
ANNEXES FOR GEM ZAMBIA FINAL EVALUATION REPORT

Contents

ANNEX 1: INDIVIDUAL QUANTITATIVE SURVEY	2
ANNEX 2: COMMUNITY QUALITATIVE METHODS	5
ANNEX 3: SYSTEMIC CHANGE INTERVIEWS	8
ANNEX 4: REFLECTIONS ON FIELDWORK AND DATA ANALYSIS	13

ANNEX 1: INDIVIDUAL QUANTITATIVE SURVEY

The quantitative survey component for this evaluation consists of two type of analysis. The first is based on a quasi-experimental impact evaluation analysis, investigating the impact of the project on small holders farmers involved in the soya bean value chain, by using propensity-score matching (PSM). The second is based on summary representation of small holder farmers involved in the dairy value chain. This annex cover the survey sampling, the matching process, and an overall risk of bias table to assess the confidence in the results.

Sampling

Records from the project indicate that the project directly supported 2,927 small holder farmers (1,927 Heifer and 1000 SAP), of which 2,779 supported on the soya value chain, and 148 supported on the dairy production.

The survey identified an intervention group of 300 respondents directly supported by the project in the soya value chain (named the intervention group); and 400 respondents identified in neighbouring camps where Heifer and SAP did not conduct any activity (named the comparison group). A third group was identified consisting of 100 respondents were randomly selected from producer groups that were directly supported by the project in the dairy value chain. Considering women's empowerment as one of the main outcome indicators for the GEM project, interviews were conducted with women only. Information referring to income and production were collected at household level, while information referring to empowerment will be collected at personal level.

In the intervention group the sampling strategy consisted in a stratified two-stage random sampling. Out of the 2,779 farmers supported on the soya value chain by the project, 1,879 are women. To select 300 respondents, the two partners (Heifer and SAP) were given equal weight, and the clusters (in this case the villages) were selected according with a Probability Proportional to Size. Within the selected clusters, the respondents were then randomly selected for interview.

For identifying the comparison group, the evaluation team obtained a comprehensive list of all the farmers living in the four districts where the project was implemented. The evaluation team worked with staff from the partner organisations in identifying the camps in which they conducted activities. This included not only GEM activities, but also other activities conducted by the two partner organisations with similar projects. A total of 25 clusters with 5,983 women small holder farmers were identified as potential for comparison. A cluster for the comparison group was defined as the camp, or zone if available. From these 5,983 farmers, 400 were selected from a two-stage random sampling process. 20 clusters were selected with Probability Proportional to Size, and respondents within the cluster were then randomly selected.

As the lists of farmers provided by the DACO are updated on an annual base. It is possible that these lists might have included farmers that were not farmers in 2013, when the project activities started. To obviate this problem, the questionnaire had a screening question asking if the respondent was a farmer in 2013. If the response was negative, the enumerator was instructed to not proceed with the interview.

Propensity score matching

A central challenge presented in designing impact evaluation of social programmes is to identify *what would have been the case* without that programme having been carried out.

In impact evaluations based on a counterfactual approach, which are commonly used in the evaluation of programmes that involve a large number of units (whether individuals, households or communities), the evaluation design is based on a comparison between units that were subject to the programme and units

that were not.

In order to increase confidence that women in communities where the project was not implemented provide a reasonable counterfactual, the evaluation adopted a 'quasi-experimental' evaluation design based on propensity score matching. Where, women involved in the project were 'matched' with women with similar characteristics in non-project villages or comparison group.

Individual women were matched on the basis of a variety of characteristics – including household size, marital status, educational level, age, health status of the respondent and the head of the household, household wealth, amount of land owned, crop diversity, source of income, and group participation. Since some of these characteristics may have been affected by the project itself, matching was performed on the basis of these indicators *before* the implementation of the project.

Table below compares the intervention and comparison group before matching and after matching. It provides evidence that on average, *before matching*, households in the intervention group are among other things, more likely to be poorer, own more land, have greater crop diversification and more likely to being involved in groups. After matching there is no evidence that the two groups differ on any of these characteristics.

