



Australian Government
Australian Centre for
International Agricultural Research

Impact evaluation of natural resource management research programs: a broader view

ACIAR IMPACT ASSESSMENT SERIES

84

Research that works for developing countries and Australia

Impact evaluation of natural resource management research programs: a broader view

John Mayne

Adviser on Public Sector Performance and
Adjunct Professor, University of Victoria, Canada

Elliot Stern

Emeritus Professor of Evaluation Research
Lancaster University, UK

Visiting Professor, School of Policy Studies
University of Bristol, UK



ACIAR

Research that works for developing
countries and Australia

aciarc.gov.au

The Australian Centre for International Agricultural Research (ACIAR) was established in June 1982 by an Act of the Australian Parliament. ACIAR operates as part of Australia's international development cooperation program, with a mission to achieve more productive and sustainable agricultural systems, for the benefit of developing countries and Australia. It commissions collaborative research between Australian and developing-country researchers in areas where Australia has special research competence. It also administers Australia's contribution to the International Agricultural Research Centres.

Where trade names are used this constitutes neither endorsement of nor discrimination against any product by the Centre.

ACIAR IMPACT ASSESSMENT SERIES

ACIAR seeks to ensure that the outputs of the research it funds are adopted by farmers, policymakers, quarantine officers and other beneficiaries. In order to monitor the effects of its projects, ACIAR commissions independent assessments of selected projects. This series of publications reports the results of these independent studies. Numbers in this series are distributed internationally to selected individuals and scientific institutions, and are also available from ACIAR's website at <aciar.gov.au>.

© Australian Centre for International Agricultural Research (ACIAR) 2013

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced by any process without prior written permission from ACIAR, GPO Box 1571, Canberra ACT 2601, Australia, aciar@aciar.gov.au

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 pp.

ACIAR Impact Assessment Series – ISSN 1832-1879 (print); ISSN 1839-6097 (online)

ISBN 978 1 922137 40 1 (print)

ISBN 978 1 922137 41 8 (online)

Editing and design by Clarus Design

Printing by Elect Printing

Foreword

Natural resource management research (NRMR) has a key role in improving food security and reducing poverty and malnutrition. NRMR programs seek to modify natural systems in a sustainable way in order to benefit the lives of those who live and work within these natural systems—especially in rural communities in the developing world.

Evaluating the effectiveness of NRMR through the usual avenues of impact evaluation has posed distinct challenges. Many impact assessments focus on estimating net economic benefits from a project or program, and often are aimed at providing evidence to investors that their funds have been well spent. They have tended to focus on a specific causal evaluation issue: to what extent can a specific (net) impact be attributed to the intervention?

While many evaluations of NRMR programs and their projects will continue to use an impact assessment perspective, this report lays out a complementary approach to NRMR program evaluation. The approach focuses more on helping NRMR managers and stakeholders to learn about their interventions and to understand why and how outcomes and impacts have been realised (or, in some cases, have not).

Thus, a key aim here is to position NRMR impact evaluation as a learning process undertaken to improve the delivery and effectiveness of NRMR programs by developing a new framework for thinking about and designing useful and practical evaluations.

The emphasis on learning follows from the view of NRMR as operating under dynamic, emergent, complex and often unpredictable human and ecological conditions. In such a setting, adaptive management informed by careful responses to new information and understanding is essential for building and managing more-effective programs and interventions. This is highlighted by examining some specific examples: the CGIAR Research Program on Aquatic Agricultural Systems (led by Worldfish), CGIAR's Ganges Basin Development Challenge, and CSIRO–AusAID's African Food Security Initiative.

The alternative approach presented here is another tool to use in the search for understanding of how and why impacts occur in a research, development and extension environment. We hope that the learning-orientated evaluation described will help elucidate more soundly based explanations that will guide researchers in replicating, scaling up and improving future programs.



Nick Austin
Chief Executive Officer, ACIAR



Stephen Hall
Director General, WorldFish

Contents

Foreword	3
Abbreviations	7
Acknowledgments	8
Executive summary	9
Introduction	11
The context for evaluating NRMR programs	12
An overview of the report	12
Terminology	13
Evaluating NRMR programs	14
A focus on learning and refocusing accountability	15
The timing for impact evaluation	16
The unit of analysis: what to evaluate?	17
Causality and contributory causes	18
The challenge	18
Concepts of causality	19
NRMR programs as causal packages	20
Probabilistic causes	21
Fish farming as an example of a causal package	22
Theory-based approaches in evaluation	23
Useful theories of change	23
Theories of change incorporate causal packages	26
Theories of change can evolve and can be at different levels	26
Using theories of change ex ante and ex post	27
Key features of NRMR program theories of change	29
Generic NRMR results	29
Issues around NRMR theories of change	32
Key features of generic NRMR theories of change	34
Evaluation approaches: an overview	36
Evaluation questions	37
Key issues and concerns	37
A framework for NRMR evaluation questions	37
Designs for NRMR programs	40

Relating tools, methods and designs to the evaluation questions	42
Addressing rationale	42
Addressing implementation	42
Measuring results	43
Demonstrating and explaining causality: causal designs.....	43
Addressing sustainability.....	45
Addressing generalisation and transfer	47
Summary of implications for addressing evaluation questions	49
Relating program attributes to methods and designs	51
Application to existing CGIAR research programs	55
Exploring a broad NRMR program: the CGIAR Program on Aquatic Agricultural Systems	55
Exploring a case: the Ganges Basin Development Challenge	57
Exploring a case: the African Food Security Initiative (AFSI).....	61
A general evaluation framework	64
Clarifying evaluation purposes.....	64
Identifying program characteristics	64
Elaborating an initial theory of change.....	65
Reviewing data availability and quality	65
Summary recommendations for improving NRMR evaluation	66
Appendix 1. Evaluation questions for NRMR programs	67
Appendix 2. Example of a mixed methods statistical design evaluation	69
References	71
Figures	
1 The natural resource management research program causal package	20
2 Displaying a theory of change.....	26
3 A theory of change for the fish farming intervention.....	27
4 Theory of change for informing international and national level policy processes.....	28
5 An indicative theory of change for natural resource management research programs.....	33
6 Considering impact evaluation designs.....	36
Tables	
1 Generic natural resource management research program activities, target groups and results.....	30
2 Design, variants and causal inference	41
3 Likelihood of 'transfer' success: high, medium and low.....	48
4 Summary of tools, methods and design implications for impact evaluation questions.....	49
5 Natural resource management research (NRMR) program attributes and designs.....	52
6 Evaluation designs tied to questions and attributes: possible application to the case of CGIAR Research Program 1.3: Aquatic Agricultural Systems	58
Boxes	
1 Possible subjects for evaluation in natural resource management research programs	17
2 A framework for evaluation questions.....	38
3 Strengthening the Ganges Basin Development Challenge monitoring and evaluation.....	61
4 Strengthening the African Food Security Initiative monitoring and evaluation.....	63
A1 A framework for evaluation questions.....	67

Abbreviations

ACIAR	Australian Centre for International Agricultural Research	M&E	monitoring and evaluation
AAS	Aquatic Agricultural Systems Program (CGIAR)	QCA	qualitative comparative analysis
CSIRO	Commonwealth Scientific and Industrial Research Organisation, Australia	NRM	natural resource management
		NRMR	natural resources management research

Acknowledgments

This paper was written with support from the Australian Centre for International Agricultural Research (ACIAR) and the CGIAR Research Program on Aquatic Agricultural Systems. The NRM Impact Community of Practice acted as an advisory committee to the authors. Its membership includes leaders, evaluators and NRM researchers from five CGIAR Research Programs, the CGIAR Institutional Learning and Change Initiative, ACIAR, the Centre for International Economics and CSIRO. The significant contributions of Boru Douthwaite and Charles Crissman of WorldFish Center are also gratefully acknowledged.

Executive summary

Natural resource management research (NRMR) has a key role in improving food security and reducing poverty and malnutrition in environmentally sustainable ways, especially in rural communities in the developing world. Demonstrating this through impact evaluation poses distinct challenges. This report sets out ways in which these challenges can be met.

NRMR combines technological innovation with real-world changes in agricultural practice that involve many stakeholders at farm, community, scientific and policymaking levels. These programs generally seek to integrate multiple inputs or interventions—scientific, institutional, human and environmental; engage participatively with beneficiaries and other implicated parties; and mobilise stakeholders, both to support innovative programs and to carry lessons learned into the future.

Simple attribution of productivity and socioeconomic outcomes to NRMR interventions is difficult when NRMR itself is a ‘package’ of different actions adapted to diverse settings by farmers and other stakeholders, often over extended periods.

This report outlines impact evaluation strategies that accept that NRMR is likely to be a ‘contributory cause’ rather than the sole cause of program results. It builds on recent reports that demonstrate that, in many development settings, impact evaluation should be seen as contributing to an adaptive learning process that supports the successful implementation of innovative programs. Change is nearly always the result of a ‘causal package’ and for an NRMR intervention to make a contribution it must be a necessary part of the package. This contrasts with an ‘impact assessment’ perspective that is mainly concerned with forms of accountability that measure and attribute impacts to particular programs or interventions. Starting from a

learning perspective, impact evaluation still addresses accountability by demonstrating that NRMR programs make a difference by contributing to outcomes and impacts, and improve performance through continuous learning.

The proposed evaluation strategy pays special attention to the causal links between NRMR programs and intended outcomes. As these programs are expected to produce generalised answers that can be replicated and scaled up to tackle global problems, evaluation also has to be able to explain why and under what circumstances programs are effective. This is why the proposed evaluation strategy includes approaches to explanation, and why theories of change are an essential part of the proposed approach. A theory of change both helps to unpick the assumptions about how programs bring about change and takes into account the way programs are implemented. Such a theory-based approach also allows programs to be tested against what is known from wider research literatures and, at the same time, allows evaluation results to contribute to these literatures.

Against this background, an overarching evaluation framework is put forward that aims to answer impact evaluation questions by selecting appropriate evaluation designs that take into account NRMR program ‘attributes’ or characteristics.

The report argues that, in a complex program setting, an evaluation must begin with appropriate evaluation questions that interest policymakers, donors and other stakeholders. Key evaluation questions should be about what difference the program is making (i.e. the contribution being made), about understanding the progress being made and why results are occurring, and about the learning that is taking place. This is distinguishable from the kinds of evaluation questions

that are appropriate for more straightforward interventions such as: ‘Did our program cause the intended change?’ The evaluation questions to be considered are broader than those dealing solely with causality, and include questions of rationale and implementation, and of measuring results, in terms of both their sustainability and transferability.

The report suggests a framework for defining evaluation questions that takes account of both the outcomes and processes of change, and tries to explain how change occurs in different settings and can be generalised or scaled up.

A broad range of different evaluation designs and methods is considered, including theory-based, case-based and participatory approaches. However, although not specifically discussed in this report, more traditional approaches such as experimental and statistical methods are not dismissed—they will often be valuable as part of an overall ‘nested’ evaluation strategy.

The attributes of NRMR programs also pose evaluation challenges and have consequences for impact evaluation

design. These challenges and consequences are reviewed. For example, multi-stakeholder programs require methods capable of assessing collective action, and time-extended programs require iterative and longitudinal methods.

The approaches laid out in the report have been ‘walked through’ and refined in relation to several specific programs including: the CGIAR Research Program on Aquatic Agricultural Systems, the CGIAR Challenge Program on Water and Food’s Ganges Basin Development Challenge, and the CSIRO–AusAID African Food Security Initiative.

The report proposes a ‘general evaluation framework’ that would allow the evaluation design principles outlined to be turned into an overall operational plan, and suggests what activities are necessary to put together such a plan.

It concludes with summary recommendations, appendixes giving sample evaluation questions and an example of a mixed methods statistical design evaluation, and details of literature cited.

Introduction

This report considers the evaluation of natural resource management research (NRMR) programs and similar interventions. While the report focuses on CGIAR, the findings are more broadly applicable to other research and development (R&D) agencies such as the Australian Centre for International Agricultural Research (ACIAR).

NRMR has been defined (Kelley and Gregersen 2005) as:

[NRM research encompasses] research on land, water and biodiversity resources management that is focused on producing knowledge that results in technology options, information and methods or processes that enhance the productivity and stability of ecosystem services.

While the above definition does not include reference to research on policy or institutions, the reality is that NRMR programs clearly intertwine natural resources and human endeavours and are multifaceted, complex, dynamic and uncertain. Key delimiting characteristics would seem to be that they involve¹:

- *natural resource management* aimed at strengthening natural resource management in an area or region
- *a mix of intervention strategies* employing a range of often participatory intervention activities including the production of knowledge through research, and a variety of efforts to use this knowledge to change behaviour of individuals, households, institutions, markets and policies
- *a number of stakeholders and partners* engaged in collaboration needed to bring about change

- *multi-level, multi-place activities* operating at farm, landscape, regional and global scales.

In other words, NRMR programs implement innovative agricultural research through partnerships with beneficiaries and stakeholders that combine in practical ways the creation of scientific knowledge and community experience so as to influence farm practice, institutional arrangements and policy priorities, with the aim of strengthening and sustaining the management of natural resources.

However, the ultimate aim is not to improve natural resource management per se, but to bring about or contribute to broader impacts. For example, NRMR in CGIAR should contribute to the CGIAR system-level outcomes of:

- increasing food security
- reducing rural poverty
- reducing under-nutrition
- sustainable management of natural resources.

These are very much in line with the mission of ACIAR to achieve more productive and sustainable agricultural systems, for the benefit of developing countries and Australia, through international agriculture research partnerships.

While this paper focuses on NRMR and the attributes that are specific to it (related to social–ecological systems dynamics), what we learn from the complexity of such programs is applicable to many other R&D programs that share some complex attributes that pose challenges to evaluation.

¹ Additional attributes of NRMR programs are discussed later, when relating program attributes to designs.

The context for evaluating NRMR programs

The context in which NRMR programs work is often quite complicated. They are trying to enhance the management of water, land and biological resources by improving the various human interactions with those resources. Natural resource processes are dynamic, not completely understood, and change in them often takes a long time and may not be linear. Woolcock (2009) discusses the importance in evaluation of taking into account the shape of impact trajectories. Human processes are usually focused on a shorter term, even while the effects of human activities on natural systems are not always evident until well into the future; nor are they always predictable. Nevertheless, the interventions involved in NRMR programs are trying to modify these natural systems to make them sustainable and to benefit those who live and work within them.

Clearly then, learning how well and under what conditions these interventions are or are not working towards these aims is essential for improving the effectiveness of these interventions. Thus, a key aim of this report is to position NRMR impact evaluation as a learning process undertaken to improve the delivery of NRMR programs and the effects they achieve, by developing a framework for thinking about and designing useful and practical evaluations. As will be discussed, the framework is also useful for helping plan such programs and interventions.

Given the complex context and history of assessing NRMR programs, the following questions are critical to their evaluation:

- What can realistically be said about the causal links between NRMR program activities and the related outcomes and longer term impacts?
- How does a learning perspective fit with a focus on accountability and the results-based management ethos of major stakeholders?
- From a learning perspective, what are the key questions that the evaluation should address?
- Given that NRMR programs are complex with numerous components and possible perspectives, what should be evaluated?

- How to best assess replication and scaling up of results?
- In this light, what types of evaluation tools, methods and designs are most appropriate?

A starting point for our discussion is the report Stern et al. (2012), which considers a range of designs and methods for evaluating the impacts of development interventions. Much of the discussion in that report can be applied to the evaluation of NRMR programs. Here we apply its thinking and ideas to the world of NRMR, but go beyond it in our focus on evaluation for learning.

An overview of the report

This introductory section concludes with a discussion of the key terminology used in the report.

The second section discusses the general kinds of issues faced when evaluating NRMR programs.

An important issue covered in this report is that of causality. The subsequent sections discuss concepts of causality and theories of change, the latter being the key tool needed when non-experimental approaches are being used.² This is followed by a discussion of the fundamental features of theories of change for NRMR programs.

The main framework is then introduced and discussed: whereby evaluations designs depend on the evaluation issues addressed and the attributes of the program being evaluated. This is followed by a discussion of how evaluation tools, methods and designs relate to the evaluation questions posed and to the attributes of the program being evaluated.

The report then looks at several specific examples: the Aquatic Agricultural Systems program, the Ganges Basin Development Challenge and the African Food Security Initiative.

The report concludes with a discussion of a general evaluation framework for NRMR programs and makes a number of recommendations and suggestions for improving the evaluation of NRMR programs.

² Theories of change are also needed to explain why results have occurred in all impact assessments, including those when experimental designs are used.

Terminology

The evaluation domain is plagued by terminological inconsistency. The definitions of the concepts used appear to change constantly, reflecting the evolving trends in thinking and practice. Glossaries exist, such as that in OECD–DAC (2002), but are not consistently used. In this report key terms are defined as follows:

Program—a generic term used to denote a *broad* set of activities and strategies being undertaken to achieve an end. Thus the term ‘program’ can refer to broader programs such as the Ganges Basin Development Challenge, or still broader groupings such as the CGIAR Research Program on Aquatic Agricultural Systems. A capital ‘P’ program will refer to a specific NRM program.

Interventions—another generic term used to refer to a specific set of activities being undertaken to achieve a goal, such as in a project or in particular activities within a project. A larger program would consist of a number of interventions. When talking about general evaluation issues, concepts or ideas, the term ‘intervention’ is also used as a generic term to describe all types of deliberate actions taken to achieve a goal.

Research or change partners—those persons, groups or institutions with whom the NRM intervention is working to bring about change and is aiming to influence.

Results—this term is used here to denote any of the sequence of effects emanating from an intervention. Outputs, outcomes and impacts are specific types of results.

- **Outputs**—used here to denote the first-level results from an intervention, the information, goods or services delivered by the intervention that research partners are provided with or expected to respond to. Outputs are under the sphere of control of the intervention.

- **Outcomes**—the effects and changes that occur outside the intervention, often labelled as immediate and intermediate; the effects and the consequences of the actions taken by the research partners due to responding to the outputs, frequently focused on behaviour changes manifest as changes in practice, institutions, policy and capacity. Outcomes are largely expected to be in the sphere of influence of the intervention.
- **Impacts**—the positive and negative, primary and secondary, intended and unintended, long-term effects on beneficiaries that result from a development intervention, and are in the sphere of interest of the intervention. In NRM, we are interested in impacts on rural poverty, food security, nutrition and health, and sustainable management of natural resources.

Target groups—those persons, groups or institutions that the intervention is aiming to influence or benefit.

*Theory of change*³—a model showing how the intervention is expected to work, by laying out and explaining the various steps in the causal chain from the intervention activities to outputs, outcomes and impacts. A sound theory of change recognises the role of the key stakeholders along the causal chain and the assumptions about what has to happen for the intervention to work. In addition, the perceived risks to the intervention working, and other external factors influencing the expected outcomes and impacts, are made explicit. Further, a fully developed theory of change should consider unintended effects and rival explanations.

These terms are broadly consistent with those in OECD–DAC (2002) and draw on Stern et al. (2012) and Kivima et al. (2007).

³ The terms ‘impact pathway’ and ‘outcome pathway’ are often used more or less synonymously with theory of change, although theories of change are better thought of as building on impact/outcome pathways by making more explicit, and challenging, the underlying assumptions and risks.

Evaluating NRM programs

This report sets out approaches and directions that can be used in a learning-orientated evaluation of NRM programs. The paper is expected to act as a background and source document for a number of initiatives being undertaken on evaluation of NRM programs by the NRM Impact Community of Practice.⁴

In CGIAR and ACIAR, impact assessments are undertaken primarily from an accountability perspective. Accountability driven evaluations focus on the results achieved and associated causal processes; assessing whether programs ‘produced’ impacts and their magnitude. CGIAR has been evaluating NRM projects and programs for some time (Shiferaw and Freeman 2003; Fujisaka and White 2004; Maredia 2009; de Janvry et al. 2011; Independent Science and Partnership Council 2011). Kelley et al. (2008) discuss CGIAR experience with impact assessment. ACIAR has also assessed a number of its NRM projects (e.g. Harris 2004, 2006; Corbishley and Pearce 2007; Saunders et al. 2012). In the cases of CGIAR and ACIAR, the approaches used have largely been consistent with CGIAR (<http://impact.cgiar.org/methodology>) or ACIAR (Davis et al. 2008) impact assessment guidelines.

Many impact assessments focus on valuation—estimating net economic benefits from the project or program—and often are aimed at providing evidence for CGIAR investors that funds have been well spent. They have often, although not always, focused on a specific causal evaluation issue: To what extent can a specific (net) impact be attributed to the intervention?

⁴ The NRM Impact Community of Practice explores the ways in which NRM can better leverage developmental outcomes and impact. Its members include leaders, evaluators and NRM researchers from five CGIAR Research Programs, the CGIAR Institutional Learning and Change Initiative, ACIAR, the Centre for International Economics and CSIRO.

To learn, we need to be able to understand how and why impacts occur. A learning-orientated evaluation therefore also focuses on explaining why these results occur. It is through explanation that we can hope to replicate, scale up and improve future programs

In NRM, the complexity of programs gives particular emphasis to implementation processes as part of any explanation. These programs generally seek to integrate multiple inputs or interventions—scientific, institutional, human and environmental; engage participatively with beneficiaries and other implicated parties; and mobilise many stakeholders to support innovative programs and to carry forward lessons learned into the future.

The strategies used to integrate inputs, engage with beneficiaries and mobilise stakeholders—the *implementation theory*—are themselves part of the ‘explanation’ of what is or is not achieved. This is distinct from the many programs in which implementation is more straightforward and is less likely to be locally customised. The importance of embedding broader principles into local contexts is the driver behind customisation and is hence central to implementation in NRM.

The other side of explanation is *program theory*; understanding just how it is that the activities undertaken are expected to lead to the desired outcomes and impacts. For this insight, we argue the need for a different perspective on causality from the usual counterfactual interpretation of what would have happened without the intervention; one that recognises that the program is but one of several causal factors that bring about impacts. Causality is discussed in the next section.

What then are the big issues of concern in evaluating NRMR programs? An initial articulation might be as follows:

1. Can we, and how can we, causally link NRMR program activities to outcomes and the broad development goals, such as the CGIAR system-level outcomes or the Australian Government's ongoing food security policy priorities?
2. To what extent have expected NRMR outputs, outcomes and impacts been realised?
3. Through what means and mechanisms have the NRMR outputs, outcomes and impacts been achieved and development goals influenced?
4. Has the NRMR program been designed and implemented in a way that enhances or makes intended results more likely?
5. What ways have been found by programs that successfully combine different strategies and outcomes: genetic, farm systems, environmental, institutional, livelihoods etc.?
6. What would be done differently and what would be repeated?
7. What can be generalised or transferred from an NRMR program, given the often site-specific nature of these interventions?

These issues are further discussed below.

A focus on learning and refocusing accountability

Many evaluations of NRMR programs and their projects will continue to be carried out using the impact assessment perspective. This report is intended to develop a complementary approach to evaluating NRMR programs; one more focused on helping NRMR managers and stakeholders to learn more about their interventions and to understand why and how outcomes and impacts have or have not been realised. This emphasis on learning follows from the view of NRMR as operating in a dynamic, emergent, complex and often unpredictable human and ecological system. In this setting, adaptive management informed by careful responses to new information and understandings is essential for building and managing more effective

programs and interventions. There can nonetheless be potential synergies between these two approaches.

To distinguish them from 'conventional' impact assessment, learning-focused approaches to linking research activities to impact will be referred to here as *impact evaluations*. Because of the strong learning focus, impact evaluations provide insights on how to better plan, implement and manage research programs so that they can more closely adapt to conditions and context, and deliver their intended outcomes and impacts.

There are several aspects to learning in the NRMR setting. There is an interest in using evaluation insights to learn how to better implement and steer current and future NRMR programs. 'Better' implementation in the context of an impact evaluation is concerned with implementation that improves the prospects for the successful achievement of outcomes and impact.

Research knowledge generated by NRMR programs is expected to be used by others and to lead to positive actions. Interest here is on which type of research knowledge and its use best facilitate getting to results—about what works, and what programs can positively contribute to.

