



ARTICLE

Received 2 May 2016 | Accepted 20 Feb 2017 | Published 6 Apr 2017

DOI: 10.1057/palcomms.2017.17

OPEN

Evaluating policy-relevant research: lessons from a series of theory-based outcomes assessments

Brian Belcher^{1,2}, Daniel Suryadarma² and Aidy Halimanjaya^{3,4}

ABSTRACT The increasing external demand from research funders and research managers to assess, evaluate and demonstrate the quality and the effectiveness of research is well known. Less discussed, but equally important, is the evolving interest and use of research evaluation to support learning and adaptive management within research programmes. This is especially true in a research-for-development context where research competes with other worthy alternatives for overseas development assistance funding and where highly complex social, economic and ecological environments add to evaluation challenges. Researchers and research managers need to know whether and how their interventions are working to be able to adapt and improve their programmes as well as to be able to satisfy their funders. This paper presents a theory-based research evaluation approach that was developed and tested on four policy-relevant research activities: a long-term forest management research programme in the Congo Basin; a large research programme on forests and climate change; a multi-country research project on sustainable wetlands management, and; a research project of the furniture value chain in one district in Indonesia. The first used Contribution Analysis and the others used purpose-built outcome evaluation approaches that combined concepts and methods from several approaches. Each research evaluation began with documentation of a theory of change (ToC) that identified key actors, processes and results. Data collected through document reviews, key informant interviews and focus group discussions were analysed to test the ToCs against evidence of outcomes in the form of discourse, policy formulation and practice change. The approach proved valuable as a learning tool for researchers and research managers and it has facilitated communication with funders about actual and reasonable research contributions to change. Evaluations that employed a participatory approach with project scientists and partners noticeably supported team learning about past work and about possible adaptations for the future. In all four cases, the retrospective ToC development proved challenging and resulted in overly-simplistic ToCs. Further work is needed to draw on social scientific theories of knowledge translation and policy processes to develop and further test more sophisticated theories of change. This theory-based approach to research evaluation provides a valuable means of assessing research effectiveness (summative value) and supports learning and adaptation (formative value) at the project or programme scale. The approach is well suited to the research-for-development projects represented by the case studies, but it should be applicable to any research that aspires to have a societal impact. This article is published as part of a collection on the future of research assessment.

¹ Royal Roads University, Victoria, BC, Canada ² Center for International Forestry Research, Bogor, Indonesia ³ Sustainable Development Goals Centre, Padjadjaran University, Bandung, Indonesia ⁴ School of International Development and Tyndall Centre, University of East Anglia, East Anglia, UK

Introduction

The trend toward results based management approaches in public administration (Behn, 2003; Meier, 2003; Sjøstedt, 2013) is manifest in the research world as increased demands from research funders and research managers to assess, evaluate and demonstrate the quality and the impact of research. This trend has four main drivers, referred to by Morgan and Grant (2013) as the “4 As”: accountability, allocation, advocacy, and analysis. Funders are seeking ways to increase and improve accountability and better information to help allocate resources effectively. Researchers feel the need to demonstrate the value of research to advocate for continued funding. And there is a need for analysis, to understand whether and how research contributes to society.

Much discussion has focused on the accountability and allocation aspects, and there is concern that emphasis on accountability has perverse effects (Martin, 2011; Chambers, 2015). Less discussed, but equally important, is the evolving interest in and use of research evaluation to support learning and improved design, as a means to advance research practice. In their review of current research impact evaluation methodologies, Morgan and Grant (2013) highlight “analysis” as an area that needs more attention to advance the impact agenda. Researchers and research managers, as well as research funders, need appropriate criteria and methods for evaluating research design and implementation. They need improved tools and methods to assess whether and how their work is contributing to its intended outcomes and impacts. As Carter (2013: 1) put it, the impact agenda “... is about enabling, understanding, and describing the beneficial outcomes of research.”

The learning objective of research evaluation is especially important as research approaches evolve, crossing disciplinary boundaries and engaging a broader range of actors in the quest to improve effectiveness (Gibbons *et al.*, 1994; Clark and Dickson, 2003; Wickson *et al.*, 2006; Carter, 2013; Belcher *et al.*, 2016). An analysis of the 2014 UK Research Excellence Framework (REF) found that many cases of societal impact resulted from multi-disciplinary work (Kings College, 2015), reflecting the ways that universities have engaged with a range of public, private and charitable organisations and local communities. Researchers are working deliberately not only to produce knowledge, but also to promote and facilitate the use of that knowledge to enable change, solve problems, and support innovation (Clark and Dickson, 2003). Transdisciplinary approaches transcend disciplinary and institutional boundaries to contextualise research around the interests of stakeholders and actively involve the users of research, to foster more socially robust knowledge (Gibbons and Nowotny, 2000). Sustainability science has emerged as a new discipline, grounded in the belief that knowledge needs to be co-produced through close collaboration between scholars and practitioners (Holling, 1993; Clark and Dickson, 2003; Berkes, 2009). The literature on linking knowledge to action recognises the importance of non-linear, dialogical, discursive and multi-directional approaches, acknowledging that knowledge is socially constructed and not a unidirectional producer-to-consumer process (Van Kerkhoff and Lebel, 2006). New approaches stress the importance of “boundary work”, in which knowledge creation is done in conjunction with the worlds of action and policy making (Clark *et al.*, 2011; Kristjanson *et al.*, 2009). Researchers and research organisations need to learn systematically from this experience to improve their effectiveness in achieving intended results in complex, policy-relevant research environments. We still lack appropriate tools and methods to meet this research evaluation challenge.

Guthrie *et al.* (2013) and Morgan and Grant (2013) review the strengths and weakness of the current main research evaluation approaches: bibliometrics, case studies, economic analysis, and peer review. They note that different disciplines and different

evaluative purposes require different approaches; there is no universal framework. But case study emerges as the best approach to get nuanced understanding of the range and kinds of impacts. Likewise, Donovan (2011), discussing what constitutes state-of-the-art methods for assessing the impact of research, argues strongly that best practice combines narratives with relevant qualitative and quantitative indicators to gauge broader social, environmental, cultural and economic public value.

The need for better ways to evaluate research effectiveness is well recognised. The Canadian Federation for the Humanities and Social Sciences (2014), Nutley *et al.* (2007), and the British Academy (2004) have identified key challenges and underlined the lack of adequate models, indicators, and impact measurement tools to assess research and to facilitate adaptive learning.

In the research evaluation field, the Payback Framework developed by Buxton and Hanney (1996) and Hanney *et al.* (2004) was one of the first research evaluation tools that incorporated both academic outputs and societal impact as criteria for assessment. It still one of the most widely used approaches for the assessment of the impacts from health research (Canadian Academy of Health Sciences, 2009; Banzi *et al.*, 2011; Greenhalgh *et al.*, 2016). It uses a logic model and an outcome-based retrospective, narrative, case study approach to assess five pre-defined categories of research benefits: knowledge production; research targeting, capacity building, and absorption; informing policy and development (broadly defined); health benefits; and broader economic benefits (Hanney *et al.*, 2004; Donovan and Hanney, 2011). An alternative and very different approach, known as societal impact assessment, focuses on “productive interactions” between researchers and research users (Spaapen and Sylvain, 1994; Spaapen and van Drooge, 2011; Mollas-Gallart and Tang, 2011).

There is emerging consensus around the use of “theory of change” (ToC) approaches in programme evaluation, with valuable lessons for research evaluation. A ToC is an explicit depiction of the relationships between initiative strategies (an intervention) and intended results (outcomes and impacts) (White, 2009; Coryn *et al.*, 2011; 53Stern *et al.*, 2012). In the most basic form, a theory of change considers a series of stages, from inputs through outputs, outcomes and impact. More sophisticated and realistic models include both short- and longer-term outcomes and also reflect changes at different levels, as individuals, organisations, and communities respond to interventions and interact within complex systems. A theory of change anticipates and indeed plans for how a project will interact with and influence its stakeholders and other audiences to achieve developmental outcomes and impact. The intended pathway(s) to impact are made explicit and therefore testable (Coryn *et al.* 2011; Gamel-McCormick, 2011). This hypothesis testing aspect makes the approach particularly attractive for research evaluation.

But the theory has been ahead of the practice (White and Philips, 2012; Mayne and Stern 2013; Mayne *et al.*, 2013), especially as applied to research evaluation. We have not had well-tested methods available for assessing the outcomes and the impacts of research-for-development.

This paper presents and assesses the use of theory-based research evaluation by comparing, contrasting and assessing completed evaluations that explicitly tested theories of change in four research-for-development projects. It provides an overview and analyses of each case study and draws lessons for further improving and refining the approach. The following section provides a brief context of the research environment of the case studies and defines relevant terms and concepts accordingly. Notably, research outcomes and impact are defined more

precisely than is common in academic research evaluation. The next section presents an overview of the research evaluation approach, highlighting key similarities and distinctions with other approaches such as the Payback Framework and Societal Impact Assessment. We then present the method and criteria for our assessment followed by overviews of each of the cases. The results and discussion section highlights lessons learned and assesses the approach against the criteria developed earlier. The conclusion focuses on the main lessons and needs for further development.

The research context

The case studies reviewed here evaluate the outcomes of research projects done by the Center for International Forestry Research (CIFOR), a research centre within the global research partnership on agriculture and natural resources management known as the CGIAR. CIFOR is an international non-profit, scientific organization that conducts interdisciplinary research on forest and landscape management as a means to improve human well-being, protect the environment, and increase equity, primarily in less developed countries. The research aims to help policymakers, practitioners and communities make science-based decisions (CIFOR, 2016). CIFOR research projects are typically large, multi-partner, and international, with durations of three to five years. The research is often interdisciplinary, with strong biophysical and social science components. It seeks to influence policy and practice of conservation and development organizations, private sector resource managers and governments at sub-national, national and international levels.

CIFOR's research funding comes mainly from overseas development assistance budgets, directly through project funding and indirectly through the CGIAR system. These funders have all increased emphasis on results and results based management in recent years. A reform in the CGIAR system that began in 2011 emphasises that research centres and their programmes have a responsibility to do high quality science and a "shared responsibility" for achieving development outcomes (CGIAR, 2015). Research centres and their scientists are committed to "... producing, assembling and delivering, in collaboration with research and development partners, research outputs that are international public goods which will contribute to the solution of significant development problems that have been identified and prioritized with the collaboration of developing countries." (CGIAR, 2011). An overarching CGIAR Strategy and Results Framework (CGIAR, 2015) aims for three main high-level impacts: reduced poverty; improved food security and nutrition security for health, and improved natural resources and ecosystem services. This reform has created pressures and incentives analogous in many ways to those created by the new emphasis on "impact" in the UK REF.

