Netherlands Enterprise Agency

Endline report Impact evaluation of the Food for All Project in Kenya (F4APK)

Commissioned by the ministry of Economic Affairs and Climate Policy

>> Sustainable. Agricultural. Innovative. International.

Endline report

Impact evaluation of the Food for All Project in Kenya (F4APK)

Commissioned by the Netherlands Enterprise Agency (RVO.NL)

Contributors: Youdi Schipper, Daniel Rodriguez Segura, Thijs Yan

FINAL VERSION

08 December 2021

aighd | AMSTERDAM INSTITUTE FOR GLOBAL HEALTH & DEVELOPMENT

Visiting Address
Ahtc building, tower C4 Paasheuvelweg 25
1105 BP Amsterdam The Netherlands

Postal Address
P.O. Box 22700
1100 DE Amsterdam The Netherlands

t+31 (0)20 2103960
info@aighd.org

Table of contents

Table of figures

Table of tables

Abbreviations

AI	-	Artificial Insemination
AIGHD	-	Amsterdam Institute for Global Health and Development
ATT	-	Average Treatment Effect on the Treated
CAPI	-	Computer Assisted Personal Interviewing
CSR	-	Corporate Social Responsibility
CF	-	Contract Farming
DiD	-	Difference-in-Differences
F4APK	-	Food for All Project Kenya
FDOV	-	Facility for Sustainable Entrepreneurship and Food Security
GAP	-	Good Agricultural Practices (certification)
HCD	-	Kenyan Horticultural Crop Directorate
HPI	-	Heifer Project International
IPM	-	Integrated Pest Management
ITT	-	Intent to Treat Effect
M&E	-	Monitoring and Evaluation
MG	-	Meru Greens
MRL	-	Minimum Residue Levels
PF	-	Promoter Farmers
PO	-	Producer Organization
PSD	-	Private Sector Development
PWC	-	PriceWaterhouseCoopers
RCR	-	Results Chain for Reporting
RP	-	Resource Persons
RVO	-	Netherlands Enterprise Agency
SECAEC	-	Solidaridad Eastern and Central Africa Expertise Centre
ToC	-	Theory of Change
TOT	-	Training of Trainers
YPR	-	Young Plant Raisers

Summary, conclusions and lessons learnt

This is the endline report for an impact evaluation of the Food for All Project in Kenya (F4APK). F4APK was implemented with support from the Facility for Sustainable Entrepreneurship and Food Security (FDOV), a Dutch government-funded grant aimed at improving food security and private sector development in developing countries.

Food for All Project in Kenya is a five-year project that started in 2015. F4APK targets 48,500 farm households in 5 counties in the Eastern region of Kenya as beneficiaries. F4APK's primary aim is to increase land and labour productivity in horticulture and dairy value chains, as a pathway towards improved food security. F4APK targets smallholder farmers in dairy and horticulture value chains.

F4APK includes three main types of activities: (1) Hardware/Organisation investments; (2) Training; and (3) ICT interventions. The F4APK investments and ToC focus on the supply-side pathway and the private sector development pathway. The central outcomes of interest are farm and crop level profit, determined by productivity, input choices and marketing channels; and, consequently, farm household income and household food security.

This report provides an impact evaluation of F4APK. The report focuses on effectiveness of the interventions at the level of farmers, but also provides findings regarding relevance, additionality and sustainability of the project. The evaluation finds that the project was successful in reaching implementation targets, translating in positive treatment-control differences in output measures such as training participation, even though the control producer organizations (POs) also received support. The report finds substantive differences in engagement with the target products between the two value chains: farmers in dairy POs are more focused on dairy production than horticulture farmers are on horticulture production. Consistent with this difference, we find some evidence that dairy POs provided a more fertile environment for the F4APK theory of change. Both the project and the evaluation were affected by the COVID-19 pandemic. The report presents "post-COVID" results, i.e. including the effects of COVID on farmers (both control and treatment). For a few variables the report relies on long recall periods to isolate pre-COVID program impact and finds some evidence that, as a result of COVID, F4APK impact was diluted.

Summary of findings

1. Relevance: Is the intervention locally relevant?

Policy relevance is high, beneficiary relevance is determined by value chain. At a strategic level, the **policy relevance** of F4APK is high: the objectives of F4APK are in line with the overall development and growth priorities set out by the Kenyan Government in a range of policy documents, including the Vision 2030 documents and the Agricultural Sector Transformation and Growth Strategy 2019-2029 (ASTGS; Ministry of Agriculture, Livestock, Fisheries and Irrigation, 2019). Shared priorities include increasing small-scale farmer incomes, raising the contribution of agro-processing while increasing added-value, and reducing the number of food-insecure Kenyans. At the project design level, according to stakeholders, relevant components included the strong and concrete market orientation; the infrastructure investment; the connection to and innovations in existing extension networks; and the ICT platform in the context of fast growing (smart) phone ownership.

The picture is more mixed for **beneficiary relevance**. The core composition of F4APK, the strong market orientation and connecting existing producer networks to markets, after quality upgrading, is

highly relevant. However, the expected benefits depend on the strength of the producer organizations, which varies. In particular, dairy treatment POs have been established for much longer than horticulture POs, and their members have higher SES and welfare indicators.

Farmer participation in dairy production POs is high, creating high beneficiary relevance. Dairy PO farmers are very focused on dairy production, as shown by their engagement indicators (Section 4.1.2). The higher engagement levels for dairy farmers make the F4APK package a more directly complementary and therefore more relevant offer for these farmers.

Farmer participation in production of the horticulture crops targeted by the project is limited. Almost half of the targeted farmers do not grow any of the target crops at baseline. Despite that fact that Food for All only selected horticulture farmers with availability of water for irrigation, access to water may well be a limiting factor for these crops, especially since irrigation costs need to be self-financed. Among farmers that grow F4APK target vegetables, irrigation is used in only about half of crop decisions.

2. Additionality: To what extent were the projects additional (according to the DCED definition)?

We find evidence for both input-additionality and development additionality in F4APK. Input-additionality is reflected in the reported narrow surpluses (and sometimes losses) of Meru Greens, the main commercial partner, indicating the need for additional resources to make the project investment. Development additionality is particularly reflected in the explicit attention to female participation in project documentation and reporting, as well as the high reported female representation in project trainings and employment (Section 2.7).

3-4. Effectiveness: To what extent are the projects effective in reaching their outcome and impact objectives? What are the key determinants for inducing or hampering the intended and unintended effects?

We summarize our effectiveness findings in Table 1. Our assessment of "success" is summarized in the last column of this table. It is based on F4APK Annual Report findings for program outputs as described in Chapter 2 (bottom panel); and for short-, medium, and long-term outcomes it is based on a comparison of outcomes, over time, between the F4APK supported farmers and the control group farmers (described in Chapter 4). As expected, farmers in the control group also receive support, from non-F4APK sources, and so we measure F4APK impact against other program support, not zero support.[1] Furthermore, the COVID-19 pandemic has influenced the reality of both treatment and control farmers, and it is not easy to disentangle the effect of this shock on the project effectiveness.

[1] The horticulture control farmer POs were selected from Wards adjacent to the treatment counties Embu and Machakos, with very similar altitude and climate characteristics. For dairy, the program partner Heifer generated a list of groups from two dairy hubs outside the F4APK program. See Chapter 3 for details on the sample and methodology.

Table 1 - Evaluation Summary

Impact level	Judgement criteria	Indicator	Success	
			Horticulture	**Dairy**
(a) Program Outputs (Annual Report data)	Infrastructure & organizations built (Results 2)	Processing unit	++	
		Dairy hubs with hardware		++
		Young Plant Raisers	++	
		Collection centres	+	
		Contracts signed	++	
	Farmer trainings delivered	Number of farmers trained	+++	+++
	Soil testing	Number of farmers with soil tested	+/-	
	ICT platform	Farmers registered	+	++
(b) Short-term outcomes (increase relative to control group)	Use of knowledge	Use of recommended inputs	0	+/0
	ICT / Phone messages	Received message	0	0
		Changed practice bc message	0	0
	Soil testing	Received test	0	
		Changed practice after test	0	
	Training participation	Received training	--	0
		Females trained	++	0
	Increased access to markets	Access to collection centres	++	
(c) Medium-term outcomes	Increased Sales	Certification status	-/0	
		Price levels	0/+	0
		Share of output sold	0	+
	Increased productivity	Productivity	0/-	0 (+)
		Pre-COVID comparison	+/-	
(d) Long-term outcomes	Food shortage	Quantity indicators	0	0
	Increased income and/or profits	Farm income of targeted vegetables and dairy products	0	0

Notes: (1) The top panel (a) gives an absolute assessment for treatment POs, based on Annual Report descriptions; this panel should be read as follows: +/- delivered but below expectation; + delivered, some issues; ++ delivered; +++ delivered above target. (2) The three lower panels (b,c,d) in this table provide assessments based on difference-in-difference regressions that compare changes in treatment POs with changes in control POs. Here 0 indicates a non-significant DiD; + a positive significant DiD; and – a negative significant DiD; non-significant differences are represented as (+) or (-). Mixed assessments using "/" indicate effects that differ by input or output type. There are a few empty cells in the dairy farmer column, indicating that the outcome was not measured for these farmers.

Project implementation was successful: A first necessary step in any theory of change is high quality project implementation. In this regard, F4APK was certainly successful: as described in Chapter 2 of this report, across a range of project results, the project achieved many of the output goals and sometimes exceeded them by a large margin.

Training participation is at a higher and more sustained level in dairy POs. There is, over the study period, also a remarkable difference between dairy and horticulture in terms of training intensity. For the F4APK horticulture POs, the share of farmers trained over the last 12 months reduced strongly, both compared to the horticulture control group and compared to F4APK dairy farmers. We know that

F4APK training implementation continued over 2020 (see Chapter 2), and so this result is consistent with the trainers moving to other communities and POs over time. In contrast, for dairy farmers we find that a) dairy training intensity is at a much higher level; and b) dairy training intensity does not decrease over the evaluation period; and c) dairy treatment farmers are more likely to receive training than the dairy control farmers. Reach of the ICT platform has grown since baseline but quality rating of the messages is medium.

Value chain engagement is much higher in dairy POs. Dairy PO farmers are much more specialized than horticulture PO farmers: 99 percent of dairy PO farmers own cows and 99 percent are engaged in dairy production at endline. These farmers also become more specialized during the study period: at baseline, F4APK dairy treatment farmers use 22 percent of their production acreage for dairy production on average, and this share increased to 35 percent at endline.

For horticulture PO farmers, the baseline value chain engagement was lower and has decreased during the study period along several dimensions: the share growing any F4A crop, importance in terms of value and acreage devoted. In interviews, based on their experience, project management agreed with these study findings on the levels of engagement and the movement out of horticulture.

Effect on inputs and techniques is mixed. The use of inputs on F4APK target crops among horticulture treatment farmers does not increase relative to control farmers. A large majority of dairy treatment farmers use inputs and techniques already at baseline; relative to the control group they significantly increase their purchases of fodder, consistent with the training.

Productivity did not increase significantly relative to the control group. Productivity increases are found in milk production (dairy POs), but these are not significant. For dairy, the productivity level at baseline in the treatment group is 8.5 liters per cow per day and the productivity increase in this group is three quarters of a liter per cow per day, but this not different from the control group. The data do not show qualitatively different trends for the pre-COVID period.

For the main F4APK target crops we do not find positive productivity effects, except for French beans and bananas in the pre-COVID period. The comparison with the pre-COVID seasons provides some evidence of a negative effect of COVID on productivity growth: especially for bananas the data suggest productivity improvements in the treatment group and that COVID reduced these positive productivity effects.

Commercialization of dairy production is very successful. The data show an impressive increase in the share of milk produced that farmers sell to dairy cooperatives: relative to the baseline share we find an increase by 40 percent. Treatment farmers caught up and overtook the control farmers during the study period in this respect. This result is a success for the F4APK theory of change and links the supply side (productivity) pathway to the private sector development pathway.

Horticulture sales via collection centers fall. In the horticulture treatment group, French beans had a substantive share (62 percent) of sales through the collection/distribution center at baseline. This share is reduced at endline to 44 percent, while the share sold by control farmers through this channel increases. This finding is consistent with the annual reports showing that none of the newly installed F4APK collection centres were functional in 2020. Treatment farmers have replaced this channel partially with the “local market”, but overall, the centers remain the main channel for sales of French beans. Non-utilization of the centers has been linked to distrust among farmers caused by non-payment for produce by Meru Greens. For bananas a major shift took place from “other channels” to the “local market” (possibly under the influence of COVID-19 travel restrictions).

Prices for horticulture output fall on average, but more strongly in the control group leading to a positive DiD effect. However, this effect is not driven by the distribution centers but by price changes in non-program market channels (local markets, traders). Dairy treatment farmers report a substantial milk price increase, but we do not find a significant increase relative to the control group.

The overall level of **certification** reported by farmers is low at baseline; for French beans there is a clear drive towards increased levels of certification, both in the treatment and control group.

Profit levels (not project effects) are positive for dairy production and mostly negative for the horticulture target crops. Project staff mentioned recognizing these findings from their experience and observations. An unexpected finding is the strong growth in *dairy* sales and profit for the horticulture POs. This suggests that the observed decrease in horticulture engagement was matched by a successful engagement in dairy, that more than compensated the lower profit income for horticulture crops. There is a general reduction in the value of home consumption and strong sales growth for the staple crops maize and beans, across PO types.

The trends in self-reported food access and life satisfaction indicators are very similar for treatment and control POs in both value chains. Interestingly, these indicators were more positive for dairy farmers at baseline but are not far apart from the horticulture farmers' levels at endline. These movements are similar for control and treatment farmers and therefore appear not to be related to the interventions, or their effects.

5-6. Sustainability: To what extent do the benefits of the project (outcome & impact level) continue after FDOV-funding ceased and how was this influenced by the business case and/or revenue model? Did the project/ intervention lead to systemic change and/or was the intervention scalable?

