SPIA Activities Update

Prepared for ISPC 10 (15 – 17 Sept 2014) Meeting
Ceremonial Hall (Festsalen), University of Copenhagen, Copenhagen K., Denmark

This progress report provides a brief background and update on SPIA activities since the ISPC 9 meeting held at IFPRI HQ in Washington DC, March 2014. As SPIA's activities are now synonymous with its program of work under Strengthening Impact Assessment in the CGIAR (SIAC) which commenced in early 2013, the report is organized by the Objectives spelled out in the SIAC program of work.

Doug Gollin, who chairs SPIA (since July 2012), will present a more detailed account of these SIAC activities at the ISPC 10 meeting on Tuesday, 16 September. SPIA has two other members: Bob Herdt who joined SPIA in February 2014 and JV Meenakshi who joined SPIA in May 2014 (see http://impact.cgiar.org/about). Karen Macours and Erwin Bulte serve as Research Activity Coordinators. Full-time staff at the ISPC Secretariat in Rome comprise Tim Kelley, James Stevenson, Lakshmi Krishnan and Ira Vater. SPIA partners with a number of academic researchers and institutions including Michigan State University, Virginia Tech, and the World Bank LSMS-ISA team.

SIAC Objective 1: Develop, pilot and verify innovative methods for collection and assembly of diffusion data

Underpinning this objective is the development of a robust set of methods for routinely tracking adoption of CGIAR-related technologies in a cost-effective manner. Such information is a prerequisite for achieving the highest quality assessment of outcomes and impacts. A set of activities will test innovative ways of assessing the adoption of improved varieties of crops, livestock and fish technologies, agronomic and natural resource management interventions, with the goal of eventually embedding protocols derived on these tests into large-scale surveys carried out by others.

Progress since March, 2014:

This objective is managed by Michigan State University. There has been significant progress on activities 1.1 (new methods for crop germplasm improvement adoption data collection) and 1.2 (new methods for NRM adoption data collection). Three DNA fingerprinting experiments, comparing a gold standard of DNA fingerprinting to alternative methods for varietal identification, are in the process of being analysed following data collection in 2014. The crop-country combinations are cassava in Ghana, beans in Zambia and maize in Uganda. The Bill and Melinda Gates Foundation hosted a workshop in August 2014 where all three projects were presented as work in progress, alongside other DNA fingerprinting initiatives funded by the Foundation. This peer exchange helped to clarify some issues regarding protocols for sample collection and looked at the potential for integration with large-scale household surveys, such as those described below in activity 2.4. A brief workshop report is attached as annex 1 to this update.

Following a competitive call for proposals in 2013, two experiments for new methods for collecting data on NRM adoption are being implemented by IRRI (remote sensing for alternative wetting and drying) and CIMMYT (a cell phone app for monitoring improved nutrient management practices) and will be completed in early 2015. A feasibility study is being prepared by ICRISAT on using remote sensing to track improved tank bund management in India.

Workplan for 2015:

In 2015, following completion of the analysis of the three field experiments comparing DNA fingerprinting to alternative methods for varietal identification (cassava in Ghana; beans in Zambia; maize in Uganda) and completion of the two case studies testing innovative protocols for tracking diffusion of NRM technologies (hyperspectral signature analysis for tracking adoption of alternative wetting and drying of rice in Gazipur Bangladesh; mobile phone based applications in tracking adoption of improved nutrient management in India), MSU will hold a brainstorming meeting (mid-to-late 2015) to assess results and outcomes and prepare a document on best practice

guidance for gathering data on the diffusion of agricultural technologies. In addition, MSU will be exploring on a pilot scale new alternatives for outsourcing the collection of data on a routine basis that will allow the CGIAR to track adoption of major agricultural technologies in developing countries.

SIAC Objective 2: Institutionalize the collection of the diffusion data needed to conduct critical CGIAR impact evaluations

The objective here is to compile and make available the best information on outcomes that are at least plausibly attributable to CGIAR research outputs, and on a large-scale. This is where a key bench-marking function for the CRPs is most obviously fulfilled by this program. Large gaps in existing adoption databases for genetic improvement technologies (activity 2.1), natural resource management technologies (activity 2.2) and policy-oriented research (activity 2.3) will be filled for priority regions. In addition, under activity 2.4, the World Bank Living Standards Measurement Study-Integrated Surveys of Agriculture (LSMS-ISA) team and the CRPs will work together with NARS partners and statistical agencies to see how some of these processes can best be integrated into existing surveys to reduce cost and increase frequency of data collection.

Progress since March, 2014:

Activity 2.1: Crop germplasm improvement, filling gaps in adoption estimates. An initial planning workshop led by Michigan State University and a number of regional workshops, led by CGIAR Centers were carried out in order to launch the data collection effort for missing crop-country combinations in South, South-East and East Asia. Contracts with the Centers to implement the NARS training and data collection activities are almost all finalized. Training sessions in the field have commenced in specific countries (Nepal, Laos).

Activity 2.2: Natural resource management, filling gaps in adoption estimates. This was initially part of the Michigan State University sub-grant but it was agreed in Jan 2014 that SPIA would manage this part of the program. Subsequently, SPIA contracted a consultant to analyze annual reports and old Performance Management System (PMS) data as a basis for consultations with relevant Centers about prioritisation, methods and data availabiilty, for collecting national estimates of adoption for key practice-country combinations. In the second phase, a two-stage call for proposals will be issued (end of September 2014) to identify a consultant who will build this database of adoption outcomes.

Activity 2.3: Policy-oriented research (POR) influence claims. A consultant with expertise in POR outcomes was hired by SPIA to undertake the first phase of this activity – assembling an initial database of POR outcomes drawing on information extracted from earlier CGIAR PMS data files from 2006 through 2010. This represents a wealth of information of all types of "outcomes" initially proposed by the Centers and subsequently vetted and culled by the ISPC using external reviewers. The initial POR Outcome Database will be shared with the relevant Centers (those associated with the outcomes reported) to verify that the earlier presented information is still accurate, or to update. The initial POR Outcome Database will in itself represent a useful source of information for CGIAR stakeholders wanting to see in one place a compilation of significant credible POR related outcomes from CGIAR research. The second phase (proposed) would consist of building on this database by, for example, searching relevant documents published since 2010 or soliciting directly from the Centers/CRPs. Relevant outcomes that this SIAC Activity would seek to document in the POR Outcomes Database include:

- Agricultural and relevant macro, trade and nutrition/health policies, that can have a large impact on economic incentives in agriculture, as well as modulating the poverty and nutrition impacts of some new technologies
- Monitoring and evaluation work that substantially improves the targeting and/or delivery of social safety net programs (e.g., conditional cash transfers, productive safety nets programs)
- Management practices/protocols/agreements adopted at national or international levels
- Levels and types of investments in agricultural research, roads, markets and other infrastructure
- Expansion of training and institutional capacity (e.g., through farmer field schools)

- Major international conferences / workshops around a highly relevant theme, e.g., IFPRI's 2020 Vision conferences.

Activity 2.4: Institutionalising adoption data collection. SPIA is now in partnership with the World Bank Living Standards Measurement Study – Integrated Surveys of Agriculture (LSMS-ISA) team, through two research associates that SPIA has hired to work in LSMS-ISA countries over the period mid-2014 to mid-2016. These research associates are the focal points for all efforts to improve the way that agricultural technologies are uniquely identified in the surveys of 8 Sub-Saharan African countries (Ethiopia, Malawi, Uganda, Tanzania, Nigeria, Niger, Mali, Burkina Faso). Michigan State University is also working in parallel with the same objective with a focus on Mozambique, Zambia and India.

Workplan for 2015:

Data collection on the adoption of improved varieties will take place for the most significant crops (in terms of area of production) in Bangladesh, India, Iran, Nepal, Pakistan, Sri Lanka, Cambodia, China, Indonesia, Laos, Malaysia, Mongolia, Myanmar, Philippines, Thailand, Viet Nam (activity 2.1). In addition, SPIA will competitively commission a two-year process (2015-16) for collecting data on the adoption of a set of the most important natural resource management practice – country combinations. This activity (2.2) will be managed by an outside entity, following a competitive call for proposals in the final quarter of 2014, and will be independent of the CGIAR centers. The data collection will comprise a combination of expert opinion methods, survey estimates, and possibly remote sensing or other frontier methods. For policy-oriented research outcomes (activity 2.3), a series of steps (still to be defined) are envisioned in going from assembling a larger body of recently submitted potential CGIAR POR outcomes to a reviewed and verified smaller set of "credible" POR outcomes to add to the POR Outcome Database[1]. For the partnership with the LSMS-ISA, SPIA has two Research Associates working full-time to support the institutionalisation of CGIAR adoption data into these surveys (activity 2.4). By the end of 2015, SPIA will have preliminary data from a large-scale experiment in four districts in Malawi in which cassava DNA fingerprinting will be integrated with a multi-topic house survey. This will also help with the integration of future improvements on natural resource management or other crops in Malawi. In Ethiopia, SPIA will have facilitated the integration a number of amendments into the way the Central Statistical Agency runs its surveys, in order to generate data that is more relevant for CGIAR adoption monitoring and impact assessment.

SIAC Objective 3: Assess the full range of impacts from CGIAR research

While work under Objectives 1 and 2 pave the way for future *ex post* impact assessment studies, Objective 3 activities are focused on carrying out a number of impact assessments of CGIAR research and development initiatives along the entire chain of causation - from research investments to the System-Level Outcomes. Since this causal chain is long and complex, SPIA is approaching it from a number of different perspectives: case studies that focus on measuring the impact of CGIAR research on health and nutrition (activity 3.0); long-term large-scale studies of impact for major areas of CGIAR investment (activity 3.1); sets of short-term micro-scale impact studies using experimental and quasi-experimental methods (activity 3.2) to provide evidence on the impact of CGIAR research-derived technologies to adopting households; studies of a number of under-evaluated areas of research (e.g.

[1] One possibility envisages Centers and CRPs being invited once a year to submit good examples showing how their research has influenced policy. These would take the form of short case-studies according to a standard format to be developed by SPIA in consultation with the Centers/CRPs. For example, this could comprise a description of the nature of the policy (e.g., a policy was changed or a negative policy change was averted or a new mechanism was set-up for investments), a description of the theory for how the research influenced this policy (e.g., the impact pathway), and description of all the possible sources of evidence that document this influence. SPIA would appoint an independent evaluation individual or team (having expertise in this area) to evaluate the credibility of the case-study reports, giving each a rating in terms of the significance (scale, poverty of the people affected, etc.) and attribution established (credibility of the theory suggested / comprehensiveness of the evidence provided). Highly rigorous evidence of influence is not expected at this stage (more appropriate under SIAC Objective 3 – Impacts) but some level of minimum evidence to establish either credibility would be expected (and needs to be defined). All case-studies above a certain quality rating threshold will be added in the POR Outcomes Database.

irrigation and water management; livestock and impact types (activity 3.3); a system-level meta-analysis of *ex post* IA of CGIAR research (activity 3.4).

Progress since March, 2014:

The call for proposals, issued July 2013, on impacts of agricultural research on nutrition and health (activity 3.0) entailed two competitive stages of proposals with internal and external reviews. Final selection and approval by SIAC Project Steering Committee occurred in April 2014 which led to 5 new studies being funded and led by: Virginia Tech (high iron beans in Rwanda); Columbia University (irrigated horticulture in Senegal); CIMMYT (maize-legume intercropping in Ethiopia and Malawi); ILRI (dairy hubs in Tanzania); and MIT (NERICA in Sierra Leone). A launch workshop was organised by SPIA at Wageningen University in July 2014, attended by 20 participants comprising scientists from the 5 projects as well as representatives from A4NH, London School of Hygiene and Tropical Medicine, SPIA and Wageningen. A short report from the workshop is attached to this activity update (annex 2).

