Designing and evaluating support programmes in Lower and Middle Income countries

A handbook for Martin Bauer Group

WAGENINGEN
UNIVERSITY & RESEARCH

the nature network®
Martin Bauer Group

Koster, T., Bosselaar, J.M., Waarts, Y., 2020. *Designing and evaluating support programmes in Lower and Middle Income countries; A handbook for Martin Bauer Group.* Wageningen, Wageningen Economic Research, Report 2020-061. 38 pp.; 11 fig.; 2 tab.; 9 ref.

This document has been prepared for Martin Bauer Group and the authors received valuable contributions of Jan von Enden and Joscha Reichold. However, it reflects the views only of the authors, and Martin Bauer Group cannot be held responsible for any use which may be made of the information contained therein.

This document is available at https://doi.org/10.18174/527060 or at www.wur.eu/economic-research (under Wageningen Economic Research publications).

P.O. Box 29703, 2502 LS The Hague, The Netherlands, T +31 (0)70 335 83 30, E communications.ssg@wur.nl,
http://www.wur.eu/economic-research. Wageningen Economic Research is part of Wageningen University & Research.

Economic Research Report 2020-061 | Project code 2282500400

Photos: Shutterstock (front page) and Martin Bauer Group

Handbook by Wageningen University & Research

This handbook provides guidance to Martin Bauer Group to design, monitor and evaluate support programmes within their tea and botanicals supply chains.

The process consists of four phases as illustrated in the figure:

1. *Needs assessment.* In this phase the goal is to make an inventory of the most pressing needs and desires for improvement for the workers or community where the project will take place. Before this phase starts, a decision is already made on the location and value chain of the project in the Country Risk/Opportunity Assessment, which is beyond the scope of this hand book.
2. *Project design.* This process starts with identifying a project that contributes to the identified needs in phase 1. In this phase the activities and the desired impact are mapped and the planning is made.
3. *Impact monitoring.* During the implementation of the project, impact monitoring is used to track and understand the developments. If needed, the project can be adjusted based on the findings. Project implementation itself is outside the scope of this handbook.
4. *Impact evaluation.* After the project, or a certain period, the effects of the project are evaluated. The results can reflect both on the impact, but also how and why this impact is reached.

Per phase the following topics are discussed:

- The process of the phase
- The essentials for high quality project planning, impact monitoring and evaluation
- Basic tool descriptions that can be used in this phase

In the Appendix more detailed information is provided on specific elements of this process, such as Theory of Change (ToC) development, research methods, cooperating with local researchers, reporting on findings and quality assurance. It also includes a road map of how the different MBG documents and processes are linked to each other and some useful resources. Within each chapter you will find links to the relevant Appendix.

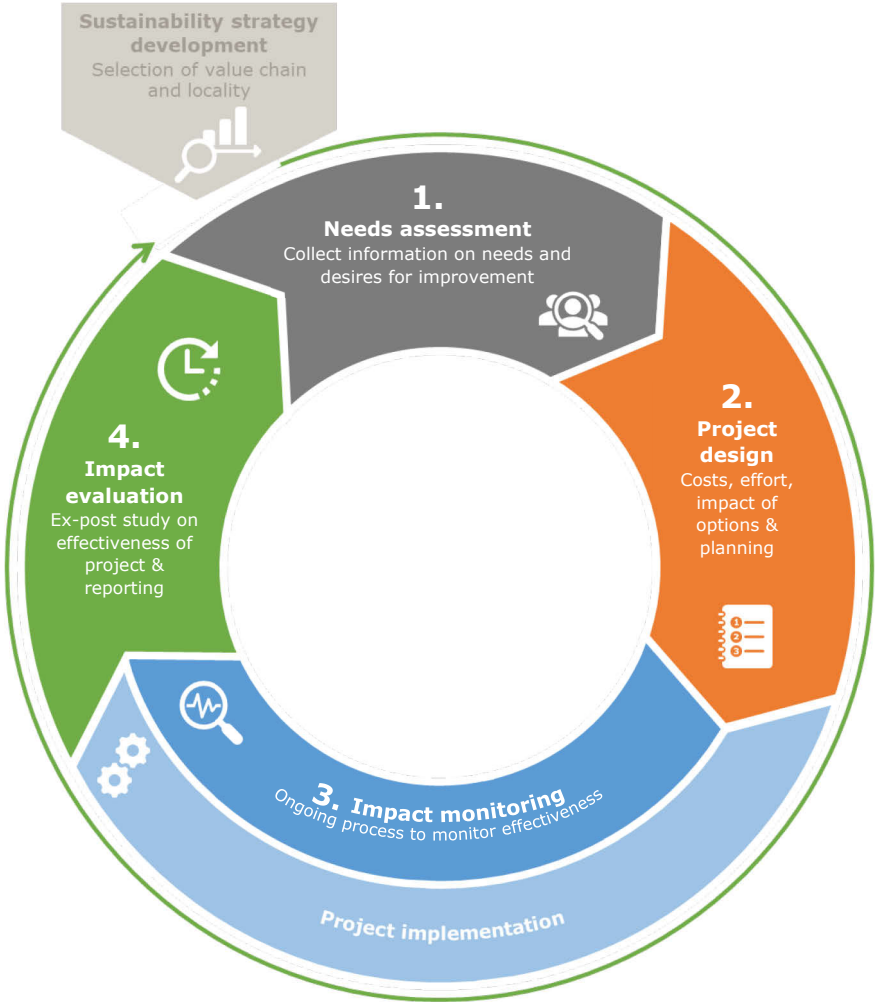


Figure 1 WUR's sustainability management cycle tailored to Martin Bauer Group's supply chain sustainability strategy. The figure is interactive, click to jump to the corresponding section.

Table of Contents

Introduction

Introduction

This document provides the building blocks for developing Martin Bauer Group's supply chain sustainability strategy based on solid field level data collection, and for monitoring and evaluating project and programme implementation.

It is a handbook which can be applied to the various types of projects Martin Bauer Group could implement, and evaluate. Martin Bauer Group intends to implement a variety of projects within their sustainability strategy to address sustainability challenges in various supply chains. This handbook, developed for and together with Martin Bauer Group, contains guidelines for how to best design sustainability interventions, and how best to monitor and evaluate the projects within the sustainability programme.

Supply chain stainability strategy development needs a proper fact-finding phase for the activities to be relevant, effective and efficient.

Such a proper fact-finding phase consists of: selecting the supply chains and locations where interventions are most likely required based on existing information (hotspots), connected with information on volumes sourced and the possibility for Martin Bauer Group to have an impact through its support. It also includes discussions with factories, communities, collectors and producers to obtain their perception on what challenges are important to be addressed and how.

Martin Bauer Group supported projects are based on the results of a needs-assessment and can therefore involve a wide range of topics.

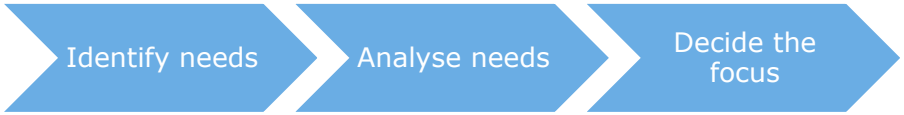
For now, Martin Bauer Group distinguishes two broad types of projects: giving training and providing investment. Each of these types of projects can involve all kinds of different topics. For example, the topic of a training can range from effective financial management to good agricultural practices or hygiene. We take into account the foreseen scale of the projects, which is now considered to be quite small. Therefore, it is unrealistic to focus on large and costly evaluation methods (e.g. large *n* quantitative assessments or even randomised control trials).

Proving contribution and attribution (accountability) is often very challenging, but this does not mean an evaluation cannot provide interesting findings for learning purposes.

Showing the contribution of a project enables to conclude that the project helped to cause the observed outcomes. Attribution entails proving the project caused the observed outcomes. The latter is often a challenge in environments where multiple developments and interventions take place at the same time. And it requires a lot of evidence, and therefore often a large sample size including a control group, which incurs high costs. We therefore advise Martin Bauer Group to focus the evaluation activities on assessing their contribution to observed changes. Besides, we advise Martin Bauer Group to select certain projects for thorough impact evaluation, for example because they want to scale up successful projects. Such an impact evaluation can then explore whether the project really was that successful and why. Another example could be to design a pilot project to test whether this indeed is the most effective type of intervention, before deciding what project to scale up, or to use pilot projects to compare the cost-effectiveness of different interventions. What is crucial here is to consider evidence-based pilot projects not just for accountability purposes, but as a useful tool to improve overall effectiveness.


1
Needs assessment

Needs assessment



The goal is to understand the needs and desires for improvement of the workers and/or the community selected in the Country Risk/Opportunity Assessment to design an effective, tailor-made project in line with Martin Bauer Group's sustainability strategy.

Let us start with explaining what is meant with a 'need' or 'desire for improvement'. It refers to the difference between the current situation and a desired situation. This sounds simple, but people are often inclined to confuse needs with solutions. For example, 'we need more staff' refers to a solution and not a need. The reason why the additional staff is wanted, for example 'to have more time to spend at home with my family', is the real need. The process starts with identifying the needs and prioritising them. Hence, what needs are most important to people, and should be addressed? The second step is to analyse the need. This means that you have to understand what the current situation is, why this need is so important and why the desired outcomes are not yet achieved. Based on this analysis a decision can be made on the need the project will address. The project design is further discussed in the next chapter.

Martin Bauer Group conducts needs assessment on two different levels; the Community Needs Assessment and the Supplier Sustainability Assessment, which mainly rely on qualitative research methods.

For a qualitative needs assessment it is crucial to really understand the needs of the target group. This means that questions should not be leading and that respondents should have the ability to express the reasons for the needs, why the desired situation is not yet achieved and for prioritising their needs. Qualitative research methods give the opportunity to ask follow-up questions and zoom in on the given responses. In focus group discussions (FGDs) respondents have the opportunity to discuss the answers in the group, which challenges them and shows possible conflicts of interest or synergies. A risk of FGDs is that not all respondents may contribute equally. Therefore, interviews can be considered for respondents that you feel could be especially dominant or passive in a certain group setting. It is good to have a clear and specific goal for the conversation. If the conversation is too open there is a risk of discussing many topics, but nothing in detail. So define focal discussion points. What the points of discussion should be can for example be inspired by earlier interactions with the target group or a small survey on various topics handed out to the respondents prior to the conversation. Be aware in this process that your own needs do not become leading. In this process, be aware to set aside your own needs (and solutions) to investigate the needs of the target population as objectively as possible. Then, you can assess the potential overlap between their needs and your priorities.

For more information on qualitative research methods, see the chapter on methods in the Appendix. Click here for the sampling guidelines.

It is important to talk to the right people.

Even though your project might address a very specific target group, that does not mean that everyone in this group has the same needs. Make sure that those different needs are identified and that you understand if and how a future project will affect different people in different ways. Furthermore, it is important to acknowledge the position of whoever executes the needs assessment and collects the data. In some situations, it is not advisable for someone representing Martin Bauer Group to do the data collection, and a more impartial person might be a better pick. Read more on inclusive research and facilitators in the chapter on quality assurance.