A learning focus in an evaluation has several implications. One is that the evaluation questions posed need to get at understanding why and how a program is or is not contributing outcomes and impacts. This includes attention to implementation, and a structured way to assess implementation progress and challenges. Another is that the more dynamic and uncertain the context, the more the need for adaptive management—frequent feedback on progress based on monitoring and real-time evaluation information. There is a need for ongoing feedback, reflection and learning, and consequent revision of activities being undertaken. A structured annual reflection and revision process is good practice.

This learning focus does not mean that accountability is ignored. Findings from these impact evaluations will provide considerable information for reporting on a wide variety of program accomplishments, and full advantage should be taken of this. They will also demonstrate to donors that the NRMR is learning and improving, thereby making more effective use of their funds. Assessing the value of this learning will also demonstrate the usefulness of such learning-focused evaluation.

Indeed, it can be argued (Auditor General of Canada 2002; Mayne 2007; Perrin 2007) that, given the focus of projects and programs on outcomes and impacts over which an intervention has influence but not control, there is a need to revise ideas about what accountability in such systems should be concerned with. Interpreting accountability as demonstrating the extent to which results can be attributed uniquely to an NRMR-type program is not realistic.

Mayne (2007, p. 68) has argued that meeting accountability requirements for outcomes/impacts entails:

1. providing information on the extent to which the expected and other outcomes [and impacts] were attained, and at what cost
2. demonstrating the contribution made by the activities and outputs of the program to the outcomes
3. demonstrating the learning and change that have resulted
4. providing assurance that the means used were sound.

The key point here is focusing accountability on demonstrating contribution and showing that learning is occurring, rather than just on attributing outcomes/impacts. This refocusing of the concerns of accountability is much more consistent with the context and complex realities of most NRMR interventions and programs, and directly supportive of the need for a focus on learning. It argues that donors need to recognise the challenges of trying to influence longer term, high-level results when there are numerous partners and other factors at play, and hence that it is essential that funds be used to learn about how best to influence the realisation of these results.

The timing for impact evaluation

It is often suggested that a distinctive feature of NRMR is its extended timescale. Despite a common definition of ‘impact’ as taking place over the long term, many impact assessments and evaluations take place at a point in time quite early in a program cycle. On the other hand, very few evaluations take a true long-term

perspective, looking back from today’s observable results to decisions and policies made 15 or 20 years ago. And, indeed, the passage of time makes such a perspective quite challenging.

The challenge of when to evaluate depends partly on what is being evaluated, insofar as our understanding of time structures the entire ‘object’ that is to be evaluated. For example, a more radical understanding of the time dimension is that the policy-framed logic of programs and their evaluation is flawed or difficult to apply to NRMR programs. Stakeholders in this field are often dealing with dynamic processes that not only take far longer than policymakers might wish but also are dealing with processes over which we have relatively little firm knowledge and control. Programs and their interventions are only a small part of these fundamental processes of change in human and natural systems. This way of thinking is especially prevalent in natural resource management (NRM) and related environmental fields such as sustainability, climate change and resilience.

However, similar ideas can be found in many other policy domains, such as economic development, social inclusion, strengthened governance etc.—wherever it is recognised that any short-term program is embedded in long-term social and cultural processes. So, NRM is not unique in this regard. One temptation when taking on board this more radical view of time in NRM and NRMR is to challenge the possibility of evaluating ultimate impacts. The best that can be expected is to focus on outcomes and impact-related processes—how we implement and steer, what barriers are encountered and how adjustments to strategies are made and justified.

An intermediate position might be to:

- recognise that the timing of an impact evaluation will be an important design consideration and that each case needs to be judged on its merit—some will be prospective, some early in a program cycle and some a few or many years later, depending on overall purposes.
- take seriously the dynamic long-term and process character of NRMR, and that any impact evaluation will inevitably be partial, focusing on a time slice rather than an entire set of processes. It should still be possible to ensure that the selected time slice is a coherent and appropriate unit of analysis.

- take account of the long-term and to some degree unpredictable nature of these programs by adopting an iterative/staged design that can be revised as new problems or understandings emerge. This will also make easier, evaluative inputs to any mid-course corrections that might be needed.
- at the time of an impact evaluation, based on current data and understanding, thoughtfully and critically reassess the theory of change of the program as to the likelihood of future impacts. Are the current strategy and activities still most likely to lead to the desired impacts? Or is a more fundamental rethink needed?

However, it may not be useful to think of impact evaluation as an action at one point in time. On the one hand, monitoring should be seen as an essential element of the evaluation package, providing data for evaluation and for adjustments to implementation. Further, the more complex the setting, the more useful it will be to look to more real-time evaluation approaches that regularly gather and analyse data, perhaps through special studies or as part of monitoring. In this perspective, evaluation is an ongoing process that can still include an impact evaluation or impact assessments at appropriate points in time. The monitoring and evaluation plans for the interventions in the African Food Strategy Initiative discussed later have several built-in learning mechanisms, including an annual review process to reassess the partnerships involved.

The unit of analysis: what to evaluate?

There remains an outstanding question: How to break down big NRM programs for evaluation purposes? This would probably involve selecting a subset of program activities that fall within a particular period of time, but then what? Key is probably a methodological foundation for the way the unit of analysis is selected. Box 1 lists some of the possibilities.

Projects or groups of projects are often taken as the unit of analysis. This may be appropriate for some

evaluations, depending on the questions to be asked. But there are other useful ways of cutting up a big program that should also be considered, such as looking across the set of projects at:

- the impacts on spatial areas or population target groups or research partners
- the specific results attained from different types of intervention strategies
- how different intermediate outcomes were brought about.

In the discussions later, on NRM evaluation designs generally and the specific cases examined, examples of these different possible perspectives on what to evaluate are discussed. The idea is that building theories of change from one or more of these perspectives would allow for insight to be gained on how and why interventions are working or not.

In all cases, the intent is to lead to some kind of causal inference, such as developing a comprehensive theory of change for influencing a target group. The findings would be about what works; for example, for which target groups or research partners in which settings in terms of the getting to outcomes and impacts, perhaps exploring the mechanisms at work. However approached, any breakdown of a program should begin from a theory-based and methodological justification rather than analytic distinctions or typologies alone, with the aim of understanding how and why results are occurring.

Box 1. Possible subjects for evaluation in natural resource management research programs

- Projects or groups of projects
- Spatial areas (hubs)
- Types of intervention strategies
- Target groups
- Intermediate outcomes

Causality and contributory causes

The multiple factors and stakeholders involved in NRM highlight the complexity of the causal relationships between the varied activities of an NRMR program and the results sought. To gain insight on these relationships, the need to have some understanding of the cause–effect relationships between the activities of the NRMR program and subsequent outputs, outcomes and impacts is crucial. Without such an understanding, it is difficult to know whether activities are delivering the intended result and what actions to take when things are not going well, or to make credible claims about what is being accomplished by the program. But in the context of an NRMR program, there are many factors at work, and cause–effect linkage is not straightforward. Indeed, we argue the need to develop a different perspective on causality to better tackle causal questions in interventions like NRMR programs, different from the usual counterfactual interpretation.

The challenge

All development interventions endeavour to make a difference, and to demonstrate that they are doing so. They undertake activities and produce outputs that are expected to lead, through a sequence of effects, to specific improvements in the wellbeing of beneficiaries. However, making the causal link between the activities and outputs, and subsequent outcomes and impacts, can be challenging, for the following reasons:

- The causal path between the activities/outputs and the impacts can be quite distant and complicated, involving a long causal sequence of immediate and intermediate results and often a long timescale.

- Events and conditions outside those of the intervention can influence the extent to which the impacts are brought about.
- There may be a number of causes, including other concurrent interventions, contributing to the realisation of the impacts, in addition to the influence of the intervention. The intervention is not working alone.
- The direction and strength of causality can change as time passes. Thus, factors that have a positive influence at a certain time can become a hindrance later on.⁵
- In complex processes, similar results can be obtained with different interventions and, conversely, similar interventions in slightly different sites or periods can have very different outcomes.

NRMR programs have many of the above characteristics. The causal path between the research on the natural world and the impacts sought is often distant. For example, many factors influence food security and rural poverty, and there are likely other government, donor and non-government organisation interventions at play influencing these impacts. Additional challenges can include complex biophysical changes and the fact that benefits (or costs) can occur at a considerable distance from where the outputs were adopted.

In these circumstances, what can be said about the causal relationship between the intervention and the observed results? We expect that the intervention, as a ‘cause’, has indeed had some effect and we want to

⁵ For example, application of fertiliser can improve the management of crops and pastures, but run-off can contaminate watersheds. At low levels of fertilisation, the first effect is likely to dominate, but at high levels of use the second may become dominant.

be able to make some credible causal claim about the intervention, such as:

- the intervention caused the impacts
- the intervention made a difference
- the intervention contributed to the impacts.

Are these all the same? What type of causal claim makes sense? And how can we go about making the claim?

Concepts of causality

Can we credibly say a research intervention ‘caused’ the impacts?—that, for example, research and related activities on farming methods themselves caused a reduction in rural poverty. Clearly we cannot. There will be a number of ‘causes’ explaining any observed impacts and we hope that the intervention is one, and perhaps a significant one. Saying the intervention alone caused the impacts is much too simplistic and scarcely credible.

Causality involves relationships between events or conditions, and is often discussed in terms of necessary and sufficient conditions. Is, then, the intervention a necessary cause of the impacts? Most of the time, the answer will be ‘no’. Necessity means that the impacts can be realised only if there is the specific intervention. Yet, most desired impacts such as better health or education, reduced poverty, improved environmental conditions, and greater food security can potentially be realised through a variety of different types of interventions, and not only the specific intervention of interest. It would be presumptuous to say that your intervention is the only possible way to bring about the desired impacts.

Might we instead be able to say that the intervention is sufficient? Again, the answer is a clear no. We are assuming that there are other factors, perhaps many, also at work. So, on its own the intervention is not sufficient. In the NRMR case, it is widely recognised that, for the desired impacts to be achieved, many other events and conditions will have to be in place in addition to the research.

On the other hand, we do expect that the intervention *along with the other influencing factors* is indeed sufficient; that, collectively, this set of actions and

conditions including the intervention are expected to bring about the impacts. And, indeed, when we say X causes Y in everyday discussions, sufficiency is probably what we usually mean: that X did indeed produce or lead to Y.

There is, of course, a large and longstanding literature on causality, going back centuries, and these issues and concerns have been repeatedly explored. Causes that are neither necessary nor sufficient are called *contributory causes*. Thus, for example, smoking heavily is a contributory cause of lung cancer. It is not a necessary cause, since there are other sources of lung cancer, and it is not a sufficient cause, since not all smokers suffer from lung cancer. In the philosophy literature, a contributory cause is called an INUS cause: an Insufficient but Necessary part of a condition that is itself Unnecessary but Sufficient for the occurrence of the effect (Mackie 1974).

Stern et al. (2012) reported that thinking about and conceptualising development interventions as contributory causes was especially helpful. Many interventions are in fact designed to be part of a package of other causal events and conditions⁶ and, even when not so designed, need to take these other influencing factors into account. Cartwright and Hardie (2012) call these *supporting (causal) factors*; other events and conditions that need to happen in order for the intervention to work—that is, to make a difference.

From this perspective, the key causal question becomes:

Was the causal package of the intervention plus its supporting factors sufficient to produce the desired result?

Here it is recognised that there could be other ways that the desired result is brought about, and hence that the particular causal package in question is not necessary to achieve the result; that is, it is not the only way that the desired results could be achieved.

But we would also want to know if the intervention was an essential part of the specific causal package. Perhaps the desired result could be realised with just the supporting factors without the intervention. The causal package with these two characteristics—sufficiency of

⁶ As discussed below, this is very much the case with NRMR programs that are designed to take into account the actions needed from a number of change partners.

the package and necessity of the intervention as part of the package—is the *intervention causal package*.

If these conditions hold, then the intervention is a *contributory cause* and hence has ‘made a difference’. That is:

- the intervention causal package was sufficient to produce the observed result
- the intervention was a necessary element of the causal package.

This is a useful operational definition of ‘making a difference’.

NRMR programs as causal packages

An NRMR program comprises a number of diverse activities carried out by the program personnel. Many of these activities involve efforts aimed at influencing the actions of the several groups and stakeholders associated with the program—the research or change partners. For the program to work—that is, to bring about the desired results for the beneficiaries—appropriate actions by these other partners/stakeholders are seen as essential. Those actions are all supporting (necessary) causal factors, as are actions by the beneficiaries. That is, NRMR programs

require interventions that target central issues of capacity development, policy influence and institutional change. These supporting causal factors, plus any other necessary conditions, comprise the *NRMR program causal package*, and the expectation is that the NRMR program is a contributory cause; that is, it is an essential component of this (hopefully) sufficient causal package (Figure 1).

As with all agricultural research, identifying and understanding these supporting factors is clearly important in designing and assessing NRMR-type programs. These changes in the thinking and practice of the research partners are some of the necessary events and conditions that NRMR interventions rightly focus on to try to bring about larger changes. Getting results, then scaling up and out, requires that a range of events and conditions have to be aligned, well beyond the research and knowledge that are being produced by the intervention. A well-designed NRMR intervention will work actively to identify and to positively influence as many of these supporting causal factors as possible.

There is an obvious need to be practical in identifying supporting factors. Conceptually, there are many events or conditions that are needed for an intervention to work, including that there are no revolutions and that the sun rises every day. A priori, we need to identify reasonably plausible and *relevant* supporting factors. Ex post, in trying to explain what has happened, the

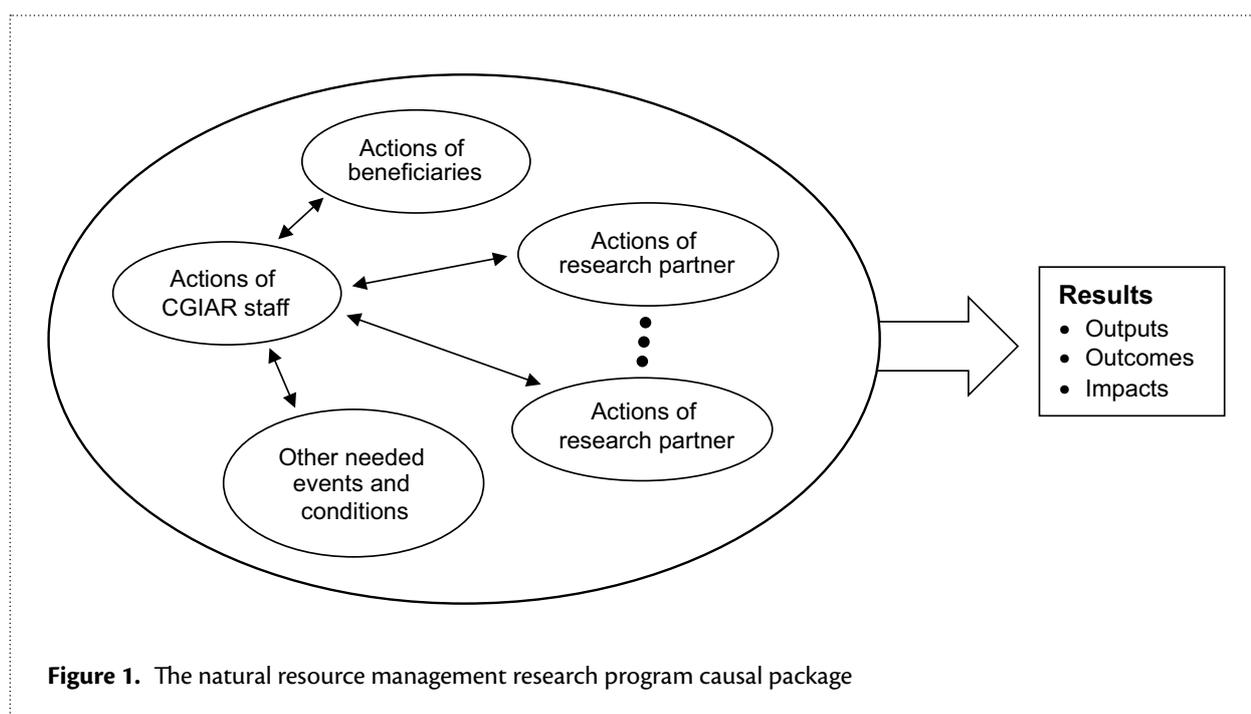


Figure 1. The natural resource management research program causal package

task is easier as the other factors that led to change are more easily identifiable. We know the sun rose every day and there was no revolution, and can find out what else happened that we need to take into account.

As a contributory cause, an NRMR program is one among several ‘causes’. Yet our interest is on the program as an instrument of change; a collection of activities deliberately undertaken to get or continue change happening where adequate change was not happening before. We would like to know what *role* the intervention played in bringing about impacts.⁷ We may expect that, at a minimum, the program acts as a *trigger* to start the causal chain. And, in the more complex contexts we are considering, the program may also involve other actions subsequently taken along the causal pathway to *sustain* the pathway. Thus, we would like to assess whether programs were triggering and/or sustaining contributory causes. In such cases, a program can be said to be a *principal contributory cause*. In other cases, the program might see itself as playing a more modest supporting role, joining others in an already ongoing process and *enhancing* a change process already underway so that better or more timely results are achieved.

For NRMR, we would usually expect that the intervention is a principal contributory cause of falling rural poverty and rising food security in some area; that is, that the research plus a number of other supporting factors were sufficient to realise the expected impacts and that the research played a principal role in bringing about the change.

Thus, a (strong) causal claim about an NRMR intervention would be that the intervention was a (principal) contributory cause of the relevant observed results. That is:

The intervention was a necessary component of a package of causal factors that together were sufficient to produce an observed result; in other words, the intervention made a difference. (In addition, the intervention played a key role; it was the trigger that initiated the chain of events that led to the observed results.)

In summary, the causal claim is that *the NRMR causal package was sufficient to bring about an observed result and the NRMR intervention was a necessary part of the package: the NRMR intervention was a contributory cause.*

Probabilistic causes

The discussion of contributory causes has here so far been in deterministic terms; that is, a causal package is either sufficient or it is not. However, the discussion needs to reflect the probabilistic nature of causality for many phenomena. Mahoney (2008, p. 421) argues that ‘a treatment is a cause when its presence raises the probability of an outcome occurring in any given case ...’. Following on from Mahoney, in terms of the intervention causal package, the term *likely necessary* can be used to describe the supporting causal factors—factors that almost always have to be present for the outcome to occur—and *likely sufficient* to describe the sufficiency of the intervention causal package, meaning that, in this case, the causal package most likely produced the observed result. For an intervention being evaluated, these are more realistic interpretations of necessary and sufficient conditions as discussed earlier.

With this perspective on causality, the key causal questions related to an intervention are:

1. Is it likely that the intervention has made a difference?
 - Is the intervention likely a contributory cause of the result?
 - What role did the intervention play?
2. How and why has the intervention made a difference?
 - How did the causal factors combine to bring about the result?
 - What context was relevant and which mechanisms were at work?
 - Has the intervention resulted in any unintended effects?

⁷ Cartwright and Hardie (2012, p. 95) talk about *salient* causal factors.

Fish farming as an example of a causal package

Consider a specific example of an NRM intervention. The ACIAR–World Vision intervention reported by Harris (2010) was aimed at improving the diet, food security and income of low-income farmers in northern Thailand by encouraging and helping them adopt freshwater fish farming. It involved NRM issues of improving land and water management. Selected farmers participated in the research to develop low-cost fish food from local materials. They were provided with initial funds for equipment and for fingerlings, and gave and provided advice on fish-farming methods to the researchers and their peers. In addition to improving their own lives, others were expected to adopt fish farming through example.

The initiative components were:

- participatory research on low-cost fish food
- start-up funding
- advice on improved fish-farming methods.

Other supporting factors mentioned in the report were⁸:

- an adequate number of initial farmers convinced to try fish farming
- an adequate market for the fish surplus to family consumption

- an adequate supply of affordable healthy fingerlings
- support by the farmer's family for the additional work involved
- improvement of the lives of adopters over time, and visible to neighbours
- availability of adequate cash to buy fish food and other supplies.

The intervention causal package is this set of activities and supporting factors. The intervention is clearly not sufficient on its own, and neither is it necessary: an alternative approach to achieving the intended results would be, for example, to provide set-up funds and training for the farmers' households to start cottage weaving businesses, or to provide resources and training to increase yields of existing food and cash crops.

It is likely that the initiative made a difference if (1) the expected results were achieved as a result of the initiative and its supporting factors, and (2) the initiative was an essential component of the causal package.

If the expected results have not occurred, then one would want to try to understand why. Did some of the supporting factors fail to occur? Were there other factors that now appear to have been needed but did not occur? Were the initiative activities poorly implemented, inadequate or unsound? Is the causal model sound? And if there were unexpected results—either positive or negative—we would also try to understand why they occurred.

⁸ One can imagine other possible supporting factors, such as the availability of an adequate supply of water and the absence of fish diseases.

Theory-based approaches in evaluation

Explanation is impossible without theory, insofar as we cannot observe causality: it has to be inferred and theory allows us to see the sequence of steps between cause and effect. A learning-orientated evaluation will rely heavily on various forms of theory to explain findings and support causal inferences. Theories will include the substantive theory of the program and broader bodies of theory within the same domain. It will also have to include ‘implementation theory’—an articulation of the strategies by which the program is implemented and delivered. In a participatory program that relies on action-research and includes in its goals institutional innovation, new practices, knowledge utilisation and new forms of governance, both implementation and program theory will be important.

This report puts ‘theory-based’ evaluation at the heart of an impact evaluation approach looking for explanations, in particular those approaches that make use of theories of change. Theory-based approaches also include ‘theory testing’ approaches that assess a substantive theory through some comparison of what happens and what theory predicts. Qualitative comparative analysis (QCA) explores cause–effect, building a ‘theory’ of the collection of contributory causes, but does not need a sequential, staged theory of change. Realist approaches⁹ (Pawson and Tilley 1997) build a form of theory of change, focusing on the context, mechanisms, outcomes

and configurations behind the intervention, trying to understand what it is about the intervention that makes things work. Stame (2004) and Blamey and Mackenzie (2007) compare and contrast different theory-based approaches. More recently, Coryn et al. (2011) have reviewed general theory-driven practice in evaluation during 1999–2009.

Useful theories of change

To get at explanation and address many of the evaluation questions of interest, there is a need to discuss and represent the relationships among the activities, outputs, outcomes and impacts of an NRMR program or intervention. This is what a program theory of change sets out to do. Vogel (2012) reviewed the use of theories of change in international development. Stein and Valters (2012) also reviewed theories of change in international development and note that what a theory of change entails and when and how it is used varies.

Developing a theory of change for an NRMR program is useful for several reasons:

- Developed at the outset, a theory of change helps in the design of the program intervention and in identifying indicators for monitoring. Ex ante, a theory of change can also be used to assess the likelihood that the intervention will work.
- Reviewed annually, a theory of change assists in assessing progress and in delivery adjustments—a tool for adaptive management.

⁹ Realist evaluation is a form of theory-driven evaluation. It assumes that programs are complex, adaptive systems designed to alleviate problems that have multiple, entwined roots. It is distinctive in the way that it articulates program theory, assuming that program outcomes are generated by the collective choices of program stakeholders which are in turn constrained by the social, political and cultural context in which the program is embedded.

- Developed or reviewed at the time of an evaluation, a theory of change helps develop the design and tools for the evaluation, such as surveys and interview guides, and can be used as the basis for understanding and making causal claims about the program.

Ideally, a theory of change is developed when a program is being designed, or revised on an ongoing basis as understanding and events unfold, and is used as a key element in the design of any evaluation. The theory of change should be developed using the perspectives of key stakeholders, beneficiaries, prior evaluations and the existing research on the substantive area.

In relation to both program and implementation theory, the main purpose of a theory of change is to provide a structure for the collection and subsequent analysis of data. The theory of change will:

- identify critical links (preconditions, stages, decisions etc.) in the program's planning, implementation and delivery
- identify critical conditions (assumptions/supporting factors) that will be needed for links in the theory of change to be realised, and threats to the links (risks)
- identify alternative hypotheses of change, some of which will extend to hypotheses not considered by program implementers, for each of these critical links
- assess different contexts of program implementation and how they might influence processes and mechanisms of change
- provide for the testing of hypotheses about program goals and implementation by eliminating alternative hypotheses and confirming those that inform the initial program theory.