Documenting long-term, large scale impacts (activity 3.1) from CGIAR research remains a challenging area of IA, but donor demand for these assessments remains strong. Recently SPIA issued a call for expressions of interest for proposals to document such cases. The basic idea behind this work is to generate studies that credibly document the impacts of successful CGIAR research adopted at scale and over the long term using best available methods. http://impact.cgiar.org/sites/default/files//pdf/Adoption%20of%20Innovations_Call%20for%20EoI_final.pdf
The call describes the aims and the various types of studies (deliberately left fairly open) that might be relevant here.. SPIA hopes to commission about 3 or 4 large studies over the course of the next two years.

The first stage of the two-stage call for proposals on experimental impact evaluations of CGIAR research (activity 3.2) announced in April 2014 is completed. SPIA invited impact evaluations that addressed one or more of the following themes:

- studies designed to convincingly show the effect of adoption of a CGIAR technology on productivity (downward shift in marginal cost of production or upward shift of production function) – and ideally also on measures related to household welfare. Studies that credibly show the lack of impacts are of equal interest.
- studies that help understand under which conditions adoption of a CGIAR technology leads to welfare impacts (on income /nutrition / good or bad environmental outcomes) – this includes studies that evaluate potential complementary interventions.
- studies that help understand other constraints that limit the effectiveness or sustained adoptability of current CGIAR research, this can include questions related to technology production process itself.
- studies with potential to make a significant and innovative contribution to the economic literature linking agricultural productivity to rural poverty.

21 expressions of interest (EOIs) were received and reviewed in person by Karen Macours and Doug Gollin, and SPIA Secretariat. 8 proponents have been contacted and have submitted full proposals on 10th September, 2014. SPIA intends to make a selection of 3-4 proposals in late October / early November, following independent external review of the full proposals and endorsement by the Project Steering Committee.

As an initial step in our work on under-evaluated areas of CGIAR research (activity 3.3), SPIA commissioned Doug Merrey, independent consultant, to conduct a critical review of the impact assessment work to-date on irrigation and water management research. Merrey has submitted his draft report which is now undergoing external review. This desk study includes IA work done within and outside the CGIAR, with the goal of evaluating how comprehensively and effectively these assessments cover the fields. The review focuses on research since about 1990, In addition to identifying the strengths and limitations of the existing IAs in irrigation and water management research (in terms of scale effects, rigor of causal relationships, or how close the impact indicators of the studies correspond to the System-Level Outcomes of the reformed CGIAR system), the review seeks to identify the major constraints and limitations (e.g., methodological, data-related, resource-related, etc.) which would in turn highlight potential for new work. New initiatives may emphasize targeting intermediate impacts (e.g. estimates of the impact

on water-use efficiency), changes in irrigation management policy or simply adoption of research outputs, rather than ultimate, CGIAR system-level outcomes and impacts.

Workplan for 2015:

3 or 4 studies of long-term or large-scale impact will be competitively commissioned in the first quarter, following the two-stage call for proposals issued in September 2014 (activity 3.1). For the experimental impact evaluations call (activity 3.2), 3 or 4 new studies will be started following the two-stage call and review process and SPIA will organise an inception workshop for these in January 2015. The studies will run for up to 24 months, starting January 2015. The portfolio of 5 studies on nutrition and health impacts (activity 3.0), commissioned in 2014 following a competitive call, will be running throughout the year in 2015, but results will not be available until 2016. The background review of irrigation and water management research impact assessment will lay the groundwork for a subsequent scoping study report to assess the potential for utilizing state of the art approaches and possibly new data for launching a series of IAs of specific improved irrigation/water management interventions or policy actions deriving from CGIAR research. Ultimately, this and other critical reviews of past research IA studies (e.g., for NRM research, livestock management, etc.) and accompanying scoping study reports will form the basis for the SIAC Project Steering Committee recommending to the Fund Council Committee on Evaluation and Impact Assessment some specific areas for further IA work under the SIAC program that has good potential for generating large scale, long term economic, social and environmental impacts from under-evaluated CGIAR research. Activity 3.4 will not start until 2016.

SIAC Objective 4: Support the development of communities of practice for ex-post impact assessment

The CGIAR will benefit from a structured attempt to support the existing capacity and some emerging collaborations on *ex post* impact assessment. Information-sharing and regular interaction are important in enabling the kinds of dialogue that can raise standards of impact assessment in the CGIAR, as well as ensuring that individuals have the skills that they need to be successful in their work. Activities towards this objective include a small grants program (activity 4.1); a targeted program of capacity-building using competitive calls for collaborations with advanced research institutes / universities (activity 4.2); conferences and workshops on impact assessment (activity 4.3); support for independently reviewing and publishing quality ratings of impact assessment studies carried out by CRPs and Centers (activity 4.4); maintenance and enhancement of the impact website (http://impact.cgiar.org) (activity 4.5).

Progress since March, 2014:

The small grants program was launched and 5 small grants were funded, but at a heavy transaction cost due to FAO's administrative processes. SPIA wound this up in February 2014, rolling the funds over into activity 4.2.

In late 2013, following a competitive call for proposals for strengthening IA capacity within the CGIAR (activity 4.2), a sub-grant was issued to Virginia Tech to work with CIP and CIFOR on building their capacity to strategically choose topics for impact assessment and then implement them. Workshops took place in May (CIP) and July (CIFOR) between scientists, senior CRP and Center staff and management, and Virginia Tech; a number of cases have been identified for development as impact studies. Also under this activity, SPIA has been working with ICRISAT to help merge aspects of two proposals received from ICRISAT into a consolidated, coherent single sub-grant which is in the process of being contracted using funds from the Bill and Melinda Gates Foundation that remained unspent at the close of the DIIVA project.

Two SPIA-organized workshops (activity 4.3) were held recently in Minneapolis just ahead of the annual AAEA meetings; both are part of an overall strategy to strengthen the CGIAR impact assessment community of practice: (i) Measuring the poverty impacts of agricultural research; and (ii) Impact assessment focal point meeting. There was an excellent turnout for both events. Brief draft reports of the meetings are attached as annexes 3 (poverty impacts) and 4 (impact assessment focal points meeting). A paper exploring methods and models for documenting the

poverty impacts of agricultural research, drawing on relevant literature, recent case studies and discussions at the poverty impact workshop, is currently under development by the SPIA Chair.

SPIA is in the process of launching its on-line IA study quality review system (activity 4.4) as a key mechanism for ensuring high quality assessments of impact (and hence credibility) by the CGIAR. This is expected to also give Center- and CRP-based economists the leverage they need to argue more effectively for required resources for implementing more impact studies.

A new CGIAR impact website (activity 4.5) was launched on May 13th 2014, and includes new sections like Key References (SPIA recommended IA publications targeted at donors), a blog (open to IAFPs contributions as well), and a map of publications

Workplan in 2015:

Progress is expected to continue on each of these fronts in 2015: Virginia Tech working with CIP and CIFOR in strengthening *ex post* IA activities related to planning/prioritizing, data collection protocols, and identification of appropriate methods and implementation of specific case studies; and University of Illinois engaging in similar activities with ICRISAT; SPIA will be organizing a side-event meeting with the Center/CRP Impact Assessment Focal Points at the 2015 Science Forum meeting of the ISPC; the new website launched in 2014 but this will require frequent updating and maintenance to ensure effective utilization; key areas for further development include: better tagging of publications; developing the publications map; and blogging and tweeting more often.

SIAC Program Management Issues

M & E: Survey of donor demands for impact assessment:

The SPIA Secretariat is conducting a re-survey of its earlier 'Donor Demand for ex post IA' study. This study aims to understand:

- How/whether use of epIAs has changed/evolved over time; current perceptions of CGIAR epIAs
- Are/how some of these changes in use/perceptions relate to actions (e.g. publication of impact briefs) taken by SPIA on the issues identified in 2005
- Additional actions required of SPIA, to feed into SPIA strategy. And use this opportunity to solicit feedback from donors on quality ratings of CGIAR epIAs, methodology and parameters proposed

While the response rate for the IA survey is low, we have received what appears to be thoughtful responses from major donors to the CGIAR. Below is a quick sampling of some of the interesting responses.

- Overall, the purpose of epIAs does not appear to have changed between 2005 and now: use epIAs to illustrate impacts on SLOs and IDOs in order to justify continuation of funding to the CGIAR.
- There is some signaling that Center-commissioned epIAs are not as reliable as SPIA commissioned ones. In this instance, it would be helpful to understand the why (independence, better quality – fewer studies, etc.).
- Increased crop, fish and livestock productivity; and improved dietary intakes are 2 indicators from a list of 11 indicators that were clearly rated H (high) priority by 85% of respondents; increased carbon sequestration was rated M (medium) by 82%; and decreased degradation of soil, water, and air rated H by 72% of respondents.
- There is at least one donor out of 14 (fully completed) respondents who wasn't aware of DIIVA and the adoption database. How could we have communicated this better? What more should we do? Or are these adoption metrics (area – hectares under an improved variety) not what they are looking for when they think of adoption outcomes (x% of farmers in a country have adopted an IV).

SPIA intends to launch a phase 2 component of the Donor Demand for IA information study, building on the results of the donor survey conducted and analyzed in late 2014.

Project Finance and Administrative Support:

In late 2013, SPIA recruited a consultant to help support the legal contracting and financial management of SIAC, who has been working full-time since Jan 2014. The Project Steering Committee has met twice thus far in 2014, and will be meeting again in late September 2014.

Research support for SPIA Chair:

The ISPC Secretariat has recently offered short term contracts to two MS graduates of Oxford University. They are expected to provide research and technical support to the Chair of SPIA under the SIAC program of work. Their specific tasks include:

- Reading and critiquing completed impact studies
- Preparing a literature review on poverty impacts of agricultural research
- Assist in drafting and editing comments and commentaries related to impact assessments
- Managing routine administrative matters related to impact assessment studies
- Drafting /co-authoring blog posts
- Attend the workshop in Minneapolis (26th July 2014) on poverty impacts of agricultural research, take notes, and summarize presentations
- Contribute to SPIA synthesis and presentation to ISPC
- Assist in putting in place a quality rating system for impact assessment studies in CGIAR
- Incorporate comments on draft impact assessment criteria
- Identify and modify (as appropriate) software to use for the quality rating system
- Initiate and manage review process

Annexes:

1 – Report from DNA fingerprinting workshop, Seattle, August 2014

2 – Report from nutrition impacts workshop, Wageningen, July 2014

3 – Report from poverty impacts workshop, Minneapolis, July 2014

4 – Report from impact assessment focal point meeting, July 2014

Annex 1:DNA fingerprinting workshop: Brief report for SIAC Program Steering Committee

Bill and Melinda Gates Foundation, Seattle, 4 - 5 August 2014

Convened by Greg Traxler and Mariana Kim from the Foundation, this two-day workshop of 40 participants (see annex 1) had three objectives:

1) Review findings from Diffusion and Impact of Improved Varieties in Africa (DIIVA) study

2) Share insights from four sets of DNA fingerprinting pilots:

- Rice in South Asia (Rice monitoring system, with IRRI)
- Aflatoxin and maize in Kenya and Tanzania (ACIAR)
- Cassava in Ghana, Beans in Zambia, Maize in Uganda (MSU leading, under SIAC program)
- Wheat and maize in Ethiopia (Gates Foundation working with the Ethiopian Central Statistical Agency)

3) Collaborate and engage in scale-up design and next steps, including brainstorming around applications of internal and external uses of DNA technology with agriculture

Objective 1 was largely met through a lunch-time presentation by Doug Gollin, Arega Alene, Tom Walker and James Stevenson, to a group of around 60 people (40 participants + 20 Foundation staff / visitors). Greg Traxler introduced the DIIVA talk by outlining the Foundation's motivation for funding DIIVA. Varietal adoption is central to how the CGIAR has an impact but data on adoption has been neglected since the late 1990s when Evenson and Gollin last pulled together aggregate adoption estimates for improved varieties. At the time DIIVA was commissioned, the methods proposed were "state of the art", but in the intervening five years, the technology of DNA fingerprinting has come on stream, coinciding and reinforcing an awareness of the uncertainty in the data coming from expert opinion or household survey methods. The audience were very bright, asking the right questions of the panel when it came to discussion – questions about: characterisation and standardisation of data collection; how the same data on lack of varietal turnover can be interpreted in different ways (as a good thing or a bad thing); defining an improved variety, etc.

