INDEPENDENT IMPACT ASSESSMENT REPORT No. 41

Workshop on Best Practice Methods for Assessing the Impact of Policy-Oriented Research: Summary and Recommendations for the CGIAR

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

CGIAR RESEARCH PROGRAM ON Policies, Institutions, and Markets
Led by IFPRI

CGIAR Independent Science and Partnership Council

About IFPRI

The International Food Policy Research Institute (IFPRI), established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. The Institute conducts research, communicates results, optimizes partnerships, and builds capacity to ensure sustainable food production, promote healthy food systems, improve markets and trade, transform agriculture, build resilience, and strengthen institutions and governance. Gender is considered in all of the Institute's work. IFPRI collaborates with partners around the world, including development implementers, public institutions, the private sector, and farmers' organizations. IFPRI is a member of the CGIAR Consortium.

About PIM

The CGIAR Research Program on Policies, Institutions, and Markets (PIM) leads action-oriented research to equip decisionmakers with the evidence required to develop food and agricultural policies that better serve the interests of poor producers and consumers, both men and women. PIM combines the resources of CGIAR centers and numerous international, regional, and national partners. The program is led by the International Food Policy Research Institute (IFPRI). www.pim.cgiar.org

About ISPC

The CGIAR Independent Science & Partnership Council (ISPC) is a standing panel of eminent scientific experts, whose aim is to strengthen the quality, relevance, and impact of science in the CGIAR. The ISPC's overarching purpose is "to provide independent advice and expertise to the funders of the CGIAR through services to the Fund Council and the Funders Forum. It also serves as an intellectual bridge between the funders and the Consortium." http://ispc.cgiar.org/

About IFPRI Independent Impact Assessment Reports

The Independent Impact Assessment Reports are the product of externally conducted impact assessment studies of IFPRI's research. These studies are organized and overseen by an external impact assessment coordinator appointed by IFPRI who arranges for external experts to conduct the studies and who oversees a peer review of the draft reports by at least one internal and one external reviewer. This particular study was commissioned jointly by IFPRI and the CGIAR Research Program on Policies, Institutions, and Markets (PIM) led by IFPRI. While it followed the standard peer review process for all IIARs, the PIM impact assessment team collaborated in organizing and overseeing the study. This work was funded by PIM. PIM is in turn supported by these donors: [see http://www.pim.cgiar.org/donors/]. Any opinions expressed are those of the author(s) and do not necessarily represent the opinions of IFPRI or PIM.

Series Name Change Announcement

The Impact Assessment Discussion Paper (IADP) series has been renamed "Independent Impact Assessment Report" beginning with report #36, and the numbering for this series will continue from the IADP series.

Workshop on Best Practice Methods for Assessing the Impact of Policy-Oriented Research: Summary and Recommendations for the CGIAR

Frank Place
Peter Hazell

Independent Impact Assessment Report No. 41
December 2015

International Food Policy Research Institute
Washington, DC

Copyright © 2015 International Food Policy Research Institute (IFPRI).

All rights reserved. Sections of this report may be reproduced for personal and not-for-profit use without the express written permission of but with acknowledgment to IFPRI. To reproduce the material contained herein for profit or commercial use requires express written permission. To obtain permission, contact the Communications and Knowledge Management Division at ifpri-copyright@cgiar.org.

The views expressed here are those of the authors and do not necessarily reflect the policies or opinions of IFPRI.

AUTHORS

Frank Place, IFPRI
Senior Research Fellow (PIM)
Email: f.place@cgiar.org

Peter Hazell
External Coordinator of Impact Assessment
Email: p.hazell@cgiar.org

Published by
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
2033 K Street, NW
Washington, DC 20006-1002 USA
Tel.: +1-202-862-5600
Fax: +1-202-467-4439
Email: ifpri@cgiar.org
Web: www.ifpri.org

Contents

LIST OF ACRONYMS ..... **VII**

ABSTRACT ..... **IX**

1. OVERVIEW OF WORKSHOP OBJECTIVES AND SUMMARY OF SESSIONS **1**

INTRODUCTION 1
OVERVIEW OF WORKSHOP OBJECTIVES 1
SUMMARY OF SESSIONS 2
Introduction *2*
Taking Stock—What We Know About Assessing the Impact of Policy-Oriented Research *3*
Non-economist Perspectives on Impact Assessment of Policy-Oriented Research *4*
A Typology of Policy-Oriented Research *5*
Methods for Assessing Impacts from Policy-Oriented Research *6*
Indicators and Approaches for Tracking Policy Change *9*
Working Group Session *10*
Summary Session *12*
Donors' Perspectives *13*

2. CHALLENGES OF PORIA: BREAKING DOWN THE ELEMENTS **14**

POLICY OUTCOMES AND IMPACTS 14
Policy Outcomes *14*
Impacts of Policy Outcomes *16*
TYPOLOGY OF POLICY-ORIENTED RESEARCH 17
THEORY OF CHANGE (TOC) 18
Outputs—Evidence Feeding into Policy Processes *19*
Science-to-Policy Linkages *19*
Learning from the TOC *20*
SELECTION OF CASE STUDIES 20
INFLUENCE / CONTRIBUTION / ATTRIBUTION 21
INDICATORS 22
METHODS 24
Overall Approach to PORIA *24*
Steps in an Approach *25*
COUNTERFACTUALS 26

3. RECOMMENDATIONS FOR THE CGIAR **29**

IDENTIFYING TARGET POLICY OUTCOMES AND IMPACTS 29
DEVELOPING A THEORY OF CHANGE 29
TRACKING OF OUTPUTS AND DISSEMINATION OF ACTIVITIES 30
TRACKING OF OUTCOMES 30
COLLECTING EVIDENCE OF OUTCOMES 31
COLLECTING EVIDENCE OF RESEARCH INFLUENCE ON OUTCOMES 31
MEASURING IMPACTS OF POLICY OUTCOMES 32
CONDUCTING PORIA STUDIES STARTING FROM A POLICY CHANGES 32
ADDITIONAL RECOMMENDATIONS 32
SOME CAUTIONS 33

REFERENCES **34**

ANNEX 1: LIST OF PARTICIPANTS **36**

ANNEX 2: AGENDA **38**

LIST OF ACRONYMS

3ie	International Initiative for Impact Evaluation
ACIAR	Australian Centre for International Agricultural Research
BMGF	Bill & Melinda Gates Foundation
CCAFS	Climate Change, Agriculture and Food Security program of the CGIAR
CIFOR	Center for International Forestry Research
CRPs	CGIAR research programs
ERS	Economic Research Services
IA	impact assessment
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IFPRI	International Food Policy Research Institute
IMPACT	International Model for Policy Analysis of Agricultural Commodities and Trade
IPGs	international public goods
ISPC	CGIAR's Independent Science and Partnership Council
MCC	Millennium Challenge Corporation
MSU	Michigan State University
MTID	Markets, Trade and Institutions Division
NCSU	North Carolina State University
ODI	Overseas Development Institute
PIM	CGIAR Research Program on Policies, Institutions, and Markets
POR	Policy-oriented research
PORIA	policy-oriented research impact assessment
PROGRESA	Programa de Educación, Salud, y Alimentación
RCTs	randomized control trials
SPIA	Standing Panel for Impact Assessment
TOC	theory of change
UC	University of California
USAID	U.S. Agency for International Development
USDA	U.S. Department of Agriculture

ABSTRACT

Methods and practice for the impact evaluation of policy-oriented research (POR) lag behind many other research sectors, such as agricultural technology research, making it difficult to prioritize investments in POR and retarding institutional learning about how to make POR more effective. To address these issues within the context of the CGIAR, a workshop on "Best Practice Methods for Assessing the Impact of Policy-Oriented Research" was cosponsored by the International Food Policy Research Institute (IFPRI), the CGIAR Research Program on Policies, Institutions, and Markets (PIM), and the Standing Panel for Impact Assessment (SPIA) of the CGIAR's Independent Science and Partnership Council (ISPC), and held at IFPRI headquarters in Washington, DC, November 11–12, 2014. This paper provides a summary of the workshop discussions and draws on this, together with findings from the literature, to derive guidelines for assessing policy-oriented research in the CGIAR.

1. OVERVIEW OF WORKSHOP OBJECTIVES AND SUMMARY OF SESSIONS

Introduction

A workshop on "Best Practice Methods for Assessing the Impact of Policy-Oriented Research" was cosponsored by the International Food Policy Research Institute (IFPRI), the CGIAR Research Program on Policies, Institutions, and Markets (PIM), and the Standing Panel for Impact Assessment (SPIA) of the CGIAR's Independent Science and Partnership Council (ISPC), and held at IFPRI headquarters in Washington, DC, November 11–12, 2014. Details about the workshop, including the agenda, background papers, and presentations, are available. This paper provides a summary of the workshop discussions and draws on this, together with findings from the literature, to derive guidelines for policy-oriented research in the CGIAR system.

Overview of Workshop Objectives

A key rationale for the workshop was that methods and practice for the impact evaluation of policy-oriented research (POR) lag behind most other research sectors, such as agricultural technology research. Policy-oriented research is defined here as research intended to result in new or improved policies, regulations, and institutions (and their management) that enhance economic, social, and environmental welfare (Raitzer and Ryan 2008).

Although some types of POR can be subjected to rigorous impact assessment (IA) using randomized trials and econometric assessments, research on a huge range of policy issues is not amenable to the same standards of quantitative assessment. At a time when some key investors are demanding greater evidence of impact and 'value for money', this can place POR at a disadvantage when competing for funds with research sectors that can more easily quantify their impact. A lack of acceptably rigorous methods also discourages greater investment in assessing POR, as evidenced by the small amount of funding spent on assessing POR compared to other research activities in the CGIAR (Byerlee and Bernstein 2013) and the small number of journal papers evaluating specific POR investments. Walker et al (2010) found only 24 IAs of POR undertaken by the CGIAR, as compared to more than 700 IAs of technology. This under-investment, in turn, may be retarding institutional learning about how to make POR more effective.

There is a clear need for the further development of IA methods for POR. This need is reinforced by the growing importance of POR within the CGIAR system (now accounting for some 15–20 percent of its total US$1 billion/year budget), and the need to demonstrate that this investment has impact within the 2015–2024 strategic results framework, newly adopted by the CGIAR. As a step in this direction, the workshop was organized to:

1. Define the scope of POR within the context of the CGIAR and the challenges this poses for IA.
2. Take stock of what is already known about policy-oriented research impact assessment (PORIA) in terms of methods, case studies, etc. and of promising new developments in terms

of approaches and methods. Identify the key challenges for conducting PORIA and the knowledge gaps that need to be filled.

3. Develop a typology of POR that can help differentiate methods and approaches that are relevant in different cases. Possible criteria to consider are type of POR, type of policy process, and type of audience. Present good practice or current state of the art for IA in the different types of POR.
4. Explore implications for the selection of intermediate indicators that can help ensure that POR is on track to have impact and contribute to credible evidence bases for subsequent ex-post assessments.
5. Prepare a clear statement that defines realistic expectations for what can and cannot be achieved in evaluating the impact of different types of POR. An important audience is the donor community.

While the workshop scope included a broad range of types of POR, a key focus was on developing findings and recommendations for the types of POR carried out by the CGIAR system and its partners, which is directed toward food security and poverty reduction in developing countries.

The workshop convened about 45 people, including evaluation experts and participants from the academic community, key development assistance agencies, and donors (Annex 1). It was structured around short presentations based on background papers made available before the event, panel presentations, and working group discussions (Annex 2).

Summary of Sessions

Introduction

On behalf of IFPRI's Director General Shenggen Fan, Maximo Torero, director of IFPRI's Markets, Trade and Institutions Division (MTID), officially welcomed the participants to the workshop and highlighted its importance by emphasizing that there has not been as much progress in assessing impacts from policy research as compared to other interventions. More concrete ways of measuring POR influence are needed.

