

---

# ANNEXES FOR GEM BANGLADESH FINAL EVALUATION REPORT

---

## Contents

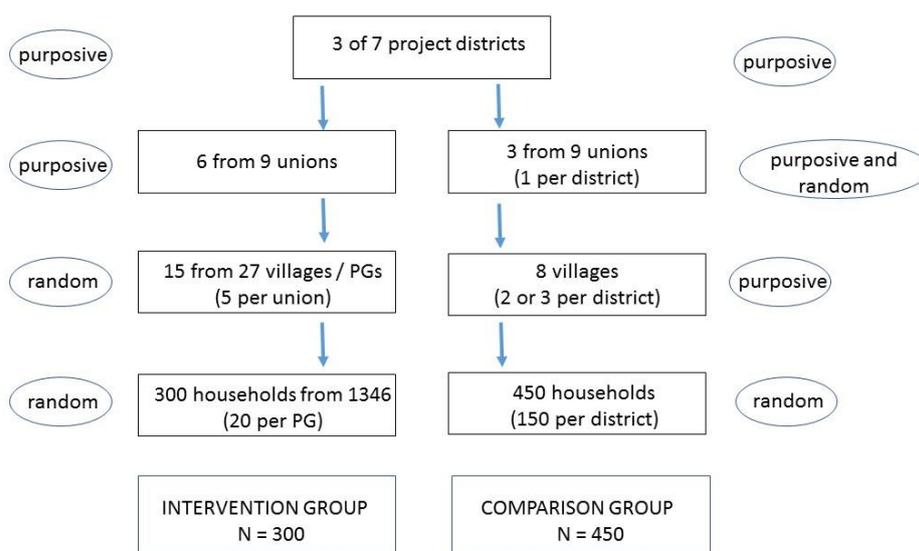
ANNEX 1: HOUSEHOLD SURVEY AND QUASI-EXPERIMENTAL ANALYSIS .....	2
ANNEX 2: FOCUS GROUPS .....	8
ANNEX 3: SYSTEMIC CHANGE INTERVIEWS .....	9
ANNEX 4: ADDITIONAL TABLES OF RESULTS .....	14
ANNEX 5: LIMITATIONS OF FIELDWORK AND ANALYSIS .....	21

# ANNEX 1: HOUSEHOLD SURVEY AND QUASI-EXPERIMENTAL ANALYSIS

## Sampling

Sampling was undertaken using a two-stage stratified random sampling method. Stratification was undertaken through purposive discussion of various criteria to establish broad fieldwork sites (at sub-district level) in three districts. Households were randomly selected from a clearly established sample frame. The sampling strategy took careful account of local conditions, notably the effects of heavy flooding over previous months, the potential for further weather events as it was the monsoon season still, and the expectation that a team of 10 female enumerators would be in the field and the particular security and logistical needs the fieldwork team would need to be mindful of. An early decision was made to interview female participants only in the GEM evaluation, and in Bangladesh this implied excluding 2% of project participants who were male. Project eligibility criteria was determined by having one or two adult cows in 2014, and this was used as a screening question for the comparison group. The figure below summarises the sampling decisions taken and whether the process was based on random sampling or purposive screening and selection.

Figure A1. Sampling strategy overview



## Matching: constructing an appropriate counterfactual

The household survey ultimately involved 747 households, 300 from the intervention group (40%) and 447 from the comparison group (60%), who were approached by the survey team during September 2017. The main analysis for the quasi-experimental component is conducted by comparing outcome variable means between the two households groups using propensity-score matching (PSM).

The propensity score is calculated using the following outcome indicators:

- Household size in 2013
- Education of the household head
- Education respondent
- Marital status respondent
- Crop diversity in 2013

- Number of cows owned in 2013
- Land owned in 2013
- Household income engagement in 2013
- 1[household in the 1<sup>st</sup> quintile wealth index in 2013<sup>1</sup>]
- 1[household in the 2<sup>nd</sup> quintile wealth index in 2013]
- 1[household in the 4<sup>th</sup> quintile wealth index in 2013]
- 1[household in the 5<sup>th</sup> quintile wealth index in 2013]

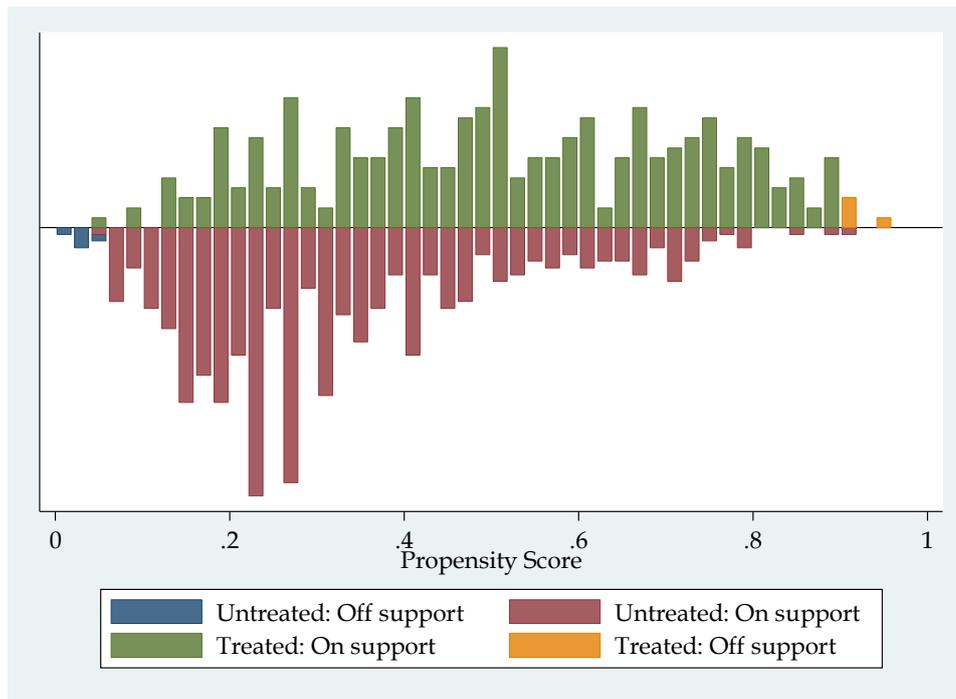
Table A1 below compares intervention and comparison group before matching, providing evidence that on average, *before matching*, households in the intervention group have a larger number of household members, were likely to be poorer, and have greater crop diversification. There were no differences between the two groups on the education level of the household head, the respondent's marital status, the number of cows owned, the area of land owned, or involvement in other sources of income other than crop farming.