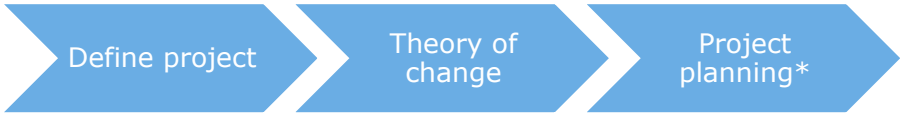
The World Bank gives a good overview of tools and techniques for data collection and decision making in the needs assessment process.

In Section 3 of A guide to Assessing Needs by the World Bank (2012) many tools and techniques are discussed which can be used for a needs assessment. Which tools are most suitable depends on the context, the goal of the exercise and the experience of the facilitator (among other factors).

2

Project design

Project design



Define a project that addresses the identified need, is likely to make significant impact and is executable.

Brainstorm on how a variety of strategies that addresses the identified need could look like. This includes activities, collaborations, incentives, investments and timing. Also think of strategies where you can strengthen existing programmes or finance activities by others. This step is best done in collaboration with relevant stakeholders including representatives of the target group.

Considering the list of possible strategies based on the previous analysis, you will need to make a choice. Consider the following selection criteria (from DAC-OECD, also further explained here):

- Relevance
  - What is your organisation particularly good at, compared to other development partners? How could you make a difference?
- Coherence
  - Possibility to cooperate with others to enlarge impact
  - What are other projects and organisations not yet addressing and could therefore make an impact?
  - How does the project fit in the overall portfolio of your organisation?
- Effectiveness
  - Have similar projects elsewhere proven to be effective?
- Efficiency
  - Cost-effectiveness of the project
    - What are the required investments (time and resources)?
    - Are the resources, time and staff available?
- Impact
  - Likelihood to make significant impact: Does the project address the root causes of the issue?
  - Different potential impact to different groups in the target community (inclusiveness)
- Sustainability
  - How likely is it that the impact made will continue to exist?

Draft a Theory of Change (ToC) to map how activities are expected to lead to different kinds of impact.

A ToC is a visual overview of how a project is perceived to lead to desired changes. For example, how do you expect a training on farming practices to lead eventually to a higher income of farmers? And what are the assumptions and risks underlying the change logic? A ToC helps to decide which activities should be included in the project. Plus, the ToC is a tool that helps monitor and evaluate the impact of the project. More information and an explanation on how to make a ToC can be found in the Appendix.

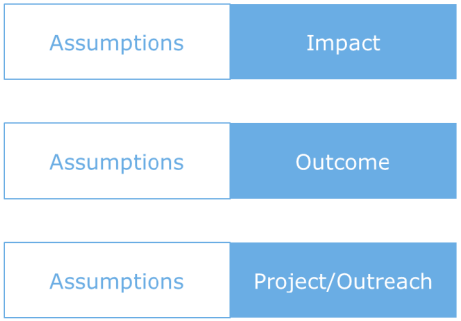


Figure 2 Simplified Theory of Change

**Project planning is outside of the scope of this document*

3&4

Impact monitoring and evaluation

Impact monitoring and evaluation

Impact monitoring and evaluation usually serve two goals: accountability and learning.

Questions of accountability focus on testing a Theory of Change: does the project work as it is expected to work? Through such assessments, the size of the impact and the cost-efficiency (value for money) of the project are established. Learning questions go beyond only testing the direct links in the ToC, and look into the mechanisms behind them. Learning questions include questions like: 'Which project variant works best?', 'Which type of people should we target?' or 'Should the current project be scaled up?'. Combining these two goals provide insights that are most adequate for planning, managing and implementing projects more effectively.

Impact of projects can only be measured by the question: 'What would have happened without the project?'; by looking at the counterfactual.

Imagine that a project has the goal to improve household income of the target group. A logical question to study the impact is to ask the question 'How did household income change over the course of the project period?' However, with this data alone, it is not possible to distinguish between the effects of the project versus effects of other events. Maybe there were market price fluctuations or climatic events which also affected the household incomes. Therefore, the central question in any impact study is 'What would have happened to the target population if the project would not have taken place?'. This alternative scenario – without the project – is called the counterfactual. The most common form is to ask the same questions to a comparison group. This is a group that is relatively similar to the target group, but was not part of the project. Other options are comparing the trends within the target group with general trends (e.g. income estimations by a governmental organisation) or to ask respondents to what extent they feel the project has contributed to the change. In every case, think of alternative explanations for the observed change and study to what extent each event could have caused what kind of impact.

The difference between impact monitoring and evaluation is mainly the timing.

With both impact monitoring and impact evaluation the goal is to measure the effectiveness of the project. In both processes similar research methods can be used. The difference is that monitoring takes place as a continuous process during the project and evaluation after the project or a certain moment in time.

Impact monitoring is a continuous process. This means that during the project it can be assessed to what extent a project is successful in generating change and why a project is or is not working. If needed, adjustments to the project can be made in a relatively early phase to optimise the impact. However, impact monitoring requires a continuous information flow and input of time and resources during the project. It is context dependent if and to what extent impact monitoring is considered useful and cost-effective. Furthermore, impact monitoring does not give insights into long-term impact, but only on immediate outcomes. If you continuously implement impact monitoring over a longer period, this could extend to intermediate or ultimate outcomes, but long-term impact often only happens later, after the project has finalised.

With impact evaluation only after a certain amount of time (mostly after a minimum of one year) or at the end of the project the effectiveness is measured. The impact may therefore be more clear as there has been more time for change. However, it gives less opportunity to adjust a project if immediate outcomes are not realised as foreseen. It is also possible to combine a form of impact monitoring with impact evaluation, as was done in WUR's PRIME project.

For a good impact study, it is key to always monitor project implementation.
Project implementation plans and reality often deviate. Plans get delayed, circumstances require last minute changes, etc. It is key to keep track of how the implementation of a project works out in reality. When a project does not reach the perceived impact, there is a need to know if this is because a failure of the project implementation or a failure of the change theory.

Any good research starts with a clear and fitting research question.
The research question is the basis for any decisions in the research design including indicators, research methods and tools, sampling and drawing conclusions. We therefore advise to spend sufficient time in defining the question. The chapter Defining the research question in the Appendix explains what this process entails.

As the project implementation may differ from the initial plan, the Theory of Change might need revision before doing the impact study.
With the impact study the change logic in the ToC is tested. So it is important to work with a ToC that accurately reflects the activities of the project. The process of making a ToC is described in the Appendix.

The Theory of Change leads to the research indicators.
In order to study the research question, indicators need to be chosen. Indicators are operational definitions to measure a concept. So, for example, if you want to know if the quality of a product has increased, an indicator refers to the ways you measure the concept of 'increased quality'. The project ToC gives the structure you use to choose the right indicators. How this exactly works is explained in the Appendix. For the projects of Martin Bauer Group, an indicator list is created as a guide in this process. However, as the projects of Martin Bauer Group are very diverse, the indicators that can be used are not limited to this list. How to use this indicator list is described in the Appendix.

The best methods and tools depend on the context and project goal.
There are many ways for doing impact studies. The decision on the best method should be heavily based on the research question, but also take into account other aspects of the project, such as the goal behind the question, the timeline, etc. For example, if the goal is to study the satisfaction of workers, there is a need for a different approach than when the price of product needs to be measured. We give an overview of different research methods and tools in the Appendix. This section explains which method is used for specific types of questions and indicators.

Sampling is another key component of a research design.
Sampling refers to the choice of which and how many people are chosen as respondents for the impact study. The choice of sampling depends (among other things) heavily on the research question. In the chapter on sampling the advantages and disadvantages of various sampling methods are discussed.

Well trained, high quality data collectors are the fundament of a good data collection.
Collecting data is a profession and it stands or falls with the quality of the data collectors. Data collectors are often local researchers, which need to be trained. In the Appendix we explain why there is a need for local researchers and how you can train them.

Smooth and useful data collection requires the right tools and timing.
In the Appendix there is an overview of considerations regarding data collection tools and timing.

The research design influences what conclusions can be drawn and reported.
In the Appendix we explain how to conclude on the contribution of Martin Bauer Group/the projects to outputs, outcomes and impact based on the methods and tools presented in this handbook.

Appendices

Appendix 1 Research methods and tools

Needs assessment studies and evaluation studies serve different purposes.

The research methods and tools described in this chapter can both be used for the needs assessment and for the impact studies. However, which tool or methods is suited for which case depends both on the goal of the study and the context. For the needs assessment studies the goal is to uncover the most pressing need or needs of the target group. Therefore the perception of the respondents is most relevant. These type of studies rely most on qualitative research methods, such as interviews and focus group discussions. This can however be combined with small surveys or other quantitative data collection, to get insight into the representativeness of those needs among a larger group. For evaluation studies, the research method and tools depend on the research question and the related indicators. In Table 1 we give an overview of the relation between the type of indicators and research method.

Table 1 Indicator types and research methods

Type of indicator	Method requirements	Type of data	Data collection methods
Observable	- Quantifiable/ observed data - Larger numbers of observations	Quantitative	Any
Self-reported	- Relatively objective questions to respondents - Larger number of respondents	Qualitative	Survey Focus Group discussion
Perception	- Questions on perceptions of respondents - Smaller number of respondents	Qualitative	Focus group discussion Interviews

Make sure the tools generate data/information as objectively as possible.

It is challenging to frame questions in a survey, interview or focus group discussion, in such a way that results in relatively unbiased and objective data. Hence, it is important to formulate all the questions beforehand and critically assess them. The most import criteria are: questions should not be leading and the order of the questions should not be leading.

Leading questions are those that already imply an answer. So for example, 'Do you think the labour standards are too low?' Instead, make this an open question 'What do you think of the labour standards?'. Or, in case of a survey, 'How satisfied are you with the labour standards?', giving the options from not satisfied to very satisfied.

Second, the order of the questions can be leading as well. This means that the answer to a previous question can influence an answer to the next. Therefore, if different questions address the same or similar topic, it is a rule of thumb to always start with the most general questions. So first 'What do you think of the overall labour standards?', and then the more specific questions 'What do you think of the holidays arrangements?' If you ask it the other way around, the answer to the general question is often influenced by the answers to the specific question and therefore less objective.

Make sure everybody understands the questions in the same way.

Questions to respondents often include concepts that can be interpreted in different ways. So, for example, what exactly is meant with 'labour standards' in the example above? Sometimes, then, the concepts of questions need to be explained first, so that everybody gives the answer to the same question. In explaining the concepts in the question, one should be just as careful not to 'push' the respondents in a certain direction as in phrasing the original question, for example while giving examples.

This is even more so when a study is undertaken in another country or another culture. Therefore it is crucial to work closely with your local researchers and facilitators and to practice with the question beforehand to be sure that they are understood the same by everyone. More on this can be read in the chapter on quality assurance in the appendix.