Karen Brooks, director of the CGIAR's Policies, Institutions, and Markets Program, provided a rationale for the workshop in the context of the CGIAR. She noted that the workshop comes at an important time within the CGIAR and similarly in the arena of development-oriented research, where accountability is increasingly demanded. Researchers want clients and partners to understand the impact of policy research so they can better position future research. As a manager, she stressed the value of impact information to steer the investment portfolio in promising directions. In terms of workshop outputs, she was hoping to learn about typologies of POR, recommended approaches for the different types of POR, feedback from the donors on what meets expectations from their side in terms of methods and evidence in PORIA, and how to operationalize recommendations in future research programs.

Peter Hazell, external coordinator for impact assessment at IFPRI, closed the introduction by reviewing the objectives and the agenda of the workshop program.

Taking Stock—What We Know About Assessing the Impact of Policy-Oriented Research

Mitch Renkow presented a background paper he coauthored with Derek Byerlee titled "Assessing the Impacts of Policy Oriented Research: A Stocktaking."[1] A conceptual framework was presented depicting how research may contribute to policy formulation, potentially resulting in a policy outcome and only later in impact. Research itself is multifaceted and may include products such as analyses, capacity building, and workshops. Other inputs to policymaking may by more influential than evidence from research. Furthermore, it is important to recognize that there are many types of agricultural policy outcomes, including laws and regulations, creation of institutions, government investment priorities (e.g. programs), and operational and management innovations.

The next part of the presentation focused on the identification of key challenges for PORIA. A major challenge is ***attribution:*** how much of a policy outcome can be attributed to a specific research program or organization? Attribution issues arise over: (1) apportioning credit among multiple sources of POR outputs, (2) sharing ownership of POR outputs with sponsoring agencies that were seeking to confirm prior views, and (3) unraveling the role of research outputs versus other factors leading to policy outcomes. All are context specific. A good result for attribution might be: "They were part of the conversation." Another difficult challenge is the creation of a ***counterfactual***: What would have happened without the POR? Stakeholder interviews are the main method of assessing the contribution of POR to policy outcomes, but more rigorous methods such as experimental methods and model simulations are sometimes available for quantifying impacts from policy outcomes. ***Time*** is another challenge for PORIA. There may be long lead times before POR outputs lead to policy outcomes, or before policy outcomes lead to impacts. Impact assessments need to allow for these lead times, but the more time that passes, the more difficult it will be to collect firsthand accounts from participants in the policy debate that led to the change. Finally, capturing ***spillovers*** across countries is a challenge when trying to assess the international public goods produced in the CGIAR.

Renkow then reviewed some of the prior experience with PORIA. Two organizations, the CGIAR and the Australian Centre for International Agricultural Research (ACIAR), have conducted quantitative impact evaluations of POR, while IFPRI has conducted a series of primarily qualitative studies of influence. For the most part, these are "supply-led" assessments that start from a particular body of policy research and trace its influences inductively through the impact pathway. This reflects the primacy of the challenge in assessing attribution. "Demand-led" approaches—that is, assessments that start from the impacts of a specific change in policy and work backward to the confluence of all the various causal factors (including the research) that influenced it—are only feasible if consensus exists regarding the degree to which a particular body of POR leads to a particular policy outcome.

Studies that have attempted to quantify impacts show substantial cost-benefit ratios. These benefits accrued within individual countries, and do not capture the much larger benefits that might accrue at international scales from POR, which have yet to be successfully measured.

Renkow concluded with several observations: (1) there are two phases to PORIA, pre-outcome (influence) and post-outcome (impacts); (2) narrative evaluations are seemingly imperative for the pre-outcome phase; (3) small-scale outcomes are likely to be more readily quantified; (4)

[1] https://sites.google.com/a/cgxchange.org/poria-workshop/background-papers

quantitative impact assessment in the CGIAR seems to require an organized push; and (5) well-articulated and attributed outcome assessments may be sufficient.

Discussion Highlights

- Several participants thought it is better to focus on well-developed narratives about policy influence and outcomes than to worry too much about quantifying impacts.
- Outcomes and impacts in the private sector should be considered for several types of POR.
- An important outcome of some POR may be prevention of bad policy choices, in which case developing the counterfactual is especially difficult. A related problem is that some POR could lead to bad policies (the "poison well" problem).
- To understand impact, we need to understand the underlying processes and the role of the CGIAR. Outcomes may differ depending on whether researchers are embedded within country processes and institutions or POR is undertaken from outside.
- The process of conducting POR (how, with whom, communications strategy, etc.) may be as important—in some cases more important—as the research itself for achieving impact. POR that contributes to local capacity for evidence-based policymaking may have powerful longer-term impacts.
- Attribution of ideas is easier if an idea is unique.
- Sometimes influence is direct, but in other cases POR outputs are picked up by others without any effort by, or sometimes even the knowledge of, the POR researchers, making it very difficult to trace influence.
- Impact assessments of POR should feed back to the design of future research.

Non-economist Perspectives on Impact Assessment of Policy-Oriented Research

Robert Paarlberg, a political scientist, made a presentation on lessons from his own evaluative experience and studies from political science. His main arguments are: (1) do not assume social welfare is a priority of governments—their top priority is to remain in power, (2) do not assume knowledge is a priority for governments—they use knowledge only when it is convenient for their preexisting purposes, but active involvement with policymakers can be valued more than the research itself, (3) do not rely on supply orientation to PORIA and instead start with the policy change and examine the likely causes of it, (4) time spent seeking to establish attribution is more useful than establishing a counterfactual, (5) reserve ambitious quantification efforts for site-specific applied research—quantifiable counterfactuals are more easily constructed under these circumstances, and (6) cherry-pick case studies for IA with a clean conscience—selection bias is unavoidable as there is not a universe of cases and assessing the big ones will be most informative.

Ruth Meinzen-Dick, a rural sociologist, followed with a brief commentary based on insights from a workshop on "Approaches and Methods for Policy Process Research," held at IFPRI in Washington, DC, November 18–20, 2013.[2] Impact assessment was not a major topic at the workshop, but some insights were gained. Policy processes are often complex and interactive ("wicked"), in

[2] Workshop details and presentations are available at http://www.pim.cgiar.org/2013/12/03/making-policy-research-conducive-to-policy-change-no-simple-recipe/.

which case it is important to be engaged along the way through action research. From an impact perspective, it is also important to document changes along the way. The impact community noted the usefulness of conducting assessments of changes in knowledge, attitude, and practice as a result of POR. Specific tools can be useful, including discourse analysis and stakeholder analysis, which may reveal improvements in policy processes as a result of research. Historians also have useful methods for documenting changes and causes. It is important to remember that there are multiple realities, which creates challenges in IA studies, and it is important to interpret these different views. Assessing influence is also difficult because policymakers may wish to keep aspects of their decisionmaking confidential and may not want to be quoted. Finally, Meinzen-Dick noted that there is no one-size-fits-all approach to IA. Policymakers may prefer narratives while other stakeholders, such as donors, may give more weight to quantitative results.

Discussion Highlights

- If POR is not influential, then why do donors fund it? Donors are often more influenced by POR than policymakers, and hence may fund POR to develop or confirm their own policy positions. If donor pressure influences policy changes, then the attribution to POR remains indirect, but is potentially very important.
- Policymakers often do not want to reveal how they arrived at a decision. The best evidence of attribution may never be available to a researcher—only to those in the room when decisions were made, often over lengthy periods of time. A successful POR may be one where policymakers take full ownership of the ideas or findings but do not acknowledge the source of influence.
- The best opportunities for policy changes often arise when there is a change of government. In these circumstances, quick advocacy and communications based on existing knowledge may be more influential than starting new POR. Having products ready when such changes take place is important, but predicting them is not easy.
- Cherry-picking studies versus representative samples. One view is that POR is like oil exploration—we should expect several failed wells for each gusher. If so, then it does not make sense to study many failures, but rather to focus on the winners and hope they are successful enough to justify the entire investment program. Studying failures is important for learning about how to do POR more effectively, but it is less important for assessing cost-benefit ratios; we just need to include the cost of the failures. Another approach is to undertake strategic cherry-picking of projects, but as part of a more comprehensive evaluation of a research program.

We need to pay more attention to identifying where the big wins might be through ex ante analysis of new POR investments. It is a matter of planting more cherry trees so there will be cherries to pick later.

A Typology of Policy-Oriented Research

Roger Slade presented a background paper that he coauthored with Peter Hazell titled "Policy Research: The Search for Impact."[3] He highlighted that there are many different types of POR and methods of evaluation need to be adapted to each situation. Some types of POR are amenable to evaluation through randomized trials and econometric approaches, such as some local program

[3] https://sites.google.com/a/cgxchange.org/poria-workshop/background-papers.

implementation mechanisms, whereas others, particularly POR on many sectorwide and macro policy issues, cannot realistically be evaluated with statistical rigor. To help clarify these different types of POR for IA purposes and to guide investor expectations about what is possible when evaluating POR, it is helpful to have a typology of POR based on criteria that capture the challenges involved in assessing their impact. Such a typology could help define the discernable limit of IA for each type of POR (that is, acceptable rules of evidence) and then identify best-practice methods for each type of POR, as well as promising new ones for further development.

POR can be classified in many ways, including by theme, method, subject, and outcome. A focus on the evaluation of outcomes and impacts suggests a typology of POR that is driven by the nature of the outcome sought. Slade suggested a typology that sorts POR into classes based on the intended reach of the POR. Initially this yields three broad categories: country, region, and global. Within these categories, policy-oriented studies may take many forms: **country**—project or program, sector or subsector, and nation or macroeconomy; **region**—multicountry studies, regional policy and policy harmonization, and multicountry networking and capacity building; and **global**—conferences and advocacy, modeling and international agreements, and thematic research.

Each of these types of POR have their own theories of change. Articulating these will assist in identifying an appropriate PORIA approach. Three major challenges in PORIA are: (1) assessing the outcome, impact, and attribution to POR, (2) specifying an appropriate counterfactual, and (3) assessing the welfare change resulting from a policy. Within the CGIAR, the attribution problem may be unsolvable given w that, if one entity did not engage in POR, another would; thus it will be more constructive for the counterfactual to reflect the evidence and not the organization. Further, attribution ought to be based more on plausible arguments than rigorous proof, which is often not available. In terms of impacts, measures of costs should be included, not just benefits. Slade then offered a checklist that can be used for selecting methods. Lastly, he indicated that he is currently reviewing available PORIA studies to look for patterns between the depth of evaluation (in terms of how far down the impact pathway the evaluation proceeds) and type of POR. He had expected to find declining depth to the evaluations as one moves up from POR to support site-level projects or programs, through POR aimed at country, regional, and global issues. But early results show more complex patterns.

Methods for Assessing Impacts from Policy-Oriented Research

Past Case Studies

Jere Behrman presented a case study of an evaluation of IFPRI's contribution to the PROGRESA conditional cash transfer program in Mexico. IFPRI's main role was undertaking impact evaluation studies of PROGRESA interventions. Behrman's evaluation posed questions related to the influence of IFPRI's evaluation of the program—to what extent did the evaluation inform or underpin the new government's decision to continue the program, why was IPFRI selected to conduct research on PROGRESA, how did the results contribute, and were there spillover effects of the research beyond PROGRESA? Challenges faced in this assessment included determining how far downstream benefits should be assessed and whether IFPRI's role should be examined in relation to the policy environment or the policy itself. As many as 39 key informants from various stages of the outcome pathway were interviewed, but not everyone responded or agreed to be interviewed. Obtaining responses from key stakeholders is critical to assessing influence; in this case, it was fortuitous that one of PROGRESA's early leaders was also a researcher and had written documentation citing IFPRI research. In terms of quantifying the impacts, simple cost-benefit analyses were carried out using

very conservative estimates of attribution to IFPRI, which still showed enormous benefits to the research.

Vince Smith provided insights based on a number of assessments he has participated in. For retrospective IAs, the first question to address is why do them. Research organizations want these IAs to help set their own priorities. Pure cherry-picking can be misleading—where practiced, the cost denominator should be spending on all POR, not just the cherry-picked study. Donors want IAs to justify their investments in policy research, which itself may be funded for many reasons ranging from contributing to knowledge creation to having an impact on welfare. Smith listed a sequence of steps normally involved in an assessment, followed by some comments on particular challenges. For attribution, for example, one can employ a variety of methods including interviewing policymakers and other key actors and accessing media information that reflects a project's relevance. He argued that while smaller, location-specific outcomes may be easiest to assess and quantify, more investment is needed in PORIA on broader policy outcomes, such as those from international public good outputs of the CGIAR, challenging as those may be.