Table A1: Descriptive Statistics Before Matching	Intervention mean	Comparison mean	Difference
HH size in 2013	4.247	3.776	0.470***
1[HHH has education greater than primary]	0.180	0.206	-0.026
1[Respondent has education greater than primary]	0.180	0.206	-0.026
1[Respondent is married]	0.933	0.953	-0.020
Crop diversification 2013	3.503	1.987	1.517***
Number of cows owned in 2013	2.857	2.875	-0.018
How much land did you own in 2013?	41.417	37.023	4.394
1[HH involved in sources of income other than crop in 2013]	0.777	0.792	-0.015
1[HH in the 1st wealth quintile]	0.253	0.226	0.027
1[HH in the 2nd wealth quintile]	0.190	0.145	0.045
1[HH in the 4th wealth quintile]	0.190	0.206	-0.016
1[HH in the 5th wealth quintile]	0.120	0.253	-0.133***
Observations (for each variable)	300	447	

In order to control for the observable differences, Propensity Score Matching was implemented. As a result, 9 observations were excluded as being outside the common support area (4 from the intervention group and 5 from the comparison group). Observations are weighted accordingly with their propensity score.

<sup>1</sup> Wealth index calculated using Principal Component Analysis based on household asset ownership and housing characteristics based on information referring to 2013. Questions 201 – 217, and 327 – 328 in Bangladesh survey refer.

Figure A2. Common support area for propensity score matching



## Balancing test

Following the matching process, none of the matching variables were significantly different as set out in the table below.

Table A2: Descriptive Statistics After Matching

Variable	Un-matched/ Matched	Treated Mean	Control Mean	%bias	%reduct bias	t-test t	t-test p>t	V_e(T)/ V_e(C)
hysize2013	U	4.23	3.80	26.1		3.44	0.001	0.73*
	M	4.23	4.11	7.2	72.3	0.88	0.381	0.8
hhh_edu	U	0.18	0.20	-4.2		-0.56	0.574	0.94
	M	0.18	0.18	0.9	78.8	0.11	0.911	1.04
resp_edu	U	0.18	0.20	-4.2		-0.56	0.574	0.94
	M	0.18	0.18	0.9	78.8	0.11	0.911	1.04
maritalstatus_1	U	0.94	0.95	-5.9		-0.79	0.429	1.27*
	M	0.94	0.95	-2.7	53.9	-0.32	0.75	1.1
cropnum2013	U	3.45	2.00	77.7		10.56	0	0.89
	M	3.45	3.42	2	97.4	0.22	0.824	0.91
asset_selected2013_1	U	2.87	2.89	-0.8		-0.11	0.912	0.9
	M	2.87	2.81	2.5	-194.2	0.31	0.753	1.15
hh_land2013	U	39.80	31.13	13.8		1.84	0.067	0.93
	M	39.80	34.67	8.2	40.8	0.98	0.327	0.9
othersinc2013_b	U	0.77	0.80	-5.5		-0.74	0.46	1.06
	M	0.77	0.80	-5.6	-1	-0.68	0.498	1.13
wealth1_2013	U	0.26	0.23	6.6		0.88	0.379	1.09

	M	0.26	0.25	2.6	60.6	0.31	0.756	1.05
wealth2_2013	U	0.19	0.15	11.3		1.52	0.13	1.27*
	M	0.19	0.21	-5.7	49.2	-0.65	0.516	0.92
wealth4_2013	U	0.19	0.21	-3.9		-0.52	0.606	0.95
	M	0.19	0.17	6.6	-69.8	0.84	0.403	1.06
wealth5_2013	U	0.12	0.24	-32.1		-4.17	0	0.65*
	M	0.12	0.11	2.4	92.7	0.34	0.734	1.11

\* 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	0.143	142.3	0	16	6.2	94.9*	1.25	33	0
Matched	0.004	3.15	0.989	3.9	2.6	14.6	0.91	0	0

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

## Project exposure

Before considering the project's effect on outcomes, it is important to examine whether the respondents reported having participated in the activities implemented under this project. The data confirmed that respondents had experienced a number of training activities spanning farming practices as well as business skills and to support leadership, negotiation and assertiveness techniques, and awareness of women's rights. As shown by the table below, there were statistically significant differences between the intervention and comparison groups for each type of GEM project activity.

