



Evaluation of the “Strengthening Impact Assessment in CGIAR” (SIAC)

Project Phase 1, 2013-16

Volume 2 - Annexes

December 2016



Julia Compton
Timothy J. Dalton
Sophie Zimm



Independent
Evaluation
Arrangement

This evaluation has been commissioned by Standing Panel on Impact Assessment (of ISPC) and is managed by the Independent Evaluation Arrangement (IEA) of CGIAR.

The Independent Evaluation Arrangement (IEA) of CGIAR encourages fair use of this material provided proper citation is made.

Correct citation: CGIAR-IEA (2016), Evaluation of the “Strengthening Impact Assessment in CGIAR” (SIAC), Project Phase 1, 2013-16, Volume 2 (Annexes). Rome, Italy: Independent Evaluation Arrangement (IEA) of CGIAR.

<http://iea.cgiar.org/>

Cover image credit: IRRRI

Table of Contents

Annex 1: List of Stakeholders consulted	1
Annex 2: Evaluation Team composition	4
Annex 3: Timeline of key SIAC events	5
Annex 4: What is “impact assessment” and why is terminology important?	7
Annex 5: What do donors want from impact assessment in the CGIAR?	9
Annex 6 Brief findings on the organisation of IA in CRPs and Centers	13
Annex 7: Independence, impartiality and overcoming bias in impact assessment	15
Annex 8: Progress against SIAC outputs	17
Annex 9: Science quality assessment	26

Annex 1: List of Stakeholders consulted

Name	Organisation	Position	Role in SIAC	Interview type
Dr.Tahirou Abdoulaye	IITA	Agricultural Economist	IAFP	FGD
Dr.Ramadhani Achdiawan	CIFOR and FTA	Research Specialist	IAFP	FGD
Dr. Andrew Alford	ACIAR	Research Program Manager for impact assessment program	Peer and Donor	VOIP
Prof. Julian Alston	UC Davis	Distinguished Professor, Agricultural and Resource Economics	PSC member	VOIP
Prof. Jeff Alwang	Virginia Tech	Professor, Department of Agricultural Economics	Principal Investigator across multiple studies	Informal F2F
Dr.Aminou Arouna	Africa Rice	Agricultural Economist - Value Chain Specialist	IAFP	FGD
Dr. Nick Austin	CGIAR Consortium		Interim Executive Director of CGIAR (till October 2016)	VOIP
Dr. Aden Aw-Hassan	ICARDA	Principal Agricultural Economist and Director of Social, Economics, and Policy	IAFP	FGD
Prof. Brian Belcher	CIFOR	Senior Associate (CIFOR) and Professor Royal Roads University (Canada)	CGIAR scientist	VOIP
Dr. Rob Bertram	USAID	Chief Scientist	Donor	VOIP
Dr. Sara Boettiger	n/a	Independent consultant, former Deputy Director at BMGF		VOIP
Dr. Bas Boumann	IRRI/GRiSP	CRP Director	Center or CRP management	VOIP
Dr. Marie-Charlotte Buisson	IWMI	Researcher - Economics, Impact Evaluation	IAFP	F2F
Prof. Erwin Bulte	Wageningen University	Professor	SPIA activity leader	VOIP
Dr. Derek Byerlee	n/a	Independent consultant	Peer	VOIP
Dr. Richard Caldwell	BMGF	Senior Program Officer	PSC observer	F2F, VOIP
Dr. Bruce Campbell	CCAFS	CRP Director	Center or CRP management	VOIP
Dr. Andrew Clayton	DFID	Social development adviser within DFID's Research and Evidence Division	PSC member	VOIP
Dr. Ruben Echeverría	CIAT	Director General	Center or CRP management	F2F
Dr. Olaf Erenstein	CIMMYT	Director of CIMMYT's Socio-Economics Program	Impact Assessment specialist	VOIP
Dr. Peter Gardiner	CGIAR Consortium	Director of Science	Former ISPC ED, CO Science Director	VOIP
Dr.Marcel Gatto	CIP	Post-doctoral researcher	IAFP	FGD
Prof. Maggie Gill	ISPC	Chair	ISPC	VOIP

SIAC Evaluation - Report

Name	Organisation	Position	Role in SIAC	Interview type
Prof. Doug Gollin	University of Oxford	Professor of Development Economics in the Department of International Development at Oxford University	SPIA chair	VOIP
Dr. Guy Hareau	CIP	Acting Global Science Leader, Social and Health Sciences and Innovation Systems	Principal Investigator across multiple studies & IAFP	FGD
Dr. Peter Hazell	IFPRI and PIM	Economist	IAFP	F2F
Dr. Robert Herdt	Cornell University	Advisor	SPIA member	VOIP
Dr. Emily Hogue	USAID	Team leader, Monitoring, Evaluation and IA	Peer and Donor perspective	VOIP
Mr. Albin Hubscher	CGIAR Consortium	Director of Finance and Corporate Services	PSC observer	VOIP
Dr. Karl Hughes	ICRAF	Head of Monitoring, Evaluation and Impact Assessment	Principal Investigator across multiple studies & IAFP	VOIP
Dr. Nancy Johnson	IFPRI	Agricultural economist	Principal Investigator & IAFP	VOIP
Dr. Timothy Kelley	SPIA	SPIA Secretary	SIAC management	F2F
Dr. Michael Kidoido	ILRI	Economist-Impact Assessment	IAFP	FGD
Dr. Enoch Kikulwe	Bioversity-Uganda	Associate Scientist	IAFP	FGD
Dr. Holger Kirscht	GIZ and IITA	Anthropologist	Donor	VOIP
Mr. Victor Kommerell	CIMMYT/Wheat	Program Manager	Center or CRP management	VOIP
Ms. Lakshmi Krishnan	SPIA	Agricultural Research Officer	SIAC management	Informal F2F
Dr. Ricardo Labarta	CIAT (and CCAFS)	Senior Scientist	IAFP	FGD
Dr. Leslie Lipper	ISPC	Executive Director	ISPC	F2F
Prof. Karen Marcours	PSE & SPIA	Associate Professor (Economics) and Researcher (INRA)	SIAC activity leader	VOIP
Prof. Mywish Maredia	Michigan State University	Professor, International Development	SPIA activity leader	F2F
Dr. Paswel Marenya	CIMMYT/MAIZE AFS CRP	Socioeconomist	IAFP	FGD
Dr. John McDermott	IFPRI/A4NH	CRP Director	CRP Director	VOIP
Dr. Dawit Mekonnen	IFPRI and WLE	Research Fellow	IAFP	FGD
Dr. Samarendu Mohanty	IRRI	Head, Social Sciences Division	IAFP	FGD
Dr. Conrad Murendo	ICRISAT	Special Project Scientist (Monitoring and Evaluation)	IAFP	FGD
Dr. Robert Nasi	FTA	CRP Director	Center or CRP management	FGD
Prof. George Norton	Virginia Tech	Professor	SPIA project leader	VOIP
Dr. Lesley Perlman	USAID	Program Analyst, Monitoring and Evaluation	Peer and Donor perspective	VOIP
Dr. Frank Place	IFPRI	Senior Research Fellow	IAFP	VOIP