Tim Kelley presented a case study that was conducted as part of a SPIA-coordinated study on the impacts of POR.[4] The case study, produced by CIFOR, was on the topic of reducing natural forest degradation by exposing corrupt practices and policies within the pulp and paper sector in Indonesia. The steps used in the methodology were: identify impact hypotheses, determine the overall chronology of events, assess CIFOR's influence and construct counterfactuals through 31 key informant interviews, develop a valuation framework for the impacts resulting from the policy change (an economic surplus approach plus some environmental benefits), and attempt to apportion the benefits to CIFOR. To feed into the PORIA, CIFOR documented media attention to their output. Nongovernmental organizations that used the research results in their own advocacy work were also able to document some changes. In terms of measuring impact against a counterfactual, the study assumed that the main benefits were due to bringing forward a policy change that was likely to happen anyway, resulting in less loss of natural forest and an earlier investment in private plantations.

Challenges of Assessing Influence

Andrew Selee provided insights from working at a policy think tank that were captured in a recent book he authored, *What Should Think Tanks Do? A Strategic Guide for Policy Impact.* A major aim of think tanks is to influence public discourse through ideas. Assessing their influence is challenging from the perspective of defining the output. Ideas themselves are difficult to attribute as ideas typically build on others and are therefore subject to multiple owners. Equally, from a downstream perspective, by the time an idea gain enough traction to influence discourse or policy, it will have been championed by various people along the way. Selee advocated for a flexible approach to assessing impact. One can think like a scientist where it is possible to quantify impacts. In other cases, such as the impact of ideas, it is better to think like a lawyer and put together a story about reasonable doubt or probable cause. This can be aided by the formation of a reference group that adds credibility.

John Young then made several remarks based on his own experiences and those of the Overseas Development Institute (ODI). He noted that ODI has worked on methods for IA and has published some guidelines. Qualitative methods are likely to be important in assessing influence of

[4] http://www.fao.org/docs/eims/upload/256853/Brief%2025(CIFOR)-pr4_enabl.pdf.

research. He provided an example from recent research with a CGIAR center for which an historical timeline was constructed for each different stakeholder, including the research period; then a form of a social network analysis was used to examine how different actors or actions affected others. In another instance involving the CGIAR, targets had been established at the beginning of the research. To establish an inventory of outcomes, researchers used many methods, such as collaborative outcomes reporting, which relied partly on data trawling and social inquiry, followed by integration and analysis. An outcomes panel was formed and a summit workshop held. In terms of quantifying influence, Young mentioned that though this can be done in some cases, it is preferable to demonstrate contribution rather than attribution.

Ruth Meinzen-Dick rounded out this session by noting a few lessons she had gathered over the years. One is to be careful in public language around the issue of influence. While policymakers may want to show that they are considering evidence, they rarely want to admit that they were influenced by external organizations. Therefore policy outcomes should always be viewed as owned by the policymakers. A second lesson is that debates and discussions, apart from direct meetings with policymakers, are important because they increase demand for policy change from different stakeholders. Third, it is important to develop lasting partnerships for policy research and dissemination with national organizations because governments come and go. Further, researchers are rarely the best people to take on the continual advocacy or support roles often needed to see through reforms. Fourth, institutional development and reform, not just policies, are important. Meinzen-Dick echoed some of the previous speakers in advocating a mixed methods approach to constructing a plausible case for influence and use of the term contribution rather than attribution.

Promising New Methods

Annette Brown presented the 3ie approach to impact evaluation and how it is used to make decisions. 3ie uses counterfactual-based evaluation methods in developing countries. One approach is to measure influence, identify a counterfactual, measure against that, and then add up the total impacts. While assessing attribution is challenging, activities to promote influence, such as outcome planning, can make this easier. It is important to build the evaluation into the POR from the beginning. Multiple outcomes as well as multiple products contributing to the outcomes will require more complex evaluation designs. 3ie partners use evaluations to make decisions about whether to expand or close programs or to change designs. Results of the evaluations also influence other programs. In terms of a counterfactual, it may be possible to extrapolate results from well-designed evaluations from pilot projects before scaling up. However, Brown noted that the validity of that approach may not hold if the scaled-up program leads to economywide consequences that were not revealed by the pilot evaluations, or if the management of the program is different, for example if more intense effort is devoted to the pilots. Given these problems, an impact evaluation at the scale of the program, that is, covering its intended geographical reach, is preferable. Finally, although more case studies are needed, better use of systematic reviews for shaping policy is also needed.

Karen Macours began her remarks by noting that there is very little hard evidence on the performance of many technologies or institutional approaches once they are being used by the intended beneficiaries. In many cases, principles of randomized control trials (RCTs) can be used to provide more rigorous evidence on the impact of these interventions. In principle, the heterogeneity of performance can be assessed by stratifying and then randomizing the receipt of the intervention within strata. Macours gave three examples of ongoing studies that integrate randomization the IA design. The first is testing the effect of prices, information, and availability of improved seed on adoption of improved seed in the Democratic Republic of Congo, stratified by gender and market

access. The second is a study of the effect of integrated soil-fertility management on crop yields in Kenya, combining an RCT trial with a researcher-managed trial. The third is testing the use of cattle feed resulting from exposure to a volunteer farmer-trainer model in Uganda, where exposure is randomized. Increased interaction between evaluators and policymakers has been found to increase the salience of research and to feed evidence into policy decisions.

Berk Ozler talked about new methods being used to deal with key challenges in evaluation studies. The first challenge is how to handle spillover and general equilibrium effects of policies or other intervention treatments, effects that are not encountered when conducting small pilot evaluations. The use of different treatment intensities can be used, but this creates more treatment groups, reducing the power of the results for a given amount of funds. A second challenge is identifying a counterfactual for an intervention at the state or country level. The concept of a synthetic control is being tested in which composites of multiple states or countries are used as a comparison. Whether this provides a suitable comparison depends on many factors, including a matched historical trend. A third issue is measurement, which presents many challenges. Good RCT designs may be trumped by poor data quality. For example, in some cases data collection is different for control and treatment groups, which can lead to biases. Finally, meta-analyses can only be done when there is coordination to generate consistent measurements.

Maximo Torero noted that some outputs are intended for academic audiences, and there are good methods for assessing quality and impact of those, such as the use of citations. For outputs intended to feed into other pathways, many issues arise around measurement. Is it impact and/or influence that we want to measure, and at what point in time? The selection of projects to assess is challenging and subject to selection bias. While identifying a counterfactual is a problem, assessing attribution or contribution is less difficult the more unique the research output. Torero then discussed four ideas for improving IA methods: (1) use the best methods possible at small scale as they are simpler to implement at such a scale, (2) use encouragement approaches (where intensity of the information provided is varied randomly to induce different behaviors) in order to be able to assess the marginal impact of information, (3) use models to build counterfactuals and simulate effects without the policy change, and (4) conduct media analysis—mining media data based on phrases or key words to examine whether particular messages are being picked up.

Indicators and Approaches for Tracking Policy Change

Karen Brooks presented the methodology used by the Policies, Institutions, and Markets program of the CGIAR in tracking policy outcomes and impacts. She noted that the portfolio is diverse and, while each thematic area responds to specific problems (conveyed in terms of Intermediate Development Outcomes within the CGIAR), they also come together to jointly contribute to impacts. Because the portfolio is large and diverse in topic and geography, it is not easy to evaluate at the thematic level. However, Brooks recognized the importance of measuring the larger contribution to development outcomes and plans to continue to tackle that challenge. At the activity level, the program uses a template for research leaders, both to plan for and report on outcomes. The template requires researchers to respond to questions regarding what policy the research is aimed at, who used the research, and how it was used. Additionally, researchers answers questions about partners involved in achieving outcomes.

Jim Oehmke presented the indicators and monitoring systems used by the US Agency for International Development's (USAID) Feed the Future Initiative. USAID uses indicators for all grantees. Among those is a set of policy-related indicators, which tracks the number of policies along

a 'stages of policy' framework. At the impact level, USAID focuses on hunger and poverty reduction and therefore analyzes whether there are co-movements in policy indicators and the high-level indicators. A key interest is to understand if policy change is systemic. The countries in the initiative developed self-identified many priority policy reforms. So USAID has invested in "process" reforms, including architecture of policy formulation and implementation. USAID is now test-piloting methods to monitor both systemic change and policy reforms, and is also interested in learning how the agency's support of policy research contributes to policy change. USAID is currently funding many impact studies across the Feed the Future portfolio.

John Underwood discussed how the Millennium Challenge Corporation (MCC) has tried to measure the impact of policy change. MCC started by identifying countries with a good policy environment as its targets, using indicators around several thematic areas. This produced an interesting behavioral effect, namely that countries work to improve their indicators in order to qualify for MCC assistance. MCC continuously works to improve its own indicators, which are tracked by country recipients. Much of MCC's work related to IA is of an ex ante nature, to help prioritize areas of investment in countries. It uses methods such as growth diagnostics (using the Hausmann-Rodrick-Valesco framework [Hausmann et al. 2008]) and cost-benefit analyses to provide inputs to investment decisions which may include policy and institutional reforms. MCC also undertakes process evaluations and impact evaluations. Now that MCC programs have been running for more than 10 years, many studies will be undertaken. The intention is to apply sound methods with proper designs and counterfactuals.

Working Group Session

Following the presentations, working groups were formed to analyze PORIA issues and challenges from the perspective of the basic typology proposed in Slade's presentation. Groups were organized to discuss country-level, regional-level, and global-level policy-oriented research. They were tasked to answer these five questions:

1. For the type of POR assigned to your group, fill in the relevant row of the table for each of 2 or 3 important lines of research.
2. What general conclusions can be drawn from your answer to question 1 about the pathways to impact? And about the challenges facing and the best available methods for undertaking IAs for your assigned type of POR?
3. What are reasonable expectations about the standards of evidence?
4. What are the right messages to convey to CGIAR donors and investors about assessing the impact of your assigned type of POR at a time when they are asking for rigorous IA of their investments?
5. How would you use your IA design for management purposes, that is, at what point might informed decisions be made about whether or not to continue with a line of research?

The table to be filled in by the groups is shown below:

Type of Research Activity in Agriculture and Rural Development		Elements of Impact Assessment			Challenges to Measuring Impact
		Theory of Change or Impact Pathway	Expected Research Outcome or Destination on Impact Pathway	Possible Methods	
Country	Project and/or program				
	Sector (& subsector)				
	Nation (the macroeconomy)				
Region	Multicountry studies seeking IPGs				
	Regional policy & policy harmonization				
	Networking & regional capacity building				
Global	Conferences & advocacy				
	Modeling & international agreements				
	Thematic				

Report Back

There was unresolved debate about the need for a typology of POR to guide IA, although the distinction between country, regional, and global impacts was considered useful for capturing aspects of impact pathways that affect the various types of IAs. However, there was less agreement on the further subdivisions presented in the table. Some specific comments were:

- Capacity building should be included in all types of POR, not just regional.
- Greater consideration should be given to classifying on the basis of research outcomes rather than impacts, as in Renkow's presentation. Outcomes could be given more weight when disaggregating POR within the country, regional, and global levels.
- The typology works better for projects than for research programs or institutions. Not only is there an aggregation problem over mixed types of POR within a program or institution, but one might also expect variation in how far down the impact pathway individual projects are pushed for achieving impact (e.g. the more promising ones are pushed further by a program or institution).
- For ex-ante analysis, a typology might need to give more attention to the different stages of the impact pathway, such as ideas/discourse, policy change, policy implementation, and impact, as in the ODI typology presented by Young.