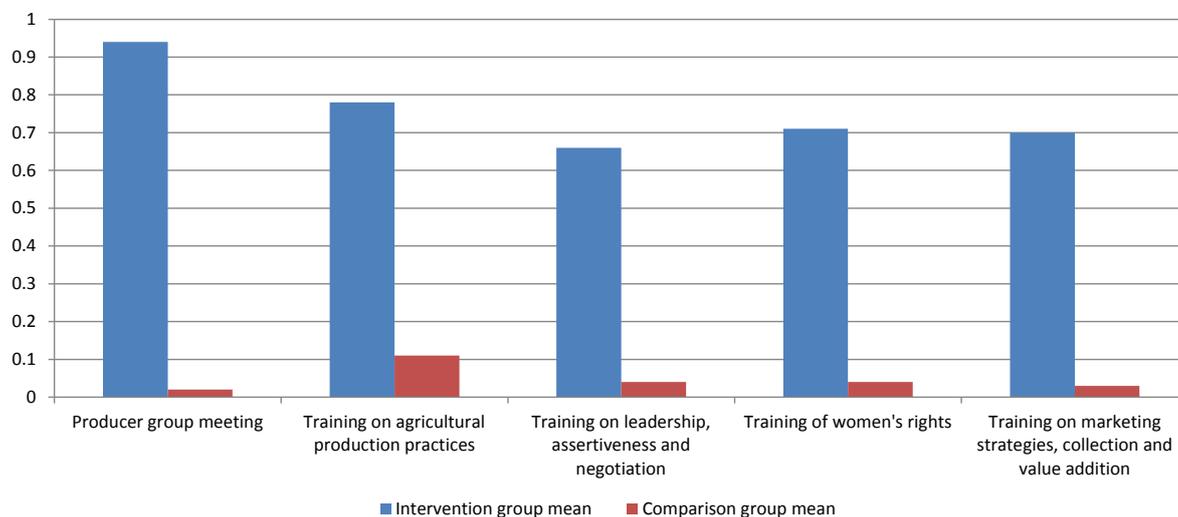
	1	2	3	4	5
Table A3: Project Exposure	Producer group meeting	Training on agricultural production practices	Training on leadership, assertiveness and negotiation	Training of women's rights	Training on marketing strategies, collection and value addition
Intervention group mean	0.94	0.78	0.67	0.71	0.71
Comparison group mean	0.03	0.10	0.03	0.04	0.02
Difference:	0.90***	0.68***	0.64***	0.67***	0.69***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Observations (intervention group)	296	296	296	296	296
Observations (total)	738	738	738	738	738

Standard errors clustered at village level in parentheses; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ ; PSM estimates are bootstrapped with 1000 repetitions.

These statistical checks confirm that the intervention and comparison groups, while broadly similar in key demographic and geographical aspects, provide a strong contrast in their exposure to the kinds of

interventions GEM sought to provide. This was the basis for the quasi-experimental analysis reported in this evaluation, allowing for the differences between the intervention and comparison groups to be attributed to the GEM project. As with all matching-based statistical analysis, the findings rely on their being minimal unobservable differences between the intervention and comparison group. The evaluation design and household survey sampling process aimed to minimise any risk of bias arising from unobservable differences, and present in this report the most robust available estimates of project impact outside of randomised control trial methodology.

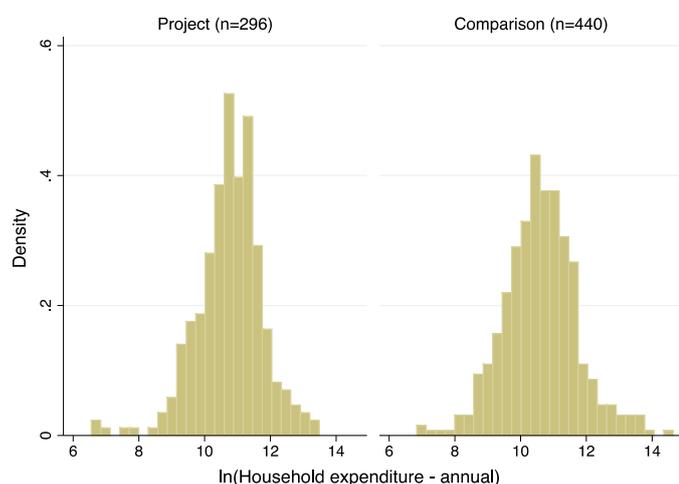
Figure A3. Exposure to GEM-type activities



## Household expenditure per annum

Annual spending was estimated by summing a number of activities that were expected to take place less frequently than those included in monthly expenditures. Expenses in the past 12 months included community events and ceremonies, health costs, rent or purchase of land for farming, building materials and repair for the home, clothes and shoe purchases, and purchase of livestock. Clear differences emerged in the average level of household spending between project and comparison groups, as set out in the main report. The distribution of spending took a non-normal form, and as per standard practice the variable was transformed by natural logarithm to present a normal distribution (figure A4). Project impact was found to be statistically significant at the 5% level using log household spending.

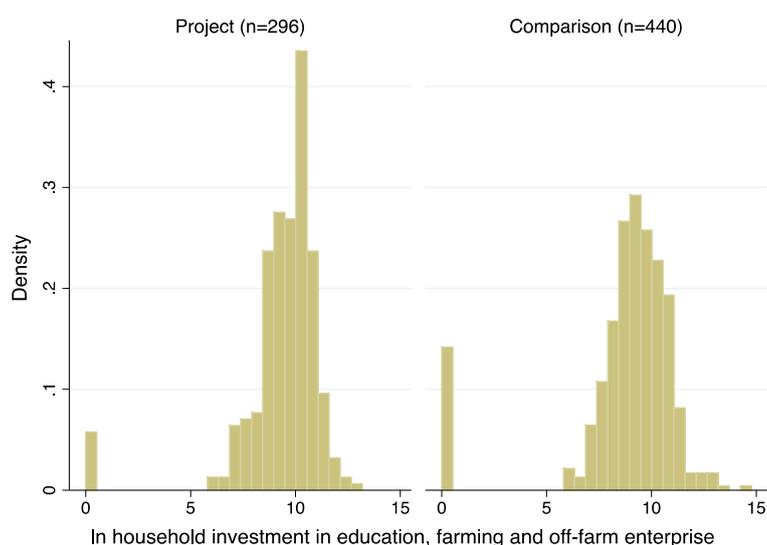
Figure A4. Logarithmic form of annual household spending



## Household investment

This variable was defined as annual household spending on three types of activity: education, farming equipment, and off-farm enterprise. Investment was measured in Taka, and statistical analysis transformed the data into logarithmic form to better understand and compare trends. In both quasi-experimental groups, a minority of households reported zero investment (3% in the project group, and 8% in the comparison group). Project impact was found to be statistically significant at the 1% level using log household investment.

