

Annex 6: Evaluation methodology and approach

1. Analytical Framework

The design of the evaluation is based on the consideration of three interlinked elements: clearly, the design must reflect the objectives as set out in the ToR, but it is equally important to ensure that the evaluation captures the programme approach and context and the implications hereof with regards to what critical issues and linkages to explore. Finally, the design must be tailored to make the most of the resources at hand, by utilizing existing data and collecting additional information in a manner that considers both the analytical needs and the strengths and weaknesses of the data environment.

Below, the evaluation methodology and approach is explained by first outlining the task given by the ToR and the overall implications for the evaluation approach, followed by an overview of how this has shaped the evaluation focus on an unpacking of the programme ToC, to be able to respond to the evaluation questions in a nuanced and credible manner. Furthermore, the assessment of the data environment is outlined, as this influenced the design and prioritization of the data collection needed for the evaluation. Finally, the actual data collection is presented, together with an explanation of the challenges and limitations.

2. The evaluation objective and focus:

According to the ToR, the evaluation should address three interlinked objectives:

- To document the achievements of the AGEP since 2013;
- To analyse the outcomes and impact of the IFMC and AFSP (components) in terms of results, based on the original theories of change, logical frameworks and results frameworks and with a particular emphasis on the adoption of the farming techniques and practices as well as marketing knowledge promoted through the components;
- On the basis of the lessons learned through the AGEP, to prepare recommendations for the future as inputs for the design of a new country programme for Danida in Bangladesh (for five years from 2021-26).

By implication, a key issue has been to ensure that the evaluation was able to investigate male and female farmers' use and application of the knowledge gained through participation in FFS training. While the FFS approach has repeatedly been shown to have effects (for instance in the 2011 evaluation), it is often less clear how particular elements in the approach are contributing to the successes. The evaluation design therefor needed to establish a more explicit assessment of the extent to which production, incomes and nutrition had improved due to the adoption of new techniques acquired through the FFS and through the establishment of FOs for crop marketing. The particular contribution of the FFS approach to women's empowerment also needed to be explicitly addressed. Thus, identifying and measuring of outcomes and impact, both at the overall level for AGEP as well as for more specific areas of FFS, was at the heart of the evaluation design, as well as unpacking the processes leading (or not) to the achievement of results, including enabling and hindering factors.

To this end, the evaluation team used the AGEP logical framework and results frameworks to establish a theory-based framework as a basis for the assessments. This overall theory-based framework constituted the core foundation of the analytical framework. The overall approach to data collection and analysis was based on a mixed-methods approach, combining rigorous quantitative data collection with qualitative methods (see further detailing of the specific methods below). Different levels of results needed to be addressed:

- At AGEF programme level, the overall assessment of achievements, with emphasis on documenting the changes to which AGEF has contributed had to be established. This is based on the assessment of component performance and results together with document review, trend analysis, etc., to be able to create a broader picture of programme contribution to change.
- For IFMC, the mentioned theory-based, mixed method approach aimed at capturing results at outcome and impact level in a credible manner as well as unpacking the enabling and hindering factors behind the performance of the programme has been key. This involves both a survey following up on the earlier baseline study, as well as qualitative field work and programme documentation review.
- For the AFSP component, the assessment builds on a combination of the recent UNDP-commissioned impact assessment (including a household survey) and a supplementary field work, to supplement and deepen the existing assessments. of the AFSP component (see below for more regarding the use of the UNDP evaluation).

The Katalyst component has not been explicitly addressed by this evaluation, in line with the ToR and dialogue with EVAL and ERG. Thus, the presentation of the analytical framework will mainly focus on the assessment of the IFMC, though the evaluation analysis will include also AFSP.

3. Unpacking the Theory of Change and identifying critical issues

3.1 The ToC for FFS in AGEF, with emphasis on IFMC

The analytical framework builds on the Theory of Change (ToC) underpinning the supported interventions. It is key both for quantitative measurement of outcomes and impacts as well as for assessing whether and how the support has worked as intended, including the role of contextual and external factors. More specifically, the analytical framework entails a mixed methods approach including both rigorous quantitative impact assessment and contribution analysis with more qualitative aspects, all informed by the ToC.¹ (for a graphical presentation of the reconstructed IFMC ToC see Annex 7).²

The ToC was reconstructed by the evaluation team,³ based on the available programme documentation.

The different elements of the IFMC support is to a large degree well established in the programme documents, including the FFS approach. While successful implementation of an approach (as expressed in the implementation strategies) is seen as fundamental to achievement of the expected results (outputs, outcomes and impact), some of these strategies are at the same time based on critical assumptions. This, for instance, relates to the content of the curriculum, the planning of the sessions and the quality of the training (which is well known to be a critical aspect of FFS implementation). Delivery of quality training assumes that programme managers are able to supervise and control the training activities. By using the ToC to understand both the concept of the FFS approach and unpacking the critical assumptions, the evaluation is better placed to distinguish between enabling and hindering factors linked to the concept and those linked to the practical implementation.

Likewise, in relation to the FOs and the market aspects, the evaluation team finds that at least two critical assumptions seem not to have been considered in the original programme framework,

¹ As contribution analysis here serves as a way of framing the work with the ToC, and as the evaluation not just addresses contribution but also attribution of results, the approach is not explained in detail. The operational implications of the approach are covered in the sections on the ToC as an analytical framework, the quantitative and qualitative data collection and challenges and limitations respectively.

² It should be noted that the high value crop module was piloted and implemented later than originally expected, and not in all FFS. It is included in the model to illustrate this part of the reasoning in the programme, but it must be stressed that this particular causal strand cannot be expected to have worked as originally envisioned.