Sustainability indicators provide a mixed picture. A potential sustainability concern for the processing unit is that the net cash-flow margins appear to be thin. This is consistent with the financial difficulties of Meru Greens during these start-up years, as reported in project documentation and interviews. The nursery business cases are financially healthy and do not raise sustainability concerns. The dairy business case shows good and stable financial reports.

The evidence on the condition of the hardware at the end of the project is mixed. The 2020 Annual Report finds that three out of six investments (representing a large share of the investment) are in good condition and functional at the end of the project, and three are not. The private-sector development pathway outputs included the establishment of organizations and contracts that are still functional. A concern mentioned by the internal project evaluation is that certain aspects of F4APK, including input access and collective output marketing, were largely organized through the project and may require continued support.

The selection of trainers from county government and private sector extension staff means the capacity is there to provide farmer training and extension support after the end of the project. F4APK supported the facilitation of trainings, and provided allowances for the extension staff to visit the producer groups. There is some concern in the internal evaluation whether the trainings will continue without project support.

The Global GAP training and certification has increased, but was mainly focused on French beans. Even though the certificates are valid for one year only, farmers have an incentive to pursue certification in the presence of the processing factory market linkage, provided that the outreach to farming communities, including through collection centres, remains.

7. Corporate Social Responsibility: What is the CSR performance of the selected FDOV projects?

The project CSR plans and measures were relevant and sensible, and appear to have had effect. These plans included attention to and/or mitigation measures for a range of concerns, including child labor, environmental quality, product quality upgrading and certification, and political risks; and appear to have had effect. For example, pesticide use among horticulture group members was lower in the F4APK treatment clusters. Furthermore, the F4APK promoted the participation of women at all relevant levels and this led to a substantial and growing share of female and youth reached, from 45 percent in 2017 to 70 percent in 2019-20.

Conclusions and lessons learnt

Project definition and design

1. **F4APK's theory of change is too ambitious and demanding for a 5-year project.** The aims of F4APK require it to (1) manage a diverse group of stakeholders to implement a broad package of interventions in two very different value chains, including the set-up of marketing and supply channels; (2) get farmers to participate in training, adopt techniques, use information and increase production; (3) get farmers to sell more, increase profits and farm incomes; and so (4) improve their food security; while (5) effecting system change. The literature shows that programs that successfully manage to achieve (2) and (3), especially at scale, are rare (see the systematic reviews in Waddington et al., 2014 and Stewart et al., 2015). **RVO should focus projects on lower level and more realistic outcomes in the results chain.**
2. **F4APK targeted a large variety of producer organizations.** On many levels, the project was more effective in the dairy value chain, possibly because of the larger experience and cohesion of the POs there. The infrastructure investments in the dairy hubs were a success, with a positive impact on sales. **RVO should consider explicitly consider the strength of POs (cooperatives) and commercial partners as a prerequisite and determinant of success** during project preparation.
3. **Previous average growth rates do not guarantee "picking winners" in private sector development.** The evaluation shows that dairy producers did well while horticulture production fell, despite the a priori understanding in the project plan (Netherlands Enterprise Agency, 2014) that horticulture ".. is one of the fastest growing sub-sectors with growth rates of between 15 and 20 percent per year." The F4APK market (PSD) orientation is very important but to be successful requires a good understanding of micro-level supply constraints in the project areas.

Implementation

4. **Despite the ambitious scope of F4APK, project implementation was successful and many output targets have been reached or exceeded.**
5. **In multi-partner PSD projects, RVO and/or project implementers should ask for functional redundancy.** F4APK faced implementation challenges because of financial constraints at the main commercial partner. Such situations may be alleviated if key functions within a value chain are played by more than one partner; although too many partners can also be a struggle.
6. **RVO and/or project implementers should recognize and plan for micro-level supply constraints.** Given the focus on horticulture, access to water and irrigation is a production requirement. In horticulture producer organizations, although the project trained and introduced farmers to water efficient technologies (drip irrigation), farmers mentioned the high cost of irrigation and/or scarcity of water as a constraint to (continued) engagement with the F4APK target crops. This reflects a tension between the business potential offered by horticulture value chains on the one

hand, and the producer costs and risks associated with these on the other (Section 2.6). **More flexibility is needed in project strategy, activities, budget to address such emerging constraints.**

7. **RVO should investigate the potential of phone-based information interventions.** These interventions are very cheap and can potentially be cost-effective and easy to scale. Unfortunately, this component had not been rolled out at the time of the baseline survey. This is an area of active research in East-Africa (see e.g. Fabregas, 2019).

Lessons for evaluation

8. **RVO should make impact evaluation an integral part of the project cycle; project implementation should not start without a complete evaluation design**. Program funders and designers should frontload the research and evaluation questions in project design. The process should include thinking about counterfactual questions and what (secondary) data are available to help construct a credible counterfactual. All this should be part of an evaluation and learning strategy, prioritizing what questions the organization needs answered most urgently and what budget to allocate to these priorities.
9. **RVO should use its funding leverage to improve evaluation design and learning.** Better evaluation should ideally improve the design of the next funded project. In this evaluation, a challenge is that there are no untreated farmers: in all three result areas where F4APK intervened, at least some intervention components were provided to control farmers as well. Even so, there are ways to improve learning, especially if elements of specific policy interest (e.g. phone-based information campaigns) can be studied in separation.
10. **RVO should ask implementation partners to provide digitized monitoring microdata.** The evaluation team found that F4APK has a wealth of M&E micro data for the treatment POs, including PO level production, productivity and price sheets. However, these data are hard to use because they are not digitized. Digitization is relatively cheap these days. We recommend planning an accessible micro data infrastructure in future projects. This will provide management with up-to-date field micro data and facilitate monitoring and evaluation.

1. Introduction

1.1 Background

This is the endline report for an impact evaluation of the Food for All Project in Kenya (F4APK). Food for All Project in Kenya is a five-year project that started in 2015 and implemented until the end of 2020. F4APK targeted 48,500 farm households in 5 counties in the Lower Eastern region (province) of Kenya as beneficiaries. F4APK's primary aim was to increase land and labour productivity in horticulture and dairy value chains, as a pathway towards improved food security.

Figure 1: F4APK project area

F4APK was implemented with support from the Facility for Sustainable Entrepreneurship and Food Security (FDOV). FDOV is a Dutch government-funded grant programme initiated in 2012 that supports public-private partnerships aimed at improving food security and private sector development in developing countries. FDOV is implemented by the Netherlands Enterprise Agency (RVO.nl). FDOV issued calls for proposals in 2012 and 2014 (and another one in 2018 under the successor facility SDGP).

As a background to the project, the F4APK project plan states: "Improving food security through intensive dairy and horticultural crop production is severely constrained by several factors. One of the major constraints is low production and productivity mainly as a result of poor crop and livestock production practices, high post-harvest losses as well as heavy losses associated with disease and pests. Dissemination of technologies to improve production is constrained by weak extension delivery systems. Most of the soils have been exhausted due to continuous farming without informed soil fertility management decisions. […] smallholder farmers are generally excluded from the market systems mainly due to poor organizational capacity to undertake collective marketing, lack of market

information and poorly developed market infrastructure. As a result, the smallholder farmers are often exploited in the market place. This reduces potential income and profitability and is also a disincentive to utilization of essential inputs to improve production."

The project activities focus on the supply-side pathway and the private sector development pathway.[2] The project activities can be grouped into (1) Hardware/Organisation investments; (2) Training; and (3) ICT interventions. The activities include (not exhaustive, see the next chapter for details):

Supply-side pathway:

1. Setting up tissue culture hardening nurseries and young plant risers for improved production inputs;[3]
2. Training farmers on agricultural and dairy production, soil testing; organising demo-plots;
3. Providing production relevant information to farmers via mobile phones.

Private sector development pathway:

1. Setting up horticulture produce collection centres and processing unit, and dairy hubs;
2. Training farmers on accessing market information;
3. Providing marketing relevant information to farmers via mobile phones, e.g. market prices.

The central outcomes of interest are farm and crop level profit, determined by productivity, input choices and marketing channels; and, further down the results chain, total farm household income and, ultimately, household food security.

F4APK was formulated by a public-private consortium consisting of two international NGOs, Solidaridad and Heifer International; two private enterprises, Meru Greens and SoilCares Foundation; and the Kenyan Horticultural Crop Directorate (HCD) for the public sector. The project application and implementation were managed by Solidaridad. The project was selected in the 2014 Call for Proposals and approved for FDOV support in 2015. In 2016 the inception phase was completed and approved. Project implementation started in 2016 and continued until 2020.

The total project budget was € 5,352,986 and was supported by an FDOV contribution of € 2,598,675 or 49 percent; the remainder was financed by the consortium, with the largest contribution coming from Meru Greens (41 percent of the overall budget).

1.2 Overview of the impact study

The evaluation assignment provides the following headline research questions (RQs), distributed as follows over five evaluation areas:

- **Relevance, RQ1:** Is the intervention locally relevant?
- **Additionality, RQ2:** To what extent were the projects additional (according to the DCED definition)?
- **Effectiveness**
  - **RQ3:** To what extent are the projects effective in reaching their outcome and impact objectives?

[2] Although foreign and domestic consumer trust is recognized as important in the project plan, there are no specific consumer level (demand side pathway) activities in F4APK. Consumer trust is addressed *indirectly* through project activities such as production trainings (e.g. regarding pesticide levels) and certification.

[3] The Young Plant Risers can in addition be placed in the PSD pathway, as they are intended to generate sustainable business income for youth groups.

- **RQ4:** What are the key determinants (both internal and external to the project) for inducing or hampering the intended and unintended effects?

- **Sustainability**
  - **RQ5:** To what extent do the benefits of the project (outcome & impact level) continue after FDOV-funding ceased and how was this influenced by the business case and/or revenue model?
  - **RQ6:** Did the project/ intervention lead to systemic change and/or was the intervention scalable? If yes, in what way?
- **Corporate Social Responsibility, RQ7:** What is the CSR performance of the selected FDOV projects?

This study is based on a mix of data sources, representing both quantitative and qualitative information. These sources are: farmer survey data; project documents and data; and interview data. The content and function of these data in the study are as follows.

1. Farmer survey data
These are primary data, collected by survey teams in a sample of both beneficiary and non-beneficiary farmers (in February 2019 and 2021). These data will serve to compare changes in outcomes between project beneficiaries (treatment) and non-beneficiaries (control). We describe the survey data and the methodology in section 3.3.

2. Project documents, reports
These include the 2014 project plan (Netherlands Enterprise Agency, 2014), the annual progress reports, and M&E data. These data will mainly be used for process evaluation purposes, that is, confirming that project tasks and activities have been implemented and/or completed as planned (regardless of their effect on outcomes).

3. Interview transcripts
These include interviews with the national project management, implementers and partners; with field implementation staff and extension workers; and with farmers. The function of these interviews is to answer (a) questions at various levels of project management and implementation; and (b) to provide perspective and better understanding of the context, program components, and survey findings.

In addition, the impact assessment and reporting are informed by articles and reports on the horticulture and dairy sectors in Kenya; on contract farming and value chain development; and previous impact studies of interventions that are similar to program elements included in F4APK. A full bibliography is included in References section.

1.2.1 Focus

In line with the inception report, the overall focus of the Food for All evaluation throughout the study is on a quantitative analysis of the effectiveness questions RQ3 and RQ4. To support this focus and as requested by the FDOV evaluation Steering Committee, the Netherlands Enterprise Agency funded two rounds of survey data collection that provides a comparison over time between farmers in treated producer organizations (targeted by F4APK) and farmers in control organizations. The other five questions, RQ1-2 and RQ5-7, are addressed in this report too, based on various sources of information.

The F4APK evaluation thus features a large survey data collection exercise that allows for the construction of a counterfactual, with the aim to provide more rigorous quantitative estimates of the impact of the project. To do justice to this investment, the F4APK reports feature extensive analysis and discussion of **effectiveness questions**. The central question is whether outcomes of interest have

improved more for farmers participating in F4APK than for comparable farmers who did not participate.

1.3 Limitations of the study

Timing

The F4APK evaluation was commissioned by the Netherlands Enterprise Agency (RVO.nl) as part of an evaluation of several FDOV supported projects. Following a competitive tender in late 2017, a consortium of three organisations (PwC, SEO, and AIGHD) was selected to carry out the evaluation of selected FDOV projects. Five projects, with project start dates between 2013-2015 and targeted project end dates between 2018-2021, were selected for evaluation. The F4APK evaluation is one of the five FDOV project evaluations and is carried out by AIGHD.

It is clear that the F4APK evaluation (and the other FDOV project evaluations) were started late in the project cycle. This timing has the advantage that there is more certainty about project viability before starting the evaluation. This was in fact one of the reasons that F4APK was selected for evaluation (another reason was that the project has many beneficiaries and was more likely to allow for a quantitative analysis). However, this gives rise to methodological drawbacks. First, F4APK is not a randomly drawn project out of the FDOV portfolio but one that has proved itself to be viable. Second, overall project implementation was advanced at the start of the study and it was impossible to do a baseline survey that was completely "pre-treatment". Fortunately, not all beneficiaries received project support at the start of implementation.

To address the methodological problem, the F4APK quantitative evaluation is focused on a sub-set of farmer beneficiaries whose program training started late (2018 in most cases). The production reports provided by these farmers are for 2018 and 2017. As confirmed by F4APK program management, for this group of farmers production impacts are likely to show in harvest reports from 2019 onwards. In addition, the F4APK ICT program components were implemented from 2019 onwards. In other words, our treatment sample had not experienced the full F4APK treatment package during the baseline survey.

A related issue is the selection of control farmers. The evaluation team was not able to create an experimental research design (for example with randomized assignment across producer organizations; or with randomized individual encouragement to participate in trainings). As a next best option, a quasi-experimental research design was created that allows a comparison of treated and control farmers over time (see Chapter 0 for details).

1.4 COVID-19

The COVID-19 pandemic, as a challenge to the implementation and evaluation of F4APK, deserves a separate subsection. The outbreak of the COVID-19 pandemic in March 2020 created substantial problems for farmers and other parties in the value chains, for F4APK implementation, as well as for the measurement of the impact of the project. We provide a brief overview of issues, based on interviews with implementers and farmers. The findings are consistent with other reports, for example the COVID impact assessment by Heifer International (Ojwang et al, 2020). We address COVID-19 as a challenge to the impact evaluation and mitigating measures in Section 3.2.4.