For **objective 2**, a series of presentations on the individual projects helped with understanding the process and methods used in DNA fingerprinting, even if there were few results to report to date. Those that did have early results, such as the wheat and maize in Ethiopia pilot, showing some worrying findings. 35% of wheat farmers and 29% of maize farmers correctly identified their varieties. 12% of households believed they were growing the wheat variety Kubsa, which is susceptible to stem rust and has been de-listed as a variety in Ethiopia. Actually, the figure is 30%. There are a number of laboratories that can provide this analysis at ever-decreasing unit costs, and the impression is that this has now become a method that can and should be mainstreamed in survey design. If the discrepancies between farmer-reported data and DNA fingerprinting data are similarly large in the pilots that have not yet reported results, then the implication seems to be that DNA fingerprinting is now an essential tool in agricultural survey design.

For **objective 3**, there were useful presentations by Talip Kilic from the LSMS-ISA, from Biratu Yigezu Gutema (CSA) on the constraints inherent in attempting to integrate these approaches in large-scale multi-purpose surveys, and from Steve Rownsley (Dow) on the power and limitations of the analyis. The participants broke into small groups and brainstormed on how further scaling-up of these pilots could be achieved in Ethiopia for wheat, Malawi for cassava, and in other countries, with SPIA looking to play a coordinating role in the next year through John Ilukor (Malawi) and Frederic Kosmowski (Ethiopia) being present in those countries.

Next steps: Pilots to continue, wrap up and report. SPIA to talk with Greg about what role SPIA might play in trying to scale-up DNA fingerprinting, on two fronts – within our current SIAC grant and activities; and planning ahead for beyond 2016 for how this might be institutionalised and made part of regular program monitoring within the CGIAR.

Annex 1 – Participant list

Last Name	First Name	Email Address:	Organization
Ehlers	Jeffrey	jeff.ehlers@gatesfoundation.org	Bill & Melinda Gates Foundation
Atlin	Gary	gary.atlin@gatesfoundation.org	Bill & Melinda Gates Foundation
Traxler	Greg	greg.traxler@gatesfoundation.org	Bill and Melinda Gates Foundation
Kim	Mariana	mariana.kim@gatesfoundation.org	BMGF
Caldwell	Richard	Richard.Caldwell@gatesfoundation.org	BMGF
Shukla	Vipula	vipula.shukla@gatesfoundation.org	BMGF
Harris	Mychal	mychal.harris@avyci.com	BMGF
Fraser	Elvis	Elvis.Fraser@gatesfoundation.org	BMGF
Beyene	Habekiristos	habekiristos@yahoo.com	Central Statistical Agency
Gutema	Biratu	ybiratu@yahoo.com	Central Statistical Agency of Ethiopia
Stevenson	James	james.stevenson@fao.org	SPIA
Debello	Moti Jaleta	m.jaleta@cgiar.org	CIMMYT
Kriticos	Darren	Darren.Kriticos@csiro.au	CSIRO
Kilian	Andrzej	a.kilian@diversityarrays.com	Diversity Arrays Technology Pty Ltd
Rounsley	Steve	steve.rounsley@gmail.com	Dow Agrosciences
Tizale	Chilot Yirga	ctizale@yahoo.com	Ethiopian Institute of Agricultural Research
Bimerew	Dawit Alemu	dawit96@gmail.com	Ethiopian Institute of Agricultural Research
Bayissa	Kefyalew Negisho	kefalg@yahoo.com	Ethiopian Institute of Agricultural Research
Kosmowski	Frédéric	fkosmowski@gmail.com	SPIA
Jones	Jared	jared.jones@gatesfoundation.org	Gates Foundation
Koo	Jawoo	j.koo@cgiar.org	IFPRI
Alene	Arega	A.Alene@cgiar.org	IITA
Kulakow	Peter	p.kulakow@cgiar.org	IITA
Labarta	Ricardo	r.labarta@cgiar.org	International Center for Tropical Agriculture
Mwanga	Robert	r.mwanga@cgiar.org	International Potato Center (CIP)
Singh	Uma S	u.singh@irri.org	International Rice Research Institute
Panda	Architesh	a.panda@irri.org	IRRI
Reyes-Padilla	Byron	reyespad@msu.edu	Michigan State University
Maredia	Mywish	maredia@msu.edu	Michigan State University
Asea	Godfrey	grasea9@gmail.com	National Agricultural Research Organization
Gollin	Douglas	douglas.gollin@qeh.ox.ac.uk	SPIA / Oxford University
Ried	Jeffrey	ried.jeffrey@gmail.com	Self employed
Ilukor	John	john.ilukor@gmail.com	SPIA
Kilic	Talip	tkilic@worldbank.org	The World Bank
Walker	Tom	walkerts@msu.edu	Tom S. Walker Consulting
Agustin	Javier	javier.agusti@plants.ox.ac.uk	University of Oxford
Ambel	Alemayehu	aambel@worldbank.org	World Bank

Annex 2: Assessing the impacts of agricultural research on nutrition and health

Wageningen, 4-5 July 2014, Inception workshop for 5 selected case-studies under SIAC Objective 3

Workshop Report (11 September 2014)

Erwin Bulte (SPIA, WUR), Doug Gollin (SPIA, Oxford), Maggie Gill (ISPC Chair), Alan Dangour (LSHTM), Inge Brouwe (WUR), James Garrett (A4NH), Ram Fishman (GWU), Rachel Glennerster (MIT), Menale Kassie (CIMMYT), Catherine LaRochelle (Virginia Tech), James Rao (ILRI), Isabelle Baltenweck (ILRI), Tim Kelley (SPIA), James Stevenson (SPIA)

1 - Introductory session

Erwin Bulte introduced the rationale for SPIA to fund a series of impact studies under SIAC. The idea is for these studies to improve understanding of the links between agricultural research and improvements in nutrition and health. This work should be complementary to (rather than duplicating) the work of A4NH. The objectives of the workshop are: 1) to help the individual projects start off in the direction most likely to deliver credible evidence; 2) Seek possible synergies across projects regarding measurement of nutrition outcomes in particular.

Maggie Gill gave a brief overview of the role of the ISPC within the CGIAR system, and summarized some outcomes from the 2013 ISPC Science Forum in Bonn, which focused on agriculture, nutrition and health. From a measurement perspective, one of the key findings from that set of meetings was the agreement that measurement across the whole diet is important for generating strong evidence of nutrition impact. Another was an appreciation of the multiple factors that constrain our ability to achieve nutrition outcomes. Two follow-on initiatives were highlighted – a meeting in DC in late September 2014 that follows up on methodological issues for evaluation and a special issue of the journal Food Security featuring papers inspired by the Science Forum.

Tim Kelley and **James Stevenson** gave an overview of SPIA's work and the key aspects of the SIAC work program up to 2016, with a focus on the opportunities for researchers to participate in future calls for proposals, workshops etc.

2 – Focus on the five commissioned studies

Catherine LaRochelle presented the project on assessing impacts of the adoption of high iron bean (HIB) varieties on iron intakes of rural vulnerable populations in Rwanda (Virginia Tech, CIAT, HarvestPlus, Rwanda Agricultural Board). An RCT focusing on boarding schools in Mexico has shown that a feeding program with high-iron beans can improve iron deficiency status after 6 months, but there are a number of unknowns regarding the process of adoption, harvest, cooking, consumption of these beans and the dose-response relationship – how much consumption is needed to see a health benefit? Rwanda has the highest per capita bean consumption in the world, and the best estimate from HarvestPlus is that one third of Rwandan farmers are growing HIB.

A number of methodological issues were raised: Hawthorne effects (changes in behaviour from the process of measurement taking place) in particular related to the recall data on consumption of the beans and the provision of specific bowls and plates with known volume; and the process of matching adopters and non-adopters and the search for a valid instrument to find exogenous differences in adoption. The only major logistical issue noted was a delay in receiving a research permit from the Rwandan government, though this is not expected to delay any aspects of the research other than the process of taking a census of HHs in selected communities. The expected adoption rate of 35% remains unverified until the survey teams can get into the field for the listing exercise.

In discussion, the issue of bioavailability was introduced by **Inge Brouwer** – that a stable isotope study had shown that rates of absorption of iron from HIB are lower even if total iron content is higher so there is no extra benefit from consuming HIB. **Alan Dangour** explained the role of phytate in limiting bioavailability – that it should be as close as possible to zero from a nutrition perspective, but that agronomic trade-offs mean that it is still present in varying levels in different bean varieties. **Rachel Glennerster** noted that the identification strategy (matching, search for an instrument) is precarious. Is there an exogenous shock that might have led to a subset of farmers purchasing more

seed (i.e. something that would represent a valid and convincing instrumental variable)? The other option being considered is a regression discontinuity around altitude given variability in adoption with altitude, though that may not be sufficiently sharp as a cut-off. Finally, there was significant disagreement among the group regarding the relevance of measuring anemia. Anemia is not equivalent to iron deficiency but may actually be the ultimate public health goal that HIB is directed towards. The nutritionists in the group argued that, ideally, the study should measure, in addition to anemia, changes in iron deficiency status, since this is the intended intermediary mechanism leading to reduced anemia. However, if a sub-sample were finger-pricked for anemia testing, the sample was powered for it, and the causal identification strategy was strong, the results would likely be persuasive evidence (to an economist – nutritionists may not be sufficiently impressed) regarding the impact of HIB on anemia.

Isabelle Baltenweck and **James Rao** presented their project on assessing the impacts of dairy hubs in Tanzania (ILRI, Emory, Tanzania NARS). There are many small-scale dairy herds in Tanzania and 97% of milk consumption is raw milk. A dairy hub can be one of two models – either simple bulking of raw milk from multiple producers to sell to traders (no new technologies used) or bulking and chilling the milk. Both types of hubs also provide producer services to a greater or lesser extent; e.g., access to inputs, market information, etc. The Tanzania hubs are all different from one another, reflecting local preferences and practices. The East African Dairy Development (EADD) project has shown that producers in hubs have higher production, and the SIAC project aims to look at the potential for nutrition impacts. The project was chosen by SPIA both because dairying has been proposed as an important channel for nutritional improvements and because the formation of hubs represents the kind of institutional innovation that a number of CGIAR centers have been exploring in recent years.