³ A ToC was not included in the AGEF Programme Document.

namely 1) the assumption that FOs will develop to serve members' interests in markets and governance" and 2) the assumption that FO members will be able to become active in markets on terms beneficial to them.

In terms of gender equality, it has been assumed in the IFMC that women will participate on the same terms as men. However, as highlighted in the Danida FFS evaluation from 2011, women's mobilization is often restricted and their actual qualitative participation in decision-making is limited. Therefore, it is not sufficient only to include income and control over use of income as indicators for women's empowerment. IFPRI's Women Empowerment in Agriculture Index (WEAI) provides some innovative thinking into this area, and the evaluation team has used the WEAI as inspiration for defining of additional women empowerment indicators.

For FFS, the scaling up – with recruitment of new staff, different management, etc., and the combining of crops, livestock and fish in one FFS – may have had implications for the quality. Thus, understanding the practical implementation, participants' experience with training, etc. has been an important element in assessing the crucial element of new learning and experience for farmers, as well as the adoption of new techniques. This also relates to the understanding of - and effects from - the particular focus on market linkages and gender mainstreaming through the FFS and FO support provided (see the section on limitations below, for more on challenges in this regard).

IFMC has included some tacit assumptions about FOs, farmers' market participation and market linkages, found in the approaches as they evolved over time and as featuring quite prominently in the phase to be evaluated. They concern the effectiveness and suitability of supporting FOs and market linkages through government agents and the effects hereof on nature and functioning of FOs and market groups. They also concern ways in which farmers use markets and how well they are able to exploit market mechanisms. As these assumptions are important in the possible future direction of the support, they will be included in the analysis, with focus on governance and distribution of benefits and power in the organisations.

Further, it was deemed important to be able to include some additional factors, which may (or may not) be of such importance that they could affect the changes experienced by programme participants, and thus should be considered by the evaluation (from remittances as a factor influencing the well-being of families to the gender and ethnicity of trainers). In many cases GoB officers have acted as trainers, field workers or master facilitators. Lead facilitators were always GoB-officers (Regional IFMC Coordinators). In the villages, farmers were the 'trainers'. Information on these issues were gathered through both the quantitative and qualitative elements of the evaluation (more on these below).

3.3 The Evaluation Questions

The 10 Evaluation Questions (EQs) from the ToR provide the overall framework for the evaluation assignment. The EQs are presented in the Evaluation Matrix with judgement criteria and indicators.

Evaluation Question	Judgement criteria	Indicators	Means of verification
EQ 1: What have been the main achievements of the AGEP? Effectiveness <i>Outcomes/ impact</i>	Fulfilment of programme targets and objectives Change processes with implication for future interventions	Programme outcome and impact (likely) indicators Institutional and/or structural changes Changes in behaviour/attitudes	Programme documentation and monitoring data National statistics and data Semi-structured interviews HH survey External reviews, studies and evaluations (incl. UNDP evaluation)
EQ2: What are the results in terms of participation in farmer field schools and the development of market linkages (both in IFMC and AFSP)? Effectiveness <i>Outcomes/ impact</i>	Fulfilment of component targets and objectives Implementation strategies (choice of approaches and modalities) have been supportive to the targets and objectives Due attention has been paid to gender equality (see also EQ 6)	Component outcomes and outputs have been achieved and documented The selection of approaches and modalities has been based on periodic and critical reflections Gender issues have been mainstreamed across the various engagements	Programme documentation Semi-structured interviews HH survey FGD/site visits External reviews, studies and evaluations (incl. UNDP evaluation report; possibly BIHS data)
EQ3: What are the costs of providing agricultural extension and training using FFS approaches? <i>Efficiency; financial aspects of sustainability</i>	Cost assessment (whether resources have been well spent, no overspending, bottlenecks addressed etc.), Comparison with FFS in other countries in the region (as far as data meaningfully allows)	Costs per FFS/farming HH Economic value of benefits per HH	Financial programme reports Literature study, UNDP evaluation report Semi-structured interviews
EQ4: Are the skills and techniques acquired by the farmers considered relevant and do the farmers adopt and use what is learnt in the farmer field schools and in the market linkages training? <i>Impact</i>	Extent to which farmers have the incentives, conditions and abilities to adapt – see ToC	New techniques from FFS have been applied also after completion of FFS Farmers are using either input or output market linkages, also after completion of FFS	Programme documentation Semi-structured interviews HH survey FGD/site visits External reviews, studies and evaluations incl. UNDP evaluation report
EQ5: What are the critical factors determining the quality of farmer field schools and training in market linkages and are the FFS and marketing development processes adequately organised and managed by the extension services (DAE, together with the livestock and fisheries services)? <i>Effectiveness/ coherence/ lessons learned</i>	Farmers attendance to training sessions Factors that make farmers feel more empowered and confident FFS Curriculum	Number of farmers (male/female) attending training sessions Share of facilitators with experience in marketing Level of practice vs. theory Quality and background of trainers Share of female facilitators Content of training; adaptation to circumstances	FFS attendance lists Semi-structured interviews HH survey FGD/site visits External reviews, studies and evaluations
EQ6: How has women's empowerment been enhanced through the AGEP? <i>Effectiveness and impact</i>	Gender issues are mainstreamed in a manner that is relevant and effective for empowerment. Women's role in decision-making processes around production and	Women's experience with participation in FFS Women's level of input in decision-making processes related to	Programme documentation IFPRI data Semi-structured

