BEAM EXCHANGE

building effective & accessible markets

IMPACT EVALUATIONS FOR MARKET SYSTEMS PROGRAMMES

CITATION

Fionn O'Sullivan (February 2016) 'Impact Evaluations for Market Systems Programmes', The BEAM Exchange, accessed from https://www.beamexchange.org. © 2016 The BEAM Exchange

Acknowledgements:BEAM Exchange would like to thank Tim Ruffer and Marcus Jenal for their support and comments.

Published by:

The BEAM Exchange c/o
PricewaterhouseCoopers LLP
7 More London Riverside
London SE1 2RT
United Kingdom

info@beamexchange.org
https://www.beamexchange.org

These guidelines have been produced with the support of:

UKaid
from the British people

https://www.gov.uk/government/organisations/department-for-international-development

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development and Cooperation SDC

https://www.eda.admin.ch/sdc

The BEAM Exchange is a programme funded by the UK's Department for International Development ("DFID") and the Swiss Agency for Development and Cooperation. It is administered by PricewaterhouseCoopers LLP, working with organisations including the Institute of Development Studies and ITAD.

CONTENTS

1. INTRODUCTION

1.1 Purpose of the guidelines

These guidelines set out the key steps and considerations for designing, planning and undertaking the evaluation of a programme or intervention that applies a market systems approach (also described as the making markets work for the poor (M4P) approach). Programmes using this approach typically implement projects targeted on different areas of a market system, with the aim of shifting the way the system works and therefore improving the lives of men and women living in poverty.[1]

Evaluation is understood here as a systematic assessment of impacts which includes positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.[2] A key motivation for commissioning this document is that many previous evaluations of market systems programmes failed to take all of these effects into account. In particular, it aims to provide practical guidance to tackle weaknesses identified in previous research,[3] including:

- The use of simple, linear theories of change which do not adequately reflect the system-changing ambitions of the approach.
- Limited attention paid to possible unintended and negative effects, particularly for poor men and women who may not be the direct beneficiaries of market changes.
- Poor data quality (such as small sample sizes and little consideration of sampling frames, statistical significance or bias).
- Weak triangulation practices, in relation to qualitative data in particular.

The guidelines are intended both for evaluators, and for those involved in drawing up the terms of reference and commissioning an evaluation (programme managers and donors).

While many market systems programmes undertake monitoring or results measurement - activities that could be considered as 'evaluative,' - these guidelines are focused on evaluations that are undertaken independently of the programme.

1 For a fuller description of what this means see, 'The Operational Guide for the making markets work for the poor (M4P) approach', (2014). https://beamexchange.org/resources/167/

2 OECD-DAC (2002), 'Glossary of key terms in evaluation and results based management' http://www.oecd.org/dac/2754804.pdf

3 See Ruffer, T. and Wach, E. (2013), 'Review of making markets work for the poor (M4P): evaluation methods and approaches'. https://beamexchange.org/resources/133/

However many parts of the guidelines should also be useful for programmes that choose to undertake evaluative work internally.

Evaluation practice is still developing in response to the specific challenges of market systems programmes. It is therefore hoped that lessons learned in evaluating programmes in this field will help the guidelines presented in this document to be developed and further refined.

1.2 Structure of the guidelines

These guidelines start by reviewing some general considerations in evaluating market systems programmes, and then move on to consider evaluation designs and methods.

Chapter 2: Defines what market systems programmes are and how they work in practice. It reviews definitions of systemic change, and identifies important implications for evaluation arising from the complexity of market systems.

Chapter 3: Explains the relevance of theory-based evaluation for market systems programmes and the central importance of the theory of change in providing the underlying framework for the evaluation. It explains how such a theory can be defined and revised during the implementation of a market systems programme.

Chapter 4: Explains how to select an evaluation design, taking into account what the results will be used for, what the evaluation questions are, which parts of the programme are actually evaluable, and which evaluation designs can be used.

Chapter 5: Considers different methods for evaluating impact using the principles of attribution and contribution, and the potential biases and other pitfalls involved for each one. It considers how results can be triangulated, which data collection tools may be appropriate, and discusses the sequencing of data collection. It concludes by explaining how the choice of designs and methods can be summarised in a simple evaluation framework.

Chapter 6: covers some outstanding issues, including the reasons for linking monitoring and evaluation, the role of the evaluator, and the reporting and communication of findings.

2. KEY EVALUATION CONSIDERATIONS

2.1 Overview of this chapter

This chapter introduces some key considerations to ensure that an evaluation adequately addresses the specific characteristics of market systems programmes. It starts by defining what a market system is, and the ways in which programmes seek to change systems in order to improve the lives of people living in poverty. It discusses the ideas of complexity and adaptive programming and outlines the implications for evaluations. Finally, it suggests key criteria for rigorous and credible evaluations.

2.2 Market systems and market systems programmes

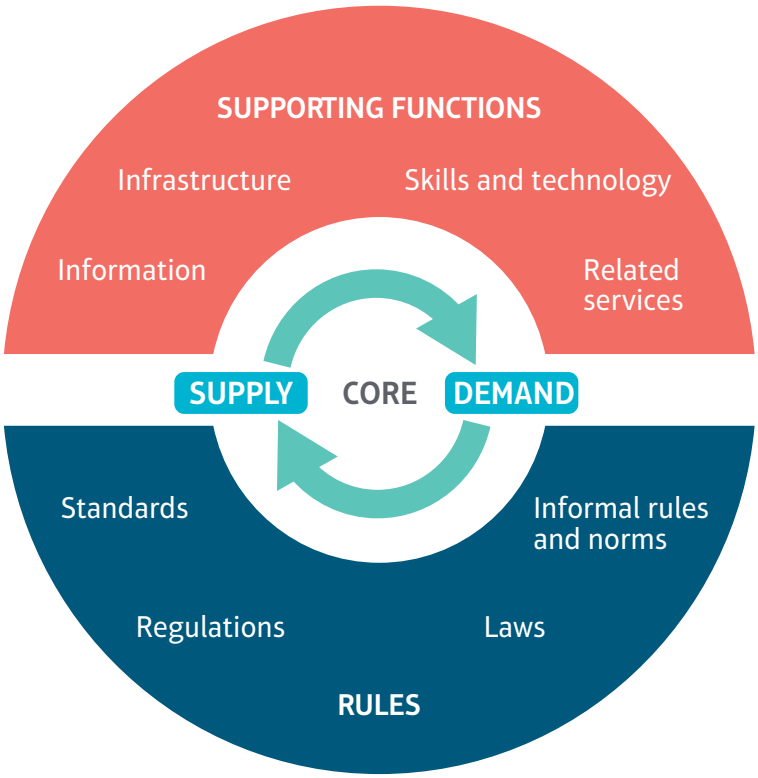
Market systems approaches aim to reduce poverty by stimulating market growth in order to provide employment and income, or to allow men and women living in poverty to access new or improved goods and services. The approach is based on an understanding that markets are not simply a collection of transactions between buyers and sellers, but function within a wider system in which both formal and informal rules and other supporting functions help determine how efficiently they operate, who can participate in them, and who benefits. These elements are summarised in Diagram 1 below.

Market systems programmes typically consist of a set of coordinated projects or interventions which are designed to address the underlying factors that result in a market not benefitting poor people, and in doing so, change the way the system works. As 'programme' and 'intervention' are key terms used throughout this document, it is helpful to define them:

- A programme is meant here as a collection of coordinated interventions which are implemented during a given time period, to achieve a specific set of goals.
- An intervention is a project which forms one component of the wider programme.

Further reading
For a much fuller explanation of what a market system is, and why a systemic approach is important see, 'The operational guide for the making markets work for the poor (M4P) approach' (2014).
https://beamexchange.org/resources/167/

Diagram 1: Elments of a market system



Source: M4P Operational Guide (2015)

Using a systemic approach to improve markets for health services in in Kenya

The Private Sector Innovation Programme for Health (PSP4H) works in Kenya to encourage companies to invest in providing health care products and services for the 'working poor'. It works with business partners to identify and pilot innovative approaches, and then to scale these up if they are successful. PSP4H is implementing interventions in seven different areas, including frontline services for eye care and maternal and child health, the supply chain for medicines, and health infrastructure such as the system of medical laboratories.

Programme teams analyse a market system (or systems) and identify the reasons for sub-optimal outcomes for men and women living in poverty. Typically they work with partner organisations to pilot and roll out new business models or innovations, improve policies or regulations etc. When these changes are diffused or taken up throughout the market it is expected that they improve the way the market system functions.

2.3 Defining different levels of change for a market systems programme

The ways in which programmes aim to achieve their desired goals has repercussions for which elements of their work can be evaluated, and in what way. This section therefore presents a schematic representation of both pilot interventions whose impacts are scaled up, and whole programmes, which function as a portfolio of coordinated interventions targeted on different parts of a market system. Chapter 4 returns to this representation in order to discuss the implications for evaluation design.

Working with importers and agro-vets to boost ginger yields in Nepal

The Ginger Disease Management intervention is one of the Samarth-Nepal Market Development Programme's ten interventions.

A central focus of the intervention has been on improving the productivity of ginger producers in the Mid-Hills region, where around 200,000 farmers grow the crop as contract farmers, and where others work as labourers, processors or traders.

Working with stakeholders, the programme team identified rhizome rot as one of the major causes of low yields.

The intervention then worked with importers and agro-vets (agricultural input shops) to introduce Trichoderma (a bio-fungicide that can eradicate the disease) into the market in a format that was affordable and easy to use.

The first step for a market systems intervention is to pilot a new product, process or business model. In the Ginger Disease Management intervention for example, the programme worked with selected partner organisations to introduce Trichoderma into the market for agricultural inputs. The intention was to reduce losses to rhizome rot, to increase output and therefore to improve incomes for farmers and others in the market ('pro-poor growth'), reducing their poverty. Other programmes (such as PSP4H in Kenya) aim to reduce poverty by increasing access to key services. The impacts of a pilot intervention can therefore be represented as follows:

Diagram 2: Initial impacts of pilot interventions. The aim is that where a new model or innovation delivers positive results, further work can be undertaken to scale it up, so that other market actors either adopt it, or adapt their behaviour, thereby changing the way the market works as a system and benefitting poor women and men. This approach, which aims to use market forces to make changes sustainable over time, is one of the defining characteristics of market systems thinking.

The issue of what systemic change means is discussed below. In relation to the ginger sector intervention mentioned however, the aim of the project was not just for selected partners to import and sell Trichoderma and for target communities to start using it. Instead, the intervention sought to induce a systemic change, by spreading use of this bio-fungicide throughout the Mid-Hills region, and engraining its use in local farming practices beyond the lifetime of the project.

Diagram 3: Possible systemic changes from scaling up pilot interventions. Where systemic change occurs, it is hoped that poverty reduction effects will be magnified as growth in jobs and income or an increase in access to services takes hold.

Diagram 4: Further impacts of systemic changes. Market systems are complex, and interventions that aim to influence the way they