CGIAR, in common with an increasing number of international non-government organisations and United Nations organisations, now frequently develops types of theories of change for their programs and projects. ACIAR is also increasingly developing theories of change for its projects and programs, although the term 'theory of change' is not always used and the underlying assumptions and risks have not always been made explicit.

The CGIAR Challenge Program on Water and Food M&E Guide¹⁰ discusses theories of change, and their development and use at the project, basin and program level. Various terms are used: theory of change, outcome pathway, impact pathway (Douthwaite et al. 2008) and outcome logic model. Theories of change are often developed to help design the project interventions and the kind of project monitoring that can be done, at basin and program level.

This report argues that a theory of change needs to go well beyond the typical results chain, logic model or logical framework. Mayne (2012a, p. 273) argues that a good theory of change would include:

- a results (causal) chain showing the basic logic of the intervention
- the underlying assumptions behind the links in the results chain
- the risks to each link occurring
- identification of unintended effects¹¹
- identification of possible alternative rival explanations.

Most important here is identifying the assumptions in the theory of change. These are the various supporting factors that are needed for the causal link in question to occur. These play a role similar to identifying the mechanisms behind causal links that are a focus of attention in realist approaches to evaluation (Pawson and Tilley 1997). Assumptions and mechanisms allow one to understand just how a causal link operates.

It can also be useful to determine the degree of influence the intervention has had or can have over the assumptions—the other necessary causal factors. The key point is that the intervention should be working to influence the realisation of the assumptions and to minimise the risks. This is typically achieved through partnerships and being part of a collective effort. This is what doing research in development—in context—is about. With this in mind, when developing a theory of change, it can be useful to label the assumptions as those which the intervention can directly influence, those which can be only indirectly influenced and those

¹⁰ Accessible at <<http://monitoring.cpwf.info/background/theory-of-change>>

¹¹ Ex post, these would be possible unintended effects.

which are beyond the influence of the intervention (Mayne 2011).

Figure 2, adapted from (Mayne 2012a), illustrates the various components of a theory of change. Mayne (p. 274) notes:

The theory of change is displayed deliberately as a quasi-linear process, but, as shown, allows for feedback loops as needed. A 'sort of' linear theory of change facilitates both arriving at causal claims and communicating the performance story of the intervention. The assumption boxes can be used to reduce the number of explicit links that might otherwise be needed in a theory of change. Other explanatory factors (rival explanations) may be different for different links or may apply to the overall causal logic of the intervention. The vertical 'activities and outputs' box allows for an implementation theory to be shown, i.e., the activities and outputs that are going to be delivered, perhaps over time, to implement the intervention.

What could be added in a specific case is labelling of each assumption as to the degree of influence the intervention has or seeks to have.

Figure 2 generically shows both the *program theory* causal pathway from outputs through outcomes to impacts and the *implementation theory* causal pathway showing how activities and outputs of the intervention interact with the program causal chain (Weiss 1997). Given that the focus of this report is on learning, some attention will be given to implementation theory.

A well-developed theory of change is embedded in the context of the intervention. A 'full' theory of change may be best arrived at through the iterative development of theories of change that initially do not comprise all of the features listed above. But it needs to entail more than a logic model of boxes and arrows. The trick is to build a 'good enough' theory of change, with an adequate focus on the arrows.

For communication purposes, it is useful to have two versions of each theory of change. The first is a text version, a *theory of change narrative* or *outline*, describing in a sentence or two how the intervention is intended to work. This version 'explains in a straightforward manner how the intervention is supposed to work' (Mayne 2011, p. 69). The second version is a diagram model as set out in Figure 2.

Figure 3 illustrates a theory of change narrative for the earlier fish-farming example. The assumptions are the supporting factors listed earlier.

By developing cheap fish food from local ingredients, and providing training and education to local farmers, freshwater fish farming becomes a viable livelihood option and will be adopted by farmers and their neighbours. This will both raise their incomes and increase the nutritional value of their diet.

Figure 4 is an example of a theory of change in the climate change component of CGIAR Research Program 6.

Tree-based carbon sequestration and reduced deforestation and forest degradation in rural landscapes (e.g. implementation of agroforestry, improving forest management, forest conservation etc.) offer significant opportunities for developing countries to reduce their national greenhouse gas emissions. Appropriate incentives can be developed nationally and internationally to influence land use decisions, conserve forests and promote sustainable development. This research will contribute by analysing the experience of REDD+¹² policy formulation and implementation in a series of case studies to learn how to achieve effective, efficient and equitable outcomes. The aim is to ensure that policymakers and practitioner communities have the knowledge, information, analysis and tools they need to ensure effective and cost-efficient reduction of carbon emissions and enhancement of carbon stocks with equitable impacts and co-benefits, including poverty reduction, enhancement of non-carbon ecosystem services and protection of local livelihoods, rights and tenure.

It is understood that it is a highly political process and the theory of change requires that the research results be effectively communicated to and used by appropriate audiences/forums; that appropriate policy reforms will be made; and that those policies will create incentives and reduce disincentives for increased management of ecosystems for carbon sequestration, in a way that benefits local stakeholders.

¹² Reducing emissions from deforestation and forest degradation; a United Nations program

Theories of change incorporate causal packages

We have seen that theories of change need to incorporate the various supporting factors involved in an intervention. Thus, *theories of change are models of causal sufficiency*—they set out a model of how the intervention is expected to contribute to the desired results. Theories of change not only incorporate causal packages but also set out the expected relationships between the intervention and the supporting factors (assumptions), as well as identify the risks (the confounding factors). ‘A theory of change is a model of how the intervention is expected to act as a contributing cause’ (Mayne 2012a, p. 7).

To reiterate, impact pathways and outcome pathways can be considered theories of change or, if not fully developed, can serve as a basis for a theory of change. However, in

this report, the term ‘theory of change’ is used to describe models of causality that explicitly include assumptions, risks, unintended effects and rival explanations.

Theories of change can evolve and can be at different levels

Theories of change are statements about how an intervention is expected to contribute to desired outcomes and impacts, based on current knowledge and understanding. As a result, they should evolve over time as evidence and understanding are gained. At the outset they represent the expectations based on previous experience of similar interventions elsewhere. As monitoring and other evaluative data and information are acquired, it is useful to regularly revise the theory of change.

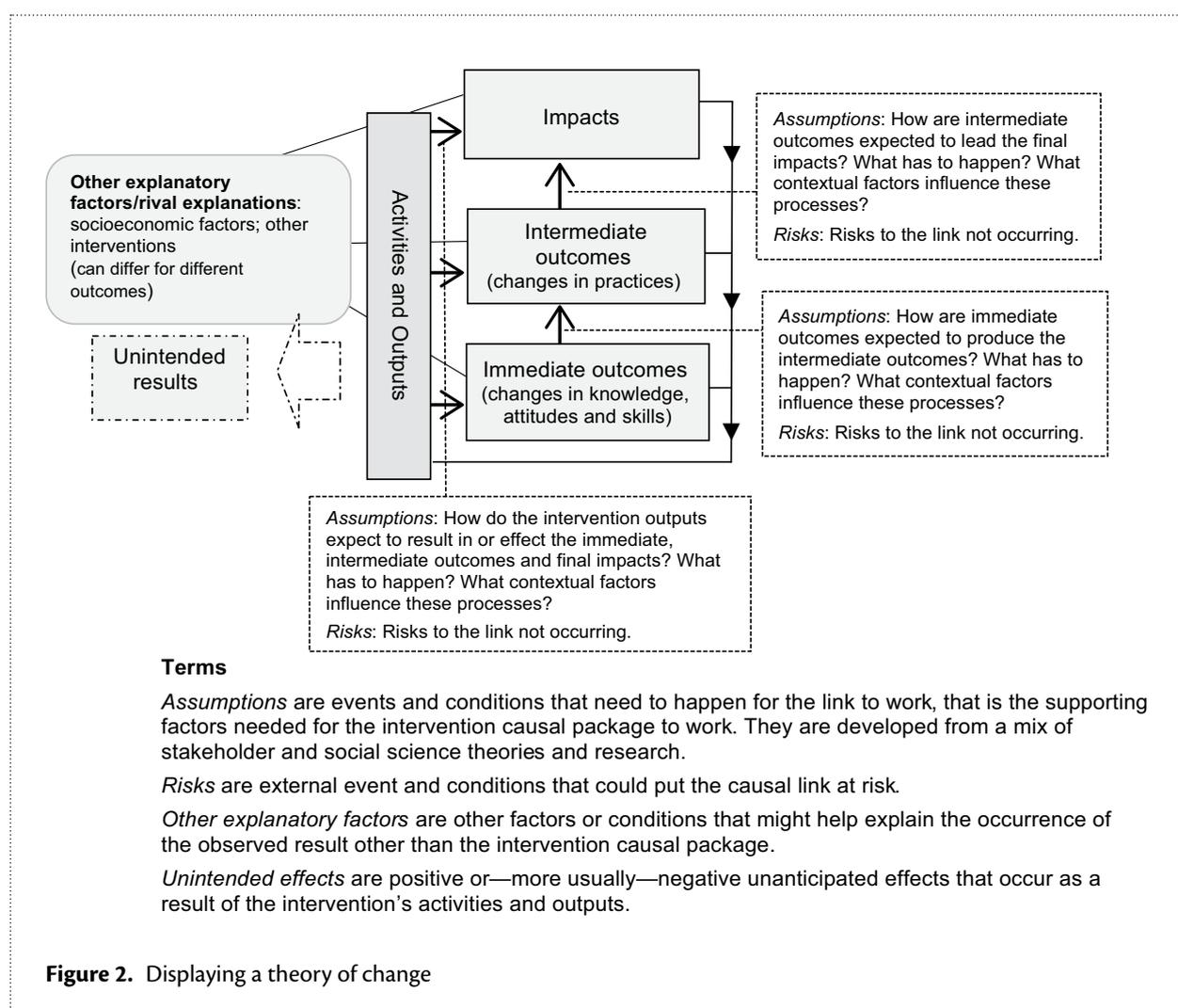


Figure 2. Displaying a theory of change

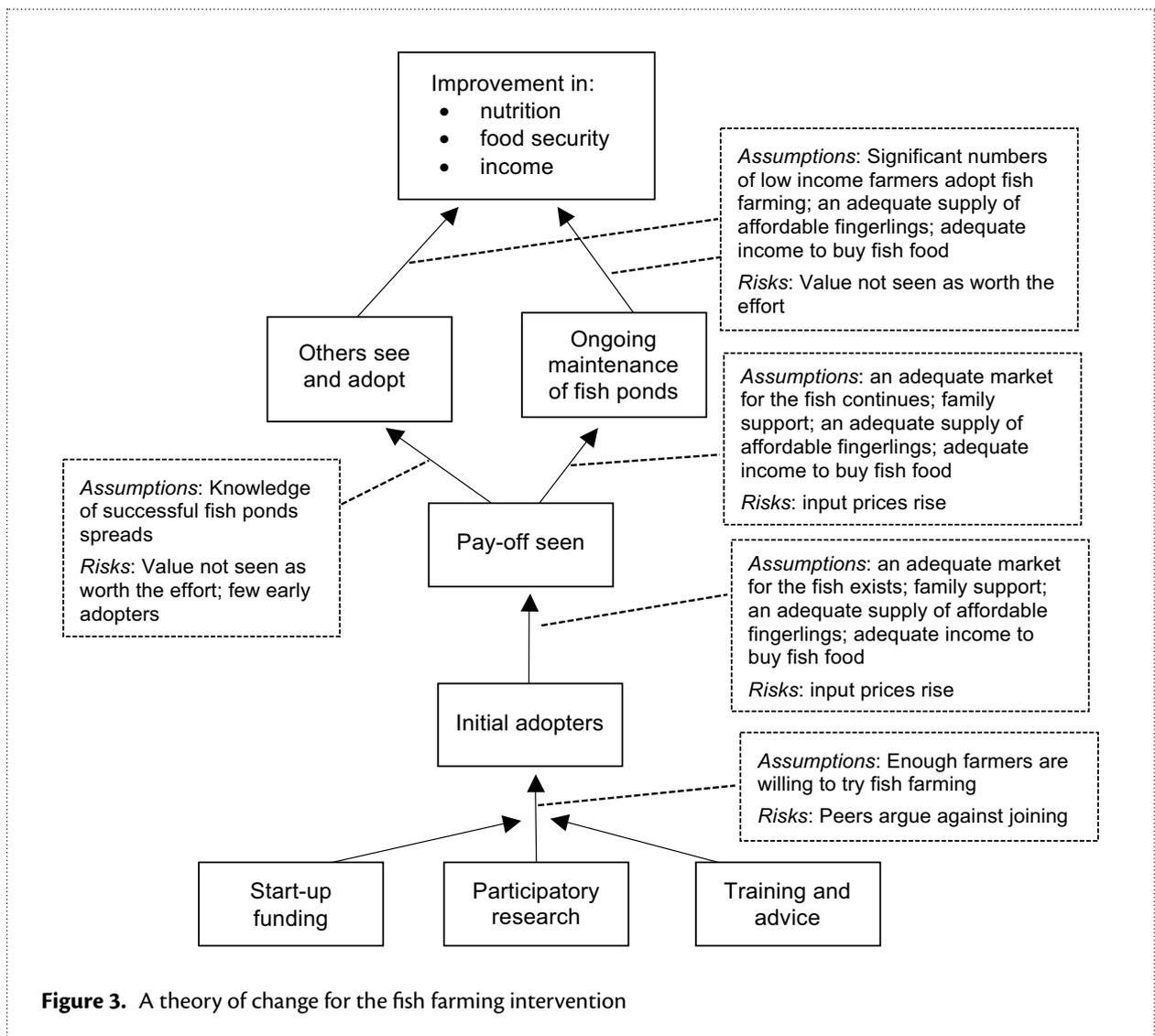


Figure 3. A theory of change for the fish farming intervention

Further, theories of change can be developed at different levels and from different perspectives, such as at the project, locality and program level. All serve a purpose. Theories of change could also be developed from different perspectives, such as for the different types of interventions used in a program: research, engagement or capacity building of the different target groups whose behaviour the program is trying to change, such as farmers, households, decision-makers etc. Thus it can be expected that in the evaluation of many NRMR programs there will be a need for a set of theories of change; for *nested theories of change*. Stein and Valters (2012) discuss the development and use of different levels of theories of change in evaluating development interventions. In the next section of this report, nested theories of change are discussed for NRMR programs and later for the Aquatic Agricultural Systems Program in particular.

Using theories of change ex ante and ex post

Theories of change developed before an NRMR intervention is implemented are very useful tools for assisting in the design of the intervention. In an ex-ante mode, they can be used to:

- identify the supporting factors that are needed for the intervention to work; that is, identify the various partners that need to be engaged with and how, and other factors that may need to be monitored during implementation
- assess the likelihood that the intervention will work—approaches to do this based on logical analysis, look-alike interventions and synthesis of

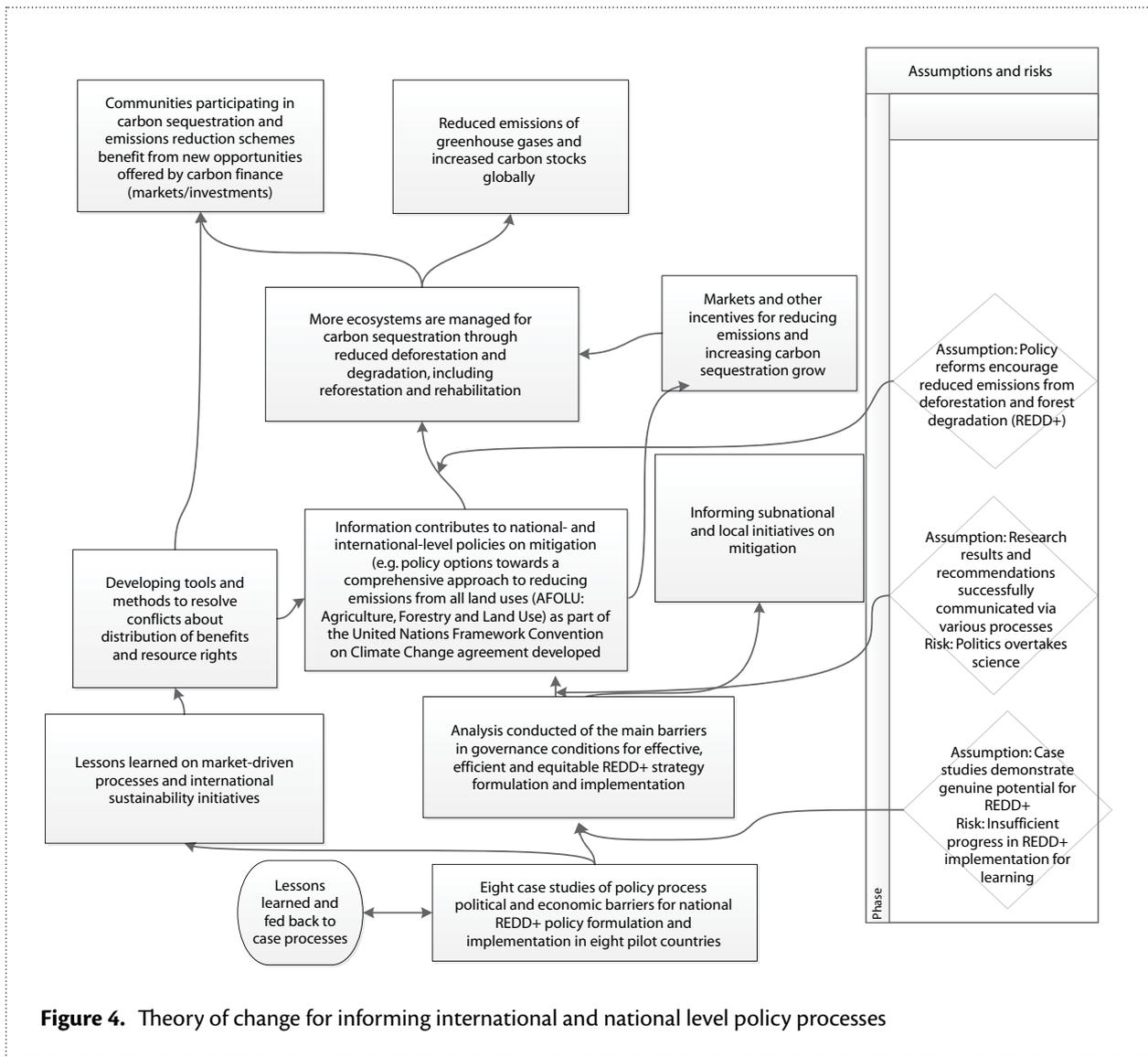


Figure 4. Theory of change for informing international and national level policy processes

relevant intervention literature are being formalised (Brouselle et al. 2009; Brouselle and Champagne 2011; Leeuw 2012)

- identify what needs to be monitored to help steer the intervention and support credible evaluation in the future.

As noted earlier, a challenge in building theories of change ex ante is identifying a reasonable set of plausible and relevant assumptions. Good practice would be to annually review and update the theory of change as increased understanding of the intervention and its context is acquired. Updating may also suggest the need to adapt the implementation of the intervention activities.

When an impact evaluation is being done, the theory of change in this ex-post perspective provides a basis

for designing the data collection and analysis, looking for evidence to confirm or refute the theory of change. It provides, as has been discussed, the basis for making causal claims about the intervention. Just how to do so will be discussed later.

Identifying assumptions from an ex-post perspective is somewhat more straightforward than in the ex-ante case. The evaluation can explore if the expected assumptions came about and if those involved in the intervention have evidence of any other factors that occurred that were needed for any observed results to be realised. As a result, once again the theory of change can be updated.

From both ex-ante and ex-post perspectives, theories of changes can be an essential tool for designing and evaluating NRMR interventions.

Key features of NRMR program theories of change

Considerable attention has been paid over the past few years to setting out versions of impact pathways for NRMR programs. Walker et al. (2008) set out guidelines for developing impact pathways. Maredia (2009) presents a generalised impact pathway for agriculture interventions aimed at food security. The Independent Science and Partnership Council (2012) sets a number of generic impact pathways for CGIAR research activities. And some programs, such as the Ganges Basin Development Challenge, have developed quite specific outcome pathways for their projects.

This report was asked to consider key features of generic theories of change for NRMR programs. We have attempted to stand back and take a look at these past efforts and to consider what might be characteristic of theories of change in this field. To do this, we have tried to first identify the generic types of results—outputs, outcomes and impacts—associated with NRMR programs, discuss the need to set out key assumptions (supporting factors) and then identify what such key features might be for the purpose of evaluation.

It should be noted, of course, that not all NRMR programs engage in all the activities and results described below.

Generic NRMR results

Table 1 sets out generic results for NRMR programs, which are discussed below.

Activities

Three types of intervention activities appear to be being made:

- NRMR *research* activities cover a wide spectrum of scientific and social research. Sometimes this research is undertaken in a participatory manner or as action research.
- *Capacity-building* activities are aimed at increasing the capability of the wide variety of actors that the interventions are trying to influence.
- *Engagement activities* cover the range of activities undertaken to communicate with, work with and/or influence a variety of partners through meetings and seminars, having discussions, co-learning, demonstrating good practices etc.

Outputs

These activities generate two generic outputs:

- *New information and understanding* is created through the various research activities undertaken and the capacity building done. There are, perhaps, two types of information and understanding outputs:
 - research information and technologies
 - new understanding for action.
- *Engagement*, the other output, covers the various advocacy and capacity development events held, publications distributed etc. undertaken to make use of the new knowledge created to inform and influence the research groups of interest.

Table 1. Generic natural resource management research program activities, target groups and results

Type	Specific examples
Activities	
Research on	<ul style="list-style-type: none"> • germplasm • technological practices • local production systems • postharvest • management practices at the farm, community and regional level • large ecosystems • policy • markets • local institutions
Capacity building	<ul style="list-style-type: none"> • providing training courses and workshops • developing training material • supporting curriculum development • supporting 'learning by doing' through research partnerships • developing new models of research (e.g. action-research approaches) • building research and innovation networks
Engagement activities	<ul style="list-style-type: none"> • holding meetings and seminars • having discussions • co-learning • demonstrating good practices • opening institutional spaces for partnerships
Outputs	
New information and understanding	<p><i>Research information and technologies</i></p> <ul style="list-style-type: none"> • research findings • databases • new information in briefing notes, presentations, videos, training material • new prototype technologies • new germplasm varieties <p><i>New understanding for action</i></p> <ul style="list-style-type: none"> • new practice and analytical models • new models for implementing research and innovation projects • better understanding of the role of policy and of local institutions
Engagement (including advocacy and capacity development efforts)	<ul style="list-style-type: none"> • events and meetings held • training provided • publications communicated • networks built

continued ...

Table 1. (continued)

Type	Specific examples
Research partners	<ul style="list-style-type: none"> • individuals and households (farmers, fishers) • communities • regions of communities • advanced research institutions • CGIAR researchers from different centres and disciplines • National Agricultural Research and Extension System individuals/groups (local researchers, officials) • private-sector groups (seed producers, microfinance agencies etc.) • regional/national policymakers • local, national and international non-government organisations
Immediate outcomes	<ul style="list-style-type: none"> • changes in knowledge, attitudes and skills of research partners and beneficiaries • increased social capital
Intermediate outcomes	<ul style="list-style-type: none"> • changes in practices and their direct effects • new policies and their direct effects • new institutions and their direct effects
System-level outcomes/ impacts	<ul style="list-style-type: none"> • increasing food security • reducing rural poverty • reducing under nutrition • sustainable management of natural resources

To arrive at general lessons about the kinds of activities undertaken in NRMR programs, it is important to be able to identify the types of intervention strategies used. The main engagement strategies used seem to include:

- presenting and discussing evidence
- co-learning
- building capacity
- mobilising and networking with non-government organisations and perhaps civil society organisation.

For each of these, there could be a specific theory of change developed.

Research partners

These strategies are aimed at influencing a number of research partners.

The complexity of the problems that NRMR interventions within natural resource and human system are trying to address means that, as seen in Table 1, there is often a need for action by a wide range

of actors to effect real and lasting improvements. Note that here we are also de-facto defining the different scaling levels in an NRMR program as different groups, namely farms/households, communities and regions.