Variable	Un-matched Matched	Treated Mean	Comparison mean	% bias	% re-duction bias	t	p> t	V_e(T) / V_e(C)
hhsz2013	U	5.6733	6.2225	-22.4		-2.92	0.004	0.97
	M	5.6733	5.3747	12.2	45.6	1.53	0.126	1.23
maritalstatus_1	U	.72333	.6325	19.5		2.54	0.011	0.82
	M	.72333	.72162	0.4	98.1	0.05	0.963	0.99
resp_edu	U	.53	.53	0.0		-0.00	1.000	1.00
	M	.53	.49358	7.3	.	0.89	0.373	0.99
resp_abletowork	U	.99333	.9975	-6.2		-0.83	0.404	2.68**
	M	.99333	.98875	6.8	-10.1	0.60	0.552	0.56*
resp_age	U	48.097	48.445	-2.6		-0.34	0.734	0.98
	M	48.097	47.422	5.0	-93.7	0.60	0.547	0.90
hhh_male	U	.70667	.63	16.3		2.13	0.034	0.86
	M	.70667	.71223	-1.2	92.7	-0.15	0.881	1.01
hhh_age	U	55.05	54.855	1.4		0.18	0.853	0.99
	M	55.05	54.384	4.8	-241.3	0.56	0.578	0.82
hhh_abletowork	U	.91333	.9325	-7.2		-0.95	0.344	1.27*
	M	.91333	.90314	3.8	46.8	0.43	0.666	0.95
wealth1_2013	U	.27667	.1725	25.1		3.33	0.001	1.50*
	M	.27667	.27817	-0.4	98.6	-0.04	0.967	1.00
	U	.20667	.195	2.9		0.38	0.703	1.05

wealth2_2013	M	.20667	.19678	2.5	15.3	0.30	0.763	0.99
wealth4_2013	U	.21	.195	3.7		0.49	0.625	1.06
	M	.21	.2016	2.1	44.0	0.25	0.800	1.06
wealth5_2013	U	.08667	.2725	-49.8		-6.33	0.000	0.54*
	M	.08667	.08433	0.6	98.7	0.10	0.919	1.03
hh_land2013	U	6.929	4.0577	34.0		4.59	0.000	2.12**
	M	6.929	7.7091	-9.2	72.8	-0.88	0.377	0.83
cropnum2013	U	4.1467	2.745	66.9		8.97	0.000	1.39*
	M	4.1467	3.9831	7.8	88.3	0.89	0.372	1.07
othersinc2013_b	U	.52333	.535	-2.3		-0.31	0.760	1.00
	M	.52333	.52257	0.2	93.4	0.02	0.985	1.01
ngroups2013_b	U	.91333	.84	22.4		2.88	0.004	0.55*
	M	.91333	.88751	7.9	64.8	1.06	0.292	0.87

* if 'of concern', i.e. variance ratio in [0.5, 0.8) or (1.25, 2]; ** if 'bad', i.e. variance ratio <0.5 or >2

Sample	Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	% concern	% bad
Unmatched	0.187	179.23	0.000	17.7	11.7	108.9*	1.18	31	13
Matched	0.009	7.51	0.962	4.5	4.3	22.4	1.04	6	0

* if B>25%, R outside [0.5; 2]

ANNEX 2: COMMUNITY QUALITATIVE METHODS

Qualitative methods at the community-level were designed to provide insights into whether programme-related changes were occurring, particularly in difficult to measure or sensitive areas such as household decision-making, as well as how change happens over time for individuals, households and communities.

The primary method was via participatory focus group discussions with tools designed¹ and conducted separately for different groups in each community: village leaders; women producer group (PG) members; male PG members or family members of PG members (mostly husbands); and women PG members and non-members. Since men were not included in the household survey, focus groups with husbands and male family members was an important means of discovering attitudes and perceptions towards women's agency and leadership. Some tools such as the vignettes were used in discussions with both men and women to understand the range of perspectives and especially highlight areas of agreement and dissonance. The focus groups with husbands, village leaders and women who did not belong to the PGs were also designed to investigate evidence of spill-over effects from GEM activities and training amongst the wider community. In each focus group, a trained facilitator obtained informed consent from participants and then led them through a series of participatory exercises and questions, while a dedicated note-taker took detailed field notes that were later transcribed for analysis.