	<p>income generation</p> <p>Women's access to and ownership of productive capital</p> <p>Women's access to credit</p> <p>Women's time allocation</p> <p>Women's group membership and leadership role</p> <p>Women's mobility outside their homes</p>	<p>production/income (none, few, some, most/all)</p> <p>Women's level of ownership of productive capital (yes (solely), jointly, no)</p> <p>Women's role in decision-making on borrowing and what to do with the money</p> <p>Women's activity log (last 24 hours)</p> <p>Women's membership of FOs, including their role/position and meeting attendance)</p> <p>Women's permission to visit markets (yes – alone, yes – but only with husband, no permission)</p> <p>(Note: Indicators developed with inspiration from IFPRI's WEIA index)</p>	<p>interviews</p> <p>HH survey</p> <p>FGD/site visits</p> <p>External reviews, studies and evaluations</p>
<p>EQ7: What have been the impacts of the FFS approach as promoted by the DAE and by UNDP in the CHT in terms of agricultural diversification, household income, nutrition and employment?</p> <p><i>Impact</i></p>	<p>Changes in agricultural diversification, household income, nutrition and employment in the area</p>	<p>Indicators and proxy indicators for changes (UNDP evaluation report)</p>	<p>UNDP evaluation report</p> <p>Brief site/verification visit to CHT (TBD)</p> <p>Farmer FGDs in CHT (TBD)</p> <p>Interviews with key stakeholders in Dhaka</p> <p>BIHS data (TBD)</p>
<p>EQ8: What obstacles have arisen in the implementation of the AGEP and how have these been overcome?</p> <p><i>Effectiveness/efficiency/lessons learned</i></p>	<p>Extent to which the support has been flexible and adaptive</p> <p>Extent to which negative impact from political decisions has been mitigated</p> <p>Extent to which implementation approaches have been designed to overcome difficulties in the cooperation</p>	<p>AGEP has been adjusted to reflect changes in contextual factors or learning from programme implementation</p> <p>A continuous policy dialogue has taken place</p> <p>The selection of partners and modalities has been based on periodic and critical reflections</p>	<p>Programme documentation</p> <p>Semi-structured interviews</p> <p>FGD/site visits</p> <p>External reviews, studies and evaluations</p>
<p>EQ9: What are the prospects for increased employment arising from improved production techniques and better agricultural marketing, notably in terms of opportunities for young people?</p> <p><i>Impact (likely); prospects for sustainability</i></p>	<p>Trend in uptake and labour intensity of improved production techniques</p> <p>Age composition of new employees</p> <p>Changes in marketing approaches</p>	<p>Expectations to further uptake and employment</p> <p>Share of new employees below 25 years old</p> <p>Increase in volume of new market linkages established</p>	<p>National data and statistics</p> <p>FGD/site visits</p> <p>External reviews, studies and evaluations</p>
<p>EQ10 (8): What are the key recommendations for the future country programme?</p> <p><i>Lessons learned/recommendation</i></p>	<p>Evaluation conclusions</p>	<p>Lessons learned</p>	<p>The evaluation analysis</p>

These questions, together with the evaluation objectives and key issues, are the basis of the design and methodological approach for the evaluation. Both qualitative and quantitative indicators have been used, as a range of issues are multifaceted and require indicators that capture this (i.e. issues related to the implementation processes content/quality, adaptation to context, etc.). Below, an overview is provided that links the EQs with the different areas of investigation for the evaluation, indicating how the ToC has informed the interpretation of the EQs and the design of the data collection:

Evaluation Question and link to OECD-DAC Criteria	Key issues investigated	Data collection/sources
<i>Approach, content and implementation at component level</i>		
EQ5: What are the critical factors determining the quality of farmer field schools and training in market linkages and are the FFS and marketing development processes adequately organised and managed by the extension services (DAE, together with the livestock and fisheries services)? <i>Effectiveness/ coherence</i>	FFS Curriculum; content, time frames for modules etc. Quality and background of trainers; Share of female facilitators; Share of facilitators with experience in marketing Level of practice vs. theory (how much “exploratory” and hands-on training for what elements?) Number of farmers (male/female) attending training sessions Factors that make farmers feel more empowered and confident	Semi-structured interviews HH survey FGD/site visits External reviews, studies and evaluations
EQ4: Are the skills and techniques acquired by the farmers considered relevant and do the farmers adopt and use what is learnt in the farmer field schools and in the market linkages training? <i>Relevance, effectiveness, efficiency, impact</i>	Extent to which farmers have the incentives, conditions and abilities to adapt New techniques from FFS have been applied also after completion of FFS Farmers are using either input or output market linkages, also after completion of FFS	Programme documentation Semi-structured interviews HH survey FGD/site visits External reviews, studies and evaluations incl. UNDP evaluation report
<i>Results achieved; contribution to change at component level</i>		
EQ2: What are the results in terms of participation in farmer field schools and the development of market linkages (both in IFMC and AFSP)? <i>Effectiveness, Outcomes/ impact</i>	Fulfilment of component targets and objectives (outputs and outcomes) Implementation strategies (choice of approaches and modalities) have been supportive to the targets and objectives: The selection of approaches and modalities has been based on periodic and critical reflections The selection of participants has been in accordance with objectives and criteria Gender issues have been mainstreamed across the various engagements	Programme documentation HH survey – double difference analysis for assessment of outcomes and impacts. FGD/site visits External reviews, studies and evaluations (incl. UNDP evaluation report) Semi-structured interviews
EQ6: How has women’s empowerment been enhanced through the AGEP <i>Effectiveness and impact</i>	Gender issues are mainstreamed in a manner that is relevant and effective for empowerment. Women’s experience with participation in FFS Women’s role in decision-making processes around production and income generation Women’s access to and ownership of productive capital; access to credit; women’s time allocation Women’s group membership and leadership role, FOs Women’s mobility outside their homes; visits to markets etc.	Programme documentation Semi-structured interviews HH survey (double difference analysis for outcomes and impacts) FGD/site visits External reviews, studies and evaluations
EQ7: What have been the impacts of the FFS approach as promoted by the DAE and by UNDP in the CHT in terms of agricultural diversification, household income,	Changes in agricultural diversification, household income, nutrition and employment in the area	UNDP evaluation report Brief site/verification visit to CHT (TBD) Farmer FGDs in CHT Interviews with key