There was a lot of discussion about tracking policy influence and impact at country levels. Some key observations were:

- For tracking policy influence over time it helps to document 'event' changes, including who was involved, in what ways, when, etc. This can help create a plausible narrative.
- Creating a counterfactual for policy influence is very difficult. Perhaps little more can be done than undertaking a thought experiment, or in some cases picking a comparator region or country. Even project-oriented POR where impact can be assessed through rigorous methods (e.g. social protection projects) faces the same challenges when assessing the influence of the POR itself.
- Successful ex-post IAs tend to focus on specific policy changes.
- Developing an ex-ante theory of change (TOC) provides a helpful tool for tracking and assessing policy influence. It pays to think through which policies the POR is intended to influence, the decisionmakers to be influenced, how the influence is to be achieved, how progress can be assessed, etc. Many donors now insist on this kind of ex-ante analysis.
- Some participants felt strongly that comprehensive ex-ante IAs of projects were also worthwhile. Donors and research managers can use them to help establish research priorities and funding allocations, and they provide detailed benchmarks against which actual project performance can be assessed. An ex-ante analysis also provides indications of longer-term impacts that may be hard to capture within the timeframe of most ex-post IAs.
- Donors and researchers often value different indicators. For example, donors prefer direct measures of poverty reduction or income generation rather than cost-benefit ratios or internal rates of return.

On the challenge of assessing cross-country spillovers (capturing the international public goods), it was noted that impacts can sometimes be simulated with global and regional models (e.g. IFPRI's IMPACT model), which can also be used to create counterfactuals. But assessing policy influence is especially challenging, and is typically limited to aggregating narratives from a few case study countries and interviewing international stakeholders (donors, multinational agencies, and banks) about impacts on their agendas. Such narratives inevitably miss out on many spillover benefits from POR. Although difficult to measure, it was generally agreed that some types of POR have the potential to generate very significant spillover benefits, much larger than most project- or country-focused POR.

Summary Session

A panel was constituted of individuals with long experience in IA to provide some final thoughts after listening to the presentations and discussions over the previous day and a half (Tom Tomich, Jere Behrman, Keith Fuglie, and Robert Paarlberg). A few highlights from their remarks include: (1) there are well-defined rigorous methods for PORIA with respect to public projects or programs but not for national policies, so continuing to develop alternative approaches for assessing POR will help advance IA studies, (2) if we want to have successes to evaluate, we need to make sure we invest more in planning around TOC, (3) whether we use the term attribution, contribution, or influence, there is an implication of causality that requires definition and assessment, (4) we should continue to devote resources to developing sound counterfactuals, (5) researchers should track inputs, outputs,

immediate use of outputs, and policy objectives on a routine basis and then selectively measure influence of outputs on policy decisions and the effect of those decisions on policy impacts, (6) cherry-picking of case studies, whether on the basis of successes or information value, does have merit, particularly given the limited resources for PORIA, and (7) there should be more concerted effort to quantify impacts resulting from researcher-influenced policy reforms.

Donors' Perspectives

Just prior to closing, the workshop heard remarks from members of the donor community who were present at the workshop (Greg Traxler, Jim Oehmke, Rob Bertram, and Ejaz Qureshi). All were agreed on the importance of impact studies for justifying POR funding within their organizations, and all wanted to see more PORIA studies. Quantitative assessments, which allow for comparison with impacts from other types of investment such as plant breeding, would be nice. But what donors most need are evidence-based narratives that are convincing within their agencies, bearing in mind that most of their staff are not economists. Another way to convince their agencies is to have clear ex-ante impact pathways for POR projects, showing what policies are to be changed, the key partners, and how the team expects to influence the policy process and within what timeframe. Impact pathways also provide a useful reporting framework. Another area of agreement was that showing impact within countries is a top priority, and that international public goods are of little value if they do not also translate into country impacts. Other points that came up were: (1) capacity building is important to donors, and they would like to see more evaluation of capacity-building work within POR projects, (2) the donor is not the main client for many PORIA studies—national governments are key clients for evaluation of policy impacts, (3) although donors recognize that their interests in impact can distort research incentives to a short-term focus, nevertheless demonstrating impacts will remain important, and (4) good impact studies are more important than monitoring results.

2. CHALLENGES OF PORIA: BREAKING DOWN THE ELEMENTS

As a first step in drawing lessons and recommendations from the workshop for the CGIAR, this section attempts to synthesize key findings regarding the elements to be considered in an IA of policy-oriented research, starting with what constitutes a policy outcome and working through design and other methodological issues related to PORIA. Some of the challenges discussed below were also noted in the review of CGIAR PORIA by Raitzer and Ryan (2008), and therefore we try to analyze options for responding to the challenges.

Policy Outcomes and Impacts

We begin with policy outcomes and impacts because they provide the objectives that should drive both the POR and its subsequent IA.

Policy Outcomes

The term 'policy' in relation to agriculture can be taken to mean different things, so it is useful to discuss its scope for this paper. First is the issue of *whose* policies are important for the CGIAR. The most important policymakers are national governments in developing countries because they directly shape the environment of the key beneficiaries of CGIAR research—smallholder farmers and the poor in developing countries. Public agencies at the global and regional level are also important policymaking groups for POR in the CGIAR. With decentralization, policies at local levels are increasingly important; however, pathways to widespread dissemination of local policy will almost always pass through the national level.

Although the primary focus for the CGIAR should be on government policymakers, other key players should not be ignored. Many nongovernmental organizations play important roles in advocating for policy change, and many are engaged in grassroots development and relief aid. The CGIAR can help inform these activities through POR. Many organizations in the private and nonprofit sectors are key partners of agricultural research institutes in technology development and dissemination and can also be influential in the policy arena. The influence policy both through advocacy with governments and by changing their own behaviors. For example, private-sector companies may develop zero-deforestation objectives and follow up through adherence to certification mechanisms. Another example is the increasing role of supermarkets and large agribusiness firms in agricultural value chains. These companies set their own policies about quality and safety standards and the kinds of farmers they wish to provide with technologies, credit, and market access. Indeed, it can be argued that in the cases of deforestation and more integrated value chains, changing policies of multiple actors is essential for significant and lasting outcomes that benefit small farmers and the poor.

'Policy' is often defined as a rule that regulates behavior. However, a rule should not be narrowly interpreted as a law or regulation. In his presentation, Renkow noted five different types of policy outcomes of relevance to international agricultural research:

- laws and regulations at global, regional, national, and local levels

- creation of institutions
- government investment priorities and budget allocations
- operational and management innovations for government agencies and projects
- declarations, agreements, and strategies reached among parties, such as at major policy conferences

A slightly different classification of policy outcomes was introduced at the workshop and draws from Young et al. (2014):

- interest of key stakeholders; getting issues on to the policy agenda
- public opinion
- capacity and engagement of other actors
- change in discourse among policy actors and commentators
- improvements in policymaking procedure or process
- change (or no change) in policy content
- behavior change for effective implementation
- networks and systems for supporting delivery
- relationships between actors

It is not our intention to develop a definitive list, but rather to reinforce the notion that there are many types of policy outcomes that are worthwhile objectives of POR, and so we should maintain a broad definition of the term 'policy' in the CGIAR.

Documenting change in a policy or program is relatively straightforward in terms of what changed, by whom, and when. What changed can be something simple, such as a tax rate, or something more complex, such as a whole-sector reform involving several institutions and their rules. Such changes can and should be tracked by researchers themselves as they take place. It is more challenging to describe policy change in normative terms, such as whether it is more pro-poor, more gender-equitable, or more evidence-based. These types of assessments will typically require more analytical effort and sometimes a longer gestation period before evaluation can be carried out.

It is important to note that policy *influence* need not result in an observable change in policy—influence may help to retain a current policy over an alternative option that is being considered. So policy *outcome* should be taken to include a change in policy *in relation to its trajectory in the absence of the POR.* Maintaining the policy status quo can be as important and effective as changing policies. Furthermore, it is important to note that POR may produce misguided advice, for example, if implications are drawn from one context and applied in another. The assessment of impacts, discussed below, is a means by which the appropriateness of the policy advice can be inferred.

Impacts of Policy Outcomes

Understanding policy influence can address questions of relevance of POR. However, to assess the value of investing in POR, it is also important to measure the impacts that occurred after the policy outcome. Potential agricultural policy impacts are wide, covering food, nutrition, poverty, income, attention to socially disadvantaged groups, environment, and access to natural resources, to name a few. Regarding selection of policy impacts to measure, a few remarks can be made.

Most POR will be expected to have *direct* outcomes within its TOC, such as changes in laws, policies, or institutions, and these in turn will be expected to lead to *indirect* or knock-on impacts, such as changes in prices or input supplies, which in turn might lead to further rounds of impacts, such as increased food production, incomes, and consumption. A first priority should be to measure the initial direct impacts, since if these are insignificant there would be no reason to measure further indirect effects when evaluating the POR. The direct impacts may also have negative consequences. This may arise if research leads to an ill-advised policy outcome or possibly if the new policy is poorly implemented.

While evaluating research on the basis of its intended TOC is an appropriate approach, this may be more relevant for the evaluation of influence on policy (i.e. did the evidence influence the policies it intended to?) than to evaluation of policy impacts. This is because many other factors that lie largely beyond the influence of the research team can influence policy changes and their impacts. Policy decisionmakers will also have broader objectives than those outlined by researchers in their TOC.

Assuming that policy influence can be demonstrated, it may be possible to calculate the value or returns to policy research, using standard measures such as cost-benefit ratio or internal rate of return. Quantification of the impacts of a policy outcome presents a challenge, specifically generating the required data, identifying a suitable counterfactual, and defining the large number of assumptions that are required in such an analysis. Even in cases where quantification is possible, there is still the challenge of attributing a share of the benefits to the POR. In their review of IAs based on CGIAR policy-oriented research, Walker et al. (2010) found that only 3 of 24 studies attempted to empirically quantify the impacts.

Ryan (1999) was the first of these three, giving a quantitative estimate of a rice policy outcome in Vietnam, in which the emphasis was on the research contribution by a specific organization (IFPRI), which resulted in a more rapid policy outcome than if an alternative provider had later provided needed POR. The author was able to use a national-level agricultural model developed for Vietnam to contrast outcomes of the old and new policies to arrive at estimates of benefits over the limited period of time until the policy outcome would have occurred without IFPRI's POR. Babu (2000) employed the same approach of valuing IFPRI's contribution as an earlier implementation of a policy outcome in the case of a social protection program in Bangladesh. He also attempted to estimate the share of IFPRI's contribution in conjunction with estimating the overall longer-term benefits from the policy change. For both, Babu (2000) draws on a country-level model developed for analyzing policy options in Bangladesh. The final study noted in the Walker et al. (2010) review was the study by Ryan and Meng (2004) analyzing the impacts of IFPRI's research on the Food for Education Program in Bangladesh. The authors combined measurements of the effect of the program on schooling participation from an impact evaluation with the effect of school achievement on incomes to calculate the program impact. IFPRI's share was attributed based on a faster implementation of the policy decision.

Economic surplus models have been used in calculating the net economic benefits from policy change by Kaitibie et al. (2010) to analyze dairy sector policy change in Kenya, and by Ahmed et al. (2010) to estimate the effects of a fertilizer allocation program change. Raitzer (2010) applies an economic surplus model, but also captures values of environmental services from changes in deforestation from change in policy and implementation on sourcing of materials for the Indonesia pulp and paper industry. Others have used analyses from large household studies to help estimate economywide benefits accruing to smallholder farmers (for example, Templeton and Jamora 2010).

Given that resources will never be sufficient for a program or organization to evaluate policy impacts for all of its POR, the selection of case studies for such analysis becomes important. This is dealt with in Section 3.

Typology of Policy-Oriented Research

Although there was broad support at the workshop for developing a POR typology, there was less agreement on how the typology should be constructed. Difficulties arise if one attempts to classify POR on the basis of the type of research activity, since there are many ways of classifying these and none are obviously helpful for guiding IA work. A more promising approach is to classify POR on the basis of the scope of its intended outcomes or impact. Some POR is targeted at project-level or specific country problems, such as improving a safety net program or an agricultural extension system. In these cases the impact pathways are relatively short, the relevant policymakers and beneficiaries clearly identified, and the primary impacts can be captured at a local or country level. Other types of POR are targeted at multicountry or global problems, such as regional or global trade arrangements and global warming. These types of POR have longer and more complex impact pathways in terms of the numbers and types of stakeholders and beneficiaries involved, the size of potential indirect and cross-country spillover benefits, and the time it takes to realize change. Classifying POR outcomes according to whether they are discernable and possibly measurable at the country, multicountry, and global levels would therefore seem to provide a useful framework for thinking about TOC and IA methods and for clustering results and lessons from IA studies.