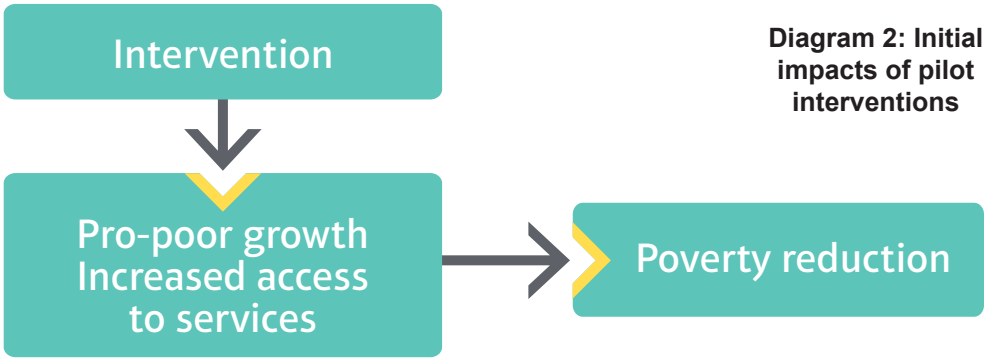


Diagram 2: Initial impacts of pilot interventions

Source: BEAM Exchange

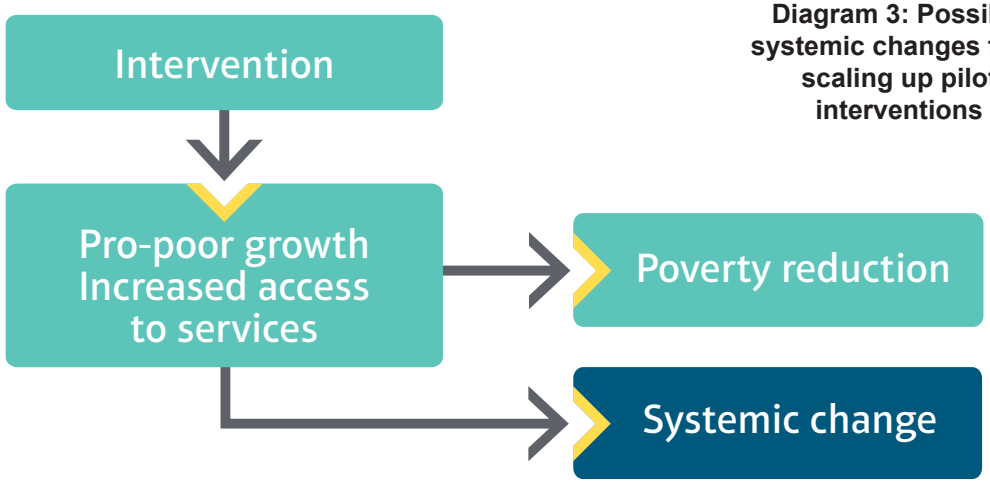


Diagram 3: Possible systemic changes from scaling up pilot interventions

Source: BEAM Exchange

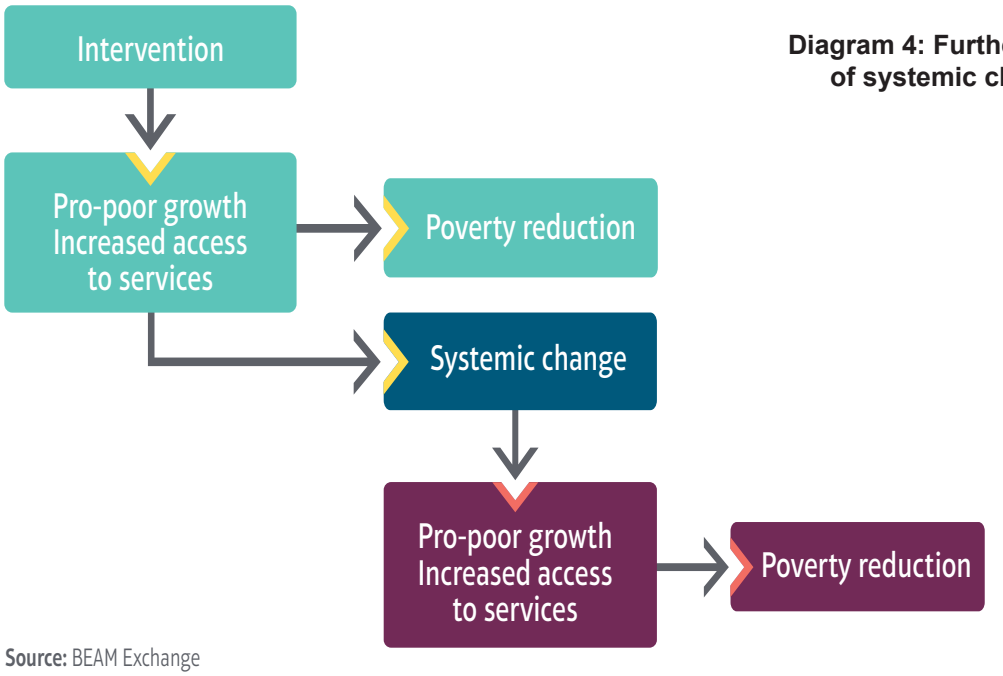


Diagram 4: Further impacts of systemic changes

Source: BEAM Exchange

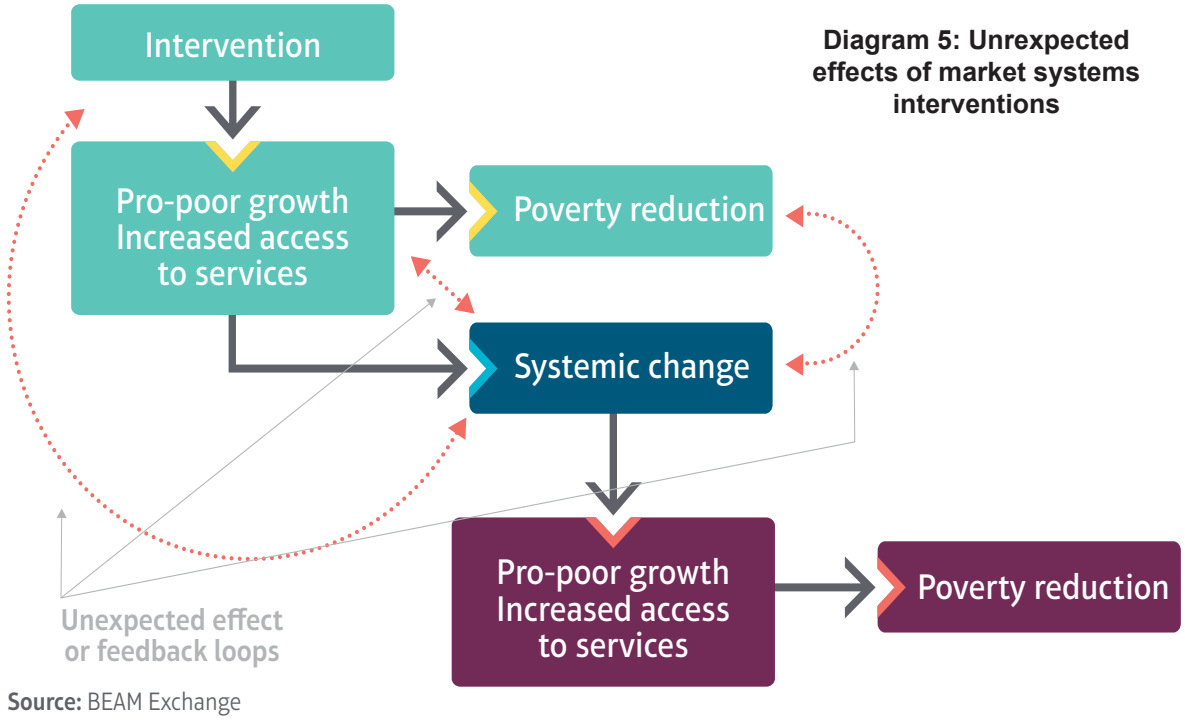


Diagram 5: Unexpected effects of market systems interventions

Source: BEAM Exchange

work will not always operare in a linear fashion. This is compounded by the fact that programmes are designed with a partial understanding of how a market system works. This suggests that unexpected effects or feedback loops between different elements of the programme will quite possibly occur in some form (indicated in Diagram 5 by the dotted red lines).

Diagram 5: Unexpected effects of market systems interventions Only measuring what a programme aims to achieve is a significant but often overlooked source of bias in evaluation. In order to be rigorous, market systems evaluations should generate an understanding of the multidimensional changes which have occurred as a result of market systems change.

This is particularly important for changes experienced by people living in poverty, even when they are not intended as primary beneficiaries. It is important then, that evaluations consider changes in relation to:

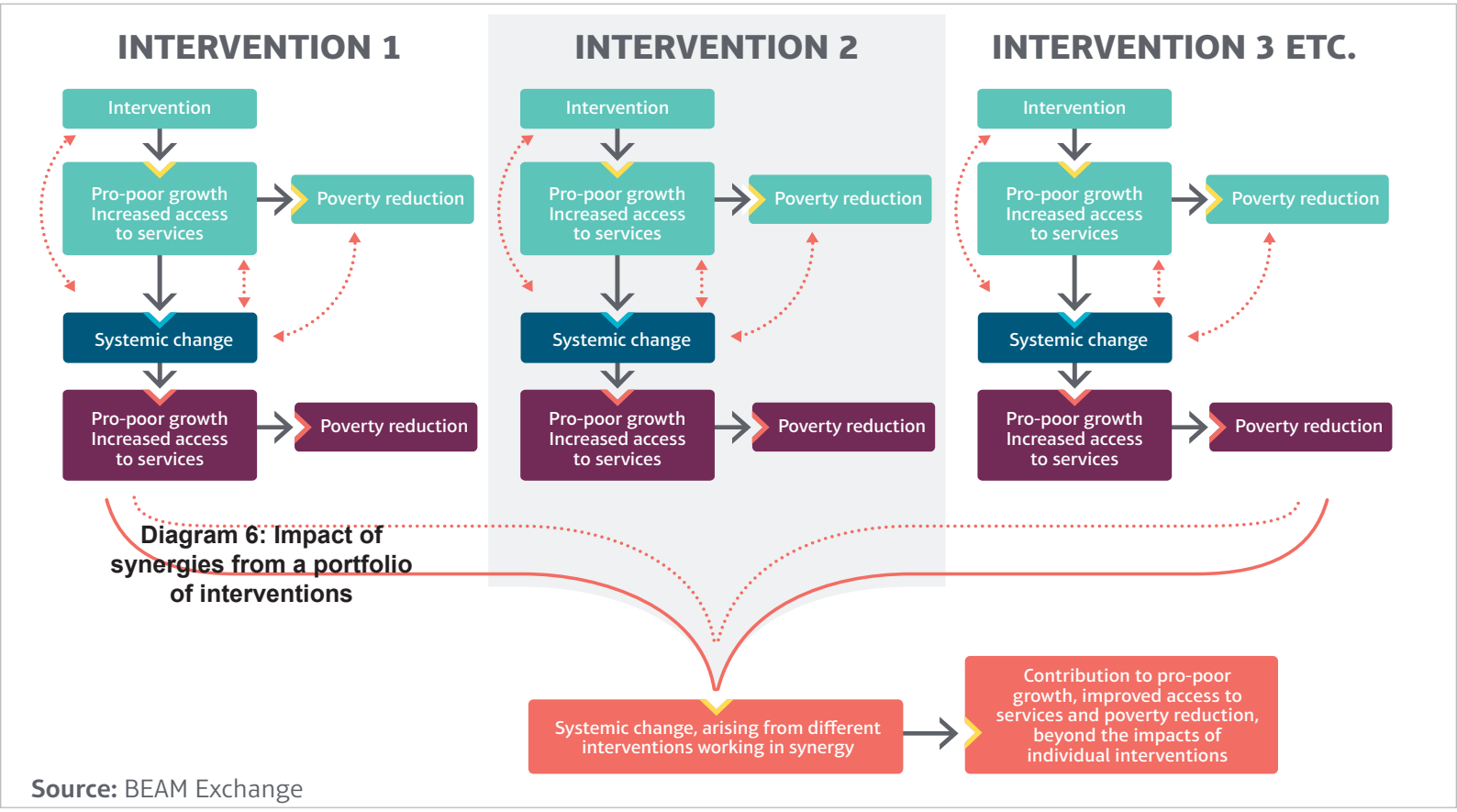
- The market actors who are expected to change their behaviour (e.g. farmers and agro-vets who start to buy and sell Trichoderma.)
- The people living in poverty who are expected to be affected by these changes (e.g. other participants in the market system such as farmers or traders, who benefit from increased production of ginger.)

Further reading
Fowler, D. and Dunn, E. (2014), 'Evaluating systems and systemic change for inclusive market development, literature review and synthesis'. https://beamexchange.org/resources/147/

- The people living in poverty who are not necessarily expected to benefit from changes, but who may still be affected (e.g. farm labourers, other members of the farm household such as the women and children of the farmers.)
- In relation to this last category, it is important to consider possible negative as well as positive effects. It is quite possible for multiplier effects to spread the benefits of successful market systems changes more widely within communities or regions. It is also possible, however, that negative impacts will occur. To provide one example: past assessments of contract farming have overlooked the implications for non-contract farmers and excluded labourers, as well as the wider implications for the local economy (e.g. higher food prices for rural households, changes in land access).

In addition to the sum total of impacts from individual interventions, evaluations of market systems programmes should also take into account the intention that synergies between different interventions work to amplify the overall impacts in terms of market efficiency, pro-poor

Diagram 6: Impact of synergies from a portfolio of interventions



Source: BEAM Exchange

growth, increased access to services and poverty reduction (as well as the possibility that they hinder or reduce each other's impacts).

2.4 What is systemic change?

A useful starting point for an explanation of systemic change is to characterise market systems programmes as aiming to, "transform the structure or dynamics of a system" in ways that lead to, "impacts on large numbers of people, either in their material conditions or in their behaviour."[1] This definition of systemic change therefore has two parts which can be considered in more detail as follows.

In relation to the first element of the definition, "transforming the structure or dynamics of a system", the Leveraging Economic Opportunities (LEO) initiative has produced a useful synthesis of different approaches and indicators that have

4 Osorio-Cortes, L., and Jenal, M. (2013), 'Monitoring and measuring change in market systems – rethinking the current paradigm.' https://beamexchange.org/resources/175/

Table 1: Synthesis of different approaches to assessing systemic market change

Type of indicator	Examples
Buy-in indicators	These measure the degree to which market actors have taken ownership of the new business models, technologies, practices or behaviour changes introduced by the intervention. Examples include: • adaptation or innovation to the original, programme-sponsored model(s) • continued, independent investment after programme sponsorship ends • repeat behaviour • satisfaction with program-facilitated changes
Imitation indicators	These measure the scale or breadth of programme-supported behaviour change within a system. Examples include: • crowding-in by other businesses that imitate programme-sponsored business models originally adopted and demonstrated by business(es) that collaborate with the implementer • copying, where market actors imitate the new practices originally adopted and demonstrated by the target beneficiaries of the intervention

been used to assess the extent to which this has occurred[2].

In relation to the second element of the definition "having impacts on large numbers of people", it is important that evaluations consider not only the numbers of people involved, but also the extent to which changes in the system that have benefitted them endure over time.

In light of this, the three elements of systemic change identified by the Donor Committee for Enterprise Development (DCED)[3] provide useful reference points that evaluations can incorporate:

- **Scale.** Systemic changes influence and benefit a large number of people who were not directly involved in the original intervention.

- **Sustainability**. Systemic changes continue to have impacts beyond the end of the programme, and are sustained without the need for further intervention.

- **Resilience.** Market players are able to adapt so that benefits continue to accrue to poor people even as the market and the external environment changes.

5 Fowler, D. and Dunn, E. (2014), 'Evaluating systems and systemic change for inclusive market development, literature review'. http://beamexchange.org/resources147/

6 Kessler, A. (2014), 'Assessing systemic change'. http://enterprise-development.org/download.ashx?id=2113

Further reading:
Fowler, D. and Dunn, E. (2014), 'Leveraging Economic Opportunities: Evaluating systems and systemic change for inclusive market development, literature review and synthesis'. https://beamexchange.org/resources/147/

2.5 The implications of complexity

One of the things that is immediately apparent from Diagram 1 is that market systems are inherently complex. Insights from systems theory and practical experience from programmes highlight the fact that interventions that aim to change the way that market systems function are unlikely to work in a linear fashion, with inputs and activities leading simply to outputs and outcomes.

Further reading:
This book provides an explanation for non-specialists of complexity theory, and its implications for development programmes in general.
Burns, D., and Worsley S. (2015), 'Navigating complexity in international development'. https://beamexchange.org/resources/665/

Non-linearity of effects may take different forms: large, well-planned interventions may founder and provide no significant or sustainable impacts; small changes may create domino effects leading to large-scale changes; nothing may seem to be happening at all, and then everything changes.

This has significant implications for programme management (see the BEAM monitoring guidance) as well as for programme evaluation. As programme interventions interact with other activities for instance, results that do emerge will have been influenced by multiple factors. While a programme intervention may contribute to an outcome, it is the broader package of causal factors that produce the intended effects. Some key considerations are:

- Market systems programmes do not 'implement' market systems change. Rather, they work to stimulate changes in the behaviour of market players. Programme effects are therefore indirect and may be difficult to predict beforehand. While a theory of change is an important tool for an evaluation, the application of a traditional linear logic model may not be appropriate for evaluation purposes.

- Systems are comprised of interconnecting elements. Accordingly, the perspectives of those working in any one part of the system will be different. Understanding how a system works requires awareness from multiple perspectives, including not only market actors and beneficiaries, but also other participants in the system who are affected by change.

- The extent to which changes to a market system are sustainable, or prove to be resilient will only become clear over time. As a result, evaluation designs need to take account of the likely pace of change. In some cases this may mean returning to evaluate impacts several years down the line.

- The more successfully a programme stimulates market system change, the greater the extent to which the changes are owned and sustained by large numbers of market actors. This makes it increasingly difficult to establish causality. It will often be

difficult to quantify the contribution to overall change of an individual intervention. It may not be possible to do so for a portfolio of interventions i.e. the programme as a whole, though some assessment of the scale of effect (negligible, minor, significant etc.) should be possible with the use of an appropriate evaluation design. (see chapters 3 and 4.)

- As systemic change in itself is not the final objective of market systems programmes, evaluations will also need to assess how this has led to growth in employment or incomes for poor people, to improvements in access to key services, and to poverty reduction.

- Any assessment of impact needs to be based on claims about the mechanisms by which change has occurred. This focus on the mechanisms of change illustrates the need for evaluations to be explicitly based on theories of change. The next chapter discusses this in more detail, and explains how a theory-based evaluation can be carried out.

Examples of 'non-linearity' of effects in market systems

Examples of large programmes leaving no significant impacts behind are unfortunately common in international development. There are often multiple reasons for this, but failing to appreciate the complexities of local environments, and trying to transplant 'best-practice' from a completely different context are common factors.

There are fewer examples of small interventions transforming market systems, but the case of M-Pesa is a good one. M-Pesa is a mobile phone based service that allows users to deposit money into an account stored on their phones, send balances using text messages to other users, and redeem deposits of regular money.

Initially developed with a grant from the UK's Department for International Development (DFID), M-Pesa was piloted in 2005, and launched as a commercial product in Kenya in 2007. In 2014 it had more than 12 million users, and was used by more than half of Kenya adults.

It has also been launched in several other countries. M-Pesa also illustrates why it is difficult to attribute impact to a single intervention: while DFID's grant was important in developing early versions of the product, the parts played by many other actors were essential to its success.

2.6 The need for adaptive programming, and the implications for evaluation

The dynamic and unpredictable nature of market systems means that there is no blueprint for how to successfully facilitate improvements in the way they function. Experimentation, learning-as-you-go, and adaptive programme management are therefore all tactics for successful programmes.

There are implications both for monitoring activities (as discussed in detail in the BEAM monitoring guidance), and for evaluation strategies. These relate to:

- How evaluations are conducted. Understanding the theory of change, and how it has been modified over time is a key task.

- How field research is carried out. There is a risk that baseline data may become obsolete if the focus or geographical coverage of an intervention shifts.

- How evaluation results are used. Evaluation findings need to provide timely information as a tool to help programmes adapt, rather than providing ex-post information for the design at a later date of the next programme cycle.

Further reading:
Wach, E. (2015), 'Towards better evidence for market systems initiatives'.
https://beamexchange.org/resources/666/

Adaptive programming – an example from Nepal

The SAMARTH Nepal Market Development Programme includes a broad set of interventions aimed to catalyse change in smallholder milk production and marketing systems. A review at the end of the first year highlighted the fact that the original programme design had focused on the formal distribution systems used mainly by large dairy cooperatives (milk booths based in markets), and neglected the more informal channels that were in fact used by most poor producers. As a result, the project team paused the intervention completely, and then undertook further analysis in order to re-focus the project to achieve greater impact.

2.7 Principles for strengthening the rigour of evaluations

The implications for the key issues reviewed in this chapter can be summarised in a set of principles developed in a separate review of evidence for market systems programmes. This proposes that for evaluations to be rigorous and credible, they should:

1. Use the theory of change to inquire into specific areas of interest.

2. Capture the complex system changes of people living in poverty.

3. Incorporate and build on the plurality of perspectives, experiences and values of beneficiaries and non-beneficiaries, with a particular focus on those living in poverty.

4. Capture change beyond what the programme team or evaluators may have anticipated. By incorporating a wide range of perspectives (point 3), and a focus on the wider systems which affect people living in poverty (point 2), assessments will be better able to identify these unexpected impacts.

3. THE IMPORTANCE OF THEORY-BASED EVALUATION

3.1 Overview of this chapter

This chapter looks at theory-based evaluation, the overarching evaluation design advocated by these guidelines. It starts by explaining what theory-based evaluation is and explains why this design should be adopted. It then explains in outline how an evaluation can be based on the key characteristics of a theory of change for market systems programmes.