nutrition and employment? <i>Impact</i>		stakeholders in Dhaka
EQ9: What are the prospects for increased employment arising from improved production techniques and better agricultural marketing, notably in terms of opportunities for young people? <i>Impact (likely); prospects for sustainability</i>	Trend in uptake and labour intensity of improved production techniques Changes in marketing approaches	FGD/site visits External reviews, studies and evaluations HH survey on increased income (link to non-farm income).
<i>Costs, Efficiency</i>		
EQ3: What are the costs of providing agricultural extension and training using FFS approaches? <i>Efficiency; financial aspects of sustainability</i>	Cost assessment (whether resources have been well spent, no overspending, bottlenecks addressed etc.), comparison against results achieved Comparison with FFS in other countries in the region (as far as data meaningfully allows) Costs per FFS/farming HH Economic value of benefits per HH	Financial programme reports Literature study, UNDP evaluation report Semi-structured interviews
<i>Programme-level issues: Implementation issues and overall achievements</i>		
EQ8: What obstacles have arisen in the implementation of the AGEF and how have these been overcome? <i>Effectiveness/ efficiency</i>	Identification and follow up to obstacles of problems identified. Extent to which the support has been flexible and adaptive; degree and role of policy dialogue; Extent to which negative impact from political decisions has been mitigated Extent to which implementation approaches have been designed to overcome difficulties in the cooperation The selection of partners and modalities has been based on periodic and critical reflections	Programme documentation Semi-structured interviews FGD/site visits External reviews, studies and evaluations
EQ 1: What have been the main achievements of the AGEF? Effectiveness <i>Outcomes/ impact</i>	Fulfilment of programme targets and objectives, as related to the evaluated components Change processes with implication for future interventions	Component results, based on the above Programme documentation and monitoring data External reviews, studies and evaluations (incl. UNDP evaluation)
<i>Lessons learned and recommendations</i>		
EQ10 (8): What are the key recommendations for the future country programme? <i>Lessons learned/ recommendation</i>	Evaluation conclusions; drawing on all of the aboard	The evaluation analysis drawing on all of the aboard

In the following, the key methods and approaches used by the evaluation to provide answers to the EQs are presented in more detail. In order to understand the needs and challenges regarding data collection and analysis an assessment of the available data based is presented, as this has implications for both the approach to data collection, and to the limitations and challenges encountered at the time of analysis.

4. The data environment and the implications hereof

During the inception phase, a first assessment of the data environment was carried out as a basis for fine-tuning the evaluation design. As expected, a mixed picture of positive and less positive elements emerged, with a large amount of relevant data available, but with important challenges in relation to

quality and coverage in certain areas. Below, considerations regarding key data sources and the implications for the evaluation design and data collection are outlined.

4.1 Baseline data

A baseline study for IFMC was carried out in 2013 and had, by a stroke of luck, been preserved by M&E programme staff. The evaluation team shares the view of the MTR that the baseline study was carefully designed and implemented, in the sense that it has collected a large amount of detailed and relevant data, focusing on both household characteristics as well as outcome and impact indicators of relevance for change processes.

The data quality is assessed to be quite high in some areas. For instance, simple questions (such as e.g. “do you own a tube well for the household? (Yes/No)”) have high response rates (almost 100% throughout). On the other hand, more complex questions (so-called combination questions, such as “did you have any off-farm income last year” combined with a question of the “amount of the off-farm income”) had much lower response rates (down to 30%), which is however a quite common finding in such surveys. It is of higher concern that only 80% of the respondents state “farmer” as primary occupation. This could indicate that up to 20% of the survey respondents may not have been part of the main target group or that farming in these cases is a secondary occupation.

Further, it was clear that extracting income per capita information from the data would be challenging, since the questionnaire includes more than 50 types of income from farm. In addition, in some cases household responses are reported in kg and in other cases in mound planted per farm product. Some households seem to have indicated the overall value of the harvested crop and others the value (per kg or mound) of the crop. The variance of the value of some crops is quite large. While standard deviations are not too big, there are cases where the same crop is rated 5 times more valuable by some households than by others. This may reflect that prices do differ a lot from start of season to end of season, even from morning to evening on the same day, and from one geographical location to another. Some farmers might have started planting early and fetch high prices at the beginning of the season, while others planted late and let their produce rot in the field because the price has collapsed. Finally, in some cases, households indicated that they were planting a specific crop although it had a value of zero to them, making it uncertain if the crops were destroyed or if that reply was missing. It should be clear this represents challenges for any questionnaire trying to capture an accurate, detailed picture of income.