This emphasis on results has promoted the development of more precise and analytical definitions of results concepts at CIFOR. In academic research impact discourse, definitions of societal impact tend to align with the REF definition as: "An effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia (REF, 2011)". CIFOR uses a logic model conceptual framework in which research activities produce a range of outputs (knowledge products and services) that are used by other actors in a system (potentially) resulting in a series of outcomes and impacts, with the following definitions (CIFOR unpublished):

An outcome is a change in knowledge, attitudes and/or skills, manifest as a change in behavior that results in whole or in part from the research and its outputs.

An impact is a change in flow or a change in state resulting in whole or in part from a chain of events to which research has contributed.

In the CIFOR context, it is expected that there will be multiple levels of outcomes. Impacts may be socio-cultural, economic, institutional or environmental.

A theory-based research evaluation approach

Each of the research evaluations presented here used a theory-based evaluation approach, with a ToC serving as the key conceptual and analytical framework (Weiss, 2007; Coryn *et al.*, 2011; Vogel, 2012). A ToC is essentially a comprehensive description and illustration of how and why a desired change is expected to happen in a particular context (Centre for Theory of Change, n.d.). It aims to show the causal relationships between a project's activities (often termed "interventions") and results, with attention to the primary pathways, actors and steps in the change process. The approach explicitly recognises that socio-ecological systems are complex and that causal processes are often non-linear, with multiple stages. A ToC sets out testable hypotheses of a change process by working back from long-term goals to identify all the conditions that (theoretically) must be in place for the goals to occur. With the right evidence, it is possible to assess actual achievements against expected outcomes at each stage.

In practice, each of the research evaluations: retrospectively documented the (previously implicit) project/programme ToC; used the ToC to identify data requirements and potential data sources to test each node in the ToC, and; collected and analysed data against the ToC. Data collection typically involves document review, key informant interviews, surveys, focus groups and sense-making workshops involving actors identified as having a role in the ToC. More detail is provided in the individual case descriptions.

The approach has elements in common with the Payback Framework (Hanney *et al.*, 2004; Donovan and Hanney, 2011) as each is based on a logic model. The Payback Framework considers direct interfaces between the research project and users and indirect influences through the "stock of knowledge" (Donovan and Hanney, 2011). Payback Framework practitioners appreciate that "By affecting the understanding of stakeholders, research can have an impact on policy-making at any stage, be it at the initial step of issue identification or at the final step of implementing a solution." (Klautzer *et al.*, 2011: 205). Our approach also has elements in common with Societal Impact Assessment (Spaapen and Van Drooge, 2011; Molas-Gallart and Tang, 2011), and differs from the Payback Framework, in that it appreciates that "knowledge" itself may be less significant than the varied and changing social configurations that enable its production, transformation and use (Greenhalgh *et al.*, 2016).

Our cases varied considerably in the nature and scale of the research activities that were evaluated and in the particular methods and combinations of methods employed, providing a rich basis for experiential learning and comparative analysis. The following sections present the approaches used in the four evaluations and the key differences between them. Table 3 summarises the main elements of each evaluation.

Methods: evaluating the evaluation approach

The four research programs were selected for evaluation as part of CIFOR's overall monitoring, evaluation and learning efforts. They represent the first examples of theory-based research evaluation done by CIFOR and among the first in the CGIAR.

Authors 1 and 3 had roles as evaluation scientists with CIFOR, with a specific focus on developing practical and effective ways to

assess and demonstrate research outcomes and both were involved as supervisors and/or team members (overall conceptualisation and design; data collection; analysis; reporting) in all four of the evaluations presented. Author 2 was involved as a team member (design; data collection; analysis; reporting) in two of the evaluations (GCS-REDD+ and SWAMP). We draw on our own experience as researchers and research managers. We also draw on our extensive interactions with the project researchers, donors and other partners (all of whom are grappling with the how to deal with increased need to assess research “impact”) during and subsequent to the evaluations.

We are reporting the key aspects of the approaches used in each case to share and to draw lessons from the experience. The analysis is qualitative and subjective—it represents the authors’ reflections on the process, informed by practical and theoretical considerations. To make it as systematic and transparent as possible, we provide a set of criteria against which we will assess the approaches.

Recent publications on research evaluation identify common methodological challenges. Morgan and Grant (2013) refer to the need to deal with: time lags from research to impact; attribution and contribution; assessing marginal differences; transaction costs of assessing research impact; and; unit of assessment. Penfield *et al.* (2013) offer a slightly different list of methodological challenges: time lag; the developmental nature of impact; attribution; knowledge creep; and gathering evidence.

The approach seeks to evaluate research projects in terms of both the knowledge production and the outcomes of the knowledge produced and associated activities to support knowledge translation. There are several audiences for such an evaluation: research teams; research managers; research users and intended beneficiaries; research funders, and; society more generally. Therefore, we also draw on criteria suggested for good evaluation practice more generally (e.g. Better Evaluation n.d.; OECD, 1991). We assess the approach used in these four cases against the following criteria:

1. Credibility of evaluation results to main intended audience (incl. dealing with time lags and attribution)
2. Contribution to the broader evidence base
3. Informs decision making aimed at improvement (formative value)
4. Informs decision making aimed at selection, continuation or termination (summative value)
5. Cost effectiveness

Research evaluation cases

This section provides a brief summary of the four case studies, with a one-paragraph summary of the research being evaluated followed by a summary of the evaluation process. Details of each research project are provided in Table 1 and specific evaluation case characteristics are provided in Table 2.

Contribution analysis of the sustainable forest management in the Congo Basin (SFM) program. The SFM Congo Basin research comprises a portfolio of research activities, rather than a discrete project or programme. The evaluation considers research done by CIFOR and the French Agricultural Research Centre for International Development (CIRAD) from 1995 to 2013 on forest governance, non-timber forest products, forest economics and impacts of the informal sector and climate change. The data, analyses, policy recommendations and operational solutions produced by the research, as well as training and capacity development, were intended to influence and support the policies and actions of international donor agencies, governments,

forestry companies and non-governmental actors and thereby to contribute to sustainable forest management in the region.

The SFM evaluation used “Contribution Analysis” (Mayne, 2008, 2012), a method that assesses whether: the expected results occurred; the supporting factors (assumptions in the ToC) have occurred and provide a reasonable explanation for the results; any other identified supporting factors have been included in the causal logic (thereby potentially revising the ToC); and any plausible rival explanations have been accounted for.

The evaluation focused on three main changes that occurred over two decades. First, forest issues became more prominent in the international policy arena, with sustainable forest management becoming the preferred approach over command and control conservation means of forest protection. Second, Cameroon, followed by other countries in the Congo Basin, began reforming their forestry laws. National forest administrations were replaced by regulation and control authorities that allocated exploitation rights to private companies, established mandatory forest management practices, and created national management standards. In 1999, a Central African Forest Commission (COMIFAC) was established, elevating forest conservation and sustainable management to a regional level. Third, timber companies began adopting sustainable forest management practices.

The evaluation aimed to assess CIFOR and CIRAD’s contribution to those changes. This was the first theory-based evaluation done by CIFOR. The methodology choice was influenced by a 2012 special issue on Contribution Analysis (CA) in the journal “Evaluation”. CA was deemed an appropriate way to deal with the complexity of the SFM research and the change process. External consultants were contracted to do the evaluation.

As with the evaluations discussed below, the method is organised around a ToC, which needed to be constructed retrospectively. Unlike the following examples, the ToC development was done mainly by the consultants, based on project documents and interviews with project personnel. It was not done using a participatory approach. The ToC focused on the three observed changes, incorporating assumptions about the main mechanisms through which the research could theoretically have contributed to those changes. A graphical representation of the ToC is provided as Fig. 1.

At each step of the ToC, four questions were asked: What happened? What were the main drivers of these changes? How did CIFOR and CIRAD contribute to these changes? What other factors should be taken into account?

Evidence was gathered from project documents, literature review and 65 key informant interviews to answer these questions. Based on this evidence, the contributions of CIFOR and CIRAD were classified into one of three categories: 1) not necessary—CIFOR and CIRAD did contribute, but their contribution was part of a package of causes and the observed changes would probably have been similar without their contribution; 2) necessary—CIFOR and CIRAD did contribute, and their contribution was necessary for the changes to be observed, in conjunction with other contributing factors; or 3) sufficient—CIFOR and CIRAD on their own caused the changes. No other factors were necessary.

The evaluation also employed a review process in which an independent peer reviewer was provided with the evidence and analysis and asked to critically assess the conclusions.