Farm milk production and hub aggregation in 2020 was affected, as most farmers had limited access to extension, inputs and animal health services and advice. In horticulture, production was severely affected during the first four months of the pandemic (March-June 2020). Towards the end of the period, many new "part-time" farmers entered the market because of the lockdown, meaning more

localized competition for the project farmers while overall production increased. Meanwhile, market and transport linkages remained broken.

For the Meru Greens factory, restrictions in movement affected the inflow of farm produce, packaging materials and outflow of the processed products, thus negatively affecting factory output. Factory employment was also negatively affected by social distancing measures.

At the project implementation level, the main effects were felt for in-person meetings and trainings, and activities requiring transport. On the ICT side, the information exchange was protected from the pandemic because of the remote implementation model. Disruption was limited, although technical and field officers could not travel for tech updates. For a number of project implementation and output measures, we find activities being reported for 2020 by farmers. For example, training activities were reported in project monitoring data and in our survey data we find that training intensity remained constant between baseline and endline for dairy farmers, but not for horticulture.

In Table 2 we present responses to questions on the COVID experiences and concerns by the farmers in our sample, provided during the endline survey (February-March 2021). This table shows considerable concern about the risk of infection, and concern about the access to health facilities. The data also show a high level of concern about economic issues: farmers experienced loss of income (dairy more than horticulture treatment farmers), and expressed concerns about possible loss of employment, the ability to sell crops and rising costs of living.

Table 2: COVID-19

Experiences COVID-19 (past 7 days, fraction yes)		**All**	**Control**	**Treatment**	**Diff**
Difficulties going to food market because of mobility restrictions	Horti	0.06	0.03	0.07	0.05**
					(0.02)
	Dairy	0.08	0.14	0.04	-0.13***
					(0.04)
Unable to buy the amount of food my household and I usually consume become my household income has dropped	Horti	0.40	0.45	0.37	-0.12***
					(0.04)
	Dairy	0.42	0.42	0.42	0.02
					(0.06)
Concerns COVID-19 (fraction yes)		**All**	**Control**	**Treatment**	**Diff**
Possible loss of employment	Horti	0.30	0.29	0.30	0
					(0.04)
	Dairy	0.20	0.15	0.22	0.08
					(0.07)
Possible infection myself	Horti	0.77	0.82	0.73	-0.08**
					(0.03)
	Dairy	0.84	0.87	0.82	-0.02
					(0.04)
Not being able to sell crops	Horti	0.23	0.25	0.21	-0.02
					(0.03)
	Dairy	0.20	0.21	0.20	0.01
					(0.05)

No access to health facilities	Horti	0.16	0.07	0.21	0.12*** (0.03)
	Dairy	0.27	0.24	0.29	0.04 (0.06)
Rising cost of living	Horti	0.33	0.29	0.35	0.04 (0.03)
	Dairy	0.45	0.47	0.44	0.05 (0.07)
N		362	142	220	362

Note: the last column ("Diff") provides the treatment-control difference estimate, conditional on covariates. The number in brackets is the standard error of the difference. Significance stars: * p<0.10 ** p<0.05 *** p < 0.01.

1.5 Organization of the report

This report has five chapters, including this introduction (Chapter 1).

Chapter 2 presents the design of F4APK and details the theory of change (ToC) of the component parts of F4APK. It also provides information on the budget, the division of labour among the implementing partners and summarizes the outputs delivered under the program. This chapter also answers research questions on project **Relevance (RQ1)** and **Additionality (RQ2)**.

Chapter 3 discusses the research focus and methodology. It highlights the measurements required to answer the key research questions and the organization of the data collection.

Chapter 4 provides estimates of impact, organized according to impact level (outputs; and short-term, medium-term and long-term effects). The tables provide means of the relevant variables by treatment-control and by baseline-endline. This chapter addresses the questions on **Effectiveness (RQ3 and RQ4)**.

Chapter 5 addresses the remaining questions on **Sustainability (RQ 5, RQ 6)** and on **Corporate Social Responsibility (RQ 7)**.

Chapter 6 summarizes our findings and concludes.

Note for readers

This is a stand-alone impact evaluation of F4APK. The report includes a few sections from the evaluation baseline report that have remained largely the same. These are sections 2.1—2.4 on F4APK program design; and Chapter 3 on evaluation design, except for the new sections 3.2.4 and 3.3.2.

2. Food for All Project Kenya[4]

2.1 Design

Food for All Project Kenya has five project result areas, which cover all the project activities and expenditure (Food for All Project Plan, Netherlands Enterprise Agency 2014). The key activities for this study fall under result areas 2, 3 and 4. These three areas account for 95 percent of the project budget and cover the following activity groups: Hardware/Organization investments; Training; and ICT interventions. Result areas 1 and 5 mainly cover the inception phase and M&E activities and will not be part of the analysis.

F4APK targets smallholder farmers in dairy and horticulture value chains. Within horticulture, the program targets seven specific crops: French beans, Banana, Watermelon, Butternut, Onion, Tomato, and Capsicum. Within dairy, the main product is milk.

We now turn to the detailed program activities and outputs, as listed in the Results Chain for Reporting (RCR). The RCR corresponds to the project-result areas, and is divided in a number of sub-results.[5] The sub-results are part outputs to be realized, part outcomes to be affected by the project. In Table 3-Table 5 we present the RCR sub-results that are project *outputs* to provide a concise summary of the project activities, by result area and by pathway. Note that in this chapter we focus mainly on project activities and outputs, not on outcomes. The tables present sub-results (outputs) that were planned under the supply and PSD pathways, respectively.

2.1.1 Hardware/Organization investments (Result Area 2)

We find two types of project activities under Result Area 2. First, building hardware and setting-up organizations has two distinct support functions in the value chains. In the supply pathway, tissue hardening nurseries (banana) and young plant raisers are set-up, to improve market availability of quality inputs that can be bought by target farmers. In the Private Sector Development (PSD) pathway investments are made in marketing channel hardware to facilitate logistics and increase returns to agricultural production. These investments include establishing collection centers and dairy hubs; the set-up of banana ripening chambers, to set higher and uniform product quality; and a horticulture processing unit. A second element is ensuring contracts are in place between the newly established organizations (young plant raisers, dairy hubs) and firms that provide inputs and services to these units.

Table 3: Hardware/Organization investments

F4A Activity	Supply Pathway	PSD Pathway
Hardware and Organizational Investments **(Sub-results 2.1, 2.2, 2.5)**	10 Tissue culture hardening nurseries 15 Young plant raisers (YPR) Personnel of these organizations trained.	5 banana ripening chambers 10 Local and export produce collection centers and 5 dairy hubs established and operational 1 processing unit (jobs created)

[4] The description of the F4APK design and implementation structure in this chapter is based on project documentation and interviews with project implementers and administrators.

[5] For each sub-result progress is monitored and verified using agreed "Means of Verification". An example of an MOV is "Photos of the established hardware and Location coordinates of the hardware established".

VALUE CHAIN SUPPORT **(Sub-result 2.2)**	Long term Contracts established between service providers (Biotechnology labs, Seed Companies and agro dealers) and nurseries/Cooling hubs.

2.1.2 Trainings (Result Area 3)

Training activities are grouped under result area 3. The training programs were developed for six different topics in 2017 (see Annual Progress Report). The headline topics and sub-topics[6] are

- Good Agricultural Practices (GAP), including sub-topics: horticultural crop nutrition, crop pests/diseases and management, GAP for each F4APK focus crop;
- Livestock production practices, including modules on breeding and artificial insemination (AI), animal feeds management, calf rearing, housing and security (for zero-grazing), disease management;
- Market information, including record keeping and marketing (channels), Global GAP certification;
- Soil testing (topic is part of agricultural practice training), including soil fertility management, soil sampling, analysis and crop nutrition;
- Governance (no sub-topic information provided);
- Group dynamics and social capital development, including accountability, sustainability, full participation, improved animal and resource management, group formation.

The training programs aim to reach 48,500 farmers and are organized through a cascaded system, with training of 525 trainers at the start of the cascade. This is followed by training of 48,000 farmers by the trainers of trainers. As part of the trainings, which follow a practice-oriented Farmer Field School (FFS) model, demonstration plots are established. Also included in this result area is soil testing, both the test implementation and the training (capacity building) of Meru Greens staff.

Table 4: Trainings

F4A Activity	Supply Pathway	PSD Pathway
Trainings **(Sub-results 3.1—3.6, 3.8)**	Six farmer and group training support programs developed 525 Training of Trainers (TOT) trained on the six key areas 48,000 smallholder farmers trained by TOT on the six key areas[7] 100 Extension staff from the Ministry of agriculture trained (sustainability) 120 Demonstration plots on Banana (15), French beans (20), Local vegetables (60), Drip irrigations (10), Fodder (15) established, soil testing done	

[6] Based on topic lists received from F4APK management; full lists are provided in Appendix B Training topics

[7] Promotor farmers were mentioned in the project plan but are no longer part of the training structure, following a request for change by the implementers.

	A Joint learning platform organized for farmers to learn on best practices in other projects
Soil testing **(Sub-results 3.7, 3.10)**	10 Meru Greens staff trained on Soil sampling, analysis and understanding results 75% (36,375) of the smallholder farmers have their soils sampled and tested by Soil Cares Ltd

2.1.3 ICT platform (Result area 4)

The main aim of the ICT platforms is to collect information on farmers and disseminate relevant information such as agroeconomic tips, weather information and prices. Therefore, the intervention is located in both pathways. The information will be used to create a dashboard that supports monitoring of farmers and short-term decision-making. The messages are sent in two forms: apps (on Android, in case of smart-phone) and regular text-message. Solidaridad plans to use farm-level data stored on the platform for M&E purposes. The platform can also eventually assist buyers to obtain information about supply and farmers to get information about demand (virtual marketplace). Practically the aim is to develop collection centres to be information hubs, through which market information is accessed, bids and offers are uploaded.

Table 5: ICT platform

F4A Activity	Supply Pathway PSD Pathway
ICT platforms **(Sub-results 4.1—4.3)**	10 producer and 5 Milk collection centers (Dairy hubs) transformed into information hubs 300 members of the collection centers have been trained on accessing information via information platform 48,500 smallholder farmers have been trained on accessing extension and market information by mobile phones Market Linkages and networks have been established (database)

2.2 Theory of change

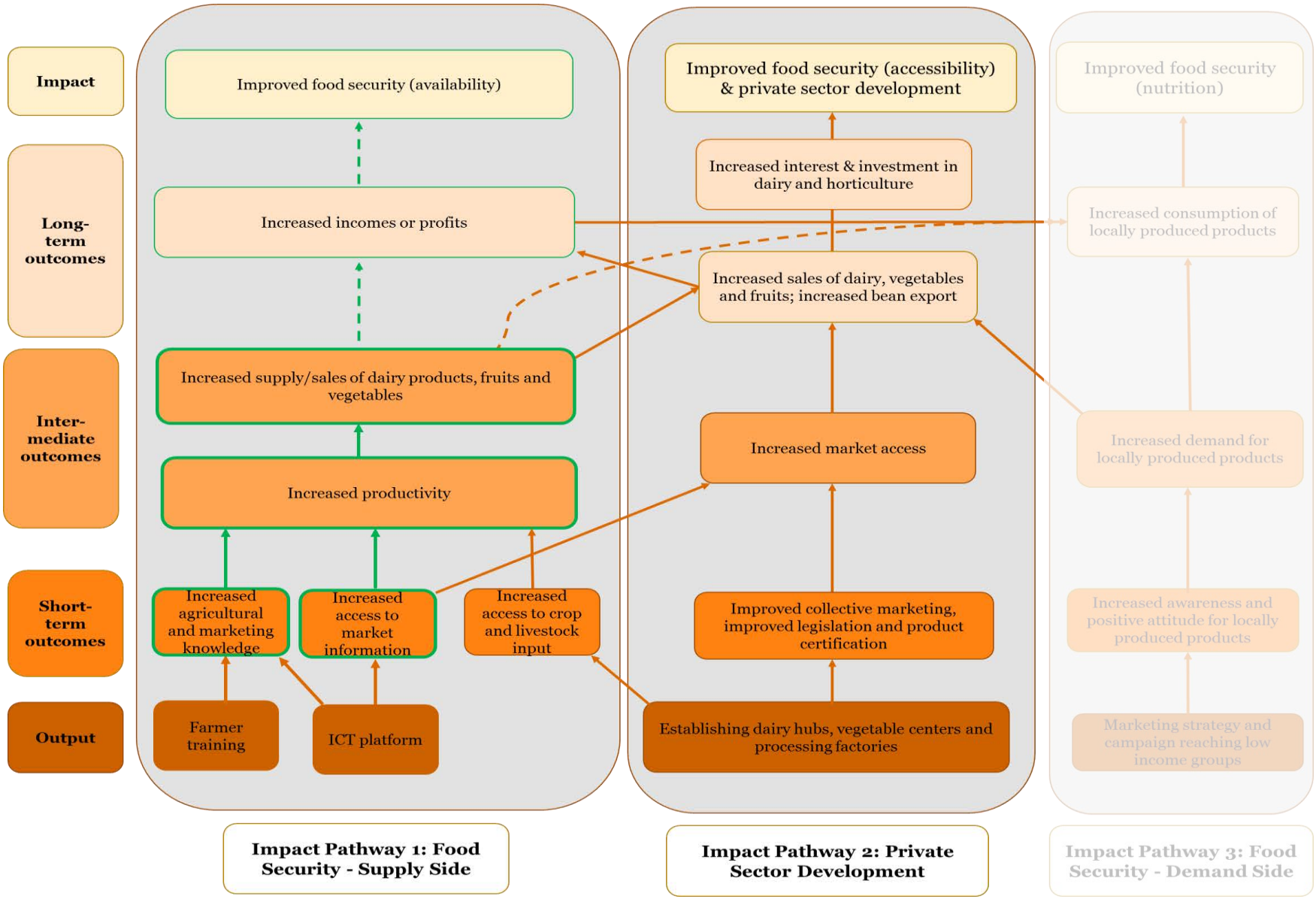
The F4APK Theory of Change is based on the situation analysis presented in the project documentation, which lists a number of constraints faced by smallholders in Kenya. Lifting these constraints motivates the design of the F4APK program components.