Due to issues of limited coverage of programme participants, it was also clear that recall questions would have to be included. As the survey still needed to be kept manageable in length, these issues led to a revised questionnaire, pursuing outcome indicators that were deemed more relevant for recall, with less emphasis on very detailed income estimates, on more emphasis on adaptation of techniques, compared to the original baseline. This was seen as important, as the questions on application of techniques in the original baseline were seen as too limited to capture different types of change in practice. While it was not feasible to include all possible techniques and potential changes in practice as promoted by FFS, it was seen as important to include a broader range of possible “new adaptations”, not least in relation to homestead gardening. The baseline survey was still used as point of departure, to get as solid a basis for comparison as possible, but with reduction of complexity, especially with regards to income, and addition of questions for investigation of additional aspects of change, as well as recall questions, in order to address the challenges stemming from limited coverage of FFS participants in the baseline survey.

The issue of the limited coverage has posed the biggest challenge related to the baseline survey: It was designed and conducted independently from the implementation of the current phase of FFS and was

aimed at covering households within the target group of landless, marginal and small farm households. This is all well. The problem arises in that there is limited overlap between the villages selected for the baseline study and villages later covered by IFMC (only 14 villages, located within 12 different Upazilas). 15 households were surveyed within each of the 15 baseline villages. Given the size of villages in Bangladesh, it was assumed that the overlap was limited, which was also found to be the case. In practice, this implied that instead of covering the situation prior to FFS interventions for both treatment and control households, the baseline survey mainly covered the control group, and additional informants participating in the FFS had to be included to reach a viable sample. This again led to the need to include recall questions, in order to gather information regarding the “baseline” situation for FFS households, as well as to include programme information regarding the situation prior to FFS.

4.2 Monitoring data

A wide range of monitoring data has been collected by the IFMC, covering both selection of participants, implementation, quality assessments and selected outcome and impact indicators. The preliminary assessment of the monitoring system found it to be well elaborated with formats, checklists, collection schedules etc. As such, there was an abundance of programme data, including the information collected regarding households at the point of initiating FFS (through the programme Household Survey Format) of use to the evaluation.

However, it has been clear from the outset (e.g. from the MTR 2017) that the reliability and quality – and thereby the usefulness - of the monitoring data was an issue to be critically assessed by the evaluation team. This risk of positive bias is well known, and it is understandable that programme staff can be eager to present efforts and results in a positive light. Nevertheless, it is of course problematic if there is suspicion that this may have created bias in the data collection and/or reporting, and if senior/central level staff have problems investigating the issue in a manner that allows for ascertaining whether it is indeed a problem or not, and the degree of the problem at hand.

With regards to the internal monitoring, attempts to follow up and validate have been made by programme staff, but with experiences similar to those presented in the review report, implying that assessing accuracy of the data was difficult. Further, examples were shared with the team of instances where monitoring data indicated progress and “business as usual”, in a manner that did not tally with actual field conditions (e.g. areas affected by severe flooding). In addition, the definition of certain indicators makes it difficult to use the information to draw conclusions on the outcomes of the programme. As an example, the team was informed that the monitoring system reflecting the key result area of “adaptation of new technologies” for FFS participants is based on a criterion of whether specific technologies have been applied “at least once”. This would mean that in case a FFS participant attempted to implement a new technology, but encountered problems (could not manage or did not find it worthwhile to repeat) this would still count positively in the assessment of difference in use of technology.

The implication of this has been to use the programme data wherever relevant, but mainly in relation to verifiable and descriptive issues such as programme roll-out. For issues regarding the quality and performance of the programme, monitoring data has not been used independently, but only with careful consideration of the risk of bias, emphasising triangulation and assessment of consistency against other sources of information.

4.3 Other data sources

In addition to the baseline and internal monitoring data, programme staff facilitated access to a range of other data sources for the evaluation. This included discussions with UNDP regarding their evaluation/impact assessment of AFSP II. This evaluation to a large degree built on a double-

difference approach, aiming to establish a counterfactual analysis. This is clearly a suitable approach, but based on the information presented in the report, the double-difference is found to have been based on a matching of treatment and control households with limited information to support it and some variation between the two groups. Further, the risk of positive bias is also relevant to the UNDP evaluation. Thus, while it has been a useful document to support the evaluation analysis, it has also been important to collect additional information from the field in order to supplement and triangulate the findings from the AFSP II impact assessment and ensure that the results are used in a nuanced and appropriate manner (see below for a more detailed assessment of the AFSP II impact assessment).

IFPRI has been carrying out a series of highly interesting surveys, including two randomized surveys for the ANGeL project, from which preliminary findings are expected available by January-February 2019. Further, IFPRI has implemented the BIHS in 2011 and 2015. Other surveys of relevance to the evaluation are underway but were finalized too late to be of use to the evaluation. The IFPRI surveys has provided additional information for trend analysis and triangulation and as well as some supplementary information on issues such as youth in agriculture.

5. Data collection

Based on the assessment of the types of issues to explore and the questions to investigate, together with the early assessment of the data environment, the data collection strategy focused on gathering the quantitative information needed to enable a quantitative assessment of the outcome and impact of IFMC-FSS, and on establishing a more nuanced and in-depth picture of the working and role of the FFS through qualitative field work.

The overall approach to data collection and analysis was based on a mixed-methods approach, combining quantitative data collection with qualitative methods (see further detailing of the specific methods below). The data collection strategy focused on gathering the information needed to enable a quantitative results assessment of IFMC-FSS, and on establishing a more nuanced picture of the working and role of the FFS through qualitative field work.