Figure A5. Logarithmic form of annual household investment



## Female-headed households

With approximately 10% of the sample reporting as female-headed, and after applying the matching methods, the sample size for analysis falls to 77 observations. Statistical power to detect significant project effects is reduced considerably. The results are therefore indicative at best. The data suggests female-headed households had higher annual dairy income of over 15,000 Taka, and greater annual household spending (both statistically significant). Effects are weaker but positive for monthly household spending. No significant effects are detected for food consumption or household investment.

Table A4: Female headed households	Food consumed over past 7 days (per capita per day, calories)	Log household investment	Annual dairy income (Taka)	Log monthly household expenditure	Log annual household expenditure
Intervention group mean (n=38)	94.89	9.15	24109	7.26	10.58
Comparison group mean (n=39)	88.69	8.26	9013	6.88	10.07
Difference:	6.20	0.884	15095**	0.378*	0.508**
(std error)	(10.13)	(0.67)	(7123)	(0.22)	(0.25)
Observations (total)	77	77	77	77	77

## ANNEX 2: FOCUS GROUPS

The evaluation incorporated valuable community-level feedback through a series of focus groups across three GEM districts. A total of four communities were targeted, and in each location four focus groups were conducted to ensure the voices of distinct stakeholders were heard: (i) women belonging to GEM dairy Producer Groups (PGs), (ii) village leaders, (iii) husbands of women belonging to GEM PGs, and (iv) women who did not belong to GEM PGs. Approximately 6 to 8 individuals took part in each focus group discussion, from 20 to 85 years of age.

Since men were not included in the household survey, focus groups with husbands was an important means of discovering attitudes and perceptions towards women's empowerment and leadership amongst the male members of their households. The focus groups with husbands, village leaders and women who did not belong to the PGs were also designed to investigate evidence of spillover effects from GEM activities and training amongst the wider community. In each focus group, a trained facilitator led participants through a series of questions and research tools including vignettes, while a second researcher took detailed field notes that were later transcribed for analysis.

The list of locations and dates of focus group meetings is set out in Table B1. Sampling was undertaken at the start of the evaluation in consultation with local partners. Taking the household survey locations as a starting point, the evaluation design team selected communities where community level qualitative data gathering would take place. Site selection for focus group discussions was based on criteria including physical access, extent of damage and recovery from the 2017 floods, and a desire to balance data gathering across river island and mainland communities. Due to timing and resource constraints focus groups only took place in GEM communities and not comparison communities.

**Table A5: Schedule of Four Community Focus Group Discussions (October 2017)**

Date	Village and District	Brief description of village
2 – 3 Oct	Panchasona, Sirajgonj	Mainland village, no boat travel required, was within 5 mins walk of a road. Some village services identified including a madrasah, small enterprise and petty trading, but overall a quiet scene. Only NGO active was MMS as part of GEM. The significant event of the past 12 months was flooding, which destroyed the rice fields.
4 – 5 Oct	Bara Ghorjan, Sirajgonj	<i>Char</i> village, required travel by boat and is not within 5 mins walk of a road. More dispersed settlement, but nonetheless was busy with enumerators identifying children and women, a doctor and school teacher. It had a modest market scene of a small grocery shop and public services appeared less well established. Significant events of the past 12 months did not mention flooding.
6 – 7 Oct	Paschim Deluabari, Gaibandha	<i>Char</i> village, required travel by boat and is not within 5 mins walk of a road. Small enterprises and petty trading was visible to fieldwork enumerators, and a primary school was located there. Only NGO active in the village was part of GEM/REECALL. Enumerators noted it was a quiet village. There had been positive changes noted in the past 12 months linked to training, livestock practice, and GEM activities. But floods were the most significant recent event, submerging cropping fields and causing damage to houses and part of the road network.
8 – 9 Oct	Taluk Shabaj, Rangpur	Mainland village, no boat travel required, was within 5 mins walk of a road, with proximity to the river. Services in the village included a high school, market and a dispensary. Only SEED was active in the village as part of GEM/REECALL project. The significant event of the past 12 months was flooding, with crops unable to be grown due to the fields being underwater.

# 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 to use 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**

GEM being a markets project, has a substantial focus on systemic change and has a strategy to work with government departments, private sector, media and civil society as well as smallholder producers to facilitate change in the dairy 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 techniques 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 is using SCIA to understand the outcomes of the project in relation to market systems change. During the Theory of Change workshop held in Gaibandha between 17<sup>th</sup> and 19<sup>th</sup> September a number of research questions were identified by project stakeholders, which will also be 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 we asked participants 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 23 change examples were collected.

### **Step 2c: Outcome trawl of project documentation**

A 3<sup>rd</sup> year MEAL report for GEM in Bangladesh was conducted in 2016/17. 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 19 outcome descriptions:

Small-scale dairy producers are spending more on inputs leading to lower than anticipated profit margins.	PG member households spending income on child education and health
Insufficient availability of livestock services at community level, i.e. AI, input and veterinary services, vaccines.	Households experiencing increased conflict between male and female members
Increased number of stakeholders providing input supply services – including feed and medicine/vaccinations and AI.	Increased workload for women entrepreneurs at household level
Cost of fodder has increased substantially	Increased decision making power for women at household level
Increase in percentage of owned cows receiving full vaccination courses and treatment	Increased social acceptance among men that women can earn and contribute towards family income
Increase in the percentage of heifer cows becoming pregnant through AI for improved breeds	PG members are collectively demanding services from government departments (District Livestock Offices)
Milk quality and production has increased	Bangladesh Government prepared draft of National Dairy Development Policy (NDDP) in 2016.
Household income has increased	Bangladesh Bank financed BDT 200 Crore (2 billion) fund for dairy farmers at 5% interest rate.
PG member households able to purchase more land	Loan scheme of Bangladesh Bank not implemented well

### Step 3b: Selecting change statements

The next step was to determine which of these 18 outcomes we wanted to investigate further. For this the main criteria was that it related to (1) the research questions identified in the Theory of Change workshop; (2) the market services stream 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.

