develop the literature, in general it is very challenging to do good impact assessment. Technology is hard to define, and may be dynamic already, without a research contribution. Poverty likewise is not static. Both poverty and technology are difficult to measure. Productivity is a function of other things than technology. We need to foster cross-centre collaboration – joint, micro-level, village level, ongoing data gathering with the centres doing the data gathering.

Alain de Janvry on a critique of the impact assessment literature: The Average Treatment Effect on the Treated (ATT) is endogenous to adoption – there is self-selection into treatment. So we have to be careful that our impact estimates correspond to the field of adoption. In addition, the value of ATT changes over time: with learning by individual farmers; as adoption beings with favoured areas and proceeds to move out to the margins; and as general equilibrium effects start to emerge at scale. This poses a problem for the best timing of impact assessment - at the very start, we have no impact; at late adoption we have general equilibrium effects making things more complicated; and at full adoption, we have no counterfactual. The best option is to measure impact at small scale, and do it very carefully. An observational method such as Propensity Score Matching does not work when choice has been exercised by non-adopters. Also, we need to carefully handle spillovers as they are likely significant in many technologies.

Per Pinstrup-Andersen on links to nutrition: We should be very careful in choosing relevant measures for nutrition impacts – for example, how the difference in digestibility of micronutrients from different sources can result in a different outcome than would be suggested by total micronutrient content. Also, it could be time to reconsider the link between diversity in agricultural production systems (particularly semi-subsistence) and diversity in the diet.

Ben Davis on contingent outcomes: We should be aware of the role of political economy in conditioning impact of agricultural technologies – including social structures and connectedness. Impact assessment should be endogenous to the implementation of the programme, rather than an after thought. We should also think carefully about where the programmes are placed within countries – for example there is no correlation between field stations of CIMMYT and a poverty map.

Meredith Soule on a donor's demands: Donors need to be able to defend the allocation of resources to agricultural research rather than to other means for reducing poverty, such as roads. USAID is very focused on delivering against the MDGs, but are the links being made between agricultural research, more food and less undernutrition? In delivering the evidence for accountability and for learning, how much money is really being reallocated away from programme implementation and towards impact assessment? Good evaluation is costly and the impact assessment community needs to be communicating it to the rest of the CGIAR.

Peter Hazell on diversification and multiple pathways to impact: With diversification of incomes by households it is a) harder to assess impact on poverty; and b) there is an endogeneity problem - raising productivity in agriculture could reduce income from other activities. This implies that we need a whole household model when it comes to assessing the impacts of agricultural research. We also need to develop a typology that links different causal pathways between agricultural research and particular types of impact, and different methods for assessing these connections.

SESSION 3: MICRO METHODS AND DATA

Elisabeth Sadoulet on experiments: Good impact assessment should be planned in advance and involves an explicit research design. There is a difference between the unit of observation and unit of randomisation – the latter can be higher than the former, and good practice suggests village level randomisation is best, in order to pick up spilover effects. Supply-side treatments (e.g., making it available through seed systems) should be used, rather than demand-side treatments (e.g., subsidies). There may be other sources of exogeneity that can be exploited, which then would obviate the need for a randomised design.

John Hoddinott on panel methods: SPIA-sponsored evaluations should use double-difference (DD) designs. If you're worried about credibility, you cannot use single-difference ex-post designs – you will not convince your critics otherwise. Some unobservables we can turn into observables – such as asking respondents the extent to which they agree to the statement: "I have the power to make important decisions that change the course of my life". By asking questions relating to personal agency and incorporating this data in econometric analysis, we can further reduce selection bias related to the presumed "unobservable" characteristic of entrepreneurship. There are also almost always differential effects on men and women – intra-household impacts – and these should be treated seriously. Downside shocks often have very significant impacts on nutrition so variance is as important as the mean.

Aliou Diagne on cross-sectional methods: The availability of the technology is the key assumption underlying the criticism in the paper by de Janvry et al. Non-adopters cannot adopt what they do not have access to. If there are restrictions in access, then access can be a valid instrument for adoption, in which case, it is possible to get clean identification without recourse to an experiment. A range of alternative estimators, each with different generating assumptions, are available to be used with care – a range of alternative estimators can be used on the same study to test for differences due to assumptions.