Table: Focus Group Discussion tools, objectives and target groups

FGD Tool	Objectives	Link to evaluation area	Who
Village entry + Social mapping	Changes in the external context / enabling environment at the village level	Rich picture of villages Understanding project's contribution	5-7 village leaders (incl. village head, head of the producer groups, teacher, other)
Ladder of Life	Changes in opportunities for smallholder farmers Availability of agricultural and market services. Who uses these? Who is excluded? What are the barriers to use for women? Benefits and disadvantages/risks of producer groups, and barriers to entry	Quality of income: stability, predictability, regularity of income Quality of market services Resilience strategies Perceptions of PGs and collective action	Women PG members Women non-PG members
Vignettes (included private voting)	Intra-HH decision-making, specifically: Women's influence on decisions on production practices Women's influence on how to spend income Women's influence on going outside the home (for community activities, for work outside) Women's influence on care responsibilities Women's influence on decisions of large investments and assets	Ability to engage in decision-making at HH and community Influence decisions on production practices Contribution to HH income and decision-making on spending income Women's decision-making power in communities	Women PG members Male partners of PG members

¹ The design of the focus group discussion tools was informed and adapted primarily from two sources: a) the GENNOVATE research initiative examining gender norms and agency in agriculture and natural resource management (methodology accessible at http://gender.cgiar.org/wp-content/uploads/2018/02/GENNOVATE-Methodology_Feb2018_FINAL.pdf); b) van Hemelrijck, A. (2017) *Governance in Myanmar: Evaluation of the 'Building equitable and resilient livelihoods in the Dry Zone' project, Effectiveness Review Series 2015/16* (<https://policy-practice.oxfam.org.uk/publications/governance-in-myanmar-evaluation-of-the-building-equitable-and-resilient-liveli-620177>).

	Relative influence over group decisions		
Power spectrum (included private voting)	<p>Meaningful decision-making at the group/community level, specifically:</p> <p>Women's influence in mixed-sex spaces at community level</p> <p>Women's power beyond their group and community</p> <p>Do women influence community decisions that will improve their well-being? How? Who?</p>	<p>Ability of women to engage in decision-making at community</p> <p>Knowledge of rights and relationship between legal interventions and well-being</p>	<p>Women PG members</p> <p>Husbands of PG members</p>

Village sampling was undertaken at the start of the evaluation in consultation with local partners. A set of village sampling criteria and village selection was informed by discussions with Oxfam staff and partners, balancing feasibility with a desire to include communities with elements of the following characteristics: range of districts where GEM was implemented; range of geographical conditions (especially related to ease of market access); representing targeted value chains (soya and dairy value chains); representing both partners (Heifer and SAP); and a range of GEM interventions (beyond the core training, such as the Rapid Care Analysis and 'I Care About Her Campaign' and use of a milk collection centre).

Table: Schedule of Community Focus Group Discussions (August-September 2017)

Date	Village name	Brief description
25-26 Aug	Mokambo	Geographical: Quite difficult to access town and markets, far from milk collection centre. Value chain: Dairy. Partner: Heifer. GEM interventions beyond core: RCA, Milk collection centre. Piloting/testing the qualitative methods (focus group discussions with a mixed group, in-depth interviews with producer groups, participant observation).
29-31 Aug	Chisangwa	Geographical: Remote, far from market access; 'traditional' village. Partner: Heifer. Value chain: Soya. Village has undergone a number of changes in the last 10 years; they have new services such as a new health clinic and marketplace, a bar, a borehole, etc. Primary school until grade 7. Land is given by a local chief, not the government. People have little to no transport options though the recent introduction of cows has enabled people to use ox carts to transport in cases of medical emergency. No ongoing development projects mentioned other than GEM/Heifer.
1-2 Sept	Kamaluko	Geographical: Moderately good road access, relatively close to the milk collection centre (cycling distance). Value chain: Dairy. Partner: Heifer. GEM interventions beyond core: RCA, Milk collection centre (not yet operating). This is a 'traditional' village, bordered by forest. Relatively good access to a main road and some jobs with local timber company. A few local schools up to grade 9.
4-6 Sept	Kamilili	Geographical: Remote, far from market access. Partner: SAP. Value chain: Soya. The village is a new farming block, which used to be a national forest reserve. The people are yet to be issued with land titles from the government. Difficulty accessing education and health services. Few jobs apart from small stores and a mine just across the border with Congo. No ongoing development projects mentioned other than GEM/SAP.
7 Sept	Mutupa	Geographical: Peri-urban 'compound' part of a larger municipality, on the roadside. Partner: SAP. Value chain: Soya. GEM interventions beyond core: Crop (soya) bulking centre with a hammer mill This is a large and high-density community with good mobile network access but not connected to the electricity grid. Easy market access and jobs in trading, as well as farming and charcoal burning. Local school up to grade 9 and a clinic. Good local transport, including frequent public minibuses. No ongoing development projects mentioned other than GEM/SAP.