Change statement	ToC Research Question	Market services stream	Not in survey or FGDs	Selected for SCIA
Small-scale dairy producers are spending more on inputs leading to lower than anticipated profit margins.		X		Yes
Insufficient availability of livestock services at community level i.e. AI, input and veterinary services, vaccines.	X	X		Yes
Increased number of stakeholders providing input supply services – including feed and medicine/vaccinations and AI.		X		Yes

Cost of fodder has increased substantially		X		Yes
Increase in percentage of owned cows receiving full vaccination courses and treatment	X	X		Yes
Increase in the percentage of heifer cows becoming pregnant through AI for improved breeds	X	X		Yes
Milk quality and production has increased	X			No
Household income has increased	X			No
PG members' households able to purchase more land			X	No
PG member households spending income on child education and health	X			No
Households experiencing increased conflict between male and female members				No
Increased workload for women entrepreneurs at household level				No
Increased decision making power for women at household level				No
Increased social acceptance among men that women can earn and contribute towards family income	X			No
PG members are collectively demanding services from government departments (District Livestock Offices)		X	X	Yes
Bangladesh Government prepared draft of National Dairy Development Policy (NDDP) in 2016.			X	Yes
Bangladesh Bank financed BDT 200 Crore (2 billion) fund for dairy farmers at 5% interest rate.		X	X	Yes
Loan scheme of Bangladesh Bank not implemented well		X	X	Yes

It was decided that of the 18 change statements, based on the criteria 10 would be taken forward for the SCIA component of the GEM evaluation in Bangladesh.

### Step 3c: Categorising change statements

Once the changes to be investigated are selected we can categorise the changes 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 six positive and four 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. The collection of change statements should be amended to better facilitate the collection of unintended changes. More time dedicated to harvesting outcomes in the Theory of Change workshop may have helped and also the framing of questions on what has changed in a way that is more likely to create a thought process about unintended changes.

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

<b>Positive Outcomes</b>		<b>Intended outcomes</b>
<b>Unintended Outcomes</b>	<ul style="list-style-type: none"> <li>• Increased number of stakeholders providing input supply services – including feed and medicine/vaccinations and AI.</li> <li>• Increase in percentage of owned cows receiving full vaccination courses and treatment</li> <li>• Increase in the percentage of heifer cows becoming pregnant through AI for improved breeds</li> <li>• PG members are collectively demanding services from government departments (District Livestock Offices)</li> <li>• Bangladesh Government prepared draft of National Dairy Development Policy (NDDP) in 2016.</li> <li>• Bangladesh Bank financed BDT 200 Crore (2 billion) fund for dairy farmers at 5% interest rate.</li> </ul>	
	<ul style="list-style-type: none"> <li>• Small-scale dairy producers are spending more on inputs leading to lower than anticipated profit margins.</li> <li>• Insufficient availability of livestock services at community level, i.e. AI, input and veterinary services, vaccines.</li> <li>• Cost of fodder has increased substantially</li> <li>• Loan scheme of Bangladesh Bank not implemented well</li> </ul>	
<b>Negative outcomes</b>		

## 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 ITAD<sup>2</sup>, 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-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

<b>Description:</b> (What was the change; when and where did it happen; who was involved in the change; and what was the consequence of the change?)
<b>Outcome identified from:</b> (Name of stakeholder, report, reference)

<sup>2</sup> Ruffer, T. & Wach, E. (2013). *Review of M4P Evaluation Methods and Approaches*. ITAD. Commissioned by UK DFID.

<b>Outcome verified by:</b> (Name of stakeholders interviewed for verification)
<b>Contribution:</b> (What did GEM do that contributed towards this outcome? What other actors were significant contributors?)
<b>Evidence of systemic change:</b> (Crowding in, copying, replication, sector growth, backward and forward link-ages)
<b>Scale:</b> (What is the estimated reach of this outcome? How many smallholder farmers/indirect beneficiaries is it likely to impact? How was this identified?)
<b>Sustainability:</b> (Static or dynamic? Commercial viability, investment by private or public sector, innovation, organisational capacity).
<b>Impact on women:</b> (How have women been specifically involved/impacted by this change?)
<b>Negative consequences:</b> (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 and October 2017 with the following key informants:

No.	Type of Stakeholder
1	SKS - GEM Partner
2	MMS - GEM Partner
3	SEED - GEM Partner
4	Artificial Insemination (AI) provider (Rangpur)
5	Feed Supplier -ACI (Rangpur)
6	District Livestock Officer (Rangpur/Gaibandha)
7	PRAN Dairy (Rangpur/Gaibandha)
8	Rupali Bank (Rangpur)
9	Department of Livestock Services - Policy Contact (Dhaka)
10	Secretary, Multi-stakeholder Forums
11	OXFAM Bangladesh

# ANNEX 4: ADDITIONAL TABLES OF RESULTS

**Table A6a: Dairy income indicators**

	5	6	7	8	9	10	11
	Produced milk in the last week	Total quantity of milk PRODUCED in the last week (litres)	Total quantity of milk SOLD in the last week (litres)	Total value of sales of milk SOLD in the last week (Taka)	Average revenues per milking cow (Taka)	Dairy income – annual (Taka)	Buyer of milk was milk collection centre for producer group
Intervention group mean	0.72	19.46	16.30	645.13	469.00	33546.83	0.20
Comparison group mean	0.65	7.25	3.73	176.73	138.22	9190.13	0.00
Difference:	0.07	12.21***	12.49***	467.30***	327.99***	24299.67***	0.20***
	(0.09)	(2.34)	(2.35)	(106.95)	(72.74)	(5736.43)	(0.02)
Observations (intervention group)	141	102	99	99	99	99	285
Observations (total)	263	184	180	180	180	180	691