We advise to use quantitative data collection – where possible – to complement the qualitative methods.

The reason why quantitative data collection by itself is in most cases not sufficient, is because the data can only indicate certain correlations, but not causal effects. If for example school rates go up after a training on helping kids with school work, there might be a correlation. However, this data alone is not enough to draw any conclusions on causal relation. There might as well be other causes, such as school fares becoming cheaper. Conclusions on causality can only be drawn after a thorough comparative study (a relative expensive and time-consuming method) or with combining the data with qualitative methods which do discuss *why* things are the way they are.

Focus group discussions – qualitative method

A focus group discussion (FDG) is a qualitative research method in which a discussion or interview with several people is facilitated. This is a method that will be valid to many of the studies done by Martin Bauer Group, as the projects are often small and the perceptions of the target group are central.

FGDs are especially effective for collecting the perceptions of people and understanding the dynamics of causalities.

This is the case for most qualitative research methods. Where a survey that is repeated over time is able to collect hard data on for example productivity levels, in a FGD one discusses the perceptions on changes in productivity. So one can ask if people experienced a change and how they would describe this change (e.g. negative change, neutral, minor improvement, significant improvement). Be careful not to ask for exact numbers (e.g. growth by 30%), as people are often not aware of this. Aa big advantage of a FGD over a survey is that you can discuss with the group *why* they experienced a certain change, or what conditions are needed to increase a positive change. Hence, the respondents can bring anything to the discussion that they feel is important to the issue discussed.

With FGDs data on multiple individuals and group dynamics is collected.

A FGD is in general done with 4 to 10 respondents per time. The smaller the number of participants, the more in-depth the discussion can be and the more each individual respondent is able to express themselves. However, it is important that in the group a variety of voices is heard, which is more likely with a higher number of participants. Plus, the more people you invite for a FGD, the more data you collect per question or discussion point. We recommend not to invite more than 12 people for one FGD, as this will make your discussion too shallow and chaotic.

Good facilitation of FGDs is crucial.

There are two things that are really key when it comes to facilitation of FGDs. First, FGDs need to be well organised and the discussion needs to be well guided. That means that the questions and activities during a FGD need to be understandable and focused on a clear defined topic. The more participants, the more important it is that the FGD is facilitated in a clear way. Hence, a good preparation, practice and a good facilitator are key.

Second, a FGD is only valuable if people feel free to participate and answer questions honestly. Therefore it is crucial to create an open and safe space. This depends much on the connection the facilitator has with the participants. Hence again, a good facilitator is key. Plus, it means that the selection of participants should be done with consideration. Having certain stakeholders in the same FGD may prevent some from speaking freely. Also, respondents can answer on different levels, for in case of inviting an academic and a farmer for the same FGD. Hence, invite participants who are expected to understand each other and have an interesting and relatively open discussion.

Collecting the information in FGDs can be done in several ways.

To gather the information of a FGD, the sessions can be recorded (although this might hinder participants to speak freely) or notes can be made. We advise to have a note taker in the meeting, which is not the facilitator itself. Or to use posters made during the FGD in which the answers of the participants are well reflected.

Many different tools can be used within a FGD.

Here we give an overview of some of them. Pick those that you feel will contribute most to answering your question. If you design a FGD that lasts

longer than one hour, we recommend combining different tools as it will provide variety and increase the attention span of the participants. In any case, limit a FGD to a maximum of 2 hours.

- **General discussion.** The facilitator asks the group a question and lets them answer. Either one by one, or in a more organic discussion. Make sure, however, that you include everyone in the conversation.
- **Pick a side.** The facilitator poses a challenging statement. Each participant has to pick a side (agree, doubting, disagree). This can be done by raising cards reflecting the statement or by moving through the room (each position is one side of the room). After picking sides, some participants are asked to explain their position.
- **Voting.** Make a list that you want the participants to vote on (either with words or visuals). Participants can vote by raising their hands. Another possible way to vote is by giving each participant an amount of items (e.g. beans), that they can distribute over the different items on the list they can vote on. In this way, they can vote for multiple options and can express how important they find each item they voted for.
- **Role play.** This method is best used when you need information on (sensitive) social issues, such as harassment at work. Ask some participants to play a role and describe the situation. Involve the group by asking, how would this person react if a certain event takes place? Or, what will happen if a person undertakes a certain action?
- **Visualisation.** Ask the participants to bring something to the focus group discussion, such as their favourite food crop. This often makes it easier for people to talk about it. Drawings or picture cards can also help.

Interviews – qualitative method

Similar to the FGDs, interviews are a qualitative research method. Therefore, many of the pros and cons are the same. However, the difference is that with interviews mainly one or two respondents are questioned at the same time.

An interview has the advantage that the respondent has more and a safe space to give in-depth answers to the questions.

Hence, especially when a respondent has more in-depth knowledge which you need or the questions are more complex, an interview is preferred above FGDs.

The answers are only given to the interviewer, without anyone else present. This is especially relevant when the questions that are asked are relatively personal and/or make the respondent vulnerable. So for example for topics that involve discrimination or harassment, illegal practices or are closely linked to power relations.

The downside of interviews is that it is not as cost-effective as a FGD when it comes to numbers of observation.

As there is only one person interviewed, many interviews need to be conducted to make real contribution claims. Therefore, this method is most likely used for the evaluation of projects of Martin Bauer Group to complement data obtained by other methods, to gain more in-depth or confidential information.

There are different tools and forms of interviews that can be used.

We discuss here some of the most common types, which we feel are useful for evaluating projects of Martin Bauer Group.

- **Semi-structured interviews.** This is the most common format for interviews. This means that the interview has an interview guide, consisting of mainly open questions. However, based on the answers given during the interview, the interviewer is able to vary the sequence of questions and to ask follow-up questions.
- **Unstructured interviews.** These interviews are often built around one question. For example, 'What do you think about the project?' This way of interviewing is meant to reveal what is most important and relevant for the respondent, without imposing one's own vision of what should be discussed. The risk - and benefit - is that the interview may take unexpected turns.
- **Conceptual drawing.** When complex issues are discussed like causal mechanisms or networks, making a conceptual drawing is a powerful tool. This can either be done by the interviewer based on an answer and reviewed by the respondent or by the respondent itself.
- **Transect walk.** Walk together with respondents through the area and discuss what you see. This tool helps discussing topics like living conditions and resource access. Besides talking about what is seen, causal relations can be discussed. This can also be done in groups.

Quantitative data collection

Quantitative data collection usually involves data collection on a larger scale. Survey data is the most common format, for example by asking all participants of a training to fill in an evaluation form on their appreciation of the training,

what they learned and if they will use their knowledge in the future. Of course, many forms of quantitative data collection exist.

There are three ways to sequence qualitative with quantitative data collection.

The first way is to collect the qualitative and quantitative data at the same time (Figure 3). So, for example collect data on farm production levels from buyers and discuss the changes in production levels in a FGD with farmers. The outcomes of both methods combined can either strengthen each other if they show the same outcomes, or challenge each other when they differ. In case outcomes differ, an additional study can be done to understand how and why the outcomes are not consistent.

The second way is to start with quantitative data collection and then use FGDs or interviews to interpret the findings (Figure 4). For example, when data has shown that the production levels did increase over time, the FGDs and interviews can be used for interpretation. So how did it change, why did it change, what were the effects of the change (for example on working hours and overall income) and did people perceive the change as positive or negative?

The third way starts with qualitative data collection, followed by quantitative research to show if findings are representative and/or transferable (Figure 5). For example, in a FGD many farmers indicated that they experienced a growth in their production due to a project. However, this was only a group of 10 farmers in one village. So to see if this change is more widespread, quantitative data can be collected for more farmers or farmers in other places (or times).

It is key to look for cost-effective quantitative data sources

As said, surveys are often not feasible for Martin Bauer Group projects. However, some quantitative data is easily collected, or already existent. Here we give an overview or sources you can think of. Of course, whether data is useful, available or collectable depends on the local context. Also note that this list is by no means exhaustive.

- Training participation
- Wage data (supplier/Mabagrown)
- Test scores
- Environmental data (Mabagrown)
- Production data (supplier)
- Etc.

For all research methods and tools, ensure that you adhere to regulations on personal data collection as well as research ethics.

More and more countries have regulations in place related to personal data collection ('Personal data are any information which are related to an identified or identifiable natural person'). Increasingly, one has to obtain a licence to collect such personal data from a local authority. Other important elements to include in data collection efforts are to ask people about their consent to participate and ensure the reporting is done in an anonymous way such that information from a report cannot be traced to the person giving the information.

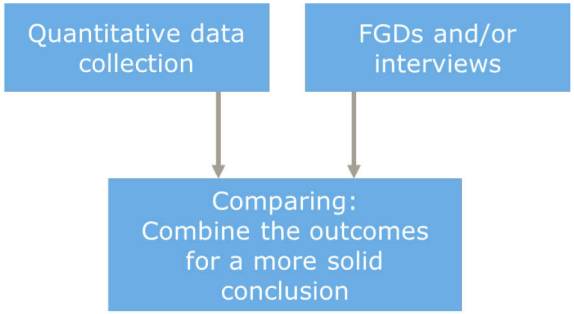


Figure 3 Comparing results

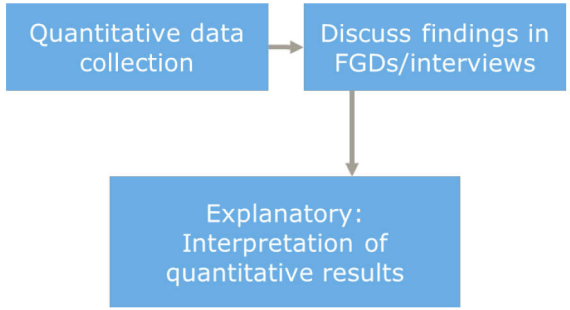


Figure 4 Interpreting correlations

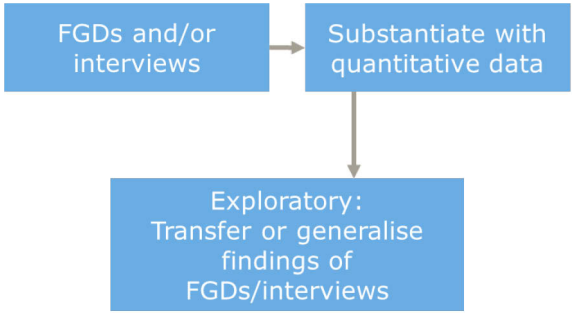


Figure 5 Transfer or generalise findings

Appendix 2 Sampling

There are different possible ways to draw a sample from a population, each with its specific advantages and disadvantages.