3.2 What is theory-based evaluation?

While different definitions of theory-based evaluations exist, they all have in common the idea that evaluations should explore the causal links in a programme theory, testing its underlying hypotheses.

They contrast, therefore, with evaluation approaches that look solely at outcomes (for example those that use experimental methods). A theory-based approach to evaluation can help establish whether the linkages between interventions and intended impacts are plausible, account for other contributory factors, and also capture unintended effects.

Core features of a theory-based evaluation approach include:

- Having two key elements: a conceptual element (developing a theory of change, or causal model, and using this to guide the evaluation); and an empirical element (testing the causal model to investigate how the programme caused intended or observed outcomes).
- Understanding the transformational relations between treatment and outcomes, as well as contextual factors.
- Opening up the 'black box' to answer not simply the question of what works, but also why and how it worked.

Further reading:
Carter, B. (2012), 'Theory-based evaluation approach', http://r4d.dfid.gov.uk/Output/192576/

Stame, N. (2004), 'Theory-based evaluation and varieties of complexity', http://www.stes-apes.med.ulg.ac.be/Documents_electroniques/EVA/EVA-GEN/ELE%20EVA-GEN%207360.pdf

Theory-based evaluation is neutral with regard to methods; which of these are chosen will depend on the specific programme and context, as discussed in Chapters 4 and 5.

3.3 Why use theory-based evaluation?

There are two fundamental reasons for advocating the use of theory-based evaluation as the key design for market systems programmes.

Firstly, market systems programmes seek to stimulate change at multiple levels (market system, growth/improved access, and poverty reduction). No single evaluation method will be appropriate for assessing change at all of these levels. A theory-based approach provides a coherent framework within which different parts of the causal chain can be observed and empirically tested.

Secondly, an important aspect of market systems evaluations is understanding why and how things work. Yields for particular crops might vary because an intervention has helped farmers understand how to reduce the occurrence of a particular plant disease. However, a number of factors, including weather or the cost of inputs such as fertilisers might also be important. The complexity of market systems means that it is necessary to understand how a programme intervention has interacted with other factors to achieve change. Theory-based evaluation provides the basis for doing this.

3.4 Developing a theory of change

Developing a theory of change is the common basis for theory-based evaluations. This theory of change sets out a hypothesis explaining how the interventions that make up a programme are expected to achieve their desired outcomes. A theory of change therefore provides the starting point for defining the questions that an evaluation seeks to answer and the methods that are appropriate for answering them. Theories of change are usually presented in a single document, with the following elements:

- A description of the long-term goal that the programme seeks to achieve.
- An explanation of the programme's context.

- An explanation of the sequence of change that is expected to achieve that goal.
- A description of the assumptions between each link in the sequence of change.
- A conceptual diagram which outlines the key causal links from programme activities through to impacts.

Further reading:
The DCED Standard provides a framework to help practitioners articulate the theory of change underlying an intervention, and to systematically set and monitor indicators which show whether events are occurring as expected. Guidance documents for adopting the Standard provide widely used tools which can help with this task. http://enterprise-development.org/page/introduction-standard

Theories of change and log frames come from the same conceptual family and both are used by development programmes. It is important therefore to bear this difference in mind.

The strategy for any market systems programme should be based on identifying where the development of a market system is likely to make the most difference to the target population group.

The theory of change should set out programme thinking on how this is expected to happen. As such, it will incorporate and explain the links between each of the following levels:

- Interventions.
- Expected changes in the wider market system
- Expected outcomes such as economic growth or improved access to particular services.
- Poverty reduction.

Diagram 7 (overleaf) sets out a simplified theory of change diagram for a horticulture intervention from the Business Opportunities & Support Services (B0SS) project[4]

It is important to note that while a well-constructed theory of change diagram can be helpful in

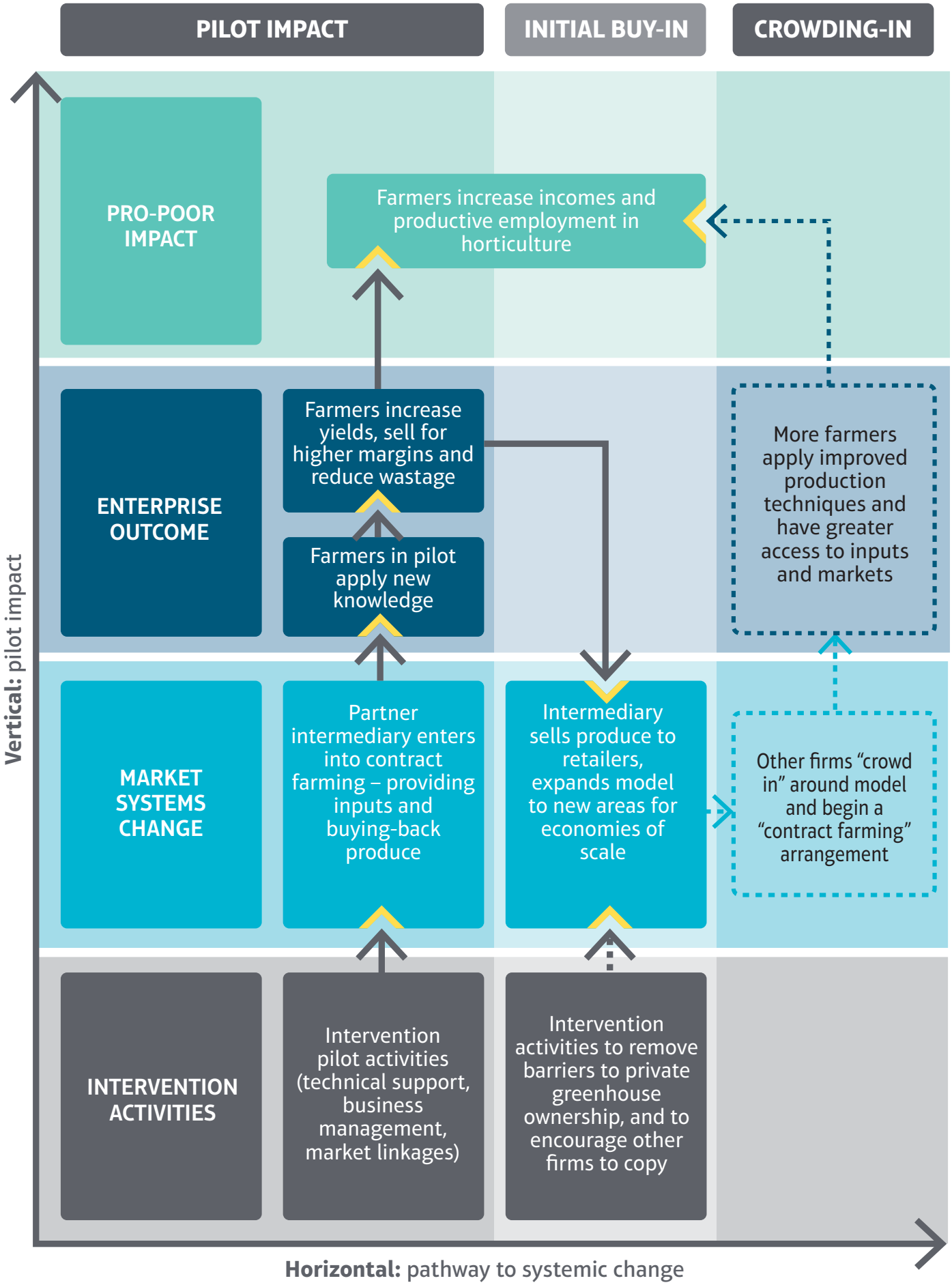
4 Source: Ripley, M. and Major, A. (2015) The BOSS project in Timor-Leste: thin markets, thick impact? https://beamexchange.org/resources/393/

Further reading
Tools for Development: "Theory of Change vs Logical Framework - what's the difference?" http;//www.tools4dev.org/resources/theory-of-change-vs-logical-framework-whats-the-difference-in-practice

Table 2: What is the difference between a theory of change and a log frame?

THEORY OF CHANGE	LOG FRAME
Gives the big picture, including issues related to the environment	Gives a detailed description, showing the progression from activities through to outputs, outcomes and impacts
Could be used to complete the sentence, "If we do X then Y will change because…"	Could be used to complete the sentence "We plan to do X which will give Y result
Is presented as a flexible diagram with narrative text	Is normally presented in a matrix format
Describes how one activity or event will lead to another, detailing the assumptions involved	Specifies the activities that a programme will carry out and the outputs and outcomes that will be achieved
Is mainly used as a tool for programme design and evaluation	Is mainly used as a tool for monitoring and for contract compliance (e.g. for a donor to check that the organisation implementing the programme is doing their task adequately)

Diagram 7: Simplified theory of change from the BOSS project [4]



Source: The BOSS project in Timor-Leste: thin markets, thick impact?, ILO

clarifying and summarising how a programme seeks to make an impact, it is not the only element.

The theory of change must also clearly define the market system to be changed (e.g. the sectors, geographies, common processes or organisations) as well as the assumptions regarding the ways in which systemic change will be realised.

The process leading to the articulation of a theory of change is important. Given that the theory of change is the key framework against which an evaluation is designed, it is critical that the evaluator is not left exposed by using one that does not fully reflect the reality of the programme.

Where a theory of change already exists, it is therefore important that the evaluator reviews this with the programme team - both to support and quality assure the process and to ensure that it is evaluable (see Chapter 4).

Where a theory of change has only partly been described or not set out in explicit terms at all, the evaluator should work collaboratively with the programme team to do this.

Working through a group process should help to create a shared perspective regarding the nature of the programme, and how it is expected to lead to impact, including identification of various intermediate steps, the roles of other actors and the contextual factors required for the programme to work. A further point is that it will be easier to use empirical evidence to test the theory of change if it was developed on the basis of research and evidence in the first place.

Further reading:
The steps in developing a theory of change is also described more fully in the Monitoring Guidance on the BEAM Exchange website. This site also provides links to further examples from existing market systems programmes. https://beamexchange.org/guidance/monitoring-overview/

Vogel, I. (2012), 'Review of the use of "Theory of Change" in international development', https://beamexchange.org/resources/350/

3.5 Characteristics of a theory of change for market systems programmes

The complexity of market systems has important implications for the way a theory of change is developed. In particular, a simple linear progression from inputs to outcomes is unlikely to occur and it will also be difficult to fully understand cause and effect at a system-wide level (at least in advance).

The theory of change therefore needs to be regularly updated and reviewed to reflect changes in our understanding on how the market system works, as well as changes in the market system itself. The knowledge to undertake this will generally come with experiences gained from programme implementation, making it important to involve programme staff, programme partners and other stakeholders in the process.

Furthermore, given the uncertainties about how the overall market system may be re-oriented to better serve people living in poverty, the identification of programme interventions will be based on initial hypotheses about the mechanisms by which change might be achieved.

It is important however to recognise that change may not happen in this way, and that some interventions will need to be adapted, terminated or scaled up during the course of implementation.

Experience using a theory of change approach to accommodate complexity and the need for adaptive management are relatively new. Some argue, for instance, that the whole notion of setting out a clear theory is counter to the concepts of being adaptive and flexible.

There is still much to learn in this field, but for the present it is recommended that theories of change are designed in a way that embrace complexity and uncertainty. Features of such a 'complexity-aware' theory of change include the following:

- In complex systems like markets, different stakeholders will have different perspectives and interpretations about what makes things work. These contexts may not be amenable to analysis with a single model. It makes better sense, therefore, to build up a picture of the context from the ground upwards, involving a full range of stakeholders and working in a consultative manner.

- Where different views about the way forward only partially coincide, or conflict, it is better to record these different views and then revisit the issue as implementation progresses, and information from the field becomes available to clarify matters.

- There is sense in setting out the theory of change in the form of a diagram. Given the uncertainties about precisely which mechanisms will work however, it makes sense to do this at a 'high level' which does not specify the full detail of the intervention (see Diagram 7 for a good example).

- In any market system programme, there will be 'unknown unknowns' which require an adaptive, learn-as-you-go approach. In conditions of significant uncertainty, it may make sense for programmes to include a range of exploratory interventions that can be scaled up, or brought to an end.

- The implementation plan for an intervention should build in short planning horizons. The time frame for such horizons may vary, depending on an initial estimate about how long it should take the planned intervention to achieve change. Where there is considerable uncertainty, the horizon should be shorter.

- Due to the adaptive nature of market systems programmes, the theory of change should be reviewed and adapted regularly to reflect emerging findings, changing hypotheses, and adjustments to programme strategy. A good practical description of how this has been undertaken is provided in this blog: http://oxfamblogs.org/fp2p/of-sasquatches-and-flexible-programming-a-genuine-sighting/

Reviewing the theory of change in light of new evidence: promoting the use of artificial insemination in Kenya

The Kenya Market Assistance Programme (MAP) is a seven year programme investing £23m (US$37m) to promote better incomes for poor people through interventions in the cotton, water, agriculture, dairy, aqua-culture and media sectors.

One of its projects has promoted the uptake of artificial insemination services for livestock. Initial analysis suggested high prices were the barrier to increased uptake of these services.

However, over time it became clear that lack of transparency and honesty among sellers of services was a bigger issue. This illustrates the need to be ready to revise the theory of change, and not to stick rigidly to output indicators (in this case, prices) once understanding of a market system improves.

Further reading:
Hummelbrunner, R., and Jones, H. (2013), 'A guide for planning and strategy development in the face of complexity', https://beamexchange.org/resources/235/
Valters, C. (2015), 'Theories of change: time for a radical approach to learning in development', http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9835.pdf

3.6 How to conduct a theory-based evaluation

Market systems evaluation should start by clarifying the theory of change and checking that it adequately describes the expected pathway to change, how programme activities will combine with other factors to achieve change, and the associated assumptions.

The next step is to undertake the empirical element of the evaluation. This involves investigating the links between each stage of the theory through field research and gathering evidence to decide whether the intervention is having the planned impact or not. Critically, the investigation should focus on why observed effects are taking place as planned, and whether this suggests that the theory is broadly correct.

Returning to Diagram 7 (page 14), evaluating the theory of change for this project would involve:

- Seeking to understand how and why technical support provided to partner intermediaries (described as an 'intervention activity' in the diagram) has stimulated the growth of new contracting arrangements with partners (an anticipated and desired change at the market systems level).

- Reviewing outcomes to see not only if farmers' yields, margins or wastage levels have improved, but also why this is the case (i.e. change at the level described as an 'enterprise outcome' in the diagram).

This raises the question of which evaluation designs and methods might be appropriate to undertake this assessment, which is covered in the next two chapters.

4. DECIDING THE DESIGN OF THE EVALUATION

4.1 Overview of this chapter

An evaluation of a market systems programme needs to start by reviewing the hypotheses about how impacts will be achieved, and then consider how evidence will be gathered to test these hypotheses. However, there are different ways in which evaluations can undertake this, and different combinations of evaluation designs and methods are possible.

This is quite a complicated subject, so the discussion has been divided into two chapters. This chapter discusses how to choose an evaluation design (or designs). The following chapter then considers the analytical methods that can be applied, and explains how the overall choice of designs and methods can be summarised in an evaluation matrix.

Under current practice, the questions covered in this chapter will often be decided by evaluation commissioners prior to the formal tender process. It is then left to evaluators to propose methods and how to apply these in practice (the subject of the next chapter). This runs the risk however of a mismatch in expectations between commissioners and evaluators, and a less coherent evaluation strategy. It is advisable therefore, for decisions on design, methods and the implementation plan to be made through an iterative process involving both the commissioners and evaluators.

One way this might be achieved is through an extended inception process, where the first inception stage is used to help draft or to revise the terms of reference for the following stage. The respective roles of commissioners and evaluators are discussed further in Chapter 6.