For example, the project plan (Netherlands Enterprise Agency, 2014) states: “The underlying pressure for change in the small holder farming systems is driven by a number of factors which include [..] poor livestock and crop production methodologies, leading to low production and productivity; weak systems for dissemination of new production improvement technologies as well as weak market systems. The poor organizational capacity of small holder farmers has led to their exclusion on the market systems, as a result of volumes and poor-quality products, which does not meet the consumer expectations, leading to low returns from their crop and livestock enterprises.”

Improvement of the situation requires “.. more intensive farming of high valued horticulture crops, fodder and adoption of improved livestock breeds ..” and “ .. farmers to shift to a market oriented commercial production of crops and livestock”. The project plan states that such shifts require good quality and affordable extension services; better marketing and quality control; and strong producer organizations: “Product aggregation and effective coordination of production and marketing activities in the small holder farming system is needed to address these changes. In addition, a strong and

empowered producer marketing organization is a key structure needed for effective coordination of production and marketing activities, linkages to markets and other service providers."

Figure 2: Theory of Change


Impact
Long-term outcomes
Inter-mediate outcomes
Short-term outcomes
Output
Improved food security (availability)
Increased incomes or profits
Increased supply/sales of dairy products, fruits and vegetables
Increased productivity
Increased agricultural and marketing knowledge
Increased access to market information
Increased access to crop and livestock input
Farmer training
ICT platform
Improved food security (accessibility) & private sector development
Increased interest & investment in dairy and horticulture
Increased sales of dairy, vegetables and fruits; increased bean export
Increased market access
Improved collective marketing, improved legislation and product certification
Establishing dairy hubs, vegetable centers and processing factories
Improved food security (nutrition)
Increased consumption of locally produced products
Increased demand for locally produced products
Increased awareness and positive attitude for locally produced products
Marketing strategy and campaign reaching low income groups
Impact Pathway 1: Food Security - Supply Side
Impact Pathway 2: Private Sector Development
Impact Pathway 3: Food Security - Demand Side

The overall Theory of Change (ToC) for F4APK is presented in Figure 2 (also provided in the FDOV inception report. Project activities (outputs) are summarised at the bottom, while envisaged outcomes over various time lags are presented from bottom to top. Improvements in farm productivity, followed by increased sales and income (intermediate and long-term outcomes) are at the heart of the ToC in the Supply Side pathway. Investments and activities that support production of improved inputs or facilitate marketing are included in the PSD pathway. Since the project does not include activities that directly target (low income) consumer demand, the Demand Side pathway will not be included in the discussion.[8]

As indicated in the inception report, the F4APK evaluation will focus on farmer level effects in the supply side pathway. These effects and outcomes are marked in green in Figure 2. PSD pathway investments are analysed insofar they affect farmer level outcomes. It is worthwhile to assess the program logic leading up to these outcomes in more detail at the level of the individual program components.[9]

2.2.1 Hardware/Organization investments

(A) Investments aiming to improve production inputs

Food for All supports improved production inputs in two indirect ways. First, it invests in organizations that produce higher quality inputs – seedling nurseries and young plant raisers. Second, during the production trainings the use of improved inputs is discussed and promoted.

F4APK does not provide (subsidies for) inputs directly. French bean farmers receive French bean seed on credit from Meru Greens, a standard practice in contract farming. Once the produce is harvested, it is expected that farmers sell produced beans back to Meru Greens. At that stage, the cost of the seeds is subtracted from the sales value. Different models are used for the other F4APK fruit and vegetable crops and milk. For the (domestic market) fruits and vegetables, the farmers cater for all production costs. The dairy cooperatives provide various extension services to their members (ranging from AI services, feeds and veterinary services), but these are paid services and not financed by Food for All. However, milk farmers may access these services on credit from the dairy cooperatives.

For improved inputs availability to have the desired productivity effects, a number of steps in this results chain are logically required:

1. Farmers in target group have information about the availability of higher quality horticulture (seedlings) and dairy (AI, veterinary services) inputs;
2. The information is new, salient and attractive;
3. Farmers start making use of the availability of these good quality inputs;
4. Improved inputs use increases productivity, sales, profits and household income.
5. Improved food security through increased income and higher consumption, both from purchased food and own production.

The availability of inputs and trainings that promote them are clearly complementary. The combination of availability and training is expected to increase the use of improved production inputs, by addressing points 1, 2 and 3. In addition, the project expected that seedling supply via Young Plant Raisers (YPR) would be an attractive innovation for farmers who before established their own nurseries. (“The idea that one can buy 100 ready seedlings to plant in their kitchen garden will be interesting.”)

[8] Note, however, that improved food security is one of the ultimate focal outcomes for F4APK and as such it is represented at the top of Figure 1 in both the Supply Side and PSD pathways. While this is a consumption related outcome, it is not exclusively part of the Demand Side pathway.

[9] Based on project descriptions and conversations with the Solidaridad implementation team.

(B) Investments aiming to improve marketing channels

Food for All investments further aim to improve marketing channels for both Horticulture and Dairy (see Section 2.1.1, Table 3). For Horticulture, these investments are:

1. Construction of a processing units for French beans (completed in 2017, in operation now), financed and managed by Meru Greens;
2. Establishment of produce aggregation/collection centers for initial grading and bulking of French beans, and facilitating logistics.
3. Establishment of banana ripening chambers, to improve quality and sorting of bananas; and facilitate marketing logistics.

For Dairy, the investment is in five Dairy hubs, or milk collection centers. These centers have milk coolers to ensure cool storage and facilitate marketing up the value chain. They will also have equipment for checking milk quality, value addition/preservation, hygienic milk dispensation (and will host the ICT platform, see below). At the hubs, farmers will also have access to AI services, extension on fodder production, and input services.

The building of collection centers is paid for by the project funds. Once completed the collection centers are owned and managed by the respective dairy cooperatives or producer organizations. These marketing investments assume that marketing channels were/are sub-optimal before the intervention. In principle these investments should improve/increase the range of output options, and the quality of the output so that higher prices can be charged by farmers. For example, the idea behind the collection centers and banana ripening chambers is to facilitate produce aggregation and value addition to improve farmers access to markets and better prices. If the additional and/or improved channels opens more attractive marketing possibilities we may expect to see shifts in the amount of produce sold through these channels; improvements in the prices obtained; or both.

Improving market opportunities is related to quality standards, particularly for French beans, an export product. As awareness of certification standards is a requirement, this is aspect is closely related to training. In the project proposal one of the deliverables is to take 20 producer groups through Global GAP training leading to certification. The groups are trained by local consultants (approved by Global GAP), when they are up to speed with standard requirements and have put in place the necessary infrastructures, certification bodies are brought in to conduct audits and award certification where it is due.

For improved marketing channels availability to have the desired effects, a number of steps in this results chain are logically required (some of these align with the training ToC):

1. Farmers in target group have information about the availability of marketing/output channels, e.g. via their farmer group;
2. The information is new, salient and attractive;
3. Farmers start making use of the availability of these improved marketing channels, potentially after making required quality improvements (certification);
4. The marketing channels increase the price obtained and/or increase sales volume and sales certainty; and these improvements are sufficient to offset any potential “disutility” associated with the channel, including added quality control and logistics, uncertainty about rejection rates, pay delays.
6. The improved price triggers an increase in production, productivity and household income.
7. Improved food security through increased income and higher consumption, both from purchased food and own production.

The marketing channel results chain is linked with the general characteristics of contract farming (CF). The advantages and challenges of CF schemes have been described by previous studies (see e.g. Strohm and Hoeffler, 2006; Ragasa et al., 2018). A well-known CF challenge is side-selling (out-selling), where contract growers sell part of the contracted harvest to alternative buyers, e.g. at the farm gate. If step 3 is not sufficiently strong, out-selling may remain widespread and can hinder a sustained increase of contract sales.[10]

In the case of F4APK, the MG contracts are with the Producer Organizations, which is a convenient, less time-intensive way of contracting. With a group level contract, however, the success of the contract partly depends on the strength of the Producer Organization (PO) management structure. A generic issue is that some local "Self Help Groups" do not have a legal status and so contract enforcement in court is unlikely. The project plan appears to have anticipated this issue, with planned trainings on "group dynamics" and "collective marketing".[11] These two training programs are not listed on the RCR (Table 4), but "social capital development" is. In addition, the collection centers reduce the distance and costs between buyer and seller in CF. This is important because transportation costs are high.[12]

2.2.2 Trainings

As shown in Table 4, F4APK developed training programs for six different topics. Two of these are directly aimed at productivity enhancing techniques (Agricultural production practices; Livestock production practices). Soil testing has a possible indirect positive effect on horticulture and fodder production.

Farmers are approached for production training by F4APK through their PO. Farmers who are member of a dairy cooperative are approached for dairy production trainings. Farmers who are member of a PO that focuses on agriculture/horticulture are approached for horticulture trainings. Typically, farmers will engage in both dairy and horticulture but will differ in their degree of specialization. In theory, a farmer might be approached for both types of training.

Training modules can take up to two months. Farmers receive the training in 6 sessions under the horticulture production component and 12 sessions under the dairy production component. The horticulture farmer training is implemented in cascaded fashion via Training of Trainers, who then host decentralized farmer training. The latter follows the Farmer Field School model, with a practice-oriented training based on demo plots to show how to apply techniques and where farmers can observe the results in their village. Installation of demo plots is also supported by F4APK, under result area 2. Dairy farmer trainings are delivered by both resource persons (RPs) and Promoter farmers (PFs). The RPs are livestock and veterinary technical experts who are selected and coordinated by Heifer, one of the project partners. The RPs train farmers on technical topics such as feeds and feeding, disease management, breeding etc. The PFs are selected and trained (by Heifer) to deliver farmer trainings on group dynamics and social capital.

[10] For example, for small farmers who grow vegetables as one of their activities, side-selling may be attractive given low sales volumes, a relatively low fixed contract price and slow payment. The farmers can get a better price at the farm gate with commercial traders, with payment over M-Pesa. For larger farmers who need to sell a lot of produce CF is more attractive because it provides security for a large volume and value.

[11] Moreover, the project plan states "The project will organize the 48,500 small holder farmers into marketing organizations (Producer marketing groups and Milk hubs) and link the organizations to markets and service providers. The organizations will be strengthened to operate as viable business units, linking the small holders to markets […]." (Netherlands enterprise Agency, 2014, p.9).

[12] According to Strohm and Hoeffler (2006), "… it is more costly to send a truck from Nairobi to Mombasa than shipping something to Belgium".

This type of intervention assumes that knowledge of agricultural techniques and inputs is a binding constraint on productivity. It aims at improving knowledge, changing attitudes, adoption of techniques, "intensification" of production and higher productivity. For trainings to have the desired effects, a number of steps in this results chain are logically required:

1. Farmers in target group attend (all) training sessions;
2. The information provided in the training is new, salient and attractive;
3. Training improves knowledge of and attitude towards techniques, inputs;
4. Training leads to adoption of intensification techniques;
5. Training and adoption increase productivity and household income.
6. Improved food security through increased income and higher consumption, from purchased food and own production.

2.2.3 ICT platform

Investments in ICT are under F4APK result area 4. The main aim of the platform is to collect information on farmers and disseminate relevant information such as agroeconomic tips, weather information and prices. The information will be used to create a dashboard that supports monitoring of farmers and short-term decision-making. The information provided can potentially support farm production (supply-side pathway) on one hand, and support marketing channel decisions (PSD pathway) on the other.

The messages are sent in two forms: apps (on Android, in case of smart-phone) and regular text-message. The messages that will be sent to farmers are envisaged to be farm-specific based on the information shared by farmers at registration.

For improved information, delivered by phones, to have the desired effects, a number of steps in this results chain are required:

1. Farmers in the target group are registered on the ICT platform database;
2. Farmers receive messages through the ICT platform on their phone;
3. The information/messages are new, salient and helpful for farmers;
4. Farmers make changes based on the information;
5. The changes result in higher farm productivity, sales and household income.
6. Improved food security through increased income and higher consumption, from purchased food and own production.

2.3 Project budget summary

The total F4APK budget is 5,352,986 Euro (F4APK project plan, 5.2). The project partners and their financial contributions to this budget are as follows (5.1: RVO-FDOV 49%, Meru Greens 41%, Solidaridad (Eastern and Central Africa Expertise Centre, SECAEC) 6%, Heifer International 4%, SoilCares Foundation 1%, Horticultural Crop Directorate 0%. The budget summary in Table 6 provides a budget breakdown over result areas and shows that Result 2 has the largest budget share (56%). Consistent with the activities in this Result Area, which include the establishment of a beans processing unit by Meru Greens (Table 3), the share of hardware in Result 2 is relatively high (81%). Budget shares for the other two main activity areas are smaller at 26 percent (Result 3, Trainings) and 13 percent (ICT). Using the targeted group of 48,000 farmer beneficiaries we calculate a project cost per beneficiary of 112 Euro over the duration of the project.

Table 6: Project budget composition

Result area	Total (Euro)	Budget share	Hardware

1 Inception	205,794	4%	0
2 Hardware/Org	2,999,709	56%	81%
3 Trainings	1,403,621	26%	6%
4 Ict	691,497	13%	10%
5 M&E (Ao)	52,365	1%	0%
Total	5,352,986	100.0%	48%

2.4 Project partners

F4APK was proposed and is being implemented by a public-private consortium consisting of two international NGOs, Solidaridad Eastern and Central Africa Expertise Centre (SECAEC) and Heifer Project International (HPI); two private enterprises, Meru Greens Horticulture and SoilCares Foundation (represented by SoilCares Ltd in Kenya); and the HCD. For more details on the implementing partners, please see the F4APK Project Plan.[13]

Table 7 provides an overview of the functions within the two value chains that F4APK targets. The functions that are supported directly through project activities are those where one or more sub-results are listed. The functions where no sub-result is listed are not directly supported by project resources, but are value chain functions that are expected to improve as a result of the project activities.