Sampling refers to the segment of the population that is selected for the research. To decide on the sampling method, the most important question is always: what are the research or evaluation questions? Also, different target population sizes require different sampling strategies. Target populations do not only consist of the direct project participants, but can include their communities, if for example spillover effects are expected (and you want to learn about those too). These sampling techniques theoretically hold for both qualitative as well as quantitative data collection, although the size of qualitative and quantitative samples will differ. The sampling strategy and sample size always needs to be tailored to the programme at hand, which makes it impossible to give absolute guidelines in terms of sample size. The sample size depend on the size of the total population, the indicator, expected effect size, desired strength of conclusions, etc. We distinguish three main types of sampling strategies: convenience sampling, random sampling and purposive sampling.

Convenience sampling:

This strategy relies on whoever is readily available, or most keen to talk. As the name implies, this is very convenient, but not always the most reliable method. There is often a reason why those people are available or so willing to share their opinion, which then often means they are not representative of the target population. However, these people can still be great sources of information if verified with others.

Random sampling:

If the population is large enough, you can be confident that a random sample of sufficient size will result in a representative sample. There are multiple options for doing this, such as simple random sampling (sampling randomly from a list) or random walk (walking along a predetermined route, second house on the left, third house on the right, etc.). Another way to sample randomly is to use stratified random sampling. In this sampling strategy, you divide the population in different groups, or strata, and sample randomly from each of those. If different results can be expected for different groups and you want to make sure they are all represented in the sample, this method is very useful. For small populations, the problem is often that simple random sampling results in overrepresentation of certain groups. Sampling within the groups of interest avoids this problem. These strata can for example be based on gender, age, farm size, or any other characteristics that is relevant to the project at hand. Another important group that can make up a strata is non-supported respondents to compare the results with.

Purposive sampling:

In purposive (or purposeful) sampling, there are many different options. For example, one can look for the most critical voice, or for special cases (e.g. early adopters, largest successes or failures). Another effective option is to look for the source who can give most additional information. In that case, you are looking for a new perspective from every additional source. For example, if everyone you have interviewed so far was part of the MB supply chain, but you also expect wider community effects, the next interviewee could be an informed outsider (who is not working in the supply chain), e.g. a school principal or village leader.

Collecting data from non-supported people/villages is not always necessary, but can be very insightful.

As described in the chapter on Impact monitoring and evaluation, building a counterfactual is central to impact studies. A common method for this is using a comparison group. Such a comparison group can take many different forms. In general, the sampling strategy guidelines as described above would also hold for selecting a comparison group. However, it is not always necessary to choose the same sampling strategy for both the target group and the comparison group. For example, if quantitative data is being collected from the target group based on random sampling, you might very well opt for holding a

focus group discussion with some purposively sampled non-supported individuals. You can use such information for 'counterfactual thinking'.

The selection of a non-supported group is based on comparability.

The two main criteria for selecting a comparison group are: comparability with the target group and absence of spillover effects from the project.
Comparability with target group means that the target and comparison groups should be as comparable as possible. And in case there are any remaining known differences, these should be made explicit. For example, if the community that receives beehives in the beehive project is very far from the main road, and you select a comparison community very close to the main town, these communities have different access to markets or public facilities to start with, and comparison on indicators such as income do not make sense.
Some examples of characteristics to take into account are:

- Location and access to infrastructure
- Socio-economic characteristics
- Income sources (and crops grown)
- Age and education levels
- Size of community
- Other projects/programmes active in the area.

Absence of spillover effects from the project means that the comparison group should not have been influenced by the project. If spillover effects (intended or unintended) can be expected at all, it is not suitable as a comparison group, as disentangling the direct and spillover effects will be impossible.

Incentives may be needed to motivate people to be part of needs assessment and evaluation activities.

It is important to consider why people who are part of the project, as well as people who are not, would participate in the research. For project participants it may be logical to take part in the study, but this does not necessarily hold for the comparison group as they have not received any support. It can thus sometimes be challenging to have people take part in research. Therefore, it is important to compensate them for their time. Examples are: providing meals during the interview/discussions, give a remuneration for transport costs, top up for mobile phones, soap, tools to be used on the farm, useful information on farm management, etc.

Appendix 3 Drafting a project Theory of Change

A Theory of Change (ToC) is a visual overview of how a project is perceived to lead to desired changes.

For example, how does a training on farming practices lead to a higher income of farmers? A ToC is implicitly already in the minds of those who draft a project. However, an explicit ToC and change logic in a visual overview helps by defining indicators to evaluate the impact of the programme. Additionally, it shows the prerequisites and risks for the project to succeed. Here we describe in four steps how to draft a ToC for a project. In Figure 6 an empty, simplified version of a ToC is illustrated. Figure 11 shows an example ToC for a project with a training on agricultural practices and financial support for buying inputs.

Step 1: Define the end goal of the project (impact)

Each project has an ultimate impact goal. In our example in Figure 11, the training and financial support are aiming at increasing financial household security. The Martin Bauer Group ToC (Figure 7-10) includes the different impact goals. In case a project is meant to contribute to more than one impact goal, they can be listed in separate boxes besides each other. When there are multiple impact goals, there will also be different pathways within the ToC to describe the paths to those goals.

Step 2: Describe the project activities (project / outreach)

For a new project, describe the activities to be undertaken to address the needs of the target group. When evaluating a programme, the activities of the programme are already known. Hence, these can be filled in at the bottom of the ToC. A ToC is not set in stone, and may change throughout the project implementation.

In case a project exists of multiple activities, they can be listed in separate boxes next to each other, such as in Figure 11. This example describes the project in quite general terms, but the more specific the ToC, the easier it is to define outcome and impact indicators later in the process.

Step 3: Define the causal links (output)

In this step you describe how you expect the project to contribute to the impact goal in the end. So in our example, how do training and financial support lead to an increase of financial household security. The example shows big steps for a quick demonstration. In reality, there can be many of these steps (or outcomes, as they are called). It is important to include the steps that are prerequisites for the project to reach the goal. For effective training for farmers this may include educating trainers or organising events to attract attention of farmers.

Step 4: List the assumptions

The last step is crucial and often overlooked. In this step you list the underlying assumptions per step of the ToC. These assumptions are often implicit and require a detailed analysis of the project. For example, when a project consists of a training for farmers, it is not a given that a training leads to better farming practices. It is assumed that a) farmers come to the training; b) they understand what is taught; c) they learn something new from the training; d) it is possible for them to put the learned knowledge into practice; and e) they will use the new knowledge. Listing all the underlying assumptions gives insight into what to measure when measuring the impact of a project.

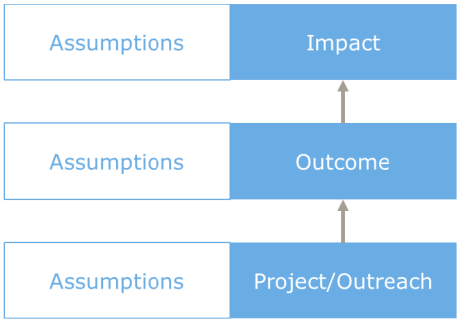


Figure 6 Simplified Theory of Change

It is important to also take into account possible external influences and unintended effects while developing a Theory of Change.

The events specified in the change logic do not occur in a controlled environment. There are other actors operating in the same area, there is a government, etc. These aspects should be addressed in the evaluation to be able to explain why an effect did or did not take place. The same holds for any unintended effect that happened. These external influences or unintended effects are very valuable learning points for future projects.

Martin Bauer Group supply chain ToC is a good starting point for drafting project ToCs.

To start this process, first assess which pathways in the Martin Bauer Group ToC best fit the programme. This pathway is the blueprint for the project ToC. From there you can add the project activities, additional and tailor made causal links and assumptions. The current version of the supply chain ToC and the corresponding narrative can be found here. The latest version of the ToC at the time of the handbook is shown below, and in three parts (for better readability) on the next pages.