**Table A6b: Household income indicators: food consumption, income from farming and change in household wealth**

	1	2	3	4
	Value food consumed in last 7 days, daily per person equivalent	Log(Value food consumed in last 7 days, daily per person equivalent)	Change in household wealth	ln(Household income from crop farming)
Intervention group mean	102.09	4.54	0.01	9.14
Comparison group mean	91.40	4.44	0.05	8.64
Difference:	10.74***	0.10***	-0.04	0.52**
	(3.53)	(0.04)	(0.04)	(0.23)
Observations (intervention group)	296	296	296	232
Observations (total)	738	738	738	429

**Table A6c: Household income indicators: income from non-farming sources and household expenditure**

	5	6	7	8	9
	ln(Income from sources other than crop)	Household expenditure - annual	ln(Household expenditure - annual)	Household expenditure - monthly	ln(Household expenditure - monthly)
Intervention group mean	6.70	73574.27	10.76	2858.19	7.35
Comparison group mean	6.84	65986.62	10.54	2779.19	7.42
Difference:	-0.15	7506.34	0.21**	57.96	-0.07
	(0.16)	(6648.20)	(0.09)	(695.83)	(0.08)
Observations (intervention group)	244	296	296	296	296
Observations (total)	608	738	738	738	738

**Table A7: Quality of income indicators**

	1	2	3	4	5
	Stability of income (Months in a year you have at least one cow giving milk)	Frequency of income (Receive income daily or weekly)	Predictability of income (Received amount of income as expected)	Timeliness of income (Received income when expected)	Sufficiency of income (Selling assets to cover household expenses)
Intervention group mean	4.99	0.07	0.43	0.84	0.46
Comparison group mean	4.03	0.01	0.43	0.88	0.42
Difference:	1.02**	0.06***	-0.00	-0.04	0.04
	(0.45)	(0.02)	(0.06)	(0.04)	(0.04)
Observations (intervention group)	231	229	232	232	296
Observations (total)	428	415	429	429	738

Standard errors clustered at village level in parentheses; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ ; PSM estimates are bootstrapped with 1000 repetitions.

**Table A8: Dairy farming practices**

	1	2	3	4
	Ever used a separate feed and water manger in the last 12 months	Ever given your live-stock concentrated feed in the last 12 months	Ever used artificial insemination (AI) services in the 3 years	Vaccinated your cow in the last 12 months
Intervention group mean	0.78	0.94	0.77	0.93
Comparison group mean	0.41	0.60	0.15	0.46
Difference:	0.37***	0.34***	0.61***	0.46***
	(0.04)	(0.04)	(0.04)	(0.04)
Observations (intervention group)	296	296	296	296
Observations (total)	738	738	738	738

Standard errors clustered at village level in parentheses; \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01; PSM estimates are bootstrapped with 1000 repetitions.

**Table A9: Participation in household decision making indicators**

	Farming practice	How to spend money made from farming	Education of children	Going to see the doctor	Who takes care of family members	Personal travel to relatives	Personal participation in community group	Family planning	Who works to earn money
Intervention group mean	0.75	0.69	0.86	0.91	0.97	0.90	0.93	0.82	0.65
Comparison group mean	0.59	0.56	0.78	0.80	0.97	0.75	0.55	0.77	0.38
Difference:	0.16***	0.13***	0.08**	0.10***	0.00	0.15***	0.38***	0.05	0.27***
	(0.04)	(0.05)	(0.04)	(0.03)	(0.01)	(0.04)	(0.04)	(0.04)	(0.04)
Observations (intervention group)	296	296	296	296	296	296	296	296	296
Observations (total)	738	738	738	738	738	738	738	738	738

Standard errors clustered at village level in parentheses; \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01; PSM estimates are bootstrapped with 1000 repetitions.

**Table A10a: Household unpaid care work indicators**

	1	2	3	4
	Men support in-creased: care of elderly or disabled	Men support in-creased: care of children	Men support in-creased: cooking, cleaning or washing	Men support in-creased: collecting water or firewood
Intervention group mean	0.48	0.54	0.27	0.31
Comparison group mean	0.28	0.32	0.10	0.12
Difference:	0.20***	0.23***	0.17***	0.19***
	(0.04)	(0.04)	(0.04)	(0.04)
Observations (intervention group)	296	296	296	296
Observations (total)	738	738	738	738

Standard errors clustered at village level in parentheses; \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01; PSM estimates are bootstrapped with 1000 repetitions.

**Table A10b: Household unpaid care work indicators**

	1	2	3	4
	Ask help: care of elderly or disabled	Ask help: care of children	Ask help: cooking, cleaning or washing	Ask help: collecting water or firewood
Intervention group mean	0.46	0.54	0.33	0.47
Comparison group mean	0.38	0.47	0.23	0.28
Difference:	0.09**	0.08*	0.11**	0.19***
	(0.04)	(0.05)	(0.04)	(0.05)
Observations (intervention group)	290	291	290	291
Observations (total)	725	724	722	723

Standard errors clustered at village level in parentheses; \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01; PSM estimates are bootstrapped with 1000 repetitions.

**Table A10c: Household unpaid care work indicators**

	1	2	3	4	5
	Acceptability of men doing care work: care of elderly or disabled	Acceptability of men doing care work: care of children	Acceptability of men doing care work: cooking, cleaning or washing	Acceptability of men doing care work: collecting water or firewood	No satisfied with the division of labour
Intervention group mean	0.50	0.52	0.21	0.28	0.02
Comparison group mean	0.43	0.44	0.09	0.12	0.04
Difference:	0.07	0.08*	0.12***	0.15***	-0.02*
	(0.05)	(0.05)	(0.03)	(0.04)	(0.01)
Observations (intervention group)	296	296	296	296	296
Observations (total)	738	738	738	738	738

Standard errors clustered at village level in parentheses; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ ; PSM estimates are bootstrapped with 1000 repetitions.