Within communities, village leaders and/or producer group members were informed of the tentative discussion schedule prior to the evaluation team's arrival to help with participant recruitment. Participation of a wide range of project participants and non-participants and community leaders was encouraged, as well as a demographic diversity (e.g. representing a range of ages). The participants self-selected often depending on capacity to reach the meeting venue and time availability, though efforts were made once in the community to actively invite harder-to-reach populations especially non-project participants. The ideal number of participants in the discussions was 8-12 but that varied depending on community, with some having many more than that.

The qualitative evaluation team also stayed with local families in the communities visited to conduct informal interviews, life histories and do direct observation, to see and experience some of the more invisible topics of the evaluation, such as changing attitudes, norms and dynamics in the household. Staying in the communities also served to enhance trust and familiarity with the community leaders and focus group participants, increase the potential that harder to reach populations (especially non-project participants) could approach them in their own time and leave open the opportunity to unveil some potential 'unknown unknowns'.

ANNEX 3: SYSTEMIC CHANGE INTERVIEWS

The Systemic Change Interview approach (SCIA) draws on Outcome Harvesting, Most Significant Change and Collaborative Outcome Reporting approaches. The overall aim is for project stakeholders to identify changes and outcomes; identify gaps in information needed to explain the changes (specifically in relation to a market systems project); and build up a list of questions for specific stakeholders based on these information gaps. Key informant interviews can then be conducted with these stakeholders to verify and elaborate upon these changes, after which the change descriptions are revised and updated.

Step 1: Identify the levels and domains of the Theory of Change to be investigated

Being a markets systems programme GEM has a substantial focus on systemic change and strategies to work with government departments, private sector, media and civil society as well as smallholder producers to facilitate change in the targeted value chains. When such change happens, it may not be limited to a specific intervention group or direct beneficiaries – the aim is to facilitate positive changes that benefit agricultural producers in targeted value chains, whether or not they are direct project participants. These changes are therefore difficult to capture from a household survey that compare intervention and comparison groups or from qualitative methods at the community level if they are only being conducted in project areas.

Techniques such as Outcome Harvesting, Most Significant Change and Collaborative Outcome Reporting are useful approaches that enable the collection of data for complex projects such as market systems programmes. If used well they can also be a means of capturing unintended and negative outcomes of a project intervention. For this reason, the GEM evaluation trialled an approach to understand the outcomes of the project in relation to market systems change. During the Theory of Change workshop held in Kitwe on August 21-22 2017 a number of research questions were identified by project stakeholders² which were collected through SCIA.

Step 2: Design questions to enable collection of change statements from project stakeholders and documentation

Step 2a: Questions to ask stakeholders to prompt thinking and discussions

At the Theory of Change workshop participants were asked to record and share their experience of what had changed in the last three years, specifically changes regarding their relations with other stakeholders. Participants were asked to give examples of one positive and one negative change. To help participants think about these changes they were given a guiding question: “In the last three years, what positive and negative changes have you experienced in working with the private sector and local agribusinesses, government departments, civil society or smallholder farmers?”

Step 2b: Get participants to complete templates for each change identified

For each of the changes they identified, participants were asked to complete a template to capture details on what the change was, when and where it took place, who was involved in the change and what were the consequences of the change. Some of the answers were vague or focussed mostly on personal change in income and farming practices, however a total of 36 outcome descriptions were collected of which 21 were taken forward into the next stage of selecting the outcomes.

² Approximately 35 male and female participants attended the workshop, consisting of representatives from the government, private sector, dairy industry, media and communities, as well as Oxfam staff, partner staff and evaluation team.

Step 2c: Outcome trawl of project documentation

A mid-term review of GEM in Zambia was conducted in early 2016. Outcomes identified in this report were also collected and added to the outcome descriptions collected in the Theory of Change workshop. The same information for step 2b was documented for these outcomes.