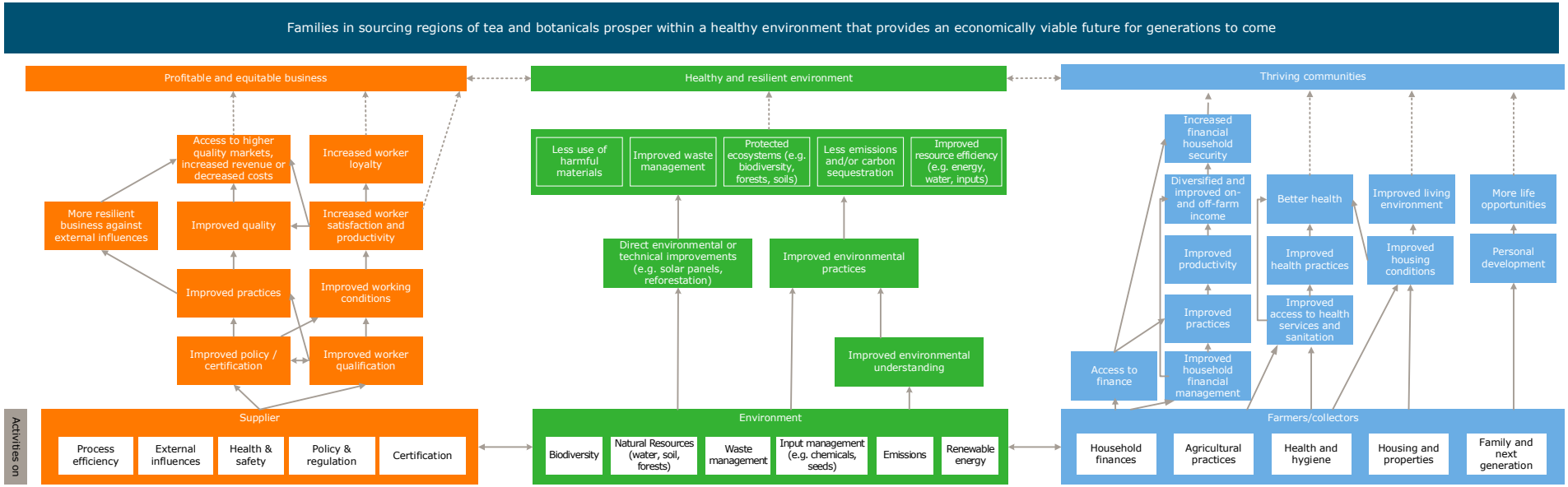


Figure 7 Martin Bauer Group supply chain ToC: Overview

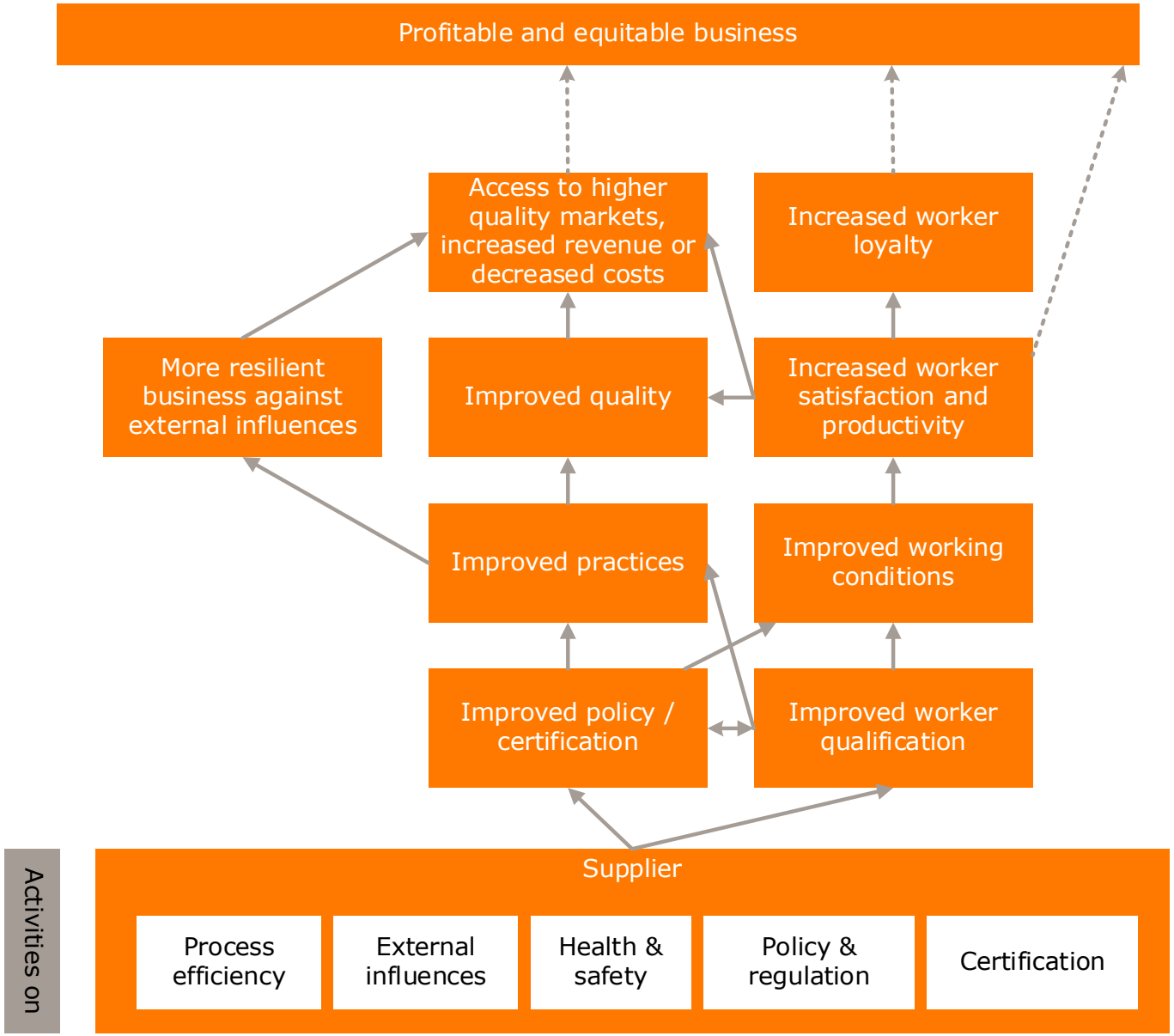


Figure 8 Martin Bauer Group supply chain ToC: Profitable and equitable business

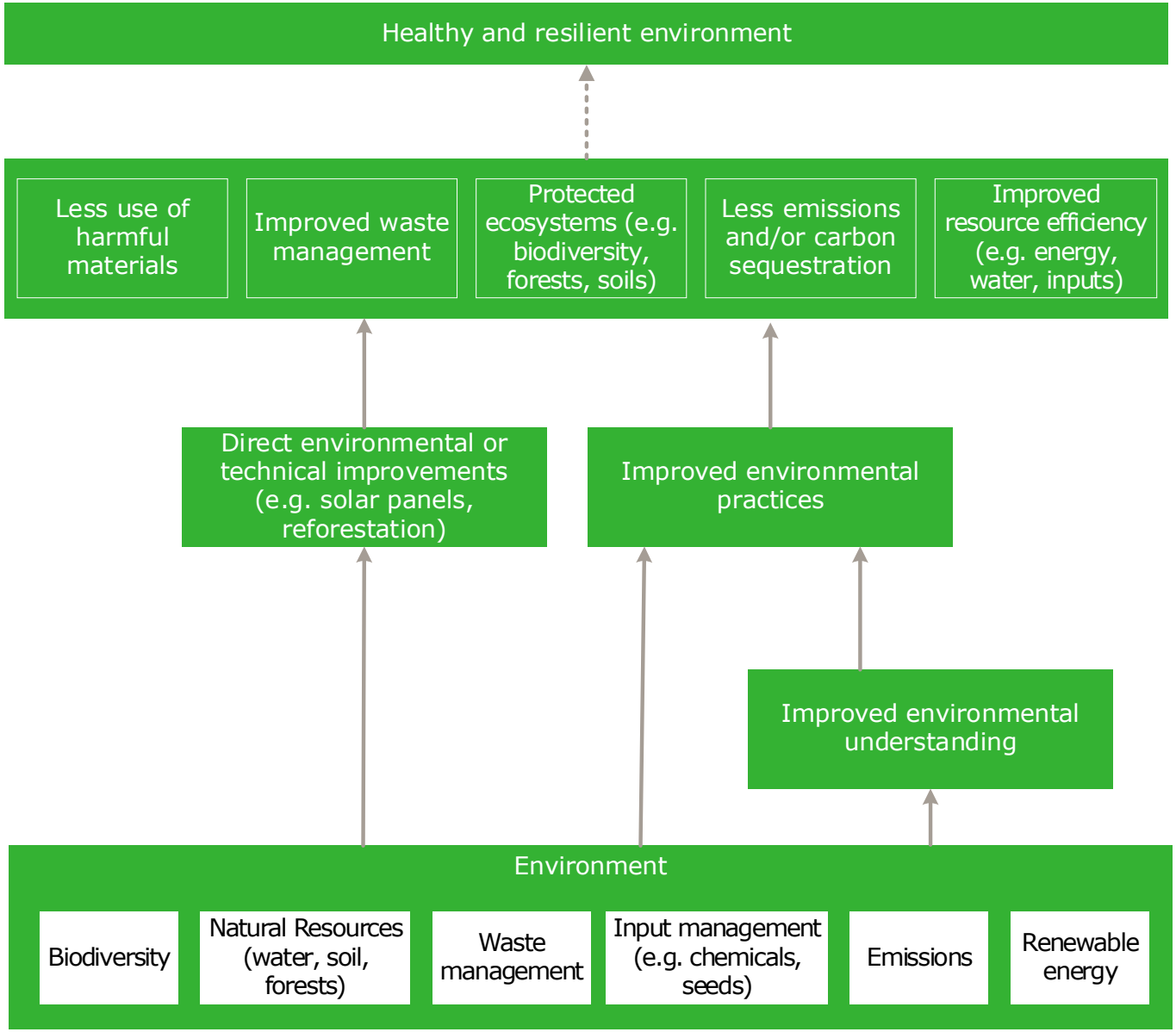


Figure 9 Martin Bauer Group supply chain ToC: Healthy and resilient environment

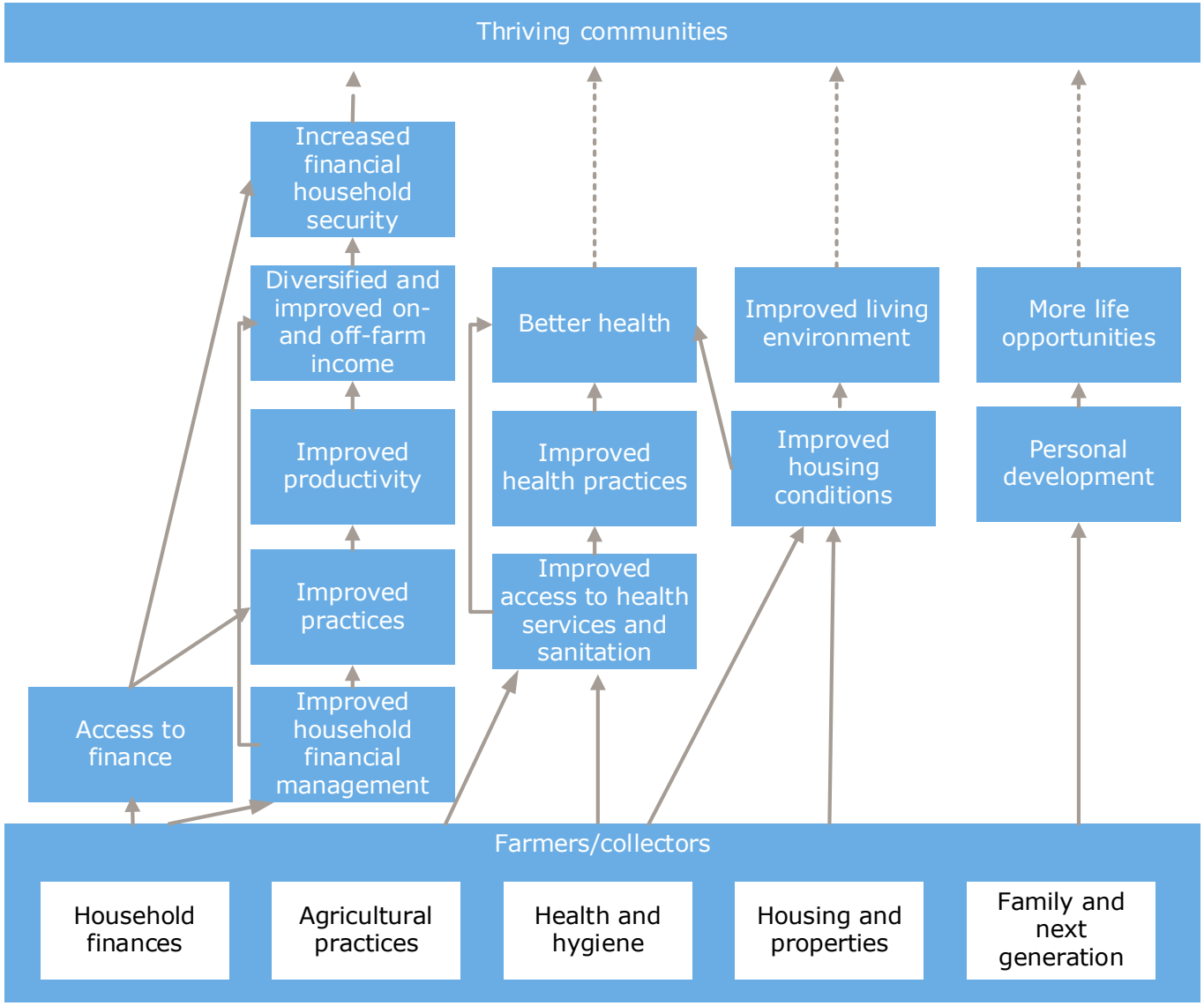


Figure 10 Martin Bauer Group supply chain ToC: Thriving communities

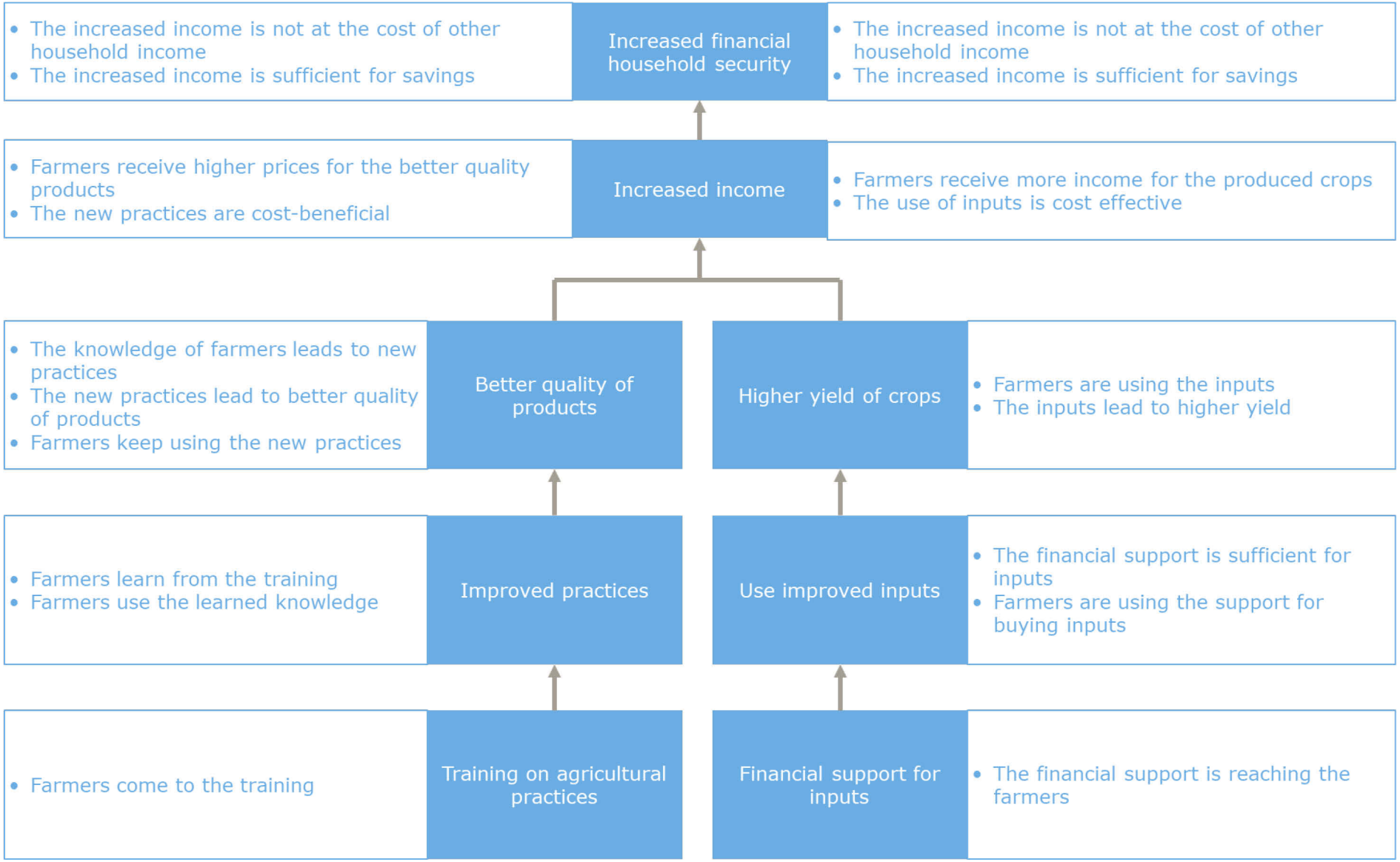


Figure 11 Example of a project Theory of Change

Appendix 4 Defining the evaluation question

A good and clear research question is key and guides the entire research.
The research question is the basis for all decisions in the research design. We therefore advise to spend sufficient time on defining the question, and to follow the four steps described below:

Step 1: Define the focus of the study.
What do you want to learn about the project? For example, there is a project on building beehives with the ultimate goal of increasing household income. In this case, the subject of the research question could be household income. To choose the subject of the study, Martin Bauer Group's ToC is a good reference point. The subject of the study and the related change pathway within the ToC are central in the choice of indicators for the study (see chapter Choosing the right indicators).

Step 2: Define the goal of the study.
A study on a project can be done for a variety of reasons. The choice for the goal of the study has implications for the methods and tools, sampling and what conclusions can and cannot be drawn. Here we list the most relevant goals of evaluation studies (based on OECD):

- **Relevance.** This questions the relevance of the project regarding the local requirements and priorities. 'Is increasing income by building beehives addressing the local needs?' This goal relates to the needs assessments done by Martin Bauer Group.
- **Effectiveness.** Are the objectives of the development intervention being achieved? 'To what extent are the beehives built and are they generating income for the households?'
- **Impact.** This goal relates to effectiveness, but goes a level higher in terms of impact. In the case of the beehives: 'Does the beehive project contribute to higher household income?' Hence, this question also takes into account that income generation by the beehives may have negative implications on other forms of income. Unintended effects are also included in such assessments.
- **Efficiency.** This refers to a cost-benefit analysis. 'How do the investments in the beehive project relate to the effects/impact of the programme?'
- **Sustainability.** The last goal looks at the expected duration of the positive effect or impact. An outcome is only sustainable if the effects or impact sustain for a relatively long period. 'How long do we expect the beehives to contribute to higher household incomes?'

Step 3: Define the scope of the study.
In general there are two different types of scope to define for a study: geographical scope and target group. Geographical scope simply refers to the area you want to study. The target group is often context specific, and could involve criteria such as: occupation, gender, age, social or cultural group, religion, socio-economic position, etc. Target groups can refer to individuals, but also to larger entities such as households.

Step 4: Define the research question: 'subject + goal, for scope'.
In case the research question is not yet clear, we advise using one of the following tools to help in this process.
Rich picture. To identify knowledge gaps, visualise the related actors, factors, and relationships affecting an issue.
Problem definition. Clarify which problem you are working on by asking 5 questions:

1. What is the key issue you are trying to address and why is it important?
2. Who is it a problem for?
3. What social and cultural factors shape this problem?
4. What evidence do you have that this is worth the investment?
5. Can you think of this problem in a different way? Can you reframe it?

This will help a group of people to focus ideas in the same direction.

Appendix 5 Building counterfactual evidence

Including a comparison group is often the strongest, but also a very costly and sometimes even unfeasible way to build a counterfactual.

One very common way of answering the question of 'what would have happened without the project' is to compare the results of the supported people to results of non-supported people. When the comparison group is chosen rightly (more information on the criteria for this in the chapter on sampling), this is very effective. However, sometimes there simply is no suitable comparison group, or it is not feasible to include them because of time or budget constraints. In those cases, there are multiple other ways to build a credible counterfactual.

Comparing people who have just joined the project with people who have joined the project earlier on.