The evaluation found that CIFOR and CIRAD research had a direct influence on the international forestry agenda and policies, NGO activities, and timber company practices. The research was deemed to have made a necessary contribution to adapting international policies to the Congo Basin context. Finally, CIRAD

Table 1 | Case characteristics

<i>Characteristic/Case</i>	<i>Congo Basin</i>	<i>GCS-REDD+</i>	<i>SWAMP</i>	<i>Furniture Value Chain</i>
Research scale and geographic scope	Portfolio of research projects from 1995 to circa 2013, focusing mainly on Democratic Republic of Congo (DRC) and Cameroon, but also including other countries in the Congo Basin.	Research programme with multiple teams and with about 60 partners internationally. Primary focus: Asia Pacific (Indonesia and Vietnam), Africa (Cameroon and Tanzania), and Latin America (Brazil and Peru). Secondary focus: Asia Pacific (Lao PDR, Nepal), Africa (DRC), Latin America (Bolivia); Tertiary focus: Asia Pacific (Cambodia, Papua New Guinea), Africa (Burkina Faso, Mozambique), Latin America (Mexico).	Research project with core research team and several research partners, with activities in: Asia-Pacific (Indonesia, Vietnam, India, Cambodia, Bangladesh, Philippines, Thailand, Malaysia PNG, Fiji), Africa (Liberia, Gabon, Mozambique, Tanzania, Kenya, Madagascar, Senegal, Côté d'Ivoire Cameroon, Sierra Leone, Gambia, Benin, Togo, Niger, and Nigeria), Latin America (Mexico, Colombia, Dominican Republic, Costa Rica, Peru, Ecuador, Brazil, Honduras, and Panama).	Research project with core team and national partners working in one district in Indonesia.
Research goals	Research activities were mainly aimed to: (i) provide data and knowledge on the forests of the Congo Basin; (ii) provide operational solutions or tools for decision makers; and (iii) build capacity related to sustainable forest management in the region.	Through GCS-REDD+, CIFOR aspires to inform negotiations toward a global REDD regime, and contributes to the design and implementation of national-level REDD schemes so that the design and the implementation are efficient, equitable and provide benefits to affected communities in developing countries.	SWAMP aims to support the development of the international REDD+ mechanism in wetlands and significantly reduce technical constraints (MRV of emissions and emission reductions) throughout the tropics through better characterisation of carbon stocks in intact tropical wetlands and in the land cover types that are replacing them.	The project aims to improve the value chain efficiency, especially to benefit small-scale furniture producers.
Intended impact pathways	Multiple impact pathways were used. These include directly influencing international and national policymakers, NGOs and timber companies who were involved in the research. In addition, the research was trying to indirectly influence international development agencies and timber companies who were not directly involved in the research.	CIFOR's GCS-REDD+ informed negotiations at the UNFCCC through a formal engagement with negotiators and donors at experts meetings and influence national-level REDD policies and strategies through collaborative research and policy engagement with CIFOR's partners to support informed decision making in national and sub-national (including project) level	SWAMP achieved its goals mainly through a direct involvement of CIFOR's scientists in the development of international official guidelines for wetlands and in joint research with partners, formal engagement events with donor and policy makers, and the improvement of researchers and government technical staff' capacity.	The project aims to increase the capacity of small-scale furniture producers by moving them up the value chain, so they are closer to the consumers, and by improving their supply of raw materials. It also aims to influence local policymakers to enact policies that support small-scale furniture producers.
Intended end-of-programme outcomes	The research projects were trying to achieve better international forestry policies and forestry agenda, especially shifting from strictly conservationist views to a more mixed approach looking at management, governance, climate change and informal sector; improved national government forestry policies on management, non-timber forest products, and community forestry; a shift in NGO lobbying from purely conservational approach to a sustainable forest management approach, and; supporting timber companies in designing and	By the end of GCS-REDD+ international negotiators and national policymakers utilised GCS-REDD+ research findings in international policymaking; GCS-REDD+ research partners promoted REDD policies that are 3E+; national and subnational policymakers used GCS-REDD+ research findings for better-informed decision; and practitioners adopted GCS-REDD+ research findings in pilot projects.	By the end of SWAMP global negotiators and policymakers utilised CIFOR research on SWAMP in international policymaking; donors used CIFOR research on SWAMP in their policies; and national policymakers were aware of and used credible scientific information on tropical wetlands.	Enhanced structure and functioning of the furniture industry to benefit small-scale producers; improved marketing by small-scale producers and their organisations.

implementing sustainable forest management plans.	Norad, AUSAID.	USAID.	Australian Government, through the Australian Centre for International Agriculture Research (ACIAR).
French government, through Ministry of Foreign Affairs and the French Development Agency (AFD), European Commission.	REDD+ policies, REDD+ sub-national initiatives, measuring carbon emissions, mitigation-adaptation synergies, multi-level governance, REDD+ benefit sharing.	Quantification of GHG emission and C stocks in tropical wetlands, development of tools to quantify GHG emissions and C stock in tropical wetlands; monitoring changes in C stock due to LULUCF and global climate change.	Wood supply, occupational health and safety, gender, livelihood, value chain, market research. All pertaining to small-scale furniture producers.
Criteria and indicators of forest management, local forest management, national policies on local forest management and sustainable forest management, impact of certification.	2008 2015	2011 2015	2008 2013
CIFOR: 1995. CIRAD: 1950 s. For the purpose of the evaluation, only projects ending up to 2013 were considered.	USD 32.8 m Peer reviewed: 246 documents (As September 2014)	USD 3.8 m Peer reviewed: 44 documents (As July 2015)	USD 800 K 34 documents
118 documents			
Project budget			
Number of peer-reviewed and non-peer reviewed publications			

indirectly influenced national forest management standards. A detailed summary of the findings is provided as Annex 1 (Supplementary material).

Outcomes evaluations. The other three research evaluations each used methods that built on the tools and concepts of Outcome Mapping (Earl *et al.*, 2001), Contribution Analysis (Mayne, 2012), Collaborative Outcomes Reporting Technique (CORT) (Dart and Roberts, 2014) and the RAPID Outcome Assessment method (ODI, 2012). Each approach has its own conceptual and methodological strengths. Combining them allowed us to better accommodate the complexity of the research programmes. We provide a detailed overview of the GCS-REDD+ evaluation, and brief descriptions of SWAMP and FVC, which used similar but less elaborate approaches.

Global comparative study on reducing emissions from deforestation and forest degradation (GCS REDD+). As the principal vehicle for CIFOR's research on forests and climate change mitigation, the GCS-REDD+ programme (2009–2015) focused on identifying challenges and providing solutions to support the design and implementation of effective, efficient, and equitable REDD+ policies and projects. It is the largest research project among all cases (see project budget in Table 2). The research involved more than 60 research partner organizations in 15 countries. It was organised in four main “modules” that: 1) documented and analysed REDD+ strategies, policies and measures; 2) documented and analysed demonstration activities (i.e. REDD pilot projects) to assess and learn lessons from experience with subnational REDD+ implementation; 3) developed and analysed approaches to setting monitoring and reference levels as a contribution to the design of measurement, reporting and verification (MRV) standards, and; 4) investigated potential synergies between REDD+ and climate change adaptation approaches. The GCS-REDD+ programme was intended to contribute to improved policy and practice in sub-national REDD+ project implementation and at national and international policy levels.

The GCS-REDD+ evaluation, a participatory evaluation led by the Overseas Development Institute (ODI), asked “How well has the GCS-REDD+ programme achieved its goals, and how could it be improved?”, with seven sub-questions corresponding to the stages if the ToC.

In this case, and in each of the subsequent cases, there were implicit and partially articulated ToCs for the overall programme and for sub-components, but there was no single, explicit, documented ToC available. We used a facilitated participatory process with project personnel to retrospectively document the ToC at the overall programme level and at the country-level for three main programme countries (Indonesia, Peru and Cameroon). The ToC was iteratively refined improve precision and clarity. The ToC was conceptualised in stages, identifying the theoretical causal links from the main research activities and outputs, tailored products and engagement processes, leading to intermediate outcomes, end-of-programme outcomes and higher-level outcomes.

The concept of outcomes is critical in the analysis. Outcomes are defined as changes in knowledge, attitudes and skills manifest as changes in behaviour. The ToC identifies actors thought to be important in the change process; it also anticipates how they would use knowledge and capacity from the research process and what they would do with that knowledge. Intermediate outcomes are changes expected to happen as a proximate result of the generation and utilisation of new knowledge and associated project activities: i.e. due to actions of direct users of research and of actors/processes influenced by direct users.