SIAC Evaluation - Report

Name	Organisation	Position	Role in SIAC	Interview type
Dr. Tom Randolph	ILRI/L&F	CRP Director	Center or CRP management	VOIP
Prof. Mitch Renkow	North Carolina State University	Professor	Peer	VOIP
Dr. Rachel Sauvinet-Bedouin	CGIAR Independent Evaluation Arrangement	Head	PSC observer	VOIP
Ms. Neha Sharma	World Bank, CLEAR	Evaluation Consultant	Peer	VOIP
Ms. Birte Snilstveit	3IE UK	Evaluation Officer, Systematic Reviews at International Initiative for Impact Evaluation	Peer	VOIP
Dr. James Stevenson	SPIA	Agricultural Research Officer	SIAC management	VOIP
Dr. Graham Thiele	CIP/RTB	CRP Director	Center or CRP management	VOIP
Dr. Philip Thornton	CCAFS	Senior Scientist / Systems Analyst	CGIAR scientist	VOIP
Dr. Alan Tollervey	DFID	Head of Agriculture Research	Donor	VOIP
Prof. Tom Tomich	UC Davis	Director, Agricultural Sustainability Institute	ISPC	VOIP
Dr. Greg Traxler	Univ of Washington	Senior Lecturer	Formerly of Gates and PSC, initiated DIIVA and SIAC	VOIP
Ms. Ira Vater	SPIA	Program Officer	SIAC management	VOIP
Dr. Jonathan Wadsworth	Fund Office	Head of Fund Office	Donor perspective	VOIP
Dr. Tom Walker	Independent	Former CGIAR economist, leader of DIIVA project	CGIAR scientist	VOIP
Mr. Eric Witte	USAID	Senior International Affairs Specialist	Donor perspective	VOIP

Annex 2: Evaluation Team composition

The external evaluation team members were Dr Julia Compton, team leader (independent consultant) and Professor Timothy Dalton, impact assessment expert (Kansas State University). Neither evaluation team member had any direct involvement in the design or implementation of SIAC.

Julia Compton's background is in agricultural research and rural development, predominantly in Africa. She then worked for ten years in the UK Department for International Development, first as a rural livelihoods adviser, and eventually as deputy head of evaluation. Since leaving DFID in 2010 she has worked as an independent consultant specialising in evaluation, agriculture and food security and rural development. Julia recently led the evaluation of the CGIAR Research Programme on Agriculture, Nutrition and Health (A4NH).

Timothy J. Dalton has extensive experience in agricultural technology adoption and his research focuses on the relationship between agricultural production, technological change and the environment. His background is in agricultural economics. He is familiar with the CGIAR and has worked with FAO, USAID and USDA especially in the context of Africa. He is currently the Director of the Feed the Future Innovation Lab for Collaborative research on Sorghum and Millet at Kansas State University.

Sophie Zimm of the IEA, an experienced evaluation analyst, has worked closely with the external team. Prior to the CGIAR, Sophie worked in the evaluation departments of UNIDO and UNODC, and also gained experience in project coordination/management in Mozambique.

Annex 3: Timeline of key SIAC events

Developed by the SIAC/SPIA team and edited by the evaluation team

Key changes in context	Year	Key SPIA and SIAC activities and milestones
IAFP meeting in Brasilia: Growing dissatisfaction with state of evidence, including on adoption of modern varieties	2008	
	2009	Start of DIIVA project quantifying adoption of CGIAR varieties, supported by BMGF
De Janvry, Dustan and Sadoulet report (De Janvry et al., 2010) raises question of quality in impact assessments (selection bias) and recommends portfolio of RCTs	2011	
BMGF and other funders show interest in supporting large-scale program addressing weaknesses in evidence base and IA capacity within the CGIAR.	2012	New SPIA Chair (Doug Gollin) SPIA I initiates dialogue with the World Bank LSMS-ISA team to pilot joint collection of data on agricultural technologies SIAC proposal developed and approved for funding by BMGF MSU sub-contracted to manage SIAC Objectives 1 and 2
Strong donor interest shown in the nutrition area	2013	SIAC starts, with BMGF funding. 2 new SIAC management staff 3 RCT grants funded – noncompetitive process. DFID and the PSC push SPIA to put as much of the SIAC portfolio through competitive processes as possible Call for Eols on nutrition studies (new area for SIAC) Call for Eols on capacity building Small grants program launched
	2014	(mid-year) DFID starts to support SIAC, channeling funding through Fund Council Window 1 and FAO Inception workshop on collecting varietal release and adoption data Small grants program closed (high transaction costs)

SIAC Evaluation - Report

Key changes in context	Year	Key SPIA and SIAC activities and milestones (SIAC activity numbers in parenthesis)
	2014	<p>Compilation of Policy-Oriented Research (POR) outcomes database (Phase-I)</p> <p>Inception workshop for nutrition studies</p> <p>Workshop on poverty impacts (70+ people) and Impact Assessment Focal Point (IAFP) meeting</p> <p>DNA fingerprinting workshop</p> <p>Call for EOIs on long-term large-scale impacts</p> <p>Policy-Oriented Research Impact Assessment (PORIA) workshop with IFPRI</p>
CRP Pre-proposals developed	2015	<p>SIAC mid-term review meeting February)</p> <p>Inception workshop for experimental impact evaluations</p> <p>Scoping study commissioned on second 'IAs of under-evaluated area of CGIAR research' – livestock research</p> <p>POR outcomes database (Phase-II)</p> <p>CRP pre-proposals reviewed by SPIA Chair and ISPC Secretariat (including 3 SIAC management staff)</p> <p>Call for Eols on under-evaluated areas of research</p> <p>Procurement of lab services for DNA fingerprinting of crop samples</p> <p>Inception workshop for long-term large-scale impact studies</p> <p>Call for Eols on NRM practices and workshop for shortlisted NRM studies</p> <p>LSMS-ISA-SPIA pilot on adding technical questions to household surveys, with fieldwork in three African countries</p> <p>Study initiated on evidence on impact of modern technology on agricultural productivity</p> <p>Initiated compilation of data on CGIAR research investments</p> <p>Approval for SIAC W1 no-cost extension to mid-2017</p>
<p>CRP full proposals developed for Phase II</p> <p>Restructuring of the CGIAR with the formation of the System Council (SC) and System Management Board (SMB) – 1 July</p> <p>SMB and SC establish two new committees on Strategy Impact Assessment and Evaluation</p>	2016	<p>BMGF approval for SIAC no-cost extension to mid-2017</p> <p>Learning for adoption workshop</p> <p>Mid-term workshop for RCTs under calls 3.0 and 3.2</p> <p>Inception workshop for under-evaluated areas</p> <p>SIAC Progress review meeting</p> <p>3 workshops: IAFPs; Long-term / large-scale studies; and adoption</p>

Annex 4: What is “impact assessment” and why is terminology important?

An important conceptual issue for this evaluation is that the term “impact assessment” can be used to cover a wide variety of studies – small and large scale, under relatively ‘controlled’ conditions and/or ex-post¹. Different people use the term in different ways, potentially leading to talking at cross-purposes, for example when discussing the division of organizational responsibilities for IA, or when comparing the investment in IA by different parts of the CGIAR. The lack of clarity can also potentially affect meta-analysis and systematic reviews, which may “add up apples and oranges”, for example if a small-scale or proof of concept study is given the same weight as an ex-post study². Finally, the debates about whether it is acceptable to “cherry pick” positive results for the purposes of impact assessment also suffer from confusion about what sort of IA is meant³.