Methodologically, this project is rather problematic in its current configuration. The study was conceived as a quantitative study monitoring outcomes for 30 villages in the project. However, power calculations have not been carried out and there is a binding constraint on the number of villages that can be included in the study, determined by the resources for the project implementation. From the discussion with the proponents, it was clear that there is an insufficient number of villages in both project and "control" (this is not a randomized design) villages to allow for a with vs without comparison – the study would be under-powered. The economists in the group suggested some changes that would allow for differences in participation within villages, so as to gain some variation in (at least the intensity of) treatment at the household or individual level. But this seems not to be possible. As a result, the methodological problem is not easily resolved – there is no way, apparently, of scaling up the intervention to a greater number of villages. The consensus of the group in the end seemed to be that it was better for the project to be largely qualitative and with detailed analysis of hub villages. The qualitative evidence can be compared to the proposed theory of change, i.e., that participation in hubs leads to increased milk sales, increased income and increased milk consumption for the household which, although not measured in this study, would often lead to improved nutrition.

Ram Fishman presented the project on assessing the impacts of new irrigated horticulture schemes in Senegal (Columbia University, GWU, MDG center Dakar). This study is a randomized control trial taking place in 70 villages chosen randomly to participate in the project. These would be compared to 70 control villages. The treatment is the installation of drip irrigation systems on under-utilised, commonly owned land in the villages. Eligible households (50 HHs per scheme, selected through a lottery) in the project villages receive a small plot of land and the benefits from the commonly-owned and commonly-managed irrigation scheme.

Methodologically, the identification issues are taken care of by the randomization, so the discussion focused on the nutrition measures and the rationale for their use. The minimum adequate diet and a measure of dietary diversity were the two recommended mainstream indicators that were recommended for use in the project. The expected impact mechanism is an increase in consumption of a range of fruits and vegetables, either from own production or via increased incomes. Women's empowerment and time use (the scheme focuses on women and there could be trade-offs over their time use) are also the subject of measurement challenges for this project, though there is emerging good practice in these areas. The only major weakness with this study is its dependence on the project intervention, which is a three-way cooperation between the governments of Italy, Israel and Senegal and already

subject to significant delays in planned implementation. Some skepticism was also expressed concerning the strength and robustness of the causal pathway from production to consumption, but this is in some sense one of the main questions of the study.

Menale Kassie presented the project on the impact of the integration of legumes into maize systems, through intercropping and rotations in Malawi and Ethiopia, on nutrition (CIMMYT, Lilongwe University, EIAR). This study is based on two rounds of data collected by CIMMYT in 2010-11 and 2013-14 from a large sample of households in both countries. A follow-up survey in 2015 focuses on nutrition outcomes, measured through recall data on food expenditures and on dietary diversity scores.

Methodologically, this study attempts to identify causal effects from legume integration by applying a variety of econometric techniques (multinomial endogenous switching regression; instrumental variables; difference in differences) and examining the sensitivity of the results to changes in these estimation strategies. Regardless of the method, there will always be some set of assumptions that needs to be made in order to identify an effect from observational data such as this. The fact that the nutrition data has not yet been collected gives an opportunity for the economics-focused team to reach out and find good advice from the right people on how best to do this.

Rachel Glennerster presented the project on the impact of NERICA rice in Sierra Leone on potentially shortening the hungry season in rice-dependent areas by bringing forward the date of the first harvest (MIT, Innovations for Poverty Action, International Rescue Committee, Sierra Leone Agricultural Research Institute). In the study area, every August of each year, 80% of the sample households regularly skip meals due to the high cost of food in the hungry season. The study is an RCT with NERICA seed offered at three prices (full-price, half-price, or free), with and without agronomic training. Early results have shown that there is a yield penalty for adopters who do not receive training (compared to local varieties). This reflects the difficulty of growing NERICA and the need for training. But early indications suggest that there may be significant positive impacts on nutrition that are related to the early maturation of the NERICA rice. The nutritionists at the workshop expressed skepticism about the mechanism and about the likelihood of finding nutrition impacts related to early duration; but the large size of the study and the randomization seem to provide strong *prima facie* evidence for some effect. Workshop participants agreed on the need to explore this further and to understand better the channels of causation.

Some concern was expressed about the proposed approach to measuring nutrition impacts. Body Mass Index is arguably not appropriate for a population of under-5s; but there was some lack of clarity over whether it is simply a noisy measure or an inaccurate one. The nutritionists in the group argued that weight for height would be the best measure to use in this context, and this advice was taken on board. Dietary diversity can only be calculated at the household level for this study, even though it was recognized that individual-level dietary diversity is strongly preferable.

3 – Open discussion

Alan Dangour made a number of general observations. Regarding the design of the individual projects, there is often not a very good fit between the original design of the studies and the desire to measure nutrition impacts as many of the studies have not been designed for that purpose. Per Pinstrup Andersen recently noted that there are now too many RCTs but the truth is that it is hard to get the right answers from other designs. DFID – BMGF now have put out call for large trials (3 – 5 years, $3 million per study) for intervention studies and this is an opportunity for a much more concerted effort to document nutrition impacts.

The study teams need to think very carefully about the pathways they are attempting to document - what will the intervention do and what can the study actually demonstrate? If, for example in the case of the dairy hubs, the rationale is to change the availability of milk then if that's what it will do, measure milk availability, rather than attempt to estimate causal effects on income or dietary diversity. Validating even small incremental steps along the hypothesized impact pathway is still very useful. Inge Brouwer and Alan were the only nutritionists in the room and it was felt strongly that the studies needed to bring in a nutrition perspective. This was intended as an observation,

not a criticism. This was supported by **James Garrett** who emphasized the need to distinguish between the different possible impact pathways as the intermediate impact indicators will likely vary in each case, e.g., if under-2 year children's nutrition is targeted then relevant indicators are mother's time use, food availability and household income.

Inge Brouwer encouraged SPIA to try and standardize the measures across these projects. Can we focus on diet diversity for women and minimum adequacy for children of particular ages? This would allow us to address intra-household allocation issues. However, **Rachel Glennerster** re-emphasized the point that not all studies have consumption data per person. This is not just an issue of data availability from surveys – in some contexts such as in West Africa where the family eating from a single plate is common, it can be hard to estimate individual consumption.

There was considerable discussion regarding the measurement of women's empowerment and the use of the WEAI (Women's Empowerment in Agriculture Index) in particular. The full version of that survey requires 3 hours of data collection and covers time use, decision-making and the feeling of agency, etc. On the other hand, the Oxford Poverty and Human Development Initiative (OPHI) modules for specific aspects could be mainstreamed into the projects relatively easily. In planning ahead, the following timelines were outlined as being the most likely scenarios:

	CIMMYT	**CIAT**	**ILRI**	**Columbia**	**IPA**
Data all in	May-15	May-15	July-15	Dec-16	Jul-14
Analysis		Sep-15	Dec-15	2017	Jul-15
Final report		Dec-15	Mar-16	2017+	Oct-15

In wrapping up, **Menale Kassie** noted that this workshop was a bit special in that having such a discussion before the project helps improve the quality of the work. **Isabel Baltenweck** found it useful to have shared ideas regarding the agreed indicators and approaches with the expectation of further communication to come. Discussion of the design issues have been helpful. **Ram Fishman** stated that the workshop had been helpful and motivating.

Next step:

SPIA to circulate this note to workshop participants, requesting short response from each project about how they will respond to the main discussion points highlighted in section 2 here.

Annex 3: Measuring the poverty impacts of agricultural research

Workshop organized by SPIA, University of Minnesota, 26 July 2014

Overview and main messages from the final session

Coming the day before the start of the annual American Applied Economics Association (AAEA) meetings, this one-day workshop attracted over 60 participants, mostly by invitation from SPIA. The meeting was attended by SIAC Program Steering Committee (PSC) members Greg Traxler and Julian Alston. The agenda was designed in such a way that we would hear briefly (10 mins per presentation) from a small group of presenters (case study leaders of SPIA-commissioned studies + other relevant papers) that have used a particular approach to assessing poverty impacts. In each session this was followed by a group of discussants that were briefed with the task of critiquing the role of the broad family of methods for the purpose of assessing poverty impacts from new technologies, rather than dwelling too much on the details of the individual papers.

This format worked well but required some steering from the Chair, overseen by Doug Gollin. The take-home messages from the first four sessions (on observational methods, micro-models, RCTs, and macro-models) was that it was a very hard task to find strong evidence with a single study that will cover the long causal chain from new technology to long-run poverty impacts. The take-home message from the final session 5 was, that despite the challenges, producing this evidence should be a priority for the CGIAR. Karen Brooks, Jim Oehmke and Greg Traxler all highlighted the importance of poverty outcomes for their institutions, and Peter Hazell and Bob Herdt offered some optimism that these significant challenges may be overcome. Nonetheless, SPIA remains concerned about a mismatch between the consensus of the group of academics gathered in the first four sessions, and the expectations outlined in session 5. Where should we draw the line on rigour for impact assessments in the context of balancing the standards for top-level academic research with the pressing concerns for evidence from governments and donors to help understand the effectiveness of their investments in agricultural research.

With this tension in mind, SPIA is currently working on a paper that will summarise and critique the literature on agricultural research and poverty, as well as outline its perspectives for good practice in this area. This is being written by Doug Gollin with research assistance from Lilli Probst and a draft will be delivered by 8th September, and presented and discussed at the forthcoming ISPC meeting in Copenhagen, on 15th September.

Introductory Session

Peter Gardiner, welcomed participants to the workshop and gave some context to the need for it. How can the CGIAR articulate what we are doing in a convincing manner? Doug Gollin, explained that this workshop is primarily about methods, and that SPIA is open-minded and ecumenical across the spectrum of possible identification strategies. Each method has its own boundaries and limitations, and SPIA believes that it is important to uphold standards of rigour and to highlight what is and is not possible. Year by year monitoring of poverty changes from agricultural research has been suggested and SPIA is concerned about these expectations.

Session 1 – Observational micro studies

Melinda Smale, Jeff Alwang and Sosina Bezu presented their recent SPIA-commissioned studies that all employ variants on observational methods – instrumental variables, fixed effects methods, correlation random effects and control function approaches. Marc Bellemare cautioned that some of the “fancy” estimators used may actually give identification from functional form – that the identification may not really be there. In addition, testing for exogeneity of your instruments is not reliably achieved using the Hausmann test, for example. Where panel data exists, why not simply use difference in differences? The assumption of parallel trends is no less credible than some of the other identifying assumptions that the other estimators require. And in cases where there are more than two time period panel data to draw on, the case for difference in difference is stronger yet.