Table 2 | Evaluation case characteristics

Characteristic\Case	Congo Basin	GCS-REDD+	SWAMP	Furniture Value Chain
Evaluation Questions	<ol style="list-style-type: none"> 1. To what extent did CIFOR and CIRAD contribute in the last 20 years to framing forestry issues and putting them on the international agenda, either directly or via other stakeholders? 2. In which ways did CIFOR and CIRAD help national governments designing and implementing relevant forestry policies? 3. How far did CIFOR and CIRAD activities contribute to shaping more sustainable practices in timber companies? 	<p>How well has the GCS-REDD+ programme achieved its goals, and how could it be improved?</p> <ol style="list-style-type: none"> 1. How has GCS-REDD+ activity contributed to its end-of-programme outcomes? 2. Are the target audiences using the GCS-REDD+ work? 3. Are the target audiences aware of GCS-REDD+ work? 4. Have GCS-REDD+ engagement and communication channels been effective? 5. Have GCS-REDD+ projects produced relevant science to achieve its goals? 6. Have the GCS-REDD+ programme and its projects been effectively integrated across scales (sub-national, national and global)? 7. Has the GCS-REDD+ utilised coherent strategies to achieve its outcomes? 	<p>How well has the project achieved its goals, and how could it be improved?</p> <ol style="list-style-type: none"> 1. How have SWAMP activities contributed to its end of programme outcomes? 2. Are the target audiences using the projects' outputs? 3. Are the target audiences aware of the project's outputs? 4. Has the project produced relevant science to achieve its goals? 	<p>How well has the project achieved its goals, and how could similar projects be improved in the future?</p> <ol style="list-style-type: none"> 1. Has the project produced relevant science to achieve its goals? 2. Are the target audiences aware of the project's outputs? 3. Are the target audiences using the outputs, and how are they used? 4. Has the project achieved its end of project outcomes? 5. How have the project's activities contributed to the end of project outcomes?
Number of Component Studies (e.g., country cases, whether it includes an impact assessment study (separate from the outcome assessment))	1. One overall contribution analysis, with three case studies including: Cameroon country-level case study; management and certification, and; non-timber forest products.	1. Outcome evaluation including: 1 global case study; 3 country case studies (Indonesia, Peru and Cameroon), 3 episode studies in countries where CIFOR is not working in (Philippines, Ghana and Costa Rica), 7 stories of change; 1 communication review	1. Outcome evaluation including: 1) case study assessing outcomes at UNFCCC discourse and policy level; 2) national (Indonesia) level case study. 2. Ex-ante impact assessment.	1. Outcome evaluation assessing the project achievements toward s end of programme outcomes 2. One impact assessment, assessing the impact of the end of project outcomes on the livelihood of small-scale furniture producers.
Evaluation team and leadership	Led externally	Led externally	Led internally	Led internally
External review was organised	Yes (one external)	Yes (two externals)	Yes (one external)	No
Direct Cost of evaluation	EUR 80,000, USD 15,000. Total: USD 105,000	£127,180 (ODI), USD 59,150 (CIFOR). Total: USD 241,000	USD 8,500 (external), USD 8,450 (CIFOR). Total: USD 16,950	USD 10,000 (external), USD 7,150 (CIFOR). Total: USD 17,150
Timeline (duration) of evaluation	July 2013–September 2014	September 2014–July 2015	July–October 2015	July 2015–March 2016
Theory of Change development	There was no collective theory of change for the overall portfolio of projects. An explicit theory of change was developed retrospectively by the evaluators, through discussions with the scientists and participation in a regional forestry workshop held in Cameroon.	GCS-REDD+ had had an implicit theory of change (ToC) included as part of its funding proposals. Explicit theories of change for overall programme and for each of three-country case study were retrospectively developed together with CIFOR scientists for the purpose of the assessment.	SWAMP had clear goals and outcome statements expressed in its funding proposals. An explicit theory of change was developed by the evaluation team in consultation with the project scientists for the purpose of the evaluation. The SWAMP ToC followed the template developed in the GCS-REDD+ evaluation.	The FVC project had clear goals in the proposal, and it had a clear target audience. The explicit theory of change was developed retrospectively the evaluation team and research team through a series of meetings and document reviews.
Difficulties encountered during implementation	<ul style="list-style-type: none"> • Lack of record and documentation. Many people have left the organization, taking records, documentations with them. 	<ul style="list-style-type: none"> • The novelty of the method made it difficult to find consultants to implement case studies in a consistent and correct way (i.e. for country case studies where the programme was not active). • Changing personnel and institutional memory that was not systematically recorded; respondents often forget what actually happened in specific moments. • Difficulties in identifying specific knowledge that is attributable to the research. 	<ul style="list-style-type: none"> • Time constraint for assessors to interview national policy makers with highly dynamic agenda. • Difficulties in identifying knowledge that is attributable to the research. 	<ul style="list-style-type: none"> • This was the first action research project that CIFOR evaluated. Therefore, the development of the ToC was not as straightforward. • The project was a continuation of a previous project, so it was difficult to ascertain how much was the sole achievement of this project.
Main evaluation findings (specifically on the validity of impact pathways, achievement of end-of-programme outcomes, unintended results)	<ul style="list-style-type: none"> • CIFOR increasingly became the research centre of reference on forestry issues, giving it more opportunities to contribute to international agenda. • CIFOR and CIRAD likely made a significant contribution to setting the international agenda. Their contribution was necessary for the agenda to evolve. 	There are firm indications that the research based evidence developed in the GCS-REDD + programme is influencing the development of systems that will achieve reductions in greenhouse gas (GHG) emissions from forests in ways that are effective, efficient, equitable, and will have co-benefits when a global agreement is reached. At the global level, CIFOR has	There is sufficient evidence that SWAMP research informed international and national policymakers through the active engagement of CIFOR's principal scientist in international knowledge creation and policy discussions and through direct interactions with national knowledge-sharing partners and government technical staff. A UNFCCC formal note <FCCC/SBSTA/2014/INF.1> acknowledges the SWAMP scientific output as the reference for	<ul style="list-style-type: none"> • The project built on a network established during a previous project, so it was able to establish itself much quicker. It is also focused on a small region in Indonesia, making the task less complex. • The project facilitated the establishment of a small-scale furniture association, which became the main vehicle through which capacity building and engagement,

- The changes above translated into new policies, to which CIFOR and CIRAD did contribute. There is sound evidence that multilateral organisations rely on these works, especially for review purposes, but unsure whether there was any influence on EU or most bilateral donors.
 - CIFOR and CIRAD were the only research centres able to provide input to national governments in the Congo Basin.
 - CIRAD had a necessary contribution to establishment of national management standards.
 - There is limited evidence on contribution to legal frameworks on forests. Significant evidence only exists in Cameroon.
 - CIFOR's work proved necessary to adapt international certification criteria to national forests in the region.
 - CIRAD's provided a necessary contribution to timber companies' implementation of forest management and certification.
- influenced the UNFCCC negotiations in suggesting a step-wise approach to setting reference levels, which was formally adopted in 2011. At the national level, the step-wise approach has been adopted in Ethiopia and Guyana as a direct consequence of the GCS-REDD+ programme.
- policy-makers to make sound decisions relating to the role of tropical wetlands in climate change adaptation and mitigation. The assessment also finds private companies use SWAMP research to support their projects on the sub-national level through an assistance of a CIFOR research partner.
- both with small-scale furniture producers and policy-makers occurred.
- As the association was the only one of its kind, understanding its role in building capacity and influencing policymaking was straightforward.
 - The association was influential in the development of the local government policy on small-scale furniture producers, so CIFOR could claim a substantial contribution to the policy change.
 - There was evidence that producers who were not members of the association also adopted the improved practices taught through the association. However, the direction of causality is not very clear.
 - However, the association became less active and weak after the project was completed and CIFOR staff no longer played a strong supporting role. This shows that in the future, projects need to have a proper exit strategy to ensure the sustainability of its activities well after project closure.

Another key concept is the end-of-programme outcome. These are outcomes the programme aims to contribute to and that can be reasonably expected to occur within the time-frame of the project. Projects are accountable for contributing to results at this level. Higher level results are also represented in the ToC, but are considered outside the realm of accountability because they require more time and because they depend on other variables beyond the influence of the project. The GCS-REDD+ ToC, presented visually as Fig. 2, includes policy change and subsequent conservation and development impacts beyond the end-of-programme outcome stage.

The ToC guided a purposive sampling approach and helped to identify data requirements and potential data sources. Key informants were identified to represent each of eight categories of ToC actors: 1) international agencies and donors, 2) national policy organisations (research partners), 3) national policy organisations (non-research partners), 4) government, 5) REDD+ proponents, 6) other relevant NGOs, 7) researchers and 8) local communities. A snowball sampling approach was used to identify additional respondents in each category.

Data were collected through six individual studies, including: a case study on the contribution of the research to the adoption of a particular approach to setting reference emission levels and reference levels (REL/RL) in international processes, and the degree to which that has been reflected in national level plans; a case study on the contribution of the research to REDD+ readiness (REDD+ policy development, procedures and capacity) in Indonesia; country case studies in Peru and Cameroon, where there was direct programme engagement; studies in three countries with no active country-level research programme to assess the indirect influence of the programme; seven “stories of change” to document particular examples of influence identified by CIFOR staff and other stakeholders, and; a communications review, including two surveys to assess the reach and uptake of GCS-REDD+ information.

Analysis and reporting were also done using participatory processes. A series of workshops engaged stakeholders and researchers to share, verify, correct and improve data and analyses. Data were systematically extracted from the various studies against the overall ToC and against key evaluation questions in two results charts (Dart and Roberts, 2014). The results charts present summaries of results corresponding to each ToC node and each key evaluation question respectively, with evidence reported as bullet points of relevant data by source. A companion evidence table collated references for all of the source material. The team reviewed the evidence table and provided general and detailed comments on a meta evidence table which consolidated evidence from all cases of GCS-REDD+. The process concluded with a final sense-making workshop involving research leaders and CIFOR research managers to discuss and further develop emerging findings, conclusions and recommendations. An independent external reference group provided oversight throughout the process. The final report including recommendations was prepared by ODI (Young and Bird, 2015). It recognized that the overall objectives of the program were limited by the international policy environment on REDD+ but that there is evidence that the program has had positive influences on capacity and on the discourse and development of improved systems for implementing REDD+ at international and national scales. These outcomes were achieved through: 1) production of high quality independent research and publications and extended outreach; 2) development of approaches and tools such as the step-wise approach; 3) provision of expert support at the international and national level; 4) hosting of international events and training; and 5) collaboration with and capacity development of national

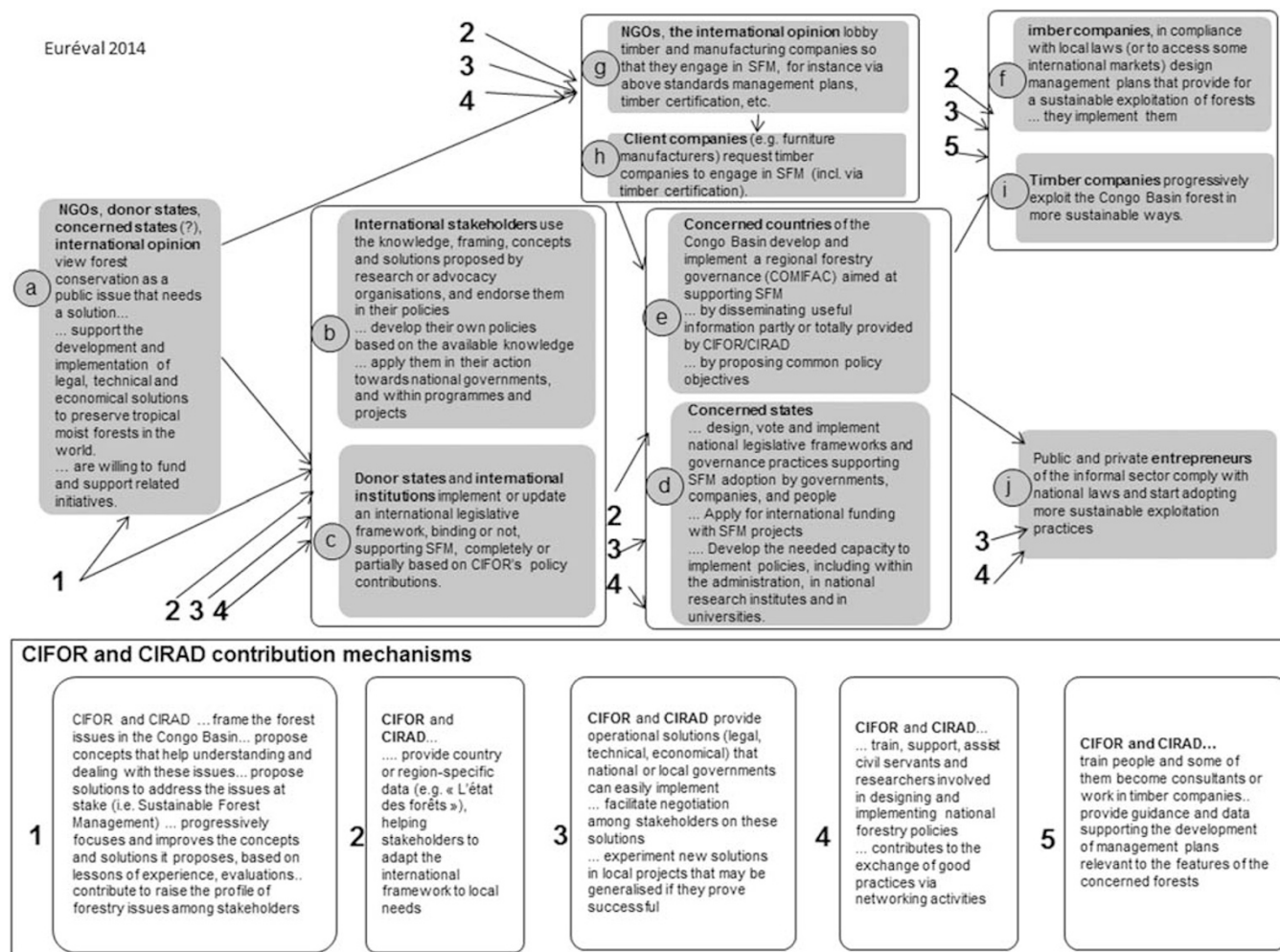


Figure 1 | Sustainable forest management portfolio theory of change. Reproduced with permission.

partners. Figure 3 illustrates the overall GCS-REDD+ evaluation process and Table 3 provides a summary of the results.