Household survey

A household survey was implemented within four different regions; two in the North (Rangpur and Rajshahi) and two in the South (Barisal and Chittagong, Feni District). The household survey effectively covered 965 households (388 FFS households and 577 control households) within the four regions:

- Rangpur: 249 households (85 FFS households);
- Rajshahi: 289 households (109 FFS households);
- Barisal: 303 households (121 FFS households);
- Chittagong (Feni): 124 households (73 FFS households)

The household survey was implemented as a follow-up to a baseline survey (implemented in 2014) with some adjustments to enhance its usefulness to the evaluation. As the baseline survey was conducted prior to identification and selection of FFS villages and households, it did not include FFS households. It therefore mainly served as a baseline for the control group, including non-FFS households within FFS villages (a within-village control group). For the FFS households (the treatment group), AGEP

programme data collected from FFS households prior to the start of the FFS (the FFS household survey) were used as a proxy baseline, although with much less detail than the baseline survey.⁴

The development of the household survey questionnaire took point of departure in the baseline survey questionnaire. By using the same questionnaire as far as possible, the comparability of findings was enhanced. Some response categories were reduced, to simplify the questionnaire as much as possible and other questions were added, in order to investigate a few issues in more depth. Further, a series of recall questions were included to identify changes.⁵

14 villages (including a total of 19 FFS groups) was identified across the baseline survey villages and the list of FFS villages. These 14 villages were located within the four above-mentioned regions. In addition, 34 non-FFS villages from the baseline survey (all located within the same four regions) were included in the control group. Care was taken to include as many informants from the baseline survey and as many FFS-households as possible, in order to ensure a large enough sample of control households and villages to match against. Through careful follow up and use of all available information, this was largely achieved. The use of two different types of control groups (control household both from within and outside FFS villages) allowed for a more nuanced assessment of variance in results patterns, including possible spill-over effects.⁶

The Evaluation has used a propensity score matching approach⁷ to carry out an econometric analysis of the collected household data, based, to a large extent, on a matched double difference approach.⁸ The information on general household characteristics (size of land, education (years), household size, number of males/females) in the data set has been used fully in the matching approach pursued. The robustness of the results from the econometric data analyses has been tested at the 1% (most significant), 5% and 10% (least significant) statistical significance level.⁹

The survey was implemented in collaboration with the Bangladesh Centre for Advanced Studies (BCAS). While the time frame is tight, BCAS has a corps of experienced enumerators, implying that they can field a larger team to allow for a speedier process. The baseline questionnaire was field tested prior to the baseline study, but care will be taken to check the revised questionnaire. Both the training of the enumerators and the initial implementation of the survey was supervised by the evaluation team.

⁴ It should be stressed that the qualitative field work raised issues regarding the accuracy and reliability of the FFS household survey, finding that the information may not always have been provided directly from the farmers and the register having been filled in later by programme staff (see below for more regarding challenges and limitations).

⁵ While use of recall is less than ideal (due to the obvious problem of accuracy), care has been taken to ensure it is as sound as possible, by drawing on research into what types of issues are most relevant for more detailed recall, and where more overall questions regarding trends have been more appropriate.

⁶ In the findings section, the group labelled “non-FFS” consists for non-FFS household in villages where the FFS has been implemented. The group labelled “control” consists of households from villages that has not been part of IFMC.

⁷ Mathematical technique used to select members of the control group that share characteristics with members of the participants’ group, through estimation of a statistical model based on matching characteristics (household characteristics).

⁸ The double difference measures the difference in the observed change between participating households/individuals and control village households/individuals, based on baseline (recall) data and ex-post data. Thus, the double difference eliminates external determinants of the outcome, in cases where these are the same for the two groups during the intervention period. The double difference approach assumes common time effects across groups and no composition changes within each group.

⁹ In statistics, a result is called statistically significant if it is unlikely to have occurred by chance. In this analysis, the significance level is used to measure the statistical strength of a data finding. The significance level is, in this case, the risk of concluding a data relationship that may not exist. Frequent levels of significance used for statistical testing are 10%, 5% and 1%. If a significance test gives a value lower than the test levels, the null hypothesis (a hypothesis that an observed difference between two data sets is random/due to chance) is rejected. Such results are referred to as being ‘statistically significant’. For example, in this report, if an observed difference between data from participating households and control village households is found to be significant at the 10% level, it means that the null hypothesis (that the observed difference is by chance/random) can be rejected with 90% certainty. The lower the significance level is, the stronger the certainty that the null hypothesis can be rejected. Cases with relatively few observations (data) and large variation, increase the uncertainty and make it more difficult to reject the null hypothesis.

Although BCAS is experienced in data processing, an additional level of data quality control was conducted in Denmark under the supervision of the subject matter specialist, Professor John Rand.

Qualitative data collection

The qualitative fieldwork was designed to be implemented *after* the preliminary results from the household survey were known. This sequencing allowed an element of follow-up on particular interesting findings and results from the survey, including more in-depth assessment of specific issues. Key Informant Interviews (KIIs) and Focus Group Discussions (FGDs) together with site inspections were applied as the key qualitative methods by the evaluation team.

Key Informant Interviews (KIIs) were conducted with key stakeholders to obtain qualitative findings on fundamental evaluation issues. These interviews were implemented as extended one-on-one exchange with key individuals in a unique and strategic positions in relation to the supported interventions. Semi-structured interview guides were applied in order to make sure that information will be gathered in a consistent manner, covering all relevant evaluation areas and;

Focus Group Discussions (FGDs), which were a key instrument in the qualitative approach. The FGDs helped to capture views and opinions from a larger group of key stakeholders on this assignment. The FGDs were also useful to complement and validate/verify findings from survey data analysis and explain how the links in the ToC have evolved. The FGDs were also designed to be able to uncover unintended outcome/impact from the supported interventions, not initially captured by the quantitative data analysis. ‘Checklists’ were used for the FGDs with different stakeholder groups to ensure that similar type of data and information was collected from the FGD sessions, and “scoring cards” facilitated that the different critical issues were considered in a nuanced but consistent manner.