Table 7: Value chain functions

Value Chain Function	Sub-results	Value Chains	
		Horticulture	*Dairy*
Marketing			
1. Export sales		Meru Greens	None
2. Domestic sales		Meru Greens, brokers, Twiga Foods, Mase Foods	Meru Central, Daima, New KCC
3. Certification		HCD, Meru Greens	None
4. Aggregation, processing	2.1, 2.2	Meru Greens: processing unit, collection centers, banana chambers (farmers contribute time)	Dairy Cooperatives, processors (Meru Central, Daima, New KCC) Heifer: Milk coolers
Extension services			
5. Managing trainings, extension services	2.5 3.1-3.6, 3.8	Solidaridad, HCD, Meru Greens: trainings, extension services, demonstration plots, irrigation kits	Heifer: building capacity, Cooperatives (extension services)
6. Inputs	2.3	Solidaridad, Meru Greens: YPRs, seedling nurseries; Local agro-dealers (Hybrid Seeds, fertilizer, pesticides)	Dairy cooperatives (veterinary services, feeds, AI), local agro-dealers
7. Soil testing	3.7, 3.10	Soil Cares	Soil Cares
8. Information	4.1-4.3	Mass media, brokerage, cell phones, local market, peers	Mass media, brokerage, cell phones, local market, peers, cooperatives
Overall project management		Solidaridad	Solidaridad

Source: provided by F4APK management.

[13] Kenya Highland Seeds (KHS) is mentioned in the project plan to manage parts of the project but eventually did not take up the intended role.

2.5 Project implementation

This section provides a summary of the activities and outputs of F4APK, based on the annual progress reports of 2017-2020.[14] We refer to the report themselves for full details of the implementation progress. We provide a timeline in Appendix A.

In the next three sub-sections we briefly discuss the overall activity and output progress using a set of figures, based on the progress reports, that illustrate F4APK's achievements. In the fourth sub-section we discuss the overall implementation progress of F4APK. In Section 2.5.4 we discuss a few specific implementation challenges.

2.5.1 Result Area 2

According to the F4APK progress reporting, a high share of the project infrastructure targets in Result 2 have been achieved, and many of these outputs were delivered early in the project cycle.

In Result 2.1, the French beans processing unit was established 2016 and has been in operation since then. For dairy, the five hubs were in operation in 2018 and all planned hardware had been delivered to the hubs (milk coolers, generators, lab equipment and dispensers).

Figure 3 shows the progress for three horticulture investments. Of the 15 planned Young Plant Raisers (YPR), 11 had been established by 2017, the remaining 4 were added in 2018. The tissue hardening nurseries were added more slowly and fully installed by 2020. Horticulture produce collection centers are the only infrastructure type not fully completed: by 2019 7 of the 10 planned centers were completed. In addition, the 2020 AR finds that the 7 collection centers were not utilized (see next section on implementation challenges).

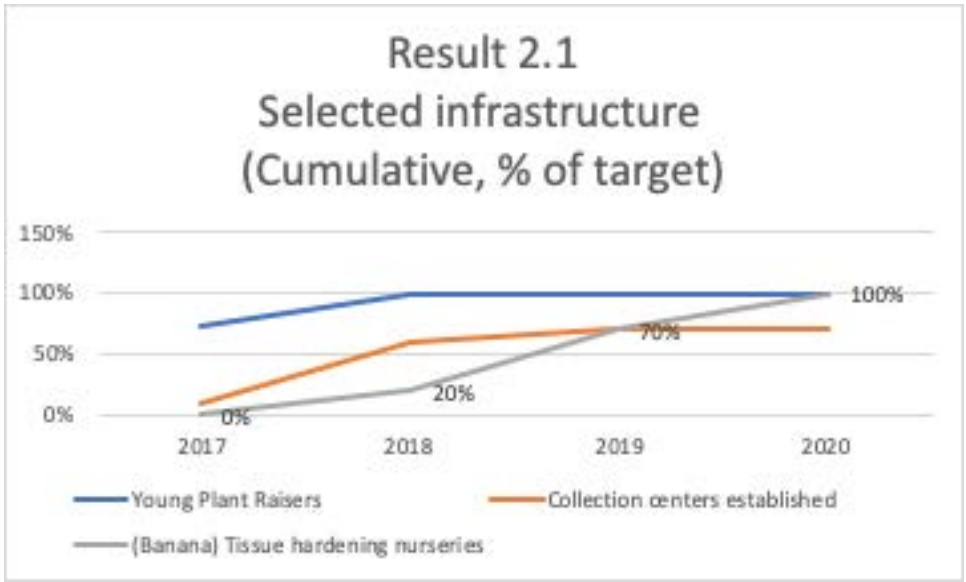


Figure 3: Infrastructure

The Annual Reports show consistent results corresponding to these investments. For example, the reports show substantial job creation (Result 2.4, Figure 4) at the Meru Greens French beans processing unit, equivalent to more than 900 person-years over the 2016-2020 reporting period.

[14] To be precise, these reports are labeled "Annex 3b: Annual Progress Report" to the Netherlands Enterprise Agency for Food for All Project Kenya, project code FDOV14KE63.

Similarly, the reports find that contracts were signed for seed deliveries to the MG processing unit, and for milk deliveries to the dairy hubs, and for other Hub inputs, including software. Consistent with the timeline of the horticulture infrastructure investments in Figure 3, the reporting for Result 2.3 shows YPR sales from 2018 onwards (averaging about EURO 4,000 per year) and tissue hardening nursery sales in 2019 and 2020 (averaging about EURO 4,500 per year). Finally, the AR data show that the cumulative number of farmers trained on seedling raising and nursery management is 1035, or twice the target of 500.

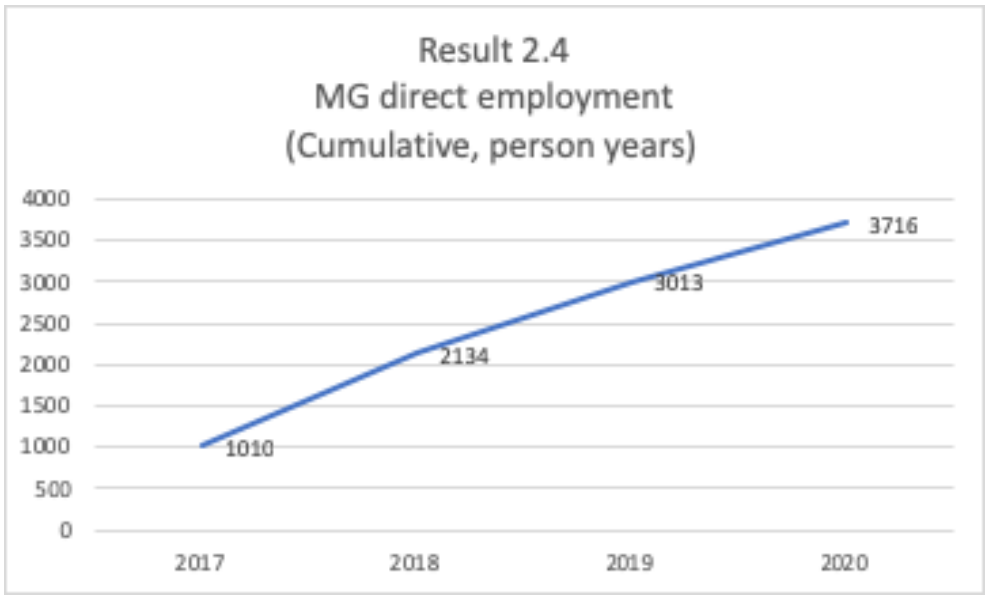


Figure 4: Employment

2.5.2 Result Area 3

In Result Area 3 (training), we find encouraging output results. Training materials were developed (Result 3.1); Trainers of Trainers (ToTs) were trained for horticulture and dairy specific trainings (Result 3.); and peer farmers were trained (3.3). As a result of this set-up, a large number of farmers was trained (3.4). In addition, Ministry of Agriculture extension staff were included in the trainings (3.5); and Meru Greens staff were trained on soil sampling and testing (3.7).

A consistent and remarkable feature of the reporting is that the cumulative number of trainees of the 2016-2020 period is larger than the target, in some cases much larger. If we take the AR numbers and add across years, the ratio of cumulative to target for the various training areas is: 265% for horticulture ToTs and 156% for dairy ToTs; 235% for peer farmers; and 164% for farmers (see Figure 5).

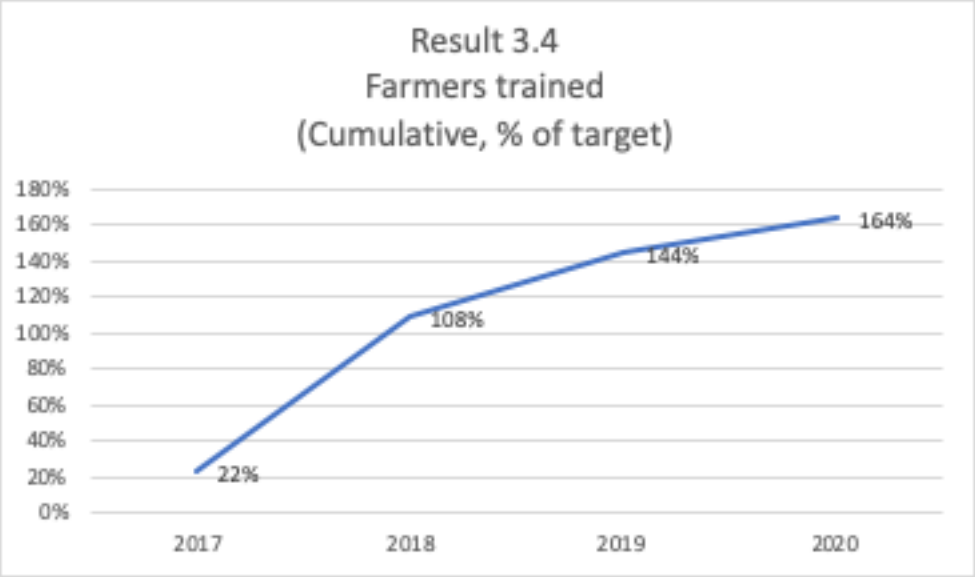


Figure 5: Farmer training

We see two explanations for these results. One is that the trainee numbers in the AR in year T include individuals that were also trained and counted in T-1, for example if a training extended across two calendar years, or if a farmer decided to enlist in a second training. The 2020 report provides a cumulative number of farmers trained of 70,645, which is about 8,000 less than when the AR numbers are added across years. Nevertheless, the qualitative conclusion is the same, as this is 147% of the target of 48,000 farmers trained.[15]

The second explanation is that targets were set too low, relative to the capacity of the various implementers to mobilize trainees. This is an important finding for a number of reasons: many food security and agricultural support programs feature farmer training and being able to predict the number of trainees precisely is important for future programs.

Under the remaining result areas, we find encouraging output results as well. Joint learning platforms (JLP, result 3.6) were organised (for farmers to learn best practices) in the last three reporting years, 2017-2020. The total number of demonstration plots (Result 3.8) reported as established, summed over the four years is 222, which is 185% of the target of 120. This number appears to include double counting in the ARs, as the 2020 report provides a cumulative total of 132 or 110% of the target.

Finally, in Result 3.10, the total number of farmers who have tested their soils since the inception of the project is listed in the 2020 AR as 13,390 (or 75% of the revised target of 17,949).

2.5.3 Result Area 4

The 2020 AR writes under Result Area 4: “A lesson learned was that the activities under this result area should have been undertaken before farmer training started under Result 2 and 3. This would have made it easy to create awareness and enroll farmers into the Farmers’ Realm mobile application. Contrary, the ICT training and farmer registration was implemented almost at the tail end of the project.”

[15] A solution against double counting is to require the AR to report both the total number of farmers trained; and the number who had been trained by the program for the first time.

This quote is consistent with the relatively slow progress reported in the baseline report of this evaluation; and with the absence of detailed reporting on the ICT platform development (Results 4.1-4.5) in the reports for 2016-2020. See section 2.5.4 for details.

Nevertheless, towards to end of the project cycle, a number of features of the ICT platforms for horticulture and dairy were reported. The first ICT platform is aimed at horticulturalists. It has the form of a mobile application for farmers who have smart phones (it also has a USSD functionality for farmers who have feature phones). The application is in Google play store by the name 'Farmers Realm'. Development of the platform backend was completed at the end of 2018, the application and USSD were completed in mid-2019. Piloting was done with about 100 trainer of trainers (TOT) over the months of July and August 2019. Official roll-out was done in September 2019 starting with training and onboarding of 125 TOTs across the 5 project counties. The TOTs were then mandated to publicize the App and train farmers on how to use the App; each TOT registering 100 farmers (anticipating a total of 12,500 farmers). Registered farmers are expected to tell their neighbors and increase the number of users.

The application has the following functionalities:

1. Live chat function: farmers can engage agronomists on live chat just like WhatsApp;

By the end of the project, F4APK reported having a database of 30,000 users, of which an estimated 50 percent are active users.

The second ICT platform that was developed and rolled out is the Dairy management system. This was with 2 dairy cooperatives (Mkulima Bora and Makueni Dairy hubs in Embu and Makueni respectively). The platform is used for management of data at the dairy cooperative including, inter alia, milk data, processing, sales, farmers, payments, input sales, services, and credit to members. The platform was rolled out in quarter 3 of 2019 and is currently being run in parallel with the manual system at the cooperatives.

The Result Areas 4.6-4.9 relate to absolute production volumes (no reporting was done on 4.10). We find that in all cases the cumulative target was exceeded by a large factor, often already in a single reporting year. A striking example is French beans (Result 4.6, Figure 6). The result is "16.5MT of conventionally produced and of organically produced beans have been produced, processed and marketed per year", so the cumulative target over 2017-2020 is 4*16.5 MT = 66 MT. The annual average volumes procured and exported by MG are 1925 MT and 1426 MT, respectively. The cumulative volumes are 7700 MT and 5704 MT; translating into 11,667% and 8,624% of the targets, respectively. It appears that the targets were set too low in these result areas and, in any case, were met by a large margin.