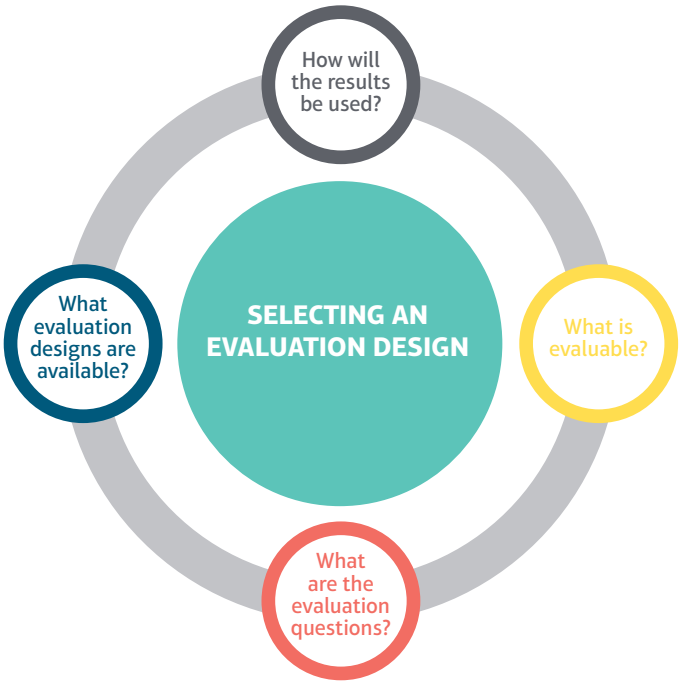
4.2 Issues to consider in evaluation design

It is important to think through evaluation design carefully and not assume in advance that any particular design is appropriate. If this stage is neglected, it will have negative consequences down the line for the relevance, validity and usability of evaluation outputs. Key issues to take into account in deciding the design include:

- How will the results of the evaluation be used?
- Can all elements of the programme be evaluated, or is it necessary or desirable to focus on particular issues or interventions?
- What are the key questions that the evaluation needs to answer?
- What designs are appropriate for evaluating an individual intervention, and for evaluating a market systems programme as a whole?
- The key issues for deciding on a design are summarised in the diagram below, and discussed in detail in the rest of this chapter.

Diagram 7: Issues to consider in selecting an evaluation design[5]

5 Based on a diagram in Stern, E. *et al,* "Broadening the range of decisions and methods for impact evaluations" 2012 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf



4.3 How will the results of the evaluation be used, and by whom?

In broad terms, external evaluation provides some combination of a 'proving' and an 'improving' function. It is important to be clear on the relative importance of each, as this will impact on the evaluation design and where resources should be focused.

Further reading:
Barnett, C. (2015), 'An evaluator's perspective: imbalances of power in evaluation design and implementation', http://europeanevaluation.org/sites/default/files/ees_newsletter/ees-newsletter-2015-12-december-r19-web.pdf

Evaluations that focus on 'accountability' / 'proving' seek to inform decisions for continuing, scaling up, or replicating programmes; and build knowledge on how and why market systems change over time. Evaluations focused on 'proving' results are typically used for accountability and cost-effectiveness reasons - they seek to provide proof of the contribution that a programme has made to observed results.

Evaluations that focus on 'learning' / 'improving' require a greater emphasis on the processes and changes in the market system (rules, incentives, relationships, behaviours, capacity, etc.) than on final impacts at the household or enterprise level. Evaluations that provide real-time information to facilitate adaptive management can help to contribute to improved programme performance.

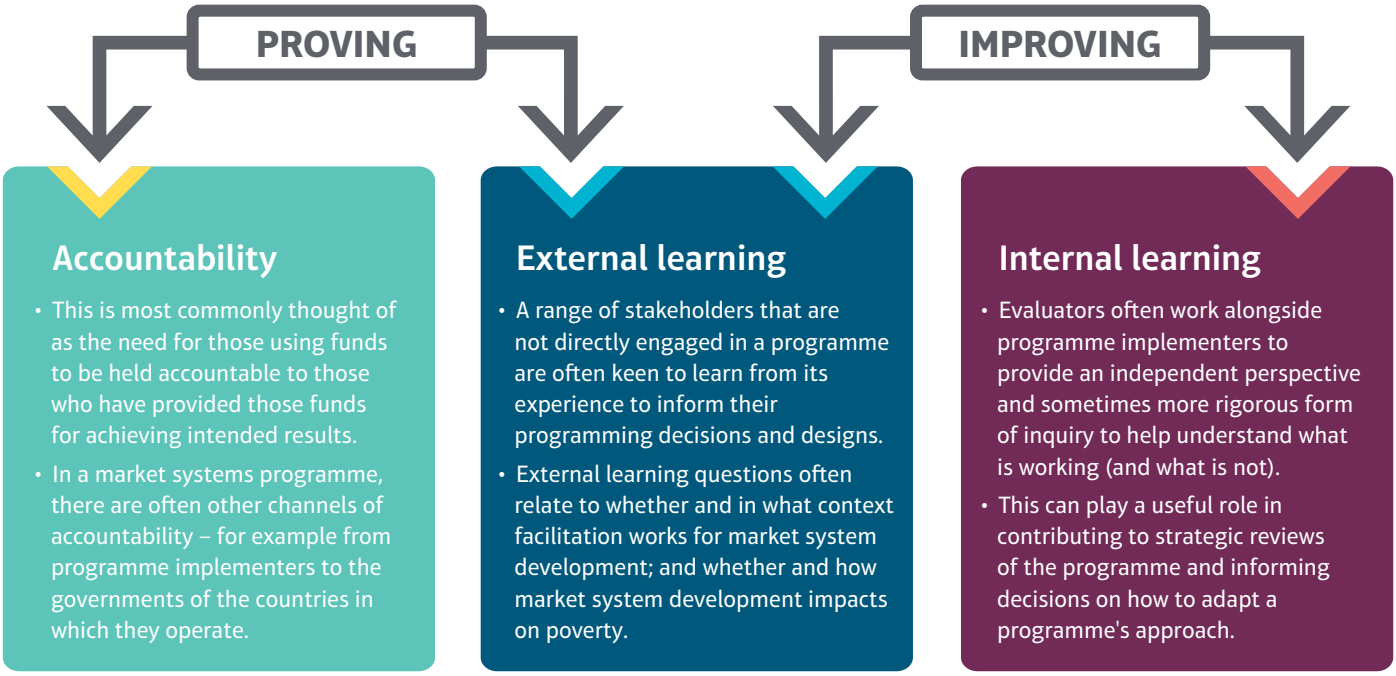
While evidence for this is important, the burden of proof may be lower than it would be if the focus was on 'accountability'. Generating information that is timely and can therefore influence investment decisions during implementation is emphasised, compared to 'accountability' focused evaluations.
These distinctions are summarised in the diagram below:

An important early task in designing an evaluation is therefore to identify the people who are intended to use the evaluation results, and how they will use them. Taking the time to consult them and making sure that their interests are reflected in the evaluation questions and other aspects of the design will help to ensure the evaluation achieves its overall aims.

Producing a matrix which maps evaluation users and uses can be helpful to summarise the results of these discussions, and therefore to draw out the implications for the design of the evaluation.

Evaluation users normally can be divided into three categories: funders, facilitators and market actors (including government and other public sector bodies). Table 3 (overleaf) provides some illustrative examples of how different users might need to use the evaluation results. In practice the 'market actors' category might include quite a diverse range of organisations and therefore need to be broken down in more detail.
The results of this exercise could have various

Diagram 8: Different uses for an evaluation



Source: BEAM Exchange

Table 3: Mapping evaluation users and uses

Evaluation user	What the evaluation might need to be used for	
	Proving / accountability	**Improving / learning**
Funder	• Hold programme team to account for results achieved	• Understand what has been achieved, and why • Inform design of a follow-up phase or the replication of the programme • Share learning with other offices and teams
Programme team	• Demonstrate to funder that they have done a good job • Holds partners to account for achieving the results for which they have been contracted	• Understand how to adapt the programme to achieve better results
Market actors	• Hold programme team accountable for offers/promises made • Have independent verification that the innovations they have piloted work well	• Understand what is happening in the market system • Understand if new business models are viable, and produce better results

implications. The exercise might, for instance:

- Highlight the need to focus on particular interventions within the programme.
- Help decide how to communicate with different stakeholders about the evaluation process and the results it produces.
- Help identify contentious issues at the outset, and what level of evidence will be required for the evaluation results to be considered credible and useful.

Discussing the evaluation with stakeholders at planning stage should also provide an opportunity to secure their 'buy-in'. This may help with various practical aspects of the work, including arranging interviews, collecting data and interpreting the results of fieldwork.

A further important step in planning an evaluation, carrying out an evaluability assessment (see below), also requires discussions with stakeholders. Where this is the case, it will make sense to combine these consultations into a single process.

Further reading:
Better Evaluation, Understand and engage stakeholders', http://betterevaluation.org/plan/manage/identify_engage_users

4.4 Which elements of the programme should be evaluated?

In an ideal world, each individual intervention within a market systems programme would be fully evaluated. In practice, resource and time constraints may require a choice to be made, raising the issue of which interventions to focus on.

The programme team should therefore develop an initial list of interventions that the evaluation will cover, taking into account the interests of stakeholders. Reasons to focus on particular interventions might include the fact that they:

- Account for a significant proportion of overall programme resources.
- Are considered likely to have made a significant impact on the way that the market system operates.
- Are particularly innovative.
- Are considered to be critical to the effectiveness of the programme as a whole.

In addition, evaluations will normally need to review the overall programme, to assess for instance if its different interventions add up to a

coherent whole, and whether they work in synergy with each other.

Once the programme areas to focus on have been identified, it is important to make sure that it is feasible to evaluate them. This task is typically carried out in development programmes by undertaking an evaluability assessment.

'Evaluability' is defined by the OECD DAC as "the extent to which an activity or project can be evaluated in a reliable and credible fashion." An evaluability assessment therefore needs to consider:

- Whether a project or programme can be evaluated in principle. It may not be possible to do so, for instance, if the theory of change is incoherent.

- Whether an evaluation is possible in practice. A central issue is whether the data required to make evaluative judgements can be obtained.

The results should provide the information to confirm which interventions or elements of the overall programme can be evaluated. In turn, this should also help to guide decisions on evaluation design, as discussed below. In summary terms, carrying out an evaluability assessment involves the following:

- Agree what the evaluability assessment will focus on, using an evaluability checklist.

- Clarify and define the programme theory of change, and the logic for those individual interventions that are of particular interest.

- Identify what data is already available from stakeholders, or obtainable through other means such as surveys.

- Interview the main stakeholders. This should include a review of the theory of change and a discussion about what data the stakeholder might be able to provide to the evaluation.

- Make recommendations in a written report, reviewing how progress towards impact can be assessed, and which assumptions in the theory of change are most in need of objective verification.

It is recommended, though not essential, that the evaluability assessment be carried out by an independent third party. The time required may vary from a few days to a few weeks, depending on the complexity of the programme. The financial resources dedicated to the task should also be proportionate to the overall evaluation budget.

Although it is good practice to carry out an evaluability assessment, in reality not all programmes have always done so. Instead they commission an evaluation, and then ask evaluators to undertake most of the work for the steps described above during its inception phase. However, doing this risks a situation where evaluators are contracted to conduct an evaluation which, in practice, can only be undertaken in part or not at all.

Carrying out a full and independent evaluability assessment as described above is therefore recommended. If this is not to be undertaken, the programme team should, at a bare minimum, carry out its own internal assessment of whether the programme can be evaluated, and ensure that a coherent theory of change can be articulated.

Further reading:
This document provides detailed guidance on what an evaluability assessments is, and how to undertake one:

Davies, R. (2013), 'Planning evaluability assessments: a synthesis of the literature with recommendations', https://beamexchange.org/resources/338/

This note (extracted from the 2013 report) provides a useful checklist for an assessment: Davies, R. (2015), 'An evaluability assessment checklist', http://betterevaluation.org/sites/default/files/An%20Evaluability%20Assessment%20checklist.doc

4.5 What questions does the evaluation need to answer?

Undertaking the tasks described so far will clarify how the evaluation will be used, define which parts of the programme to focus on, define the theory of change, and confirm the that the proposed scope is feasible. The next step is to decide the specific questions that the evaluation will answer.

In general terms, there are four broad questions that a theory-based evaluation can help to answer:

- Did the intervention make a difference?

- What specific contribution did the intervention make?
- How was the difference made?
- Can the intervention be expected to produce similar results elsewhere?

These broad issues need to be translated into a list of specific evaluation questions which relate to the content of the programme and its theory of change.

Further reading:
Better Evaluation, 'How to specify key evaluation questions',
http://betterevaluation.org/plan/engage_frame/decide_evaluation_question

The box below presents examples of questions from the evaluation of a programme that aimed to increase the access to financial services of men and women living in poverty and micro, small and medium-size enterprise in a country in Africa.

In this case, the questions were posed to identify what had changed a) at the level of the interventions b) in the market system c) the extent to which target groups could access financial services and d) the extent to which the programme had impacted on poverty.

EVALUATION QUESTIONS FOR A FINANCIAL SECTOR DEEPENING PROGRAMME IN AFRICA

Specific evaluation questions used for the impact evaluation, mapped against different levels of the financial services market system, included the following:

LEVEL	EVALUATION QUESTION
Interventions	• Have individual programme interventions helped service providers successfully identify and meet effective demand for financial services, and if so, how?
Sysytemic effects	• Has the programme's work to help pilot partners develop new financial products also encouraged their competitors to follow suit?
	• What have been the mechanisms for change at the macro, meso and micro levels of the financial services sector?
Access to services	• Has the programme made a difference to financial access for a) poor men and women b) enterprises?
Impact on poverty	• What are the linkages between improving access to financial services, and reductions in poverty?

4.6 Which evaluation designs can answer these questions?

As explained in Chapter 3, the underlying design for the evaluation of a market systems programme should be theory-based. To reprise the argument: the complexity of market systems, the fact that programmes only induce change indirectly, and that they do this through long results chains, means it is important for evaluations to 'open up the black box' and understand how impacts are achieved. A theory-based approach is appropriate for this because of its central focus on causal mechanisms.

Theory-based evaluation designs are "methods-neutral", and it is perfectly normal for them to incorporate a range of methods. It is feasible for instance, for a theory-based evaluation to include statistical analysis of the impacts of an intervention on beneficiaries, along with a survey using qualitative methods to understand how the project contributed to the overall results.

In a market systems programme, a mix of designs is normally useful, with different combinations appropriate for evaluating change and answering questions either at the one or more individual interventions level, or for the whole programme.

Apart from a theory-based evaluation, the main designs useful for impact evaluation are:

- **Statistical:** where there are large numbers of cases – populations, small businesses and so on – with measurable characteristics that can be analysed.
- **Experimental:** where the impacts of an

intervention on one group are compared with those on a similar group who have not participated or been affected by the project.

- **Case-based:** where different cases (or case-studies) are analysed and sets of case characteristics (configurations) are compared in relation to outcomes.

- **Participatory:** where the judgements and experience of stakeholders and beneficiaries are used to identify the most relevant theories of change and meaningful outcomes from among several possibilities.

A key characteristic distinguishing these different designs is their approach to establishing causality, and the extent to which impacts can be quantified. In broad terms:

- Statistical and experimental designs typically make use of an explicit counterfactual, allow outcomes to be directly attributed to an intervention, and also allow impacts to be quantified in a robust manner.

- Case-based and participatory designs use methods which allow evaluators to make a judgement that an intervention has contributed to an outcome, along with other contributory factors. While an assessment may be made relating to the scale of the effect (e.g. strong, moderate, or weak), impacts cannot be quantified.