The fieldwork covered visit to three regions (Rangpur in the North, Barisal in the South and Chittagong Hill Tracts), where the following upazilas and villages were visited during a two-weeks period:

Qualitative fieldwork in Rangpur, Barisal and CHT regions

Rangpur region			
Upazila	Village	Female farmers	Male farmers
Pirgacha Upazila	Bara Hayat Khan village	11	0
	Gunjar Khan Amintari Village	11	8
	Uttar Chandipur Village	6	7
Palashbari Upazila	Basudebpur, Bhagwanpur village	12	5
	Balarampur Village	14	3
	Purbo Gopalpur village	9	6
	Paschim Goalpara village	20	18
Barisal region			
Betagi Upazila	Dakshin Hosnabad village	14	5
	Chandkali village	11	6
	Uttar Kawnia	15	4
Chittagong Hill Tracts region			
Rangamati upazila	Borodona village	23	4
Langadu upazila	Ishaqpara village	22	9
Naniarchar upazila	Jogendrapara village	20	12
Total		188	87

The selection of upazilas and villages for fieldwork visits were based on a wish to be able to study how the market linkage element was implemented, and to be able to cover implementation of FFS activities within different provinces (rich/poor) and within different agro-ecological zones; activities that were completed some time ago (potential impact and sustainability issues), as well as more recent activities (more focus on outcomes) as well as logistics and practicability of travel.

The following group of stakeholders was covered as part of the qualitative fieldwork:

- 3 Upazila Agricultural Officers (UAOs), all males
- 3 Upazila FFS coordinators (CHT)
- 3 Sub-Assistant Plant Protection Officers (SAPPOs), all males
- 8 Sub-Assistant Agriculture Officers (SAAOs), all males
- 2 District officers/coordinators (CHT)
- 26 FGDs with FFS farmers, 188 females and 87 males
- 2 FGDs with non-FFS farmers in Barisal, 13 females and 6 males
- 7 non-FFS members, 2 males, 5 females, in Rangpur FFS villages
- 23 farmers facilitators (FFs)
- 13 executive members and Business focal Points (BFPs), 7 male/5 females
- 7 Farmer Organisations (FOs)
- 6 UNDP technical programme officers (M&E, livelihoods, training)
- 6 UNDP Master trainers

In addition to the KIIs and FGDs, the evaluation made direct observations within the visited villages of FFS technology uptake and/or any changes at village/household level as a result from FFS activities.

Interaction and facilitation of dialogue:

During the FFS village visits, emphasis was on interviewing farmers in their fields as much as possible, to be able to use observed practices as a starting point for the dialogue. However, in many cases fields were located too far from the village for this to be feasible. A main priority in relation to the qualitative fieldwork was to ensure that the team had the opportunity to study the full “chain” of selection processes in the FFS approach (i.e. from the selection of unions and villages down to selection of FFS facilitators/trainers and, ultimately, the household beneficiaries and topics for the FFS sessions), including the rationale and consequences related to these choices. A clear practical understanding of these selection aspects has been important in order not to over or underestimate the potential impacts from FFS interventions as well as for analysis of various social and qualitative aspects. As mentioned, no programme implementation was under way at the time of the field work, which implied that the investigation was based on prior experience rather than ongoing processes.

When discussing with women, emphasis was on ensuring a secure space where no male family or community members are present. This allowed women to share opinions and experiences e.g. on actual decision-making power in the household and in the community more openly without interference or consequences after the discussion, as shown by the, sometimes critical, reflections and experiences brought to the table by the women.

In order to minimize the risk of positive bias, officials were (politely) requested not to participate. If officials were seen as interfering, they were politely be invited for separate discussion later on. However, it was not possible to fully avoid the presence of officials.

6. Limitations and Challenges

As for any evaluation, there are some important limitations and challenges to consider. While the approach to data collection and analysis was planned so as to address and remedy these challenges as far as possible, there were nevertheless a range of issues that must be kept in mind.

6.1 Risk of positive bias

As outlined above, the risk of positive bias was considered up front; both in relation to the monitoring data, and in terms of “diplomatic” bias during data collection, either due to general politeness and tendencies towards confirmations; or due to risk of showcasing and prepping of informants, whether by design or unconsciously. This posed challenges for both the quantitative data collection and the qualitative fieldwork.

For the *survey*, there was little room for verification as part of the data collection, but care was taken to ensure that all decentralised offices are informed about the survey mission in an overall manner, so that enumerators can visit villages without specific prior notification. Similarly, care was taken to emphasize the independence of the evaluation. It was not possible to carry out the field work without some involvement of DAE staff, but in general, the enumerators could carry out the interviews in an unsupervised manner. It was not possible to make physical checks of for instance crops, techniques etc as part of the survey.

For the *qualitative fieldwork*, the challenge was even larger, in that was impossible to avoid prior notifications and thereby have full discretion regarding participants. Here, care was taken to include at least a few “surprise visits” to ensure that participants for FGDs and KIIs and spot checks were also selected on the spot.