Michael Carter on rigour: RCTs are powerful but in implementing them there is always scope for the introduction of bias. We should think carefully about bias and aim to bound its severity. Proponents of RCTs should be aware of the dangers of simplistic epistemological hubris – we should not just throw out the information we get from studies carried out using alternative identification strategies. Additionally, we can also get very precise estimates of the wrong thing – Santa Claus treatment effects, where interventions are handed out for free. Fully endogenous behaviour (who adopts and how) is captured by the Policy Relevant Treatment Effect (PRTE). Let's not forget about differentiation within "the poor". What good is an unbiased estimator if the confidence interval is so large? If we examine the mean square error, we can pick the estimator that is closer to the truth.

Peter Lanjouw on data sources: We should think about the interrelations between outcomes – they are often all mixed up with each other. We should also be worried about data costs when a study is implemented over a large number of villages. It is essential to be able to demonstrate the pathways through which impact occurs or does not occur for each case. Small area estimate methods now have much more data to draw on and agricultural and population censuses can provide very useful data. Poverty maps, LSMS panels, the DIME initiative (AADAPT) and the Africa-gender group – are all useful resources for these kinds of studies.

John Hoddinott on the need for theory: Look forward, not backward. There is value in the process and pathways, as well as the aggregate. We should think about modelling different behaviours – *ex-ante*, as well as incorporating them in the studies. When it comes to nutritional effects, womens' time might be the most important determinant of outcomes. The poor quality of impact assessment in the CGIAR likely reflects the revealed preferences of the DGs. A recommendation for SPIA – offering a significant sum of money, to be matched by the centre, for few but large scale assessments will likely get much better studies than many but smaller studies.

SESSION 4: MACRO METHODS

Will Martin on GTAP simulations: Vegetables represent the most significant potential contributor to poverty reduction – more than rice, maize or wheat – which might suggest a significant change from the Green Revolution era in the structure and function of global agriculture. We should validate the *ex-ante* analysis against results from specific impact evaluations. There is the potential to target poverty reduction on certain commodities. Often not critical to reach all producers – the general equilibrium effects ensure that the majority of poor people will benefit.

Clemens Breisinger on country CGE models: Recent developments are twofold: better disaggregation over last two years – different agroecological conditions, household disaggregation; and linking CGE to nutrition outcomes, both on individual level and child malnutrition. In CGE models we always assume exogenous TFP growth, something that would benefit from better micro estimates. The relationship between undernourished people and GDP per capita is strong but not as strong for child malnutrition. India and Egypt are the main problem examples where income growth is happening at the same time as child malnutrition is increasing. If we take the outcome of the CGE model and apply a nutrition elasticity, we can see how a change in agricultural productivity impacts on household's nutrition. We would then like to compare agricultural productivity increases with other policy options.

Joaquim Bento de Souza Ferreira Filho on Brazilian CGE modelling: The normal approach to evaluating technological change does not take into account vertical effects or horizontal effects (competition in factor markets). There is a potentially negative social effect: which factors go unemployed? What happens to their prices? Effects are country-specific and depend on structure of labour markets. The model is a CGE of Brazil with micro-simulation component to look at poverty. With an unbiased technological change TFP shock there is a fall in employment and wages of poorest 10% which has potential to increase income inequality. We know that technological change in agriculture in Brazil has been labour-saving. When modelling the impact of technological change in CGEs, it is important to be clear about the type of technology that is driving productivity increases.

Shenggen Fan on macro models: An exogenous TFP shock does not look at the cost side. TFP increases are not just "manna from heaven" - how can we link TFP growth to investment? A 1% TFP increase in rice is fundamentally different from a 1% TFP growth in vegetables. Maybe growth in rice has plateaued? We need to be careful about the parameters used – the elasticities between growth and poverty reduction. Can we allow these to change over time? When reporting results from macro models, are we looking at GDP per capita, poverty or nutrition? How can we take nutrition, health more seriously in our analysis? Employment is the big challenge in SSA and S Asia with 40 to 50% unemployment in some countries. Should this be included explicitly as an indicator?