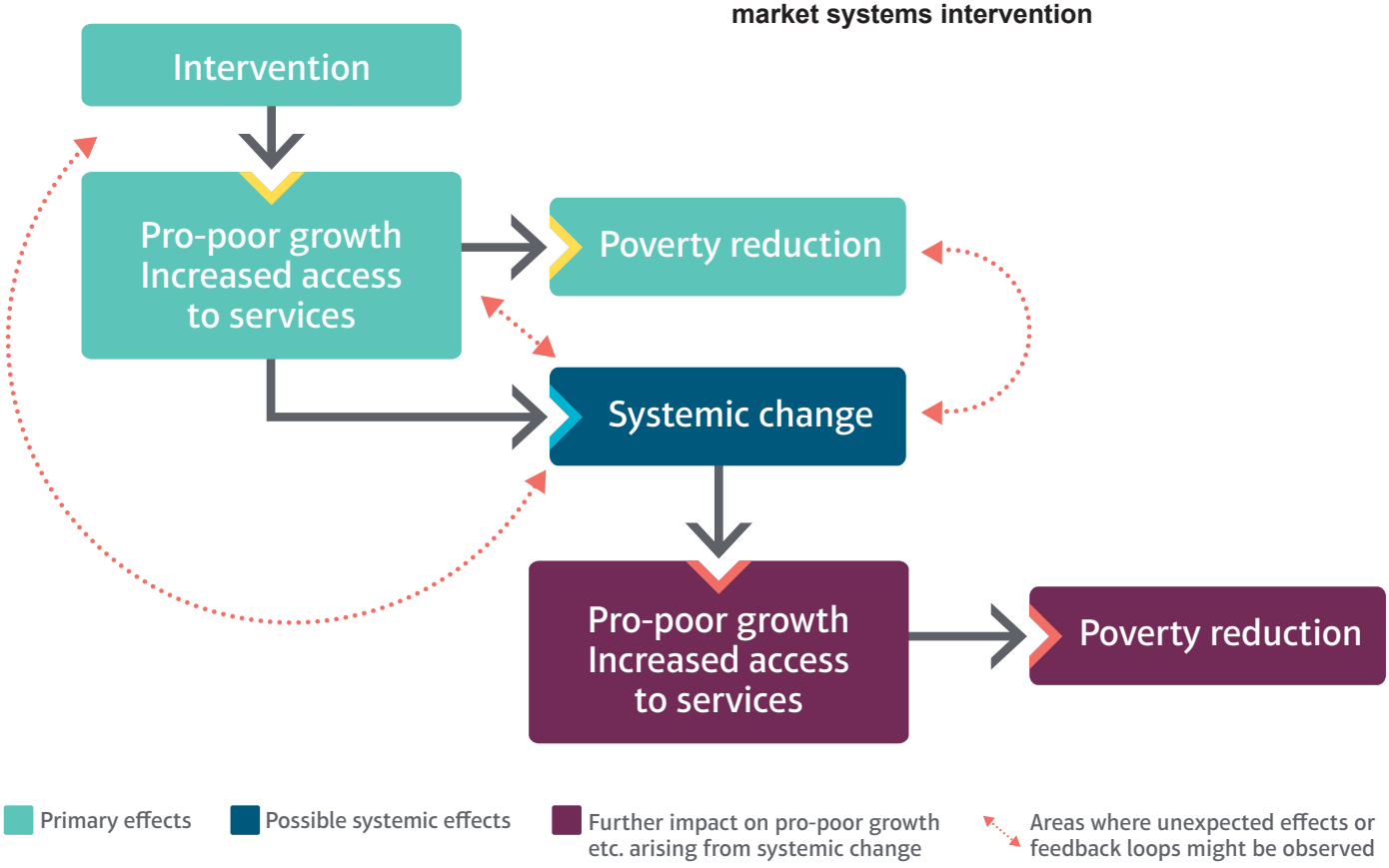
The question of which evaluation designs to use will vary according to what needs to be evaluated. Different options are appropriate if the focus of the evaluation is on an individual intervention (for instance, assessing the results of a pilot scheme), or on the programme as a whole.

4.7 Evaluating an individual intervention

In discussing the available designs for evaluating an individual intervention, it is useful to bear in mind the conceptual diagram from chapter 2 on how individual interventions within a market systems programmes aim to make an impact.

As explained in chapter 2, four broad types of effects are possible for an individual intervention:

Diagram 9: Potenttial impacts for a market systems intervention



Source: BEAM Exchange

Further reading:
This paper describes the different designs that can be used to evaluate impact in more detail.
Stern et al. (2012), 'Broadening the range of designs and methods for impact evaluations',
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf

This paper describes how a theory-based approach can be combined with an experimental and quasi-experimental design, labelling this hybrid as "theory–based impact evaluation."
White, H. (2009), 'Theory-based Impact evaluation: principles and practice',
https://beamexchange.org/resources/224/

- Primary effects (labelled '1' in the table) are the immediate impacts that an intervention seeks to achieve, such as increasing jobs and income for a target group, or enabling that group to access a particular service. These are the kinds of effects that a pilot intervention for a market systems programme will typically be seeking to achieve.

- Where an intervention produces these primary effects (e.g. successfully piloting a new business model), this may also lead to systemic changes (2) (e.g. the adoption of the business model by other market players), which may then produce further impacts (3), for instance, more generalised increase in income or access to services. These are effects that a programme would hope to see once a successful pilot intervention has been scaled up.

- In addition, there is always the possibility that unexpected effects (4) might also arise. These may be positive or negative. Insights from systems theory encourages us to expect these effects to be present, and that it is possible that they are significant, meaning that it is important that an evaluation incorporates a way of identifying these.

The available evaluation designs for assessing each of these effects is summarised in the table below:

Table 4: Evaluation designs for individual interventions

	1.PRIMARY EFFECTS	2. PRIMARY EFFECTS	3. FURTHER IMPACT FROM SYSTEMIC EFFECTS	4. UNEXPECTED EFFECTS
	Increased income? Improved access to services? Helped to reduce poverty?	Contribution to systemic change?	More generalised pro-poor growth, increased access to services, or reduced poverty, arising from systemic changes?	Unexpected effects, including negative ones?
Theory-based	YES	YES	YES	YES
Experimental	YES	NO	NO	NO
Satistical	YES	NO	NO	NO
Participatory	YES	YES	YES	YES
Case-based	YES	NO	NO	NO

Further explanation of the table is provided below:

A theory-based design is required to investigate the causal mechanisms which produce the observed changes. For type 2, 3 or 4 effects, theory-based design will need to include research methods that enable possible systemic and unexpected effects to be investigated. In practice this means using a mix of methods that provide both quantitative and qualitative information to assess such effects.

Where it is feasible to use an experimental or statistical design, this could be incorporated into an overall theory-based design. This might be undertaken, for instance, to quantify the extent to which introducing a new product or service into the market raises the incomes of a target group, and therefore reduces their poverty. However, once a target population has been affected by more than one intervention, it may be difficult to disentangle the effect of each one, meaning that these methods will often be appropriate only at the pilot stage.

Experimental or statistical designs are unlikely to be effective in identifying systemic changes or unexpected outcomes (effects of types 2, 3 or 4). This is because they work by specifying the effect to be investigated in advance, and then measure the different contributory factors for this effect. Where experimental and statistical designs are employed to evaluate a market systems intervention, it therefore makes sense for them to be combined with a theory-based approach using mix methods that can also identify possible systemic changes.

Participatory designs may be helpful to engage a wide selection of market actors, to understand their views on the progress of the intervention as it is delivered. An insight from systems theory is that because people have different perspectives depending on where they are placed within the system, no single perspective is likely to provide an accurate assessment of how the system works or is changing nor is any single perspective necessarily more correct than another one.

As explained in chapter 2, it is important therefore that the evaluation covers a wide range of perspectives, and also involves market actors and other stakeholders in defining the theory of change and the evaluation questions.
Case-based designs may be used as part of a theory-based evaluation. An intervention might provide the basis of a single case, with impacts assessed through the method of Process Tracing for instance (see the next chapter). Results from multiple cases might also be compared if the intervention had been implemented in a number of different locations.

Using quasi-experimental methods to assess the impact of a new tractor-leasing business model in Nigeria

The Propcom Mai-karfi programme in Nigeria piloted an intervention to introduce tractor-leasing, on the assumption that enabling poor farmers to mechanise their tilling and harvesting work would reduce costs and improve crop yields.

The programme team implemented a quasi-experimental study, which was successful in identifying cost savings of between 51 percent and 57 percent for farmers using tractors, compared to a control group.

An attempt to use the same approach to estimate yield improvements was unsuccessful however because of problems in identifying an appropriate control group.

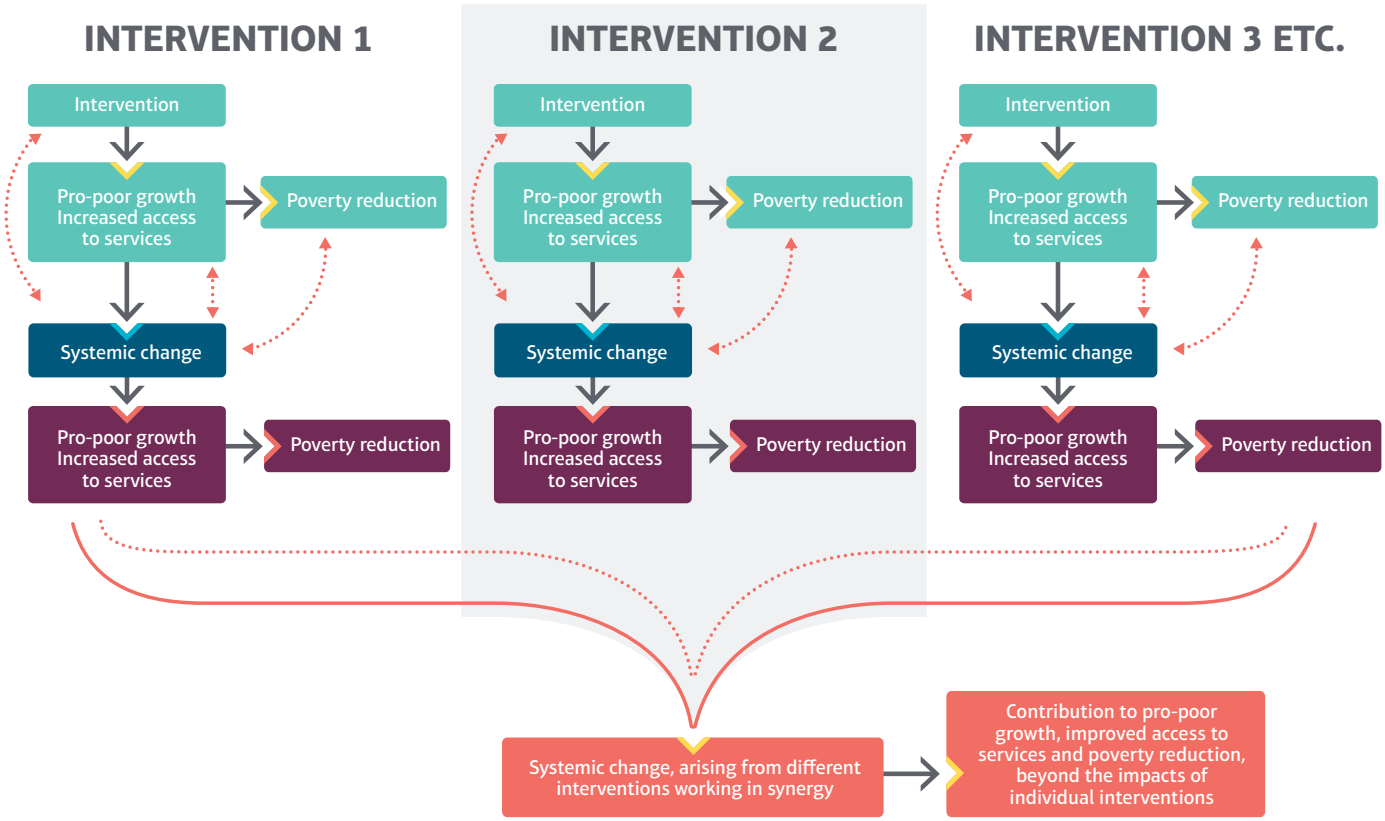
The case study will provide interesting reading for programmes intending to use these methods, not only for its explanation of the methodological issues, but also the practical ones, such as the time and resources involved and difficult working conditions (such as 13 hour days following tractor operators in remote areas).

It also mentions other possible systemic impacts that the experimental method would not have identified, highlighting the need for market systems evaluations to use a mixed methods approach.

These include the question of whether the tractor leasing business model was copied by other market actors, and the fact that tractors leased for use in the intervention area were also used in other geographical areas, raising the question of whether demonstration effects led farmers in other regions to use tractors.

Posthumus, H., and Wanitphon, P. (2015), 'Measuring attribution: Propcom Mai-karfi in Nigeria, using the comparison group method for an intervention in the tractor sector', https://beamexchange.org/resources/645/

Diagram 10: Potential impacts for a market systems programme



Source: BEAM Exchange

4.8 Evaluating the programme as a whole

In discussing the available designs for evaluating an overall market systems programme, it is useful to bear in mind the conceptual diagram from Chapter 2 on how the programme aims to make an impact (above).

The diagram highlights the fact that the evaluation of the programme as a whole will need to take into account:

- The impacts of the programme's interventions (or those selected as being of particular interest to, as discussed above).
- Whether the programme, as a coordinated portfolio of interventions, has led to additional impacts. Such impacts might be where synergies between interventions have been exploited, such that the programme effects adds up to more than the sum of the effects of its parts.
- It is also important to consider if the programme has generated negative synergies, with different individual interventions impeding each other.

Negative synergies in market systems programming

Team members from one market systems programme in Nigeria highlighted how another section of the same programme operated an innovation fund that directly funded market players, creating competition between different interventions within the same programme.

An intervention in a different programme working to encourage businesses to take a commercial approach to provide a paid-for service faced similar problems when another intervention within the same programme started to provide the same service for free.

Source: BEAM Exchange (2015), 'Practitioner's voices: how programmes in Nigeria are applying market systems approaches', https://beamexchange.org/resources/586/

The designs appropriate for evaluating each of these effects for the programme are summarised in table 5 on the next page. Further explanation of the table is provided below:

Table 5: Evaluation designs for individual interventions

	HAS THE PROGRAMME			
	(1.PRIMARY EFFECTS)	**(2. PRIMARY EFFECTS)**	**(3. FURTHER IMPACT FROM SYSTEMIC EFFECTS)**	**(4. UNEXPECTED EFFECTS)**
	Increased income? Improved access to services? Helped to reduce poverty?	Contributed to systemic change?	Led to systemic changes which then had a more general effect in reducing poverty?	Had unexpected effects, including negative ones?
Theory-based	YES	YES	YES	YES
Experimental	NO	NO	NO	NO
Satistical	NO	NO	NO	NO
Participatory	YES	YES	YES	YES
Case-based	YES	YES	YES	YES

At programme level, an evaluation needs to take into account the impacts of individual interventions (type 1 effects). Part of this assessment might be made by combining findings from a series of intervention-level evaluations.

However, a programme level evaluation will also need to consider whether individual interventions have contributed to systemic impacts in their own right, and also whether they have worked in synergy to achieve this (i.e. type 2 and 3 effects).

Experimental or statistical designs may be used to identify the impact of a single intervention, as discussed above. Where such designs have been used to evaluate several individual interventions, the results of these might be aggregated, to illustrate some of the things the programme has achieved.

However, such designs are unlikely to be effective in getting a sense of the overall impact at programme level, for the following reasons.

Firstly, the nature of market systems programmes as a collection of interventions will impede the ability of experimental or systemic approaches to evaluate overall programme impact:

- It is hard to disentangle the effects of simultaneous interventions carried out in parallel. For instance, where a programme implements a project at a local level, but also seeks complementary changes in policy at a national level, it may be difficult to decide which intervention is responsible for observed changes.

- Once a programme has been active for some time in diffusing innovations or new business models throughout the system, it may be difficult to establish a counterfactual with people that have not been affected in one way or another by this activity.

Secondly, experimental and statistical approaches work where the external environment is relatively stable, and interventions are fixed and well defined. As such, they are unlikely to be suitable when searching for and assessing unexpected impacts. An investigative approach, which includes a strong qualitative component, is more

likely to identify these (for instance, interviews with people who are active in the market).

For the same reasons that when evaluating an individual intervention, applying a participatory approach will be helpful to engage a wide selection of market actors and beneficiaries. It will also help to understand the range of programme impacts from different perspectives, including different groups of poor people.

A case-based approach might be incorporated with a theory-based design, assessing a number of different interventions each as a separate case, and using a method such as process tracing to assess causality for each one (see below).