Michael Carter noted that poverty lines are arbitrary and not easily defended – in many cases the impacts for the poorest people may be negligible whereas those that are just below the poverty may benefit just enough to put them over the poverty line. In general, greater care is needed in handling heterogeneous effects along the full income distribution. The issue of continuous vs binary treatment was also discussed, as was the possibility of taking things out of the error term through methods such as generalisable propensity score matching. Can attributes such as "eagerness", or "entrepenurial-ship", which might otherwise be in the error term, be made observable, given its likely importance to the process of adoption? Andrew Dillon also emphasized that the search for a clear source of exogenous variation, a targeting rule that can be pulled out of the data, is preferable to allow for a regression discontinuity design to be implemented. He supported the idea that certain confounding factors, that would otherwise languish in the error term and potentially bias the estimations, can be mopped up by measuring "unobservables" such as "eagerness". Finally, it is important to understand whether the process being studied (e.g. the "choice" of variety) is an active choice from a range of possible options for the farmer, or is taking place somewhat by default (such as when farmers only have access to a single variety).–

In disscussion, Jeff Alwang wondered what the right amount of money is that the CGIAR should be spending on surveys and how this might be coordinated. Doug Gollin noted that SPIA's rough approximation is that 100 – 200,000 households are visited every year by CGIAR researchers, and so the benefits from greater coordination are potentially quite considerable. Tavneet Suri supported the calls on the part of the discussants to move back to linear models. Non-linear models are all biased in agricultural applications as errors are not independent and identically distributed (i.i.d). The control function approach, for example, will identify from functional form in the absence of an exogenous instrument. Michael Carter and others made a call for a greater focus on theory around the process of adoption and disadoption.

Session 2 – Micro-models

Stein Holden, David Raitzer and John Antle gave short presentations on their respective models, all different in formulation and applied in a diverse range of contexts from maize in Africa to rice in SE Asia to aquaculture in Bangladesh. Tim Dalton opened discussion by noting that mathematical programming represents a broad range of tools that are important across the research spectrum. In testing a hypothesis using models that are not inherently statistical in nature, there is a good deal of judgment required. There are different ways of achieving a degree of peer review over this process such as the example of the community of researchers using the Global Trade Analysis Project (GTAP) model, which for over 20 years have been sharing data and model extensions with each other. There can be a trade-off between flexibility of the model and the ability to characterise the context adequately. It is desirable if the model can be constructed in such a way that behaviour emerges rather than being imposed by the model, and the solution space has to be plausible.

Julian Alston asked the question "What should we try and measure most accurately?". The yield gap in agriculture is getting bigger, there are market failures in research, and importantly, agricultural research may not be most appropriate instrument for achieving the targets that have been set for it. Cheryl Doss emphasized that the models used to assess the poverty impacts of agricultural research should focus closely on the theory of change articulated for the research programs in question: "Why do we think this research will impact on poverty? How?" Models depend on data and they are only as good as the data we put into them. There are a number of measurement issues – relating to technology, adoption and poverty outcomes. Perhaps most importantly, there are serious concerns about heterogeneity of impact – that the elasticity varies for different kinds of household. For some questions this may not matter too much in the aggregate, but for poverty it is very important. Most fundamentally, if the agricultural research succeeded, what would the world look like? Land ownership patterns are likely to be part of this story – bigger farms and out-migration?

In discussion, Stein Holden noted that the model assumptions are becoming more flexible, away from strict rationality to include features such as procrastination and prospect theory. Doug Gollin asked the question of validation of different models using back-casting exercises – which of these models and for what contexts? John Antle noted that the Agricultural Model Inter-comparison Project (AgMIP) is trying to do these kinds of cross-validation and, importantly, simulates outcomes taking a more complete system's perspective and not looking at a single farm-level technology in isolation. Both Jerry Nelson and David Raitzer noted that poverty impacts could mean two things: poverty as the dependent variable, or focusing on poor people as the population of interest, zeroing in on what's happening to them. Both are important but can be different from another.

Session 3 – Randomized Control Trials (RCTs)

Michael Carter, Tavneet Suri and Kyle Emerick all presented work in progress using RCTs to assess the impacts of agricultural research. In general, the outcomes being studied in these experiments are short-run in nature and theory suggests that they correlate well with poverty. Mywish Maredia delineated the kinds of questions that can and cannot addressed using RCTs. For example, indirect effects over long time horizons are not well suited to this family of approaches. In the sense that RCTs are excellent at examining the on-farm constraints and complements to technology adoption, they can properly be viewed as research tools rather than tools for evaluation. Mary Arends-Kuenning rehearsed the arguments from the de Janvry review on RCTs that SPIA commissioned in 2011 – that experiments should be randomised at the community level and that the technologies should be offered at market price, not subsidised. The process of deciding on which RCTs to do and what doesn't get studied seems to be somewhat arbitrary. There is growing interest in studying the mechanisms through which treatment effects are studied – the Emerick paper is particularly convincing on mechanisms. Meena Meenakshi stated that she felt it was a good thing that the agricultural community have embraced RCTs – there has been explosion of them across multiple interventions and mediating factors. More effort on studying packages of interventions, picking up externalities, and examining cost-effectiveness would add to their value to agricultural research.

In discussion, Tavneet Suri expressed her opinion that the incentives are changing for researchers and donors – that more scale-ups are being funded based on the findings from RCTs. To learn effectively from RCTs, they need to be large scale and designed so as to pick up heterogeneous effects across different regions and target groups. In most cases, RCTs may be able to measure productivity changes in the short run, but unlikely to detect significant changes in poverty levels (Carter). On externalities, the RCT structure gives you the flexibility to go back and look at externalities if there are concerns. However, Kyle Emerick argued that looking beyond the producer is important at the stage at which the RCT is designed, in order to pick up wage effects etc. Alessandra Garbero thought that greater transparency around pre-analysis plans for effects that emerge later would be step in the direction of greater rigour. Similarly, greater emphasis on heterogeneous treatment effects would be appropriate. There was some agreement that the most useful role for RCTs may be in rigorously estimating parameters for feeding into macro models, particularly related to productivity effects.

Session 4 – Macro models

Xinshen Diao, Phil Pardey and Will Martin all gave short insights into their research in this broad field that cut across country-level models (Diao), spatially-explicit modelling (Pardey) and global macro models (Martin). Terry Roe noted that the share of workers in agriculture decreases with growth over time, and as incomes grow, the share of agriculture decreases. There are strong indirect effects via the rest of the economy. We need studies that tell us the story behind this process – what are the causes of the macro processes? George Norton emphasized the links between the micro and the macro perspectives. Ideally, one needs a team to analyse these things properly. Jerry Nelson asked about the CGIAR reform process and the lack of ex-ante analysis that informs these major changes of change currently underway. The comparative advantage of the CGIAR may shift and donors say that this should happen. But important questions – should the CGIAR do systems research? What is the role of gender research in

the CGIAR? How does biophysical research interact with economics? – are currently poorly supported by spatially-explicit modelling.

Doug Gollin steered the conversation in the direction of specific methods by asking when does spatial analysis matter? When are there important interactions between farmer decisions with spatial and temporal variations? There was wide agreement that it was necessary to work together to get better data and share it effectively. The results may not be consistent across micro, meso, and macro models – there can be different magnitudes or signs, and these should explored in collaborative groups. In that context, Phil Pardey pointed out the changing nature of the market for "innovation" – agricultural R&D representing about 5% of the global market of approximately $1.2 trillion USD annually.

Session 5 - "Reducing Rural Poverty" as a System-Level Outcome for the CGIAR system

This session focused on a major question: should the CGIAR retain rural poverty reduction as an SLO? Karen Brooks started by making the case for keeping poverty reduction as a target for the CGIAR. We should step back from it if we felt that either a) the CGIAR research does not impact on poverty or b) we can't measure it. However, she feels that neither of these apply and it would send a strong negative signal to donors if we scrapped it. The well understood and often-claimed impact pathway story (ag research → new innovations adopted → productivity improvements → poverty reduction) has changed and that story needs to be re-stated and revalidated, since it is playing our differentially now. Thus,the stylized facts of direct effects, food prices and employment and wage effects need an update. Globalisation, with more open economies and the cost of trade and communication decreasing has meant that many are sceptical of a poverty-reducing role for agriculture. However, we need better geographic differentiation in updating the Strategy and Results Framework (SRF) of the CGIAR.

Peter Hazell agreed in retaining poverty reduction as an SLO. He reminded the group that much of the discussion and debate about evidence of poverty changes from agricultureal R&D already played out in the 70s and 80s. Had some of the 'early lessons' from the GR been acted upon (e.g., large farmers benefited the most), some of the biggest impacts on the poor from the GR would have been missed, because small holders eventually did adopt and many indirect effects (price reductions of major staples) accrued disproportionately to the urban poor. Hazell also made the case for country-level CGIAR impacts, such as the Diao study, but in a manner that better integrates the micro- and the macro- perspectives. There are big challenges in using poverty (#s of poor) as a dependent variable in our micro models, especially when looking at the effects of adoption of a single technological innovation, given the typically small share of income contributed by a single crop and, typically, the small percentage change in yield/profits through adoption of a new variety. Thus, studying a single intervention may not show any significiant impacts, whereas measuring the direct and indirect effect of a composite of CGIAR related interventions taken together, in a given country, may well show significant results. Also, these things take time – the dynamics of change are important and the time lags can be significant. However, it is important to distinguish attribution of impact by the CGIAR from contribution by the CGIAR – the NARS and others play important roles. This would ideally be done with cooperation across CRPs and centers. The danger is that we get very rigorous using reduced form approaches but miss information about the pathways inherent to the more structural models used in the past.

Greg Traxler outlined how the Bill and Melinda Gates Foundation broadly subscribes to the "poor but efficient" model of a smallholder. Raj Shah in 2006 decided to direct a significant portion of the $31 billion endowment from Warren Buffett towards a new agricultural development unit. However, agriculture is competing within the Foundation against global health programs (50%) and US programs (25%). The agriculture research community has to convince Bill and Melinda regarding the distribution of benefits, differential effects, gender and poverty impacts and the number of people lifted out of poverty is an important metric in this argument. Thus there

continues to be very strong demand for poverty outcome indicators – it's required to keep the funding going (from BMGF); headcount of the poor is important, along with other outcome indicators that rely on a strong theory of change. Adoption of technologies is central to the argument for funding research – we need better numbers on this as well. As a community of researchers we need to ensure that RCTs are used to address the most important and relevant questions rather than being supply-driven. On the other hand, many of the macro models are difficult to explain and describe to non-specialists. Perhaps the piece of the puzzle that is missing most glaringly is the kind of institutional analysis that Vernon Ruttan made famous.

Jim Oehmke read out an official statement on behalf of USAID's Feed The Future (FTF) initiative which has the goal of ending poverty and hunger by 2025. FTF believe that it can and has had significant impact on reducing poverty, and has reported to the US Congress on the reducing numbers of poor people in countries where USAID works. The program will be renewed for a 2nd 5-year phase. FTF needs to know the numbers of poor people but is not wedded to a firm attribution to specific interventions within the overall portfolio of sustainable intensification and agricultural transformation. We need to shift from a focus on comparative statics to comparative dynamics to enable analysis of climate smart agriculture, resilience and related outcomes. All of these perspectives are within the context of CAADP and the mutual accountability that exists for the commitments therein. Oehmke noted that outcome measures such as gross economic surplus, or consumer/producer surplus are no longer of such interest to donors; they want to know about poverty reduction.

Bob Herdt sketched out a model for how the multiple pathways from technology adoption to poverty effects intersect and interact in specific contexts. Economic transformation is a process of accelerated change with agriculture becoming less and less important, within which significant changes take place in the demand for food. Perhaps it might be instructive to also take a different approach to those outlined in the previous sessions and start with poor people and work back to look at the role that technologies play in their lives? He argued for a better understaning of the poor and their livelihoods to put into perspective how agriculture can contribute to poverty reduction.