Lack of clarity in terminology can equally affect discussions around the scope and selection of SIAC studies. As discussed in the main report, SPIA traditionally concentrated on large scale ex-post Impact Assessments, but SIAC has broadened this remit to include microstudies of technologies/ interventions prior to their wide dissemination and diffusion – types of studies that would often be considered part of “regular” CGIAR research. While there may be good reasons for SIAC to carry out such studies in some cases, we feel that it is important to develop transparent selection criteria and to discuss these with CRGIAR research leaders. Using clear terminology, rather than using “IA” to cover all such studies, would facilitate such discussion.

Unfortunately, there is no accepted and systematic international terminology which fully distinguishes different types of impact studies. Different terminology is used by different organisations. A particularly useful example which could perhaps be adapted for use by the CGIAR is from EvaluATE⁴, which not only defines different types of impact studies (efficacy/proof of concept, effectiveness/early adoption, and scale-up) but also provides standards and checklists indicating the technical qualities

¹ “Ex post IA takes place after the program’s or project’s investment has generated the intervention, and sufficient time has elapsed and experience accumulated to assess the intervention’s performance in terms of longer term economic, social, and environmental consequences” Science Council and CGIAR Secretariat (2010): Instructions for the Reporting of Performance Indicators for CGIAR Centers

² There is even a risk that more weight is given to early stage proof of concept studies – which can be seen as more rigorous as they can take place under more controlled conditions - than to studies in real world conditions.

³ Our view is that it is fine to “cherry pick” positive results for **ex-post IA** if the specific purpose is to show that an overall research portfolio has been a worthwhile investment (a good example is Raitzer and Kelley 2008). Research is inherently a risky business and not all lines of research will lead to outcomes, but investors (such as CGIAR funders) need to be sure that the overall portfolio is giving reasonable returns (Perrin 2002). However, investors also need to be sure that the portfolio is managed responsibly; that researchers are making good decisions as to the direction of the research based on impact evidence. Such management requires different types of impact evidence, often collected earlier in the impact pathway that will look at both successful and failed research lines.

⁴ EvaluATE is the evaluation resource center for the USA National Science Foundation’s Advanced Technological Education program. http://www.evaluate.org/resources/cg_overview/

expected in a good-quality study of each type. **We suggest that SIAC/SPIA consider issuing a glossary and standards for IA as part of its work on building capacity.**

In the meantime, however, we refer to impact assessment or IA indiscriminately throughout this report, and only distinguish particular types of studies when it is important for the argument.

Annex 5: What do donors want from impact assessment in the CGIAR?

SPIA has long seen CGIAR funders as a principal audience for its work, and has commissioned workshops and two surveys (2003? Or 2005?) and 2014) to find out “what donors want” in terms of impact information (Raitzer and Kelley, 2008; SIAC, 2016; Watson, 2003). We were asked to investigate this area further in this short evaluation. We examined the results of the most recent survey and also interviewed 10 individuals from five major CGIAR donors (BMGF, UK-DFID, USAID, Australia-ACIAR and Germany-GIZ) as well as two CGIAR senior managers who work closely with donors. The results are summarized below, with appropriate caveats on the limitations of this short piece of work⁵.

The main requirements that funders have for impact information are threefold:

- a) **Justifying further funding to the CGIAR**, when writing funding proposals and defending decisions to senior managers and politicians. There are many competing calls on donor funding, and the agriculture sector -- and in particular agricultural research -- is often seen as a high-risk, slow-payback area. For many donors (half of those in the 2014 SIAC survey), the demand for evidence on outcomes and impacts has increased over the past ten years, since the global rise in attention to “impact evaluation” and results-based management. The perceived under-investment of the CGIAR in this area was indeed one of the motivating factors for the SIAC project.
 - “*Big numbers*” (of beneficiaries) are in demand, to compete (implicitly) with numbers coming from other sectors, such as health and education. Large-scale adoption and outcome studies are therefore a high priority, as are meta-analyses across the CGIAR system⁶. Two interviewees stressed the benefits of *investing in new methods to bring down the cost of collecting large-scale adoption and outcome data*, including validating methods and proxy indicators that can provide “good enough” precision.
 - Average data on adoption and economic benefits is not enough, however. Funding agencies require information on the *distribution* of benefits, in particular to the poor and women, as well as the impacts on a *range of indicators* (e.g. nutrition and resilience to climate change). Moreover, an earlier donor focus on smallholder farmers (e.g. see IFAD in Watson, 2003)) has broadened dramatically, for example to include value chains, urban and rural food consumers and the natural environment. All this implies a very ambitious agenda for outcome and impact assessment in the CGIAR, which cannot be satisfied by SPIA alone. Several interviewees also

⁵ Funding agencies vary in their interests and requirements, and a given agency may change its requirements over time, due to political pressures and aid fashions. Moreover, individual funding agencies are not monolithic: different parts of the agency often have different interests, and individuals do not always provide a corporate view.

⁶ Asked for examples of particularly useful studies, several interviewees cited positively an ACIAR-funded study of IRRI’s contribution to rice varietal yield improvement (Brennan and Malabayabas, 2011), as well as an earlier meta-analysis (Alston et al., 2000).

commented that these broader objectives required a broader skill set in impact assessment (beyond economics, and especially on the qualitative side).

- The reputation of the CGIAR is also very important in making the overall case for donor support. This has two implications. First, the *credibility of adoption and impact data* is important. Several interviewees thought it would be useful for SPIA to set clear standards for IA across the system. Second, *communications can be more powerful than data in shaping perceptions* (Pritchett, 2002). “Overclaiming” can undermine donor confidence, especially if claims are ‘debunked’ by later impact studies.

b) Reporting benefits from investments already made in the CGIAR.

- Broadly similar to (a) in the data demanded.
- Many bilateral donors are required to “demonstrate” the specific contribution of their funds to outcomes. The result is a plethora of individual studies and baseline/endline surveys on bilaterally-funded projects and donor ‘areas of influence’⁷, overload on national statistical systems, and even ‘beneficiary fatigue’ with surveys in particularly popular areas. International agencies and donors are aware of this problem and are slowly taking some practical steps to tackle it.⁸ (SIAC’s work with LSMS-ISA was seen as a step in the right direction.) Another challenge is timing: adoption of technologies typically takes 10-20 years, while most donors are under pressure to demonstrate the impact of current investments, typically in a time frame of 3-5 years.
- Some funders need to report on specific monitoring indicators. For example the USAID Feed the Future programme regularly reports against indicators such as “Individuals who have applied new technology or management as a result of U.S. Government (USG) assistance” and “Number of agricultural and nutritional enabling environment policies analyzed, consulted on, drafted or revised, approved and implemented with USG assistance”, and has defined how to collect and report on each indicator (USAID, 2016). Harmonising a core set of reporting indicators (linked to the SRF and IDOs) is an aim of the CGIAR, and there is currently a working group looking at this, which includes SIAC/SPIA representation. It is clearly in SIAC’s interest to use harmonised indicators in its studies (e.g. for adoption) wherever possible. SIAC/SPIA

⁷ Some donor data such as the Feed the Future household surveys are published as open data (Malawi example in <https://www.usaid.gov/data/dataset/7eb4c66f-3f68-4ac1-9ce9-b2240e81d66f>)