Greg Traxler on priorities for the future: SPIA has two uses of impact assessment. Evidence (for accountability) and making allocation decisions. How much of SPIA's obligation is about resource allocation decisions? There has been an impressive increase in the sophistication in simulation models in the past few years, as evidenced by the presentations. The leadership of the Gates Foundation doesn't care about agriculture for its own sake, so the pressure is there to justify the choice to investment in agriculture relative to opportunity costs. The incorporation of micro analysis is much better – much more credible than the CGE models used to be. The finding that the strongest pathway for poverty reduction is through reduced food prices and increased real incomes poses difficult question for the Gates Foundation, which has focused on smallholder productivity. Does this need to change?

David Raitzer on priority-setting and sector-specific models: We should be careful to examine the cost of achieving different alternatives and what the probability of success would be. Cross-checking with *ex-post* studies is a useful place to start. There are distributional implications of price effects – this can challenge the focus on poverty benefits to poor producers. For example, an 8.5% supply shift in rice over 25 years leads to 9 – 17% domestic price reduction. In the Martin paper, major countries are omitted from the pool – China, India and Brazil

which could explain some slightly counter-intuitive results. If we construct a poverty-weighted value of production, rice is ranked first, maize tenth. In addition yield gaps can be quite significant. Can research help reduce these gaps?

SESSION 5: LINKING MICRO AND MACRO APPROACHES

Peter Hazell on integration across scales: Micro, meso and macro effects can all be working in different directions. In this case, the researcher can pick the scale and timescale that proves whatever you want to prove. In the early studies on the Green Revolution we can't gloss over the fact that there really were losers. Sociologists, geographers focusing on the micro, still emphasise this fact. Impact assessments need to try and understand the whole mosaic and do need to say something about processes – if you don't even think about it in the impact assessment design, you're not going to convince the critics. Can we provide a framework / typology for the kinds of thing that should be included in a comprehensive impact assessment? This should be collective action across the CGIAR, with centre scientists and IFPRI scientists.

Elisabetta Gotor on integrated micro and macro modelling: We get two orders of results at the same time. If we divide a regional agent into 10 social strata (rich to poor), and then analyse shocks to a CGE model we can look at the impact on income and expenditure among these 10 strata. Micro-macro modelling is useful for *ex-ante* assessments, but wonder about whether it is desirable *ex-post* on the centre-by-centre basis. Is it more useful to do macro modelling for the CGIAR as a whole? In addition, the CGIAR's activities go beyond improved technologies – capacity building, conservation of diversity. We need to find methodologies to capture externalities that contribute to poverty outcomes.

Per Pinstrup-Andersen and Bob Herdt on the need for collective action: Need to look beyond the centre to the system as a whole, organising ourselves to do this as the CGIAR. We need to know where the technologies are adopted and how that is playing out through different impact pathways.

Meredith Soule on metrics: If we are talking impact of agricultural research on poverty and undernutrition – how to measure these things? Currently the CGIAR is inconsistent, even within the same study. The same observation holds for "hunger".

Will Martin on a typology: We need a range of different approaches but these should be combined in a typology that recommends types of method.

SMALL GROUP WORK: 3 TYPES OF CASE-STUDY

- Group 1 – Technologies at early phase of adoption
- Group 2 – Technologies at late phase of adoption
- Group 3 – Country-level CGE modelling assessing the impacts of multiple CGIAR interventions over time

Group 1 – Early phase adoption

Elisabeth Sadoulet outlined 5 Steps for the group to follow: 1) define adoption; 2) define intervention; 3) what is scheduled for implementation; 4) possible counterfactuals; 5) outcomes of interest and time frame over which they would be observed (i.e. data collection required).

Cases from Africa Rice (Participatory Learning and Action Research, #2 in the list of 26 studies), Harvest Plus (Biofortified Beans in Rwanda, #15), IRRI (#20, Sub-1 Rice) and World Agroforestry Centre (#25, Fodder Shrubs in East Africa). The group went in detail through the five steps for Sub-1 Rice in India and Biofortified Beans in Rwanda (see Annex 1).