Step 3: Grouping, selecting and categorising change statements

Step 3a: Grouping change statements

At this point, a number of the changes identified were overlapping so it was necessary to group and categorise these outcome descriptions and collate the information into one outcome description, this led to the formulation of 17 outcome descriptions:

Increased diversification of income opportunities for smallholder farmers.	Increased quality of milk received by Parmalat.
Increased number of SHF in dairy and soya value chains in the Copperbelt Region.	Small-scale milk producers unable to utilise full production capability due to inadequate storage and bulking facilities.
Increase in time and labour spent on expanding land area under production.	Lack of negotiation power in dairy value chain due to dealing with only one buyer.
Decrease in price of soya in 2017.	Private sector more willing to engage with SHF due to improved organisation.
Soya now being collected within the communities by buyers rather than SHF having to arrange and pay for their own transport.	Increased collaboration of actors and networking model enriched relations among market players.
Increased negotiating power of soya producers.	Businesses expanding to new areas in Copperbelt as a result of interaction with other stakeholders.
There is a not enough soya seed to meet demand of SHF.	Women in PGs able to influence government and private sector.
Increase in weather information available for SHF.	Increase in extension services specifically from private sector.
PGs doing value addition of soya products to produce High Energy Protein Supplements for babies in local communities.	

Step 3b: Selecting change statements

The next step was to determine which of these 17 outcomes we wanted to investigate further. For this the main criteria was that it related to either (1) the research questions identified in the Theory of Change workshop; (2) the market services domain of the Theory of Change; or (3) was a systemic change that wouldn't be captured from the household survey or community focus group discussions. It was decided that all 17 research questions met at least 1 of the criteria so in this instance none were excluded. N.B. Between step 2c and 3a a number of outcomes were excluded when writing the outcome descriptions. It would be more transparent to include all the outcome descriptions identified and then exclude during step 3b.

Step 3c: Categorising change statements

Once the outcomes to be investigated are selected we can categorise the outcomes into four groups: positive intended; positive unintended; negative intended and negative unintended. You can see in the table below that the majority of outcomes are positive intended and that there is a split of eight positive and five negative outcomes. This categorising process helps to check if we have collected a range of outcomes and that we are not only checking on positive outcomes. Ideally in this process we would have identified more unintended (both positive and negative) outcomes. More time dedicated to harvesting outcomes in the Theory of Change workshop may have helped.

Figure 1: Outcomes categorised into positive, negative, intended and unintended

Positive Outcomes		
Unintended Outcomes	<ul style="list-style-type: none"> PGs doing value addition of soya products to produce High Energy Protein Supplements for babies in local communities. 	<ul style="list-style-type: none"> Increased number of SHF in dairy and soya value chains in the Copperbelt Region. Soya now being collected within the communities by buyers rather than SHF having to arrange and pay for their own transport. Increased negotiating power of soya producers. Increase in weather information available for SHF. Increased quality of milk received by Parmalat. Businesses expanding to new areas in Copperbelt as a result of interaction with other stakeholders. Women in PGs able to influence government and private sector. Increase in extension services specifically from private sector. Increased diversification of income opportunities for smallholder farmers.
	<ul style="list-style-type: none"> Increase in time and labour spent on expanding land area under production. Decrease in price of soya in 2017. There is a not enough soya seed to meet demand of SHF. 	
Negative outcomes		Intended outcomes

Step 4: Drafting initial details of the change statement description, identifying information gaps and associated informants

Step 4a: Detailing the outcome description

This step involves adding more detail to the existing outcome description. Currently the description will be limited to the information collected in step 2; what the change was, when and where it took place, who was involved in the change and what were the consequences of the change. Based on a review of a paper written by ITAD³, a number of failings in market programme evaluations were identified. These generally referred to concepts that were missing or lacking from M4P evaluations including scale, sustainability and systemic change. These areas have therefore been incorporated into a template for the outcome description so as to ensure capture of relevant information in these areas. Below is an outline of the sub-

³ Ruffer, T. & Wach, E. (2013). *Review of M4P Evaluation Methods and Approaches*. ITAD. Commissioned by UK DFID.

headings in the outcome description template. Content is added under each sub-heading by the evaluator based on what they already know about the project, from what they have heard or read in the workshops and documentation.