⁸ FAO has been addressing such issues through the "Global Strategy to improve agricultural and rural statistics" (GSARS) programme, an umbrella effort working to enhance the capacity of developing countries to produce and use agricultural and rural statistics and to strengthen statistical governance mechanisms. USAID and FAO have just (7 Sept 2016) announced a new \$15M project for agricultural integrated surveys (AGRIS) which will “capture improved annual data on agricultural production, but also broader and more detailed structural information relating to farms, including ... production costs, farming practices, and environmental impacts. It will incorporate recent innovations like remote sensing, GPS, mobile technology and various uses of “big data.” These tools will introduce more objective approaches to measuring agricultural performance, in some cases replacing traditional, more expensive methods.... promote the integration of disparate data sources, improve data timeliness and usability, and cut data collection costs.” <http://www.fao.org/news/story/en/item/430779/icode/>

could also possibly have an important role (along with specialist CRPs, see main text) in testing and validating some key indicators.

c) Decision making on where to invest in the CGIAR

- Several donors commented on the academic focus of most IA work, and the consequent difficulty of using the results in a “business case for investment” for particular areas of research. While there was some support expressed (by two donor interviewees) for the ‘public goods’ nature of some IA research, donors are generally more interested in IA that is designed to generate a body of evidence to answer specific decision questions. Small, situation-specific trials, and adoption studies that fail to answer distributional and ‘why’ questions, were seen as less useful.
- Donors are also concerned that their research money will be well managed. It is important to the donors interviewed – especially those funding Windows 1 and 2 - that the CRPs and Centers are seen to both generate good-quality data on outcomes and use this to test links in their impact pathways and make decisions about research management and priorities⁹. The respective roles of Centers/CRPs and SIAC/SPIA in generating and managing this data is discussed in the next section.
- It is equally important to donors that the CGIAR is able to set overall research priorities, based on a critical analysis of available information, rather than basing funding on historical allocations or unrealistic projections. SIAC/SPIA is not itself concerned with priority setting and ex-ante analysis, but it has a potentially important feedback link to ex-ante analysis.
- As documented earlier by Raitzer and Kelley (2008), many factors affect donor decisions to invest in a particular area of research. Moreover, it is difficult to rigorously compare the likely benefits of different areas of work (livestock versus forestry for example). Impact assessment (and other) data is therefore often used to justify funder decisions on broad areas of work, rather than to guide them. There is therefore a particular demand for good outcome and impact data for certain areas of research that donors feel are important, but lack large-scale evidence on outcomes and impacts, such as much NRM work. This is a long-standing issue – see e.g. Kelley and Gregersen (2005).

The **presentation of information** is also important to donors, and this has implications for SIAC/SPIA communication.

- Donor staff are typically under time pressure and do not want to trawl through many documents (however short) or a website in search of data. They need a quick way of checking what evidence there is for particular crops, research programs and indicators (such as nutrition, climate resilience or gender)¹⁰.

⁹ The CRP evaluations conducted by IEA normally review this area.

¹⁰ One format that has been well-received is the “Evidence Gap Maps” of 3iE (see paragraph 20). In contrast, despite a promising first impression, the SIAC evidence map on the website <http://impact.cgiar.org/publications/map> makes it hard to find evidence for specific research topics and indicators, even for CGIAR studies. SIAC is planning to improve it.

- Key data also needs to be presented in a format that allows donors to make clear statements (with appropriate qualifications on the strength of the evidence), rather than in what was described by one interviewee as an “overly scientific, cagey” style¹¹.

¹¹ Two interviewees specifically mentioned the DIIVA study results as being not ‘user-friendly’, so we checked the DIIVA short briefing paper (ISPC-SPIA, 2014). We agree that it presents considerable challenges for a non-specialist in translating the findings into statements of the benefits of funding research. The main findings are submerged in technical explanations and qualifications. The only mention of the CGIAR is in a short paragraph split between pages 4 and 5 (easy to miss), and the conclusions section focuses on technical methods for future adoption studies, rather than to the big implications for future varietal research.

Annex 6 Brief findings on the organisation of IA in CRPs and Centers

As mentioned in the main text, we believe that a full evaluation of the role of SIAC/SPIA in strengthening the IA system of the CGIAR would need to start with a detailed assessment of the roles and functions of other entities carrying out IA in the system. In our view, this should be a priority for the forthcoming evaluation of ISPC/SPIA. In the meantime, we offer the following observations, mostly based on interview data. These support the conclusions from the recent review of CRP evaluations (CGIAR-IEA-2016).

- Centers and CRPs are aware of the demand by funders to demonstrate what one IAFP termed “immediate and sustained impacts and value for money”.
- Some CRP flagships are engaged in systematically testing their impact pathways using impact evidence (see Annex 4) — and indeed the objective is that this should be done by all CRPs. Impact studies are also used as an integral part of CRP research to identify constraints to large-scale adoption that are then debated with national stakeholders (for example Aw-Hassan, 2016).
- However, the level of investment in IA, and the number of IA specialists, varies considerably (see Figure 5 in the main text). Some Centers have only a single IA specialist.
- The location of IA specialists varies: some work directly for the DDG(R) or CRP Director, some are located in a broader M&E team, and others are located within a “socio-economic” research team. This partly reflects whether the IA is seen as part of the evaluation function for the CRP/Center work, or whether IA studies are an integral part of the research – or both.
- Funding for IA is a big constraint. Much of the funding for IA comes from bilateral donors who want to conduct IAs of ‘their’ research projects. Even when a CRP includes plans for IA, they are frequently not fully funded. Many IA specialists described themselves as lacking core funds for their work and making opportunistic use of bilateral projects to collect data¹².
- Many CRPs have carried out one or more household surveys as baselines either for the whole CRP, or for projects¹³ within it. There does not appear to be any harmonisation of relevant indicators or methods between these surveys or collection of the data at a central CGIAR

¹² As an example, one of the better funded CRP IA plans presented in the 2016 IAFP meeting was that of RICE/GRIISP, which has budgeted in Phase 2 for \$0.5M from the lead Center (IRRI), \$0.5M from the CRP, \$0.75M from bilateral projects within the CRP, and \$0.3M from other sources, including calls from SIAC and 3iE (Mohanty et al., 2016).

¹³ Organising baseline and endline surveys for bilateral projects reportedly consumes a lot of the time of some IA specialists, and some IAFPs said that baseline surveys are not always well-planned and can end up as unanalyzed, unused data. This supports the findings of CGIAR-IEA (2016): “Much of the IA work is being promoted by donor requirements for IA of bilateral projects. This type of IA has previously been strongly criticized by the ISPC review of social science as being focused on short-term local impacts and often of low quality. Moreover, the requirement for detailed baseline surveys in many bilateral projects is drowning IA economists in data much of which are not analysed or used for IA (e.g., GRIISP, MAIZE).”

level. The degree to which all these CGIAR household surveys are coordinated with national and international (e.g. SDGs, World Bank-LSMS) efforts is also variable (often they appear to be stand-alone). Most CRPs are making an effort to follow the CGIAR open data policy and are sharing datasets, but as one IAFP said, “we need a better mechanism to bring the relevant data together into one global dataset” (see footnote 8 for recent developments in this area).