Group 2 – Late phase adoption

Three cases were discussed, from Bioversity (Home Garden Conservation Strategy in Nepal, #5); ICARDA (Lentil variety Alemaya, #16); and CIMMYT (Zero-tillage in Indo-Gangetic Plains, #9). In the Bioversity case, it was agreed that further clarity was required to identify a particular intervention that could be studied – was it going to be a subsidy by the government? Just a set of guidelines at district level? In the ICARDA case, a full proposal has already been developed using village-level Social Accounting Matrices (SAMs), but there was discussion about whether an econometric approach could be used. In the CIMMYT case, no panel data for the IGPs exists, although there is a good opportunity for using double-difference methods based on district-level administrative data, given the large scale of adoption.

Group 3 – Integrated country-level modelling

Ethiopia was used as a possible case as there are a number of CGIAR centres involved. Possible successful interventions to look at include maize, potatoes, wheat, lentils. Those with mixed track record include barley, livestock, and policy interventions. CGIAR centres would need to work together to establish a comprehensive database on what has been done. Getting data on households' revenues in the micro studies would best ensure complementarity between micro and macro level estimations. A framework is presented in Annex 2.

In each of the three groups, the case-studies were discussed only as examples. This whole exercise aims to be inclusive, despite the fact that not all CGIAR centres were represented at the workshop. Full consideration will be given to all centres who submit proposals next year. There was an excellent response to this initial request for ideas to develop our thinking, with 13 centres submitting a total of 27 ideas of possible cases, and there wasn't time to discuss all of them in detail in the workshop.

DISCUSSION ON FILLING DATA GAPS AT THE SYSTEM LEVEL

We are missing large scale adoption data and information on differential effects – which gains are accruing to poor populations? Poverty maps overlaid with maps of different crops is one approach. Income and expenditure data on non-rice crops are missing when it comes to IRRI being able to assess the impact of IRRI's research on poverty. There is a big data gap that would need to be filled to link crop production to expenditure / poverty. We will also need to collect data on a lot of other variables if we want to avoid omitted variable bias. World Bank LSMS data are available but lack information on varieties etc. to allow link to production. There is the CRP number 2 on policies / social science. How can we coordinate while still meet the individual needs of the centres?

Annex 1 – Details of possible strategies for two micro-level early adoption / impact studies

	Sub-1 Rice in India (Raitzer, IRRI): Yield advantage over existing improved variety for submergence events of between 7 and 17 days duration	**Biofortified Beans in Rwanda (Birol, HarvestPlus):** Beans with same agronomic characteristics as existing improved varieties but iron-rich
Define Adoption	Farmers planting Sub-1 rice in the year after they receive free seed	Farmers planting iron-rich varieties in the year after they receive free seed
Define Intervention	Government of India's roll-out of 5kg seed mini-kits	Roll-out of seed + information materials about nutrition benefit at the level of cooperatives (1 or 2 per village). Everyone in the cooperative gets the seed.
Implementation schedule	Roll-out over entire Swarna area with Swarna Sub-1 in Andhra Pradesh by 2013	Roll-out: 36 villages in 2011; 36 more in 2012; 36 more in 2013 (108 villages total)
Possible counterfactual / "Strategy"	Risk exposure design. Stratify a sample based on: a) phasing of roll-out (need to know where it will be happening and when); b) probability of submergence. Then match communities with the same probability of being flooded and same adoption status – some will experience the flood shock, and some won't. It is a comparison among the adopters - we assume that we will reach high levels of adoption quickly.	RCT, phase-in over treatment villages with a maintained control group. Village-level randomisation. Rounded up to a roll out to 120 villages for simplicity of exposition. Define 10 groups of 24 villages which are paired up (10 x 12 x 2) = 240 villages total. One from each pair is randomly allocated to treatment (T); the other to control. Within the 12 T villages from each group, 4 are randomly allocated to each of phase 1, 2 and 3. Data collected from random sample within each village – to enable measurement of spillovers.
Outcomes / time frame	Restricted farm-level profits; Income poverty; Intensification response	Continued adoption over time; blood testing for anaemia; z-scores; restricted farm-level profits?

Annex 2 – Framework for integrated country-level modelling of CGIAR research impact