**Table A11a: Women's empowerment and leadership indicators**

	1	2	3	4
	Women's economic empowerment <i>(If a woman earns more money than her husband, it's almost certain to cause problems)</i>	Acceptability of market role <i>(On the whole, men make better entrepreneurs than women)</i>	Care responsibilities <i>(If the wife is working outside the home, the husband should help her with household chores)</i>	Collective action <i>(It is better for women to work together to solve problems than working alone)</i>
Intervention group mean	0.62	0.33	0.95	0.94
Comparison group mean	0.67	0.47	0.82	0.89
Difference:	-0.06	-0.14***	0.13***	0.05*
	(0.04)	(0.04)	(0.04)	(0.03)
Observations (intervention group)	296	296	296	296
Observations (total)	738	738	738	738

Standard errors clustered at village level in parentheses; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ ; PSM estimates are bootstrapped with 1000 repetitions. Empowerment measured positively when respondent says 'disagree' or 'strongly disagree' to first two statements, and 'agree' or 'strongly agree' to last two statements.

**Table A11b: Women's empowerment and leadership indicators**

	1	2	3	4
	Perception of self-esteem <i>(I feel I have a number of good qualities)</i>	Confidence <i>(I continue to work on hard tasks even when others oppose me)</i>	Business skills/confidence <i>(Even when my farm is doing well I keep my eyes open in case I find a way to improve it)</i>	Leadership skills <i>(I feel comfortable speaking out at a meeting of men and women to help decide on infrastructure (roads, water supplies, well) to be built in my community)</i>
Intervention group mean	0.91	0.64	0.84	0.66
Comparison group mean	0.66	0.39	0.75	0.39
Difference:	0.24***	0.24***	0.09**	0.27***
	(0.04)	(0.05)	(0.04)	(0.04)
Observations (intervention group)	296	296	296	296
Observations (total)	738	738	738	738

Standard errors clustered at village level in parentheses; \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01; PSM estimates are bootstrapped with 1000 repetitions. Means reported on proportion of women who 'agree' or 'strongly agree' with statement.

**Table A12: Gender based violence indicators**

	1	2	3
	NOT accept any violence	Received any violence in the last 12 months	Reported, in case of violence
Intervention group mean	0.39	0.63	0.34
Comparison group mean	0.35	0.84	0.19
Difference:	0.06	-0.21***	0.14**
	(0.06)	(0.05)	(0.06)
Observations (intervention group)	160	158	101
Observations (total)	388	381	278

Standard errors clustered at village level in parentheses; \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01; PSM estimates are bootstrapped with 1000 repetitions.

**Table A13: Exposure to 2017 floods indicators**

	1	2	3	4	5
	Affected by floods to some extent or severely in 2017	Affected by floods SEVERELY in 2017	Sufficiency of income: selling assets to cover household expenses	How much cash did you raise from all sales? (Taka)	Preparedness: No seek weather info to help prepare for potential floods
Intervention group mean	0.92	0.44	0.46	49983.48	0.09
Comparison group mean	0.72	0.24	0.42	39872.43	0.10
Difference:	0.19***	0.20***	0.04	9183.51	-0.01
	(0.04)	(0.04)	(0.05)	(5684.17)	(0.02)
Observations (intervention group)	296	296	296	135	296
Observations (total)	738	738	738	278	738

Standard errors clustered at village level in parentheses; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ ; PSM estimates are bootstrapped with 1000 repetitions.

**Table A14: Consequences of 2017 floods indicators**

	Evacuated home temporarily	Number of days evacuated	Anyone in the household suffer illness or injury due to the flood water	Crop or milk production was disrupted	Lost cows due to the flood	How much cash compensation did you receive? (Taka)	Value of any compensation you received in kind (food, clothes) (Taka)
Intervention group mean	0.35	8.89	0.45	0.59	0.03	116.97	197.44
Comparison group mean	0.23	16.84	0.42	0.46	0.05	17.71	185.20
Difference:	0.12***	-5.80**	0.04	0.14***	-0.03	102.21**	7.95
	(0.05)	(2.45)	(0.05)	(0.05)	(0.02)	(48.90)	(43.74)
Observations (intervention group)	271	95	271	271	271	271	271
Observations (total)	597	171	597	597	597	597	597

Standard errors clustered at village level in parentheses; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ ; PSM estimates are bootstrapped with 1000 repetitions.

# ANNEX 5: LIMITATIONS OF FIELDWORK AND ANALYSIS

All fieldwork faces limitations, and this evaluation exercise was no exception. The timing of data collection coincided with the monsoon season and, following heavy rainfall and flooding earlier in the year, this made for logistical challenges in travel and reaching pre-determined survey locations. As far as possible, the survey team relied on local partners' knowledge of the area and transportation links to reach project participants even in some difficult to access villages. This enabled reasonable coverage of diverse geographical areas, notably the river islands (*char* villages). A pre-agreed sampling technique for alternative locations in comparison unions (sub-district level) also ensured minimal bias where sampled villages and households either could not be reached or contacted. The use of tablets minimised risks to fieldwork operations, ensuring secure collection and transmission of data and promoting extensive travel to enhance coverage of the three districts chosen for the evaluation exercises.

This evaluation exercise deliberately relied on a combination of quantitative and qualitative methods and data, in order to enhance both our understanding of the results and context, and the validity of findings. The analysis has relied on quasi-experimental methods in order to allow for causal inference and determine the impact of the project on those it reached. The statistical rigour of these methods depends to a great deal on the quality of data and the robustness of the matching approach used. Household data was subject to a series of quality assurance checks both during fieldwork and once the data was collated. The matching approach also underwent a number of statistical procedures to select the right matching algorithm, and replication materials can be made available on request. Ultimately, the statistical analysis relies on the comparison group being an appropriate counterfactual for the project intervention, and the matching process ensures as far as possible this is the case based on observable characteristics. There remains the possibility of unobserved variables introducing bias and reducing the validity of results. However, the triangulation of statistical analysis provides an important source of triangulation to sense-check and corroborate findings. Further investigation of the 'risk of bias' is set out in table A15.