Annex 7: Independence, impartiality and overcoming bias in impact assessment

Lack of institutional independence has raised the question of whether Centers and CRPs should be allowed to conduct their own impact assessments and adoption studies, or whether these can be conducted in a way that can give users confidence that they are not biased. SPIA cannot do every impact assessment and adoption study that is needed in the CGIAR, so we think it is important to look carefully at the question of independence in evaluation, and how to promote it. The discussion in this annex draws on the ideas of Gakusi and Sindzingre (nd); Picciotto (2013) and White (2014).

Independence has a number of aspects, for example:

- Institutional independence - specifically, freedom from interference in choosing the questions asked, the data examined, and the conclusions and recommendations; SPIA has institutional independence from the CRPs, for example.
- Behavioural independence – avoidance of bias on the part of the evaluators. Bias can arise, for example, from a fear of offending powerful people or friends, a desire for future contracts, technical/disciplinary biases, or a personal wish to see an intervention/ technology succeed.
- The perception/ confidence of partners, enumerators and those being evaluated that the evaluators are completely independent from those responsible for the intervention (for example, by avoiding arriving in the village in a vehicle belonging to the programme being evaluated).
- Two other aspects -- that are sometimes harder for external than for internal evaluators -- are: Ability to ask the right (probing) questions; and Free and complete access to documentation and individuals for interview, without bias in selection.

Institutional independence from the organization or program being evaluated is insufficient on its own to ensure all the above aspects, although it may be helpful. Moreover, institutional independence at the level of SPIA/SIAC does not necessarily guarantee institutional independence at the level of SIAC subcontractors.

Other means of promoting impartiality and reducing bias include:

- Using a rigorous, replicable method: this was seen by most of our interviewees (including donors) as a more important guarantee of independence than the institutional location of the research. However, qualitative research (e.g. as part of PORIA) was one area where institutional independence was seen as important, due to the risk of (and lack of visibility of) bias in many existing “small n” methods (White, 2012; White and Phillips, 2012) — although not all interviewees saw this as a concern.
- Contracting external independent individuals to do the work, either alone or jointly with program staff: Many CRPs and Centers contract external IA specialists, working alone or with Center staff. However, one potential problem is that contracting others effectively not

only requires different skills than doing the research yourself, but also is not attractive to many IA specialists in the CGIAR, because it takes time away from their own professional interests and career progression that depends on developing their own publication record. (This is also potentially an issue for SPIA research staff.)

- Transparency on the questions being selected for evaluation, together with consultation with stakeholders with a range of views
- Oversight by independent committee: for example the Center Board or the CRP steering committee
- Appropriate implementation details to reduce the risk of perceived bias
- Open data, following the CGIAR open data policy
- Transparent and external peer review: used by some Centers and CRPs but mainly ex-post








In our view, SIAC/SPIA can best promote impartiality and lack of bias in IA by **setting clear standards for IA right across the CGIAR** (including its own studies), which **include the above aspects in the criteria**, and by **examining these aspects of independence as part of any quality assurance**¹⁴.

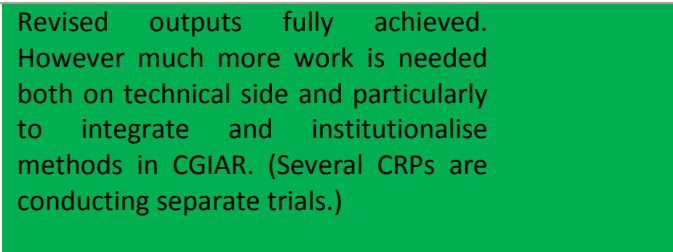
¹⁴ The current draft of SIAC's quality rating system for ePIAs does not cover this.

Annex 8: Progress against SIAC outputs

The table below presents information on planned SIAC output areas and a summary of activities and outputs for each. It also presents (in traffic light form) the evaluation team’s judgement on progress, together with SIAC’s own self-assessment (far right hand column). Generally these coincide. In a few cases, we have judged that although an output has been fully achieved in the strict sense of the word, the ‘spirit’ of the original output has not. For example, while the SPIA website has been redeveloped, the website is not yet a ‘one-stop shop for IA information in the CGIAR’ as planned.

The evaluation team had originally requested a self-assessment of progress against the specific output and outcome targets outlined in the SIAC agreement with BMGF¹⁵. However, many of these specific targets have been quietly dropped, and we think this is reasonable given that the project is still in a first phase and learning how best to approach many of these complex areas.

KEY TO TRAFFIC LIGHTS	
	Completed
	On track to complete by Phase 1
	Progress made, but finalising outputs will last beyond Phase 1
	Some setbacks (see comments)
	Discontinued activity
	New or modified activity
	Too early to tell

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
1: Develop, pilot and verify innovative methods for diffusion data			
1.1 Methods for crop varieties ¹⁶	Comparison of identification methods: DNA fingerprinting vs farmer and expert opinion. 12 substantial surveys (including 4 organised under activity 2.1 and 1 under 3.2) covering 6 crops/1 fish and 10 countries.	 Revised outputs fully achieved. However much more work is needed both on technical side and particularly to integrate and institutionalise methods in CGIAR. (Several CRPs are conducting separate trials.)	

¹⁵ A full list of these is available in the project proposal and also the Evaluation Inception Report, Annex 7.

¹⁶ Original wording: a) Validate and verify existing data on crop genetic improvement and b) Design and test new protocols for collecting data on diffusion of crop genetic improvements

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
1.2 Protocols for diffusion of NRM technologies	2 pilot studies on lower-cost adoption survey methods: phone surveys, satellite data.	This activity received little funding. It has been supplemented by methods work integrated into Activity 2.2 below but this will still likely require a second phase to develop further. Planned SIAC workshop on this topic in 2017 will synthesise learning.	
1.3 New institutional approaches for diffusion data	3 pilots of outsourcing adoption studies to national private firms to reduce costs.	Planned outputs achieved, but less than initial ambitious proposal to "Experiment with alternative institutional arrangements and new technologies for collecting data". Considerably more work is needed to standardise data collection methods and reduce design and management inputs.	
1.4 Disseminate best practices learned	Workshops on methods for tracking adoption mid 2015 to discuss early learning.	Making progress, with more dissemination planned for 2017, but unlikely to meet the over-ambitious outputs envisaged in the project proposal - this would need more time and investment.	
2. Institutionalize the collection of the diffusion data needed for IA			
2.1 Institutionalise collection of adoption data	Database of 12 crops and 17 countries, based on expert opinion surveys, with 130 crop x 'country' combinations (counting states in India and regions in China). Managed by MSU, subcontracted to Centers and NARS. (Training workshop for NARS.)	Exceeded original output target for 109 crop x country combinations. However, a one off collection managed externally may not lead to institutionalising the collection of such data across the CGIAR, without extra efforts.	
2.2 Case studies of NRM claims Original wording: "Organise and institutionalise the collection of natural resource	Database (compiled by consultant) with adoption and impact claims from 14 CGIAR Centers. Studies delayed as this activity was moved back from MSU to SPIA management in 2015. 9 adoption studies now underway covering 6 main NRM practices (Agroforestry / conservation agriculture (CA) / CA and agroforestry/ CA and	The activity is making progress toward recording adoption of some of the major CGIAR NRM claims, and testing comparative methods. Outputs from these studies are expected in 2017. However much more work will be needed to build on this and to develop cost-effective methods for tracking NRM adoption and institutionalise this.	