Figure 2: Sub-headings from outcome description template

Description: (What was the change; when and where did it happen; who was involved in the change; and what was the consequence of the change?)
Outcome identified from: (Name of stakeholder, report, reference)
Outcome verified by: (Name of stakeholders interviewed for verification)
Contribution: (What did GEM do that contributed towards this outcome? What other actors were significant contributors?)
Evidence of systemic change: (Crowding in, copying, replication, sector growth, backward and forward link-ages)
Scale: (What is the estimated reach of this outcome? How many smallholder farmers/indirect beneficiaries is it likely to impact? How was this identified?)
Sustainability: (Static or dynamic? Commercial viability, investment by private or public sector, innovation, organisational capacity).
Impact on women: (How have women been specifically involved/impacted by this change?)
Negative consequences: (What are the actual or potential negative impacts of this change? Do these differ for men and women?)

Step 4b: Identify information gaps and associated informants

For ease of explanation, steps 4a and 4b are separated here but in practice these steps were combined. While detailing the outcome description (step 4a) a number of gaps in the information become apparent and it is in this step that we record all these gaps and questions we have about the outcome and determine where we can get this information from. Some of the missing information may be answered from the community level qualitative research and some from the household survey but we will also need to interview other project stakeholders to better understand these outcomes and to understand their perspective.

Step 5: Drafting Key Informant Interview questions

Once the information gaps have been identified in step 4 it is then possible to group these gaps/questions according to the associated informant that is required to be interviewed. The information gaps will need to be structured into an interview checklist

Step 6: Conducting and documenting key informant interviews

Key informant interviews were conducted by the qualitative research consultant in-person and via telephone. Interviews took place in September 2017 with the following key informants/organisations:

No.	Type of Stakeholder
1	Dairy Association Zambia (DAZ)
2	Radio Chimwemwe
3	Radio Icengelo

4	District Agricultural Community Officer - Kitwe
5	Heifer Projects International – GEM Partner
6	Sustainable Agriculture Programme (SAP) – GEM Partner
7	Agricrops
8	ASTRAL
9	Bimeda
10	Parmalat
11	SEBA Foods
12	Local Seed Grower

ANNEX 4: REFLECTIONS ON FIELDWORK AND DATA ANALYSIS

Quantitative methods

The survey field team provided several important reflections on their experience and suggestions for improving for similar work in the future, including:

- The household lists for both the intervention and comparison areas were not up to date and contained people who had died, or relocated. The sample therefore, contained some of these households. The large number of absent respondents especially in the comparison areas affected the sampling. The absent farmers were replaced by other farmers from the within the community using strictly random methods. In the intervention areas it appeared that a number of respondents had relocated or died. The suggestion therefore, is that before the sampling exercise is done in future exercises, an up-to-date register is used for the sampling frame so as to minimise on the number of replacements. The same should be done in the comparison areas. The farmer register remains the most credible data base but this should be revised and made as updated as possible.
- There was resistance by extension staff in Ndola to participate in the data collection exercise and they ended up referring the team to contact farmers. The reluctance is borne out of the established norm in the department that financial considerations should be paid to the extension officers even when they are working in their camps as long as the work involves other stakeholders other than government. The team had to use contact farmers in Ndola and had to visit the district on three occasions with the fourth visit being on the last day of the data collection exercise.
- The two cars assigned to the quantitative survey team for the first two weeks of the exercise were not enough as this limited the number of sites the enumerators could go to each day. As such instead of achieving an average of 80 interviews per day, the team was only able to achieve 60 at the most until the third vehicle was made available in the last week of the exercise. In our proposal we had indicated three vehicles for an exercise such as this. The third vehicle helps to mobilise and move team members around and meet targets. The results of the availability of the third vehicle showed that we were able to interview 80 respondents per day except for instances where the farmers simply didn't show up in areas where the exercise was arranged.
- One major lesson learnt from this exercise is the need to provide for sufficient time to notify the target survey communities and arrange to have accurate village and household information. While this was done by SAP and Heifer project and field staff to a certain extent, as CSK we would have liked to have been in the forefront of mobilisation of communities. This would ensure that the sample selected is much more apt and also provide for more practical scheduling of activities. Sufficient time also allows the survey team to clearly explain the objectives of the survey to the target communities to avoid instances of misinformation.
- Another lesson is the need to ensure full participation of stakeholders such as project field staff and extension officers especially in the comparison areas. Engagement with these officers before data collection would help enlist their full participation and also address the other expectations such the those of financial nature.