Sustainable wetlands adaptation and mitigation programme (SWAMP). The Sustainable Wetlands Adaptation and Mitigation Programme (SWAMP) started with a recognised need for more and better scientific information about the role of tropical wetlands in carbon storage, carbon emissions and climate change processes. It aimed to provide policy makers with credible scientific information needed to make sound decisions relating to climate change adaptation and mitigation strategies. The project developed methods to measure wetland forest and mangrove carbon stocks and quantified carbon stocks in below ground biomass in mangroves, tidal salt marshes, and seagrasses in 25 countries across Asia Pacific, Africa, and Latin America, as well as built capacity and informed policymakers and the general public on wetland forest and mangrove inventories. Prior to SWAMP, mangrove forests, which occur along the coasts of most major oceans in 118 countries, adding ~30–35% to the global area of tropical wetland forest over peat swamps alone, were overlooked in international climate change mitigation strategies (Donato *et al.*, 2011); wetlands as high carbon reservoirs had not been included in the UNFCCC agenda.

The SWAMP evaluation asked; “How well has the programme achieved its goals, and how could the programme be improved?” with similar sub-questions as in the GCS-REDD+ evaluation.¹ The evaluation followed a similar approach but with a

substantially smaller scope, budget and project team. The evaluation team developed a draft ToC based on project documents. Project scientists then helped revise and fine-tune the ToC and identified data sources, including potential key informants. The data collection and data integration were done by the evaluation team. The data were gathered from 16 interviews and document reviews of 45 CIFOR project documents and publications and relevant UNFCCC documents, media reports and government policy documents. Unlike the GCS-REDD+ evaluation, there was no final sense-making workshop due to resource constraints.

The evaluation found that the SWAMP programme achieved its goals and had positive unexpected influences that benefit local communities and the private sector. Areas for improvement were also identified. For example it was suggested that SWAMP could more effectively reach local governments and local communities by presenting findings in local languages. The improvement of these local audiences’ capacity for implementing sustainable wetlands management is a crucial condition for its implementation.

SWAMP engaged different target audiences and knowledge-sharing partners through joint research, formal discussion and presentations at policy events. SWAMP’s work on carbon quantification in wetlands has been influential in the climate change mitigation debate, demonstrating that systems covering just 3% of the earth’s surface store around 30% of the earth’s carbon stocks.

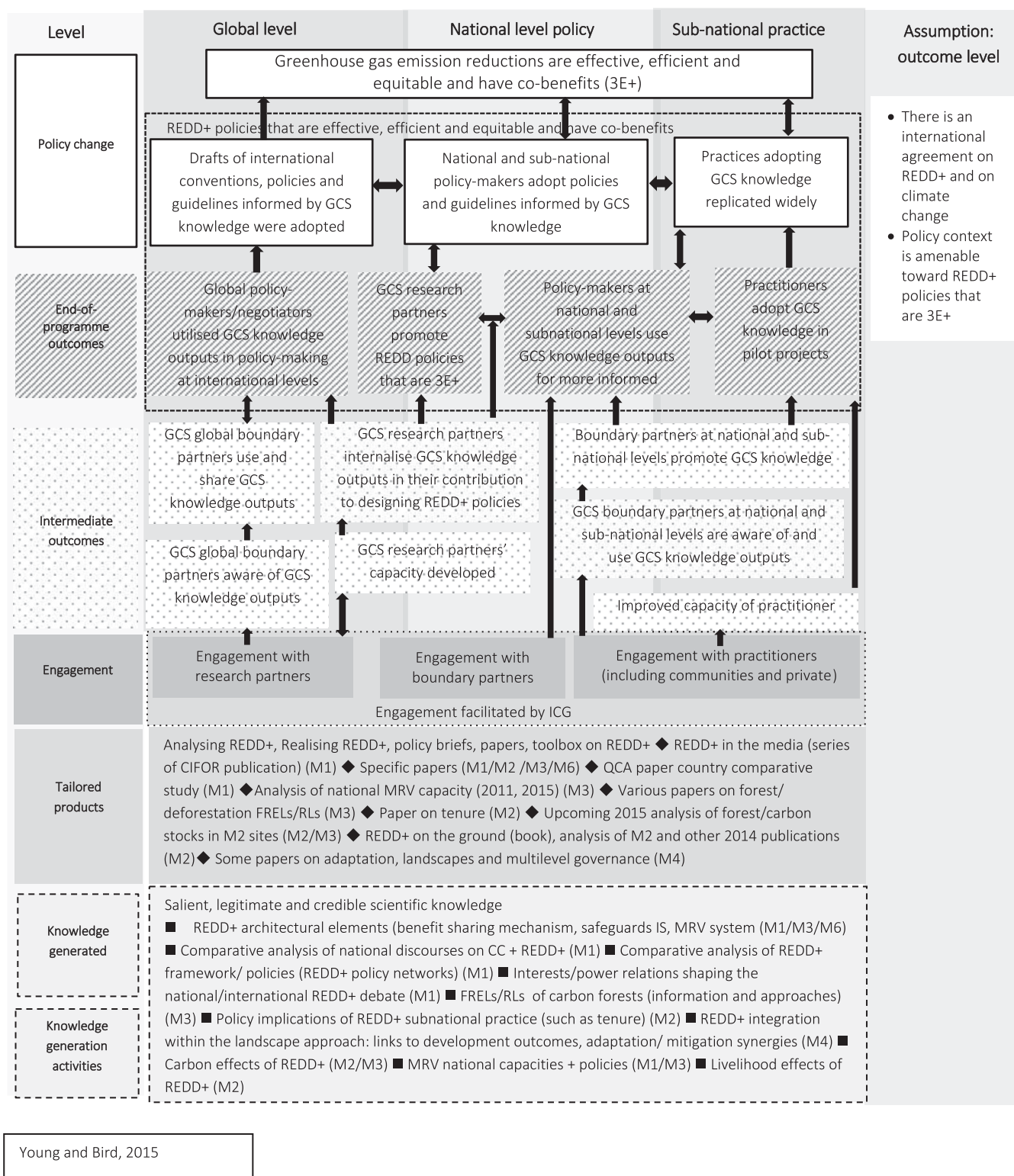


Figure 2 | The GCS-REDD+ ToC. Reproduced with permission.

Furniture value chain project (FVC). The FVC project was an action research project that aimed to improve value chain efficiency and enhance livelihoods of small scale furniture producers. The project focused on Jepara District in Central Java, Indonesia, which is home to around 18,000 small scale furniture makers. CIFOR researchers collaborated with national government and university researchers and with the Jepara Furniture Multi-

stakeholder Forum (FRK) and the Jepara district government to identify information needs, conduct relevant research, and organised events and processes to disseminate and share knowledge and recommendations. The project supported small scale furniture producers to: 1) move up the value chain, including organising participation in furniture tradeshow in Jakarta and internationally, establishing a furniture maker association, and

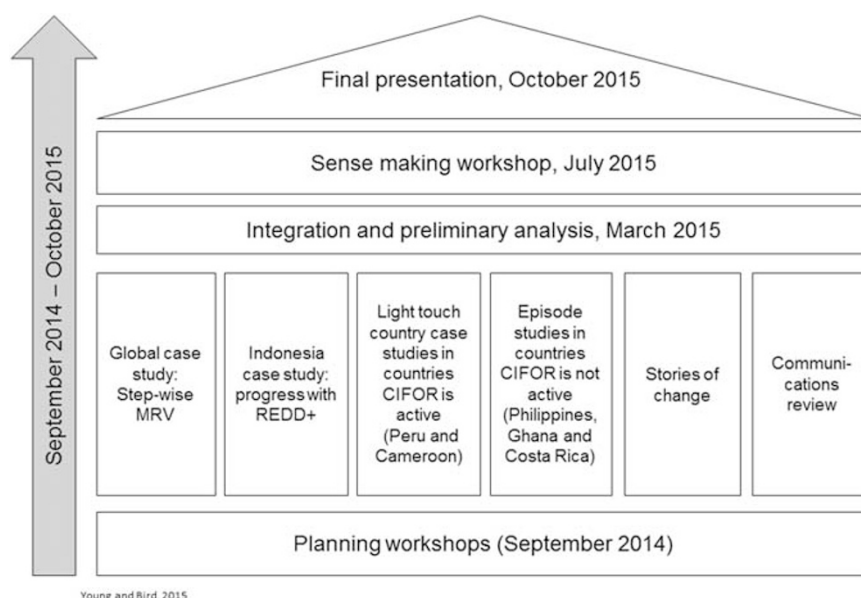


Figure 3 | GCS-REDD+ Evaluation process. Reproduced with permission.

constructing a web-based selling system; 2) secure timber supply; and 3) qualify for the Timber Legality Assurance System. The research also informed local policy processes to the extent that CIFOR was asked to write a “Jepara Furniture Roadmap” which was ultimately enacted into law.

The FVC evaluation followed the GCS-REDD+ and SWAMP evaluation model, but with a further reduced scale and budget. We were interested in trialling whether the approach could be implemented more economically. In addition, we wanted to test the approach on an action research project.