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
management research (NRM) related direct outcomes".	microdosing fertiliser/ Alternate Wetting and Drying / Integr. soil fertility management) and 17 NRM practice/'country' combinations. 4 adoption studies compare methods (e.g. remote sensing and panel surveys).		
2.3 Organise the collection of Policy-Oriented Research (POR) outcomes	Database (compiled by consultant) of policy outcome claims and assessment of evidence supporting them. 61 cases initially assessed as strong and others being followed up. Workshop on Methods for POR IA with IFPRI and PIM.	Original wording: "Organise and institutionalise..."Phase 1 outputs fully achieved. However more work will be needed to institutionalise this effort, so that the collection and validation is systematic and regular and not just a contracted-out exercise. So far the emphasis has been on the impacts of policy change and less on measuring the contribution of the CGIAR -an area of great interest to Centers/CRPs. Division of labour between SPIA, IEA and its CoP (which has done some training on methods in qualitative/contribution studies) and PIM/IFPRI in this area (and how best to support the rest of the CGIAR) deserves further study.	
2.4 Institutionalise the collection of adoption data [into existing national and international surveys]	Integration of adoption-related questions into national and/or World Bank LSMS-ISA survey rounds in Ethiopia, Uganda and Malawi (SPIA) and Zambia (MSU) on a pilot basis. Data being analysed. Initial discussions with India and Mozambique (MSU).	This activity is making good progress in four African countries but is still in a pilot stage and will need further time to embed and extend. Sustainability and national leadership are important questions to consider (as pilots were partially financed by SIAC).	

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
3. Assess the full range of impacts from CGIAR research			
<p>IA on nutrition and health Not included in original project proposal</p>	<p>5 diverse studies underway: high iron beans in Rwanda, NERICA rice in Sierra Leone, Crop diversification in Malawi and Ethiopia, Dairy hubs in Tanzania and Irrigated horticulture in Senegal.</p>	<p>New area added in response to donor interest in nutrition. Outputs are on track. However the diverse nature of studies and the fact that such questions are normally part of mainstream research effort raises questions about added value of SIAC managing these (instead of supporting CRPs to manage them) unless there are clear shared objectives (e.g. methods). Division of labour between SPIA and A4NH/IFPRI as a center of expertise on methods for IA in this area (and how best to support the rest of the CGIAR) deserves further study.</p>	
<p>3.1 Long term/large scale ePIAs</p>	<p>7 studies underway. 4 are testing large scale CGIAR adoption claims together with some outcome/impact data: C88 potato variety in China (Yunnan); Tilapia fish in Philippines and Bangladesh; lentils in Bangladesh and cassava in Nigeria. 2 studies use modelling approaches to estimating large scale impacts: CGE modelling to estimate the impact of CGIAR technologies on poverty; and Using global datasets to estimate impacts of CGIAR's modern varieties on agriculture, demographic and health and economic indicators. The final study uses qualitative approaches to investigate the effect of IFPRI's research on intrahousehold analysis on international NGO policy and practice.</p>	<p>On track for outputs. The three non-adoption studies are exploratory and should help inform SIAC planning by highlighting data gaps and developing methods.</p>	

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
3.2 Experimental/quasi exp studies	Five diverse RCTs/quasi exp studies on questions in uptake pathway: (a) rainwater harvesting techniques and cash transfers on yields and soil quality in Niger (b) bundling drought tolerant maize and weather index insurance in SSA (c) tailored extension recommendations and cash grants on fertilisation, precision sowing and outcomes in Mexico.(d) Drought tolerant rice and extension method (demonstrator) in Bangladesh and (e) Social networks and other methods for agri extension in Nepal.	Outputs are on track. However the diverse nature of studies and the fact that such questions are normally part of mainstream research effort (not ePIA) raises questions about added value of SIAC managing these (instead of supporting CRPs to manage them) unless there are clear shared objectives (e.g. methods). Division of labour between SPIA and IFPRI/PIM (Cluster 1.3 and Harvest Choice) as a center of expertise on methods for IA in this area (and how best to support the rest of the CGIAR) deserves further study. The methods study (on farmer behaviour) could be more widely applicable depending on design/external validity.	
3.3. Under-evaluated areas	Reviews of outcome/IA in a) irrigation and water management (inside/outside CGIAR) and b) IA evidence on CGIAR work in livestock. 4 ongoing studies: on a) forest comanagement in Guinea, 8 years post-project b) Alternate Wetting and Drying water management for rice in Philippines c) Adoption and impact of introducing Brachiaria forage cvs in Latin America, mainly Colombia d) Wealth and land health impacts of agroforestry in Kenya	The original plan was to do reviews of the evidence in six understudied areas before commissioning studies, but in the event these ran parallel, with only two reviews commissioned (and they took different approaches). The SIAC team recognises that there is more to do in this area. The competitive call process did not generate enough high enough quality proposals, so many evidence gaps remain uncovered. Needs more thought on how to select priority research questions and generate good-quality proposals – and a more systematic approach to reviewing evidence gaps, linked to the CGIAR SRF.	

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
<p>3.4 Meta analyses at system level*</p> <p>Original wording: Undertake a 'meta-analysis' of all recent large scale and credible CGIAR ePIAs and estimate different overall B-C scenarios</p>	<p>Studies supported include: a) economic rates of return to agricultural R&D (co-funded with Harvest Choice) the most ambitious review to date in coverage, synthesising nearly 3000 evaluations from nearly 500 studies; A related review on “The Returns to CGIAR Research, 1970-2015” is expected by end 2016. b) Historical review of expenditure by the CGIAR (unpublished). c) Impact of new technology on agricultural productivity (a re-examination of an existing review with a narrow scope). A synthesis review is planned for late 2017. Also SIAC/SPIA contributed comments to a separate meta-analysis of the returns to CGIAR varietal work by IFAD.</p>	<p>This is a key output area for SIAC with high external demand and also important in helping prioritise the other work and ensure that methods used will enable future meta-analysis. SIAC has commissioned some useful building blocks: for example the review of CGIAR expenditure has important lessons for future CGIAR monitoring and the re-review on evidence on agricultural technology has some lessons on methods and also identifies evidence gaps. However the impression given by the range of studies is that SIAC has been rather opportunistic rather than clearly prioritising work in this area.</p>	

4. Support Community of Practice for ePIA with CGIAR/partners

<p>4.1 Small grants</p> <p>Original wording: Small grants allocated on request to support communities of IA practice within the CGIAR</p>	<p>4 small grants totalling 30k allocated to projects on: a) IWMI, electricity and water pump policy in India b) ILRI, indicators for pastoral value chains in Senegal; c) CIMMYT 3. CIMMYT, agri. tech. package in Malawi and Ethiopia: analysis of dataset; d) Bioersivity, Home Gardens evaluation in Nepal: approaches to measurement and evaluation of gender impacts.</p>	<p>Mechanism discontinued after a few months due to high transaction costs related to FAO admin requirements. Small grants can be a useful mechanism (e.g. to support concept development) and might be easier to manage if funding was administered differently in phase 2 (see main text). However, the four projects are small additions to diverse ongoing CGIAR research, and why SIAC should be the ‘additional’ donor is not always clear.</p>	
---	---	---	--