Further reading:
UNEG (2013), 'Impact evaluation in UN agency evaluation systems: guidance on selection, planning and management', http://www.uneval.org/document/detail/1433

5. METHODS, DATA COLLECTION TOOLS AND SOURCES

5.1 Overview of this chapter

In a theory-based evaluation, evidence is gathered to decide whether change is happening through the expected mechanisms, while also scanning for unexpected effects. A range of different methods that work with the designs already described might be used to collect this evidence. This chapter summarises some of the methods that can be used with different designs. A key consideration is also to provide guidance on strengthening the validity of evaluation findings. A review of market systems evaluations found many weaknesses in the reliability and robustness of results[6]. These included small sample sizes,

6 Ruffer, T., and Wach, E. (2013), 'Review of making markets work for the poor (M4P): evaluation methods and approaches', https://beamexchange.org/resources/133/

Table 6: Methods appropriate for use with different evaluation designs[7]

DESIGN	WHAT WILL THE METHODS HELP TO MEASURE?	EXAMPLES OF METHODS	SUMMARY OF METHOD	BASIS FOR ASSESSING CAUSALITY
Experimental	Counterfactuals and the co-presence of causes and effects	Randomised control trail	Assigns participants randomly to a treatment group (who participate in the intervention) or a control group (who do not), and then compares outcomes for each	Attribution
		Difference in difference analysis	Compares changes in an outcome over time between treatment and comparison groups	Attribution
Statistical	Correlation between different variables, allowing the influence of multiple causes on a single effect to be quantified	Longitudinal studies	Identifies changes in quantitative indicators, before and after an intervention	Attribution
		Econometrics	Uses statistical tools such as regression analysis to assess the contribution of different factors to an observed outcome	Attribution
Theory-based	Identification and confirmation of theories of change and results chains, taking into account the context and other factors	Contribution analysis	Sets out to demonstrate a plausible association between a programme and observed outcomes, using the weight of evidence to show each step in the chain between programme inputs and outcomes	Contribution
		Process tracing	A qualitative method that uses probability tests to assess the strength of evidence for specified causal relationships	Contribution
		Outcome harvesting	Evaluators, funders, programme staff and others identify, verify and interpret 'outcomes' where relations of cause and effect are not fully understood	Contribution
Participatory	Validation by participants that the effects they experienced were 'caused' by programme	Ethnographic studies	Participant observation during field research. The ethnographer becomes immersed in the culture and records extensive field notes	Contribution
Case-based	Comparison of causal factors across different cases	Case studies	Examines specific situations and determines the different factors that influence the way the situation developed	Contribution

and little consideration of statistical significance in quantitative studies. Issues for qualitative studies included weak data collection practices and limited attention to possible bias. This chapter therefore suggests practical guidance for addressing these issues.

5.2 Different approaches to assessing causality

The two main classes of approaches to establishing causality for market systems programmes are those based on the principles of attribution, and those based on the principle of contribution.

Approaches that allow impacts to be attributed to an intervention compare specified impacts to a counterfactual, that is, an assessment of what would have otherwise happened. These approaches allow strong causal claims about interventions as the cause of impacts, and the quantifying of effects. However, as explained in chapter 4, in the case of market systems programmes they are likely to be usable only for the evaluation of individual interventions.

Approaches based on the principle of contribution set out to make a plausible argument for causality, identifying outcomes and then tracing the mechanisms through which interventions may have influenced these, while paying careful attention to the context. Evaluations at programme level will certainly need to include methods based on this approach, because they can be applied in an investigatory way to identify the unexpected effects which are to be expected in system-changing initiatives.[7]

5.3 Methods for attributing observed impacts to an intervention

Experimental, quasi-experimental and non-experimental methods potentially offer significant value in providing robust and quantifiable assessment outcomes (e.g. change in performance of a firm), and of impact (e.g. change in poverty levels). However, they will usually also be expensive, requiring large surveys and specialist skills for those implementing and assuring the quality of the overall study.

Randomised Controlled Trials (RCTs) are an experimental research method in which beneficiaries are randomly assigned to either the treatment or control group. While RCTs have been used successfully in some market systems programmes, there are several challenges.

Further reading on RCTs

This blog explains how the Business Innovation Facility (BIF2) programme used an RCT to identify the benefits of a pilot intervention with garment workers in Myanmar.
Harrison, Tom. (2015), 'RCTs and market systems: an opportunity to gain insights', https://beamexchange.org/community/blogs/2015/1/22/tomharrison/

This webpage explains the basic steps in implementing a RCT.
Better Evaluation, 'Randomized controlled trials', http://betterevaluation.org/plan/approach/rct

In particular, in an intervention which pilots a new product or service with the aim that this is then diffused into the wider market, those who benefit from the intervention will not have been selected randomly, but will have selected themselves. In this sense they are qualitatively different from other actors, making it difficult to avoid selection bias.

Further reading on quasi-experimental methods

This brief from UNICEF provides a more technical explanation of when a quasi-experimental approach is appropriate, and some of the practical and ethical limitations in doing so.
White, H. and Sabarwal, S., (2014), 'Quasi-experimental design methods', https://beamexchange.org/resources/608/

In quasi-experimental methods, the two groups are not selected randomly, but by making use of some other characteristic in order to control for observed differences. Difference-in-difference analysis is then used to measure the difference in outcomes for the control and treatment groups. While this approach may be more appropriate than a RCT, the risk of self-selection and other biases will still be present.There are some common challenges in implementing both experimental and quasi-experimental methods in a market systems context.

For example, the objective of encouraging the diffusion of an innovation into the market is inherent in the design of a market systems programme. The implication of this however is that what evaluators term 'spill-over' (when members of the control group are affected by

7 Based on a table in Stern, E., et al. (2012), 'Broadening the range of designs and methods for impact evaluations', https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf

the intervention) is an explicit objective of the programme. This can compromise the results of an experimental design.

If a statistical design comparing post-intervention outcomes for selected groups against baseline data is planned, programme managers and evaluators need to be confident that the baseline will remain relevant (i.e. that programme interventions will not adapt to the extent to which the treatment group changes). It is useful to consider where uncertainty about what might change exists, for instance, in relation to:

- Which sectors to intervene in?
- What type of intervention will be used?
- In what geographical area will the intervention take place?

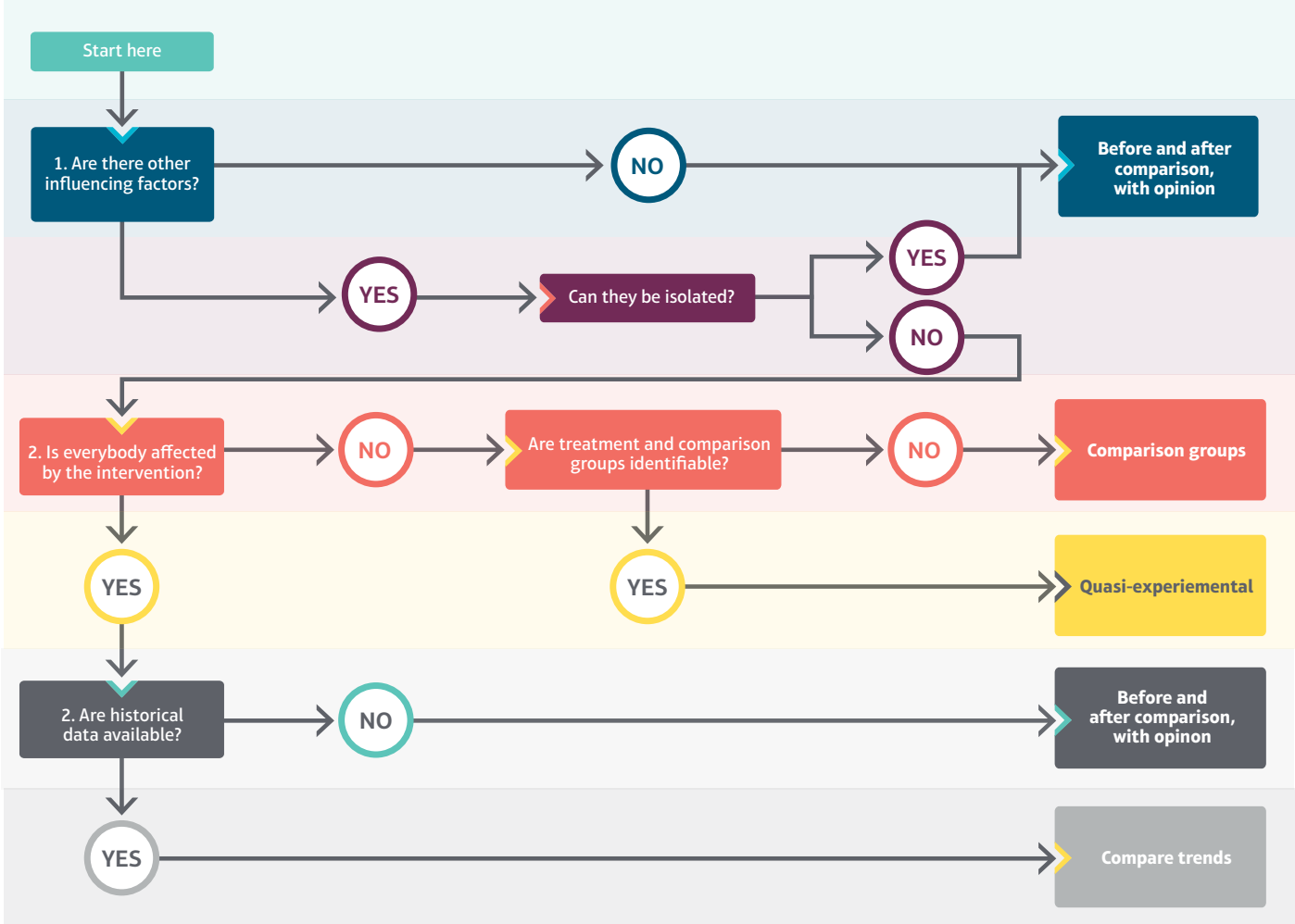
As time goes on and plans begin to become clearer about precisely what the programme will do, each of these elements are likely to become more certain. This should allow the programme team to make an informed guess about when baselines can be researched. For example, the Education Sector Support Programme in Nigeria avoided this problem by not carrying out a baseline until there was a greater degree of certainty about how it would work with local private education providers.

Non-experimental approaches also compare the results of an intervention with a counterfactual. However, these approaches use hypothetical predictions about what would have happened in the absence of the intervention to establish the counterfactual, rather than a control group.

How to choose between different experimental and statistical methods?

This paper provides an overview of different experimental and statistical methods for attributing impact to individual interventions. Posthumus, H. and Wanitphon, P. (2015), 'Measuring attribution: a practical framework to select appropriate attribution methods, with cases form ALCP in Georgia, MDF in East Timor, Propkom Mai-Karfi in Nigeria and Samarth-NMDP in Nepal', https://beamexchange.org/resources/641/

The following diagram from the paper illustrates how to chose methods, based on the characteristics of the intervention and the context in which it is implemented.



Source: Measuring Attribution, Hans Posthumus and Phitcha Wanitpho

Further reading on non-experimental approaches
This paper explains how a non-experimental approach was undertaken in evaluating the impact of leather sector interventions in Ethiopia. The modelling approach embodied in this design specifically aimed to address some of the complexities of systemic change.
Derwisch, S. and Lowe, P. (2015, 'Systems dynamics modelling in industrial development evaluation', https://beamexchange.org/resources/609/

As the name suggests, 'before and after' comparisons look at the situation before and after an intervention has been implemented to assess change. The problem with such approaches is that many factors other than the intervention might be responsible for the change, so this would not be a robust approach except in very simple circumstances (for example where a new crop is grown on lands that were previously fallow).

5.4 Contribution analysis

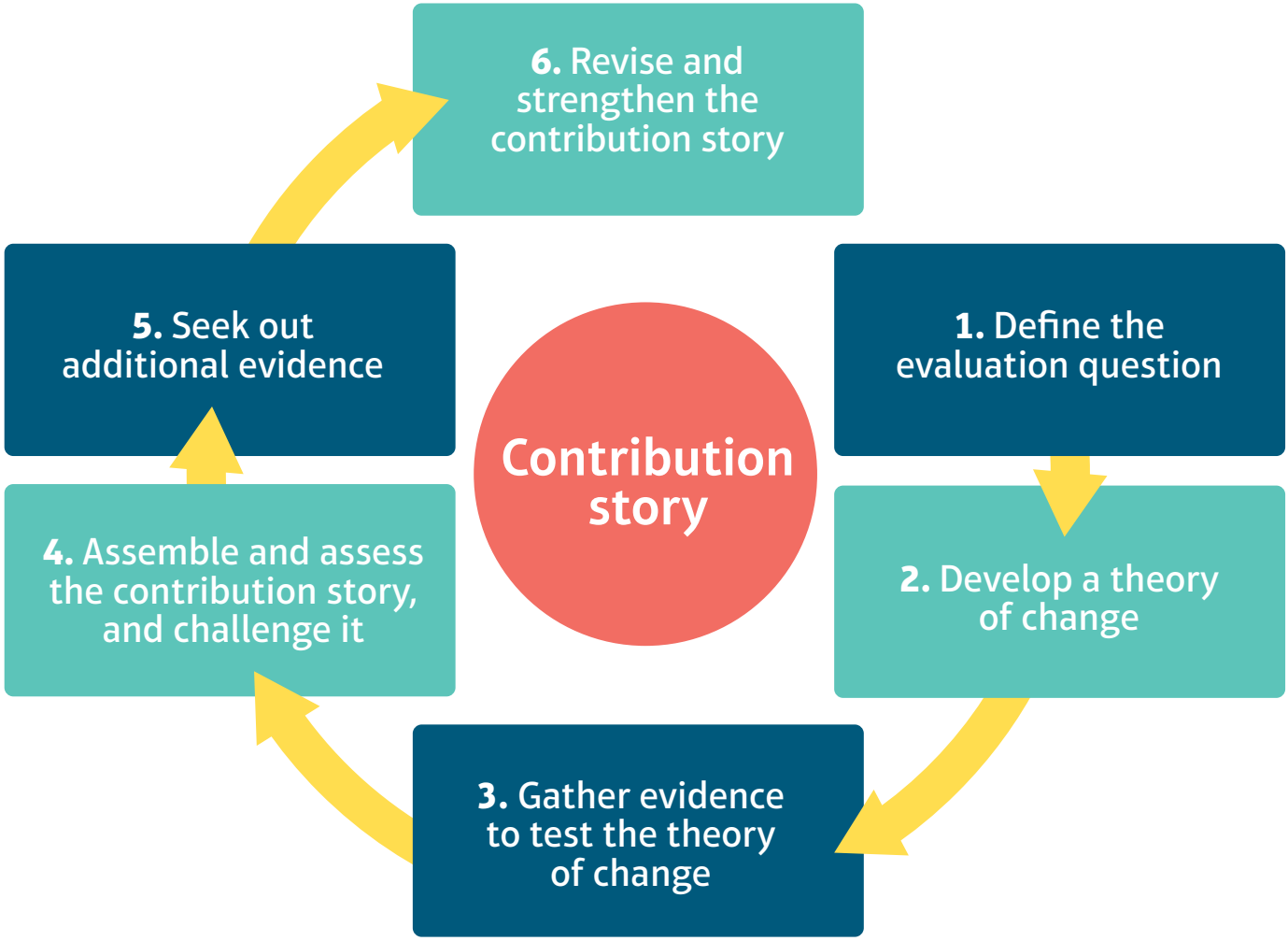
As discussed, methods that allow impact to be quantified and attributed to an intervention are unlikely to work in assessing systemic effects, or in identifying unexpected impacts. Alternative approaches will therefore be required.

Contribution analysis is a broad approach based on the consideration that changes in a market system are produced by several causes at the same time, none of which might be necessary or sufficient for impact. Examining why the intervention (or the programme as a whole) has made a difference will therefore require looking at a combination of causes.

Contribution analysis sets out to identify those outcomes identified by the theory of change. Once outcomes have been identified, the task of the researcher is to marshal and review the evidence in order to explore what contribution an intervention (or the programme as a whole) has made to these.

In each case, researchers need to review a wide range of qualitative and quantitative evidence.

Diagram 11: Steps for contribution analysis



Source: Itad

They should then work backwards from the outcome to decide whether a plausible argument can be made about the contribution of the intervention to the outcome.

An equally important step in the process is to consider other possible causes for the same outcome. The key steps in the approach can therefore be summarised in the diagram below. Steps 1 and 2 of this process should have been carried out in the planning stages of the evaluation (see chapter 4). Steps 3 – 6 involve an iterative process of gathering information and then developing a story about the contribution of the intervention (or lack of one) compared to other factors. A range of sources and data collection methods might be used, to illustrate the links between interventions and observed changes.

Contribution analysis can be used to identify and assess both expected and unexpected outcomes:

- The starting point for reviewing expected outcomes is to identify the evidence that they have occurred (without considering at this moment the issue of causality). This may be established by reviewing some of the secondary data sources described above, or through the use of primary data (e.g. information from surveys of beneficiaries or market actors).

- One strategy to identify unexpected outcomes would be to undertake a number of qualitative interviews or workshops with beneficiaries and market actors (including some who have not been involved in the programme), in order to seek the views of a range of people with in-depth knowledge of different elements of the market system. A purposive (rather than random) sampling strategy might therefore be useful to do this, with initial interviews being used to identify further groups who should be canvassed to help understand how the market system has changed.

- As discussed in chapter 2, the emphasis should be on including a wide set of stake holders, and ensuring that interviews and group discussions are carried out in a way that stimulates the emergence of other issues occurring in the lives of people living in poverty.