The evaluation began with a participatory retrospective construction of the theory of change, involving the evaluation team and the project team. The ToC was then used to determine testable hypotheses and interview questions and to identify respondents to be interviewed. Twenty-one interviews were conducted with respondents representing six types of actors in the ToC, over a period of two weeks. The evaluation also analysed local laws and district budget plans from 2013 to 2015, seeking evidence of influence from the research.

The evaluation found that, as an action research project, the FVC’s main impact pathway was through the establishment of a furniture association. The association became the main platform for training and facilitation activities, and served to attract the attention of the local parliament. In addition, a number of association members became champions, liaising with the local government and with the association of large-scale furniture makers. Given that the association was the only one of its kind, understanding its contributions to building capacity and influencing policymaking was straightforward. Moreover, the research team was invited to draft a “Jepara Furniture Roadmap” which was adopted almost verbatim into a local law. However, the evaluation also found that the association became less active and weak after the project was completed and CIFOR staff no longer played a strong supporting role. It calls for future projects to have a proper exit strategy and substantial adoption by local stakeholders to ensure sustainability.

Key lessons learned from four theory based evaluations

The following section presents the authors’ reflections on the four research evaluation cases based on our experience with the cases

and on discussions and feedback from stakeholders during and after each study.

Using theory of change. Each of these four research evaluations was based on ToCs that were developed retrospectively by the evaluation team alone (in the case of SFM) or using a participatory process and/or consultation with the research team. This was challenging; staff changes, incomplete recollections, and the natural evolution of ideas and understanding about how the work contributed to change processes made it difficult to develop an agreed ToC.

The process of articulating the ToC proved useful in and of itself, revealing substantial differences of understanding and opinion within research teams, including among principal investigators and between principal investigators and other team members. It also helped focus attention on the change process and challenged current, conventional understanding and expectations. This experience helped motivate a new policy at the organisation level (CIFOR) to develop explicit ToCs as part of all project planning and development. The consultant-led ToC development was less effective as a learning tool for project teams because there was less interaction and genuine engagement by researchers in the process, but also because this evaluation covered a set of research projects implemented by a larger number of researchers over a much longer time period.

The ToCs developed in these cases are still somewhat crude in their assumptions about the mechanisms of change and about external conditions. This reflects the fact that our understanding and ability to model knowledge translation, policy change, and social change generally, is still not well developed. Moreover, most of the scientists working on these research projects were not experts in policy or knowledge translation processes. One of the main values of the approach is that it made the assumptions about change processes, and about how the research contributed to those processes, explicit and testable. It is expected that future research projects and future ToCs will become more sophisticated and increasingly accurate in their assumptions and in their interventions as a result.

The ToCs were useful as analytical frameworks. They focused attention on the main evidence needs and sources. As mentioned

Table 3 | Summary of GCS-REDD+ evaluation results

<i>Evaluation question</i>	<i>Answer</i>
How well has the GCS-REDD+ achieved its goals, and how could it be improved:	The main goal of the GCS-REDD+ programme is hugely ambitious for a research organisation like CIFOR, especially in such a highly contested policy arena, and clearly remains out of reach in the absence of a global agreement on REDD+. There is evidence, however, that the GCS-REDD+ programme is influencing the development of systems that are effective, efficient, equitable for when a global agreement is reached. CIFOR has influenced the United Nations Framework Convention on Climate Change (UNFCCC) to adopt a step-wise approach to setting reference levels internationally. The approach has also been adopted in several countries and CIFOR has demonstrably increased the capacity and influenced the behaviour of national policy-makers and other relevant actors.
How has GCS-REDD+ activity contributed to its end-of-programme outcomes	CIFOR has achieved these outcomes in five main ways: 1) production of high quality independent research and publications and extended outreach; 2) development of approaches and tools such as the step-wise approach; 3) provision of expert support at the international and national level; 4) hosting of international events and training; and 5) collaboration with and capacity development of national partners.
Are target audiences using the GCS-REDD+ work?	The GCS REDD+ ToC identifies six categories of actor expected to use CIFOR research: 1) national research partners; 2) proponents (national organisations involved in pilot projects); 3) national practitioners (operational agencies and practitioners, including communities, the private sector and the media; 4) national policymakers; 5) international research partners; 6) and international policy actors (e.g. Intergovernmental Panel on Climate Change (IPCC), bilateral and multilateral donors). The assessment found strong evidence that all of these stakeholders are using CIFOR research in their work.
Are the target audiences aware of GCS-REDD+ work?	There is widespread awareness of CIFOR's work among the people contacted as part of this assessment. This is not surprising, however, since the assessment used CIFOR mailing lists for the surveys and interacted largely with those working with CIFOR staff for the country studies. However, the fact that many people are using the work indicates that they must be aware of the work.
Have GCS-REDD+ engagement and communication channels been effective?	GCS-REDD+ uses a very wide range of channels for communication and engagement including digital (web and social media), publications, events and conferences, research collaboration, personal engagement, formal engagement with national governments and practical engagement with practitioners. The assessment found strong evidence that all of these channels are being used effectively. CIFOR has a particularly strong digital strategy to reach global audiences, CIFOR publications are well regarded and frequently consulted, international events and conferences are well known, well attended and attract high level participants. Respondents in the national case studies and the Indonesia Country Study workshop asked, however, for more national events. Research collaboration is clearly a very effective element of CIFOR's engagement work. CIFOR also engages effectively with a wide range of practitioners at the national level, including proponents (organisations testing REDD+ approaches at the field level).
Have GCS-REDD+ projects produced relevant science to achieve its goals?	The GCS-REDD+ programme has produced a vast range of relevant and useful science including five books and 84 book chapters, 157 journal articles, 48 working and occasional papers, 48 policy and information briefs, and five doctoral theses.
Has the GCS-REDD+ programme and its projects been effectively integrated across scales (sub-national, national and global)	The GCS-REDD+ was designed from the start as an integrated programme with interlocking components. Management and coordination mechanisms have been flexible and responsive and have evolved appropriately, as the programme has developed, but this seems to work best at international level among staff based in Indonesia, whereas some country-level staff have described not being fully aware of what other components are doing.
Has the GCS-REDD+ used coherent strategies to achieve its outcomes?	The original project design has been strengthened by the development of a global ToC identifying specific boundary partners and uptake pathways, but there remains some confusion among lower-level staff about the theory and practice of ToC.

Source: Young and Bird (2015: 7-8)

above, they guided the hypotheses that were then tested during the evaluations. In addition to assessing whether a particular change occurred, the framework facilitated analysis of causes and effects in the change process.

One of the main benefits of using a theory-based approach is that it promotes the formulation and testing of hypotheses about change processes and about the role of research/knowledge

so they are explicit and testable. Case studies are better suited than other research evaluation tools to provide nuanced, in-depth understanding of outcomes (Morgan and Grant, 2013). Structured as an hypothesis testing exercise, a theory-based research evaluation provides empirical evidence and analysis of the role of research in social change processes.

Data collection and management. The ToCs provided good frameworks for determining data needs and sources. However, data collection was challenging. The evaluations used mixed methods approaches. Data for early stages of the ToC were sourced primarily from project personnel and project documents. This is straightforward, but with larger and older programmes not all unpublished documents could be located. Staff turnover and the passing of time limited recollection.

The concept of “tailored products” encompasses the idea that particular kinds of information can and should be collated and delivered to particular audiences using appropriate media in a timely way. The evaluations found that the audiences needed to be more carefully distinguished and delivery media needed to be more effectively tailored. Data for assessing later stages in the ToC were collected from published and unpublished documents and from interviews and surveys of key informants. Again, as time elapses it becomes more difficult to identify, locate and source grey literature, and memories fade.

More challenging perhaps is that typical respondents (i.e. government, private sector or civil society actors) working in or observing policy processes do not necessarily understand or analyse those processes in terms of the role or influence of knowledge. Therefore, they may not consider the research project interventions important or refer to them in their own narrative of the change process. Even if they do, they may not be concerned about the source of knowledge used. It is inherently difficult to trace the source and the influence of ideas and knowledge, especially in processes that are by definition political. This is consistent with the observations of Klautzer *et al.* (2011) that policy makers found it difficult to recall specifics of research inputs.

Separate from this challenge, there were also weaknesses in survey design and implementation in some cases. Some questions were poorly worded and inexperienced interviewers introduced potential biases. This learning is a natural part of developing and testing new methods.

The methods, analytical framework (the ToC), data, analysis and limitations were well-documented and transparent in each case. The results charts used in the GCS REDD+, FVC, and SWAMP evaluations, which logs all relevant data supporting or contradicting each step in the ToC, is an excellent way to summarise and present data as evidence and it helps establish the credibility of the process. The SFM Congo Basin evaluation used documentary analysis and developed a corpus of interviews. The results chart approach is more systematic and transparent.

Data analysis. The SFM evaluation was primarily backward-looking. It considered the changes that had taken place and looked for evidence that the research had contributed to those changes. It explicitly considered causality, asking if the same outcomes would have been observed without the research activities and outputs, with a methodical assessment of the merits of alternative explanations. The conclusions are plausible and well documented, but they are not conclusive and will not be fully convincing to some. Less attention was paid to the inherent quality of the research and outputs; the focus was on whether and how it influenced the policy process.

The three outcomes evaluations (GCS-REDD+, SWAMP and FVC) paid attention to all stages of the research-to-impact process, with evaluation questions to address each stage, from research design and implementation, relevance and actual use of outputs, through the hierarchy of the results of uptake and use. Identification and use of “end-of-programme outcomes” was useful conceptually and practically. It helped identify reasonable targets within the time and resource scope of the project being

evaluated, while still theorising subsequent outcomes and impacts. This provides a partial solution to the problem of time lags.

Theoretical causal links were tested at each stage, marshalling evidence from various sources and triangulating wherever possible. This was easier to do and more successful in the FVC case with its limited geographic and sectoral scope. In addition, it was straightforward to eliminate alternative explanations for the policy change that occurred in that case; the law copied a document produced by the project almost verbatim, so the causality between the research and the outcome was clear.