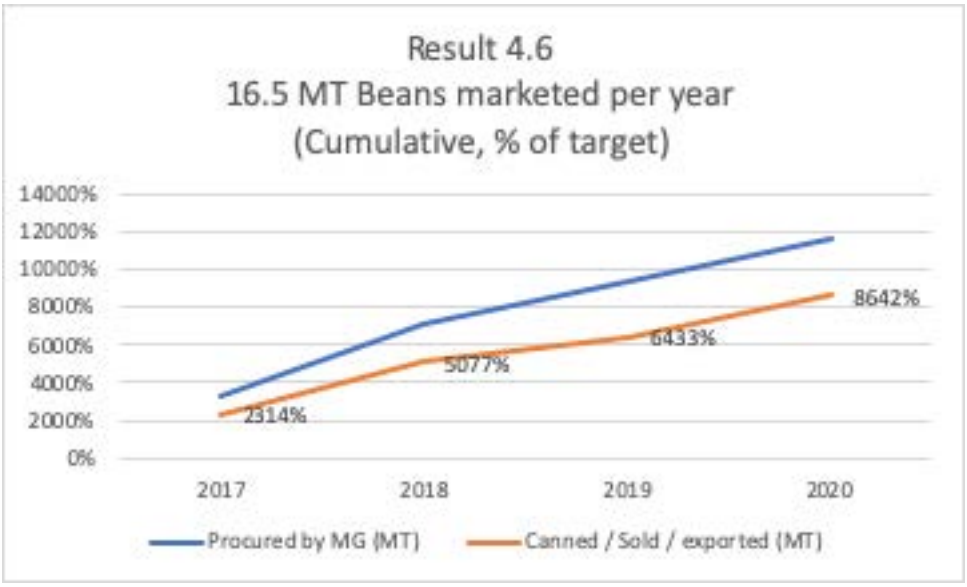


Figure 6: French beans

2.5.4 Specific implementation challenges

F4APK encountered a number of specific implementation challenges; we describe five sets of these problems, in descending order of estimated severity. These challenges were reported in the Annual Reports to the Netherlands Enterprise Agency (RVO), as well as in interviews with key stakeholders.

The first implementation challenge, the outbreak of the COVID-19 pandemic in March 2020, and its implications on the implementation, production and management side have already been discussed in Chapter 1.

A second set of implementation challenges relates to unfulfilled commitments by Meru Greens, the largest private sector project partner and responsible for market uptake of French beans. This problem was described by several key informants during interviews, and includes: delayed farmer payments for French beans and banana seedlings, causing farmer dissent and withdrawal (2017-18, but also reported in the 2020 AR); produce collection centers reportedly not being used in 2020 because of pending payments; failure to install the remaining three produce collection centers (see a.o. Figure 3); and failure to start operating the only modern banana ripening chamber afforded by the project (2020 AR). The interview respondents also mentioned that Meru Greens made large positive contributions to the project. We will return to these challenges in the conclusion.

A related challenge, mentioned several times during the evaluation, is the uneven organizational strength of the two subsectors in Food for All, dairy and horticulture. In particular, it was mentioned that on the horticulture side, the level of organization and coherence of producer organizations is low compared to dairy cooperatives. Despite this imbalance, solid implementation output progress was made in both subsectors as shown in the previous sections.

Third, the soil testing component of Food for All faced a number of issues. Due to financial constraints and change in price (from Euros 50,000 to over 100,000) Meru Greens failed to purchase and install a Lab In A Box (LIAB) from Soilcares as envisaged in the project plan. The revised soil testing approach uses 20 handheld gadgets, which required more time to reach the target number of soil tests. In addition, following the bankruptcy of Soilcares (resulting in laying off key staff spearheading the soil

testing component), progress on the soil testing strategy of the project was slowed down; and this was compounded by the anti-COVID contact restrictions in 2020.

Fourth, the F4APK training approach was revised during the project. The Training of trainers (TOT) commenced in 2017 with training of about 348 TOTs. The trainees were selected by Meru Greens from among their lead farmers. At the end of 2017, the project decided to drop the TOTs trained earlier in the year because of concerns regarding the quality of content delivered to the rest of the farmers. In addition, a request for change (RfC) on the training structure was submitted to RVO (section C in the 2017 annual report). The revised TOT program commenced in early 2018 after the approval of the RfC. For these reasons, the training program was behind schedule for about one year.

Fifth, the relationship with the Meru County government has continued to be challenging throughout the project period. The county Agriculture department expected direct funding from the project which was not possible due to the project structure. As such, unlike the other project counties, the Meru county government extension staff did not participate in the project as TOTs.

2.6 Relevance (RQ1)

Project relevance is the extent to which project design is relevant to the local context in that (a) it is consistent with government priorities and policies of the host country; and (b) it addresses the needs of end beneficiaries. We address both types of ex-ante relevance here.

2.6.1 Policy relevance

The importance of agriculture and agricultural productivity growth in the Kenyan economy are illustrated by a number of parameters (FAO Kenya, 2019; World Bank, 2018). Agriculture contributes 26 percent of Kenyan Gross Domestic Product (GDP) and accounts for some 60 percent of employment and 65 percent of exports. The poverty head-count rate in Kenya is still substantial at 35.6 percent of the population[16], with most of the poor living in rural areas. Good news is that the percentage of children below age five that is stunted[17] has decreased from 40 percent to about 24 percent. However, the Kenyan population is projected to grow to 81 million in 2039 (from 39.5 million in 2011). This means that low and stagnant agricultural productivity will result in continuing food deficits, trade deficits and the possibility of worsening nutritional outcomes. Moreover, climate change and adverse weather shocks are seen as a real downside risk for the rural economy of Kenya.

In other words, the productivity improvements targeted by F4APK are key to improving rural welfare and reducing food insecurity. Within F4APK emphasis is placed on the promotion of improved horticulture and dairy production technologies for sustainable intensification of production. The goal of the project is: “Improving the livelihoods and resilience of households, through improving access and availability of quality food and livestock products, thereby reducing food and nutrition insecurity in the Eastern region of Kenya.”

Overall, Food for All Project Kenya (F4APK) appears to be well in line with the overall development and growth priorities set out by the Kenyan Government, set out for example in its Vision 2030 documents: increasing productivity, commercialization and competitiveness of agricultural commodities and developing and managing key factors of production. In particular, F4APK is consistent with the recently released Agricultural Sector Transformation and Growth Strategy 2019-

[16] Defined as population with daily consumption expenditure below the international poverty line of US$ 1.90 in 2011 PPP.

[17] Stunted means with a height-for-age z-score that is more than two standard deviations below the median of a World Health Organization reference population.

2029 (ASTGS; Ministry of Agriculture, Livestock, Fisheries and Irrigation, 2019). The strategy prioritizes three anchors to drive a 10-year transformation: (1) increase small-scale farmer incomes; (2) increase agricultural output and value add, while raising the contribution of agro-processing; (3) increase household food resilience, reducing the number of food-insecure Kenyans.

The F4APK emphasis on the promotion of vegetable crops and dairy products, with the potential to reduce the rate of malnutrition in the Eastern region of Kenya, potentially contributes towards the strategic food security goal (3). At the same time, the support to the supply side, e.g., through adoption of improved agricultural techniques, is intended to increase productivity and incomes, with a focus on small scale farmers (goals 1 and 2). F4APK also emphasises private sector development, e.g. through supporting market sales to processors, which reflects the second goal.

Finally, F4APK has specific targets for the inclusion of social groups among its beneficiaries, including women and youth. These social groups are also prioritized in the ASTGS, which states: "The path to achieving these outcomes must address the unique challenges and opportunities for women and youth in the sector by incorporating tailored opportunities for these groups as an integral part of delivering the ASTGS".

2.6.2 Relevance for beneficiaries

In this section we use the notion of project relevance in an ex-ante sense, with respect to producers: that is, we assess whether the project design was consistent with the needs of the beneficiary farmers, based on the knowledge available before the project. In the assessment, we distinguish between the ex-ante design elements that figure in the theory of change, and the viability of these elements in practice. The assessment is based on project design, project reporting and key informant interviews.

Relevant design components included the strong and concrete market orientation; the infrastructure investment; the connection to and innovations in existing extension networks; and the ICT platform in the context of fast growing (smart) phone ownership. According to key stakeholders interviewed, the original structure of F4APK was solid and relevant for the beneficiaries. The core idea of "having a clear path to market, through MG and cooperatives … was a well thought out plan"; without this any such project was deemed likely to fail. Several respondents praised the private sector involvement as a unique feature. In addition, the infrastructure support was seen as a strong point and as the foundation for the business cases in the supported value chains. Furthermore, improving farmers' productivity through training, extension work, demonstration plots and improved inputs connects the project to a long-standing agricultural support system (in Kenya and the region), with the capacity to implement. Finally, the idea to improve the exchange of information through the ICT platform, is consistent with the widespread use of (smart) mobile phones and the importance of information for modern dairy and horticulture management. In other words, the basic elements appear to be relevant indeed.

An important point for intervention relevance for beneficiaries, mentioned by implementers and reflected in the survey data, is the difference between the producer organizations in the two targeted sectors, dairy and horticulture. At baseline, dairy farmers were member of their PO for, on average, 10 years against 3 years for horticulture farmers. This difference was mentioned in discussions with implementers, who describe the dairy sector as more "stable"; they also mentioned that in some cases, horticulture POs did not exist before and were initiated to join F4APK. Overall, the expected relevance of the project will be larger when POs are well established and stable units.

A number of further comments were made about the design, with hindsight and based on implementation experience. First, the horticulture component was dependent on one large market player. A design with a few more market players would have provided some redundancy, in case one

of the parties was unable to deliver. In this respect, the dairy experience is a positive contrast, as the project here works with five different hubs as market linkages. A second design comment relates to the large number of value chains in F4APK (milk plus seven products in horticulture), which is costly in terms of focus. One interviewee notes that it would be better to focus on one or two value chains. As a third comment, project implementers note that food security is important but that they lack the management data to monitor this in F4APK. This remark is consistent with the relatively minor role for food security and consumption in the project (design) documents.

Finally, we address the relevance of the overall focus on vegetable value chains in the targeted horticulture producer groups. Our survey data show that a fairly large share (46 percent) of the farmers in the horticulture producer groups do not grow any F4A target crops at baseline (and this share goes up over the evaluation period to 54 percent at endline). This is consistent with the problem analysis in the design document, describing lack of smallholder production of vegetables despite high export growth potential. The document suggests that improved quality and organic farming may potentially lift existing constraints for vegetable sales to export and domestic markets (page 4-5).

The document also notes a number of program risks and mitigation strategies, but these are not discussed in depth. For example, the design document notes that the "underlying pressure for change in the small holder farming systems is driven by a number of factors which include [..] recurrent droughts in the lower zones as a result of climate change." The intervention Risk assessment (section 2.8 of the design document) mentions "Long term drought affecting production under rain fed agriculture" as a risk, mitigated by "Climate smart agriculture technologies demonstrated and adopted; Irrigation, soil and water conservation methods introduced".

In terms of project relevance, drought and irrigation are points of attention, especially since irrigation investments would need to be financed by the farmers because the program does not provide (subsidies for) inputs directly. Among farmers that grow any of the F4APK target vegetables, irrigation is used in only about half of crop decisions. In half of all cases where farmers do not use irrigation, the high cost of irrigation or scarcity of water is mentioned as a constraint. These concerns are also reflected in the 2020 Annual Report, which notes (page 28): "Climate change has become the greatest challenge to sustainable agricultural production. Water for irrigation is increasingly becoming scarce in the project areas. The project trained and introduced farmers to water efficient climate smart irrigation technologies and other farming methods geared towards climate change adaptation."

2.7 Additionality (RQ2)

In this section we briefly review two additionality concepts in relation to the project subsidies provided by FDOV to F4APK.

Input additionality is a criterion that public funders use to avoid subsidizing projects that have access to private capital markets. In practice input additionality can be judged by checking the following minimum requirements: "The company cannot self-finance the project (within a reasonable time frame); it does not have the knowledge or skills to the implement the project activities alone; and/ or it is unwilling to implement the project because it perceives the costs or risks to be higher than the benefits." (Heinrich, 2014) In addition, the project cannot access commercial bank funding or advisory support of similar quality.

There appears to be a case for ex-ante input additionality. This case is made in the project document: "In an ideal situation, the income generated from the project should be able to pay for the investment, operational and other costs. The project incurs high investment and operational costs in Year 1, when the hardware's are being established. This gives a deficit of EURO 2.1 Million, against projected

expenditure of Euro 1.71 million. The cumulative cash flow breaks even in year 5. [..] It's therefore apparent that the project partners will need extra funds to manage this project. The future looks sustainable, since once the infrastructural mechanisms such as marketing, reduced post-harvest, farmers are more enlightened and they are cohesive, the project becomes sustainable."

This case becomes even more clear with the benefit of hindsight. As explained later in this report (Section 5.1.1), the net cash-flow margins of the main commercial partner (Meru Greens) were thin (1.5-2 percent) and were negative in two of the five book years. For the smaller private partners, positive surpluses are reported.

An argument against input additionality for the horticulture (French beans) component is that part of the FDOV contribution financed the training and extension network of horticulture farmers supplying Meru Greens. Meru Greens is a long-time supplier in a fast growing but competitive markets for beans and other vegetable crops. From studies of these value chains (e.g. Strohm and Hoeffler, 2006) it is clear that the growth of demand requires suppliers like Meru Greens to constantly increase their produce purchases, by adding smallholders to their supply chain. It is not clear that without the FDOV contribution, Meru Greens would not have found ways to increase their supply – as they have done successfully since the mid 1990s. Since expanding the network of suppliers inherently carries costs and risks, the FDOV subsidy will have helped Meru Greens to reach new suppliers, at reduced commercial risk.