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
<p>4.2 Strengthening IA capacity in the CGIAR through new partnerships</p> <p>Original wording: Training courses offered for CGIAR scientists</p> <p>4.3 Biennial conference on ePIA</p> <p>Original wording: Biennial CGIAR conference on ex-post impact assessment results and methods, held at a CGIAR Center</p>	<p>A number of different approaches have been used: a) biennial meetings for Impact assessment Focal Points to discuss experience and exchange approaches b) funding partnerships for capdev (chosen competitively) between 2 US universities and 3 Centers c) training workshop in RCT approaches for 38 CGIAR scientists d) methods/approaches/progress workshops for participants and partners under activities 2.1, 2.2 and 3.1 e) joint meeting on assessment of poverty impacts of agricultural research, with 62 participants, 18 from CGIAR Centers</p> <p>Planned for 2017.</p>	<p>Although many activities have taken place in this area, it is difficult to tell if outputs have been achieved as they have not been clearly and consistently specified. Capacity development has been a bit ad hoc in Phase 1 and has not benefited from an overall analysis of needs and SPIA's niche in capacity development.</p> <p>Output on track. Content and participation will be critical issues.</p>	<p></p>

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
4.4 Quality star rating for CGIAR IA studies	SPIA agreed quality criteria and launched an online system in April 2016. However it has not yet been taken up.	A quality system has been developed but (so far) no-one has taken up the opportunity to have SIAC/SPIA quality assure their products. From our interviews, there is skepticism in some quarters about value added and concerns about the reputational risks of a bad score, especially if applied by people with "over-narrow disciplinary approaches". This will require further investment in a more consultative approach to increase ownership (including joint analysis and agreement on quality criteria).	
(Original Activity 4.5) Facilitate interactions with regional research organizations on ePIA and provide support services to RROs and NARs	SIAC has not done this formally although NARs have participated in some workshops and specific projects.	Discontinued and never formally addressed in the project. This area is worthy of more thought for Phase 2 as national and regional organisations often have leading roles to play e.g. in household surveys and adoption studies. Contracting underfunded national organisations to carry out individual studies is not a sustainable approach.	
4.5 CGIAR Impact Website Original wording: Maintain and significantly enhance the CGIAR impact website as a one-stop shop on impact assessment activities	The CGIAR impact website was re-launched in May 2014 and now has additional elements e.g. a map of key evidence, a blog section, a community of practice section..	The technical output has been accomplished, but the 'spirit' of the output (being a 'one stop shop for IA in the CGIAR) has not yet been achieved. For one thing, the website does not contain comprehensive information on what IA is being carried out across the CGIAR. Becoming a 'one-stop shop' will require further investment not only in the website and in consultation on information needs, but also in the membership and management of the Community of Practice.	

Planned outputs	Summary of progress (as of August 2016)	Evaluation team's traffic light judgement on outputs (see key) and comment	SIAC self-assessment
<p>4.7 Support and capdev to Consortium</p> <p>Original wording: "At least three new studies launched based around a post-doc or pre-doctoral student or professor on sabbatical, funded through a competitive process"</p>	<p>Informal support given as a core function of SPIA, for example contributing to the Monitoring, Evaluation and Learning Community of Practice</p> <p>Dropped as an objective, but two post-docs hired as consultants under Activity 2.4 and several others involved in various strands of the work.</p>	<p>Discontinued as a formal SIAC project objective. System reforms and staff turnover in the Consortium have made this a challenging area.</p> <p>Discontinued as a formal SIAC project objective.</p>	

Key: (for details see Annex 1) ■ Outputs completed ■ On track to complete in Phase 1 ■ Progress made, but outputs will not be complete in Phase 1 ■ Some setbacks ■ Too early to tell ■ Discontinued activity ■ New or modified activity

Annex 9: Science quality assessment

Quality of science was evaluated following suggested guidelines contained in the Independent Evaluation Arrangement's January 2015 document "CGIAR Standards for Independent External Evaluation." While the document proposes reviewing processes of assuring science quality, inputs into the research process and outputs, we were provided with only a subset of this information, namely a few project outputs and annual reports plus a nearly exhaustive set of research proposals. This information was complemented with discussions with a number of the PIs who are in charge of conducting these projects and this information was complemented by attending meetings surround the AAEA meetings held in Boston from July 28 to August 3, 2016. We do have good information on processes and inputs with much less information on outputs since we are evaluating a relatively young program.

The review examined all projects to the extent that information was available. We were provided with a list of thirty-seven project agreements¹⁷ and a list of contracts amounting to \$9,260,780 of funding disbursed under the SIAC project and allocated to primary recipients in the 2016 SIAC Annual Report.

Processes for Assuring Science Quality

The primary metric used to evaluate "Processes for Assuring Quality of Science" is through a review of the calls for research or "Expressions of Interest" (EOI) for conducting a research project and the evaluation of the review process put in place by SPIA or MSU. Overall, both institutions have used two-stage calls to solicit research projects with short "Expressions of Interest" followed by submission of a "Full Proposal" from a subset of those submitting EOIs.

Procedures for evaluation and selection have evolved over the SIAC project to include workshops to cultivate better full proposals from those EOIs promoted to full proposals and also to promote collaboration between potential partners. One call (RCTs) also included a capacity building day where unsuccessful proposals received feedback and were also able to learn from presentations about successful proposals. Although we did not have time in this evaluation to investigate the benefits and costs of such workshops, our impression from past experience is that this is a useful approach to build better research and should be supported.

The calls for research proposals have been managed by several individuals at SPIA and their partners (e.g. Wageningen University, Paris School of Economics), and are not consistent in format. Because of this, it is suggested that a similar or standard format be developed for future calls for research as is done with other research funding organizations such as the U.S. National Institute of Health, Department of Agriculture or the National Science Foundation, where appropriate. In the first phase, there was a mix of research activities that required different evaluation criteria but overall there was considerable overlap of the criteria. In order to reduce ambiguity in expected research, it is suggested that the section on research methods or approach be standardized. For example, different

¹⁷ Annex 1. SIAC Program Report, 22nd July 2016. This misses out some large sub-agreements (notably those of Michigan State University) that were indirectly reviewed due to incomplete information.

terminology such as “causal relationships,” “indicators,” “methodology,” “approach,” “technical merit... feasibility” and “methods” were used to describe how a project would operationalize their proposal and investigate scientific questions. Some of the phraseology is more explicit than others and we recommend being as explicit as possible in solicitation by requiring proposals to formulate testable hypotheses and then provide specifics about what is required to rigorously test these hypotheses. This provides a check of the quality of thought and clarifies expectations on what information SPIA will receive through these studies.

Overall, the calls for proposal are explicit and provide clear criteria on what is being used to evaluate the EOIs and proposals. In most instances, the weights applied to each section are explicit but not on all, so this should be made explicit in all future calls.

In only a few of the calls was there a request for an explicit statement on data availability and research subject safety (ethics) made by the researcher. These issues are indeed often part of contract documents written between FAO and the implementing institution. However, this part of the contract is often reviewed and signed by institutional representatives and hence at arm’s length from the implementing researchers, who may not have thought through all the implications. Best practice¹⁸ (and a legal requirement in some places, e.g. the USA) is that open data and research safety should be a part of the research proposal rather than relegated to contract administration.

Inputs into the Research Process

Inputs into the research process is the second category of information that was evaluated to assess quality of science. In this review, the primary input that was reviewed where the proposals that were funded under each of the project’s objectives. We were able to review all projects that were funded under calls for an expression of interest but we were not able to review all project proposals that were administered under the MSU managed activities. We were unable to evaluate the research activities that were conducted under the Objective 4.2 “Capacity Development.” Thirty-two proposals were reviewed and qualitatively scored according to criteria established by the review team.