It is important to remember that not all quasi-experimental impact evaluations are the same. Choices made during sampling, selection of the comparison group, and at the analysis stage are crucial in assessing the overall level of confidence in the results. The table below helps to assess the risk of bias against ten predetermined parameters defined specifically for ex-post quasi-experimental impact evaluations. Lower overall risk provides higher confidence in the results.

Table: Risk of Bias Assessment

	Title	Description	Risk assessment	Description
Sampling				
1	Random sampling	<p>Score LOW risk if: Sampling is conducted using probability random sampling methods on a clearly established sample frame.</p> <p>Score MEDIUM risk if: Sampling is conducted using probability random sampling methods at geographical level (e.g. village level), and use random walk to select respondents within the geographical area.</p> <p>Score HIGH otherwise.</p>	LOW	<p>Respondents are selected from clearly defined sampling frames.</p> <p>The sample frame for the intervention group comes from the list of project participants. The intervention group is selected using a stratified two-stage random sampling. Partners are stratified with equal weight, and cluster (villages) selected by Probability Proportional to Size. Respondents within the cluster are then randomly selected.</p> <p>The sample frame for the comparison group comes from the comprehensive list of all the farmers living in the four districts, provided by the District Agricultural Coordinators in each district.</p>
2	Representativeness of project participants	<p>Score LOW risk if: Project participants have been involved for the entire duration of the project and have been involved in the project with the same level of exposure. Project participants have been exposed to a variety of different activities, some may have dropped out from some activities, but sampling is conducted on the entire list of project participants.</p> <p>Score MEDIUM risk if: Project participants have been exposed to a variety of different activities. Sampling is conducted only among those project participants that have been enrolled for the entire duration of the project or that have been enrolled in all the activities. These are not less than 80% of the</p>	LOW	<p>Sampling is conducted on the entire list of project participants, regardless on the intensity of the intervention.</p>

		entire list of project participants OR it is clear the results apply only to a particular group of project participants. Score HIGH otherwise.		
Selecting comparison group				
3	Potential for contamination (spillovers)	<p>Score LOW risk if:</p> <p>The units for comparison group are selected in geographical areas where it is not reasonable to expect for the project to have had spillover effects.</p> <p>The project also implemented some activities (which are not considered the most relevant under analysis) which are expected to have had an impact also in the comparison group. (e.g. the project implemented campaigns using radio and other digital media, but these are only a minor component of the activities implemented). The report makes clear which impact is assessed (added-value of other components, taking into account exposure to those minor components)</p> <p>Score HIGH risk if:</p> <p>Units for the comparison group are selected within the same geographical area as the intervention group, and it is reasonable to expect that project activities had spillover effects. (e.g. comparison observations within the same village, for awareness raising projects)</p>	MIX	<p>Unit for the comparison group are selected within the same district. If the project was successful, some activities at market system level might have been positively affected the comparison group.</p> <p>The analysis defines which outcome indicators would have been affected (system level indicator), and which would remain at household and community level.</p> <p>For those indicators at system level (e.g. access to markets), the analysis does not attempt to compare them between intervention and comparison, but rather assess the prevalence on the entire sample.</p>
4	Self-selection of project participants	<p>Score LOW risk if:</p> <p>The comparison group is exploiting an experiment or natural experiment.</p> <p>Units are randomly selected at community level both in the intervention and comparison group.</p> <p>The selection process for the comparison group is mimicking the same selection process used by the project.</p>	MEDIUM	<p>Small holder farmers directly involved in the project are farmers that decided to enrol in cooperatives or producer groups. It might be possible that there are unobservable characteristics that are making these farmers different (e.g. aptitude). The matching procedure will attempt to correct for this by controlling for group participation at baseline.</p>