In discussion, Julian Alston pushed back on the idea of national accounting for CGIAR impacts – that it would represent a fallacy of composition. How to represent what is happening in the rest of the world at the same time? Doug Gollin noted that this is an example of a wider issue in international development policy – there might be other things that reduce poverty more than investments in agricultural research. Tom Tomich reflected that the concept of poverty used in the workshop was narrower than he might have expected – that asset control and capabilities had received scant attention, similarly the role of agricultural research as a human enterprise. Michael Carter endorsed this perspective of poor people as agents rather than failures and that this has emerged as an issue in the second generation of cash transfer programs.

Ruben Echeverria noted that we struggle to make good ex-ante predictions, but do we need the word "rural" in the wording of the System-Level Outcome (SLO)? Tavneet Suri argued that you can't reliably compare across instruments for reducing poverty – that there aren't datasets available that would allow you to do such a thing. Karen Brooks stated that she is happy to be identified as an advocate for agricultural research – that we have endured 17 years of divestment from agriculture which contributed to the food crisis. Peter Hazell returned to the theme of cross-CRP and center tracking of progress in perhaps 8 – 10 countries with the most significant CGIAR investments. Greg Traxler encouraged SPIA to keep pushing to measure poverty impacts, or the space will be "claimed" by the M&E folks and the work will be less rigorous. Jim Oehmke urged the CGIAR to make the case for those aspects of its portfolio that will make the most difference show how it fits in a broader development agenda. Bob Herdt's take-home message was that we should take seriously the portfolios of the poor and try and understand the lives of poor people .

Anita Regmi and Peter Gardiner offered closing remarks on the importance of the workshop in the context of the CGIAR reform process, on behalf of the Consortium and ISPC, and Doug Gollin thanked everyone for their active participation, highlighting that SPIA will draft a paper that reflects on the workshop and circulate it for comments in late September.

Annex 1 - Agenda

08.30 Welcome and Introductory Remarks (Doug Gollin, Peter Gardiner)

08.40 - 10.10 Session 1 (Paul Glewwe, chair) What can we learn from observational micro studies?

- Lessons from SPIA-commissioned studies (Melinda Smale, Jeff Alwang, Sosina Bezu; 10 mins each)
- Methodological challenges, identification problems, data needs and survey design (Marc Bellemare, Michael Carter, Andrew Dillon; 10 mins each)
- Open discussion (30 mins)

10.25 - 11.55 Session 2 (C. Ford Runge, chair) What can we learn from model-based approaches drawing on micro data?

- Lessons from SPIA-commissoned studies (Stein Holden, David Raitzer, John Antle; 10 mins each)
- Methodological and data challenges; model validation; concerns with heterogeneity (Tim Dalton, Julian Alston, Cheryl Doss; 10 mins each)
- Open discussion (30 mins)

11.55 - 13.05 Session 3 (Alessandra Garbero, chair) What can we learn from RCTs?

- Lessons from SPIA experience and discussions (Michael Carter, Tavneet Suri, Kyle Emerick; 10 mins each)
- Cost effectiveness; applicability to large-scale and long-term impacts; challenges (Mywish Maredia, Mary Arends-Kuenning, J. V. Meenakshi; 5 mins each)
- Open discussion (25 mins)

13.45 - 15.00 Session 4 (Shenggen Fan, chair) What can we learn from meso, macro and cross-country studies?

- Lessons from SPIA-commissioned studies and related research on the CGIAR (Xinshen Diao, Phil Pardey, Will Martin; 10 mins each)
- Issues of data quality and model quality, sensitivity to assumptions, identification, validation (Terry Roe, Jerry Nelson, George Norton; 5 mins each)
- Open discussion (30 mins)

15.15 - 16.45 Session 5 (Doug Gollin, chair) "Reducing Rural Poverty" as a System-Level Outcome for the CGIAR system

- What have we learned about documenting and measuring impact? Is impact assessment feasible? What can we hope to find? (Panel discussion: Karen Brooks, Peter Hazell, Greg Traxler, Jim Oehmke, Bob Herdt; 10 mins each)
- Open discussion (30 mins)
- Responses from panelists (2 mins each)

16.45 - 17.00 Closing remarks and wrap-up (Doug Gollin, Anita Regmi, Peter Gardiner)

17.00 - 18.00 Reception organised jointly with the University of Minnesota, Department of Applied Economics and Agricultural Education

Annex 2 – Participants

1	Tahirou Abdoulaye	IITA
2	Julian Alston	UC Davis
3	Jeff Alwang	Virginia Tech
4	Robert Andrade	Univ. of Minnesota
5	John Antle	Oregon State University
6	Mary Arends-Kuenning	University of Illinois
7	Aminou Arouna	Africa Rice
8	Aden A. Aw-Hassan	ICARDA
9	Katherine Baylis	University of Illinois
10	Marc Bellemare	Univ. of Minnesota
11	Sosina Bezu	Norwegian Univ. of Life Sciences
12	Karen Brooks	IFPRI
13	Marie-Charlotte Buisson	IWMI
14	Michael Carter	UC Davis
15	Tim Dalton	Kansas State University
16	Xinshen Diao	IFPRI
17	Andrew Dillon	Michigan State University
18	Cheryl Doss	Yale University
19	Ruben Echeverría	CIAT
20	Kyle Emerick	Tufts University
21	Shenggen Fan	IFPRI
22	Monica Fisher	CIMMYT
23	Steve Franzel	ICRAF
24	Alessandra Garbero	IFAD
25	Peter Gardiner	CGIAR ISPC Secretariat
26	Paul Glewwe	Univ. of Minnesota
27	Doug Gollin	University of Oxford
28	Guy Hareau	CIP
29	Peter Hazell	IFPRI
30	Bob Herdt	Independent Consultant
31	Stein Holden	Norwegian Univ. of Life Sciences
32	Meenakshi J. V	University of Delhi
33	Tim Kelley	CGIAR ISPC Secretariat
34	Lakshmi Krishnan	CGIAR ISPC Secretariat
35	Catherine La Rochelle	Virginia Tech
36	Ricardo Labarta	CIAT
37	Lucy Lapar	ILRI
38	Ravinder Malik	IWMI
39	Mywish Maredia	Michigan State University
40	Will Martin	World Bank
41	Sam Mohanty	IRRI (IAFP)
42	Jupiter Ndjeunga	ICRISAT
43	Jerry Nelson	Independent Consultant
44	George Norton	Virginia Tech
45	James Oehmke	USAID
46	Phil Pardey	Univ. of Minnesota
47	Frank Place	ICRAF
48	Lilli Probst	University of Oxford
49	David Raitzer (Skype)	Asian Development Bank
50	Anita Regmi	CGIAR Consortium
51	Byron Reyes	Michigan State University
52	Terry Roe	Univ. of Minnesota
53	Ford Runge	Univ. of Minnesota
54	Melinda Smale	Michigan State University
55	James Stevenson	CGIAR ISPC Secretariat
56	Tavneet Suri	MIT
57	Daniel Suryadarma	CIFOR
58	Tom Tomich	UC Davis
59	Greg Traxler	BMGF
60	Roberto Valdivia	Oregon State University
61	Ira Vater	CGIAR ISPC Secretariat
62	Di Zeng	Virginia Tech

Annex 4: Meeting of the CGIAR Impact Assessment Focal Points (IAFPs) & the Standing Panel on Impact Assessment (SPIA) of the CGIAR Independent Science and Partnership Council
Meeting at the Hilton Marquette, Minneapolis, USA, 25 July 2014

The *Impact Assessment Focal Point Meeting* was jointly organized by SPIA and the IAFPs of the CGIAR Research Programs (CRPs) and Centers, in the spirit of similar meetings held in the past. The objectives of the meeting, similar to previous meetings, were to:

(1) Enhance communication/interactions between the Centers/CRP IAFPs on *ex post* IA related issues
(2) Give and receive feedback on Center/CRP *ex post* IA study plans and strategies
(3) Update IAFPs on selected SIAC activities and elicit feedback
(4) Discuss areas of complementarities and overlap between SIAC work and overall M&E and IA work at CRPs and Centers

The program agenda, list of participants and individual presentations can be found here on the CGIAR Impact site http://impact.cgiar.org/meetings-and-events/iafp-usa

INTRODUCTIONS

The SPIA Chair Doug Gollin welcomed the participants to the 2014 IAFP-SPIA meeting, and the Secretariat, particularly Ira Vater, for helping organize the meeting along with IAFPs. In attendance were IAFPs representing 12 Centers and 14 CRPs; a representative each from BMGF and the Consortium Office; representatives from MSU and Virginia Tech (Activity Leaders in the SIAC program); a SIAC Project Steering Committee (PSC) member; the SPIA Chair and his Research Associate , two SPIA members, and four ISPC/SPIA Secretariat staff.

The sessions were organized around CRP themes (commodity improvement focus, NRM and the environment improvement focus, systems analysis and improvement focus, policy oriented focus), followed by updates on SPIA's current work and reflections from the donor agency. The proceedings that follow merely highlight the presentation, and summarize discussions following each presentation and during the general discussion at the end of each session.

SESSION 1: Crop Improvement Focus
1.1 CRP Wheat and CRP Maize: Presentation by Monica Fisher

- Monica Fisher presented metadata (study type, impacts assessed, data used, geographical focus etc.) on Impact Assessment (IA) studies done at Wheat and Maize CRPs between 2010 and (June) 2014.
- An overwhelming majority of such studies is *adoption studies* – only 3 epIAs were done. These studies use a variety of empirical approaches in documenting adoption and impact, including RCTs like that of CASFESA and NuME.
- They would like to develop a core set of adoption and impact indicators in the future, and are deeply interested in collecting panel data (SIMLESA, DTMA) as well as making use of secondary datasets (e.g. LSMS). For the latter, they are aware of discussions to add (maize) variety related questions to the LSMS questionnaires.

Doug Gollin observed that a lot of the studies are based on observational/cross-sectional data around a project. Such studies pick places where adoption (high levels) has already occurred or has high potential for adoption – in that sense, by necessity, this is opportunistic data collection. Fisher agreed that while some of the projects are restricted to the population where intervention has occurred, they are attempting to collect data that is more generalizable. For instance, in case of DTMA, one of the (two) surveys is across 13 countries where adoption is occurring (they ask breeders and seed companies where dissemination has occurred, and this constitutes the sampling frame).

This year, however, they have selected 3 DTMA countries and will be able to generalize to districts where maize is a very important crop. Greg Traxler (BMGF) wanted to understand how many surveys CIMMYT was doing this year, and how many farmers were involved. Fisher thought this would be around 20, and for the projects she is working on in Uganda and Iran, the surveys target 800 households each. Across the CGIAR, James Stevenson (ISPC Secretariat), observed that is likely to be around 100-200K households.

1.2 CRP Roots, Tubers and Bananas for Food Security and Income: Presentation by Guy Hareau

- Guy Hareau highlighted RTB's six step priority setting assessment exercise done at the request of the Fund Council and ISPC to align RTB with SLOs and IDOs.
- He also spoke about their impact assessment strategy for 2014-15, and highlighted some specific IA studies at each RTB partner Center (Bioversity, CIAT, CIP, and IITA).

A general question on the extensive priority setting exercise (that included survey of 1,681 experts across 5 RTB crops and 4 regions) was how the demand from final beneficiaries was assessed. Hareau noted that they had not, and that this was a downside of the process. Daniel Suryadarma (CIFOR) asked how they link *ex ante* assessments to *ex post* assessments, and Hareau responded that they were attempting to do this: while RTB has a planning horizon of 2 years, the *ex ante* assessments (slide 12) referenced in the presentation look 6 years into the future. Hareau also mentioned that they were milking the DIIVA dataset to conduct additional studies on determinants and constraints to adoption in some countries of Sub-Saharan Africa (SSA).