If a project has been rolled out in phases, or people were able to opt in at different moments, this is a very effective way of constructing a counterfactual. Many effects can only be expected after a certain period has passed, so information from the people who recently joined can serve as some kind of baseline. However, it is important to carefully think about and/or measure potential (unobservable) differences between those groups. The early adopters were maybe more entrepreneurial or risk-loving, whereas the late adopters first wanted to see results for their neighbours before they chose to join. Or maybe the project deliberately targeted the poorest people first and is now letting in others as well.

Asking recall/change questions.

This is especially useful when the project has not yet been going on for a long time. In that case, people will most likely remember their situation before the project quite vividly. Questions can include something like 'What was your situation on topic x like two years ago?', 'Did you experience any changes in topic x in the past two years?' (and why/how) or the more general 'What were the main changes you experienced in the past two years?'.

Asking about factors contributing to a change as well as discussing what would have happened without the project.

This can be used in combination with asking recall or change questions. If respondents answer that they have experienced changes, you can ask what factors contributed to this change and why this was the case. Something to note in this case is that you should avoid asking about the contribution of the project directly, as this is a 'leading question'. The respondents might answer such 'leading questions' more positively than they actually feel, because they want to please Martin Bauer Group (especially if they are working in the supply chain). A more open version is to ask questions about change and what contributed to this change. Then, based on this information, one can conclude whether a project or initiative contributed to that change. An important addition is to ask people as well about 'what would have happened if the project would not have taken place' to assess additionality.

Using existing data to put your own information into perspective.

It is not always necessary to collect additional data to build a counterfactual. Sometimes, existing or even public data can do the job. A counterfactual constructed on existing data is strongest if the data is as locally specific as possible. This is also where Mabagrown monitoring data may come in very useful, because using existing information at country level is for example often useful to get a first idea, but to really be able to say something about the counterfactual situation of the target group, something more local and specific is needed. It will rarely be possible to find public data from the exact same region/context as where the project is taking place. Therefore, the information stemming from this method of constructing a counterfactual should be analysed taking into account the possibility that the locality in which the project is implemented differs from the context from the counterfactual data. Some potentially useful data sources:

- World Bank development indicators (global)
- FAOSTAT (global)
- Census or household survey data from national statistical agencies (country-specific)
- There are various sources on wages and incomes. Specifically regarding minimum wage, living wage and living income benchmarks. Some examples are: Global living wage coalition, The living income CoP and Align.

Appendix 6 Choosing the right indicators

Carefully chosen indicators operationalize studying the research question.

Indicators are operational definitions to measure a concept. For example, if the questions is whether the quality of a product has increased, an indicator refers to the ways one can measure the concept of 'increased quality'. One concept often could be measured in various ways. For product quality one can think of:

- Shelf life of a product
- Quality of nutrients within a product
- Market prices of different quality grades
- Farmer perceptions of product quality

Hence, it depends per project and product what type of product quality is meant, and how this can be measured.

WUR created a general indicator list that can be used to select indicators.

Wageningen University and Research developed a general indicator list for Martin Bauer Group, based on their general ToC. This list can be used as a guide to select the indicators.

The indicator list consists of four sheets: one general sheet and one per ultimate development theme of Martin Bauer Group (profitable and equitable business, healthy and resilient environment and thriving communities). Hence, depending on the impact goal of the project, each impact study will include indicators from the general sheet and one of the thematic sheets.

Step 1: Select the goal of the project.

Each project has a certain ultimate impact goal, which is related to the overall ToC of Martin Bauer Group. An overview of these goals is portrayed in Table 2. In the indicator list you select the theme and filter on the goal of the project. This provides a relatively short list of possible indicators that can be used for measuring the impact of a project.

Step 2: Use the project ToC to refine the indicators.

The goal 'Better health' for example includes the indicator improved health practices, with the sub-indicator knowledge on health practices.