		<p>Score MEDIUM risk if</p> <p>The self-selection is corrected during matching procedure (e.g. controlling for group participation at baseline)</p> <p>Score HIGH risk if:</p> <p>Project participants were selected or self-selected based on idiosyncratic or unobservable characteristics, and the selection of comparison respondents is done randomly from neighbouring geographical sites.</p>		
5	Other interventions in the comparison group	<p>Score LOW risk if:</p> <p>There are no other actors in the area (e.g. INGOs, NGOs, governmental programmes)</p> <p>Other actors are conducting activities which are not linked to the project's theory of change</p> <p>Score MEDIUM risk if:</p> <p>Other actors are conducting similar activities in both the intervention and the comparison group</p> <p>Score HIGH risk if:</p> <p>Other actors are conducting similar activities, partially related in the comparison communities</p> <p>Other actors are conducting activities in the comparison communities, which are not the same, but are partially related to the project's theory of change.</p>	LOW	Discussions with the DACOs did not reveal other organisations working in the area. Data from the village survey does not reveal other significant organisations working on similar thematic areas.
Analysis				
6	Representativeness	<p>Score LOW risk if:</p> <p>During analysis or matching procedure less than 10% of the sample in the intervention group is excluded.</p> <p>Score HIGH risk if:</p> <p>During analysis or matching procedure more than 10% of the sample in the intervention group is excluded.</p>	LOW	Only 7 of respondents (less than 1%) were drop.

7	Robustness checks	<p>Score LOW risk if: Magnitude and statistical significance of the results are approximately consistent with different econometric models.</p> <p>Score HIGH risk if: Results are not consistent with different econometric models and sub group analysis.</p>	N/A	Robustness checks have not been conducted.
8	Triangulation	<p>Score LOW risk if: Results are triangulated and consistent with other evaluation methods within the same evaluation. Results are triangulated and consistent with other data on the same project but from different evaluations.</p> <p>Score HIGH risk if: Results are not consistent or triangulated with other evaluation methods.</p>	LOW	Qualitative and quantitative results are providing similar results
9	Multiple hypothesis testing	<p>Score LOW risk if: Multiple hypothesis tests apply Benjamini or Bonferroni tests. The evaluation drafted a pre-analysis plan prior data analysis, and followed the plan.</p> <p>Score HIGH otherwise</p>	LOW	A pre-analysis plan was drafted and followed, except for household decision, where the variable specification was changed to be consistent with the variable construction for Tajikistan and Bangladesh.
10	Clustering	<p>Score LOW risk if: Clustering is applied Clustering was tested but rejected as providing higher standard errors than non-clustering estimates.</p> <p>Score HIGH otherwise.</p>	LOW	Clustering was tested but rejected as providing higher standard errors than non-clustering estimates. In order to provide more conservative estimates the analysis was conducted without clustering.

Qualitative methods

In general, the combination of household survey with in-depth community group discussions and stakeholder interviews, building on a participatory theory of change workshop to elicit local priorities for evaluation questions, was a viable evaluation model. It does, however, require resources and time in order to deliver high quality outputs, and this may not be feasible or cost effective for all programmes.

The qualitative analysis does not attempt to allow for causal inference, but does provide valuable context, triangulation, and nuance to the statistical analysis. Further, the qualitative data has shed light on important aspects of programme implementation and the policy environment in which Oxfam and local partners operate. This data has been essential to understanding how and why the project leveraged impact (or failed to). Focus group data was designed to ensure the lived experiences of project participants was adequately captured in the evaluation, both as a unique source of knowledge and reflection on how the programme delivered improvements (or did not) and in order to reflect Oxfam's values through the co-production of insight and learning for future programming. The quality of focus group and interview data necessarily relies on highly skilled facilitation and the use of research tools to engage participants in a meaningful and authentic way and the application of vignettes for community-level discussion was one attempt to do so.

For the community qualitative component, the field team was meant to stay in the communities to provide time to conduct the focus group discussions, observe and reflect and spend time discussing with other community members (including non-project participants). It was acknowledged that this could have potential risks and the field team was encouraged to only do so if accepted by the community and if they felt safe and comfortable. In one community, the field team chose not to stay due to feeling unsafe.

For the key informant interviews, the plan was to interview at least 16 different stakeholders ranging from the project teams, government officials and representatives of the key private sector partners. While effort was made to contact all the planned key informants, it was not possible to interview those from MRI, SeedCo and Varum. The main contacts from MRI and SeedCo that were familiar with the GEM project were out of the country and thus it was not possible to identify the appropriate replacements. There were suggestions for other officials from, for example, Seedco, that were suggested and met, but were unable to respond to any of the questions of the interview checklist.