The qualitative analysis does not attempt to allow for causal inference, but does provide valuable context, triangulation, and nuance to the statistical analysis. Further, qualitative data has shed light on important aspects of programme implementation and the policy environment in which Oxfam and local partners operate. This data is 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.

Lessons learned from this evaluation exercise include the following. 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.

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. This document provides a framework to assess the risk of bias against ten predetermined parameters, specifically for ex-post quasi-experimental impact evaluations. Lower overall risk, provides higher confidence in the results.

**Table A15: Risk of Bias Assessment**

	Title	Description	Assessment	Description
Sampling				
1	Random sampling	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>Sampling is conducted using probability random sampling methods on a clearly established sample frame.</li> </ul> <p>Score MEDIUM risk if:</p> <ul style="list-style-type: none"> <li>Sampling is conducted using probability random sampling methods at geographical level (e.g. village level), and use random sampling to select respondents within the geographical area.</li> </ul> <p>Score HIGH otherwise.</p>	Low	<p>Sampling was conducted using probability random sampling methods on a clearly established sample frame for both intervention and comparison households.</p> <p>The sample frame for intervention households came from the project lists for the three districts. The sample frame for comparison group households came from union-level administrative lists sourced from local government, for the unions selected as appropriate counterfactual regions. Household level eligibility was determined by criteria replicating original project selection criteria (including owning a cow in 2014).</p>
2	Representativeness of project participants	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>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.</li> <li>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.</li> </ul> <p>Score MEDIUM risk if:</p> <ul style="list-style-type: none"> <li>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 entire list of project participants OR it is clear the results apply only to a particular group of project participants.</li> </ul> <p>Score HIGH otherwise.</p>	Low	<p>Broadly speaking there has been good continuity with project participation. Some participants may have joined after the project start date but they appeared small in number when reviewing project listings. Sampling was conducted on the entire list of active participants in 2017.</p>

	Title	Description	Assessment	Description
Selecting comparison group				
3	Potential for contamination (spillovers)	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>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.</li> <li>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)</li> </ul> <p>Score HIGH risk if:</p> <ul style="list-style-type: none"> <li>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)</li> </ul>	Low	Units for comparison group were selected in geographical areas where it is not reasonable to expect for the project to have had spillover effects. Radio-based interventions were a minor component of the activities implemented.
4	Self-selection of project participants	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>The comparison group is exploiting an experiment or natural experiment.</li> <li>Units are randomly selected at community level both in the intervention and comparison group.</li> <li>The selection process for the comparison group is mimicking the same selection process used by the project.</li> </ul> <p>Score MEDIUM risk if</p> <ul style="list-style-type: none"> <li>If the self-selection is corrected during the matching procedure (e.g. controlling for group participation at baseline)</li> </ul> <p>Score HIGH risk if:</p> <ul style="list-style-type: none"> <li>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.</li> </ul>	Medium	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.

	Title	Description	Assessment	Description
5	Other interventions in the comparison group	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>There are no other actors in the area (e.g. INGOs, NGOs, governmental programmes)</li> <li>Other actors are conducting activities which are not linked to the project's theory of change</li> </ul> <p>Score MEDIUM risk if:</p> <ul style="list-style-type: none"> <li>Other actors are conducting similar activities in both the intervention and the comparison group</li> </ul> <p>Score HIGH risk if:</p> <ul style="list-style-type: none"> <li>Other actors are conducting similar activities, partially related in the comparison communities</li> <li>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.</li> </ul>	Low	Project exposure indicators suggest comparison group did not receive interventions from GEM partners or other NGOs.
Analysis				
6	Representativeness	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>During analysis or matching procedure less than 10% of the sample in the intervention group is excluded.</li> </ul> <p>Score HIGH risk if:</p> <ul style="list-style-type: none"> <li>During analysis or matching procedure more than 10% of the sample in the intervention group is excluded.</li> </ul>	Low	Less than 10% of intervention group sample is excluded.
7	Robustness checks	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>Magnitude and statistical significance of the results are approximately consistent with different econometric models</li> </ul> <p>Score HIGH risk if:</p> <ul style="list-style-type: none"> <li>Results are not consistent with different econometric models and sub group analysis.</li> </ul>	Med	Triangulation with qualitative data and other variables suggest results are consistent. However, no alternative econometric models have been reported (hence not assessed as low risk).
8	Triangulation	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>Results are triangulated and consistent with other evaluation methods within the same evaluation.</li> <li>Results are triangulated and consistent with other data on the same project but from different evaluations.</li> </ul> <p>Score HIGH risk if:</p> <ul style="list-style-type: none"> <li>Results are not consistent or triangulated with other evaluation methods.</li> </ul>	Low	Triangulation with qualitative data from focus groups and key informant interviews.
9	Multiple hypothesis testing	<p>Score LOW risk if:</p> <ul style="list-style-type: none"> <li>Multiple hypothesis tests apply Benjamini-Hochberg or Bonferroni tests.</li> </ul>	Med	No controls applied for multiple hypothesis testing. But pre-analysis plan was prepared and followed.

	Title	Description	Assessment	Description
		<ul style="list-style-type: none"> <li>The evaluation drafted a pre-analysis plan prior data analysis, and followed the plan.</li> </ul> Score HIGH otherwise		
10	Clustering	Score LOW risk if: <ul style="list-style-type: none"> <li>Clustering is applied</li> <li>Clustering was tested but rejected as providing higher standard errors than non-clustering estimates.</li> </ul> Score HIGH otherwise.	Low	Clustering applied at village level.
Other				
11	Other	Any other issue reported by the evaluator.	Evaluation took place during an exceptional year with severe flooding. As far as possible the evaluation design tried to minimise bias but validity of results must be seen in light of the circumstances around which data gathering took place. The impact of floods is judged to have had a more severe effect on intervention households, so would make it more difficult to detect project impacts. As such, the results can arguably be seen as conservative estimates of the impact of GEM activities.	