Table 2 Overview of Martin Bauer Group themes and goals

Theme	Goal
Business	- Increasing worker loyalty
	- Access to higher quality markets, increased revenue or decreased costs
Environment	- Protected ecosystems
	- Healthy and resilient environment
Thriving communities	- Better health
	- Improved living environment
	- Increased household security
	- More life opportunities

When, for example, evaluating a training that addresses a certain health practice, e.g. sanitation, you need to refine the indicator in a way that is specific for the project. This means that the sub-indicator becomes knowledge on health practices related to sanitation. The more specific the project, the more specific the indicators.

The project ToC gives structure to choosing the right indicators. For every step and assumption in the ToC, you need indicators for the steps and the assumptions. For all these points you have to ask yourself: 'How can we know whether...'

- How can we know whether the knowledge of farmers leads to new practices?
- How can we know whether the new practices lead to better quality of products?
- How can we know whether farmers keep using the new practices?
- How can we know whether the quality of the products increased?

Step 3: Make sure nothing is missing.

As the support programmes of the Martin Bauer Group are very diverse and the ToC is under constant revision, it might be that the indicator list is missing out on some relevant indicator for your project. Hence, have a close look at all the key steps and assumptions in the project ToC and make sure that there is at least one indicator for each.

There are three types of data that can be used to measure an indicator: observable, self-reported and perception.

Not every indicator can be measured in the same way. Therefore we make a distinction between three types of measurement: using observable, self-reported and perception data. In the indicator list, the possible types of measurement are given for each of the sub-indicators. Additionally, there is a short list of possible questions that can be answered with this type of measurement.

Observable data refers to data which you can count or observe directly. For product quality, this applies to shelf life for example. The days before the product gets rotten can be counted. Other forms of observable indicators are for example test scores, observed data on farm management practices, yield and price information. Observable data is often trustworthy (when measured correctly), but in many cases the data collection is not very cost-effective. This is because there is a need for many data points or many site visits. Hence, these indicators are especially useful if they are already available somewhere (for example price data) or can be collected easily (for example test scores after a training).

With **self-reported data** you ask respondents to report on their observations. Hence, you try to remove as much bias and perception of the respondent as possible. In our example, you can ask a selection of people how long the shelf life of a product is. If you ask this question at different times, you can measure the change in a rather exact way. However, in a less exact way than by doing observations yourself, as the information is based on the experience of others.

The last way of measuring indicators is by looking at **perceptions**, which is also the most subjective method. Respondents are asked directly how they have experienced change or what they think of a certain issue. For example: 'Do you feel the shelf life has changed over the past five years?' This question will never lead to an exact answer, as nobody is able to tell with how many days the shelf life has increased. They can however indicate whether there was a positive change and whether this was, in their opinion, minor or significant. Perceptions are often used when there is only one moment of measurement (so you ask for the perception of change over a past period) or when measuring concepts that are not observable, such as worker satisfaction.

The indicators and measurement type define the research method

In the chapter on research methods and tools in the appendix, we describe different qualitative and quantitative research methods. In Table 1 of that chapter there is an overview of how the types of indicator measurements relate to the methods.

The measurement types influence the reporting.

Conclusions can be supported in different ways when using observable, self-reported and perception data to measure indicators. In the chapter on reporting and concluding in the Appendix we explain these differences.

Appendix 7 Quality assurance

Involve relevant stakeholders throughout the design and evaluation process.

It is important for the quality of the information that relevant stakeholders are involved in all steps of the processes: this to ensure that the right questions are asked (different people have different stakes, roles, experience and focus), that the right people are spoken with and that the right people have a change to validate the findings of the data analysis to strengthen its conclusions.

Data collectors: why, when and who?

We advise to almost always use local data collectors for gathering information. Besides that collecting your own data is often not feasible, you are not a neutral data collector, which increases the chance of biased data.

Using local data collectors will benefit the quality of the information collected in nearly all cases.

However, it depends on the tool or method chosen, and the source of the information whether it is necessary to use local data collectors. It is important to always take into account the position of the people in the field. Sometimes, it is not only important to have local facilitators but:

Local data collectors should definitely be used when:

- Data needs to be collected in the local language.
- Information is collected from people in remote areas.
- Subjective topics need to be discussed. These can be very culturally dependent.
- You are looking for critical voices.

Local data collectors are not necessarily required when:

- Collecting 'objective information', such as existing financial or audit reports from Martin Bauer Group's suppliers or any other actors/persons with existing relations.

Data collectors should be well-trained and from the appropriate cultural and social background.

No matter whether collecting quantitative or qualitative data, it is crucial that the data collectors speak the local language, and are from the same, or at least very similar, cultural background as the interviewees. Depending on the context, speaking a different dialect or being from a different ethnic background may harm the data quality. This can be because certain questions may be misunderstood, as concepts may mean different things in different contexts, or because of trust issues. It is very important to check this in advance with local partners or contacts, who usually have a good grasp of the importance of these issues.

When specifically targeting certain groups, such as women or youth, data collectors should also be part of these groups.

It is always preferable, and oftentimes absolutely crucial, that focus groups consisting of for example only women are led by female facilitators. Again, local contacts or partners will be able to provide some information on how important this is in the local context of the project. The data collectors should know what their role and position is regarding the respondent. And take that into account in the interaction with the respondent. The interaction between a data collector and respondent is influenced by their backgrounds (where they come from, which group they belong to), jobs they have, roles and experiences of both persons amongst many other things. During the training, the data collectors should be sensitise about it so they can make sure their interaction with the respondent is as smooth as possible.

University students are often good data collectors.

It is useful if data collectors have experience with doing research, and know something about research methods themselves. Therefore, university students often make good data collectors. Depending on the project, it is also important that data collectors understand the contents of the project. This is especially the case in the more technical projects, e.g. on agriculture or the

environment. Local university students in those study directions are then a good option as well.

Independence of data collectors results in the most objective information.

When data collectors work for a company or cooperative buying or bulking & selling produce from the respondents, the respondents may not be entirely transparent and honest in answering the questions. For instance, they may, if asked about volumes produced and sold, mention only the columns sold to the buyer or through the cooperative. And not mention volumes sold elsewhere. Also, they are more likely to answer positively when queried about the project or programme, because they have a stake to keep the relationship well. Such bias is decreased when independent data collectors collect the data.

Training local data collectors

Collecting data requires certain skills and good understanding of study objectives. Therefore we advise to always train the data collectors.

It is very important to properly train any local data collectors.

If the tools for data collection, e.g. FGD and interview guides are developed elsewhere (e.g. Martin Bauer Group HQ), data collectors should get very familiar with them. All questions and topics should be very well-defined and discussed thoroughly. In case there are multiple data collectors, it is also very important they all understand every question in exactly the same way, and they would translate every question in exactly the same way. Therefore, a training should discuss each and every question. In particular, subjective concepts such as 'life opportunities' should be discussed thoroughly and reflected upon. Data collectors also provide a crucial local perspective on the tools. If they feel a question really does not make sense in the local context, it is most often a wise decision to change it. They know the context best, and know what works and what does not.

The training of data collectors should always include a test of the data collection tools.

No matter how well-prepared or thought through tools may be, they should always be tested. A first test should be done by data collectors themselves, for example in role plays. Data collectors should practice until they are completely familiar with the tool. In general, there is a second test as well, the so-called pilot testing. This test includes interviewees that are very similar to the target population ('real' interviewees). For example, if the target population consists of ginger farmers in village x, you might be able to do a test with either a few farmers growing a different crop in village x, or ginger farmers in village y. However, there are a few things to take into account while choosing your pilot test population:

- Your test population should be comparable enough to your target population to properly test your tools, but should not be part of your target population.
- Think about compensating people for their time with something they use, like soap or rice. These people do not benefit from the project after all.

During and after data collection

A good selection and training of data collectors is not enough. It is also key to keep in contact during and after the data collection, to develop good tools for data collection and to collect the data at the right times.

Data collectors are your eyes and ears in the process, and should therefore report back any issues.

When others are collecting the information, they are your main source of information on the data collection process. Therefore, they need to be well aware of this additional role, and understand the importance of reporting back any issues. These issues may include:

- **Sensitivities:** there might be (unforeseen) sensitive topics. Obvious sensitive topics such as income may be more difficult to discuss than expected, or topics that were not anticipated to be sensitive (e.g. number of children in the household) may give difficulties after all. This might influence the data quality without you knowing if the data collector would not tell you.
- **Unclarities/misunderstandings:** some questions might be hard for the interviewees to understand. Or, some questions might not be asked in the most appropriate way. For example, people might find it hard to calculate the total costs of inputs, but rather know the costs per package and how many packages they used.

Well-developed tools help a lot to harmonise data entry, and enable smooth analysis and interpretation.

Both qualitative as well as quantitative data collection demands well-developed tools. Martin Bauer Group already has a few of these tools, such as the needs assessment. Especially when multiple people will collect data, clear tools are crucial in harmonising the process. More information on what types

of tools to use can be found in the chapter on Research methods and tools. Additionally, to make sure the tools are supporting the data collection process the best they can, they should at least include:

- **Questions to be asked rather than broad topics to be discussed.** Different questions around the same topic can give very different results. This is especially important in quantitative data collection, but also helpful in qualitative data collection. However, in the latter, there can be examples of questions to guide the data collector (and remind him/her of the training of course); instead of one exact question to be asked. Of course, in the case of qualitative methods, these questions are not set in stone and can (and should be) adapted to the discussion/conversation or local context.
- **Clear instructions** on how to report the information. This can include many things, such as answer categories or instructions on whether to report how many participants of a FGDs agree with a certain answer. Key is that it is clear and understandable for the person entering the information, for which examples are always helpful.
- **Room for information that does not fit in the tool directly.** It is impossible to anticipate on every possible scenario when designing the tools. There might be something going on that you did not know about and that does not fit into the tool, which still might be very important information. For example, you are assessing whether the beehive project contributed to improved income diversification. The designed tool includes all kinds of questions on various income sources and many other things. However, there is no part asking about the reasons for changes in other income sources. Then, the resulting information might just show that income from one other income source, say: cassava production, has decreased tremendously. You will then not have information on why this happened. This is undesirable, because maybe it happened because of a pest or disease, or maybe because people are spending all of their time on the honey production, which might be an unintended negative effect of the project. Therefore, there always needs to be room to report information that seems relevant, but does not fit into the tool perfectly. Tools are there to help, not to limit.

Do not solely rely on the tools, but always speak to the data collectors too.

Tools are very important and helpful for data collection and reporting back to who is doing the analysis. However, relying on the tools as the only way of reporting back will often omit important information. The people who are collecting the data have been on the ground, and have therefore, besides collecting data based on the tools, been observing. These observations often provide valuable information. Observations are not only things that can be seen, but also the small conversations in between the real data collection processes, with drivers, restaurant owners, etc.

Availability, seasonality and timing of previous data collection should be taken into account when planning the data collection.