In the absence of FDOV, it may have been possible to attract non-commercial (public) funding from another donor or subsidy program. In our survey data (Chapter 4), we find that the control group farmers are not without support: many of these farmers receive different types of support that are comparable to components offered by F4APK, including collection centres, training and information services. We therefore believe that at least some project components could have been financed from other non-commercial sources.

Development additionality is defined as the extent to which public resources contribute to changes in development-relevant results that would not have materialized without them. This is a counterfactual question, that is not easy to answer at all levels. One way to pursue the question is by looking at development priorities as reflected in project monitoring documents. An instrument for funders to focus attention of grantees on particular themes is to require monitoring data with respect to certain priorities.

We find evidence of development additionality in the project. Female participation is a development relevant priority that is mentioned throughout the project documentation and reporting. For example, the project results and M&E data explicitly mention female participation, including in training activities. This development priority appears to have affected implementation. As explained in more detail under RQ7 (see Section 5.2), the share of females (and youth) reached by the project climbed steadily, from 45 percent in 2017 to around 70 percent in 2019-20. 64 percent of all trainees in 2020 were female.

3. Evaluation and survey design

3.1 Effectiveness focus

The central questions in this study relate to the effectiveness of F4APK:

- To what extent are the projects effective in reaching their outcome and impact objectives? (RQ3 of the headline research questions);
- What are the key determinants (both internal and external to the project) for inducing or hampering the intended and unintended effects? (RQ4).

The evaluation approach and outcomes of interest were defined in the FDOV evaluation inception report for this study (see section 2.4 of the inception report). Specifically, this study uses the inception report evaluation matrix, which is reproduced as Table 8. The outcomes of interest are primarily part of the supply-side results chain and include (access to) agricultural knowledge; farm productivity, prices and sales; income and food security status.

A special feature of this study is that it relies on farm level survey data among both beneficiary (or treatment) farmers and among control farmers, at two points in time (2019 and 2021). This design allows for a comparison of changes among "treatment" and "control" farmers over the study period.[18] This design implies that the study can control for (1) macro factors that affect both beneficiaries and non-beneficiaries (such as weather) and (2) initial differences between these two groups. The study design allows for an estimate of the net contribution of F4APK on the outcomes of interest.[19]

As discussed in Chapter 2, the supply side pathway consists of a number of stages. These stages are represented by the labels in column 2 of Table 8 (outputs, short-, medium- and long-term outcomes). The stages have a logical order, in the sense that outputs and outcomes earlier in the chain need to be realized before outcomes later in the chain can be expected. For example, if training does not lead to adoption of improved techniques, then it is not likely that productivity will improve via this specific route. The analysis therefore starts by asking questions such as: did farmers participate in trainings; how did they rate the quality of the training; and did the training result in adoption of techniques and inputs. A second step in this pathway, at outcome level, is the translation of knowledge and inputs into productivity. These two steps are considered crucial determinants of the effect of agricultural intensification projects. If these stages are successfully achieved, the medium- and long-term outcomes of increased sales and income are more likely to result.

Note that the evaluation matrix and, by extension, our survey work focuses on farmer training, adoption, production, sales and income. A (detailed) consumption survey is not part of our data collection efforts. This means that we address food access only qualitatively; we do not present formal food security measures and cannot analyze questions of nutritional balance.

Results are presented in logical order along the pathways, from outputs to long-term outcomes. The two effectiveness research questions RQ3 and RQ4 are therefore closely related: the answers to both questions will be provided by tracing the outcomes along the results chain.

[18] This is a quasi-experimental evaluation design known as a "difference in differences" or DiD.

[19] DiD estimation requires the assumption that in the absence of the intervention, the change in the outcomes of interest would be the same for both groups (parallel trend assumption).

A list of output and outcome variables that the impact analysis will focus on (corresponding to column 4 of Table 8) is presented in the next chapter. The remainder of this chapter provides details of the study design.

Table 8: Inception report evaluation matrix

Research question	Impact level	Judgement criteria	Indicator
RQ 3,4,5	**Long-term outcomes**	Increased income and/or profits	Prices of dairy products and vegetables sold (KES) Farm income of targeted vegetables and dairy products (KES, %)
RQ 3,4	**Medium-term outcomes**	Increased quantity and quality of supplied vegetables, fruits and dairy products	Change in quantity and certification status of vegetables and dairy products supplied (%, kg) Change in the quantity exported (%, kg)[20]
		Increased productivity	Production (%, kg) Yield (kg/ha) Use of inputs (yes/no) Productivity (kg/ha)
	Short-term outcomes	Increased agricultural and marketing knowledge	Increased agricultural and marketing knowledge (self-reported knowledge; adoption yes/no)
		Increased access to market information	# of text messages sent
RQ 3,4	**Outputs**	Establishing vegetable centres, dairy hubs and processing factories	Access to recommended/improved crop and livestock input (availability of key inputs, yes/no)
		Improved inputs to farmers	
		Farmer trainings delivered	Number of trainings followed; rating of training quality
		ICT platform	Socioeconomic characteristics of registered farmers (demographics, education) Frequency and topic of distributed messages

3.2 Impact evaluation design

3.2.1 Treatment effect definition

The F4APK program targets Producer Organizations (POs). It registers farmers and communicates program content to them, e.g. informing them about improved crop farming techniques and marketing channels. Not every farmer member of each PO will participate in the program, but every farmer member active in the F4A value chains is eligible to participate.[21] The survey samples randomly from PO member lists, resulting in a farmer sample that includes both eligible participants and eligible non-participants.

A basic and widely used approach for impact evaluation compares outcomes for the complete PO member sample with a counterfactual. The resulting estimate is called the *Intent to Treat Effect (ITT)*: this is the effect of being made eligible for treatment. Non-participation will dilute the

[20] Export information is not available at the level of farmers or POs (who mostly function as aggregators of crops for individual farmers), and so impact on this indicator will not be estimated. Any information on exports will be descriptive.

[21] The main reason given by farmers for not participating in the training was “not aware of the program”. Less frequently mentioned reasons include: not interested and no time because of farm and non-farm business.

treatment effect in an ITT analysis, because the effect of treatment will be calculated on an eligible treatment sample, where some members have chosen not to participate.

An alternative is the *Average Treatment Effect on the Treated* (ATT), which focuses on those farmers that actually participated in the program; in particular, those that report having received F4APK training. Our study will use both ITT and ATT to estimate impact. ITT typically provides a lower impact estimate but is useful for funders as it accounts for non-participation and is possibly closer to the *Average Treatment Effect* (ATE), the effect on a randomly selected farmer.

3.2.2 Difference in differences

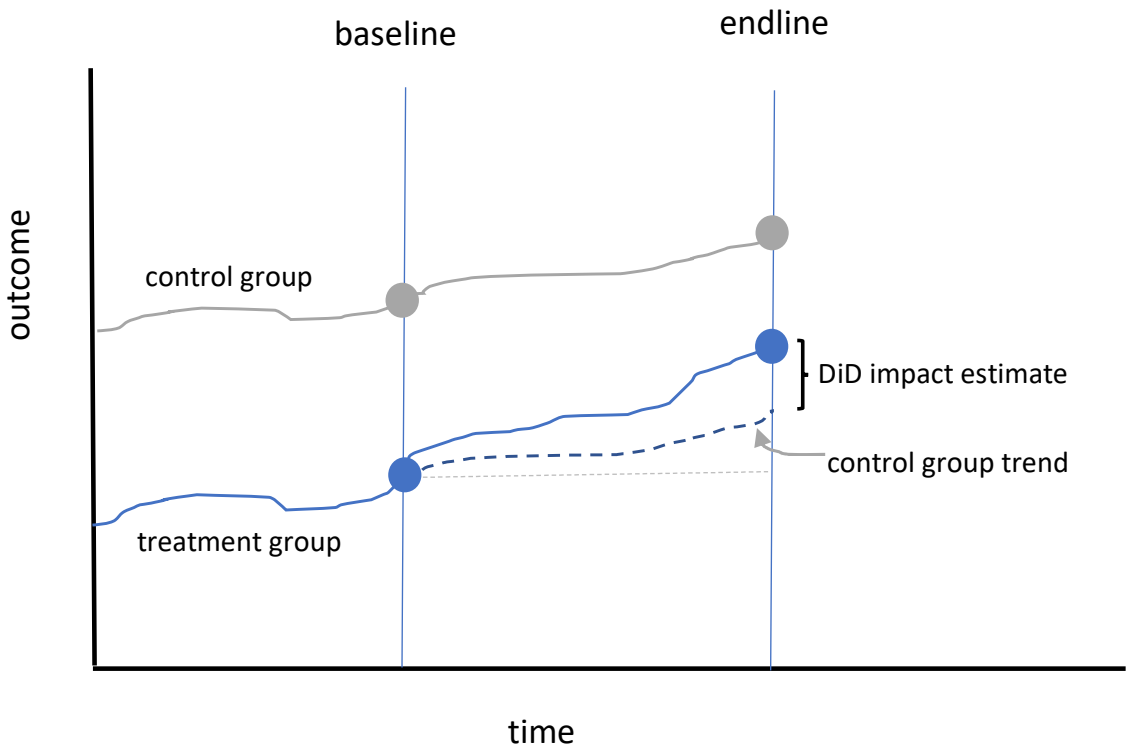
Our main tool for the quantitative analysis is *difference in differences* (DiD) estimation. This is a comparison, between treated and non-treated farm households, of changes between baseline and endline outcome measurements on the relevant indicators.

Figure 7 provides an illustration of the DiD estimation principle. A treatment group and a control or comparison group have been identified and the lines C-D and A-B represent the mean for the outcome of interest, plotted over time, for the control and treatment group respectively. In reality we can only observe the outcomes at baseline and endline for these groups, based on the survey measurements.

In this example, the outcome has a higher baseline level than the treatment group, but that is not an obstacle for DiD analysis, as long as we can assume that the trends are parallel in the absence of treatment. With that assumption, the DiD data structure allows us to identify the impact of the treatment on the outcomes of interest. In the graph, the DiD impact is the vertical difference between the treatment group endline level (B) and (E), which represents the treatment group baseline level plus the control group trend or change.[22]

Figure 7: Difference-in-differences design

[22] The vertical distance EB equals vertical distances AB minus CD, which is the double difference.



In this report we present baseline and endline means for the control and the treatment groups, so points A—D in this graph, plus the DiD impact estimate. The analysis will estimate the difference BE as depicted in Figure 7, while controlling for other covariates, e.g. household characteristics, in a multiple regression framework.

We make an important caveat. Any DiD estimate can only be interpreted as a causal estimate under the assumption that the pre-treatment trends are parallel (as depicted in the figure). This is a generic point for DiD estimates, not specific for this study. If pre-treatment (time-series) data are available for treatment and control units, it is possible to test the assumption. In our dataset, however, we do not have these data. As we cannot test the parallel trends assumption, we remain cautious and do not claim to identify causal effects.

3.2.3 Timing

A general challenge for the impact analysis is that the evaluation was commissioned relatively late in the project cycle. From a project perspective it is desirable to have most of the project activities delivered well before the end of the implementation period (2020). From a project evaluation perspective, starting an evaluation after project activities have been completed is not ideal. Generally, to determine the impact of a program it is desirable to conduct a baseline survey before major effects of the program are expected to show up in the survey measurements, that is, just before or at the same time as the roll-out of interventions. This means that effects between the baseline

As described in Chapter 2, F4APK was granted FDOV funding in 2015, completed its inception phase and started implementation in 2016. The FDOV evaluation consortium finalized its evaluation study inception report mid 2018, after which the survey preparations and contracting started. The baseline survey took place in February-March of 2019.

However, there are a number of factors that create favorable “before treatment” conditions for our baseline survey of February 2019. First, the full F4A package has not been implemented before 2019,

because of the delays in the ICT component. Full roll-out of the ICT and phone-based information interventions was expected only in 2019.

Second, the key driver of impact is enhanced production and productivity. For baseline, this is measured by production over 2018, i.e. largely precluding the impact of interventions implemented during 2018. In addition, the impact of the F4APK should be expected to come from a combination of program components: the infrastructure and improved inputs that the program supports will have an effect only if targeted farmers have been trained about it and so are able and willing to use it. In other words, one can expect considerable *complementarity* between the training components (R3) and the infrastructure component (R2). As will be illustrated in Chapter 4, we find that the reported start date of F4APK training for the farmers in our sample is on average in 2018. In other words, for this cohort of trained farmers production impacts are likely to show in harvests from 2019 onwards; and so, do not show up in the 2018 recall production measurements.

Conversations with F4APK program management confirmed that the first production effects of the intervention were not expected to be observable before quarter 1 of 2019. In addition, the question of when to implement an endline survey was discussed with F4APK management. The argument to allow at least two years for treatment effects to become observable was accepted, and so the endline survey was implemented in quarter one of 2021.

That said, as discussed with the FDOV steering committee, it is useful to distinguish between effects that are likely to manifest themselves in a relatively short time after program participation starts and those that take longer. The impact analysis uses a hierarchy in timing of effects throughout.

3.2.4 COVID-19 and impact analysis

The DiD approach can filter out "macro shocks" assuming that these affect treatment and control farmers (on average) to the same extent and do not completely disrupt production. An extreme shock, however, such as a locust attack or plant disease, can completely destroy a crop. If such a shock happens between baseline and endline, it can potentially obscure any treatment effects of the program. It is hard to say whether COVID-19 represents a more balanced shock or an extreme shock. From the implementers we do know that F4APK (in-person) implementation was disrupted by COVID-19, but this will also have affected extension work in the control group.

When designing the endline survey, the evaluation team decided to include longer recall periods for a number of questions. The idea was to enable a "pre-COVID" impact analysis based on a limited set of outcomes for the period before March 2020. We will discuss these estimates in the main presentation of results in Chapter 4. Note, however, that these estimates are based on long recall for the main crop season only. This means that in some cases (French beans) we have to work with relatively small samples.