Some of these groups are ‘next users’—both within and among these groups—who directly make use of, or are influenced by, the outputs, and some are ‘end users’ who are less directly influenced by the outputs, and are influenced more directly through next users. Theories of change could be developed for some or all of these research partner groups, setting out just how it is expected that they will be influenced, and to what end.

Outcomes

Working with next and end users, the outputs generated are expected to lead to a sequence of *outcomes*:

- *Immediate outcomes*—In the first instance, NRMR interventions are trying to change the *knowledge, awareness, skills and opportunities* of individuals or groups. As a result, *capacity* is increased.

- *Intermediate development outcomes*—Immediate outcomes are expected to result in two generic intermediate outcomes: *changes in practices* of the groups, and the *direct effects* of those practice changes.

Practice and behavioural changes are expected to result, in turn, in a number of further generic *end outcomes*:

- *new policies and policy instruments*
- *new or better functioning institutions*, such as new or improved governance arrangements (management systems, networks, planning bodies etc.) and/or markets
- *increased productivity* for the beneficiaries
- *improved distribution of benefits*.

The ultimate aim of NRMR interventions is to achieve social, economic and environmental benefits. For example, CGIAR describes its impact in terms of its beneficiaries enjoying:

- increased food security
- reduced rural poverty
- reduced under-nutrition
- sustainable management of natural resources.

Issues around NRMR theories of change

As might be imagined there are many ways to display these various groups of results in a theory of change. One way is shown in Figure 5, which presents an indicative theory of change for NRMR programs. It is not intended to be a complete theory of change, but rather the purpose is to give an indication of the key elements in theories of change associated with NRMR programs and the characteristics that need to be considered.

Figure 5 assumes that the main theory of change for an NRMR program is the sequence of events and conditions that affect the intended beneficiaries—the poor farmers and fishers. This beneficiary *program theory of change* is spelled out in a little more detail, with an indication of the kinds of key assumptions (supporting factors) needed for change to occur

and result in increased productivity, which will lead, in turn, to increased food security, reduced rural poverty, reduced under-nutrition and more sustainable management of natural resources.

As argued earlier, a theory of change needs to be more than a series of boxes and arrows. The key is what is behind the arrows. What are the hidden *assumptions*; the events and conditions needed to happen for a link in the results chain to work? These are the *supporting factors* in the discussion on causality; that is, necessary events and conditions for the NRMR intervention to work. Identifying assumptions is not straightforward, but essential if one wants to build a theory of change and to explore whether the intervention is a contributory cause; that is, made a difference.

The other major intermediate outcomes are shown here as, in essence, supporting factors: not ends in themselves but activities needed to sustain results at the farm level and to scale up results to the community and regional levels. They are presented here as part of the *implementation theory of change*.

The different levels of results are shown in Figure 5. They are shown in a linear sequence for ease of discussion but there may be *feedback loops* between different levels, as indicated by the double-ended arrows. Further, the *trajectories of change* are not shown—the time lines—but would likely be quite different among the various target groups. That is, change will occur at different times for different groups.

Result trajectories and interactions among results levels will be affected by a number of *external factors*. Factors such as larger commodity markets, natural disasters and diseases, broad government policies, and social trends can all significantly affect how an expected theory of change will play out in practice.

Figure 5 illustrates that, when considering the theory of change for a complex NRMR program, there is a need for a series of nested theories of change; that is, separate theories of change to capture how the interventions are expected to work at different levels of implementation.

There are clearly limits to how much detail can be usefully set out in a program-level theory of change, especially for a large and multifaceted intervention like an NRMR program. A complete theory of change would be too complex and not useful for presenting a

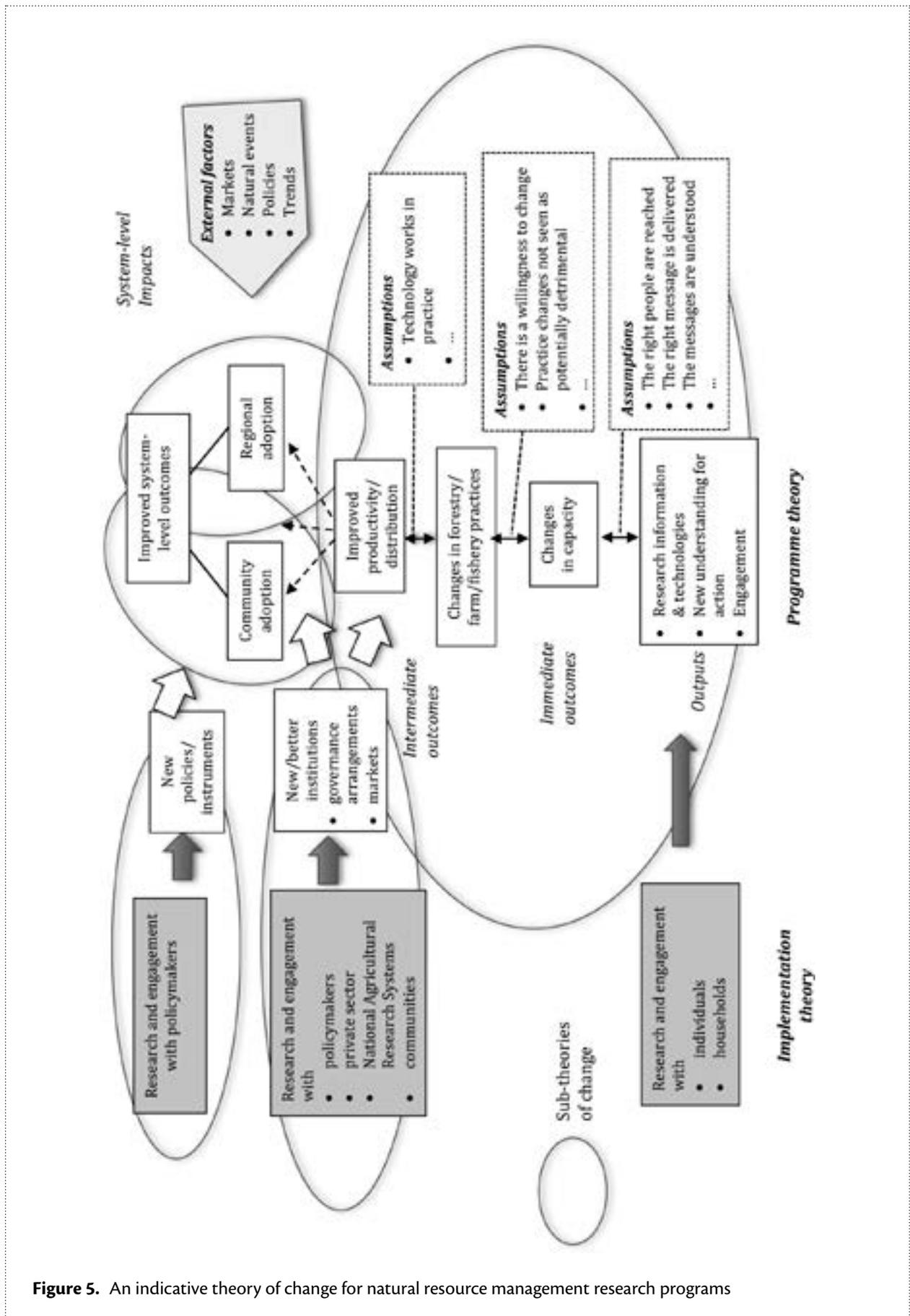


Figure 5. An indicative theory of change for natural resource management research programs

picture guiding an evaluation. A program-level theory of change should show the programmatic added-value. For example, the Aquatic Agricultural Systems program-level theory of change is about influencing how other NRMR programs conceptualise, plan and implement NRMR to maximise impact. It does so by providing communication and other support to the hub-level programs-of-work, to help create an enabling environment for them.

What is needed then, is a number of more detailed nested theories of change to describe the various levels of the program's interventions. The oval lines in Figure 5 indicate several such nested theories of change. Others are also possible. These relate to various ways of cutting-up the program, as discussed earlier (see Box 1). In the top left of Figure 5, two possible theories of change are indicated, focusing on the efforts undertaken to bring about new policies and new institutions. Depending on the specifics of the intervention context, these could be further broken down into the efforts to influence specific research partners.

At the top-middle of Figure 5, possible theories of change to set out how scaling up to the community and to the regional level are intended to work are indicated. The large oval in the lower middle-right of Figure 5 is the partially developed theory of change aimed at improving the livelihoods of beneficiaries.

Where theories of change in NRMR programs are often not spelled out is in linking intermediate outcomes to the longer term impacts of increasing food security, reducing poverty and under-nutrition, and ensuring sustainable natural resource management. The 'theory' is simply that a combination of the right new policies, new institutions, new practices and increased income will bring about the desired impacts. And at a generic or general level, this may be all that can be said. In a specific sub-theory of change, it will be reasonable to expect that a more detailed path to impact can be laid out.

However, in many cases, it is expected that these end impacts will not be realised until some time—often many years—after the termination of the program. The need therefore might be to undertake studies in the future to explore the extent to which and by what means any end impacts came about. These studies are on the boundaries of research and evaluation and would provide an evidential basis for articulating the upper levels of NRMR theories of change as well as longer term outcomes.

Focusing on how intermediate outcomes are expected to be brought about is a quite reasonable perspective to take, and for which theories of change can be developed. It is also a perspective that seems to be gaining credibility in CGIAR, as indicated by the discussion of 'intermediate development outcomes' (Independent Science and Partnership Council 2012). And these are areas where a substantial literature is often available. For example, *evaluating advocacy* has a large literature¹³ that could serve as the basis of the theories of change looking at realising new policies and new institutions through engagement. And within this, there is a large literature on *research and new knowledge influencing policy*, as well as a related literature on *getting from knowledge to action*. Further, engagement here includes *capacity-building* efforts, on which there is also a large and separate evaluation literature¹⁴.

Key features of generic NRMR theories of change

What are then some of the key features of these theories of change for NRMR programs? The following are suggested:

- A program-level NRMR theory of change can be indicative of only the big picture theory behind the program. That is still useful as an overall picture.
- There will usually be a need for developing a number of nested sub-theories of change around different intervention strategies and/or different target groups operating at different levels. For example, the Aquatic Agricultural Systems program works at three levels—program, hub and project—and theories of change are likely needed at each level.
- Theories of change for the range of engagement activities will be an essential element of NRMR theories of change.
- Within engagement activities with research partners, theories of change for how research can

¹³ Some references on advocacy evaluation are provided in the discussion of the Ganges Basin case.

¹⁴ Some references evaluating capacity development are provided in the discussion of the African Food Security Initiative.

influence policy is an area on which there is much literature, as there also is for evaluating capacity building.

- To make use of the perspective that NRMR interventions are contributory causes, any theories of change developed for or within NRM programs need to at least include the underlying assumptions—the supporting factors—in the intervention causal package.
- Result trajectories for NRMR theories of change are unlikely to be linear in either time or direction.

Evaluation approaches: an overview

In addition to the more experimental approaches to evaluation often argued for in CGIAR (see, for example, Maredia (2009)), there is a broad range of approaches that can be considered. Many of these will be discussed in this report. The challenge is to know when to use which approach or approaches.

A number of the arguments and frameworks used in this report have been adopted from the Stern et al. (2012) study. Figure 1 from that study is given as Figure 6 here. The figure illustrates the fact that selecting an evaluation design depends on:

- the evaluation questions that are to be addressed
- the attributes of the program and its context being evaluated
- the evaluation designs and methods available.

This is a fairly obvious observation, but evaluations have too often been methods driven; that is, the evaluation adopts a particular method, then sets out relevant evaluation questions, adjusting the method(s) as needed

to fit the context and the program. This report follows the heuristic in Figure 6. Key characteristics (attributes) of NRMR programs were identified in the introduction to this report. The next section discusses evaluation questions that might be addressed in an evaluation of NRMR programs.

Of course, there are at least two other practical considerations that need to be taken into account in choosing a design: timing and resources. In aiming to make evaluations useful, getting the evaluation done 'on time' to be available for input into other decision-making is important. Timing, especially short time frames, may limit the available possible designs. The extent of resources available for the evaluation is another obvious constraint. Some designs will be more costly than others.

Timing and resource constraints may result in trade-offs being made on credibility and coverage of evaluation issues. Unrealistic timing and resources will necessarily result in weak evaluations.

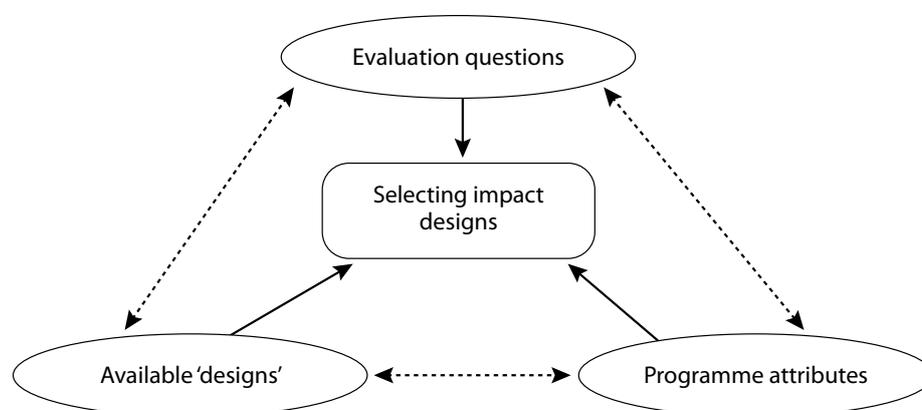


Figure 6. Considering impact evaluation designs

Evaluation questions

By their very nature, NRMR programs endeavouring to accomplish high-level and longer term impacts will need to work with beneficiaries and partners in a complex, dynamic and evolving socioeconomic context. While reporting on high-level accomplishments is important, perhaps even more important—given the ambitious goals being sought, and the fact that many of these efforts are complex and perhaps new territory for the research centres involved—is learning about how and to what extent these programs and their subprograms are working in terms of the more direct NRMR outcomes accomplished. Further, most of the time, linking higher level impacts to the NRMR activities cannot be done credibly without an understanding of the intermediate steps along the causal pathway.

This leads to a large variety of possible evaluation questions. What is important is that attention is paid to just which evaluation questions are to be posed in an evaluation and how the specific questions are articulated.

Key issues and concerns

Stern et al. (2012, p. 37) discuss four generic questions focused on assessing different aspects of the cause–effect question:

- To what extent can a specific (net) impact be attributed to the intervention?
- Did the intervention make a difference?
- How has the intervention made a difference?
- Will the intervention work elsewhere?

As noted earlier, impact assessments with a primary aim of providing relevant information for accountability

purposes tend to target the first of the above questions. Our earlier discussion on causality focused on the second and third of these questions.

A framework for NRMR evaluation questions

The intent here is to set out a range of questions that could be addressed in an evaluation of NRMR programs and interventions focused primarily on learning¹⁵. The scope is therefore beyond just looking at the impact the program has had and may have. The range of issues needs to include consideration of:

- the continued rationale for the program
- the implementation of the program
- to what extent the program worked; that is, had the intended results,
- why the program works the way it does
- whether the program will continue to work
- the extent to which the program would work elsewhere.

Note that how evaluation questions are articulated depends in part on when the evaluation is being done; ex ante when the program is being designed, or ex post¹⁶ some time after the program has been operating (such as mid-term) or has been completed.

¹⁵ As noted earlier, a primary focus on learning does not exclude attention to accountability interests.

¹⁶ We are using the term ex post here to include not only the time soon after the program has ended but also cases where the program is still running or perhaps mature. Whatever the case, often many of the impacts (and perhaps some outcomes) may not yet be realised.

Using this six-part framework, a framework for NRMR evaluation questions is set out in Box 2 and elaborated below. There are both ex-ante and ex-post evaluation questions that can be raised. Appendix 1 provides further details on the kinds of sub-questions that can be asked in each group of questions.

Box 2. A framework for evaluation questions

Ex ante:

1. Should it work?

Ex post:

1. Should it still work?
2. Has implementation worked?
3. Did it work?
4. How and why does it work?
5. Will it continue to work?
6. Will it work elsewhere?

Should it work?

The basic ex-ante question is:

- Is the rationale for the program and its design sound?

The program is expected to be neither a random collection of activities nor just a collection of past activities, but rather a well thought out response to perceived problems, taking into account the current state of knowledge and experience, as expressed in a theory of change. At the outset, the rationale for the intervention as expressed in a theory of change needs to be developed, based on existing evidence and stakeholder views and experience.

CGIAR and ACIAR have experience in doing this, using impact pathways. Impact pathway models can be built on to identify the assumptions and risks behind the pathways, and the external factors at play. The result would be a mature and robust theory of change that would provide answers to the questions:

- How is it expected that the intervention will make a difference? For whom?

- On what evidence is the implementation design based?
- Are there any potential unintended effects that should be watched for?

Looking at an NRMR program some time after it has been in place provides the opportunity for re-examining the rationale for the intervention and its design in the light of more recent information and experience on its context and the problems being tackled. From an ex-post perspective, with the passage of time and the knowledge and understanding gained, the question becomes:

- Is, or was, the rationale for the program and its design still sound?

Has implementation worked?

As noted, NRMR programs can be quite complex and it would be expected that considerable learning has occurred about implementation practices such as collaboration and coordination with partners and beneficiaries, and that there is a better understanding of the different influences at work. This is important information to gather and assess to improve current design and delivery. It can also be useful information if the intervention is to be implemented elsewhere.

- What has been learned about how the NRMR program has been implemented?
- How the implementation has contributed to the results?
- Can the implementation lessons learned be transferred elsewhere?

Did it work?

A key interest is, of course, to know if the intended results (outputs, outcomes and impacts) of the intervention were realised, and if there were unintended results. There are two main questions here: Which results happened? Were the NRMR results ‘caused’ by the intervention? The first entails measuring the results that are observed and the second causally linking specific results with the specific intervention.

- What results have been realised?

This is a measurement question—attribution or contribution is not assessed. The measurement

encompasses outputs, immediate outcomes, intermediate outcomes and impacts.

- To what extent can a specific (net) impact be attributed to the intervention?¹⁷

These types of causal questions are behind experimental designs trying to attribute results to interventions based on regularity and counterfactual causal perspectives. They are what a number CGIAR impact assessments address, and are not discussed in this report.

- Did the intervention make a difference?

These types of causal questions are behind theory-based and case designs trying to assess the contribution the intervention is making to the observed results.

How and why did it work?

If we want to explain why a program is or is not working, we need to understand and confirm its theory of change, especially understanding the assumptions behind the theory of change, the mechanisms at work and the relevant context factors.

- How has the program made a difference? If not, why not?
- For whom has the program made a difference?
- Has the program led to any unintended results?

Here these questions are being asked ex post, with the expectation that the answers will be based on empirical evidence from the specific intervention.

Will it continue to work?

We may want to know not only if the program worked, but also whether it will continue the work in the future, or under what conditions:

- Are the program and its benefits sustainable?
- What are the future expected benefits?

Will it work elsewhere?

Finally, we may also be interested in assessing the extent to which the program or elements of it would work if it were implemented elsewhere; that is, which elements are very context bound, and which have more potential to be useful elsewhere?

- Can the program or elements of it be transferred elsewhere?

Considering the NRMR evaluation issues

Quite a few potential evaluation issues have been set out. In planning for a specific evaluation of an NRMR program, it would be necessary to focus the evaluation on fewer key issues. A common shortcoming of many evaluations is trying to answer too many questions. The result is often shallow treatment of many issues, and findings that many do not find credible. On the other hand, impact assessments often focus on a quite limited number of standard questions. Finding the right balance to examine the key issues of concern is critical.

Considerations in selecting evaluations issues would include:

- information needed to resolve current concerns about the program
- specific information needs for current and upcoming decision-making
- resources and timing available for the evaluation.

It is often crucial to avoid the desire to address 'nice to know' issues. Further, credible answers to many of these questions depend on being able to make reasonable claims about the causal links between the program and observed results.

¹⁷ Note that trying to attribute results to the specific NRMR intervention—and not to the actions of others as well—could be seen as damaging to the partnerships that have been built up.

Designs for NRMR programs

The general paradigm used in this report is adopted from Stern et al. (2012). It proposes that, based on the evaluation question to be addressed, the attributes of the entity being evaluated, the methods and designs available, and the time and resources available, an appropriate approach to the evaluation can be selected. In this report, the evaluation *approach* will refer to the overall methodological plan for conducting the evaluation; that is, the data collection *tools* and analysis *methods* to be used, and the *designs* used to make a causal claim about the intervention and observed outcomes and impacts.

Stern et al. (2012) set out a quite comprehensive list of designs that can be used in evaluations. They are summarised in Table 2, adopted from that paper.¹⁸

CGIAR impact assessments have tended to focus on experimental and statistical designs, with some use made of theory-based approaches, particularly impact pathways for explanation. Experimental and statistical designs certainly have their place in evaluation. But they also have significant limitations, especially for evaluations of NRMR programs beyond the farm level and for focusing on learning.

For example, experimental designs (e.g. randomised control trials) work best when:

- the main purpose is to explain the effects of a particular intervention in a specific setting rather than establishing generalisable knowledge or ‘laws’
- expected outcomes and the intervention itself are clearly understood and specifiable
- there is a consensus about likely causes and effects among key stakeholders

- it is possible to manipulate inputs and to limit interference (other influences) or confounding factors coming from the proximal context or other settings
- there is a likelihood of one primary cause and a primary effect
- the primary causal factor is not interdependent on other causal factors
- there are sufficient numbers (beneficiaries, households etc.) for statistical analysis.

There are CGIAR personnel who recognise the limitations of experimental and statistical designs. Derek Byerlee in the foreword to a recent report (de Janvry et al. 2011, p. vi) discusses the use of experimental designs and notes that:

In this respect, the experimental approach may have more relevance for evaluation in the early adoption stage for pilot testing the economic and social impacts of a new technology on a relatively smaller and well defined scale, than for large-scale ex-post impact assessment.

Smith et al. (2012) report on an evaluation of the impact of CARE’s SHOUHARDO¹⁹ program in Bangladesh. The evaluation faced context similar to NRMR programs—in particular the need to address structural issues as well as providing direct nutritional supplements—and accepted the limitations of experimental designs. Appendix 2 discusses how the researchers used a variety of statistical and comparative approaches to reach credible causal claims.

As did Stern et al. (2012), this report will be focusing on theory-based, case-based and participatory designs.

¹⁸ Chapter 3 in Stern et al. (2012) discusses each of these designs and variants.

¹⁹ Strengthening Household Ability to Respond to Development Opportunities

Table 2. Design, variants and causal inference

Designs	Specific variants	Basis for causal inference
Experimental	<ul style="list-style-type: none"> • Randomised control trials • Quasi experiments, • Natural experiments 	<ul style="list-style-type: none"> • Counterfactuals; the co-presence of cause and effects
Statistical	<ul style="list-style-type: none"> • Statistical modelling • Longitudinal studies • Econometrics 	<ul style="list-style-type: none"> • Correlation between cause and effect or between variables, influence of (usually) isolatable multiple causes on a single effect • Control for 'confounders'
Theory-based	<ul style="list-style-type: none"> • <i>Causal process designs</i>: theory of change, process tracing, contribution analysis, impact pathways, • <i>Causal mechanism designs</i>: realist evaluation, congruence analysis 	<ul style="list-style-type: none"> • Identification/confirmation of causal processes or 'chains', • Supporting factors and mechanisms at work in context
'Case-based'	<ul style="list-style-type: none"> • <i>Interpretative</i>: naturalistic, • grounded theory, ethnography • <i>Structured</i>: configurations, qualitative comparative analysis, within-case analysis, simulations and network analysis 	<ul style="list-style-type: none"> • Comparison across and within cases of combinations of causal factors • Analytic generalisation based on theory
Participatory	<ul style="list-style-type: none"> • <i>Normative designs</i>: participatory or democratic evaluation, empowerment evaluation, • <i>Agency designs</i>: learning by doing, policy dialogue, collaborative action research 	<ul style="list-style-type: none"> • Validation by participants that their actions and experienced effects are 'caused' by program • Adoption, customisation and commitment to a goal
Synthesis studies	<ul style="list-style-type: none"> • Meta analysis, narrative synthesis, realist based synthesis 	<ul style="list-style-type: none"> • Accumulation and aggregation within a number of perspectives (statistical, theory-based, ethnographic etc.)