Classifying POR outcomes does not always map neatly into different types of POR activity. For example, while some types of research activities are conducted within individual countries, these may be part of larger research programs that use a set of country case studies to extrapolate findings to multicountry levels. In these cases, the POR should not only generate outcomes in the case study countries, but should also lead to spillover benefits to other countries facing a similar problem. Another example is POR that addresses regional or global problems. The primary objective of such POR is usually to generate regional or global outcomes (international public goods), but additionally, this POR may lead to trickle-down outcomes that can be assessed within individual countries. As a result, many POR activities really need to be evaluated at two or more levels, although sadly this is rarely done in practice.

Further subdivision of POR should reflect the purpose of the IA. For example, a research manager might want to subdivide on the basis of type of activity or type of output, because knowing the likelihood of each to have an impact may be helpful in allocating staff and financial resources. Donors might be more interested in subdivisions based on outcomes, since knowing which outcomes have the greatest impact could help prioritize the types of POR useful for supporting their other investments. If the typology is to be used for ex ante purposes, then it may be useful to subdivide on

outputs and outcomes and to develop a TOC that shows how the activities and outputs will lead to the desired outcomes.

In short, further subdivision of POR beyond country, regional, or global focus needs to be tailored to the objectives of those undertaking IAs.

Theory of Change (TOC)

A TOC begins with an objective in mind and can be defined as "a specific and measurable description of a social change initiative that forms the basis for strategic planning, on-going decision-making and evaluation (www.theoryofchange.org)". Although similar to an impact pathway, the use of a TOC normally implies a more detailed description of the process(es). A typical impact pathway, as used in the CGIAR, indicates the target impacts, the required outcomes to achieve the impacts, the required outputs to achieve the outcomes, and the required activities to achieve the outputs. A TOC will go further to explain *how* the connections between outputs, outcomes, and impacts are to be achieved (that is, what is contained in the arrows of an impact pathway) and will identify nonlinearities in the pathway.

While many organizations or other stakeholders in the policy process may develop explicit or implicit TOCs for their own missions, the important TOC in the context of assessing the impact of policy-oriented research is the one developed by the research team. Ideally, this TOC will explicitly indicate the policy outcome and its intended impacts; the key decisionmakers involved in the policy outcome; the behaviors or attitudes that need to be changed; the partners that need to be engaged; the capacity that needs to be strengthened; the evidence that is most useful to the process; and the communication methods best suited to disseminate the evidence. The more explicit the TOC, the easier it is for an evaluator to identify the stakeholders to interview and the questions to ask.

A TOC should not be viewed as a rigid set of hypothesized relationships and activity plans but one that is dynamic, responding to reviews of activities and their effects (Young et al. 2014). This acknowledges that linearity assumptions embedded in a typical impact pathway diagram rarely reflect reality, as pointed out by Renkow and Byerlee (2014). A revised and final version of the TOC should be the one used in an evaluation. The demonstration of an evolution of the TOC would itself be useful evidence for an evaluator looking at the management of the POR.

The starting point for a TOC is the POR objectives. These can take on any timeframe, but for the TOC to be meaningful as a management tool, at least some of the objectives should be achievable within the planned period of engagement of the project, program, or organization. Objectives may be based on a certain geographical area (e.g. country), policymaking body (e.g. ministry of agriculture), or thematic area (e.g. land policy). Often, a particular project may jointly specify all three of these dimensions.

No matter how well developed and planned are expectations about policy processes and opportunities for influence, Renkow and Byerlee (2014) remind us that unexpected opportunities for influence will always arise and should be seized. This has implications for the POR strategy adopted by an organization or program. For example, a strategy may suggest continual updating of datasets and models with which to react quickly when opportunities arise or maintenance of partnerships with policy influencers/shapers who are in better positions to link research findings into policy processes.

Outputs—Evidence Feeding into Policy Processes

A fundamental assumption underlying POR is that scientific evidence has a value in influencing policy decisions. This was called into question on several occasions at the workshop, and it was pointed out that political science analysis does not often find scientific evidence to be a significant factor in policy formulation (see also Contandriopoulos et al. 2010 for a study in the health policy sector). Whether the evidence will have an effect may depend on factors such as the quality of the research, its timeliness, and the way in which it is communicated. The latter two issues are addressed in more detail in the following section on science–policy linkages, but here we take up the issue of science quality. Within the CGIAR and as discussed in the workshop, three dimensions of research quality are important: its salience, its credibility, and its legitimacy. All three are important in all research endeavors, but for POR the legitimacy dimension assumes a high degree of prominence.

Salience, or relevance, has several interesting nuances with respect to POR. Salience in POR is taken to mean that the research specifically aims to address the questions and concerns of policymakers. Other dimensions of salience are also important, such as the scale and scope of the research vis-à-vis the policy question. Case studies in a few villages are not as salient and useful as nationally representative studies for national policy decisions. Studies on impacts of policies in one country will be more relevant to other countries with similar contextual factors. Salience need not always be judged by policymakers—salience of a research topic to current debates in the media is also important when the POR goal is to inform discourse and influence agendas.

Credibility is an obvious characteristic of quality. Credibility means that the evidence should be generated using appropriate research methods of data collection and analysis. Although academic publications are a poor way to communicate to policymakers, they are important in the policy-influencing process by forcing a review of methods, results, and interpretations on the part of researchers.

If salience refers mainly to the question of *what* and *when* research is produced, and credibility to the question of *how* it is produced, then legitimacy refers more to the question of *who* produced that evidence. Although research organizations all claim to be objective and neutral, they are not always perceived to be so by all stakeholders in policy processes. As a result, evidence meriting high marks for salience and credibility may nonetheless be marginalized in policy debates due to concerns over legitimacy. This naturally has implications for partnerships in POR and is one of the topics discussed in the section below on science-to-policy linkages.

With this background, PORIA evaluations should devote attention to the assessment of research outputs and pay attention to quality including aspects of salience, credibility, and legitimacy. These can be assessed both from interviews and through quantitative methods such as citation and download counts (e.g. for credibility). This part of the evaluation helps in the assessment of the influence of the research in a policy outcome. It also is important in that it may lead to the identification of other influences from the research, for example citations in later research by others, which may in turn support policy processes elsewhere. Evaluations should also delve into the inputs used to produce those outputs in order to assess how well resources are managed to achieve the POR outputs (Hovland 2007).

Science-to-Policy Linkages

A key goal of PORIA evaluations is to identify recommendations on how research could have had more influence in the past and how it could be more influential in the future. In cases where scientific

evidence is used in policy decisionmaking, research quality would be hypothesized to be positively related to influence. However, the methods used to forge linkages to policymakers (or influencers) as well as the types of dissemination or outreach activities can also be very important in achieving influence. These types of activities inevitably involve active partnerships, building capacity, and using effective communications. A well-constructed TOC will specify these plans and activities.

The PORIA workshop did not devote much time to the various methods of policy outreach, but affirmed its importance in influencing policy. Recent studies have analyzed the effectiveness of different approaches and activities of researchers, including Sylva et al. (2007), Young et al. (2014), and Masset et al. (2013). This area of research is rather limited, with very few explicitly designed tests of alternative partnership, capacity, or communication methods.

Learning from the TOC

The most important use of a TOC for a research manager is to provide a framework for learning, by continually assessing whether planned actions are leading to the intended consequences. If not, the research team should consider adjustments to improve effectiveness in reaching various milestones in the policy support/influence process. The TOC is also important for longer-term evaluation of the influence and impact of a policy research program as evaluators will structure their review around the outputs and actions described by the TOC.

Selection of Case Studies

An important issue not adequately resolved at the workshop is how research areas or projects should be selected for IA. There are basically two approaches. One is a cherry-picking approach that focuses on the most successful projects, because these might be enough to justify an entire research program or institution, as with oil exploration where the occasional gusher justifies the cost of all the trial wells. With this approach, the impact of failed projects does not need to be assessed because there will not be anything to measure. However, it is important to include the costs of all the research projects in a portfolio, including the failures, when calculating a cost-benefit ratio or rate of return.

An alternative approach is to select a representative sample (not necessarily random) of projects. Arguments in favor of a sampling approach include:

- POR is not like drilling for oil—chance should play less of a role in choosing successful POR than in finding oil wells. Good ex ante impact analyses can help ensure that more POR projects are successful.
- While it is reasonable to expect some POR to fail, it is useful to know why failures occurred (was it due to poor design, poor management, or changed circumstances?). Success or failure, picking a representative sample can make good sense.
- While cherry-picking may be adequate to justify overall investment in POR, it will probably not be sufficient to inform decisions about which types of POR to invest in, that is, which are more likely to lead to policy outcomes and impacts.

When evaluating an entire research program or institution, a pragmatic compromise is to assess the relevance, outputs, and influence of a high proportion of projects, and then select a few

case studies for in-depth analysis. This is better than reporting only a few cherries. It does not rule out a cherry-picking approach, but at least the success stories will be put in a broader context.

Influence / Contribution / Attribution

It was clear from the workshop discussion that the terms influence, attribution, and contribution overlapped conceptually. Definitions of 'influence' and 'attribution' vary but almost always include a reference to causality. 'Contribution' was more ambiguous, meaning the same as attribution to some but implying a weaker notion of "providing input" or "playing a role" to others. There was agreement that making attributions to specific organizations does not serve partnerships well and good practice would be to jointly recognize the inputs that all stakeholders provide to the policy outcome. Furthermore, strong claims of influence may not be received well by the policymakers themselves who rightly will guard against claims of caving to external interests. In the jargon of the CGIAR, the term 'contribution'—in the sense of playing a role—is increasingly preferred to describe the involvement of research organizations in influencing policy outcomes. Later in this section, we discuss the issue of the importance of the contribution.

There was also discussion about whether we are seeking to assess the contribution to a policy outcome of an organization or of research output. Both can be important, and both are challenging because often multiple organizations or interest groups are active in promoting multiple pieces of evidence or advocacy in a policy process. However, for learning and replication by others, the more important unit of observation is arguably the evidence, because it is the evidence that can be replicated by others to influence policy. Focusing on the evidence rather than the organization or individual putting forth the evidence is useful in that references embedded in evidence can show transparent links to earlier work or ideas that were produced by different organizations or emerged through interactions. The question of whether to focus on the organization or the evidence also has implications if counterfactuals are to be used in assessing the contribution of research. In the emphasis is on evidence, the counterfactual is the absence of the evidence and not the organization.

In assessing influence, an important distinction should be made between the influence of research on the *policy process* resulting in a policy outcome (often in the form of a changed policy, regulation, or government program) and the downstream impact of the *policy outcome* (on developmental objectives such as reduced malnutrition). Workshop participants agreed that researchers should be accountable for making a contribution toward policy outcomes, but that that it makes less sense to assess influence of research on impacts farther downstream. In the case of the policy impact, influence would be traceable back to the policy outcome and no further. The evaluator then would attempt two separate assessments, the first to uncover the influence of policy research on a policy outcome, disentangling this from other influences, and the second to discern the contribution of the policy outcome to downstream impacts, disentangling this from other factors affecting those impacts. There is no need to show direct linkages between research and downstream impacts as those are often separated by many intermediate steps and long periods of time. In fact, if influence can be demonstrated in a timely way, that would inform the selection of suitable policy outcomes, in terms of impacts, for subsequent assessment.

The accountability and learning aspects of IAs should be highlighted. Accountability entails collecting evidence to demonstrate to funders and other global stakeholders that investments in POR are leading to the expected tangible policy outcomes. The learning component is largely internal, helping CGIAR researchers understand how policy research and dissemination can be

improved in order to increase its contribution to policy discourse and change. Typically, the learning element is ongoing and involves data collection and analysis by the researchers themselves. Accountability may follow the same track, but at some point, the objective assessment of influence on policy outcomes will require external assessment or validation.