Using contribution analysis to evaluate a financial inclusion programme in Kenya

The evaluation of FSD Kenya's financial inclusion programme used a form of contribution analysis to assess the impact of the programme's work to help the Equity Building Society turn itself into a bank. FSD provided technical support for the conversion process, mobilising a high level of professional expertise over time to tackle the legal and administrative issues involved. The conversion process was important for several reasons, including the fact that the new Equity Bank then launched new cheaper bank accounts, which subsequently became the industry norm, thereby expanding financial services significantly for poorer, rural Kenyans.

Through interviews with key informants in the sector, the evaluators were able to develop a view that FSD had indeed made a significant contribution to the conversion process. In the evaluators' judgement, the most likely counterfactual scenario was that Equity would have transformed into a bank without FSD support, but much later on and with greater difficulties. The evaluation used various pieces of evidence to support this view, including the fact that no building society in Kenya had transformed into a bank before, and that the later transformation of a different building society, without FSD's help, took more than twice as long, despite following in Equity's footsteps.

A key point is that contribution analysis was used here to build an argument, using mainly qualitative evidence from key informants. The analogy for this type of evaluative approach is the process in a court of law, where evidence is reviewed and a judgement then made on the basis of whichever competing explanation is most convincing.

5.5 Process tracing

Process tracing follows the same general approach as contribution analysis, but works through a clearly defined method for assessing causality. Where contribution analysis states that 'links should be verified', process tracing indicates in more detail how to go about verifying them.

Process tracing analyses a single case, with emphasis placed on identifying the precise process and temporal sequence of events through which an outcome is achieved. A set of logical tests is then applied to decide whether the proposed explanation is plausible. The steps in the method can be summarised as follows:

- Define the outcome to be investigated.
- Define how the intervention is expected to have caused that outcome.
- Describe the processes or events that link the hypothesised cause and the outcome.
- Apply the logical tests to assess the strength of the argument (these are described in the literature as “straw in the wind”, “smoking gun”, “hoop” and “doubly decisive” tests – see further reading below for details.)

The theory of change or results chain for the programme or intervention should provide the basis for providing answers for the first two steps. If either the outcome or the mechanism through which it was supposed to be achieved is not clear, then the evaluator should work with the programme team to reconstruct the logic for a particular intervention.

The third step requires considering information from a range of sources to describe the process by which the intervention contributed to the outcome. It is important to be clear that one particular outcome might not be the result of one particular intervention but a result of the interplay of different interventions and potentially other changes that happened at the same time.

Further reading:
Punton, M. and Welle, K. (2015), ‘Straws-in-the-wind, hoops and smoking guns: what can process tracing offer to impact evaluation?’, http://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/5997/CDIPracticePaper_10.pdf?sequence=1

Punton, M. and Welle, K. (2015), ‘Applying process tracing in five steps’, http://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/5997/CDIPracticePaper_10.pdf?sequence=1

5.6 Triangulating results using an ‘inwards-out’ and ‘outwards-in’ perspective

Initial conclusions about the extent to which an intervention has had the desired effect might also be strengthened by considering this issue from different perspectives. This can be considered as a form of triangulation (in the sense of using different perspectives or sources of data to support conclusions).

The approach described so far can be described as providing an ‘inwards-out’ perspective. In other words, one which takes as its starting point the intervention or programme and then moves outwards to examine changes in the wider market system. Even with rigorous application of methods it is possible for bias to enter the assessment (in particular, ‘self-importance bias,’ i.e. over-estimating the importance of the programme in producing outcomes, and ‘confirmation bias,’ i.e. being particularly sensitive towards evidence that confirms one’s own hypotheses while ignoring evidence that disproves it).

To gain a sense of the broader context it is also worth using an ‘outwards-in’ perspective, and comparing this with the inwards out conclusions.

An outwards-in approach starts by observing changes at the impact level (changes in income, or poverty) and then considers whether there is a relationship between these and changes in the wider market system, and whether the programme and interventions might have contributed towards these.

The steps in this process (step one: inwards-out; step two: outwards-in, and step three: the drawing of conclusions from both perspectives) are summarised in the diagram overleaf.

5.7 Outcome harvesting

Outcome harvesting is a particularly appropriate method for undertaking ‘outwards-in’ analysis. Under outcome harvesting, evaluators collect evidence of what changed, and then work backwards to determine whether and how an intervention or programme contributed to this change.

A range of data sources can be used including face-to-face interviews or workshops, surveys and written documents. An important proviso is that the participation of those who influenced the outcomes needs to be considered.

Outcome harvesting can be implemented in six steps, each of which can be undertaken iteratively (i.e. as more information is collected it is appropriate to revisit decisions made in earlier stages).

1. **Design the outcome harvest**. Evaluators and users define questions to guide the process, agreeing what information is to be collected.

2. **Gather data, and draft outcome descriptions**. Evaluators search for information on ways that individuals or organisations have changed their behaviour as a result of the intervention. In a market systems context, this might include firms or other market actors that have changed their business practices after an intervention was carried out.

3. **Engage with 'change agents'**. Evaluators talk with individuals or organisations that influenced the outcome to review the draft outcome descriptions and identify additional outcomes. In a market systems context this might include the programme's intervention teams, or intervention partners.

4. **Substantiate the results**. Evaluators obtain the views of independent individuals knowledgeable about the outcomes and how they were achieved. Depending on the context, these might include government officials, staff from major players active in the market, business associations etc.

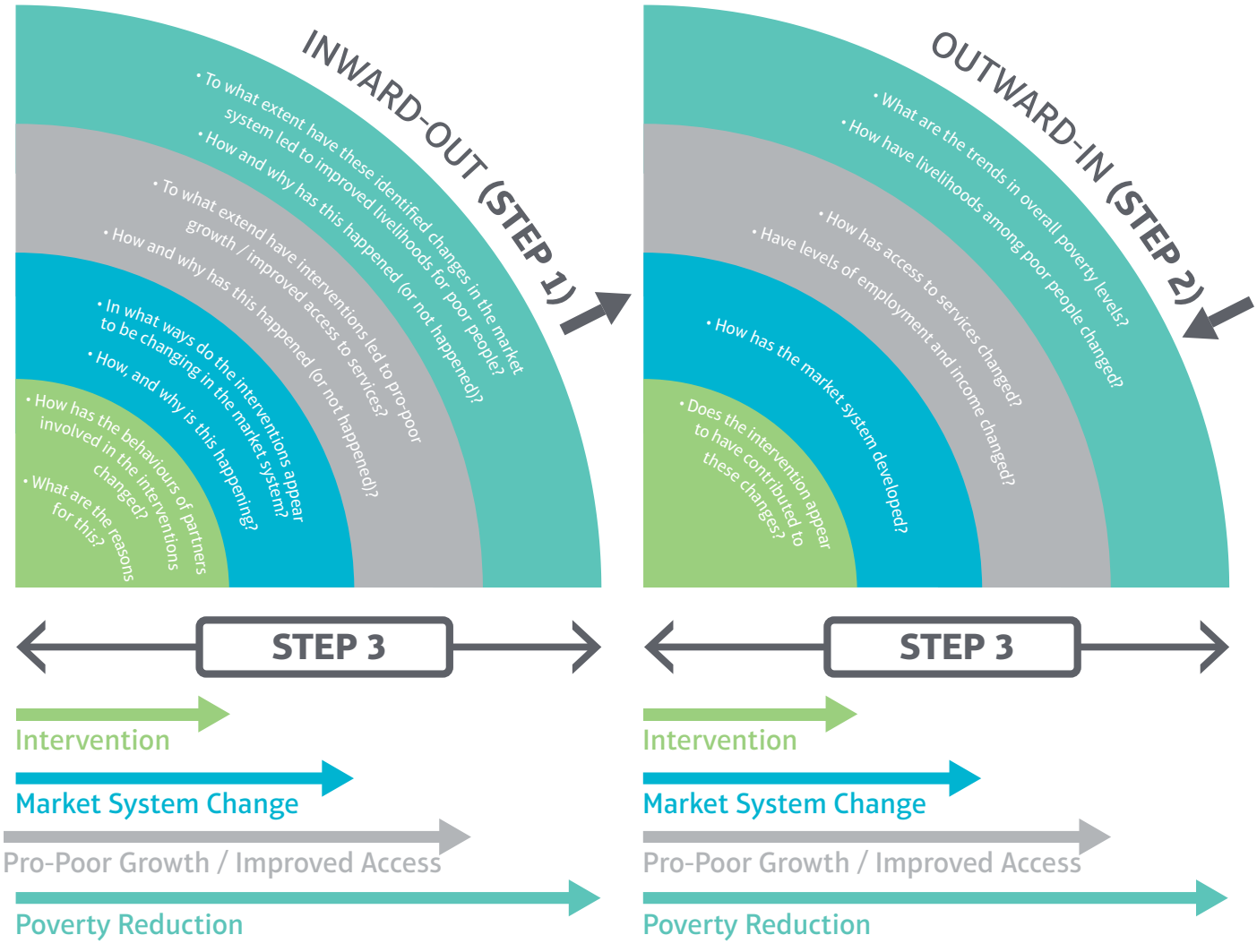
Further reading:
Britt, H., and Wilson-Grau, R. (2012), 'Outcome Harvesting', https://beamexchange.org/resources/610/

5. **Analyse and interpret the results**. Evaluators organise the outcome descriptions in order to make sense of them and develop answers for the questions defined at the outset.

5.8 Bias in quantitative studies

Experimental methods assess impact by comparing the results for similar groups which did, and did not participate in an intervention. The main source of bias in these studies is selection bias, which arises where participants for either the treatment or control groups are selected with a characteristic apart from the actual treatment that has an influence on the outcome. For instance, participants are richer or more entrepreneurial (tending to make the intervention look like

Diagram 12: triangulating results using an 'inwards-out' and 'outwards-in' perspective



Source: Itad

it had more impact than it really did), or are interested in participating due to a lack of success elsewhere (therefore making the results look less favourable).

In theory, random sampling can help to ensure that treatment and control groups are statistically equivalent and that the results are internally valid – i.e. that no other factors other than the intervention caused the observed outcome. However, the fact that market systems interventions usually identify particular partners or groups to take up a new business model or innovation piloted by the programme, means that random selection may not be possible in practice.

Quasi-experimental designs commonly construct a control group by identifying non-participants in the intervention who are matched with those who do participate on the basis of a few observed characteristics believed to influence outcomes. The matching method used (for instance propensity score matching) can help to reduce selection bias (though may not eliminate it completely). Carrying out this matching process will require a large data set from which to select a control group, alongside specialist statistical skills in designing the study and analysing the results.

In either case, careful attention to the way sampling is carried out is essential, as is analysis of statistical power. For instance, a sample size that is too small will lead to an underpowered study, with a high probability of missing an effect that is real.

Sample designs vary from the simple to the complex (for instance those that are stratified and weighted). The design will depend on the type of information required, and the way the sample is selected. In simple terms, higher levels of precision and more complex designs require larger sample sizes.

A key consideration in addressing these issues is to think about how to deal with them during the initial design and planning stage. Final evaluation reports should also discuss how potential bias has been dealt with. Where quantitative methods are used it is important to explain how the research was carried out, and to be transparent about methods and the basis for estimates.

Sample sizes and sample selection methodology should always be included in evaluation reports, along with precision and confidence intervals for the estimates. In general it is important to be clear about the limitations of results presented, and to what extent they can be said to accurately reflect the characteristics of the wider population (i.e. to what extent are they externally valid).

Further reading:
PACT (2014), 'Field guide for evaluation: how to develop an effective terms of reference', Chapter 5: Sampling, http://betterevaluation.org/sites/default/files/Field%20Guide%20for%20Evaluation_Final.pdf

5.9 Bias in qualitative analysis

Qualitative approaches are also subject to various forms of bias, including:

- Self-importance bias, where respondents over-estimate the contribution of an intervention or programme in which they are involved. The possibility of this is likely to be accentuated where agencies are under pressure to justify their work.
- Allegiance biases, where attachment to a particular hypothesis leads other explanations to be discounted.
- Similar person bias, where researchers trust what they are told by people they see as similar, or those with high status.
- Courtesy bias, where respondents tend to focus on issues or explanations of events that they think will interest evaluators.

There are ways of reducing these biases. A first step is to acknowledge that bias is possible, and to think systematically through how this might be addressed in the planning stage of an evaluation. Without a systematic assessment it is quite easy to cherry-pick data which supports a preferred argument, or to record evidence with fits with pre-conceived ideas (i.e. "confirmation bias").

What does rigorous qualitative research consist of?

"Cherry-picked anecdotes to supposedly 'prove' a predetermined position come across as what they are: argumentative advocacy, not evidence. But the systematic, intentional and careful recording of purposefully sampled anecdotes (stories) can become evidence when rigorously captured and thoughtfully analysed." [1]

1 Patton, M.,"Qualitative research and evaluation methods" (2015). See http://betterevaluation.org/blog/anecdote_as_epithet

Preparing an evaluation framework which specifies in advance the nature and sources

of data, design of instruments such as surveys and the methods for analysing results will help with this, as described at the end of this chapter. This should help reduce the temptation to "go fishing," where evidence is gathered not to test a hypothesis, but to define and confirm a new hypothesis.

An important difference between statistical methods and the contribution analysis and process tracing methods described above, is that the latter use single observations in a context-sensitive manner to draw evaluative conclusions, rather than being based on the number of pieces of evidence collected.

As such, they may make use of purposive, rather than random sampling techniques. This would occur for instance, where initial interviews with market actors, which have taken part in an intervention, are used to identify other possible interviewees who have not.

Even so, it is still important to ensure that views from all key stakeholders are taken into consideration and represented in a sampling frame. Careful mapping of stakeholders at the planning stage of the research can help to ensure that interviews are carried out to explore different parts of the causal chain. It is also important that interviews are carefully planned, documented and recorded, and that results are triangulated from different sources.

Other considerations include being conscious and transparent about what the researchers themselves felt and expected during the research process, and to think what influence this might have had on their observations.

Further reading:
White, H., and Philips, P. (2012), 'Assessing attribution of cause and effect in small n evaluations: towards an integrated framework', https://beamexchange.org/resources/195/

5.10 Data collection tools

It is helpful to consider where evidence may come from, and then what techniques can be used to assess the issue of impacts and the contribution of an intervention. Many different sources of information (both quantitative and qualitative) may prove useful.

These might include primary sources such as large-scale surveys of market actors, beneficiaries and stakeholders, using either questionnaires or a programme of interviews.

A wide range of surveys is possible and provides the basis for both quantitative and qualitative techniques, so the detail of how the surveys are designed and implemented is important.

Smaller scale surveys of key informants, in other words, people with an in-depth knowledge of the market system, will almost certainly be required for an investigatory approach that aims to identify unexpected impacts. Key informants might include staff that work for implementing partners, market actors with an overview of the sector, industry representative organisations or business associations.

Secondary sources can also provide useful sources of information, and may include government statistics, press articles or reports from consultancies, market research organisations, NGOs or market regulatory bodies.

These are likely to be the starting point for an 'outwards-in' assessment of impact. This starts with observations of change in the sector, and then considers how individual interventions or the programme itself might have influenced these.

It is sensible therefore to review what information from secondary sources is available at the start of the evaluation, and also to be aware that some sources of information may become available as the study progresses.

Information provided by market actors is of key importance in evaluating interventions and understanding what is happening in the wider market system. The incentives of partners and other market actors to share information will be constrained however by commercial or privacy considerations.

It is important therefore, to be clear about what data will be required from partners when contracting with them at the start of the intervention. Some balance between the interests of different parties will be required, such that partners are content to provide information.

Practical issues may also create challenges, such as making sure that partner staff understand how to collect and store data in a systematic way, so it is important to make sure this is resolved at the start of the process.