Source: Table 3.3 in Stern et al. (2012)

But the key point is that, in evaluations of most NRMR programs, the overall design will involve a mix of a number of specific designs to address aspects of the larger program and/or different evaluation issues. Thus, there will be occasions within an overall impact evaluation design when experiments will be a useful way to answer particular evaluation questions.

Relating tools, methods and designs to the evaluation questions

The range of evaluation question being considered here is broader than those dealing with causality, and includes questions of rationale, implementation and measuring results. In these cases, unlike for the causal-related questions, there is no causal design employed. All approaches use a variety of data-collection tools and analysis methods.

Addressing rationale

- *Is the rationale for the program and its design still sound?*

These questions are challenging the overall aims of the program, the interpretation of the problems being tackled and the design of the program—its theory of change—in light of the current state of understanding. These issues would be investigated using a combination of:

- surveys and interviews of stakeholders (staff, partners, beneficiaries) and subject-matter experts
- reviews of program documentation, including monitoring data available
- reviews of relevant literature on efforts elsewhere to solve these types of problems
- assessments of the context of the program
- logical analysis of a proposed theory of change.²⁰

²⁰ See Brousselle and Champagne (2001) for a discussion

The underlying question is whether the current design of the program and its theory of change still makes sense, or has new insight been acquired that argues for a different articulation of the issues being addressed or the way they are being addressed.

Addressing implementation

- *What has been learned about how the NRM program has been implemented?*
- *How has the implementation contributed to the results?*
- *Can the implementation lessons learned be transferred elsewhere?*

While the rationale issues look at the program theory, here the implementation theory would be examined. Implementation theory (Weiss 1997; Blamey and Mackenzie 2007; Rogers 2007) articulates the various activities of the program and links them to the results they are expected to influence. If not developed as part of the theory of change, a specific implementation theory would need to be developed. Assessing it would use the same tools as those used in looking at rationale, but with a focus on implementation.

Measuring results

- *What results have been realised?*

Measuring the results associated with an intervention can be quite challenging. The full range of social science and economic tools can be employed. Results can be measured when the program starts (a baseline), during the life of the program (perhaps through monitoring), at the end of the program, or several years after the end when it is expected that the impacts of interest will have been realised.

Challenges can be:

- Results of interest may involve concepts for which there is no agreement on methods of measurement, such as food security or good governance. However, in almost all cases, some efforts have been made to measure such concepts.
- Measuring may be quite expensive, such as when large surveys are needed, and hence can be done only selectively.
- It may be difficult to get reliable measures.

To be able to draw conclusions on causal issues, the remaining sets of evaluation questions identified earlier need, either explicitly or implicitly, a (causal) design.

Demonstrating and explaining causality: causal designs

- *To what extent can a specific (net) impact be attributed to the intervention?*
- *Did the intervention make a difference?*
- *How has the program made a difference and, if not, why?*
- *For whom has the program made a difference?*
- *Has the program led to in any unintended results?*

The literature on causality discusses several alternative perspectives on how to interpret and think about it.

There are at least four ways of conceptualising and addressing causality (adapted from Stern et al. (2012, pp. 16–17):

- *regularity frameworks* that depend on the frequency of association between cause and effect—the basis for statistical approaches to making causal claims
- *counterfactual frameworks* that depend on the difference between two otherwise identical cases—the basis for experimental and quasi-experimental approaches to making causal claims
- *comparative frameworks* that depend on combinations of causes that lead to an effect—the basis for ‘configurational’ approaches to making causal claims, such as qualitative comparative analysis (QCA)
- *generative frameworks* that depend on identifying the causal links and mechanisms that explain effects—the basis for theory-based and realist approaches to making causal claims.

These different logics aim primarily at answering different types of causal questions and are the basis for different designs and methods that can be used in evaluation. Regularity and counterfactual approaches address questions of the form: To what extent can a specific (net) impact be attributed to the intervention? CGIAR impact assessment has been exclusively focused on making causal claims using counterfactual-based designs and, indeed, defines impact assessment in those terms (Science Council Secretariat 2006).

This report makes a distinction between counterfactual logic as a basis for designs and methods, and counterfactual thinking as a general way of thinking—imagining how the world might be if interventions or programs did not exist. In many circumstances where the necessary preconditions for counterfactual designs do not hold, it is still useful to ‘imagine the counterfactual’ even though such designs are not appropriate.

For example, questions of the form ‘has the intervention made a difference?’ may be better addressed by ‘comparative’ rather than counterfactual-based designs. (An example of this is discussed below.)

Generative frameworks also target this causal question. In addition, by focusing on the mechanisms that connect a cause and effect, they can help explain how and why the results were brought about. Generative

designs are probably of most interest for NRMR programs, which often depend on explanations to steer program implementation over extended time periods. Generative causality is often how we interpret causality in everyday life. If one is trying to find the cause of an event, it is typical to trace back what happened before it to see what ‘caused’ it. These designs are the focus of this report.

Qualitative comparative analysis

QCA is an alternative approach to examining contributory causes that uses a comparative framework perspective. Ragin (1987) has developed a set of methods that use Boolean algebra to reduce the number of causal combinations and so produce reduced sets of combinations that can be considered to represent sufficient causal configurations. The approach sees individual cases as comprising sets of attributes (supporting factors). Cases are coded as having membership in a set of causal conditions.

Consider the fish-farming example discussed earlier. Here a case would be a particular farming household. The case *outcomes* could be ‘adopt freshwater fish farming’ (Yes/No). The *attributes* of each case (the earlier supporting factors) might fall into two wide groupings relating to the intervention itself and to context:

Intervention attributes:

- provided initial funds for equipment (yes/no or amount)
- provided fingerlings (yes/no or amount)
- provided advice on fish farming methods (yes/no or amount).

Context attributes:

- number of other farmers who were convinced by the intervention and/or, by example, to try fish farming
- an adequate market for the fish surplus to family consumption is within x km (yes/no and km)
- there is an adequate supply of affordable fingerlings (yes/no)
- there is support by the farmer’s family for the additional work involved (yes/no)

- the farmer can see visible improvement of the lives of adopters over time (yes/no)
- the farmer has adequate cash to buy fish food and other supplies (yes/no).

The QCA then compares households that have adopted fish farming, and their attributes, and households that started but dropped fish farming, and their attributes. The result of the QCA could be that, for example:

IF farmers had x interventions AND y context attributes were present THEN they become adopters.

QCA provides a credible causal claim that the intervention is a contributory cause. Note that it is a claim of association; that a certain causal package is logically associated with a particular effect. QCA can answer the question ‘Did it make a difference?’, but not the question ‘Why has the result occurred?’ For those types of evaluation issues, to be able to understand how the factors and conditions combined to bring about the result, a theory of change is required.

Theory of change approaches to showing a contributory cause

A generative approach to demonstrating causality usually involves some form of theory of change. One application of this thinking is the use of impact pathways (Douthwaite et al. 2008)—and related theory of change approaches—that are developed to trace out how it is expected that the intervention will lead to the anticipated impacts. Showing that each step along the pathway was caused by its predecessor and can be explained, is using a generative view of causation to get at the causal link between the intervention and the impact. We argued earlier that what is needed for this approach is a full theory of change that lays out the various assumptions involved. Theory-based approaches to evaluation (Weiss 1997; Stame 2004; Rogers 2007; White 2009; Funnell and Rogers 2011) such as contribution analysis (Mayne 2008, 2011) use these models of causation to make causal claims.

In particular, contribution analysis can be used to demonstrate that an intervention is a contributory cause. This is a generative theory-based approach to show that the theory of change occurred, is plausible and that rival explanations have been accounted for. In a recent special issue of the journal *Evaluation*, several authors discuss the practice and concepts around

contribution analysis (Mayne 2012b). Contribution analysis confirms:

- that the expected result occurred
- that the causal package is sufficient
 - the supporting causal factors—the assumptions for each link in the theory of change—have occurred and together provide a reasonable explanation for the occurrence of the results
 - any other identified supporting factor that occurred has been included in the causal package, revising the theory of change as needed
 - important plausible rival explanations have been accounted for
- that the intervention is a necessary part of the causal package
 - without the activities and outputs of the intervention, the supporting factors alone are not sufficient to bring about the results.

It was noted earlier that the assumptions might be no more than likely necessary conditions. Thus, in a specific case, not all the assumptions in a theory of change may have occurred, in which case an assessment is needed to determine whether the collection of actual supporting factors (assumptions) provides a reasonable explanation for the occurrence of the observed result. This analysis, plus the assessment of plausible rival explanations, allows for the causal inference to be made as to whether the intervention causal package (for the link) was a contributory cause. If it was, and all the other links in the causal chain are also confirmed, then the theory of change itself has been confirmed.

Data and evidence for the contribution analysis would come from applying logic and critical thinking, and the results of previous research, and asking relevant stakeholders about each link and whether they believe that there were other causal factors beyond the package at work. If there were, one would need to seek evidence for their belief. Note that the links in a theory of change should comprise relatively proximate cause and effects, thereby making judgment and the use of logic and critical thinking easier.

Consider the fish-farming intervention described earlier: how could we explore whether the intervention was a contributory cause and hence made a difference?

We would first need to observe whether or not the key expected results occurred; in particular, in terms of initial adopters maintaining their fish farms over time, and whether neighbours started and maintained fish farms. If this did happen, then we would like to confirm that setting-up the initial adopters was a contributory cause. If not, we would then need to ask why the theory of change work did not work as intended—at what point did the theory break down? In either case, one would want to explore if the supporting factors came about and whether other factors not in the causal package played a role and, in particular, if there were rival explanations for the observed results. Possibilities might be a general rise in interest in fish farming, or a significant rise in the price of commercial fish. Conversely, if the initiative did not work, which supporting factors were not realised? Finally, in this case, the intervention would appear to have played a triggering role. The early adopters did so as a direct result of the intervention activities, starting the causal chain.

Addressing sustainability

- *Are the program and its benefits sustainable?*
- *What are the future expected benefits?*

The second of these questions is one that impact assessments typically approach using economic modelling approaches. This report focuses on the first of these two questions.

The notion of sustainability is now widely diffused and applied. For the purposes of this report, two understandings of sustainability are important:

- the sustainability of the outcomes and longer term impacts that result from NRMR programs
- the evaluation of the environmental/ecological effects of NRM programs.

Sustainability of program results

This meaning of sustainability matters because, in the NRMR field: (a) many programs aim for results that can be fully achieved only in the long term; (b) uncertainties in programming, knowledge, and policy mean that existing assumptions may become outdated before

intended results are realised; and (c) many programs are themselves long term.

These attributes of NRMR programs pose severe evaluation challenges. For example:

- How can we know that initial or short-term program results will continue after the program ends?
- How should we structure an evaluation so as to capture different time slices within an extended program or intervention?
- What methods and criteria are suited to radical uncertainty over the medium and long term?

These kinds of challenges suggest a number of possible solutions:

- *Ongoing and staged evaluations that can capture results at different points in time.* It is probably unwise to commit all evaluation resources at a single moment in the program life cycle. The challenge is to decide which different times are sensible and how an evaluation design should be modified at each stage. From a real-world perspective, those who fund evaluations are usually orientated to the shorter term. This suggests that setting up a longer term monitoring system should be a priority even in the short term. It may also be useful to consider long-term evaluations as a task that overlaps with research and planning them as such.
- *As part of any proposed program, undertaking a risk analysis that identifies the factors that are likely to sustain or undermine the program's sustainability.* There have been a number of efforts to identify the factors that sustain program results, in particular within international development and environmental studies. A useful summary of sustainability in the development literature is in CIDA (2002). It suggests items to be considered when planning for sustainability that are also applicable to its evaluation. Other useful information has been produced by agencies such as the Organisation for Economic Co-operation and Development (OECD 2011, 2012) and, from time to time, UN institutions such as the United National Environment Programme.
- *Sustainability may need to be conceptualised as more than continuity.* One implicit assumption

in discussions of sustainability is that it can be equated to continuity in a static form. The literature on diffusion of innovation would suggest that a broader conceptualisation is needed. For example, sustainability could involve:

- the maintenance of particular solutions, arrangements or practices
- an adaptation of solutions to match changing circumstances
- the absorption of particular program 'results' into broader institutional and policy norms.

An assessment can be made of the likelihood that future outcomes and impacts will be realised, given current understanding and information. The theory of change for the intervention sets out what is expected to happen in the future. At a point later in the life of the intervention when an evaluation is being done, the theory of change can be reassessed to see if it still seems reasonable in light of what has been learned in the interim about the context, assumptions and what has been realised to date. One might then conclude that, yes, it is still quite likely that the future expected benefits will occur, or that the current reality suggests otherwise.

Environmental and ecological sustainability

Sustainable development has always focused on the interconnections between three pillars—environmental, economic and social development (UN 1987). While it may be difficult to separate out the environmental 'pillar' of sustainable development, there are distinctive environmental priorities, although the scope and variety of these are daunting. Many environmental priorities are associated with both the inputs (water, fertiliser, resource competition etc.) and outputs (soil erosion, pollution, habitat destruction etc.) of agriculture. Similarly, proposals for sustainable policies often focus on agriculture—as with sustainable fisheries, conservation agriculture, forestry, agro-forestry and aquaculture.

As in other aspects of NRMR, timescale and duration are key factors. Environmental consequences may become apparent only after damage has become irreversible. This places a high priority on anticipatory, ex-ante, evaluation (modelling, scenario building and prospective impact assessment) as well as monitoring and ex-post evaluation.

Given the multilevel nature of NRMR, an added complication is linkages between environmental and other consequences at farm, regional and global levels. A positive effect for a particular farm system (e.g. productivity gains from intensification) may have negative effects at a regional level (e.g. in terms of overuse of water). Similarly, 'positive' environment consequences at a global level may have negative effects on the livelihoods of poor farmers. This suggests that one operational question relates to how far it is possible to use a common system of targets or indicators at different system levels.

In relation to environmental sustainability, the key evaluation issues for this report are:

- how to anticipate and track the environmental consequences of particular NRMR programs
- how to capture and analyse the interconnections between environmental, social and economic consequences of NRM
- how to track the connections between different levels of NRMR programs.

Addressing generalisation and transfer

- *Can the program or elements of it be transferred elsewhere?*

Scientists and policymakers are keen on generalisation, so that 'laws' can be established and findings scaled up and applied elsewhere. Especially in an accountability-driven evaluation, being able to vouchsafe that a particular approach that has been shown to 'work here' will also 'work everywhere' is attractive. In reality, most such generalisation claims are limited to only some settings. For example, one of the main incentives to supplement experimental impact evaluations with theory-based approaches is because it has been found that it is often difficult to extend even strong findings from one setting to another (White 2009).

We would suggest that in a learning-orientated evaluation the key question is not 'can evidence of the impact of an innovation be generalised?', but rather, 'under what circumstances will what kind of

innovation or practice be successfully transferred?' A useful approach to answering this second question is to distinguish between two aspects of an innovation process: the characteristic of the innovation that is to be 'transferred'; and the characteristics of the context or setting into which it is hoped to make the transfer.²¹

This approach tends to focus more on 'transferability' as a core concept than on 'generalisability'. The shift in language is rooted in the action emphasis that informs most learning-orientated evaluation rationales: a wish to take an innovation further and apply it elsewhere. However, there is no reason that the same broad approach cannot also inform arguments about 'generalisability'.

One problem with a simple notion of transferability (or generalisation) is to know what it is that one hopes to transfer. There can be many attributes of the 'object' to be transferred but one important distinction is the specificity of the innovation itself. One may be seeking to transfer either a very specific or a far more general innovation; for example, one might want to transfer:

- a specific new governance arrangement for a water basin that involves local participation through existing local community entities and that has been shown to be effective
- a broadly similar participatory approach but through a variety of available channels for participation that might include an administrative structure and non-government organisations as well as community entities
- the broad idea that some improvement in water basin management is needed but remaining open to which approach (participatory, institutional, market-based etc.) should be applied in any site for transfer.

The generic distinction is between a specific innovation; a similar type of innovation; and an innovation that, although different in many details, is built on similar principles.

A second problem is to know how to characterise settings. Everyone acknowledges that context is key—frequently associated with the use of the word

²¹ This framework was developed by Stern and others as part of a program of evaluations of the transfer and learning of policies and practices between European Union member states; see Commission of the European Communities (1996), Stern (1996) and (Turbin 2001).

‘customisation’. Again there are many ways in which ‘transfer-settings’ can be characterised but, in broad terms, a ‘similar or different’ scale of some sort seems to be useful. This of course leaves open which aspect of similarity or difference is the basis for the scale. To pursue the above example, we can imagine three types of ‘transfer setting’:

- one virtually identical to the setting in which a successful innovation has been successful—that is, scale and complexity of water basin and community structures and norms that are very similar
- broadly similar settings in terms of both nature of water basin and community structures
- a mainly different setting—although retaining some similarity in terms of problems of water management, the presence of a water basin, some kind of local population and a local administrative structure.

The aim of this kind of typology is as a heuristic guide to aid decisions about how to approach transfer and justify generalisability. Table 3 combines these distinctions and suggests what might be reasonable expectation of transferability within this kind of framework.

The answer to the framing question posed at the beginning of this section—under what circumstances will what kind of innovation be successfully transferred?—is that, depending on the similarity of the ‘transfer setting’ (or context), the innovation to be transferred will have different likelihood of success. In an identical setting, the selfsame innovation is most likely to be transferable—generalisability could here be described as high. It is not that transfer of a highly specific innovation is impossible in more diverse settings but more barriers are likely to be present. Hence, the likelihood of successful transfer in more diverse settings would have only a medium chance of success; and if

settings are highly diverse, the likelihood of successful transfer would be lowest. In a similar but not identical setting, a similar type of innovation is likely to have a ‘high’ chance of successful transfer in both identical and similar settings, but could be expected only to be ‘medium’ across highly diverse settings. Finally, an innovation that is built on similar assumptions and principles—which may be operationalised in flexible ways—stands a ‘high’ chance of successful transfer in both identical and similar settings. Furthermore, even the chances of transfer to highly diverse settings are helped by the malleability of the innovation itself, so that even across such settings the chance of transfer success can still be argued to be ‘medium’.

A concomitant of this way of framing transferability is that the larger, more complex and multilevel an innovation, intervention or program, the more likely it is that what can be transferred will be at the ‘principle’ end of the spectrum. This follows from the *improbability* that all aspects of a transfer-setting and ‘transfer-object’ in a large complex program will be morphologically identical.

As is implied by the framing question, neither transferability nor generalisability is understood in absolute terms: the same innovation that ‘worked’ in one setting cannot be expected to work in all settings. This also highlights the need to extend the scope of an impact evaluation to seriously investigate issues of transfer and generalisability. At the very least, there needs to be an effort in a learning/action mode to analyse both transfer-settings and innovation-objects to decide what can be transferred; and in an accountability mode, what generalisability claims are likely to be justified. This will be partly dealt with in any ‘nested’ design: a general analysis of institutional and agricultural starting-points in one setting will provide a useful guide to what it is that can be transferred to another setting. The discussion about transfer and

Table 3. Likelihood of ‘transfer’ success: high, medium and low

Innovation to be transferred	Similarity of ‘transfer settings’		
	Identical setting	Similar setting	Mainly different setting
A near identical innovation	High	Medium	Low
A similar type of innovation	High	High	Medium
An innovation built on similar underlying principles	High	High	Medium

generalisability also re-emphasises that, within a nested design, there will be different theories of change for what at first sight might appear to be similar processes that are expected to ‘make a difference’.

Summary of implications for addressing evaluation questions

Table 4 summarises the relationships between evaluation questions and evaluation tools, methods and designs for use in learning-focused impact evaluations.

Table 4. Summary of tools, methods and design implications for impact evaluation questions

Key evaluation question	Related evaluation questions	Underlying assumptions and requirements	Suitable tools, methods and designs
Is the rationale for the intervention and its design still sound?	To what extent are the goals of the program still relevant? Does the program design and implementation continue to be realistic and supported by current evidence and practice? Is the theory of change still sensible? Are there alternative strategies that should now be used?	The program comprises a coherent set of activities with common aims.	Surveys/interviews Document review Literature review Context analysis Logical analysis
What has been learned about implementation?	What has been learned about how the natural resource management research program has been implemented? How has the implementation contributed to the results? Can implementation lessons learned be transferred elsewhere?	There was a strategy behind implementation. Implementation was modified as circumstances and understanding changed.	Surveys/interviews Document review Literature review Context analysis Logical analysis
What results have been realised?	What outputs have been delivered? What related outcomes have been observed? What related impacts have been observed?	The different levels of results can be reliably specified and measured.	Surveys/interviews Document review Data base review Observations Monitoring data

continued ...

Table 4. (continued)

Key evaluation question	Related evaluation questions	Underlying assumptions and requirements	Suitable tools, methods and designs
Has the intervention made a difference?	<p>Was the intervention likely a contributory cause?</p> <p>What role did the intervention play?</p>	<p>There are several relevant causal factors that need to be disentangled.</p> <p>Interventions are just one part of a causal package.</p> <p>Supporting factors can be identified.</p>	<p>Experimental and statistical designs</p> <p>Theory-based evaluation designs, e.g. contribution analysis</p> <p>Case-based comparable designs, e.g. qualitative comparative analysis</p>
How and why has the intervention made a difference?	<p>How have the impacts come about?</p> <p>For whom has the intervention made a difference?</p> <p>Has the intervention resulted in any unintended impacts?^a</p>	<p>Interventions interact with other causal factors.</p> <p>An adequate theory of change for the intervention can be developed.</p> <p>Supporting factors can be identified.</p> <p>There is understanding of how supporting and contextual factors connect intervention with effects.</p>	<p>Theory-based evaluation designs, e.g. 'realist' approaches and contribution analysis</p> <p>Participatory approaches</p> <p>Case studies</p>
Will the intervention continue to work?	<p>Is the intervention and its benefits sustainable?</p> <p>What are the future estimated benefits from the intervention?</p>	<p>The benefits from the intervention will continue to be realised.</p> <p>Future benefits can be reliably estimated.</p>	<p>Scenario approaches</p>
Will the intervention or elements of it work elsewhere?	<p>Can this intervention as a 'pilot' be transferred elsewhere and scaled up?</p> <p>Is the intervention sustainable?</p> <p>What generalisable lessons have we learned about impact?</p>	<p>What has worked in one place can work somewhere else.</p> <p>There is generic understanding of contexts, e.g. typologies of context.</p> <p>Innovation diffusion mechanisms exist.</p>	<p>Participatory approaches</p> <p>Natural experiments</p> <p>Synthesis studies</p> <p>Scenario studies</p>

^a Agriculture research has devoted some attention to spillover effects from such research; that is, the effects of the research beyond the target area or group; see, for example, Deb and Bantilan (2006). Methods for addressing these spillover effects are not covered in this report.

Relating program attributes to methods and designs

In the introduction to this report, we set out the key characteristics of NRMR programs as:

- using a mix of types of intervention strategies
- involving a number of stakeholders
- having a focus on natural resource management issues
- operating at farm, landscape, regional and global levels.

The diverse attributes of NRMR programs have been mentioned in the foregoing discussion. They can be summarised as follows:

1. *Complex system.* Complex ecosystem interactions mediating social and ecological systems relationships.
2. *Market failures.* Frequent absence of market-based coordination of activities around the use (and conflict resolution in that use) of natural resources.
3. *Multiple stakeholders.* Multi-stakeholder participation and coordinated action in socio-ecological systems.
4. *Multilevel.* Operating at several levels (farm, landscape, regional and global)—often quite localised interventions are seen as contributing to more ambitious goals at a higher system level.
5. *Uncertain, multilevel and lengthy trajectories for impact.* A long and uncertain developmental trajectory. As well, market variables can change very rapidly while landscape variables usually change over decades.