1.3 CRP GRiSP: Presentation by Sam Mohanty, Arouna Aminou, and Ricardo Labarta

- Sam Mohanty presented a highlight of GRiSP IA studies : 16 adoption studies/IA related to germplasm enhancement, 20 related to management enhancement, and 4 on both are ongoing or have been completed in the last 5 years.
- Mohanty pointed out that they try not to have IRRI researcher as the lead researcher in the study - that this increases the credibility of adoption studies and IAs. It is not necessarily the case that IRRI-led studies overstates the outcomes/impact – for instance, the ACIAR-funded study (slide 4) estimates higher total benefits and higher average annual benefits for Indonesia and Philippines from varietal development and releases than an IRRI study (even as management showcases the ACIAR study).
- Aminou highlighted AfricaRice and GRiSP work in 41 hubs (23 countries), and noted how M&E and impact assessment are both being done in these hubs.

Bob Herdt (SPIA Member) raised a question on the type of intervention/information provided in the Philippines rice crop manager (smartphone app) IA. Mohanty responded that this was primarily information to farmers on nutrients – quantity of fertilizer that can be applied – based on cropping information and soil data they provide. Frank Place (ICRAF) was interested in the AfricaRice effort to collect data from 19,885 postharvest actors given the limited CGIAR experience, and challenges involved. Aminou briefly spoke about the sampling protocol they developed. For instance, in case of rice, there are a number of such actors (parboiler, miller, trader, etc.). And to develop a sampling frame of traders, they go to the markets and do a complete listing of traders before sampling from such a list (taking care of attrition from refusal to participate).

1.4 ILRI and Livestock and Fish CRP: Lucy Lapar

- Lapar presented a highlight of ILRI IA studies stating that they focused on *ex post* and *ex ante* assessments, as well as IA strategy, key constraints to effective IA of Center and CRP portfolio, and challenges for IA in moving from Center to CRP.
- She had to go back a decade (1998) to find a compilation of IA studies (Thorton and Odero ed. 1998), and there hasn't been such a compilation since.

- There is recognition that M&E does not tell them what would have happened in the absence of their program – so they are moving in that direction, and are funding at least one IA in a year.
- In speaking to challenges in conducting effective IA of CRP and Center work, Lapar pointed out that CRP funding for IA is restricted to CRP work. In that sense, CRPs dictate the kind of IA that will be implemented at the Center level, and will not support IA initiatives that fall outside the CRP mandate (e.g. work prior to CRP era). She also pointed out that CRPs are evolving and may not be ripe for IA.

Doug Gollin, in responding to Lapar's summary of challenges, flagged a question for general discussion on how Centers and CRPs prioritize which technologies to evaluate with RCTs (considering RCTs are expensive). In L&F's case, Lapar responded that they are attempting to identify best bests – within each of the 8 value chains, they are looking at *ex ante* assessments to identify best bets that can later be subject to IA.

SESSSION 1: General Discussion

- Tim Kelley (ISPC Secretariat) noted that it is clear that M&E is getting pretty systematic treatment (e.g. at IRRI). For IA, it appears that there is funding but it is unclear where the demand is coming from – where is this demand coming from? How is all this going to get prioritized? And to what extent is CRP going to leave all previous work to ILRI?
- Sam Mohanty (IRRI, GRiSP) responded that at GRiSP the IA budget became the M&E budget, and there is no longer any IA funding. Unless they get initial funding (e.g. from SPIA or BMGF) to start IA work, there is no dedicated IA budget for GRiSP even if there is interest and in-kind support from project staff.
- Guy Hareau (RTB, CIP) agreed with Mohanty that there is more money for M&E in RBM sites than for IA.
- Abdoulaye Tahirou (IITA) pointed that CRPs tend to hire M&E specialists, and assume that IAs will done by Center economists.
- Aminou Arouna (AfricaRice, GRiSP) agreed with these perspectives – that because many are focusing on M&E, at AfricaRice, they are attempting to illustrate how there is complementarity between M&E and IA, but that they are different.
- Speaking to prioritization, Nancy Johnson (IFPRI, A4NH) noted that they were using Theory of Change (ToC) to understand if IA is required (again, like ILRI/L&F, identify best bets). The need for epIAs comes from looking at what evidence they need going forward, not necessarily because they identify what is working/good.

Steve Franzel (ICRAF) spoke to the issue of collecting adoption data, and extrapolating from small adoption surveys to larger areas. For crops such as maize that get sufficient attention from central statistical agencies, and extension systems are pushing seeds, this is relatively easy. At the other extreme, in case of sweet potatoes, napier grass, or agroforestry, seed is not distributed formally, there is plenty of farmer-to-farmer dissemination, and there is a lack of prioritization of products – hence, adoption is patchy and few adoption surveys are done. Is there some means of exchanging ideas or information to tackle some of these issues?

Bob Herdt noted that across the system, a large number of farmers are being surveyed – with any one survey is focused on a single commodity, and impact assessments focused on the project areas. For e.g. in case of Western Kenya, there are a lot of commodities and markets in question, and hence they are essentially covering the same geographical areas through multiple surveys. Is there is a better approach to conduct these surveys? Do large-scale ones that are not project-oriented and can give us a sense of technologies that are being adopted in the agricultural system? In responding to this, Doug Gollin asked Centers and CRPs how much they coordinate with each other on surveying,

and if this is systematic or haphazard/opportunistic. If the latter, there could be a useful role for SPIA to play as a clearing house for information on survey initiatives by various Centers and CRPs.

- Abdoulaye Tahirou (IITA) agreed this might be a good idea, but noted that when IITA, AfricaRice and ICARDA attempted to collaborate, the survey lengthened to 45 pages – that it may not be always practical and may increase time required of farmers.
- Monica Fisher (CIMMYT, Wheat, Maize) also agreed that this would be a good initiative, and in responding to Tahirou's concern thought that duplication was taking place regardless (i.e. surveys already ask about many other crops besides the focus of Center/CRP), and that one could first evaluate how much of the collected data is necessary for analysis in order to reduce survey length.
- Ricardo Labarta (CIAT) agreed that such a coordination effort requires a lot of investment at the beginning, but is worth the effort. He pointed that they had an excellent experience with DIIVA where they coordinated with ICARDA for potato and barley (typically grown by the same farmers). However, in his opinion, it was easier to coordinate with Centers than CRPs (despite the nature of CRPs being a collaborative effort).
- Sam Mohanty (IRRI and GRiSP) spoke of IRRI's experience in South Asia – how resource scarcity necessitated data collection collaboration between ICRISAT, CIMMYT and IRRI.
- Marie-Charlotte (IWMI, WLE) thought such collaborations could also improve the quality of IA by (potentially) providing baseline data for the areas they are interested in.

Given donor (for e.g. BMGF) interest in adoption outcomes, Doug Gollin asked the Centers/CRPs to speak to adoption studies (other than DIIVA) done on a regular, systematic basis. While it was clear that CIMMYT (SIMLESA), GRiSP (moving towards panel data every 3 years), CIAT (Bolivia, Peru, Rwanda), IITA (building on DIIVA), and CIP (Peru) are all doing some data collection, it isn't always nationally representative or systematic (beyond the project).

SESSION 2: NRM and the Environment Improvement Focus

2.1 IWMI and CRP Water, Land and Ecosystems: Presentation by Marie-Charlotte Buisson

- Marie-Charlotte Buisson used examples of IAs done by IWMI to illustrate their approaches to project and policy IAs.
- She also spoke specifically to challenges in IA of the research conducted in IWMI and WLE, particularly attribution, baseline, IA of program and IA of project/activity, and sustainability aspects.

Peter Hazell (IFPRI) asked if IWMI/WLE constructs counterfactuals with and without policy change.

2.2 CIAT: Presentation by Ricardo Labarta

- Ricardo Labarta spoke about how CIAT is rebuilding its IA capacity, and the challenge of conducting IA work across the 11 CRPs CIAT engages with.
- He also highlighted recently completed CIAT studies, including the one that builds off DIIVA data, and on-going epIAs.

2.3 WorldAgroforestry, CIFOR, and Forests, Trees and Agroforestry CRP: Presentation by Steve Franzel and Daniel Suryadarma

- Steven Franzel presented highlights from on-going IAs at ICRAF, and spoke about how establishing baselines for epIAs is a high priority in research for development (R4D) projects.
- Daniel Suryadarma spoke about IA at CIFOR – as an institution that engages in policy research, CIFOR gets to impact through influencing policy and practice. He also spoke about how PORIA is challenging due to a multiplicity of issues (selection into treatment, uncertainty in policymaking timelines, context dependence etc.).
- CIFOR is attempting to improve the rigour of PORIA research methods, and is looking for collaborators to develop and test such methods.

In responding to a epIA of agroforestry practices in Malawai, including of fertilizer tree technology, Tim Kelley (SPIA) asked if there is an aggregate estimate of the extent to which fertilizer trees have spread. Frank Place (formerly of ICRAF, IFPRI) responded that conservation farming group in Zambia was disseminating them to 10,000s of farmers. But he hadn't worked with major development organizations that have disseminated this in the recent years, and this is something they don't have a good sense of yet.

SESSSION 2: General Discussion

Doug Gollin asked if the challenge in measuring NRM outcomes is that they cannot be measured at the household level. Steve Franzel (ICRAF) responded that this wasn't an issue of measurement at the household level, but one of patchy adoption and extrapolation: diffusion is frequently based on where development agencies work or is dependent on farmer-to-farmer dissemination. James Stevenson (SPIA) added that adoption does appear to be a relatively rare event in nationally representative surveys like LSMS-ISA. Bob Herdt (SPIA) countered that some of the NRM work has been occurring for 20-30 years – for instance, fodder shrubs in maize cropping systems, and that if diffusion has indeed occurred, we would have found that by now. Mywish Maredia (MSU) noted that a community survey could be helpful in instances where household surveys may not be efficient.

Peter Hazell (IFPRI) asked if and what measures of environmental impacts were being captured. Frank Place (IFPRI, formerly ICRAF) responded that for soil (work at Centers like ICRISAT and CIAT) – soil carbon and soil moisture content using high resolution imagery is measured. On the vegetation and tree cover front, similarly, very high resolution mapping and spectral techniques are starting to come up. In response to Anita Regmi's (Consortium Office's) concern about the lack of discussion on environmental impacts, Doug Gollin and Tim Kelley (SPIA) confirmed that SPIA is interested in NRM as a research category and impact area, even if there isn't substantial track record (data isn't available for reasons Frank Place pointed out or attribution is difficult).

There was also a general discussion on the demand for measuring gender impacts. Nancy Johnson (IFPRI) noted that the challenge in gender research isn't in data collection (disaggregated data can be collected), but the gap at the Theory of Change level about how gender influences research (and Abdoulaye Tahirou (IITA) agreed).Lucy Lapar (ILRI) also agreed with Nancy – for instance, one needs to examine if the preference for varietal characteristics is gender dependent - and, stated that an approach where every questionnaire requires gender-disaggregated sections may not be necessary, even as one recognizes that separate interviews may be needed in specific contexts (e.g. in Malawi, without asking the woman/women in household, complete information on income cannot be assembled).