Many, if not all, of the target populations of Martin Bauer Group projects are heavily depending on agriculture or wild picking, both sectors with a lot of seasonality in terms of work load and income generation. This should be taken into account when collecting information from this target population. There is not one formula to do this, but there are several aspects to take into account:

- **Availability.** During harvest or picking season, people will likely not have time to answer many questions. When a respondent is in a rush to get back to work, this might harm the quality of the data collected.
- **Previous data collection.** If similar data is collected year after year, this should be done in the same time of the year. Different answers might be given during different times of the year, due to seasonality of income and work load. If information is gathered at a different moment each year, you might be measuring seasonality instead of change.
- **How much time it takes before a change can be assessed.** Some types of interventions can lead to changes happening quickly, other effects can take some time, sometimes even years, to materialise. Take this into account when planning data collection; it is a waste of money if data is collected too soon and the effect has not materialised yet.

Seasonality of data. Think about the potential seasonality of the information you want to collect. For example, if you want to learn about the type and amount of agrochemicals that are used, and they are only applied once a year, it is probably best to collect the information directly after the application period. Farmers will then best remember what they used. This will be much harder when it has been almost a year that they last applied agrochemicals. Therefore, if the focus of the data collection is seasonal, take this into account while planning the data collection.

Appendix 8 Reporting and concluding

What you report depends on the evaluation question and the timing of the evaluation activity.

Generally in evaluations the findings of the evaluation are reflected upon related to their relevance, effectiveness, impact, efficiency and sustainability. But this also depends on the timing of the evaluation. Early monitoring up to midterm evaluation for instance focus on presenting information on the relevance and effectiveness of the project, to be able to steer the interventions in another direction if required. While after the project has finalised, reporting focuses on concluding on the impact, efficiency and sustainability of a project.

If you report on effectiveness or impact, present findings on the status before and after the project and the change that happened since the start.

This should be done in an absolute and relative way, so people can judge whether a '30% change' is in reality a large increase or rather small. This is also possible in reporting on perceptions by indicating the % of respondents with a certain perception. It is important to always indicate the number of respondents, group discussions, or interviews in every visual and/or in the text, and also mention the number of people in the entire target group, to avoid people to misjudge the information presented. For instance if you present information from 10 interviews in 2 communities, while the project is implemented with 2000 people in 8 communities the reader judges the information in a different way than if you would have spoken with 100 people in 4 communities (if the respondents are sampled in a similar way).

Report on all factors contributing to change in indicators, reflect on your sphere of influence in creating change, and report on the representativeness of the findings for the whole target group.

Connect findings on changes in indicators between the different ToC levels, and explain the factors that influenced the results for different indicators/indicator levels, ranging from the contribution of the project to the influence of external factors outside the sphere of influence of the project.

Impact indicators are generally less within the sphere of influence of a project than outcome indicators. This should be reflected upon in the report, as well as the representativeness of the findings. For instance, mint farmers in Punjab, India, earn well below the living income benchmark. About 30% of their total household income comes from mint production. This means that a project to increase their income focusing on improving mint production is not likely to ensure an increase in income above the living income benchmark.

Triangulate the findings from different information/data sources to be able to make stronger conclusions on the contribution of the project to change.

Conclusions are more robust if they are supported by different sources of evidence. Your monitoring data may tell you that people who have participated in trainings indicated during the trainings that they have adopted certain practices. If such adoption rates are confirmed by interviews taking place at a later date this increases the strength of conclusions. This is especially the case when the respondent explains why and how this took place, and why the project (and not other factors) were the key reasons contributing to change.

Validate the findings with the local stakeholders to enhance the robustness of the results and facilitate learning.

An important aspect of the reporting process is to present draft results and conclusions with the local stakeholders, to ensure they can reflect upon the results and give feedback. This also facilitates learning by those stakeholders, as well as by the project implementers, which can then be processed in the final report.

Do not report on very exact numbers when this does not make sense, especially when they are based on qualitative research or quantitative research with short surveys and a small sample.

Reporting on numbers in much detail, for instance including decimals, may give the reader the idea that the information is very exact, while this is not necessarily the case. In reporting numbers (for instance income) based on qualitative research methods or in quantitative research methods with a small sample, round off those numbers broadly, and present it including the word 'about' so the reader can properly judge the exactness of the information. Also, explain in the methodology why you do this. A fictional example, based on 20 minute interviews with 10 people (target group is 2,000 people), is: 'The ten farmers interviewed in the state of Punjab India earn about USD 2 per household member per day, which is about 50% of the living income benchmark of USD 4 per household member per day'. If you would have presented the 'about USD 2' as 'USD 2,34', this would have seemed a very exact number, whereas such exact numbers can generally only be given by using a large and representative sample. It is therefore also important to be honest about the sample and the representativeness thereof, such as 'the ten farmers interviewed in the state of Punjab India' in the example. The reader will understand that interviewing ten farmers in an entire state in India does not give a representative picture of the whole state.

The way conclusions can be supported depends on the way indicators are measured and the counterfactual.

What conclusions can be drawn exactly depends on the research methods used, the way questions are asked and by whom and how often. Another factor is the way the counterfactual is build. Attributing the changes found in the evaluation to the project (whether observed changes, self-reported or perceived changes) is the main challenge in evaluations. An example conclusion, supported in a different way by each type of measurement is:

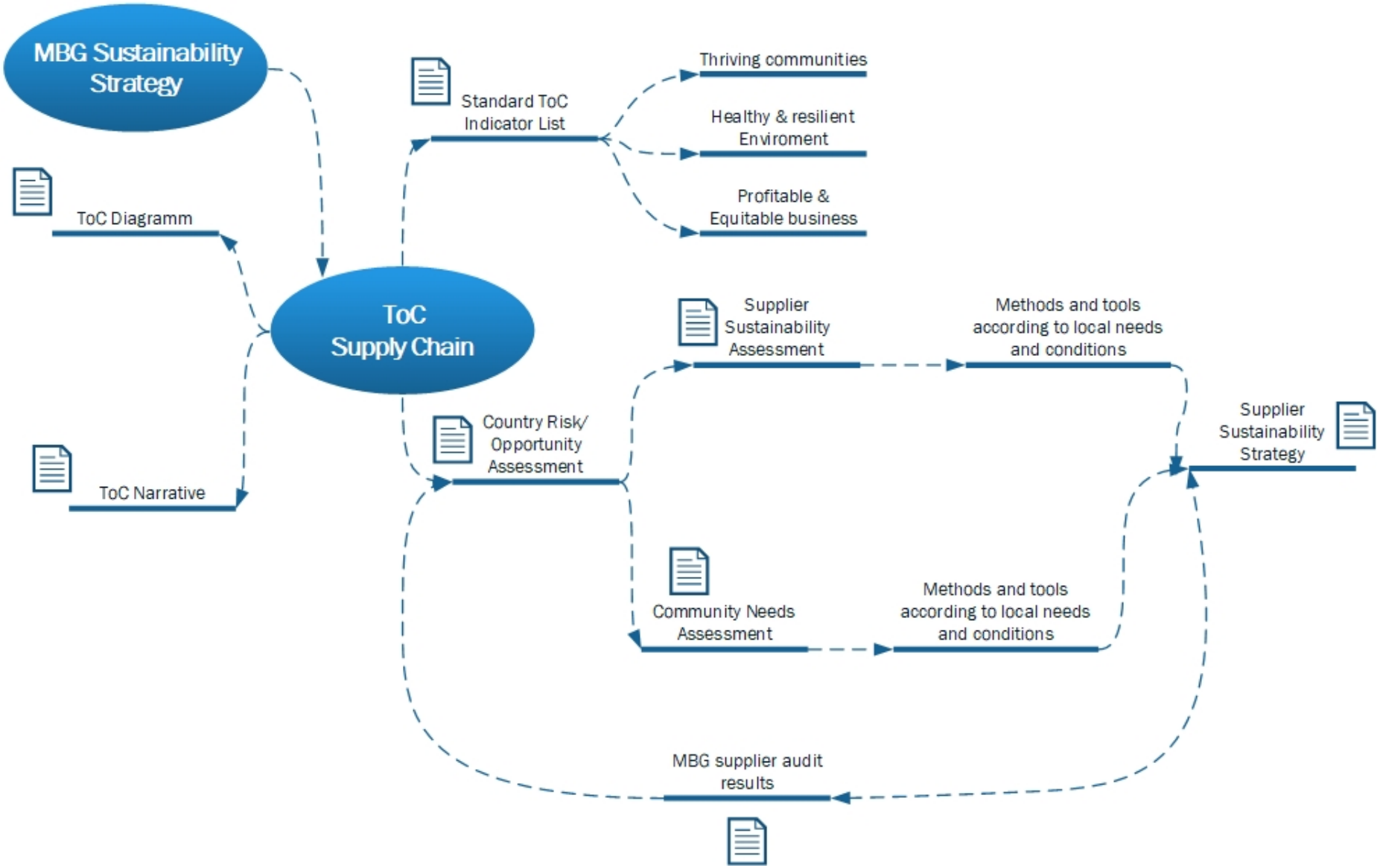
'The project has contributed to the implementation of pruning practices'

- Observable data: '20% more farmers implement pruning practices compared to two years ago'.
- Self-reported data: '20% more farmers report implementing pruning practices compared to two years ago'.
- Perceptions: '60% of the farmers perceive an improvement of their income levels compared to two years ago'.

However, the potential conclusions and supporting statements do not depend on the way the indicator is measured only, as the way the counterfactual is built also plays an important role. A few examples:

- Comparison group: 20% of the farmers implemented pruning practices more often compared to two years ago, this is 10% more than among comparison farmers (taking into account difference x, y, z)'
- Existing data: '20% of the farmers implemented pruning practices more often compared to two years ago, this is 10% more than the study by XX in year XX (taking into account difference x, y, z)'
- Self-reported contribution: '20% of the farmers implemented pruning practices more often compared to two years ago, 15% indicated this mainly resulted from the project'

Appendix 9 Overview of Martin Bauer Group documents and tools



Useful resources

There are many links to useful resources throughout this hand book. Furthermore, these are some additional resources that might be useful:

A guide to Assessing Needs by the World Bank (2012): Elaborate document from the World Bank on all aspects of needs assessments.

Better Evaluation: a lot of information on evaluating projects, from sampling, defining research questions and tools to reporting.

Bryman, A. (2016). Social research methods. Oxford university press.: Very detailed book on social research methods.

FAOSTAT: global data on agriculture-related indicators.

OECD Evaluation Criteria: The OECD's 6 evaluation criteria: relevance, coherence, effectiveness, efficiency, impact and sustainability explained.

M4SDI: A website promoting an integrated approach to planning, monitoring and evaluation. A lot of information on methods and approaches for all phases of M&E.

MSP Guide: A website on managing multi-stakeholder partnerships. Especially the tools and methods section is useful.

NCVO Know How – How to write an evaluation report: Very clear and hands-on guide to writing evaluation reports.

World Bank development indicators: global data on different country level development indicators, such as child mortality, literacy, etc.

Wageningen Economic Research
P.O. Box 29703
2502 LS The Hague
The Netherlands
T +31 (0)70 335 83 30
E communications.ssg@wur.nl
www.wur.eu/economic-research

Wageningen Economic Research
REPORT
2020-061