With regards to the use of *programme monitoring data*, the risk of positive bias has had the implementation that programme data has been used wherever relevant, but mainly in relation to verifiable and descriptive issues such as programme roll-out. For issues regarding the quality and performance of the programme, monitoring data has not been used independently, but only with careful consideration of the risk of bias, emphasizing triangulation and assessment of consistency against other sources of information

6.2 Programme stand-still

An additional challenge relates to the fact that the FFS interventions was not ongoing during the evaluation mission. A next phase is still being planned, and this implied that the evaluation could not observe implementation. For instance, it could have been highly relevant to observe training with the changes in curriculum and number of training sessions (as was the case in the 2011 FFS evaluation), assuming it had been possible to do this, without sessions being “showcased”. The situation also came with logistical challenges related to ensuring hard copies of programme information from the field prior to shut down – a challenge programme staff was most helpful in trying to remedy – and in the availability of regional programme staff for interviews.

Thus, it should be noted that the evaluation investigated a programme, which was not under implementation, which posed limitations on the both observations of practice and the dialogue regarding for instance selection processes, use of manuals and guidelines. Further, when considering the information gathered it the possibility that considerations of changes of modality for a possible next phase may influence the perception of the actors involved. The issue of programme stand-still thus creates limitations both with regards to the information that the evaluation could collect regarding programme implementation, and, as a second order issue, challenges in relation to the possibility for

triangulation and substantiation of findings: With less access to implementing staff than would have been the case with a running programme, and without the chance to discuss for instance selection and use of guidelines or observe actual training, there possibility of following up on critical issues and compare practices in a fully systematic manner has been limited. This does in no way imply that follow up or cross-site comparison has not taken place, but simply that there have been avenues of exploration that it has not been possible to follow, as a consequence of the programme stand-still.

Baseline data and “real world” limitations for the quantitative impact assessment

As explained above, the quantitative impact assessment built on a baseline study, which only covered the treatment group to a limited degree. This required the use of supplementary data sources, such as the FFS household survey information, and the incorporation of recall question in order to capture the baseline situation for FFS participants. This implies that information based on for instance recollection of conditions several years back may be less accurate than had the same information been obtained by a baseline survey with better coverage. The fact that quantitative analysis has been able to identify statistically significant differences, and the way the quantitative and qualitative findings are mutually supporting implies that the survey has been able to detect relevant differences and tendencies. However, for specific details care should be taken not to over interpret responses, for instance in relation to specifications of amounts. The key message to be taken from the survey are the broader lines of changes and results.

It should also be stressed that the results of the analysis are not representative in a sense that allows for general conclusions for the whole IFMC. With a limited number of observations, especially for participants, it will not be possible to claim full representativeness for all of FFS. Rather, the quantitative analysis will explore the effects of FFS with focus on the sample and aim at establishing a credible, externally validated assessment of the results and impacts. The added value thus lies in the use of a counterfactual to address the issue of attribution of results and establish an independent assessment. The assessment of whether the sample of FFS represents “special cases” or is indicative for the FFS in a more overall sense will not be part of the quantitative analysis per se, but rather be based on programme information, contextual analysis, comparison with other studies etc. The inclusion of four geographically diverse districts may also facilitate this assessment, albeit not in a statistical sense. Thus, rather than being able to state categorically that the findings within the sample villages can be taken to hold for the whole of IFMC, the analytical point is that the analysis of the selected villages highlights the effects achieved here, and illustrate the potential effects of FFS more broadly, based on the assumption that the sample villages are not “special cases”.

Further, with a relatively small sample, there may be nuances and effects which do not come across as significant in the survey, simply because they are hard to detect. It is stressed that it is not only the statistical results in themselves that are important, but also the interpretation of the results. In outlining the findings, care will be taken to explain the strength of the different findings, both as it comes across in the quantitative analysis, and with consideration of the qualitative field work and other data sources. Especially will the interplay with the qualitative investigations be used to clarify the issue of “evidence of absence” or “absence of evidence” of results.

Limitations to field work coverage

With regards to the qualitative data collection, a key challenge is the limited amount of time that could be spent in the field. As outlined elsewhere, care was taken in planning and selection of site visits, to gain as relevant a coverage of different contexts within the limited time available. Thus, coverage was limited, and in many cases, it was not possible to visit the fields with the farmers.

Dealing with the challenges: Triangulation, validation and planned interplay

A key element in remedying the challenges has been to build the analysis on several data sources, and to work with method and data triangulation.

For methodological triangulation, the interplay between the quantitative analysis on one hand and the qualitative fieldwork and mixed-methods contribution analysis on the other have been important to ensure that results are interpreted in a relevant and adequate manner. Both methods have their strengths and weaknesses, and especially in light of the challenges surrounding the survey, it has been important to not just take either quantitative nor qualitative findings at face value separately, but to compare and contrast in order to be able to explore consistency (or lack thereof) in the findings.

The qualitative data collection was sequenced so that the fieldwork was implemented after the survey. While this is both important to focus the contribution analysis on the issues where additional exploration is most relevant, it has also been helpful by allowing a degree of spot checks and validation of the quantitative findings. It was acknowledged from the outset that there would be limits to the number of spot checks, and an additional limitation was encountered in relation to going to the fields, in the cases where they were located too far away for farmers to go there during interviews. While the evaluation cannot claim to have solved the issue of positive bias or limited certainty, the sequencing and interplay between the data sources has been essential in qualifying the interpretation of the information at hand to reach valid conclusions.

For data triangulation, care has been taken to consider triangulate any important findings by considering other pieces of information, whether from other sites, actors or secondary sources. This has been used for careful consideration of the consistency and analytical solidity of the findings and has further been reflected as explicitly as possible in the report, so as to clarify the basis and strength of the evidence.