One exercise in the GCS-REDD+ evaluation built on an approach used by Redstone (2013) to estimate the contribution of the research to changes in six conditions considered essential for the effective implementation of policy change: functioning institutions; responsive and accessible supporting research; a feasible, specific and flexible solution; powerful champions in the key institutions; a well-planned, led and supported campaign; and a clear implementation path (Redstone, 2013). The exercise was valuable as a thought experiment. However, the results were subjective and biased, with only researchers in the room. A similar exercise with broader representation of stakeholders is recommended.

Independence and objectivity. In contrast to typical requirements that evaluations should be independent, in these cases involving the researchers in evaluation design and analysis was more conducive to learning. For example, in the GCS-REDD+ evaluation, the researchers were involved in a one-day “sense making” workshop, where they examined the evidence together with the evaluation team. The researchers themselves developed a list of lessons learned and recommendations for future research. It seems safe to assume that those lessons will be better internalised and applied than a set of recommendations produced by an external evaluator. The approach is still open to potential criticism of lack of independence and objectivity. This is answered, at least in part, by the careful and transparent documentation of methods, the ToC and the results chart. Also, reputable external organisations were contracted to lead the evaluations, with additional review by a “reference group” (GCS-REDD+) and/or peers (FVC, SWAMP, SFM).

Reception by key audiences. These evaluations had three main audiences: 1) researchers involved in the projects; 2) research managers (CIFOR management); 3) research funders and the broader scientific and development communities. Each group has different interests and expectations.

As discussed above, project researchers found the experience valuable for learning. They developed new conceptual and methodological understanding and tools, and learned directly from successes and failures in their own projects. This is already being translated into revised and improved research design and implementation in new and ongoing projects. Being familiar with the case, and with transparent presentation of the analytical framework (ToC), data, analysis (results chart, sense-making workshop) and limitations, researchers are able to assess the evaluation independently and, in these cases, considered the analysis credible.

The research evaluations have been valuable to CIFOR management in two main ways: providing learning about how to design and implement research-for-development (formative value), and helping to demonstrate the impact and value of research to funders (summative value). Management appreciation of the approach is illustrated by the fact that CIFOR has developed and endorsed a new “Planning, Monitoring and

Learning Strategy” built around a theory-based approach to research design and evaluation. As with the project researchers, they have knowledge and skills to assess the quality of the evaluation and, in these cases, found the analysis credible.

Funders have expressed appreciation for each of these evaluations and have used them in their own efforts to marshal resources for research funding. They recognise the difficulties of trying to prove outcomes of policy-oriented research and of efforts to try to influence conservation and development in complex socio-ecological systems. The ToCs have proven useful as tools for communicating with donors and developing shared understanding of how research contributes to change, and of what expectations are reasonable. The transparent analysis and reporting provides defensibility. However, the lack of a true counterfactual and quantitative analysis compromises credibility for some audiences in the broader scientific and development communities. In these cases, project funders have been satisfied with the results. The SFM evaluation was done by an independent external consultant with an additional independent peer review. This independence may bolster credibility for some audiences.

Assessing the approach against criteria. Based on this experience, including feedback and ideas from stakeholders (researchers and their partners, research managers, funders) and our own reflections, we present a summary assessment of the overall theory-based research evaluation approach as employed in these four cases against the five criteria, with a brief discussion of the strengths and weaknesses, following from the discussion in section 6 above. We also provide recommendations to improve the design and implementation of the approach. (Table 4)

Conclusions

A theory-based research evaluation approach that focuses on outcomes and on the pathways and mechanisms by which research contributes to change processes proved to be practicable and useful in these four case studies. The research cases ranged considerably in scale and geographic scope, from a large portfolio of forestry research conducted over nearly two decades in several countries of the Congo Basin to a relatively small action-research project in one district of Indonesia. The research evaluations likewise ranged in size and scope, but all used a theory-based

Table 4 | Assessment of theory-based research evaluation based on four case studies

<i>Criterion</i>	<i>Assessment of Theory-based approach</i>	<i>Recommendations</i>
Credible to main audiences	<ul style="list-style-type: none"> • Project researchers—high • Research managers—high • Research funders, scientific and development communities—Medium 	<ul style="list-style-type: none"> • Use results charts for transparent data management and presentation • Carefully assess alternative hypotheses for change processes (Mayne, 2012) • Develop and refine hypothetical counterfactual analysis (e.g. Redstone, 2013)
Contribute to the broader evidence base	Theory-based approach involves formulation and testing of hypotheses about change processes and about the role of research/knowledge so they are explicit and testable. The case studies provided empirical evidence and analysis of the role of research in policy and social change processes. Some case-specific challenges were experienced in execution.	<ul style="list-style-type: none"> • Build on social-scientific theory to add nuance and detail to theories of change • Improve training for researchers to ensure high quality execution in data collection
Inform decision making aimed at improvement (formative value)	Demonstrated learning in process by research teams; building a series of comparable case studies can support systematic learning about research contributions to change to inform research design and research management at the organisation level	<ul style="list-style-type: none"> • Develop explicit ToC as part of project/programme design to facilitate monitoring, adaptive management and evaluation of research outcomes • Develop series of comparative cases studies
Inform decision making aimed at selection, continuation or termination (summative value)	This form of research evaluation is highly context-dependent and focused on assessing achievements against purpose and intent of an individual project or program. It is possible to compare relative achievements but not possible to make absolute comparisons. With an explicit theory of change in place from the start of each project, it will be possible to assess “adaptive management” (i.e. responses to problems encountered during implementation), not only achievement of original plans and expectations.	<ul style="list-style-type: none"> • Develop explicit ToC as part of project/programme design to facilitate assessment for continuation • Develop standard protocols for theory-based research evaluation to facilitate comparison
Cost effective	The cost of implementing these research evaluations ranged from 0.4% to 2.4% of the project budget. Although more theory-based evaluations need to be done before their costs can be measured more consistently, the current range falls well within the proportion of project budget usually allocated to monitoring and evaluation (ITAD, 2014). As with case studies more generally (Morgan and Grant, 2013), the transactions costs are relatively high compared to other research evaluation approaches. It is expected that costs can be reduced, especially with explicit ToC development performed as part of project planning to streamline the process.	<ul style="list-style-type: none"> • Develop standard protocols for theory-based research evaluation to facilitate comparison

approach, with a theory of change as the main conceptual and analytical tool.

The approach as applied in these cases shares the main strengths and weaknesses of case studies more generally: it supports nuanced, in-depth understanding and covers a range of kinds of outcomes using a mixture of qualitative and quantitative data in a contextualised way, but the costs are high and generalisability and comparability are limited (Penfield *et al.*, 2013; Morgan and Grant, 2013). Some of the main advantages realised in the case studies are:

1. The use of an explicit ToC as the analytical framework helps identify and test hypothesis about how the research contributed to change
2. The methods as applied in these evaluations used well-organised evidence bases; we recommend the use of results charts and evidence table for the systematic and transparent management and presentation of evidence
3. The approach facilitates learning at the project or programme scale and provides a base for generalisable learning about how research contributes to outcomes and impacts and about how to design research to be more effective
4. The clear delineation of end-of-programme outcomes helps to manage expectations about the kind and extent of “impact”. The approach has been highly valuable as a learning tool for researchers and research managers and it has facilitated communication with funders about actual and reasonable research contributions. Evaluations that employed a participatory approach with project scientists and partners noticeably supported team learning about completed work and about possible adaptations and improvements for future projects.

Theory-based research evaluation is well suited to the research-for-development types of projects represented by the case studies. Such research tends to be inter- or transdisciplinary with an explicit focus on impact beyond the academic. It should also be applicable to other research that aspires to have a societal impact. As demonstrated in the case studies, it is useful to have explicit and deliberate planning for knowledge translation. A ToC provides a good framework for research evaluation; making it deliberate and explicit also influences and potentially improves research design and implementation.

There is still need for improvement in the design, especially in terms of assessing causality and in the implementation of data collection. Further work is also needed to draw on social scientific theories of knowledge translation and policy processes and to further test more sophisticated ToCs. Overall, this theory-based approach to research assessment generates a substantial and credible body of evidence for research outcomes and effectiveness, and supports learning and adaptation within research programmes. The approach is valuable as part of a system in which the intended contributions of research are deliberate, explicit and testable, which improves our ability to gather evidence, assess and communicate our outcomes and impacts for enhanced accountability, and our ability to learn from our experience.

Notes

- 1 A separate impact assessment to estimate the magnitude of the contribution of SWAMP to GHG emissions reduction is underway at the time of writing.

References

- Banzi R, Moja L, Pistotti V, Facchini A and Liberati A (2011) Conceptual frameworks and empirical approaches used to assess the impact of health research: An overview of reviews. *Health Research Policy and Systems*; **9** (26).