Generating evidence on the dissemination of outputs to policy discourses and processes is often relatively easy to do. Researchers themselves often monitor this through documentation of meetings held with policymakers, tracking downloads of policy studies and briefs, and the like. With this documentation, it is a straightforward task for researchers to articulate how they engaged with a policymaking process. However, describing how research was used, and hence taking a measure of its importance, can be quite challenging. As noted by Kydd (2015), the policymaker's or other users' encounter with evidence is rarely observed and is difficult to reconstruct. It may be particularly challenging to quantify this influence beyond a classification of whether the evidence had a major, moderate, or minor influence.

People involved in policymaking processes may sometimes make unsolicited remarks recognizing the contribution of certain evidence or of organizations providing the evidence in interviews or presentations. More commonly, such evaluative statements need to be prompted by an interviewer or surveyor. The assessment of influence or importance of evidence is subjective and, to avoid partiality, the interviewer or evaluator should not be part of the policy research team or their organizations.

Indicators

Although not much discussed at the workshop, we provide here some brief remarks on the identification of indicators and their ease of measurement of the various stages or elements of a TOC, from research outputs to policy impacts.

Outputs. These include completion of POR studies, briefs, and workshops to name a few. These are straightforward to track and are typically monitored by checking them off when completed (a simple yes or no). However, the proper tagging of outputs to particular research projects and programs is often overlooked. For assessment of research focusing on the impact of projects or programs, this gap makes it difficult for evaluators to assemble relevant bodies of work.

Quality of outputs. The quality of outputs is harder to assess and monitor. However, mechanisms such as citations and impact factors exist to assess the credibility aspect of quality. For salience, citations are somewhat useful, but salience and legitimacy aspects of quality are subjective in nature and their evaluation may have to rely mainly on assessments by evaluators and interviewees. Furthermore, some aspects of quality are best judged only after time passes, as part of a wider body of outputs.

Use of outputs in policy discourse and formulation. These are likely to be binary indicators identifying which outputs have been used in specified ways and/or by specified users. This can be easy to document in cases where researchers are present and capture the proceedings of policy discussions. Documenting the minutes of meetings or workshops and participants is fairly straightforward. In addition, background policy papers used in discourses that research outputs are valuable sources of evidence. But outputs may be used in discourse or meetings where researchers

are not present or research outputs not cited, and it is not simple to ensure that third parties routinely document the use of outputs.

Policy outcome/change. Describing the specific policy change is straightforward where the former policy and current policy (or strategy, regulation, or program) are both published. It is more difficult if there is not a neat mapping from the old position to the new one or if the policy outcome was not a formal change in text, as is the case with the prevention of a change in policy (in which case the old is the same as the new) or a change in mindsets or processes. Even where text has changed, it may not be easy to access. This can be true for government policymakers as well as for other organizations such as private-sector firms. There are many ways to document outcomes, ranging from simple outcome or impact logs (Young et al. 2014) to more structured documentation, as was used previously in the CGIAR and described in Section 3 below.

Influence of outputs on policy outcome/change. Influence of evidence is difficult to assess for several reasons: (1) influence of specific research may not be acknowledged by policy stakeholders, either because they knowingly wish to conceal the information or because they are unaware of the original source of information they do receive (Kydd 2015); (2) measures of influence are subjective and thus may be viewed differently by different stakeholders; and (3) the assessment of influence by a single stakeholder may also vary over time as he or she learns more about the decisionmaking process. Influence must be assessed by an external evaluator for the results to be credible. Influence can be assessed by aggregating perceptions of numerous key stakeholders, perhaps as narratives (qualitative) or averaging ratings or scores (quantitative). There are other useful methods for assessing influence, contribution, or attribution, including contribution analysis and general elimination methodology (see Mayne 2012 and Young et al. 2014).

Normative evaluation of policy outcome/change. Even when it is possible to show the changes in a policy, it is not always easy to describe whether or to what degree the changes are aligned with particular objectives. For example, whether and by how much a new policy better addresses the needs of smallholders may not be straightforward to discern when the policy change is multifaceted. Nor is it clear that policy researchers need or should be trying to measure change along such normative lines. However, the degree of policy alignment with POR recommendations may be a critical assessment in determining influence or use of evidence. If policy change appears to be moving in directions contrary to the scientific evidence provided, that would imply changes are needed to POR for it to be effective.

Special case of capacity as a policy outcome. Some workshop participants noted that strengthened capacity could be an important longer-term POR outcome. Capacity strengthening outcomes may include both researchers and research organizations with enhanced national capacity for design and implementation of policy research, and institutions and individuals with greater interest and capacity in the use of evidence for policymaking. These outcomes are usually a secondary aim of research whose primary objective is a shorter-term policy achievement. If capacity is strengthened, then greater medium- to long-term effects on policy outcomes and impacts are possible. Of course, measuring capacity of any institution is tricky. In policymaking institutions, measuring capacity is particularly challenging given high rates of turnover of the more senior staff in those agencies.

Policy implementation. Implementation of a policy outcome is an important consequent outcome to assess. In cases where the aim of policy research is to stimulate implementation of policy rather than a change in the policy itself, the PORIA will focus on changes in implementation rather

than in policy. Implementation could be measured by appearance of new programs or initiatives or increased government expenditure in existing programs, among other indicators. Measuring implementation change varies in level of difficulty. For example, government spending by broad category is easy to access, but not so for finer details, such as the level of spending on specific commodities or natural resource management practices.

Policy impacts. In measuring impacts, workshop participants agreed that it is useful to summarize policy impacts into an overall quantitative measure of "benefits." This measure, along with a measure of costs, can be used to derive indicators such as net present value, cost-benefit ratio, or internal rate of return. Specific policy impacts, such as poverty reduction or changes in agricultural production or food consumption, may also be appropriate to measure depending on the objectives of the POR.

The scope or scale and timing of expected policy impacts will affect the choice of assessment indicators. The consequence of changes in national-level policies may be monitored at least partly through existing mechanisms for tracking of indicators at country level. Policies that may affect only certain geographies, such as dryland areas, may require new data collection representative of that geography. Policies directed at certain population segments, such as rural populations or women, also may require the creation of new datasets.

As for the timing of effect, some policies may be targeted to affect immediate consumption, such as safety net programs, and have immediate impacts; on the other hand, policies designed to reduce land degradation, such as tree planting programs or conservation agriculture, will have a much longer timeframe. For these impacts with a longer timeframe, it is useful to identify milestones that can show progress toward the ultimate objective within a timeframe that is useful to the research program. In the example given above, the adoption of sustainable land management practices would be such a milestone towards an ultimate indicator on land degradation.

Methods

The workshop presented and discussed advantages and limitations of different PORIA methods, some novel techniques for overcoming particular challenges, and ways to combine various methods and approaches for two key segments of PORIA—assessment of the research-to-policy-outcome linkage (influence) and the policy-outcome-to-impact linkage.

Overall Approach to PORIA

There was a broad consensus that PORIA is best done using mixed methods. Using qualitative methods that assess research influence is essential, and combining qualitative and quantitative methods that assess policy impact is desirable. Given the variation in research outputs, intended outcomes, and availability of evidence, there is no single impact assessment approach that can be applied to all cases. As Masset et al. (2011) emphasize, "There is no standard practice for the evaluation of research projects and every evaluation strategy should be designed on a case-by-case basis."

The table below shows how qualitative and quantitative methods can be combined at various stages of an impact pathway to enhance the final PORIA study. Useful methods throughout the process include: (1) developing of a TOC, (2) monitoring the use of POR outputs, (3) documenting the details of policy outcomes, (4) engaging third parties in assessing influence, (5) collecting

quantitative baselines on policy impact variables where policy outcomes have occurred, and (6) combining several methods to understand policy impacts and the impact pathway. The table below was designed for a PORIA that will assess the impact of a defined set of research outputs. PORIA studies that take a policy outcome as a starting point will require a different framework, but will include many of same the elements. For example, contribution analysis will attempt to assess influence and attribution of many actors and ideas in addition to scientific organizations and outputs, and in doing so will seek evidence of a TOC (Mayne 2012).

Steps in an Approach

Stage of POR / PORIA	PORIA method
Start of POR research	Determine major outcomes and impacts to be targeted Develop a theory of change
During POR research	Track outreach activities Accumulate evidence for use of outputs in policy decisionmaking (compile documents, conduct strategic interviews) POR may itself be an evaluation of a pilot policy intervention and results may be useful to ex post evaluation
Policy outcome	Document the change that had been made External professional to assess influence of research Select policy outcomes to be followed up by impact analyses
Post policy outcome	Identify key indicators for the targeted impacts to measure Collect or compile baseline measures of indicators (before the policy outcome takes effect)
Policy impact	Measure quantitative impacts from the policy outcome, likely through use of modeling Use qualitative methods to assess certain types of impacts and the contribution of policy outcome to the impacts

One of the motivations for the workshop was that many more qualitative assessments of the influence of policy research have been undertaken by the CGIAR than studies that quantify the net benefits of the research using cost-benefit ratios or internal rates of return. Yet quantitative studies could be very helpful in setting research priorities within the CGIAR as well as demonstrating the value of POR to donors.

However, as noted throughout this paper, quantification of policy impacts is challenging. Perhaps the greatest difficulty is the development of a counterfactual against which impacts can be measured. Although the topic of counterfactuals is handled in the next section, their implications for measuring impact are worth highlighting here. Because it is very difficult to find a clear counterfactual to the longer-term impacts of a policy outcome, evaluators must turn to second-best methods for generating quantitative measurements of policy impacts, namely the use of models. For some situations, existing models may be appropriate for quantifying ex post ranges of impacts of actual policy change. In others, IAs from pilot tests or early roll-outs of policies or programs, such as a social protection delivery mechanism, may provide a useful point of comparison. A pilot-stage

impact evaluation can provide useful information about eventual larger-scale impacts, especially if the evaluation was rigorously designed (for example, an RCT) and representative in its coverage (that is, the pilot mirrored the population of the larger scale-up).

Counterfactuals

Key questions for an evaluator of POR are, what would have happened without the POR (the counterfactual)? And, how much of the difference in benefits between the actual and the counterfactual can be attributed to the POR? In thinking about counterfactuals, we need to recognize three possible situations. In the first situation there was going to be a policy change anyway, and the POR may have led to a more informed change with a better outcome. In this case, the relevant counterfactual is the level of benefit obtained with the policy change that would have occurred in the absence of the POR. A special case arises when the POR leads to earlier implementation of a policy change that would have occurred anyway, in which case the counterfactual is the later implementation of the same policy (for an example see Ryan's [1999] evaluation of IFPRI's contribution to the rice marketing reforms in Vietnam). In a second situation, the policy would not have changed without the POR, as for example if the POR played an important diagnostic role in identifying problems with the old policy. In this case the relevant counterfactual is the existing policy. In a third situation, the POR convinces policymakers not to make a planned change to the existing policy, thus preventing a worse outcome. In this case, the relevant counterfactual is the new policy that would have been put in place.

To clarify these counterfactual relationships and ways in which attributions to POR can be made, we introduce some notation.

Let:

R designate the new evidence generated by POR

A = Attributed share of the benefits to R

B_1^{R1} = Benefits following a policy change informed by POR

B_1^{Ro} = Benefits following a policy change not informed by POR

B_o = Benefits if existing policy is maintained.

Benefits (B) could of course be negative.

Case 1: There would have been a policy change anyway

We begin with the case in which there would have been a policy change anyway, and the POR may have informed that change and led to a better outcome. In this case, there are two alternative ways to measure the benefits attributable to POR. The first is a direct measure between the benefit B obtained with a new policy informed by R and a new policy not informed by R. That is:

(1) $B^* = B_1^{R1} - B_1^{Ro}$

An alternative measure is the total gain in benefit between the new policy informed by R and continuation of the old policy, multiplied by the attribution share A. That is:

(2) $B^* = (B_1^{R1} - B_0)\,A$

Assuming (1) and (2) are equivalent, we can solve for A, the attribution share to POR, as follows:

$$B_1^{R1} - B_1^{Ro} = (B_1^{R1} - B_0)\,A$$

hence

(3) $A = (B_1^{R1} - B_1^{Ro}) / (B_1^{R1} - B_0)$

Equation (3) shows that an attribution share can be measured entirely in terms of benefit shares rather than in terms of "influence" on the underlying policy process. However, this requires that empirically we can quantify the two counterfactuals B_1^{Ro} and B_0. The first requires that we can calculate what would have happened if the policy change without R had prevailed, and the second assumes we can project what would have happened if the old policy had been retained. In practice, it is often easier to quantify B_0 rather than B_1^{Ro}. If an independent estimate of A can be obtained from other sources to use in equation (2) then we can avoid the need to estimate B_1^{Ro}. One way to estimate A independently is through stakeholder interviews, another is by assuming a plausible range for A (for example, Renkow's [2010] evaluation of the impact of IFPRI's public investment research on rural road investments in India).