6. *Systems integration.* Interconnectedness and integration among different fields of knowledge such as farm productivity, institutional innovation and environmental concerns—between which there is often a trade-off.
7. *Contextualised knowledge.* A high level of contextualisation—the specific context and history matter.
8. *Emerging outcomes.* Outcomes that are unpredictable and subject to change.
9. *Uncertain knowledge.* Operates in areas of limited/little previous or reliable knowledge.
10. *Institutional concerns.* Impacts that are often institutional—such as governance and markets.

All provide challenges to both managing and evaluating NRMR programs and must be taken into account when developing an evaluation plan. Table 5 outlines and discusses possible implications for evaluation designs to deal with these attributes.

Table 5. Natural resource management research (NRMR) program attributes and designs

Attributes	Evaluation challenge	Design implications
Complex ecosystem interactions mediating social and ecological systems relationship	Traditional (non-NRMR) evaluations can often simplify the role of ecosystems in defining the impact of particular research. In the case of NRMR, however, these ecosystem interactions are likely to be crucial to: the means by which the research has an impact; the nature of that impact; the magnitude of the impact; the causality involved; and the stability (or longevity) of the impact. Ecosystems are often subject to complex, nonlinear and threshold-driven responses to particular interventions.	This has substantive implications for: the theory of change underlying the evaluation; the understanding of causality in the system (even the conventional ‘counterfactual’ approach becomes more complex here); the nature of data collections; and the role that explicit analysis of uncertainty needs to play in the evaluation. One overriding challenge will be to incorporate the scientific knowledge of many relevant disciplines in the evaluation process.
Frequent absence of market-based coordination of activities around the use (and conflict resolution in that use) of natural resources	In traditional evaluations, market prices often form the starting point for estimating value. The absence of markets (and in some cases associated property rights) provides a challenge to valuation and the processes by which research outputs are adopted, since market prices are a common signal of adoption in many other forms of research.	Evaluation design needs to account for the ways in which property rights over resources have been traditionally defined and the associated ‘institutions’ that mediated resource use in the communities affected. Put another way, NRMR will take place within an existing, complex dynamic of methods for resolving resource use issues. A range of different forms of data collection will be needed. Participatory approaches, and understandings of collective responses may become relatively more important. ^a
Multi-stakeholder participation and coordinated action in socioecological systems	Multiple stakeholders and beneficiaries need to coordinate their behaviours and policies to implement programs and to sustain impacts in socioecological systems. The processes of achieving collective action as well as the outcomes need to be evaluated.	The evaluation will require inputs from beneficiaries and stakeholders. Methods that evaluate collective action are also needed—probably (following Poteete et al. 2010) focusing on trust, informal relationships, networks, incentives, information and ownership. The challenge will be to link these processes to the sustainability of non-material outcomes such as new forms of governance and their value for conflict resolution.
Multilevel (operating at farm, landscape, regional and global level)	In multilevel programs with socioecological interactions across scales, the outcomes and impacts at each level have to be evaluated with appropriate methods for that level as well as aggregating for global level impacts.	A ‘nested’ design deploying methods appropriate to each level will be needed. For example, this could include different theories of change at different levels; a comparative or experimental design at farm-level; comparative case-studies at landscape level; and a statistical analysis at regional and global levels. Understanding the links between these different levels may require a further set of ‘systems’ designs, including modelling.

continued ...

Table 5. (continued)

Attributes	Evaluation challenge	Design implications
Uncertain, variable and interacting trajectories for impact	Due to the interaction between social and ecological systems, NRMR programs deal with huge variations in the impact trajectories of the systems they engage in. Furthermore, implementation trajectory changes need to be tracked rather than assessed at a single moment in time.	Tracking change over time is likely to require non-standard monitoring and evaluation approaches. These could include longitudinal methods; e.g. longitudinal case studies, panels, time series data etc. There will also need to be opportunities to revise initially formulated theories of change.
Systems integration required for resilience and sustainability (related to 4 and 5)	NRMR programs often combine research on genetic technologies and farming systems/institutions with assessments of environmental and livelihood consequences. The success of NRMR is often understood as trade-offs between production, environmental and social effects. For sustainability, a holistic approach is required to see the longer term impacts for resilience and sustainability.	A balanced evaluation will need to assess how all the elements are combined—there is a tendency to focus on one element only. Framing in terms of ‘innovation systems’ may be appropriate; so too will be methods and models that assess trade-offs and can provide holistic understanding.
Contextualised knowledge is vital	NRMR programs are often place-based, focusing on a particular ecosystem and population interacting with it. Different ‘starting-conditions’ will shape the implementation and potential results of programs. Contextual characteristics may also include history of previous initiatives. Challenges arise in evaluating how generalisable and replicable the program is.	Even though contexts are not standardised they are likely to fall into certain types. Contexts should therefore be clustered into typologies to achieve limited generalisation—a strength of using ‘realist’ evaluation approaches. This also implies building a comparative element into program selection and design. When the elicitation of local knowledge is critical, assessing the elicitation process and how this knowledge informs design and implementation will be important. Knowledge elicitation usually depends on participatory engagement and model development (as for expert systems). Local histories will be useful for identifying previous related initiatives and endogenous developments.
Unpredictability and emergent outcomes (related to 6)	The complex interactions of social and ecological systems in NRMR mean that outcomes cannot be predicted. The challenge is to be able to capture the unexpected outcomes and impact.	For elements of interventions where this is the case, designs built on developmental approaches (Patton 2011) and use of real-time evaluation with frequent feedback are needed to learn what is happening.

continued ...

Table 5. (continued)

Attributes	Evaluation challenge	Design implications
Operates in areas of limited/little previous or reliable knowledge	NRMR research programs operate on scientific frontiers. New knowledge is an important output of NRMR and is equally important to make 'impact' more likely.	Baseline efforts to systematise existing knowledge and 'knowledge in use' should be followed-through with tracing the use of new knowledge in practice by different stakeholders. The evolving knowledge base partly explains why not all decisions about evaluation design can be taken at the outset, reinforcing the need for an iterative or staged evaluation design.
Institutional concerns	Changes are expected not only in individuals but also in institutions.	Include institutions relevant to system change from the outset. Pay particular attention to barriers to sustainability and conduct repeat case studies at critical junctures in the implementation process.

^a Pearce and White (2012) discuss incorporating environmental and natural resource values within the Australian Centre for International Agricultural Research's economic impact assessments.

Application to existing CGIAR research programs

In this section a number of current CGIAR NRMR programs are reviewed to further illustrate how the approaches and design principles proposed in this report could be applied in practice.

Exploring a broad NRMR program: the CGIAR Research Program on Aquatic Agricultural Systems (CGIAR Research Program 1.3)

The Aquatic Agricultural Systems (AAS) case provides a good basis to consider how to disaggregate large ‘multi-hub’ programs, and how a theory of change approach can be applied in a case where some selectivity is unavoidable given the potential scope of an impact evaluation. This case also explores how the ‘attributes’ of a program have direct implications for methodological design. Quotes are from the WorldFish Centre (2011) proposal.

The AAS goal is:

... through our partnerships approach and targeted investment, the Program seeks to improve the lives of 15 million poor and vulnerable people over the next 6 years. By further expanding and disseminating the learning derived from this effort, we expect to increase that number to 50 million by 2022.

Two pathways are being employed to do this. The first is:

Multi-dimensional, integrated programs that combine social, economic, scientific and environmental ‘themes’²², drawing on state of the art scientific and social scientific knowledge through inclusive

processes of participation and dialogue with farming communities, partners and policy makers, will be able to identify and embed innovative platforms²³ that will support sustained improvements in productivity, income, nutrition and environmental outcomes for local, regional and global populations.

The second pathway is to influence how others do research *in* development:

Pursuing our work in this way will challenge the CGIAR to move beyond traditional circles and change the way we do much of our research. By emphasizing approaches that call for research in development – rather than research and development or research for development – we will pursue a conscious change in emphasis and mind set, one that can help the CGIAR to conceive and deliver our research differently. We therefore envisage the Program as an exemplary vehicle for implementing the fundamental changes in ways of working that the CGIAR reform process foreshadowed and the Global Conference on Agricultural Research for Development (GCARD) has endorsed.

The program is a collection of activities being undertaken in several countries, with common high-level goals. The problems addressed in each country, and within each hub in each country, vary, as do the types and nature of activities undertaken.

Because of the spatial and temporal complexity of this program, the AAS should be evaluated through a series of disaggregated monitoring and evaluation activities, resulting in a set of nested impact evaluations. Then, if

²² See WorldFish Center (2011, p. 25)

²³ Contextualised knowledge, innovative practices and new management and institutional arrangements

required, it may be possible to draw overall conclusions about the program and its intent.

As previously discussed, there are several ways to disaggregate a AAS program that might be fruitful to explore as outlined in the earlier section ‘The unit of analysis: what to evaluate?’ (see Box 1 on p. 17). The obvious disaggregation is to undertake separate evaluations of the activities in each of the five countries, further disaggregated into evaluations of each hub. These could be done at similar times, and some synthesis of findings carried out. The designs used would reflect the attributes of each hub or country subprogram.

In a spatially, or otherwise differentiated, program, more than one theory of change will be needed, even though an overall theory of change can stand as a template for other ‘nested’ theories of change and to identify the links between different parts of the program; for example, how farm-level innovation is scaled up and out to wider partnership settings.

In the AAS program there are three ‘levels’:

- local and national (hubs) which is where research takes place with farming communities
- regional (Learning Alliance)
- global (Global Coalition).

As well as an overarching theory of change —perhaps along the lines of the generic theory of change illustrated in Figure 5—specific theory of change will probably be needed at each of these levels. There could be even further useful disaggregation by specific target groups, such as for the local hubs farmers/fishers, households and communities.

A similar disaggregation would be by type of outcome produced, such things in the AAS Program as:

- increased crop productivity/AAS technologies
- fisheries, crop and land management
- community development (involvement, diversification, action research)
- governance structures for improved policies and institutions
- emerging markets.

Then, within a hub, within a country or, most interestingly, across the AAS Program, one probably could aggregate

along these lines, identify which types of activities aimed at the different target groups work (or not) and why. In each case, a theory of change could be developed to set out and understand how program activities delivering outputs connect to (influence) immediate and intermediate outcomes and, ultimately, impacts.

However, in a program as large and complex as AAS, there will need to be selectivity in what can be evaluated and which theories of change can be developed, even within a more focused evaluation that emphasises ‘impacts.’ Such selectivity requires criteria setting and prioritisation, which will need to draw on both bottom-up and top-down priorities. For example, bottom-up priorities could include:

- confirming that income-generating activities are sustainable
- identifying capacity needs for more effective governance of a water basin
- confirming that new seeds are able to resist several cycles of drought or flood
- understanding how to ensure women continue to have a voice once initial research-led innovation is over.

Top-down priorities could include:

- demonstrating that participatory action-research makes a contribution to longer term outcomes or impacts
- confirming that coalition and alliance arrangements are leading to benefits for large numbers of poor farmers
- demonstrating greater resilience for communities affected by climate change
- confirming that new NRM-related tools and data bases are useful in other CGIAR Research Programs or elsewhere in CGIAR research.

Clearly, the particular priorities of an impact evaluation need to be clarified and clustered together alongside the preparation of theories of change. The importance of prioritisation means that it needs to be built into the planning and feasibility stages of programming. Among the criteria that can be useful in this regard are:

- priorities of farmers, non-government organisations and government policymakers

- recent crises and disasters that highlight urgent needs
- new scientific knowledge or applications developed for use elsewhere that appear to be useful
- strategic choices for policymakers at local, country and regional levels
- areas where high-cost investments are being contemplated and alternative options are on the table.

Particular designs for the AAS Program would need to bring together program goals and activities with suitable evaluation questions and designs. Table 6 provides one example of how these pieces of the puzzle might fit together in the program.

Exploring a case: the Ganges Basin Development Challenge

The case of the Ganges Basin Development Challenge (GBDC) is considered in terms of how policy advocacy and engagement with stakeholders and beneficiaries can be evaluated in a multipartner setting. The case also describes the different ways in which theories of change can be used and in which different evaluation questions can be formulated.

GBDC, now part of the Khulna hub in the Aquatic Agriculture Systems Program, is a project to improve resource productivity and increase the resilience of agriculture and aquaculture systems in brackish coastal areas of the Ganges. Phase II of the program will run over the period 2011–14.

Phase II is a multi-institutional (there are 17 partner institutions involved) and interdisciplinary research-for-development initiative focused on increasing the resilience of social and ecological systems through better water management for food production. The challenge is seen as multidimensional, needing to embrace issues of agriculture productivity, farm and fishing practices, water management, adequate markets and supportive government policies and regulation.

GBDC has five component projects (G1–G5) directed at different aspects of the challenge:

- G1: To establish a geo-referenced database for the coastal zone of Bangladesh and to facilitate

out-scaling of technologies through identification of target domains and land-use planning

- G2: To develop and introduce resilient agriculture/aquaculture production systems in the coastal zone of Bangladesh and India for the benefit of poor households
- G3: To improve water governance and management for resilient production systems
- G4: To assess the impact of anticipated external hydrological changes on water resources in the coastal zone of the Ganges
- G5: To enhance impacts in Bangladesh and India through stakeholder participation, policy dialogue and effective coordination among other government, NGO, CGIAR and donor-sponsored projects and programs in the Ganges BDC research program.

The Annual Basin Report (CPWF 2012, p. 4) notes:

The goal of the Challenge is to reduce poverty and improve livelihood resilience by better utilizing the brackish water of the coastal Ganges – through improved water governance and management and by intensifying and diversifying agricultural and aquaculture systems.

GBDC involves research on land use, seed productivity and related seed producers and microfinance agencies, water-use governance and hydrology scenarios. It also involves a variety of engagement activities aimed at influencing the behaviour of key actors in the relevant social and ecological system, using the research knowledge acquired. Thus, while some of the research knowledge created is disseminated through the literature to the wider agriculture development community, the main effort of GBDC is in using this knowledge to advocate for changes in the actions of its partners.

Monitoring and evaluation of GBDC

Evaluation of the impacts of engagement is a considerable challenge. The activities undertaken are multiple and often responsive to changing circumstances, they are ongoing throughout the period of a project and they are trying to influence behaviour in a direction different from what the actors involved are undertaking. The actors involved typically are faced with a variety of influences in addition to those being advanced. The engagement efforts are clearly only contributory causes to any changes in behaviour that come about.

Table 6. Evaluation designs tied to questions and attributes: possible application to the case of CGIAR Research Program 1.3: Aquatic Agricultural Systems

Program focus	Activity/intervention	Expected results	Evaluation questions	Overarching designs	Possible approaches
Contextualising scientific knowledge	<ul style="list-style-type: none"> Village meetings Local surveys Farm schools Distributing handouts and videos 	<ul style="list-style-type: none"> Knowledge is disseminated, understood and used in local settings Integration of new practices with local knowledge and practices 	Which methods work best in what settings?	<ul style="list-style-type: none"> Experimental Case studies 	Comparative experimental and quasi-experimental designs building on observation, monitoring and case study data
Mobilising local communities and farmers	<ul style="list-style-type: none"> Eliciting local knowledge and experience Consultation and participatory action research Setting up new networks Identifying cooperation/collective governance needs 	<ul style="list-style-type: none"> Identification of problems that are taken seriously Commitment to specific pilots and interventions New forms of governance are identified and appear sustainable 	What difference does participation make to production, environmental and institutional outcomes?	<ul style="list-style-type: none"> Theory of change Participatory Case studies 	<ul style="list-style-type: none"> Participatory case studies Comparing communities with different community profiles Interviews and focus groups Baseline data and follow-up surveys
Environmental benefits and trade-offs	<ul style="list-style-type: none"> Developing new farm practices Developing new monitoring arrangements Developing new water-basin governance Developing new environment and natural resource services Capacity building for local farmers and intermediaries 	<ul style="list-style-type: none"> Less fertiliser use Adaptation to floods and droughts Use of services and data Sustainable governance arrangements 	What have been the environmental consequences of the program and what potential is there or scaling up and out?	<ul style="list-style-type: none"> Statistical 	<ul style="list-style-type: none"> Economic valuation of costs and benefits Participatory scenario building Trade-off models Geospatial databases and environmental indicator monitoring
Gender equality and social inclusion	<ul style="list-style-type: none"> Enhancing women's role in decision-making Gender mainstreaming Livelihood assessments Nutritional assessments Engagement with the poorest in focal communities 	<ul style="list-style-type: none"> Women's access and use of new technologies Voice of poorest present in innovative programs Improved family nutrition and health 	How has greater social inclusion contributed to intended technological, environmental and human outcomes?	<ul style="list-style-type: none"> Statistical Theory of change Participatory Case studies 	<ul style="list-style-type: none"> Longitudinal gender mapping Disaggregating indicator data by gender Participatory feedback and assessment Thematic workshops on livelihoods and ecology facilitated by local non-government organisations

Further, even if the engagement can be shown to have contributed to changes in practices (such as promoting supportive policies and regulations, creating a market for new seeds, undertaking more evidence-based land-use planning, or putting into practice better governance structures for community water management), there is still a long path to the ultimate impacts sought—improving the livelihood of the poor in the target areas. Nevertheless, showing that the activities of the projects have indeed influenced targeted actions does indicate progress and that the projects are making a difference. And understanding which engagement strategies worked best and which not so well, is valuable information for improving the effectiveness of GBDC, and possibly provides lessons for other similar efforts.

There are thus several components of impact evaluation that can be considered:

- impact evaluation during the time period of the challenge, assessing the results of the engagement activities undertaken and the usefulness of the new knowledge created
- impact evaluation after the challenge efforts have been completed, to assess the extent to which behavioural changes (practices) that the challenge has contributed to have led to impacts on the broader social and ecological systems and, ultimately, on the livelihoods of the target population.

Evaluating engagement

While a significant challenge, evaluating different sorts of engagement activities has some basis to work from. In particular, evaluating advocacy has a large and growing literature; see, for example, Patton (2008), (Coffman 2009), Mansfield (2010) and Innovation Network (n.d.). There is a companion literature on what works in advocacy and hence on many engagement activities; see, for example, Ringsing and Leeuwis (2008) and Darnton (2008). This report does not review either literature, but we do note some key aspects:

Using theories of change. Evaluating engagement (advocacy) typically involves the development of and use of a theory of change to represent the engagement expectations. The GBDC projects have all developed outcome pathways focussing in particular on the actors they are trying to influence and the practice changes

they expect to bring about. This familiarity with outcome pathways can be built on.

A focus on specific partners. Evaluations could be undertaken for each of the GBDC projects. Each is well defined with a leader and a team. But that might not be the most useful focus for trying to better understand the results these efforts are bringing about. Engagement efforts target specific actors whom they are trying to influence; the GBDC projects clearly identify these actors.

Theories of change could be built for each key partner group, capturing all the efforts directed towards these groups across the five projects. Then exploring these theories of change would allow for an assessment of how different strategies for influencing them have worked. Of course, at the higher levels of results, these different theories of change would intersect, since the expectation is that changes in practices of the various target groups are all needed to bring about impacts.

What is probably needed is more attention to the assumptions behind the outcome pathways that would turn them into more robust theories of change. For engagement to work, some of the assumptions that are likely needed might include that:

- the engagement actions reach the right people—people who have the power to act or significantly influence those who can act
- their messages are understood (clear, logical, sound, repeated etc.), are sensible given the circumstances, are timely in hitting windows of opportunity and are actionable
- follow-up assistance is provided
- other key actors carry out their roles.

The theory of change should set out the logical sequence of steps (immediate, interim and final) between the activities and the observed practice or further impact changes, identifying the other contextual factors that are needed for the links in the sequence to work. It is on the basis of the theory of change that a claim about contribution can be made. There are two steps (Stern et al. 2012, p. 42):

The steps in this causal chain are shown to have happened *and explained* as to why they happened. This would include identification and discussion of contextual factors and other contributing causes that brought about each step in the sequence.

Plausible rival explanations of why each step in the causal chain happened are identified and discussed as to their relative significance in bringing about the impact.

Not all engagement efforts will work. It is useful to identify breakdowns in the expected sequence of impacts. Their analysis will help understand why things didn't work out as expected and what could be done differently next time.

Evidence would be gathered from:

- observations about what happened in terms of the sequence of impacts
- confirmation that the pathway unfolded as expected and that the other contextual factors indeed happened (or did not)
- surveys/interviews to understand why the steps are thought to have occurred, and if there are significant alternative explanations.

Other theory of change perspectives. An additional perspective for using theory of change might be to look at the kinds of strategies or campaigns undertaken. If across the projects, perhaps led by project G5, there had been planned efforts to try to bring about a specific change using a number of the GBDC activities aimed at getting several actors to work together to bring about change, then that engagement campaign could be looked at in a similar fashion to looking at theories of change for specific target groups. This would be a more holistic approach focused on some more specific change sought.

A still different cut would be to look at the GBDC efforts across the projects to bring about key intermediate outcomes, such as supporting policy, markets for seed new varieties or new community polder governance arrangements.

To select among these approaches—to select useful units of analysis—would require in-depth knowledge of the GBDC project and its history, an appreciation of the challenges in collecting the data needed, and an understanding of which perspectives should provide the most useful information for improving the GBDC and its projects.

A need for active monitoring. Given the nature of engagement, it is important that deliberate efforts be taken to track just which actions have been used with what immediate reactions in terms of attitudes,

skills, understanding and practices. This is the kind of monitoring already contemplated for the GBDC projects and evident in the outcome logic models. Those involved need to be diligent in recording their efforts, and the context in which they were taken—such things as the apparent a-priori beliefs of the target groups, the constraints they face, media attention etc. The approaches to evaluation being discussed are in part akin to outcome mapping where there is active monitoring of changes in partners attitudes, skills and behaviours. Monitoring can also be done to see if any follow-on results have occurred or significant events have occurred outside the program.

Using cases. The different engagement 'campaigns' used can be candidates for a variety of case studies. 'Most significant change' cases might be sought to understand why efforts seemed to work (Davies and Dart 2005). 'Episode studies' (Hovland 2007; ODI n.d.) seek to understand what led to a policy change, then assess the relative role of research and evidence. Selective case studies can be useful to gain a better understanding of what strategies and which theories of change are or are not working.

Case studies can also be useful to get at the contribution made to follow-on results after practice changes have been observed. They can be used to support or construct/revise theories of change linking outcomes to longer term impacts. Cases can be based on examples where desired impacts occurred (or clearly did not) and trace back to explain why.

Box 3 summarises the discussion above on the GBDC monitoring and evaluation.

GBDC evaluation questions

Based on the above discussion, key evaluation questions for GBDC would be:

Rationale

- Are these five projects still seen as the 'right' five?
- Do the projects still make sense?
- Are there other actions that now seem to be needed to make things work?

Implementation

- Are the projects well coordinated? How have they in fact linked and built on each other?

- How well have the projects worked with other partners?
- How well have they learned? And learned together?

Box 3. Strengthening the Ganges Basin Development Challenge monitoring and evaluation

- Review the literature on evaluating advocacy.
- Consider the usefulness of different units of analysis: partners, types of engagement activities, practice changes aimed for.
- Build theories of change (mature impact pathways) for the various types of engagement activities undertaken.
- Put in place a robust monitoring system to track key aspects of the theories of change.
- Over time, assess the extent to which these impact pathways and supporting factors have been realised and the engagement efforts effective.
- Consider undertaking specific case studies.

Does it work?

- What results have been observed—intended and unintended? Are there concrete actions by the various actors that can be identified?
- Did the projects make a difference? Have they really influenced anybody of import to do anything concrete and useful?

Why and how did it work?

- How have the projects made a difference? Which strategies worked best? (These would be strategies on engagement etc.) Why have they worked or not worked?

Will it continue to work?

- How likely are future outcomes and impacts? (This would be based on the interest shown by the various actors, and the efforts they have undertaken to move the agenda forward.)

Will it work elsewhere?

- What generalisable lessons have been learned? (Lessons, if any, would mainly be about engagement and what works.)

Realistic expectations

The GBDC is perhaps an example where expectations appear to be unrealistically high. Linking improved wellbeing of millions of households to the activities of the GBDC would be extremely difficult to do in a credible manner. The causal links, even as contributions, are very long both in terms of the sequence of events needed to establish a credible link and the time frames involved. The high-level goals are quite useful directional goals to be aiming for, but may stretch credibility for many as explicit targets to be reported against.