Research proposals were reviewed to determine 1) whether the objectives of the proposal were clearly stated, 2) whether the research proposal had a clearly stated testable hypothesis or hypotheses, 3) whether the proposal clearly presented a data collection strategy, 4) whether the data collection strategy would collect user-group disaggregated data, for example between men and women, ethnicity of caste, 5) was the analytical approach clear and would it lead to rigorous testing of the stated hypothesis, and 6) a qualitative assessment of cost in terms of whether any evidence of leveraging was provided. The first five indicators were evaluated as “Yes or present” or “No or not present” while the last question was used an ordinal rank from one to five, where one indicated “limited leverage or expensive” while five indicate “high leveraging.” Three was used to denote an average. Overall, 84% of the proposals appeared to be of average to highly leveraged value.

¹⁸ At a minimum, the researcher (not the administrator signing the contractual documents) could be required to attest that they are aware of and will abide by the CGIAR Open Data policy. A more proactive policy would require that the researcher indicate where the data will be archived, the metadata provided and the embargo period.

All but one proposal clearly stated the research objective and provided a compelling case for the research activity. Based upon the stated objective, we searched for clearly stated hypotheses that could be evaluated but also accepted statements that strongly alluded to a hypothesis. 69% of the proposals specified a clear, testable hypothesis. Most of those lacking a clear hypothesis were in the NRM call. We could often infer a hypothesis, but this is not the same as what a researcher might have in mind. 84% of the projects described their data collection strategy clearly. In many cases, but not all, authors made efforts to explain power calculations and contingencies. It is encouraged the projects be explicit in this area. By stating clear hypotheses and data collection strategies (along with appropriate analytical procedures) SPIA will have a much clearer idea of what results to anticipate from a project and thus be able to identify how the results contribute to strategic goals. Ninety percent of the proposals were clear on the analytical approaches that were to be used.

Only about one-third of the projects provided sufficient information in their proposals to determine the extent to which the data collection effort would allow for disaggregation of results to population subgroups for comparative purposes, for example between men and women, income or ethnic groups, or caste. Less than one third of the proposals gave an explicit plan for making the data open to external analysis or provided clarity that they understood the responsibility of the researcher to follow best practices with human subjects.

Regarding open data, for nearly all SIAC contracts there is a contractual clause on sharing anonymised datasets and surveys (both in English and local language where applicable). However, this does not necessarily guarantee that researchers have fully understood the implications and how to implement them. At a minimum, SPIA could require that the researcher (not the administrator signing the contractual documents) attest that they are aware of and will abide by CGIAR Open Data policy. A more proactive policy would require that the researcher indicate where the data will be archived, the metadata provided and the embargo period. SPIA could also take leadership to ensure that research subjects are protected through explicit review of human subject protocols, informed consent statements and affiliated practices rather than devolving the responsibility to Centers.

SPIA has provided leadership to strengthen the research conducted through these projects by creating workshops between the EOI and final proposal phase. This is an excellent idea to improve the focus of the project but additional attention could be applied to ensure that the research questions are rigorously formulated in the form of testable hypotheses and data collection strategies created that will provide the inferential support for evaluating the hypotheses. While this emphasis may seem pedantic, it should eliminate ambiguity about what outputs will be produced and the marginal contribution to science and SIAC's/SPIA's strategic objectives.

Outputs from the Research Process

At the time of data collection for this evaluation (August – September 2016), very few outputs from the research activities had been produced aside from interim project reports, a few conference presentations and peer-reviewed publications. This is largely an artefact of the short duration of the project and the length of time that was required to initiate research calls, evaluate proposals and contract activities. While one peer-reviewed publication had been produced (in a genetic journal), it had been in circulation for less than one year and received three citations. For this reason, we were unable to assess research outputs as an indicator of quality of science. However, there appears to be

a stream of outputs that will be finalized in the next six to twelve months that should provide ample opportunity for publication.

Annex table 9.1 provides a breakdown of the scores applied to the proposals reviewed in preparation of the quality of science section. The project listing has been reorganized to maintain anonymity. Project proposals (all had been accepted for funding by SIAC) were evaluated on six criteria: 1) whether or not the proposal presented a clearly stated objective, 2) whether the project presented a testable hypothesis or not, 3) whether the proposal described clearly their data collection strategy, 4) whether or not the data collections strategy intended to collect data that could be broken down into user segments e.g. by gender, caste, income group or ethnicity, 5) whether or not the proposal (not the contract documents) presented a strategy to ensure that the data complied with best-practice human subjects protocols and CGIAR open data policy, and 6) whether the proposal described how the data would be analyzed. These criteria were scored with a “1” if present and a “0” if absent. In certain proposal, some categories were not applicable for examples where secondary data was used. Those projects were scored with “na”.

Summary of findings on Quality of Science

The quality of science conducted under the SIAC project has been very good overall but there is high heterogeneity. This is partly related to the project investing in new methodologies, under-researched areas, application of tools less commonly used in the CGIAR and alternative approaches to data collection. However, it is also because the project has gradually been gaining experience in how to manage research calls and improve quality, including calls managed by partners. We encourage SIAC/SPIA to build on this experience and develop consistent documentation for calls for proposals which should specify inter alia the importance of disaggregated data, ethical approaches and open data.

Annex Table 9.1 Quality of science scores by project proposal

SIAC activity	Clearly stated objective	Testable hypotheses	Data collection strategy defined	User-group disaggregated data	Human subjects statement/open data	Analytical approach described
OBJ.2.2	1	0	1	0	na	1
OBJ.2.2	0	0	1	0	0	1
OBJ.2.2	1	1	1	na	na	1
OBJ.2.2	1	0	1	0	na	1
OBJ.2.2	1	0	1	1	0	1
OBJ.2.2	1	0	0	0	1	1
OBJ.2.2	1	0	0	0	1	1
OBJ.2.2	1	1	1	1	0	1
OBJ.2.2	1	0	1	na	na	na
OBJ.2.2	1	0	0	0	1	1
OBJ.3.0	1	1	1	1	1	1
OBJ.3.0	1	1	1	1	1	1
OBJ.3.0	1	1	1	0	0	1
OBJ.3.0	1	1	0	0	0	0
OBJ.3.0	1	1	1	0	0	0
OBJ.3.1	1	1	1	1	0	1
OBJ.3.1	1	1	1	0	na	1
OBJ.3.1	1	1	1	0	0	1
OBJ.3.1	1	0	1	1	0	1
OBJ.3.1	1	0	0	0	na	1
OBJ.3.1	1	1	1	1	1	1
OBJ.3.1	1	1	1	1	0	1
OBJ.3.2	1	1	1	0	0	1
OBJ.3.2	1	na	na	na	na	na
OBJ.3.2	1	1	1	0	0	0
OBJ.3.2	1	1	1	0	0	1
OBJ.3.2	1	1	1	0	0	1
OBJ.3.2	1	1	1	0	0	1
OBJ.3.3	1	1	1	na	na	1
OBJ.3.3	1	1	1	1	0	1
OBJ.3.3	1	1	1	1	1	1
OBJ.3.3	1	1	1	0	0	1
OBJ.3.5	1	1	1	na	1	1