SESSION 3: Systems Analysis and Improvement Focus

3.1 Dryland Systems CRP and ICARDA: Presentation by Aden Aw-Hassan

- Aden Aw-Hassan spoke about the work of Dryland Systems CRP across 5 regions (West African Sahel & Dry Savannas, North and West Africa, East and Southern Africa, Central Asia, South Asia, including impact pathways and relation to IDOs.
- He also highlighted some recently completed and on-going studies, including ones on zero tillage, collaboration with CRP Wheat, and supplementary irrigation.
- In speaking to challenges and constraints to IA work, he noted that a portion of the total research budget should be earmarked for IA to enable effective coverage of CRP and Center portfolio.

Frank Place (ICRAF) stated that the uniqueness of System CRPs is their integrative approach – in many senses, significant effects on poverty will only occur with such approaches, and that it would be more interesting to evaluate these innovate approaches instead of individual technologies (and this poses a significant methodological challenge).

3.2 Integrated Systems for Humid Tropics CRP and IITA: Presentation by Abdoulaye Tahirou

- Abdoulaye Tahirou also presented on the framework, theory of change, and impact pathway for Humid Tropics.
- In its approach to IA, Humid Tropics intends to conduct situation analysis and baseline studies as well as conducting epIAs in Action areas. HumidTropics has identified how each type of study under strategic studies, impact evaluation, and outcome evaluation contribute to the System Objectives.
- Speaking to challenges of conducting IA, Tahirou stated that there were difficulties in transitioning to experimental methods, data quality issues (noise in measurement of plot level yields in RTB crops), and the issue of selection bias and attribution in past and on-going programs.

Doug Gollin commenting on both systems presentations, stated that one way to think about IA is to identify a lot of possible locations of these hubs *ex ante,* and choose a subset of these over time – both in places where there are activities and places where there isn't in order to establish some kind of logical counterfactual. He raised a question on what happens in places where the CRPs are not actively engaged – how does impact correlate with the intensive space versus margins. Steve Franzel (ICRAF) observed that choosing a counterfactual could be tricky as donors (e.g. in Kenya) may choose to work in a site because of no/low CG presence.

SESSSION 3: General Discussion

Bob Herdt, in adding to the SPIA Chair's comments on counterfactual, observed how many take counterfactual as synonymous with control. For impact analysis, we should feel free to think more broadly: what information can we use/bring to bear in terms of what the innovation would or would not have done? He also emphasized that impact analysis goes beyond measurement of consumer and producer surplus.

Based on Tahirou's presentation of a target of 11 million people lifted out of poverty, Doug Gollin raised a question to other CRPs – if they were feeling pressured to specify these high level targets. He noted that it may not be feasible to measure how many people are lifted out of poverty *and* attribute this to research. Greg Traxler (BMGF) observed that that is a number that is important to donors. Frank Place (IFPRI) noted that the other issue with such number is that at CG level, there isn't one intervention that will lift people out of poverty – but it is hard to capture the effect of integrated approaches with the current approach to IA (individual studies).

SESSION 4: Policy oriented focus

4.1 Agriculture for Nutrition and Health CRP: Presentation by Nancy Johnson

- Nancy Johnson presented highlights from recently completed IAs on biofortification (RCTs in Uganda and Malawi on orange sweet potatoes), food safety, and integrated agriculture-nutrition-health programs. Additional evaluations are being planned in these areas (e.g. iron beans in Rwanda, iron pearl millet in India, integrated programs in Burkina Faso etc.).
- She also spoke about how the creation of CRPs, and work on IDOs and impact pathways is positively influencing impact orientation of research, and new approaches they are implementing as regards the same.

4.2 Policies, Institutions, and Markets: Presentation by Frank Place and Peter Hazell

- Frank Place presented a list of completed, on-going, and planning epIA studies, and some results from the completed studies.
- He then spoke to the challenge of policy oriented impact assessments, and the workshop being organized by IFPRI and PIM to identify best practices.

Aden Aw-Hassan (ICARDA) asked how the idea that policy research should be demand driven is reflected in the PIM framework, specifically where the policymaker was involved. Frank Place

responded that governments do invite IFPRI to work on programs and gave specific examples. Tim Kelley (ISPC) noted that policies often develop in ways that it can be hard to ascertain if it is a positive or a negative development – it is not always easy to perceive this beforehand, there will be winners and losers. Peter Hazell (IFPRI) spoke about how (in longer policy processes) it may be difficult to find the people who were around at the time to be able to provide insight into how the policies were made and the role of research (complicating 'contribution' analysis, attribution difficult). James Stevenson (ISPC) asked if CRPs and Centers draw on the expertise of political scientists: Peter Hazell (IFPRI) responded that IFPRI does use political scientists, and Daniel Suryadarma (CIFOR) spoke of how CIFOR includes two political scientists as researchers and a knowledge broker (who works on UNFCCC processes).

JV Meenakshi (SPIA Member) observed that there are two types of policy influences: deriving policies from evaluating programs, and creating a demand where there isn't awareness of the existence of lacunae in policy – the latter is more challenging, and perhaps a reason why there isn't much of that on the research agenda.

Doug Gollin speaking to the challenges in PORIA, agreed that there is concern with attribution and pathways of policy changes etc. However, one could approach this from the other end, and look at the reduction in number of bad policies and document this. The analogy in crop productivity is documenting an increase in productivity, and being able to attribute this to specific research agenda. Nancy Johnson (IFPRI) stated that the ToC does both, and Daniel Suryadarma (CIFOR) agreed that CIFOR/FTA does this as well – for instance, working backward on sustainable forest management in Congo (using ODI's outcome assessment).

PROGRAM HIGHLIGHTS: SIAC (Strengthening Impact Assessment in the CGIAR)
During the last session, there were presentations on Objectives 1, 2, 3, and 4 on the Strengthening Impact Assessment in the CGIAR (SIAC) program, 2013-2016 by SPIA Chair, SPIA Secretariat, and partners Mywish Maredia (MSU) and Jeff Alwang (Virginia Tech), presentations for which can be found online.

Agriculture technology adoption and poverty: Presentation by Bob Herdt
Bob Herdt spoke briefly to IA of agricultural research, and links to poverty impacts. He cautioned researchers from falling into the 'yield trap' without thinking about the concept of productivity and the idea that there are direct costs associated with yield gain. And that one should go beyond measuring productivity – measuring, at the minimum, income gains. He spoke about the importance of defining the target population for agricultural research more carefully. In reference to the fact that adoption outcomes are still poorly documented, emphasized that 'we should be talking about the number of people using specific technologies', and the distribution of adoption (disaggregate across agroecological zones, gender etc.). For instance, Swarna Sub-1 case shows that land that is inherently poor benefits more, and that is a good story. Finally, researchers need to remember that agriculture is practiced in a larger economic context, and poverty reduction may occur when people move out of agriculture.

Responses to SIAC presentations and general discussion
In speaking about the SPIA quality rating mechanism, Doug Gollin emphasized how this would be a voluntary mechanism and will function a bit like an online journal (SPIA membership would manage the process by using appropriate external reviewers). While journals may emphasize novelty and originality, SPIA is more concerned with doing something correctly and doing it well. For the star rating, given the lack of precedent for this, the rating system will be calibrated over time. He noted that the draft set of criteria was available on the website for feedback.

In response to the SPIA presentation on Quality Rating (SIAC Objective 4.3), the following questions and comments were raised

- It is important to judge relevance and context of studies – in the older system, major evaluations were compilation of many other studies, providing a better context than individual adoption studies.
- An approach to incentivizing researchers to submit their work to the Quality Rating process would be to publish rated/reviewed work as SPIA Working Paper. While some IAFPs thought that a SPIA Working Paper series might be considered more prestigious, others agreed that one would have to evaluate how this will compete with existing Working Paper series at the Centers. Overall, the success of this process would depend on how donors view the final product – whether they view these IAs as examples of good quality work.
- A frame of reference on historic studies could be the IEA evaluation of CRPs timeframe – the idea is not to review and rate all (older) publications.

Doug Gollin also noted that SPIA could broker review of concept notes in order to ensure that studies are designed more effectively from the very beginning.

During the general discussion that followed, Doug Gollin observed that there is enormous pressure on Centers and CRPs to do project-level studies, and asked the IAFPs how they got at higher-level issues (e.g. integrated approaches). In the discussion that following, the following points were raised

- Donors continue to drive demand for IA, and of specific projects. However, CRPs fund IAs out of their own interests – to inform their strategies. For e.g. at IRRI, since 2010, there is a pot of money for strategic research, including impact assessments.
- Measuring indirect effects of agricultural research – non-farm economy, price effects, labour market effects etc. are beyond the focus of an individual Center or CRP. What is missing in IA is a look at the system-wide impacts.
- Sentinel sites are a different approach to IA – they respond to the need for long-term assessments, and shifting targets.
- SPIA is issuing a call for proposals on long-term, large-scale impact assessments that will go beyond a single crop/commodity and require an integrated approach.

Donor perspectives: Presentation by Greg Traxler

Speaking as a donor representative, Greg Traxler spoke to outcome and impact indicators of interest to donor agencies, the evolution of both the CGIAR system and the quality of IA, and other perspectives on IA. He noted that DIIVA was a response to the lack of systematic collection of adoption data in the system: that BMGF had hoped that DIIVA would illustrate the feasibility and approach in regular collection of such data, but that such institutionalization does not seem to have occurred. To BMGF, credible information on the level of adoption/diffusion is critical because there is a strongly maintained hypothesis that farmers would not adopt a technology that isn't profitable, that total benefits are correlated with area under technology, and further, that the overwhelming part of benefits is through crop improvement in the CGIAR. Perhaps this lack of focus is because of incentive compatibility – for researchers, it isn't a priority to go around and collect such data. Now, MSU was taking the DIIVA effort forward (through SIAC Objective 2.1) even as the reliability of expert opinion and household survey data is in question (e.g. in Ethiopia, only 9% of farmers were able to correctly identify wheat varieties). In this context, DNA fingerprinting (something MSU is also working on) appears to hold significant promise, and he encouraged IAFPs to use the SPIA efforts to ramping up this effort and extending it.

On IA methods, he thought it encouraging that the IAFPs are using methods that are more credible. In response to some of the discussion on RCTs, he agreed that one would have to examine the instances in which it is the correct tool to use (it gives a strong power of the test but, in other instances, one is looking for big differences). He also stated that donors have a significant appetite to understand gender issues – that some serious impact work is required around that. A frequent

question that is asked internally is the numbers of people lifted out of poverty as well as distribution of benefits. He also agreed that while there are cases where a single technology is helpful and is an appropriate focus for IA, assessments of research programs are appropriate in other contexts. Donors are also increasingly monitoring impacts over a short period of time: for instance, can one observe impacts at the end of one investment period, and perhaps a maximum of 10 years (not more than that).

Close of the meeting

Prior to the close of the meeting, a discussion on the next IAFP meeting followed: suggestions included pre-conference sessions at IAAE-Milan 2015 or ISPC Science Forum 2015, and SPIA will discuss these (and other) options with IAFPs before finalizing. Doug Gollin closed the meeting by thanking the participants, particularly the IAFPs for the excellent presentations and discussions. He reiterated that SPIA should be looked upon as a resource on impact assessments, encouraged IAFPs to interact with one another through the website and blog, and that he (and others at SPIA) would love to hear individually from the IAFPs outside of the IAFP meetings on these topics.