GBDC's 2012 Annual Report (CPWF 2012, p. 13) appears to recognise the challenges involved:

There is little political will to ensure or promote success even though issues of agricultural and aquatic systems are included in formal planning documents. There is insufficient cohesion between investments of donors. Farmer fatigue is starting to feed into the equation with so many projects directed to the same areas. Good quality NGOs to work with in social mobilisation are scarce, and those that are good are overstretched.

The report goes on to discuss the challenges associated with each project.

Rather, one would expect that GBDC would be able to show that it made a positive difference in the behaviour of its target groups, and would be able to show that it has learned in the process. This would include showing that it reached the targets set with respect to its various target groups and has been modifying how the projects are being implemented as a result.

Exploring a case: the African Food Security Initiative (AFSI)

This case allows different evaluation questions to be considered in a program where good-quality research is seen as key indicators of success.

To foster a long-term sustainable improvement in African food security, the Australian Government, through the Australian Agency for International Development, has increased its investment into Africa via the Africa Food Security Initiative (AFSI). AFSI is focused on lifting food security and agricultural productivity in Africa through joint research, working with and building the capacity of African agricultural organisations, and by enhancing community resilience. The agricultural productivity component of AFSI leverages Australia's unique agricultural and scientific expertise by engaging CSIRO to help African research institutes and farmers meet their national food security challenges.

AFSI's two agricultural components are the:

- Biosciences Eastern and Central Africa (BecA) Hub – CSIRO partnership
- West and Central African Council for Agricultural Research and Development (CORAF/WECARD) – CSIRO partnership.

The BecA Hub is a shared agricultural research and biosciences platform located at the Nairobi campus of the International Livestock Research Institute (ILRI) in Kenya, through which ILRI shares its recently upgraded and expanded research facilities with a broad range of new national, regional and international partners through the BecA initiative.

The BecA partnership focuses on *building African capacity* in biosciences to solve Africa's agricultural, nutritional, food safety and environmental problems. It is providing support in order to:

- increase agricultural productivity
- improve the functioning of markets in ways that increase incomes and reduce risks for the rural poor
- strengthen social protection systems to protect the vulnerable from shocks and severe deprivation.

CORAF/WECARD will focus on *increasing crop and livestock productivity* through more efficient water and nutrient use and management, improving animal disease management and increasing services for smallholder farmers via more effective agriculture input and output markets.

While both of these partnerships are aiming at enhancing food security, their stated goals are much

more modest and quite realistic: building African biosciences capacity and improving crop productivity—fairly standard CGIAR goals.

Independent mid-term evaluations of each of these partnerships are planned for 2012–13.

Evaluating BecA

For BecA, it is stated (BecA Hub and CSIRO 2010, p. 30) that:

... the Partnership will be judged as immediately successful if relevant, good quality research is being conducted in and for a range of countries and addressing a range of priority research questions, with inputs from both African and Australian researchers. It will ultimately be judged a success beyond the project term if that research leads to the adoption of new biosciences technologies and strategies that improve livelihoods of Africans.

A detailed monitoring and evaluation (M&E) plan has been developed for assessing the partnership, with an independent evaluation expected to start in 2012.

The evaluation questions of whether quality African research and capacity are being achieved are conceptually reasonably straightforward and important to project assessment. The M&E plan also has considerable focus on learning about the partnerships and how efforts have been implemented. But also of interest would be the rationale and 'how' questions:

- Are the current modes of capacity development still seen as the best approaches?
- Have different types of research projects been better able to build specific types of capacity and in whom?
- Which delivery strategies have been most successful?

An independent assessment of the rationale questions could be quite useful. There is a large literature on evaluating capacity building that would be useful to review and build on.

On the other hand, assessing the extent to which the project 'leads to the adoption of new biosciences technologies and strategies that improve livelihoods of Africans' is considerably more problematic. Indeed, given the time frame likely involved, the modest size

of the intervention and the multitude of other factors, events and conditions at play, may not be worth pursuing. The underlying model or theory is that, as is the case elsewhere around the world, building African biosciences capacity is a useful undertaking that will, over time, lead to benefits.

In any event, if such evaluation questions are to be posed, it might be better to consider undertaking a joint evaluation involving all the donors with related activities.

Evaluating CORAF/WECARD

The CORAF/WECARD partnership has developed a similarly detailed M&E plan. Impact pathways are being developed for each project and an independent evaluation planned. Various learning mechanisms are underway or planned.

As for BecA, it could be quite useful to undertake an independent assessment of the rationale questions about the design and implementation of the partnership, given its aim of meeting longer term goals of food security. Questions such as:

- are the right projects being undertaken?
- are the right actors being engaged?
- are the engagement activities based on previous research?

The CORAF/WECARD partnership is described as being part of other related activities, such as those seeking to strengthen infrastructure funded in part by the UK Department for International Development and the United States Agency for International Development. The extent to which the partnership efforts are seen as fitting into the overall agenda is important.

The description of the partnership sets out a number of strategies that are to be adopted in implementing the initiative. They provide good questions for an examination of implementation.

The partnership is aiming to influence all actors along the commodity food chain, from farmers to planners to markets.

It would be useful if the project-to-impact pathways that are being developed included the elements

discussed earlier about useful theories of change, such as identifying and then tracking assumptions, risks and unintended results. This would allow for stronger causal claims to be made about the effects of the projects. It might also be useful, where appropriate, to consider developing pathways for influencing various key research partners that cut across the various projects. Such theories of change would allow for an examination of the different strategies being used to interact with and influence these key partners. Specific case studies could usefully complement these efforts, allowing for an in-depth consideration of efforts to influence a partner.

And again, the longer term impacts of the partnership might best be assessed as part of a larger joint evaluation.

Finally, the CORAF/WECARD partnership is aimed at building African research capacity. There is a large literature on evaluating capacity development that would be useful to explore in designing evaluations and monitoring; see, for example, Horton (2002), Mackay et al. (2002), Larbi et al. (2005) and Huyse et al. (2012).

Box 4 summarises suggestions for enhancing the AFSI M&E plans.

Box 4. Strengthening the African Food Security Initiative monitoring and evaluation

- Consider evaluating key rationale questions.
- In assessing capacity building efforts, consider the literature on evaluating capacity building.
- To evaluate the longer term impacts, consider a joint evaluation involving all key stakeholders.
- Ensure that the project-to-impact pathways being built incorporate relevant assumptions and risks, i.e. are good theories of change.
- Ensure that the monitoring systems in place support the assessment of the pathways developed.
- Consider undertaking specific case studies of capacity building and engagement efforts.

A general evaluation framework

An *evaluation framework* sets out an overall operational plan for an evaluation. It would indicate:

- the main evaluation priorities and evaluation questions
- specific evaluative activities (data collection, analysis, drawing conclusions and recommendations, reporting etc.) and when these should take place
- the program's theoretical assumptions
- the NRMR priorities of the program
- how results (outputs, outcomes and impacts) will be addressed and explained
- the division of labour between evaluators, managers, beneficiaries/those implicated, other stakeholders
- a quality assurance and ethical set of standards and procedures.

Preparing such a framework ensures that the decisions necessary to operationalise the principles discussed in this report are consistently followed. Given the nature of NRMR programs, the plan should be developed participatively, involving stakeholders, beneficiaries and others implicated, be validated by stakeholders, and offer flexibility for revision and redirection.

The planning and design activities that are usually needed to prepare an evaluation framework can be grouped under four headings:

- Clarifying the evaluation purposes
- Identifying program characteristics
- Elaborating an initial theory of change
- Reviewing data availability and quality.

Clarifying evaluation purposes

- Review the strategic interests of program stakeholders including beneficiaries and program sponsors.
- Consider the purposes and use of the evaluation and who the users will be.
- Identify the main evaluation questions that program implementers, beneficiaries and other stakeholders are interested in answers to. These are likely to cover both impacts (i.e. intended program results) and related implementation processes.
- Clarify the balance and priority to be given to the impact aspects of the evaluation; that is, causal and explanatory dimensions. (It is assumed that even in an impact evaluation there will be some evaluative activities that are not explicitly impact-related. They may, for example, be descriptive or strategic, such as preparing for change.)

Identifying program characteristics

- Review the distinctive NRMR aspects of a program and where appropriate, propose which aspects should be prioritised.
- Assess the attributes and priorities of the program concerned and consider the implications this has for evaluation strategy, methodology and data access.
- Identify and map overlapping or related programs (for example, those that have related goals, and affect the same target groups and territories).

- Identify possible losers as well as beneficiaries.
- Conduct an ethical assessment of the evaluation—confidentiality risks; effects for the less powerful; perverse incentives and moral hazards; feedback obligations; how stakeholders and others implicated will have voice.

Elaborating an initial theory of change

- Posit an initial implementation and outcome trajectory in terms of ‘shape’ (speed and extent) and time.
- Decide on an appropriate ‘time-slicing’ of the evaluative activities (what happens when) paying special attention to the first stages of an evaluation and the first iteration of activities that will be needed.
- Develop an initial overarching theory of change. This should draw on assumptions and goals of

stakeholders/program implementers/ beneficiaries and others; feasibility and planning data; and other related experience—published sources, practitioner experience, other evaluations etc.

- Attempt a first-round outline of the main ‘nested’ evaluation elements at different system levels and those elements that link different levels.

Reviewing data availability and quality

- Review available data sources, paying particular attention to data gaps and weaknesses.
- Design monitoring systems that will track change and fill in data gaps identified.
- Specify a quality assurance plan that will ensure evaluator independence, ethical monitoring, data quality and methodological rigour.

Summary recommendations for improving NRMR evaluation

In considering the evaluation of NRMR programs, this paper has covered a great deal of ground, presenting ideas and making suggestions about how evaluations might be undertaken. To summarise our discussion, we suggest that those evaluating NRMR programs should:

- see these complex programs/interventions as contributory causes, part of a causal package of events and conditions which together are expected to be sufficient to bring about the desired outcomes and impacts
- develop mature theories of change, with assumptions, risks and unintended effects—indeed, there is probably a need for several nested theories of change for a program, covering different perspectives and levels
- ensure that there is a robust monitoring system in place to track progress and revise the theory of change as experience and insights are gained, and to provide baseline and ongoing data for evaluation
- carefully articulate an essential set of evaluation questions of interest, including those focusing on the causal links between the NRMR program/intervention and the expected outcomes and impacts
- identify and understand the attributes of the specific intervention being evaluated and their implications for the evaluation
- develop appropriate mixed methods evaluation designs based on the evaluation issues to be addressed and the attributes of the intervention, keeping in mind the timing and resources available for the evaluation
- while keeping high-level development impacts in mind, focus on intermediate outcomes and making a link to the high-level impacts with logic and previous evidence
- build on the existing evaluation literature, particularly that on evaluating advocacy and capacity building, and the research–policy nexus.

Appendix 1. Evaluation questions for NRMR programs

This appendix provides further details on the evaluation questions that can be posed in an evaluation of NRMR programs. The general framework is shown in Box A1.

Box A1. A framework for evaluation questions

Ex ante:

1. Should it work?
2. How and why is it supposed to work?

Ex post:

1. Should it still work?
2. Has implementation worked?
3. Did it work?
4. How and why does it work?
5. Will it continue to work?
6. Will it work elsewhere?

Ex ante

Should it work?

- To what extent are the goals of the program relevant?
- Is the program design and implementation realistic and supported by current evidence and practice?
- Are there alternative strategies that should now be used?

How and why is it supposed to work?

- Is the theory of change sensible?
- Does it explain how and why the anticipated causal links are expected to work?
- Does it have some evidential basis?

Ex post

Should it still work?

- To what extent are the goals of the program still relevant?
- Does the program design and implementation continue to be realistic and supported by current evidence and practice?
- Is the theory of change still sensible?
- Are there alternative strategies that should now be used?

Has implementation worked?

- What has been learned about how the NRM program has been implemented?
 - Which aspects of the implementation process and strategy—the implementation theory—worked well and which not so well?
 - How and to what extent did implementation process and strategy change in response to new information and evolving issues?
- How has the implementation contributed to the results?

- Has participation strengthened the results attained?
- Has continuous learning and dialogue improved the outcomes and results?
- Which aspects of the theory of change worked?
 - What causal factors or mechanisms in what combination have resulted in the observed impacts?

- Can implementation lessons learned be transferred elsewhere?

Did it work?

Measuring results

- What results have been realised?
 - To date, were intended results achieved?
 - What unintended results have occurred?
 - What intermediate results were realised?
 - What was the trajectory of the results achieved?
 - What net impact was realised to date?
 - What future results might be expected?

Note the need to distinguish measuring the results to date (such as the extent of adoption or economic benefits) from results that are expected in the future.

Making causal claims about the intervention

- To what extent can a specific (net) impact be attributed to the intervention?
 - How much of the impact can be attributed to the intervention?
 - What would have happened without the intervention?

These types of causal questions are behind experimental designs trying to attribute results to interventions based on regularity and counterfactual causal perspectives and what CGIAR impact assessment addresses.

- Is it likely that the intervention made a difference?
 - Was the intervention likely a contributory cause?
 - What role did the intervention play?

These types of causal questions are behind theory-based and case designs trying to assess the contribution the intervention is making to the observed results, based on comparative and generative causal perspectives.

How and why did it work?

- How has the program made a difference? If not, why not?

Or, if the hoped-for impacts are not being realised, we want to know:

- Why have the impacts not been realised?
- Were the impacts not realised because of program failure; that is, failure of the theory behind the program, or implementation failure?

- For whom has the program made a difference?
 - Which target groups have changed behaviour?
 - Which have benefited?
- Has the intervention resulted in any unintended effects?

Will it continue to work?

- Is the program sustainable?
 - To what extent did (or will) the results from the program continue after donor funding ceased?
 - What were the major factors that influenced or are likely to influence the achievement or non-achievement of the sustainability of the program or project?
- What are the future expected results?
 - How realistic is the expected theory of change given the evidence to date?
 - Are there result trends that can be extrapolated?

Will it work elsewhere?

- Will the program or elements of it work elsewhere?
 - Can this program or elements of it be transferred elsewhere?
 - What conditions are essential for it to work elsewhere?
 - How can it be adapted to work elsewhere?
 - What are the critical contexts and mechanisms that account for the results achieved?
 - Can the participatory lessons learned be used elsewhere?
 - Can this program be scaled up?
 - What generalisable/transferability lessons have we learned about results?

Appendix 2. Example of a mixed methods statistical design evaluation

Admissible evidence in the court of development evaluation? The impact of CARE's SHOUHARDO project on child stunting in Bangladesh

L. Smith, F. Khan, T.R. Frankenberger and A.K.M.A. Wadud²⁵

Abstract

Experimental impact evaluation methods have recently emerged as a dominant force within the development effectiveness movement. Although these methods have improved understanding of what works, their “gold standard” status threatens to exclude a large body of alternative evidence. This paper evaluates the impact of CARE's SHOUHARDO project in Bangladesh, which employed a rights-based, livelihoods approach. Using a mixed-methods protocol, we find plausible evidence that the project led to an extraordinarily large reduction in child malnutrition. While offering valuable policy lessons, we illustrate how rigorous evaluation can be undertaken even without the randomization and control groups required by the experimental methods.

The evaluation in the above paper was examining the SHOUHARDO²⁵ project of CARE International, operating in Bangladesh from 2006 to 2010, on child malnutrition. The primary goals were to reduce child malnutrition, poverty and food insecurity and, serving a population of two million people, the program was carefully targeted to the most remote and vulnerable areas of the country and, within these areas, to the poorest households. Going beyond the usually direct malnutrition interventions, the SHOUHARDO project also looked at addressing structural causes of malnutrition, such as poverty, poor sanitation and recurrent natural disasters, as well as deeply entrenched

inequalities in power between economic classes and between women and men. In this light, it is similar to NRM programs that focused both on ‘direct’ research and on identifying the structural changes needed to ensure the research gets implemented, adopted and scaled up.

An exceptionally large reduction in malnutrition took place over the implementation period of the project. The evaluation questions posed were:

First, was the observed reduction in stunting actually brought about by the SHOUHARDO project's interventions? Second, if so, why? Specifically, did the addition of a suite of structural cause interventions have added impact over the direct nutrition interventions?

²⁴ Smith et al. (2012)

²⁵ Strengthening Household Ability to Respond to Development Opportunities

These are the same types of evaluation question argued for in the present report. As we do, Smith et al. argue against using only experimental designs in conducting impact evaluations and seeing evidence from such evaluations as the only legitimate evidence. Their evaluation took a different approach. They argued that experimental approaches raise ethical issues in the context of food aid and further that they have only limited usefulness in complex settings such as the SHOUHARDO project (Smith et al. 2012, p. 2).

The primary data used for the evaluation were collected from two independently-drawn random samples of households, one selected before the project began (for the baseline survey) and one near its end (for the endline survey). This design does not allow precise estimation of the magnitude of the project's impact. We demonstrate here, however, how a judicious and intelligent use of the available data, combined with key information about changes in the project's external environment over its implementation period and from project administrators, can provide rigorous, informative, and useful evidence regarding that impact.

The evaluation proceeded to (1) estimate the overall impact of the set of project interventions and (2) to explore the why question (Smith et al. 2012, p. 6):

For evaluating the overall impact of the project we use the available baseline and endline survey data, data from other national surveys, and information about what was going on in Bangladesh as a whole over the project implementation period. For investigating the “why” question, we make use of PSM [difference-in-difference propensity score matching] in conjunction with supplemental information collected from project administrators on the selection of households for participation in specific project interventions, as well as descriptive analysis of heterogeneous impacts across subgroups of project households. The results of all analyses are triangulated to reach the study's final conclusion regarding the impact of the project on stunting.

To estimate the impact on stunting they undertook three types of analysis: comparison of the change in stunting in the project area with national data, comparison of

the age trajectory of the stunting prevalence among project children compared to children living in rural households nationally, and an assessment of ‘trends in the underlying determinants of child malnutrition— food security, quality of caring practices for children and women, and household health environment quality ... —over the life of the project’ (p. 7).

To explore why the project interventions impacted on stunting, they investigated the impacts of specific interventions ‘... using PSM to create comparable-on-observables control groups for each intervention from among households that did not participate to serve as the counterfactual’ (p. 7). Households did not all fully participate in the maternal and child health and nutrition interventions, giving some variation that could be analysed. Other interventions that were implemented at the village or regional level were examined using more basic comparative approaches.

Smith et al. conclude that:

This paper has presented a body of plausible evidence that the SHOUHARDO project—the first large-scale nutrition-oriented project using the rights-based, livelihoods approach—had an extraordinarily large impact on child malnutrition. At the same time it has given some insight into why the project had such an impact, which is increasingly considered crucial information for rendering evaluations policy relevant and judging their external validity.

Where does this fit in the report discussion of impact evaluation approaches? It was noted that the project being evaluated had a number of the characteristics of the complex NRMR programs being discussed, and that the impossibility and limited applicability of using experimental designs especially to explore the why questions, was recognised. The evaluation did not, however, use a theory-based approach but rather undertook a variety of comparative analyses, all of which were triangulated to reach credible conclusions.

In terms of Table 2 in this report, the approach used was statistical, made possible by the extensive baseline and endline surveys and the nature of the key interventions.

References

- Auditor General of Canada 2002. Modernizing accountability in the public sector, Chapter 9 in the report of the Auditor General of Canada to the House of Commons, Ottawa. Accessible at [http://www.oag-bvg.gc.ca/domino/reports.nsf/html/20021209ce.html/\\$file/20021209ce.pdf](http://www.oag-bvg.gc.ca/domino/reports.nsf/html/20021209ce.html/$file/20021209ce.pdf).
- BecA (Biosciences eastern and central Africa) Hub and CSIRO (Commonwealth Scientific and Industrial Research Organisation) 2010. Partnership between BecA Hub and CSIRO: design document. Revised 2 August 2010.
- Blamey A. and Mackenzie M. 2007. Theories of change and realistic evaluation. *Evaluation* 13(4), 439–455.
- Brouselle A., Contandriopoulos D. and Lemire M. 2009. Using logic analysis to evaluate knowledge transfer initiatives: the case of the research collective on the organization of primary care services. *Evaluation* 14(2), 165–183.
- Brouselle A. and Champagne F. 2011. Program theory evaluation: logic analysis. *Evaluation and Program Planning* 34(1), 69–78.
- Cartwright N. and Hardie J. 2012. Evidence-based policy: doing it better. a practical guide to predicting if a policy will work for you. Oxford University Press.
- CIDA (Canadian International Development Agency) 2002. Assessing sustainability. Gatineau: Quebec. Accessible at [http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Performancereview3/\\$file/Assessing_Sus.pdf](http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Performancereview3/$file/Assessing_Sus.pdf).
- Coffman J. 2009. A user's guide to advocacy evaluation planning: Harvard Family Research Project. Accessible at <http://www.hfrp.org/evaluation/publications-resources/a-user-s-guide-to-advocacy-evaluation-planning>.
- Commission of the European Communities 1996. Report of the final evaluation of FORCE Programme. European Commission: Brussels.
- Corbishley J. and Pearce D. 2007. Growing trees on salt-affected land. ACIAR Impact Assessment Series Report No. 51. Australian Centre for International Agricultural Research: Canberra.
- Coryn C.L.S., Noakes L.A., Westine C.D. and Schroter D.C. 2011. A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation* 32(2), 199–226.
- CPWF (CGIAR Challenge Program on Water and Food) 2012. CPWF BDC Annual Report Ganges river basin: April 2011 to December 2012. Accessible at cgspace.cgiar.org/handle/10568/21586.
- Darnton A. 2008. Practical guide. An overview of behaviour change models and their uses. Accessible at http://www.civilservice.gov.uk/wp-content/uploads/2011/09/Behaviour-change_practical_guide_tcm6-9696.pdf.
- Davies R. and Dart J. 2005. The 'most significant change' (MSC) technique: a guide to its use. Accessible at <http://www.mande.co.uk/docs/MSCGuide.pdf>.
- Davis J., Gordon J., Pearce D. and Templeton D. 2008. Guidelines for assessing the impacts of ACIAR's research activities. ACIAR Impact Assessment Series Report No. 58. Australian Centre for International Agricultural Research: Canberra.
- Deb U.K. and Bantilan M.C.S. 2006. Spillover impacts of agricultural research: a review of studies. *Journal of SAT Agricultural Research* 2(1). At <http://ejournal.icrisat.org/mpii/v2i1/v2i1spillover.pdf>.
- de Janvry A., Dustan A. and Sadoulet E. 2011. Recent advances in impact analysis methods for ex-post impact assessments of agricultural technology: options for the CGIAR. CGIAR Independent Science and Partnership Council. Accessible at <http://impact.cgiar.org/sites/default/files/images/deJanvryetal2011.pdf>.
- Douthwaite B., Alvarez S., Thiele G. and Mackay R. 2008. Participatory impact pathways analysis: a practical method for project planning and evaluation. The Institutional Learning and Change Initiative, ILAC Brief 17. At http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief17_PIPA.pdf, accessed 27 April 2010.