- Behn RD (2003) Why measure performance? Different purposes require different measures. *Public administration review*; **63** (5): 586–606.
- Belcher BM, Rasmussen KE, Kemschaw MR and Zornes DA (2016) Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*; **25** (1): 1–17.
- Berkes F (2009) Evolution of co-management: Role of knowledge generation, bridging organizations and social learning. *Journal of environmental management*; **90** (5): 1692–1702.
- Better Evaluation. (n.d) *Rainbow Framework*. Available from: <http://betterevaluation.org/sites/default/files/Rainbow%20Framework.pdf>, accessed 22 April 2016.
- Buxton M and Hanney S (1996) How can payback from health services research be assessed? *Journal of Health Services Research*; **1** (1): 35–43.
- Canadian Academy of Health Sciences. (2009) *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*. CAHS: Ottawa.
- Canadian Federation for the Humanities and Social Science. (2014) *The Impact of Humanities and Social Science Research*, Working Paper, Available from: <http://www.ideas-idees.ca/sites/default/files/2014-10-03-impact-project-draft-report-english-version-final2.pdf>, accessed 1 October 2014.
- Carter IM (2013) The impact of impact on universities: Skills. Resources and organisational structures. In: Dean A, Wykes M and Stevens H, (eds). 7 Essays on Impact. DESCRIBE Project Report for Jisc. University of Exeter.
- CGIAR. (2015) *CGIAR Strategy and Results Framework 2016–2025*. <https://library.cgiar.org/bitstream/handle/10947/3746/CGIAR%20Strategy%20and%20Results%20Framework%202016%E2%80%932025%20-%20Final%20Consultation.pdf?sequence=1>.
- CIFOR. (2017) <http://www.cifor.org/about-cifor/>.
- CIFOR. (2016) *CIFOR Strategy 2016–2025*. http://www.cifor.org/publications/pdf_files/Books/CIFORStrategy2016.pdf.
- Centre for Theory of Change. (n.d.) *What is Theory of Change?* Available from: <http://www.theoryofchange.org/what-is-theory-of-change/>, accessed 26 April 2016.
- Chambers R (2015) Perverse Payment by Results: frogs in a pot and straitjackets for obstacle courses. Participation, Power and Social Change Blog, Institute of Development Studies. <https://participationpower.wordpress.com/2014/09/03/perverse-payment-by-results-frogs-in-a-pot-and-straitjackets-for-obstacle-courses/>.
- Clark WC and Dickson N (2003) Sustainability science: The emerging research program. *PNAS*; **100** (14): 8059–8061.
- Clark WC *et al.* (2011) *Boundary work for sustainable development: Natural resource management at the Consultative Group on International Agricultural Research (CGIAR)*. PNAS early edition. www.pnas.org/cgi/doi/10.1073/pnas.0900231108.
- Consultative Group on Agricultural Research (CGIAR). (2011) *A Strategy and Results Framework for the CGIAR*. http://library.cgiar.org/bitstream/handle/10947/2608/Strategy_and_Results_Framework.pdf?sequence=4.
- Coryn CLS, Noakes LA, Westine CD and Schröter DC (2011) A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*; **32** (2): 199–226.
- Dart J and Roberts M (2014) *Collaborative Outcomes Reporting*. Melbourne. Available from: <http://betterevaluation.org/sites/default/files/COR.pdf>.
- Donato D, Kauffman JB, Murdiyarso D, Kurnianto S, Stidham M and Kanninen M (2011) Mangroves among the most carbon-rich forests in the tropics. CIFOR, Bogor. Available from: <http://www.cifor.org/library/3401/mangroves-among-the-most-carbon-rich-forests-in-the-tropics/>.
- Donovan C (2011) State of the art in assessing research impact: Introduction to a special issue. *Research Evaluation*; **20** (3): 175–179.
- Donovan C and Hanney S (2011) The ‘payback framework’ explained. *Research Evaluation*; **20** (3): 181–183.
- Earl S, Carden F and Smutylo T (2001) *Outcome mapping: building learning and reflection into development programs*. Evaluation. International Development Research Centre, Ottawa.
- Euréal. (2014) Evaluation of CIFOR’s and CIRAD’s Contribution to Sustainable Forest Management in Central Africa. Final Report to CIFOR (unpublished).
- Gamel-McCormick C (2011) *A critical look at the creation and utilization process of logic models*. Doctoral dissertation: University of Delaware.
- Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P and Trow M (1994) *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. Sage Publications: London.
- Gibbons M and Nowotny H (2000) The potential of transdisciplinarity. In: Thompson KJ, Haberli R, Scholz RW, Grossenbacher-Mansuy W, Bill A and Welfi M (eds). *Transdisciplinarity: Joint Problem Solving among Science, Technology, and Society*. Birkhäuser Verlag: Basel.
- Greenhalgh T, Raftery J, Hanney S and Glover M (2016) Research impact: A narrative review. *BMC medicine*; **14** (1): 1.
- Guthrie S, Wamae W, Diepeveen S, Wooding S and Grant J (2013) *Measuring Research: A Guide to Research Evaluation Frameworks and Tools*. RAND Corporation: Santa Monica, CA.
- Hanney SR, Grant J, Wooding S and Buxton MJ (2004) Proposed methods for reviewing the outcomes of health research: The impact of funding by the UK’s

- 'arthritis research campaign'. *Health Research Policy and Systems/BioMed Central*; **2** (4).
- Holling CS (1993) Investing in research for sustainability. *Ecological applications*; **3** (4): 552–555.
- ITAD. (2014) *Investing in Monitoring, Evaluation and Learning: Issues for NGOs to Consider*. ITAD: Hove, UK.
- King's College London and Digital Science. (2015) *The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies*. HEFCE: Bristol, United Kingdom.
- Klautzer L, Hanney S, Nason E, Rubin J, Grant J and Wooding S (2011) Assessing policy and practice impacts of social science research: The application of the Payback Framework to assess the Future of Work programme. *Research Evaluation*; **20** (3): 201–209.
- Kristjansson P *et al* (2009) Linking international agricultural research knowledge with action for sustainable development. *Proceedings of the national academy of sciences*; **106** (13): 5047–5052.
- Martin BR (2011) The research excellence framework and the 'impact agenda': Are we creating a Frankenstein monster? *Research Evaluation*; **20** (3): 247–254.
- Mayne J (2008) *Contribution analysis: An approach to exploring cause and effect. Institutional Learning and Change Brief*. Institutional Learning and Change (ILAC) Initiative, Rome. Available from [http://dmeforpeace.org/sites/default/files/Mayne_Contribution Analysis.pdf](http://dmeforpeace.org/sites/default/files/Mayne_Contribution%20Analysis.pdf).
- Mayne J (2012) Contribution analysis: Coming of age? *Evaluation*; **18** (3): 270–280.
- Mayne J and Stern E (2013) *Impact Evaluation of Natural Resource Management Research Programs: A Broader View*; ACIAR Impact Assessment Series Report No. 84. Australian Centre for International Agricultural Research: Canberra, 79.
- Mayne J, Stern E and Douthwaite B (2013) *AAS Practice Brief: Evaluating natural resource management programs*. CGIAR Research Program on Aquatic Agricultural Systems: Penang, Malaysia.
- Meier W (2003) Results-Based Management: Towards A Common Understanding Among Development Cooperation Agencies. Discussion Paper for the Canadian International Development Agency. <http://www.managingfordevelopmentresults.org/documents/Results-BasedManagementDiscussionPaper.pdf>.
- Molas-Gallart J and Tang P (2011) Tracing 'productive interactions' to identify social impacts: An example from the social sciences. *Research Evaluation*; **20** (3): 219–226.
- Morgan JM and Grant J (2013) Making the grade: Methodologies for assessing and evidencing research impact. In: Dean A, Wykes M and Stevens H, (eds). 2013. 7 Essays on Impact. DESCRIBE Project Report for Jisc. University of Exeter.
- Nutley SM, Walter I and Davies HT (2007) *Using evidence: How research can inform public services*. Policy press.
- ODI. (2012) *RAPID Outcome Assessment Toolkits*. Overseas Development Institute: London. Available from <https://www.odi.org/publications/6800-rapid-outcome-assessment>.
- OECD. (1991) *The DAC Principles for the Evaluation of Development Assistance*, Glossary of Terms Used in Evaluation, in 'Methods and Procedures in Aid Evaluation', OECD (1986), and the *Glossary of Evaluation and Results Based Management (RBM) Terms*, OECD (2000).
- Penfield T, Baker MJ, Scoble R and Wykes MC (2013) Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*; **23** (1): 21–32.
- Redstone. (2013) <http://www.redstonestrategy.com/wp-content/uploads/2013/09/2013-09-30-IDRC-Helping-think-tanks-measure-impact.pdf>.
- Research Excellence Framework (REF). (2011) *Research Excellence Framework 2014: Assessment framework and guidance on submissions*. Reference REF 02.2011. UK: REF. <http://www.ref.ac.uk/pubs/2011-02/>.
- Sjöstedt M (2013) Aid effectiveness and the Paris declaration: A mismatch between ownership and results-based management? *Public administration and development*; **33** (2): 143–155.
- Spaapen J and van Drooge L (2011) Productive interactions as a tool for social impact assessment of research. *Research Evaluation*; **20** (3): 211–218.
- Spaapen J and Sylvain C (1994) Societal Quality of Research: Toward a Method for the Assessment of the Potential Value of Research for Society: Science Policy Support Group.
- Stern E, Stame N, Mayne J, Forss K, Davies R and Befani B (2012) Broadening the Range of Designs and Methods for Impact Evaluations. DFID Working Paper 38, London: Department for International Development. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf.
- Van Kerkhoff L and Lebel L (2006) Linking knowledge and action for sustainable development. *Annual Review of Environment and Resources*; **31**, 445–477.
- Vogel I (2012) *Review of the use of 'Theory of Change' in international development Review Report*, DFID, London. Available at: http://www.dfid.gov.uk/r4d/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf.
- Weiss C (2007) Theory-based evaluation: Past, present and future. *New Directions for Evaluation*; **2007** (114): 68–81.
- White H (2009) Theory-Based Impact Evaluation: Principles and Practice. 3ie working Paper 3. http://www.3ieimpact.org/media/filer_public/2012/05/07/Working_Paper_3.pdf.
- White H and Phillips D (2012) Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework. International Initiative for Impact Evaluation Working Paper 15. <http://mande.co.uk/blog/wp-content/uploads/2012/06/White-Phillips-Small-n-Impact-Evaluation-WP-version.docx>.
- Wickson F, Carew A and Russell AW (2006) Transdisciplinary research: Characteristics, quandaries and quality. *Futures*; **38** (9): 1046–1059.
- Young J and Bird N (2015) Informing REDD+: An Assessment of the Center for International Forestry Research's Global Comparative Study. Final Report to CIFOR (unpublished).

Data availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Acknowledgements

The ideas in this paper have been developed and improved through discussions with numerous colleagues including John Young, Terry Smutylo, Paul Duignan, Bethany Davies, Thomas Delahais, Boru Douthwaite, John Mayne, Julien Colomer and Robert Nasi. Individual evaluations were led by Thomas Delahais (SFM), John Young (GCS-REDD+) and the three authors (SWAMP, FVC). Shilpa Soni reviewed an earlier draft. The work of the first author is supported by funding from the Canada Research Chairs program. Funding support from the Canadian Social Sciences and Humanities Research Council (SSHRC) and from the 'Improving the way knowledge on forests is understood and used internationally' (KNOWFOR) project of the Centre for International Forestry Research (CIFOR), funded by UK DfID, are also gratefully acknowledged.

Additional information

Competing interests: The Authors declare no competing financial interests.

Reprints and permission information is available at http://www.palgrave-journals.com/pal/authors/rights_and_permissions.html

How to cite this article: Belcher B, Suryadarma D and Halimanjaya A (2017) Evaluating policy-relevant research: lessons from a series of theory-based outcomes assessments. *Palgrave Communications*. 3:17017 doi: 10.1057/palcomms.2017.17.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>