Case 2: The policy would not have changed without POR

The problem is simpler if the policy would not have changed without the POR. This situation might arise, for example, if the POR identified serious shortcomings with an existing policy and helped convince policymakers of the need for change. In this case, equation (1) becomes:

$$B^* = (B_1^{R1} - B_0),$$

while equation (2) remains:

$$B^* = (B_1^{R1} - B_0)\,A$$

From which it is evident that A = 1. In other words, POR is attributed with all the benefit gains arising from the policy change and we only need to estimate the counterfactual B_0.

Case 3: POR prevented a policy change

Sometimes a policy change is planned, but POR helps convince policymakers that the change would lead to a worse outcome than the existing policy. In this case the analysis is the same as for case 2, except that B^* or $(B_1^{R1} - B_0)$ would be negative, and the gain from POR is avoiding a negative outcome.

Note that the above analysis only establishes a counterfactual for the new evidence (R) generated by the POR. In cases where it is also desired to attribute part of the benefit from R to a specific institution (say X) that was involved in developing R, then a further counterfactual is required that asks: would some other institution (say Y) have undertaken the POR and developed the same new evidence R as did X? If the answer is yes, then little of the benefit may be attributed directly to X. In some cases, the answer might be that yes some other institution would have undertaken the POR and generated R, but at a later date. In this case the incremental gain in the benefit from an earlier policy change can be attributed to X (as in Ryan's [1999] evaluation of IFPRI's contribution to the rice

marketing reforms in Vietnam). In cases where no other institution would have stepped in and done the POR, then all the benefit gain resulting from R might be attributable to X.

A general consensus from the workshop is that qualitative methods for estimating the attribution share A are more useful than quantitative methods. Quantitative methods are more likely to be useful for calculating the impact of a policy change and for constructing a counterfactual impact. In some cases, such as implementing a cash transfer, credit, or insurance program or a market information system, the impact might be estimated through randomized trials with a control group or through sequenced roll-outs of the intervention. For changes in sector policies, such as commodity taxes or input subsidies, or macroeconomic policies, the POR may have developed useful models for evaluating the impact of the policy change as part of its own research agenda. In the CGIAR, these typically take the form of household, market, sectoral or economywide models that can be used to simulate the impacts of a policy change and compare it to a baseline 'without project' situation. For IA purposes, models really need to be validated by their ability to reproduce history rather than to make forward projections, but not all models can be adapted to this purpose. For example, econometric models can be estimated to fit historical data from a period following a policy change, whereas computable general equilibrium models can only simulate and compare hypothetical equilibrium situations. On the other hand, forward-looking simulations may better capture the long-run impacts of a policy change, especially if the evaluation is undertaken soon after the policy change occurred.

Many policy impacts are simply not amenable to quantitative analysis, either because of difficulties in valuation of benefits or in the construction of counterfactuals. For example, some social and environmental impacts are inherently difficult to value. Likewise for policy outcomes like new or enhanced institutional capacities or improved governance, there may be no rigorous basis for constructing a counterfactual. Some of the new methods discussed at the workshop may eventually help, but for many areas of POR, mixed methods and good narratives for elucidating policy impacts and their counterfactuals remain the best options available. Stakeholder interviews, including of key policymakers, can be an important source of evidence for such analyses.

3. RECOMMENDATIONS FOR THE CGIAR

The workshop highlighted the diversity of types of POR conducted within the CGIAR system, as well as the implications of this diversity for IA in terms of the types of TOC and assessment methodologies that are relevant or feasible. POR in the CGIAR system ranges from micro-oriented studies of ways to improve specific programs or projects within countries, through studies of broader sector or national policies within individual countries, to studies of regional and global problems requiring international solutions. A simple typology based on these distinctions helps elucidate differences in the length and complexity of the impact pathways involved and the limits to using quantitative and rigorous IA methods. The workshop also highlighted the general difficulties and sensitivities involved in making attributions to specific POR activities, even in cases where the impacts of the resulting policy changes can be rigorously evaluated.

These problems notwithstanding, the workshop provided several useful pointers for improving the impact of POR and strengthening PORIA within the CGIAR system. The presentation of these "good practices" is based on an approach in which policy-oriented research is planned within a TOC and the research outputs are tracked and assessed. Alternatively, PORIA studies may also begin by identifying a policy outcome and then assessing both downstream impacts and upstream influences. In either case, the CGIAR should adopt the good practices described below, with a few differences outlined below. In either case, useful PORIA studies can be conducted even if some of these good practices are not followed.

Identifying Target Policy Outcomes and Impacts

Most POR projects are more likely to be successful in achieving impact if they clearly identify the key policy outcomes they hope to influence and the likely economic, social, and environmental impacts that would follow. While the CGIAR is neither a policymaking nor an implementing agency and cannot reasonably be held accountable for final policy outcomes and impacts, nevertheless a clear vision of what is desired is a necessary first step toward developing a TOC and reporting framework to which the project can be held accountable.

Developing a Theory of Change

Outcome mapping and other TOC techniques can help in planning the achievement of these policy outcomes. Ideally the planned outcome is specific (which policy, where, by whom), and the team also specifies the research outputs, the science-to-policy activities, and key partners. However, the TOC can be expected to be more complex in terms of the policy content and actors involved as one moves up the scale from POR targeted at project/program issues, to sector, national, regional, and global issues. Given the dynamic and nonlinear nature of many policy processes, a TOC may need to be flexible and revised over time (it is the latest one that will be relevant for use in the eventual evaluation process). Practice shows that developing TOCs can be time consuming and expensive if key stakeholders are brought together for the exercise. So while everyone recognizes the importance of TOCs and that some formalization of impact pathways for the CGIAR, including for policy outcomes, is desirable, more thought is required as to the proper level of investment in TOCs for different types of POR.

Tracking of Outputs and Dissemination of Activities

The recording of research outputs—policy diagnostic studies, pilot evaluation studies of policy interventions, ex-ante IAs of policy options, and research publications—is already well mainstreamed in the CGIAR. What has been lacking within the CGIAR is more formal documentation of dissemination, outreach, and capacity-building activities, in part because these are secondary activities that may not lead to easily identified scientific output achievements and in part because researchers have had little incentive to document them. This should not become an onerous task, but just provide a basic description of the activity. For meetings, documentation could simply include the date, location, and participants, whether any specific research output was discussed, and a couple of sentences describing the meeting content, including any decisions or conclusions. For training activities, records of participants and length and type of training may suffice. One lesson that has emerged from some past evaluations is that outputs not only need to be recorded, but also tagged to specific POR projects. This can be very difficult to do ex post, especially if the project being evaluated is part of a larger thematic program.

Tracking of Outcomes

Prior to the establishment of the CGIAR Research Programs (CRPs), a performance monitoring system was used to evaluate CGIAR centers on a number of criteria, including outcomes achieved. Since then, there has not been a clear set of rules for reporting outcomes and different Centers and CRPs handle reporting in their own way.

The former system required a short submission of 1–2 pages of text for each major outcome, along with accompanying evidence, to supply the following information:

- Outcome statement
- What output/output target resulted in the outcome?
- In which Medium Term Plan was the output/output target identified as planned before its achievement?
- Where was the achievement of the output/output target documented?
- Who used/adopted or was influenced by the output?
- How was the output used or adopted?
- What is the magnitude and significance of the outcome relative to the intended recommendation domain?
- What is the evidence for the outcome?

The Climate Change, Agriculture and Food Security program of the CGIAR (CCAFS) has developed its own tool for activity leaders to submit outcomes, organized around the following headings:

- Contribution to CCAFS Outcome
- Title

- Outcome statement
- Recent research outputs used
- Research partners
- Activities (outreach/dissemination) that contributed to the outcome
- Non-research partners
- User of the outputs
- How the output was used
- Evidence of the outcome

Both of these systems are based on self-reporting, though third-party evidence is encouraged. Self-reporting of outcomes should continue as the accepted approach toward a comprehensive inventory of major outcomes. Any other approach would be too costly. However, for a sample of outcomes, it will be important to have more rigorously documented outcomes to validate the self-reporting.

Collecting Evidence of Outcomes

Evidence of the policy outcome should be provided to support all claims of outcomes (such as new regulation, new policy/strategy, declaration, program, etc.). Where the policy outcome is captured in written law or other document, this is not difficult to do. More difficult are instances where the policy outcome is prevention of a change in policy (so written policy documents have not changed) or where the outcome is more nuanced, such as an improved policy process or technical capacity in a policymaking entity. These cases will require more information gathering and analysis and third-party reports will be useful.

Collecting Evidence of Research Influence on Outcomes

For all outcomes, a self-reported narrative relates back to planned outcomes and the TOC. For this the tools above for the description of outcomes are sufficient as they ask which outputs were used, how, and by whom, and request a description of outreach or dissemination actions. In most cases, this evidence will demonstrate some degree of contribution of the research, although mainly from the viewpoint of the suppliers of the evidence.

However well researchers assess their influence, some proportion of outcomes still should require a higher level of evidence, in particular to evaluate influence from the demand side. Influence should be measured against standards of reasonable doubt or probable cause, since proving influence is often impractical. This assessment is subjective and requires skilled techniques. Furthermore, reliable and honest answers are not likely to emerge if the research team undertakes this exercise alone. Thus, third-party evaluators must be engaged to evaluate influence in a credible and legitimate way. The external reviewers should focus on major CGIAR policy research investment areas. These reviews should be conducted as early as possible, while the relevant stakeholders to be interviewed are still in place.

Measuring Impacts of Policy Outcomes

For a proportion of policy outcomes, there should be further qualitative and quantitative assessments of impacts of the outcomes. External evaluators hired after project completion cannot be expected to amass all the evidence required to assess project impacts. Impact assessments will generally need to be produced by technical experts (mainly economists in the case of policy impacts) at various stages of the POR process in order to provide external evaluators with the information they will need. The IAs should be built into the work plan by the CGIAR research team. While the team may participate in these studies, third-party involvement and participation of technically qualified researchers is essential to lend credibility and legitimacy to the results, as well as to raise the quality of such studies by bringing in a fresh perspective.

For a smaller proportion of policy outcomes, where some IAs have been generated, there should be external evaluations of outcomes and impacts by experienced evaluators. This is for the purpose of validating results of the assessments and exploring more qualitatively why impacts have or have not materialized as predicted. The evaluation may also uncover unintended or unanalyzed impacts.

Conducting PORIA Studies Starting from a Policy Changes

The good practices in PORIA described above emphasize the role of research and research outputs in influencing policy change. Where possible, they will enable a full cost-benefit analysis for an investment in POR. Another interesting approach that can generate lessons for other POR is known as an 'episode study' (Carden 20009). Episode studies take an important and relevant policy change as a starting point and then trace backward to understand the contribution research and other inputs made to the policy outcome. The major differences between this approach and the research-led approach described above is that (a) this approach more systematically analyzes the influence of non-research factors on policy outcomes, and (b) this approach looks at a specific policy change and will not necessarily include all the outcomes from a specific research initiative (which means it is less useful for conducting a typical cost-benefit analysis of a POR investment). This type of analysis is extremely useful for understanding the relative contribution of research as well as other important factors leading to policy, which can then help researchers in developing future TOC.

Additional Recommendations

Because it may not be practical to follow all the above guidelines for every single POR project, it may make more sense to track and evaluate bundles of projects from the same research themes or programs. This could still leave room for selecting individual projects within the bundle for more in-depth assessment. The workshop discussion suggested that, in this context, it may be best to cherry-pick the more successful projects within the portfolio, since these are the ones most likely to justify the project investment. Nevertheless, selecting some less successful projects may also be useful for institutional learning about how to avoid past mistakes, assuming that poor performance was due to factors over which the POR team had some control.