- Fujisaka S. and White D. 2004. Ex post methods to measure natural resource management research impacts. International Center for Agricultural Research in the Dry Areas: Aleppo, Syria.
- Funnell S. and Rogers P. 2011. Purposeful program theory: effective use of theories of change and logic models. Jossey-Bass: San Francisco, CA.
- Harris D.N. 2004. Water and nitrogen management in wheat-maize production on the North China Plain. ACIAR Impact Assessment Series Report No. 28, Australian Centre for International Agricultural Research: Canberra.
- 2006. Water management in public irrigation schemes in Vietnam. ACIAR Impact Assessment Series Report No. 43. Australian Centre for International Agricultural Research: Canberra.
- 2010. Extending low-cost fish farming in Thailand: an ACIAR-World Vision collaborative program. ACIAR Impact Assessment Series Report No. 66. Canberra: Australian Centre for International Agricultural Research.
- Horton D. 2002. Planning, implementing, and evaluating capacity development. ISNAR Briefing Paper No. 50. International Service for National Agricultural Research. Accessible at <preval.org/files/2105.pdf>.
- Hovland I. 2007. Making a difference: M&E of policy research. Working Paper 281. Overseas Development Institute: London. Accessible at <http://www.odi.org.uk/resources/docs/2426.pdf>.
- Huysse H., Molenaers N., Phlix G., Bossuyt J. and Fonteneau B. 2012. Evaluating NGO-capacity development interventions: enhancing frameworks, fitting the (Belgian) context. *Evaluation* 18(1), 129–150.
- Independent Science and Partnership Council 2011. Measuring the environmental impacts of agricultural research: theory and applications to CGIAR research. CGIAR: Montpellier.
- 2012. Strengthening strategy and results framework through prioritization. CGIAR: Montpellier.
- Innovation Network n.d.. A practical guide to advocacy evaluation, pathfinder evaluation edition. Accessible at <http://www.innonet.org/client_docs/File/advocacy/pathfinder_evaluator_web.pdf>.
- Kelley T.G. and Gregersen H.M. 2005. NRM impact assessment in the CGIAR: meeting the challenges and implications for CGIAR centres. Pp. 341–359 in 'Natural resource management in agriculture: methods for assessing economic and environmental impacts', ed. by B. Shiferaw, H.A. Freeman and S.W. Swinton. CAB International: Wallingford, UK.
- Kelley T., Ryan J. and Gregersen H. 2008. Enhancing ex post impact assessment of agricultural research: the CGIAR experience. *Research Evaluation* 17(3), 1–12.
- Kivimaa P., Mela H. and Furman E. 2007. Approaches and practices in mid-term and ex-post evaluation of environmental research programmes, SKEP ERA-NET Work Package 3: Finnish Environment Institute (SYKE). Accessible at <http://www.skep-network.eu/Home/Research/Search.aspx>.
- Larbi G., Jackson P., House S. and Ghartey A. 2005. Capacity development outcome evaluation, field-testing of the methodology. Danida, Ministry of Foreign Affairs: Copenhagen.
- Leeuw F. 2012. Linking theory-based evaluation and contribution analysis: three problems and a few solutions. *Evaluation* 18(3), 348–363.
- Mackay R., Horton D., Dupleich L. and Andersen A. 2002. Evaluating organizational capacity development. *Canadian Journal of Program Evaluation* 17(2), 121–150.
- Mackie J.L. 1974. The cement of the universe: a study of causation. Oxford University Press.
- Mahoney J. 2008. Toward a unified theory of causality. *Comparative Political Studies* 14(4/5), 412–436.
- Mansfield C. 2010. Monitoring and evaluation of advocacy campaigns literature review. Ecumenical Advocacy Alliance. Accessible at <http://www.e-alliance.ch/fileadmin/user_upload/docs/Advocacy_Capacity/2011/11._EAA_M_E_Literature_Review.pdf>.
- Maredia M.K. 2009. Improving the proof: evolution of and emerging trends in impact assessment methods and approaches in agricultural development. IFPRI Discussion Paper 00929. International Food Policy Research Institute: Washington, DC.
- Mayne J. 2007. Evaluation for accountability: reality or myth? Pp. 63–84 in 'Making accountability work: dilemmas for evaluation and for audit', ed. by M.-L. Bemelmans-Videc, J. Lonsdale and B. Perrin. Transaction Publishers: New Brunswick, NJ.
- 2008. Contribution analysis: an approach to exploring cause and effect. The Institutional Learning and Change Initiative, ILAC Brief No. 16. Accessible at <http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf>.
- 2011. Contribution analysis: addressing cause and effect. Pp. 53–96 in 'Evaluating the complex', ed. by R. Schwartz, K. Forss and M. Marra. Transaction Publishers: New Brunswick, NJ.
- 2012a. Contribution analysis: coming of age? *Evaluation* 18(3), 270–280.

- 2012b. Contribution analysis. *Evaluation* 18(3), 270–280.
- ODI (Overseas Development Institute) n.d. Episode studies. ODI Research and Policy in Development. Accessible at <http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/7072.pdf>.
- OECD (Organisation for Economic Co-operation and Development) 2011. Evaluation of agri-environmental policies: selected methodological issues and case studies. OECD Publishing: Paris.
- 2012. Farmer behaviour, agricultural management and climate change. OECD Publishing: Paris.
- OECD–DAC (Development Assistance Committee) 2002. Glossary of key terms in evaluation and results based management. OECD: Paris. At <http://www.oecd.org/dataoecd/29/21/2754804.pdf>, accessed 8 March 2008.
- Patton M.Q. 2008. Advocacy impact evaluation. *Journal of MultiDisciplinary Evaluation* 5(9), 1–10. Accessible at http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/159/181.
- Pawson R. and Tilley N. 1997. *Realistic evaluation*. Sage Publications.
- Pearce D. and White L. 2012. Including natural resource management and environmental impacts within impact assessment studies: methodological issues. ACIAR Impact Assessment Series Report No. 81. Australian Centre for International Agricultural Research: Canberra.
- Perrin B. 2007. A new view of accountability. Pp. 41–59 in ‘Making accountability work: dilemmas for evaluation and for audit’, ed. by M.-L. Bemelmans-Videc, J. Lonsdale and B. Perrin. Transaction Publishers: New Brunswick, NJ.
- Poteete A.R., Janssen M.A. and Ostrom E. 2010. Working together: collective action, the commons, and multiple methods in practice. Princeton University Press: Princeton, NJ.
- Ragin C. 1987. *The comparative method: moving beyond qualitative and quantitative strategies*. University of California Press: Berkeley, CA.
- Ringsing B. and Leeuwis C. 2008. Learning about advocacy: a case-study of challenges, everyday practices and tensions. *Evaluation* 14(14), 413–436.
- Rogers P. 2007. Theory-based evaluations, reflections ten years on. *New Directions for Evaluation* 114, 63–67.
- Saunders C., Davis L. and Pearce D. 2012. Rice–wheat cropping systems in India and Australia, and development of the ‘Happy Seeder’. ACIAR Impact Assessment Series Report No. 77. Australian Centre for International Agricultural Research: Canberra.
- Science Council Secretariat 2006. *Natural resources management research impacts: evidence from the CGIAR*. CGIAR Science Council, Standing Committee on Impact Assessment: Montpellier.
- Shiferaw B. and Freeman H.A. (eds) 2003. *Methods for assessing the impacts of natural resource management research*. A summary of the proceedings of an International Workshop, International Crops Research Institute for the Semi-Arid Tropics. Patancheru, India, 6–7 December 2002.
- Smith L., Khan F., Frankenberger T.R. and Wadud A.K.M.A. 2012. Admissible evidence in the court of development evaluation? The impact of CARE’s SHOUHARDO project on child stunting in Bangladesh. *World Development*. Accessible at <http://dx.doi.org/10.1016/j.worlddev.2012.06.018>.
- Stame N. 2004. Theory-based evaluation and varieties of complexity. *Evaluation* 10(1), 58–76.
- Stein D. and Valters C. 2012. Understanding ‘theory of change’ in international development: a review of existing knowledge. The Asian Institute and the Justice and Security Research Programme. Accessible at http://www2.lse.ac.uk/internationalDevelopment/research/JSRP/downloads/ToC_Lit_Review.pdf.
- Stern E. 1996. Evaluation issues for European education and training programmes. In proceedings of ‘Evaluation of European training, employment and human resource programmes’. FHVR: Berlin; CEDEFOP: Thessaloniki.
- 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 No. 38. Department for International Development: London. Accessible at <http://www.dfid.gov.uk/R4D/Output/189575/Default.aspx>.
- Turbin J. 2001. Policy borrowing: lessons from European attempts to transfer training practices. *International Journal of Training and Development* 5(2), 96–111.
- UN (United Nations) 1987. *Report of the World Commission on Environment and Development: our common future*. (Brundtland report). Oxford University Press: New York. Accessible at <http://www.earthsummit2012.org/about-us/historical-documents/92-our-common-future>.
- Vogel I. 2012. Review of the use of ‘theory of change’ in international development. Department for International Development: London. Accessible at http://www.oxfamblogs.org/fp2p/wp-content/uploads/DFID-ToC-Review_VogelV4.pdf.

- Walker T., Maredia M., Kelley T., Rovere R.L., Templeton D., Thiele G. and Douthwaite B. 2008. Strategic guidance for ex post impact assessment of agricultural research. CGIAR Science Council: Montpellier. Accessible at <<http://impact.cgiar.org/sites/default/files/pdf/7-Strategic-Guidelines.pdf>>.
- Weiss C.H. 1997. Theory-based evaluation: past, present, and future. *New Directions for Evaluation* 76(winter), 41–55.
- White H. 2009. Theory-based impact evaluation: principles and practice. Working Paper 3, International Initiative for Impact Evaluation (3ie). Accessible at <http://www.3ieimpact.org/admin/pdfs_papers/48.pdf>.
- Woolcock M. 2009. Towards a plurality of methods in project evaluation: a contextualised approach to understanding impact trajectories and efficacy. Working Paper 73, University of Manchester, Brooks World Poverty Institute. Accessible at <<http://www.bwpi.manchester.ac.uk/resources/Working-Papers/bwpi-wp-7309.pdf>>.
- WorldFish Center 2011. Proposal for CGIAR research program on aquatic agricultural systems. Accessible at <http://www.worldfishcenter.org/resource_center/WF_2936.pdf>.

IMPACT ASSESSMENT SERIES

No.	Author(s) and year of publication	Title	ACIAR project numbers
1	Centre for International Economics 1998.	Control of Newcastle disease in village chickens	AS1/1983/034, AS1/1987/017 and AS1/1993/222
2	George P.S. 1998.	Increased efficiency of straw utilisation by cattle and buffalo	AS1/1982/003, AS2/1986/001 and AS2/1988/017
3	Centre for International Economics 1998.	Establishment of a protected area in Vanuatu	ANRE/1990/020
4	Watson A.S. 1998.	Raw wool production and marketing in China	ADP/1988/011
5	Collins D.J. and Collins B.A. 1998.	Fruit fly in Malaysia and Thailand 1985–1993	CS2/1983/043 and CS2/1989/019
6	Ryan J.G. 1998.	Pigeonpea improvement	CS1/1982/001 and CS1/1985/067
7	Centre for International Economics 1998.	Reducing fish losses due to epizootic ulcerative syndrome—an ex ante evaluation	FIS/1991/030
8	McKenney D.W. 1998.	Australian tree species selection in China	FST/1984/057 and FST/1988/048
9	ACIL Consulting 1998.	Sulfur test KCL–40 and growth of the Australian canola industry	PN/1983/028 and PN/1988/004
10	AACM International 1998.	Conservation tillage and controlled traffic	LWR2/1992/009
11	Chudleigh P. 1998.	Postharvest R&D concerning tropical fruits	PHT/1983/056 and PHT/1988/044
12	Waterhouse D., Dillon B. and Vincent D. 1999.	Biological control of the banana skipper in Papua New Guinea	CS2/1988/002-C
13	Chudleigh P. 1999.	Breeding and quality analysis of rapeseed	CS1/1984/069 and CS1/1988/039
14	McLeod R., Isvilanonda S. and Wattanuchariya S. 1999.	Improved drying of high moisture grains	PHT/1983/008, PHT/1986/008 and PHT/1990/008
15	Chudleigh P. 1999.	Use and management of grain protectants in China and Australia	PHT/1990/035
16	McLeod R. 2001.	Control of footrot in small ruminants of Nepal	AS2/1991/017 and AS2/1996/021
17	Tisdell C. and Wilson C. 2001.	Breeding and feeding pigs in Australia and Vietnam	AS2/1994/023
18	Vincent D. and Quirke D. 2002.	Controlling <i>Phalaris minor</i> in the Indian rice–wheat belt	CS1/1996/013
19	Pearce D. 2002.	Measuring the poverty impact of ACIAR projects—a broad framework	
20	Warner R. and Bauer M. 2002.	<i>Mama Lus Frut</i> scheme: an assessment of poverty reduction	ASEM/1999/084
21	McLeod R. 2003.	Improved methods in diagnosis, epidemiology, and information management of foot-and-mouth disease in Southeast Asia	AS1/1983/067, AS1/1988/035, AS1/1992/004 and AS1/1994/038
22	Bauer M., Pearce D. and Vincent D. 2003.	Saving a staple crop: impact of biological control of the banana skipper on poverty reduction in Papua New Guinea	CS2/1988/002-C
23	McLeod R. 2003.	Improved methods for the diagnosis and control of bluetongue in small ruminants in Asia and the epidemiology and control of bovine ephemeral fever in China	AS1/1984/055, AS2/1990/011 and AS2/1993/001
24	Palis F.G., Sumalde Z.M. and Hossain M. 2004.	Assessment of the rodent control projects in Vietnam funded by ACIAR and AusAID: adoption and impact	AS1/1998/036

IMPACT ASSESSMENT SERIES <CONTINUED>

No.	Author(s) and year of publication	Title	ACIAR project numbers
25	Brennan J.P. and Quade K.J. 2004.	Genetics of and breeding for rust resistance in wheat in India and Pakistan	CS1/1983/037 and CS1/1988/014
26	Mullen J.D. 2004.	Impact assessment of ACIAR-funded projects on grain-market reform in China	ADP/1997/021 and ANRE1/1992/028
27	van Bueren M. 2004.	Acacia hybrids in Vietnam	FST/1986/030
28	Harris D. 2004.	Water and nitrogen management in wheat–maize production on the North China Plain	LWR1/1996/164
29	Lindner R. 2004.	Impact assessment of research on the biology and management of coconut crabs on Vanuatu	FIS/1983/081
30	van Bueren M. 2004.	Eucalypt tree improvement in China	FST/1984/057, FST/1987/036, FST/1988/048, FST/1990/044, FST/1994/025, FST/1996/125 and FST/1997/077
31	Pearce D. 2005.	Review of ACIAR's research on agricultural policy	
32	Tingsong Jiang and Pearce D. 2005.	Shelf-life extension of leafy vegetables—evaluating the impacts	PHT/1994/016
33	Vere D. 2005.	Research into conservation tillage for dryland cropping in Australia and China	LWR2/1992/009 and LWR2/1996/143
34	Pearce D. 2005.	Identifying the sex pheromone of the sugarcane borer moth	CS2/1991/680
35	Raitzer D.A. and Lindner R. 2005.	Review of the returns to ACIAR's bilateral R&D investments	
36	Lindner R. 2005.	Impacts of mud crab hatchery technology in Vietnam	FIS/1992/017 and FIS/1999/076
37	McLeod R. 2005.	Management of fruit flies in the Pacific	CS2/1989/020, CS2/1994/003, CS2/1994/115 and CS2/1996/225
38	ACIAR 2006.	Future directions for ACIAR's animal health research	
39	Pearce D., Monck M., Chadwick K. and Corbishley J. 2006.	Benefits to Australia from ACIAR-funded research	AS2/1990/028, AS2/1994/017, AS2/1994/018, AS2/1999/060, CS1/1990/012, CS1/1994/968, FST/1993/016 and PHT/1990/051
40	Corbishley J. and Pearce D. 2006.	Zero tillage for weed control in India: the contribution to poverty alleviation	CS1/1996/013
41	ACIAR 2006.	ACIAR and public funding of R&D. Submission to Productivity Commission study on public support for science and innovation	
42	Pearce D. and Monck M. 2006.	Benefits to Australia of selected CABI products	
43	Harris D.N. 2006.	Water management in public irrigation schemes in Vietnam	LWR1/1998/034 and LWR2/1994/004
44	Gordon J. and Chadwick K. 2007.	Impact assessment of capacity building and training: assessment framework and two case studies	CS1/1982/001, CS1/1985/067, LWR2/1994/004 and LWR2/1998/034
45	Turnbull J.W. 2007.	Development of sustainable forestry plantations in China: a review	
46	Monck M. and Pearce D. 2007.	Mite pests of honey bees in the Asia–Pacific region	AS2/1990/028, AS2/1994/017, AS2/1994/018 and AS2/1999/060

IMPACT ASSESSMENT SERIES <CONTINUED>

No.	Author(s) and year of publication	Title	ACIAR project numbers
47	Fisher H. and Gordon J. 2007.	Improved Australian tree species for Vietnam	FST/1993/118 and FST/1998/096
48	Longmore C., Gordon J. and Bantilan M.C. 2007.	Assessment of capacity building: overcoming production constraints to sorghum in rainfed environments in India and Australia	CS1/1994/968
49	Fisher H. and Gordon J. 2007.	Minimising impacts of fungal disease of eucalypts in South-East Asia	FST/1994/041
50	Monck M. and Pearce D. 2007.	Improved trade in mangoes from the Philippines, Thailand and Australia	CS1/1990/012 and PHT/1990/051
51	Corbishley J. and Pearce D. 2007.	Growing trees on salt-affected land	FST/1993/016
52	Fisher H. and Gordon J. 2008.	Breeding and feeding pigs in Vietnam: assessment of capacity building and an update on impacts	AS2/1994/023
53	Monck M. and Pearce D. 2008.	The impact of increasing efficiency and productivity of ruminants in India by the use of protected-nutrient technology	AH/1997/115
54	Monck M. and Pearce D. 2008.	Impact of improved management of white grubs in peanut-cropping systems	CS2/1994/050
55	Martin G. 2008.	ACIAR fisheries projects in Indonesia: review and impact assessment	FIS/1997/022, FIS/1997/125, FIS/2000/061, FIS/2001/079, FIS/2002/074, FIS/2002/076, FIS/2005/169 and FIS/2006/144
56	Lindner B. and McLeod P. 2008.	A review and impact assessment of ACIAR's fruit-fly research partnerships—1984 to 2007	CP/1997/079, CP/2001/027, CP/2002/086, CP/2007/002, CP/2007/187, CS2/1983/043, CS2/1989/019, CS2/1989/020, CS2/1994/003, CS2/1994/115, CS2/1996/225, CS2/1997/101, CS2/1998/005, CS2/2003/036, PHT/1990/051, PHT/1993/87 and PHT/1994/133
57	Montes N.D., Zapata Jr N.R., Alo A.M.P. and Mullen J.D. 2008.	Management of internal parasites in goats in the Philippines	AS1/1997/133
58	Davis J., Gordon J., Pearce D. and Templeton D. 2008.	Guidelines for assessing the impacts of ACIAR's research activities	
59	Chupungco A., Dumayas E. and Mullen J. 2008.	Two-stage grain drying in the Philippines	PHT/1983/008, PHT/1986/008 and PHT/1990/008
60	Centre for International Economics 2009.	ACIAR Database for Impact Assessments (ADIA): an outline of the database structure and a guide to its operation	
61	Fisher H. and Pearce D. 2009.	Salinity reduction in tannery effluents in India and Australia	AS1/2001/005
62	Francisco S.R., Mangabat M.C., Mataia A.B., Acda M.A., Kagaoan C.V., Laguna J.P., Ramos M., Garabiag K.A., Paguia F.L. and Mullen J.D. 2009.	Integrated management of insect pests of stored grain in the Philippines	PHT/1983/009, PHT/1983/011, PHT/1986/009 and PHT/1990/009
63	Harding M., Tingsong Jiang and Pearce D. 2009.	Analysis of ACIAR's returns on investment: appropriateness, efficiency and effectiveness	
64	Mullen J.D. 2010.	Reform of domestic grain markets in China: a reassessment of the contribution of ACIAR-funded economic policy research	ADP/1997/021 and ANRE1/1992/028

IMPACT ASSESSMENT SERIES <CONTINUED>

No.	Author(s) and year of publication	Title	ACIAR project numbers
65	Martin G. 2010.	ACIAR investment in research on forages in Indonesia	AS2/2000/103, AS2/2000/124, AS2/2001/125, LPS/2004/005, SMAR/2006/061 and SMAR/2006/096
66	Harris D.N. 2010.	Extending low-cost fish farming in Thailand: an ACIAR–World Vision collaborative program	PLIA/2000/165
67	Fisher H. 2010.	The biology, socioeconomics and management of the barramundi fishery in Papua New Guinea’s Western Province	FIS/1998/024
68	McClintock A. and Griffith G. 2010.	Benefit–cost meta-analysis of investment in the International Agricultural Research Centres	
69	Pearce D. 2010.	Lessons learned from past ACIAR impact assessments, adoption studies and experience	
70	Harris D.N. 2011.	Extending low-chill fruit in northern Thailand: an ACIAR–World Vision collaborative project	PLIA/2000/165
71	Lindner R. 2011.	The economic impact in Indonesia and Australia from ACIAR’s investment in plantation forestry research, 1987–2009	FST/1986/013, FST/1990/043, FST/1993/118, FST/1995/110, FST/1995/124, FST/1996/182, FST/1997/035, FST/1998/096, FST/2000/122, FST/2000/123, FST/2003/048 and FST/2004/058
72	Lindner R. 2011.	Frameworks for assessing policy research and ACIAR’s investment in policy-oriented projects in Indonesia	ADP/1994/049, ADP/2000/100, ADP/2000/126, AGB/2000/072, AGB/2004/028, ANRE1/1990/038, ANRE1/1993/023, ANRE1/1993/705, EFS/1983/062 and EFS/1988/022
73	Fisher H. 2011.	Forestry in Papua New Guinea: a review of ACIAR’s program	FST/1994/033, FST/1995/123, FST/1998/118, FST/2002/010, FST/2004/050, FST/2004/055, FST/2004/061, FST/2006/048, FST/2006/088, FST/2006/120, FST/2007/078 and FST/2009/012
74	Brennan J.P. and Malabayabas A. 2011.	International Rice Research Institute’s contribution to rice varietal yield improvement in South-East Asia	
75	Harris D.N. 2011.	Extending rice crop yield improvements in Lao PDR: an ACIAR–World Vision collaborative project	CIM/1999/048, CS1/1995/100 and PLIA/2000/165
76	Grewal B., Grunfeld H. and Sheehan P. 2011.	The contribution of agricultural growth to poverty reduction	
77	Saunders C., Davis L. and Pearce D. 2012.	Rice–wheat cropping systems in India and Australia and development of the ‘Happy Seeder’	LWR/2000/089, LWR/2006/132 and CSE/2006/124
78	Carpenter D. and McGillivray M. 2012	A methodology for assessing the poverty-reducing impacts of Australia’s international agricultural research	
79	Dugdale A., Sadleir C., Tennant-Wood R. and Turner M. 2012	Developing and testing a tool for measuring capacity building	
80	Fisher H., Sar L. and Winzenried C. 2012	Oil palm pathways: an analysis of ACIAR’s oil palm projects in Papua New Guinea	ASEM/1999/084, ASEM/2002/014, ASEM/2006/127, CP/1996/091, PC/2006/063, PC/2004/064, CP/2007/098

IMPACT ASSESSMENT SERIES <CONTINUED>

No.	Author(s) and year of publication	Title	ACIAR project numbers
81	Pearce D. and White L. 2012	Including natural resource management and environmental impacts within impact assessment studies: methodological issues	
82	Fisher H. and Hohnen L. 2012	ACIAR's activities in Africa: a review	AS1/1983/003, AS1/1995/040, AS1/1995/111, AS1/1996/096, AS1/1998/010, AS2/1990/047, AS2/1991/018, AS2/1993/724, AS2/1996/014, AS2/1999/063, AS2/1996/090, AS2/1996/203, AS2/1996/149, AS2/1997/098, CP/1994/126, CS2/1990/007, EFS/1983/026, FST/1983/020, FST/1983/031, FST/1983/057, FST/1988/008, FST/1988/009, FST/1991/026, FST/1995/107, FST/1996/124, FST/1996/206, FST/2003/002, IAP/1996/181, LPS/1999/036, LPS/2002/081, LPS/2004/022, LPS/2008/013, LWR/2011/015, LWR1/1994/046, LWR2/1987/035, LWR2/1996/049, LWR2/1996/163, LWRS/1996/215, LWR2/1997/038, SMCN/1999/003, SMCN/1999/004, SMCN/2001/028, SMCN/2000/173
83	Palis F.G., Sumalde Z.M., Torres C.S., Contreras A.P. and Datar F.A. 2013	Impact pathway analysis of ACIAR's investment in rodent control in Vietnam, Lao PDR and Cambodia	ADP/2000/007, ADP/2003/060, ADP/2004/016, AS1/1994/020, AS1/1996/079, AS1/1998/036, CARD 2000/024, PLIA/2000/165
84	Mayne J. and Stern E. 2013	Impact evaluation of natural resource management research programs: a broader view	



ACIAR

Research that works for developing
countries and Australia

aciarc.gov.au