Further reading:
The BEAM Exchange Monitoring Guide describes various tools that can be used to collect evidence, including questionnaires, different kinds of interviews, focus group discussions, case studies and diaries. https://www.beamexchange.org/guidance/monitoring-overview/module-9-designs-tools-and-techniques-monitoring/data-collection-methods/

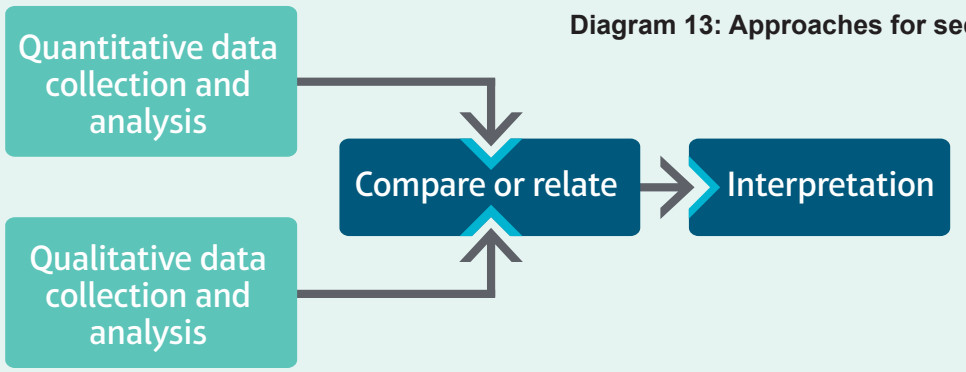
The same guidance reviews different tools that can be used to assess change in market systems, including social network analysis, Most Significant Change, and Sensemaker https://www.beamexchange.org/guidance/monitoring-overview/module-9-designs-tools-and-techniques-monitoring/techniques-understanding-market-systems-change/

Issues in obtaining information from partners, including commercial confidentiality considerations and market actors' incentives to share information are also discussed in more detail.https://www.beamexchange.org/guidance/monitoring-overview/8-collecting-information-market-actors/overview/

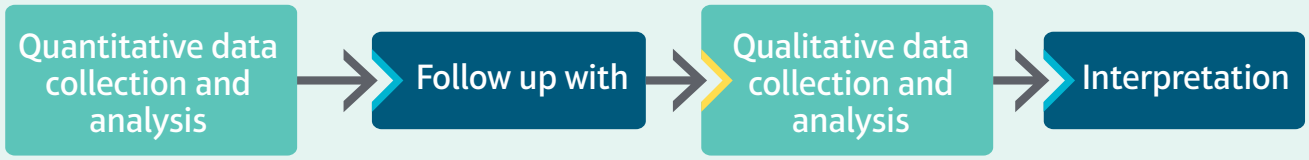
5.11 Sequencing and timing issues in data collection

In order to develop an implementation plan for the evaluation it will be useful to consider the sequence in which information is collected. In particular, data gathered using one method can shed light on how another method is put into practice. There are three broad options for sequencing, as follows:

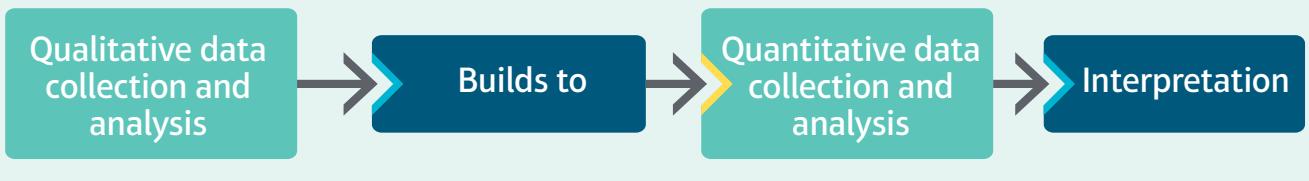
Diagram 13: Approaches for sequential data collection



An explanatory, sequential approach. Quantitative results highlights priority areas for qualitative research, and qualitative results are used to explain quantitative findings. This approach might be appropriate where a significant range of quantitative data is already available, but questions remain about how to interpret it.



An explanatory, sequential approach. Qualitative analysis can determine the design and areas of enquiry for quantitative research. This approach might be appropriate where it is expected that an approach that relies heavily on quantitative analysis (e.g. experimental or statistical) will be appropriate, but work needs to be undertaken first to understand how this approach can be implemented.



Source: BEAM Exchange

An exploratory, sequential approach: using ethnographic techniques to make sense of quantitative survey data in Ethiopia

Land Investment for Transformation (LIFT) is a market systems programme in Ethiopia that aims to increase incomes for smallholder farmers by promoting investment and productivity in agriculture. The evaluation of LIFT is using a qualitative ethnographic technique (Reality Check Analysis, or RCA) to conduct a scoping study, which will then be used to shape a quantitative baseline survey. The results of the ethnographic research will also be used to help interpret quantitative survey results. Using RCA, researchers will gather data through informal conversations to gain insight into issues including family decision making, social dynamics, power relations, and what land-holding means to people in the programme area

There is also a trade-off to be made on the timeframe for data collection, given the likelihood that systemic impacts will take time to emerge. With a short period between 'before' and 'after' surveys, impacts may not have had time to take hold and may therefore be hard to identify. At the same time, a longer time frame may also create difficulties, for instance by increasing 'interference' from other causal factors, or because attrition in the panel of respondents distorts the results.

Building the time to assess sustainability and systemic change into evaluation design

The Private Enterprise Programme Ethiopia is being implemented between 2012-2019, with a three stage evaluation (base-line, mid-term and final) being carried out during this period. An innovative element in the evaluation design is the addition of an additional survey that will be carried out in 2024, five years after the programme ends, in order to assess sustainability and the extent to which impacts are systemicrvey. The results of the ethnographic research will also be used to help interpret quantitative survey results. Using RCA, researchers will gather data through informal conversations to gain insight into issues including family decision making, social dynamics, power relations, and what land-holding means to people in the programme area

5.12 Summarising designs, methods and tools in an evaluation framework

Once all the issues discussed in this and the preceding chapter plan have been decided, it is helpful to summarise this in a matrix to provide an overall framework for the evaluation. The table should set out the key hypotheses from the theory of change, the questions to test these hypotheses, the design or combination of designs, methods, data collection tools and sources of information.

This evaluation matrix will provide a useful point of reference for managing the evaluation over time. It will also provide a check that the evaluation design is suitable for answering key questions about the programme, that there is sufficient triangulation between different data sources, and that data collection activities will gather the necessary information. Table 7 below provides an example of how the approach to testing one part of the FoodTrade East & Southern Africa programme theory of change.

Table 7: Evaluation matrix fo part of the FoodTrade East & Southern Africa programme

HYPOTHESIS FROM THE THEORY OF CHANGE	EVALUATION QUESTION	EVALUATION DESIGN	ANALYTICAL METHODS	DATA COLLECTION TOOLS AND SPOURCES
The FoodTrade programme has improved the functioning of national and regional staple food markets, by stimulating markets for inputs and services, including market information and food storage facilities	Has private sector investment and engagement in staple food value chains increased as a result of FoodTrade?	Case studies (of selected challenge fund grants and loans)	Contribution analysis	• Qualitative interviews with organisations receiving grants and loans • Secondary data on availability of food storage facilities • Programme monitoring data

6. OTHER CONSIDERATIONS

6.1 Overview of this chapter

This final chapter reviews some outstanding considerations for evaluations of market systems programmes. It discusses the need for evaluation and monitoring activities for such programmes to be integrated to a greater extent than has traditionally been the case, and reviews the implications for the role of evaluators. It also discusses how evaluation results can be reported and communicated in ways that increase the chance that they will be picked up and used by different audiences.

6.2 Linking evaluation with monitoring

Programme monitoring can be understood as work undertaken primarily by programme staff to track progress on an ongoing basis. As such, it has traditionally been distinguished from evaluation activities, which are undertaken on an occasional basis, usually by a third party.

However, market systems programmes need to be adaptive, and look beyond the scope of a traditional linear logical framework. One of the implications of this is that the conventional distinction between monitoring and evaluation may be less distinct than in other contexts.

Another is that the internal resources for monitoring and results measurement may be significant. It is likely, for instance, that internal monitoring will need to go look at the outcomes or impact that a programme is achieving. It will also require asking questions that are normally considered to be evaluative, such as how or why the result observed has been achieved.

This has implications for the role of an evaluator, and how it interacts with the programme team. The division of roles between the evaluator and programme depends on a number of factors, including the information demands and requirements of a range of programme stakeholders, and the capacity of a programme to undertake internal research and results measurement.

It is important that evaluation is used strategically and not as a substitute for effective monitoring, meaning that the evaluator should work with the programme (and funders) early in the evaluation process to clearly specify how they will interact.

This is partly due to the importance of avoiding duplication of data collection and analysis. It is also important because monitoring data is often a key source for evaluative research.

A useful tool for doing this is to develop a matrix that defines the division of responsibilities between these actors at key stages in the results measurement process, including:

- **Developing and reviewing the theory of change.** As well as being an important stage in formulating a programme strategy, developing a theory of change is a crucial starting point for a theory-based evaluation. The programme's theory of change must be 'evaluable' (see Chapter 4). Evaluation findings can be a valuable information source to inform periodic reviews of a programme's theory of change, which is good practice for adaptive programmes.

- **Intervention selection and review:** Evaluators can often provide useful insights on the relevance of the interventions selected by programmes. It is sometimes argued that commenting on the selection of interventions at an early stage of implementation runs the risk of compromising the evaluator's independence by influencing the programme strategy. However, the corollary is that an evaluator does not use their insights to influence programme decisions, which is probably less desirable.

- **Data collection:** In some cases, evaluators may rely entirely on programme monitoring and secondary data sources (sometimes quality assuring data collection processes). In other cases, they may collect additional primary data. There is no 'right answer' here: what is important is that roles are clarified, duplication in data collection is avoided, and that the implementer and the evaluator develop a shared understanding of data requirements, indicator definitions and so on.

- **Data analysis:** Both programme implementers and evaluators normally analyse data, so it is important to be clear on who is going to do what

Table 8: Possible division of responsibilities between the programme team and evaluator

	FORMULATION OF THEORY OF CHANGE AND RESUTS FRAMEWORK	INTERVENTION SELECTION AND REVIEW	DATA COLLECTION	DATA ANALYSIS	COMMUNICATE EVALUATION FINDINGS
ROLE OF PROGRAMME TEAM	Development (during inception phase) and periodic review	Develop intervention selection criteria and design interventions Develop intervention logic and monitoring framework Ongoing monitoring	Collection of monitoring data Verification of monitoring data from partners Collect market-level data from secondary sources Provide evaluator with relevant programme documentation and data	Monitoring and self-review Provide monitoring data to evaluator	Use evaluation findings to inform programme strategy Share evaluation findings with programme partners
ROLE OF EVALUATOR	Support development of log frame and theory of change and endorse from an evaluability perspective, report review	Review intervention selection criteria Case study evaluation of selected interventions Impact evaluation of selected interventions	Identify data requirements for evaluation Quality assure primary data collection design and implementation Undertake baseline and endline surveys for impact evaluation of selected interventions	Analyse impact data provided by programme Analyse baseline and midline survey data	Generate user-friendly summaries of evaluation findings

- **Communicating evaluation findings:** An evaluation is only of value in so far as it can be used. Evaluators should prepare a plan to support the use of the evaluation which takes account of its primary users, including funders, programmes, and other market actors.

An example of a completed matrix is provided above (table 8)

Further reading:
Kessler, A., and Tanburn, J. (2014), 'Why evaluations fail: the importance of good monitoring', http://www.enterprise-development.org/page/download?id=2484
The DCED also oversees an auditing system for the DCED Standard, which may help programmes to strengthen both their monitoring systems and evaluation practices.
For further information, see http://www.enterprise-development.org/page/audits

6.3 The role of the evaluator

To be effective, the evaluator needs to develop an in-depth knowledge of the programme interventions and context. Evaluators that are overly distanced from the programme often fail to gain sufficient understanding of complex and dynamic programmes and to adapt their evaluation approach accordingly.

For this reason, market systems evaluations increasingly involve a longitudinal collaboration between the evaluator and programme, which can help ensure a desirable combination of independence, relevance and utility for evaluations of market systems programmes.

The model involves an evaluator being in place from the start of a programme to identify evaluation questions, methods and data requirements early on. This enables the evaluator to work with the programme to ensure that

appropriate baselines are established, and that there is an appropriate division of responsibilities between monitoring and evaluation. Depending on the design, it can also help to provide 'real time' evaluation findings, which are particularly valuable for adaptive programming.

Evaluation: is too much independence a bad thing?

The review of completed evaluations of market systems programmes already cited[1] also showed that excessive distance between the evaluator and the programme implementer for the sake of 'independence,' negatively influenced the relevance of the evaluation findings and their subsequent usefulness.

Achieving independence should not be at the cost of ensuring a strong collaborative relationship between the evaluator and the programme. The argument for evaluation independence is often overstated for many reasons[2]:

"Institutional independence does not necessarily safeguard against biases toward positive evaluation. These can stem from poor quality evaluation designs, contract renewal bias, and 'friends' bias (an evaluator forms friendships with those responsible for the programme). These biases are best overcome by having a good quality assurance system that includes external peer reviewers.

And independence comes at a cost. If the evaluation team is outside of the agency it is evaluating it can lack access and influence. That is, it will not have the same understanding of how the agency works.

It will also not have complete access to internal project documentation for the purposes of the evaluation…And all agencies have a tendency to take reports from their own agency more seriously than they do those by outsiders, who can readily be written off as not really understanding the programme or even the institution."

1 Ruffer, T., and Wach, E. (2013) op. cit.

2 http://blogs.3ieimpact.org/is-independence-always-a-good-thing/

6.4 Reporting and communicating findings

Evaluation studies are typically presented in a formal final report, which will normally include the following sections:

- Executive summary.
- Background, including the context and baseline information, a description of the programme and interventions, and key evaluation questions.
- Description of the methodology, including an explanation of the design and methods chosen, and a discussion of potential weaknesses and limitations.
- Findings, drawn out with reference to the evaluation questions. For a market systems intervention it is important to cover not only outputs and outcomes from individual interventions, but also an explanation of the extent to which the programme has contributed to systemic change, and what unexpected impacts have emerged.
- Overall conclusions about the impact of the evaluation, learning points and recommendations.
- Annexes, providing supporting data including statistics and case studies.

A key aim of an evaluation is to see that the work makes a difference, and that findings are picked up and used to make positive changes. It is important to produce a final report that explains the evaluation process as a whole and provides the evidence for the findings. However, it is also common for evaluation information to languish in reports on donor websites, forgotten and inaccessible to many audiences who might value their contents.

The traditional output generated by evaluations is a long document that is housed on a website and emailed to contacts. However practical experience as well as more formal research raises questions about the usefulness of relying solely on this method of communication.

It therefore makes sense to think about the type of information that different audiences would value, and through which media it is most appropriate to communicate this to them.

The different users of the evaluation and the uses they will put its findings to should already have been identified during the preparatory work for the evaluation, as discussed in Chapter 4. If the evaluation process has taken a long time, it may be worth reviewing this list at the end of the process to identify if the list has changed in any way. It should then be possible to define the type of output or outputs for each group, along with methods for communicating these.

Different outputs should be prepared for different audiences. It is important that these outputs cross-reference each other. So, for example, a policymaker reading the brief should clearly be able to see how to find the longer report online.

Some examples of different outputs are provided in the table, along with some considerations what might be communicated, how this might be done and with what purpose.

Table 9: Outputs and purpose

WHAT	HOW AND WHY
Meetings to present results to stakeholder groups	Could be delivered in person or via a BEAM Grab the Mic webinar https://beamexchange.org/community/webinar/grab-mic/ Explains the findings and how they were derived Provides the opportunity to ask open questions to encourage people to discuss what they have heard
Press release	Succinct brief distributed to print, online, radio or other media which explains key evaluation findings in non-technical and easily understandable terms Evaluation may provide material that helps to diffuse a piloted innovation more widely within the market syste
Stakeholder group email	Could be sent before, during and after the process Helps maintain interest in the process and reminds people what is going on, provide the basis for them to discuss it and also integrate findings into their own decision-making cycles
Evaluation report	When the findings have been agreed, a full-length evaluation report should be published. An investment in hard copies may not be necessary
Pre-defined updates for Twitter, Facebook and other social media	A variety of Twitter-sized sample communications could be put together Enables stakeholders to share the findings and links to the report
Executive summary briefing	Findings from the evaluation report should be synthesised into an executive summary or short briefing of no more than two pages. This should be circulated in the same manner as the evaluation report An innovative approach would be to provide this also as a podcast, e.g. to allow it to be reviewed while travelling
Blog	Prepare a blog that explains the significance of the findings in an accessible manner Reach wider audiences By providing text that discusses the findings this may provide material about the evaluation report that is more easily referenced through internet searches
Infographic or data visualisation	The combination of text, visuals, and analysis can be compelling to those who are not moved by words alone. Many social media users like to share infographics which may provide the report with greater reach. An internet search for “free infographic maker” will identify various apps that can help with this
External events	Disseminate findings and lessons learned to other programmes through conferences, seminars etc. e.g. the BEAM conference or those of related networks including the Donor Committee for Enterprise Development (DCED), the SEEP network etc