The CGIAR also needs more quantitative evidence about the impact of its POR investments. The few studies that have been done to date show favorable returns, but there have been too few of them. Moreover, these quantitative IAs have focused on POR targeted at relatively small and

narrowly defined problems within single countries, where evaluation is easiest but impacts are likely to be relatively small compared to projects that generate significant international spillovers. The workshop identified difficulties and sensitivities inherent in attempting to quantify the benefits from POR. However, if POR is to compete in today's funding environment, a subset of successful POR projects should be deliberately cherry-picked for more intensive and rigorous evaluation, including some that had an international focus.

This paper has said very little about how to undertake ex post evaluations of POR projects for which a systematic evidence base has not been built up. Many evaluators of past and current POR projects find themselves in this position, as did nearly all the evaluation studies cited in the paper—it was left to the ingenuity of the evaluators to make do with what evidence they could muster. In nearly all cases evaluators followed the recommended approach of developing their own TOC and then tried to follow the evidence trail. Such studies rarely get beyond using stakeholder reviews to establish some level of policy influence (or lack thereof). There are no generic guidelines for filling important evidence gaps after the event, which is why we argue that POR teams need to build a monitoring and evaluation framework into their work from the outset to facilitate a proper ex-post IA.

Some Cautions

Finally, one should be aware of some potential pitfalls from requiring too rigid an adherence to a reporting framework for POR. Policymaking processes are typically dynamic, complex, and interactive—if researchers aim to influence these processes they must be flexible and adaptive, recognizing that first best outcomes rarely prevail. To achieve an impact, it can be just as important to be involved in the politics of policy implementation as in its formulation. Too rigid an adherence to a set of ex-ante defined outputs and outcomes may sideline the relevance of a POR project.

Another risk is that many POR outputs are intangible and hard to foresee, and long lead times are inherent to some POR impacts, for example the impacts from strengthened capacity through collaborative research or from building up data sets or economic models. Impacts are often very indirect. For example, POR outputs may influence other players, such as other researchers and advocacy agents, who in turn have influence and impact on policies. These impacts may be impossible to track—indeed the POR team may not even be aware of them. There may also be cumulative impacts and synergies from projects that focus on the same policy issue or on the same place. Overlooking these aspects in setting up a reporting framework could lead to distorted assessments of the contribution of POR projects.

Yet another risk is that requiring researchers to show direct impacts on policy outcomes could lead to greater concentration of POR on the problems of individual countries, where impacts are easiest to track. However, some of the biggest gains from POR are thought to arise from cross-country spillovers and through changing the policies of international organizations and donors who in turn influence policies in many countries (international public goods). These indirect pathways are much harder to influence and track, and hence POR projects that are weighted toward them may be less likely to be funded.

REFERENCES

Ahmed, M., K. Shideed, and A. Mazid. 2010. Returns to policy-oriented agricultural research: The case of barley fertilization in Syria. *World Development* 38(10): 1462–1472.

Babu, S. 2000. Impact of IFPRI's policy research on resource allocation and food security in Bangladesh. Impact Assessment Discussion Paper 13. Washington, DC: International Food Policy Research Institute.

Byerlee, D., and J. Bernstein. 2013. *Feed the future learning agenda literature review: Improving research and development*. Rockville, MD: Westat.

Carden, F. 2009. Understanding influence: The episode studies approach. In *The Politics of trade, the role of research in trade policy and negotiation*, edited by D. Tussie. Ottawa: International Development Research Centre (IDRC).

Contandriopoulos, D., M. Lemire, J.-L. Denis, and É. Tremblay. 2010. Knowledge exchange processes in organizations and policy arenas: A narrative systematic review of the literature. *Milbank Quarterly* 88(4): 444–483.

Hausmann, R., D. Rodrik, and A. Velasco. 2008. Growth diagnostics. In *The Washington consensus reconsidered: Towards a new global governance,* edited by J. Stiglitz and N. Serra. New York: Oxford University Press.

Hazell, P., and R. Slade. 2014. Policy research: The search for impact. Paper prepared for the workshop "Best Practice Methods for Assessing the Impact of Policy-oriented Research," International Food Policy Research Institute, Washington, DC, November 11–12.

Hovland, I. 2007. *Making a difference: M&E of policy research*. Working Paper 281. London: Overseas Development Institute.

Kaitibie, S., A. Omore, K. Rich, and P. Kristjanson. 2010. Kenya dairy policy change: Influence pathways and economic impacts. *World Development* 38(10): 1494–1505.

Kydd. J. 2015. *Ex-post evaluation study of IFPRI's research on high-value agriculture, 1994–2010*. Independent Impact Assessment Report 39. Washington, DC: International Food Policy Research Institute.

Masset, E., R. Mulmi, and A. Sumner. 2011. *Does research reduce poverty? Assessing the welfare impacts of policy-oriented research in agriculture*. IDS Working Paper 360 Brighton, UK: Institute of Development Studies, University of Sussex.

Masset, E., M. Gaardner, M., P. Beynon, and C. Chapoy. 2013. What is the impact of a policy brief? Results of an experiment in research dissemination. *Journal of Development Effectiveness* 5(1): 50–63.

Mayne, J. 2012. Contribution analysis: Coming of age? *Evaluation* 18(3): 270–280.

Raitzer, D. 2010. Assessing the impact of policy-oriented research: The case of CIFOR's influence on the Indonesian pulp and paper sector. *World Development* 38(10): 1506–1518.

Raitzer, D.A., and J.G. Ryan. 2008. State of the art in impact assessment of policy-oriented international agricultural research. *Evidence and Policy* 4(1): 5–30.

Renkow, M. 2010. Impacts of IFPRI's "priorities for pro-poor public investment" global research program. Impact Assessment Discussion Paper 31. Washington, DC: International Food Policy Research Institute.

Renkow, M., and D. Byerlee. 2014. Assessing the impact of policy-oriented research: A stocktaking. Paper prepared for the workshop "Best Practice Methods for Assessing the Impact of Policy-oriented Research," International Food Policy Research Institute, Washington, DC, November 11–12.

Ryan, J. 1999. Assessing the impact of rice policy changes in Viet Nam and the contribution of policy research. Impact Assessment Discussion Paper 8. Washington, DC: International Food Policy Research Institute.

Ryan, J., and X. Meng. 2004. The contribution of IFPRI research and the impact of the food for education program in Bangladesh on schooling outcomes and earnings. Impact Assessment Discussion Paper 22. Washington, DC: International Food Policy Research Institute.

Sylva, K., B. Taggart, E. Melhuish, P. Sammons, and I. Siraj-Blatchford. 2007. Changing models of research to inform educational policy. *Research Papers in Education* 22(2): 155–168.

Templeton, D., and N. Jamora. 2010. Economic assessment of a change in pesticide regulatory policy in the Philippines. *World Development* 38(10): 1519–1526.

Walker, T., J. Ryan, and T. Kelley. 2010. Impact assessment of policy-oriented international agricultural research: Evidence and insights from case studies. *World Development* 38(10): 1453–1461.

Young, J., L. Shaxson, H. Jones, S. Hearn, A. Datta, and C. Cassidy. 2014. *Rapid outcome monitoring approach: A guide to policy engagement and influence*. London: Overseas Development Institute.

ANNEX 1: LIST OF PARTICIPANTS

Anderson, Jock	Independent Consultant
Anisimova, Evgeniya	IFPRI-PIM
Bantilan, Cynthia	ICRISAT-IN
Behrman, Jere	University of Pennsylvania
Belcher, Brian	CIFOR/Royal Roads University
Bertram, Rob	USAID
Bohman, Mary	USDA
Boughton, Duncan	MSU
Brooks, Karen	IFPRI/PIM
Brown, Annette	3ie
Cahill, Carmel	OECD
Fan, Shenggen	IFPRI
Fuglie, Keith	ERS, USDA
Gerrard, Chris	Head of the PIM evaluation team
Gilligan, Dan	IFPRI
Hazell, Peter	IFPRI
Herdt, Robert	SPIA
Hoddinott, John	IFPRI
Kelley, Tim	SPIA
Larochelle, Catherine	Virginia Tech
Macours, Karen	SPIA, Paris School of Economics
Maredia, Mywish K.	Michigan State University
Martin, Will	The World Bank
Meinzen-Dick, Ruth	IFPRI
Oehmke, Jim	USAID
Ozler, Berk	The World Bank
Paarlberg, Rob	Williams College
Pandya-Lorch, Rajul	IFPRI
Place, Frank	IFPRI/PIM
Qureshi, Ejaz	ACIAR
Regmi, Anita	Bioversity
Renkow, Mitch	NCSU
Roberts, Stacy	IFPRI
Selee, Andrew	Woodrow Wilson Center

Slade, Roger	Independent consultant
Smith, Vincent H.	Montana State University
Tomich, Thom	UC Davis
Torero, Maximo	IFPRI
Traxler, Greg	BMGF
Underwood, John	MCC
Walker, Tom	MSU
Young, John	ODI

ANNEX 2: AGENDA

	Time	Speakers/Panel discussants
Tuesday, NOVEMBER 11		
Welcome	9:00 – 9:10 am	Maximo Torero
Session 1: **Introduction (objectives of workshop, key challenges to be resolved)**	9:10 – 9:30 am	Karen Brooks, Peter Hazell
Session 2: **Taking stock: What we know about Assessing the Impact of Policy-Oriented Research**	9:30 – 10:30 am	Mitch Renkow (***presentation***) Chair: Peter Hazell
Coffee break	10:30 – 10:45 am	
Session 3: **How social scientists other than economists approach assessment of policy-oriented research**	10:45 – 11.30 am	Rob Paarlberg (***presentation***) Comments by Ruth Meinzen-Dick Chair: Maximo Torero
Session 4: **A typology of policy oriented research**	11:30 – 12:45 pm	Peter Hazell, Roger Slade (***presentation***) Chair: Will Martin
Lunch	12:45 – 1:45 pm	
Session 5: **Methods for assessing policy oriented research**	1:45 – 5:30 pm	
(5.1) **Panel discussion on past case studies exemplifying good methods/approaches**	1:45 –2:45 pm	Jere Behrman (***presentation***) Vince Smith (***presentation***) Tim Kelley Chair: Mywish Maredia
(5.2) **Challenges of assessing influence of research in policy processes (attribution/contribution)**	2:45 – 3:45 pm	Andrew Selee John Young (***presentation***) Ruth Meinzen-Dick Chair: Karen Brooks
Coffee break	3:45 – 4:00 pm	
(5.3) **New ideas: Promising new methods; especially as relevant to CGIAR**	4:00 – 5:00 pm	Karen Macours (***presentation***) Annette Brown (***presentation***) Berk Ozler Maximo Torero (***presentation***) Chair: Rob Paarlberg
(5.4) **General discussion**	5:00 – 5:30 pm	
Welcome reception	Starts at 5:30 pm	
Wednesday, NOVEMBER 12		
Introduction to Day 2	9:00 – 9:10 am	Peter Hazell, Karen Brooks
Session 6: **Indicators/approaches for tracking policy change and contribution of research**	9:10 – 10:10 am	John Underwood Karen Brooks (***presentation***) Jim Oehmke Chair: Keith Fuglie

	Time	Speakers/Panel discussants
Session 7: **Working group discussions of the specific challenges, appropriate methods, impact pathways and indicators relevant to specific types of policy-oriented research**	10:10 – 12:30 am	
Lunch	12:30 – 1:15 pm	
Session 8: **Working groups' presentations and general discussion**	1:15 – 3:00 pm	
Coffee break	3:00 – 3:30 pm	
Session 9: **Summary session and key recommendations for assessment**	3:30 – 4:30 pm	Thomas Tomich, Jere Behrman, Keith Fuglie, Rob Paarlberg Chair: Peter Hazell
Session 10: **Discussion and donors' perspective on the way forward**	4:30 – 5:15 pm	Greg Traxler Rob Bertram Ejaz Qureshi (***presentation 1***, ***presentation 2***) Chair: Karen Brooks
Session 11: **Closing**	5:15 – 5:30 pm	Tim Kelly, Karen Brooks