3.3 Survey design

The survey data are collected from a sample of farmers. The selection of farmers is essential and followed procedures regarding sample selection, field organization and quality control, questionnaire design that are described in the literature.[23] We describe the main decisions and approaches.

[23] See for example Bamberger et al., 2012; UN Statistical Division, 2005.

3.3.1 Sample design

3.3.1.1 Treatment group listing

F4APK targets farmer groups in five counties. These counties are, ordered from North to South: Meru, Tharaka Nithi, Embu, Machakos, and Makueni. These counties form part of the lower Eastern region and lie on the southern side of Mt Kenya (three counties have part of the mountain within their boundary). The project areas are in the lower midland zones in these counties.

As a first step in the survey sampling and to reduce field costs, a selection of three counties for survey field work was done. To do this, the project management (Solidaridad) provided data on farmer group project registration from the start of the project for each of the counties, by value chain. Based on these data, three counties were selected for survey work: Meru, Embu and Machakos. Criteria for county selection included, first, representing a reasonable share of the total project population, preferably in both value chains. At the time of the county selection, the three selected counties jointly represented 70 and 79 percent of farmers on the registration lists in Dairy and Horticulture, respectively. Secondly, the selection provides geographical variation.

As a second step, the survey identified the population of farmer groups to select farmers for interviews. For the treatment group, a list of POs, their membership and PO sizes was generated during a pre-baseline visit. This list was compiled at county level with the purpose of selecting POs with limited exposure to the program. A list of F4APK beneficiary PO names, location data (and product specialization for horticulture groups) and membership size was supplied to the survey team by Solidaridad County Coordinators and the TOT. These lists do not reflect the complete population of farmers reached and registered by F4APK, but a sub-set of POs where the chance of surveying recently reached and trained farmers was considered high (see Section 3.2.3 on timing), so that exposure to the program for the sampling population was relatively short.[24]

3.3.1.2 Control group listing

The general requirements for control group members are (1) not being affected by the program and (2) providing a counterfactual to the treatment group; that is, having expected outcomes equal to the treatment group in the absence of the intervention. It was not possible to generate an experimental control group for this study, nor one based on program assignment rules (ruling out randomized assignment or a regression discontinuity design). Therefore, the control group had to be constructed based on practical arguments, as follows.

For horticulture, Meru Greens was asked for a list of PO groups outside the F4APK program in the selected counties. In Meru and Embu, all horticulture POs are targeted by F4APK and so are not suitable as a control group. In Machakos, only part of the county is targeted and so a list of the non-target POs was obtained for control sampling. To ensure sufficient numbers, POs were also sampled from PO lists from wards in Kiambu and Murang'a, two neighbouring counties. The POs were selected from Wards adjacent to the treatment counties Embu and Machakos, with very similar altitude and climate characteristics. For dairy, the program partner Heifer generated a list of groups from two dairy hubs outside the F4APK program. One of these control hubs is in Tharaka Nithi county, one in Embu.

3.3.1.3 Sample selection

Once the county PO lists were established, production details were checked. In the horticulture list, POs that mentioned specialisation in non-F4APK products (e.g. mangoes) were removed. After this

[24] Note that the PO lists used for sampling purposes are a part of the total registration lists for 2018. F4APK reports that a total of 41,232 farmers were reached and trained in 2018 (2018 progress report, result 2.1.2, page 4).

check, the population lists contained a total of 545 POs and 15,849 farmers. PO lists were then used to conduct a two-stage sampling procedure within each county for both treatment and control POs. First, a sample of POs was selected randomly from a list with the population of POs, with probability of selection proportional to the size (members) of the PO.[25] Second, a sample of six farmers was selected randomly for an interview from the list of all farmer members of the selected PO.

A minor challenge was replacement of POs. Reasons for POs dropping out of the sample include: the PO could not be reached, or was disbanded, or was listed twice but under different names, or had other (non-relevant) activities. Overall, these issues were more important in the horticulture value chain. The share of POs dropped and replaced under horticulture is 13 percent of the total number surveyed, and 4 percent for dairy (see Table 9).

Table 9: Population and Sample

	Horticulture	**Dairy**	**Total**
	Treatment and Control combined		
Population (listed)			
Number of POs	420	125	545
Number of farmers	13,562	2,287	15,849
Average size (members)	32	18	29
Sample (surveyed)			
Number of POs	158	68	226
Number of farmers	848	389	1,237
Number of POs dropped	20 (13%)	3 (4%)	23 (10%)

The baseline sample contained 1237 farmers, representing 7.8 percent of the total population of farmers on the sampling frame lists. The target (sampling) population is larger for the horticulture value chain than for dairy: dairy farmers make up 14 percent of the farmer sampling population. The sampling proportions reflect this to some extent, with dairy farmers representing 31 percent of the total sample.

During the evaluation endline phase, the survey teams revisited the same POs and farmers interviewed during the baseline visit in order to create a farmer panel. Details of the resulting full sample are presented in Table 10.

The endline sample has a size of 1115 farmers, meaning that the teams successfully tracking and re-interviewing 90 percent of the baseline sample. Respondent tracking was slightly more successful in the treatment clusters and among dairy farmers (compared to control and horticulture, respectively).

Table 10: Sample composition (number of farmers)

	Horticulture		**Dairy**		**All**	
	Baseline	Endline	Baseline	Endline	Baseline	Endline
Control	313	265	150	142	463	407
Treatment	535	488	239	220	774	708
Total	848	753	389	362	**1237**	**1115**

[25] In other words, the POs are used as primary sampling units.

The distinction between a “dairy” and a “horticulture” farmer in the study is made based on the PO they belong to. The POs are targeted by F4APK to become part of the dairy or horticulture part of the program, typically based on lists that originate from the project partners Meru Greens for horticulture and Heifer for dairy. Nearly all farmers are active to some extent in both crop farming and dairy, but they tend to specialize and their PO membership reflects this. In a sense the program accentuates their value chain specialization, since the farmers are offered a horticulture or dairy training depending on the PO they belong to. However, all farmers are asked questions about both dairy and horticulture production. Throughout the reports we refer to these two groups as dairy and horticulture farmers (or producer organizations).

A final filter on the farmer sample selection was done at the start of the interview. If a farmer indicated not being active in any of the F4APK value chains (dairy or at least one of the seven horticulture crops), the interview was ended and the farmer was not included in the sample.

3.3.2 Extrapolation

The evaluation team was asked about extrapolating the sample estimates to the target population. Given the timing and design of the sample, a number of assumptions are required to do this. We describe here what extrapolation involves, and provide a few illustrative examples. In the report we only present sample estimates.

The treatment sample selection was based on a population of POs with relatively limited exposure to the program in 2018. A random sample of POs was selected from this population and *members* (potential beneficiaries) were selected randomly in each of these sample POs. Our estimates therefore typically give the fraction of the full population of PO members who report they received a certain benefit (e.g. were trained, or had their soil tested), used a certain technique, or used a particular marketing channel. At endline, and for the treatment group, these estimates are for farmers who have been in contact with the program since at least 2018.

One way to extrapolate these estimates to the total population of farmers supported by F4APK is to multiply the estimate (fraction) with the total member population in all POs treated since the start of the project. The main (strong) assumption we need here is that our estimates for the “2018 cohort” apply to all beneficiaries, in all project counties and irrespective of when they started. This assumption implies that treatment intensity was constant over the lifetime of the project; and that results are sustained for early beneficiaries.

For the extrapolation we also need PO population numbers, by value chain because our estimates are mostly separate for dairy and horticulture POs. We need to calculate these, because we do not have a full count of POs and PO members targeted by the project. We start with the number 70,645 farmers trained over the life of the project (annual report 2020). Based on our training participation rates, we estimate the number of members in horticulture POs in the target treatment population at 114,234 (47 percent trained at baseline) and in dairy POs at 19,267 (88 percent trained). These numbers reflect the much larger number of horticulture POs in our sampling frame and their larger membership.[26]

[26] Our estimates are specific to PO type (dairy, horticulture) but the AR does not report numbers trained by PO type. We therefore estimate the dairy-horticulture split using the sampling population data we received (Section 3.3.1.3). According to these data, dairy POs represent 14 percent of farmer population and horticulture POs 86 percent. At baseline, 88 percent of dairy POs participated in F4A training and 47 percent of horticulture POs. Putting these numbers together, we estimate cumulative trainee numbers at 53,690 for horticulture and 16,955 for dairy. Using the training participation we calculate the target population from these numbers.

We now extrapolate our estimates to the full target population, under the assumption that our sample estimates apply to the full population. For example, our adoption estimates show that in horticulture treatment POs, 52 percent or 59,402 treatment PO farmers used irrigation at baseline (but this number did not increase). Similarly, we calculate that 7,514 dairy farmers used fodder at baseline. This share increased by 10 percentage points (1927 farmers) between baseline and endline.

3.3.3 Questionnaire design

The study used two farmer questionnaire types, one with a focus on dairy and one with a focus on horticulture. The choice of the questionnaire depended on the PO membership and program offered as explained above. Each of the questionnaires has ten sections (plus the filter section), covering the following areas: demographics, including the household roster; housing and assets; land usage; production and sales; expenditures on production inputs; group membership and trainings; information access and use; other income sources; food consumption and shocks; subjective well-being.[27] At endline, a sub-section on COVID-19 was added.

The questionnaires are part based on those used in impact evaluations of comparable agriculture focused programs; and part tailored to the specific components of F4APK. The questionnaire sections and individual questions were created to cover the outcome indicators in the ToC, particularly the measures in Table 8.

AIGHD provided the draft instruments for the survey. As the enumerator recruitment was going on, the survey team went through the instrument and proposed changes, based on the framing of the questions, and improving the local context for the survey questions. The team also clarified some of the responses. This process was done through an iterative and interactive process with AIGHD.

The questionnaires for both the dairy and horticulture value chains were adapted to collect information on household demographics, dwelling characteristics, land and wealth endowment, income generation, crop and dairy production practices, such as techniques, input use, and level of production, food consumption and expenditures, hunger-scale and subjective well-being. Before the commencement of the training, the questionnaire was digitized to use SurveyCTO, which was the software used for data collection in the survey.

3.3.4 Survey preparation and implementation

Both the baseline and the endline survey implementation was carried out by the Tegemeo Institute, Nairobi.[28] In preparation of the survey, Tegemeo undertook key activities which included, recruitment and training of supervisors and enumerators, finalizing the questionnaire in collaboration with AIGHD, digitizing the questionnaire and logistical planning for the surveys.

For baseline, the training of enumerators and supervisors was undertaken for seven days from January 28th to 5th February 2019 in Nairobi. The teams travelled to the field on 12th February 2019, and data collection took place from 13th February to 18th March when the teams travelled back to Nairobi. From the target of 1,338 households, 1,165 households were successfully interviewed. This represents a response rate of 87% for the total sample. To increase the response rate a team was sent back to

[27] Since the questionnaires are large, we have not included them in this report. The questionnaires are part of the dataset and are submitted to the Netherlands Enterprise Agency by the evaluation team.

[28] Tegemeo Institute of Agricultural Policy and Development, a policy research institute under the Division of Research and Extension of Egerton University. See www.tegemeo.org for details.

the field from 7th to 14th April to undertake the additional interviews. Twelve groups were interviewed, six treatment and six control groups. This exercise brought the final achieved baseline survey sample to 1237 farmers.

For endline, the training of enumerators and supervisors was undertaken for seven days from 1st to 10th February 2021 in Nairobi; followed by a pretest from the 15th to 18th February 2021 in Makueni County. The teams travelled to the field on 21st February 2021, and data collection took place from 22nd February to 24th March when the teams travelled back to Nairobi. From the target of 1,237 households, 1,115 households were successfully interviewed. This represents a response rate of 90% for the total sample.

During the endline survey implementation, all the regulations instituted by the Kenyan government aimed at reducing transmission of the COVID-19 virus were observed.

Figure 8: Location of dairy groups in sample

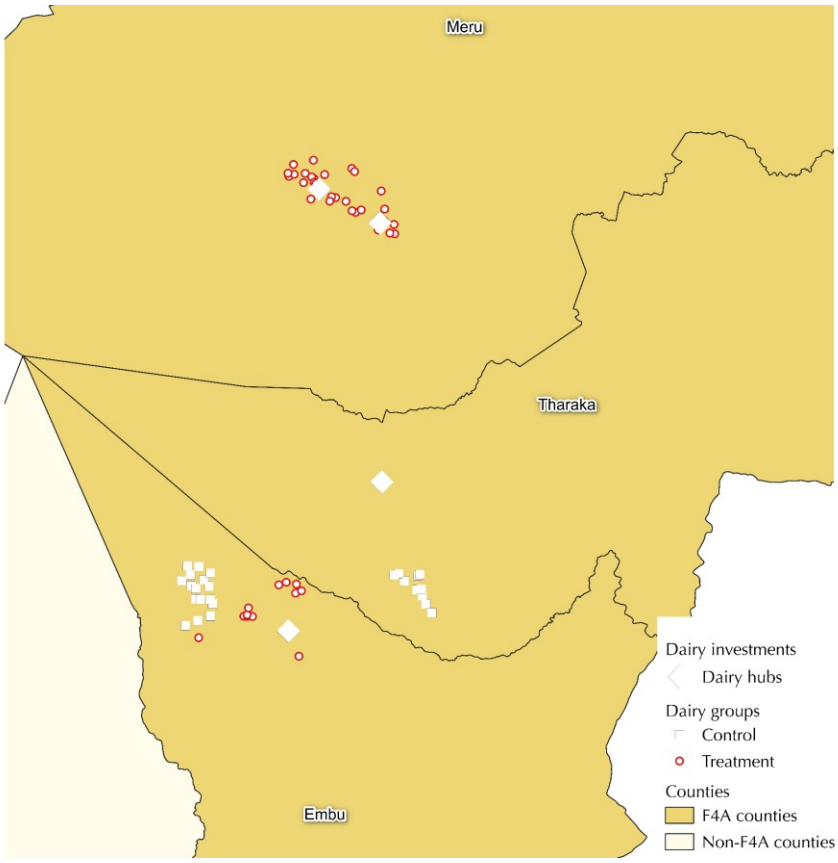
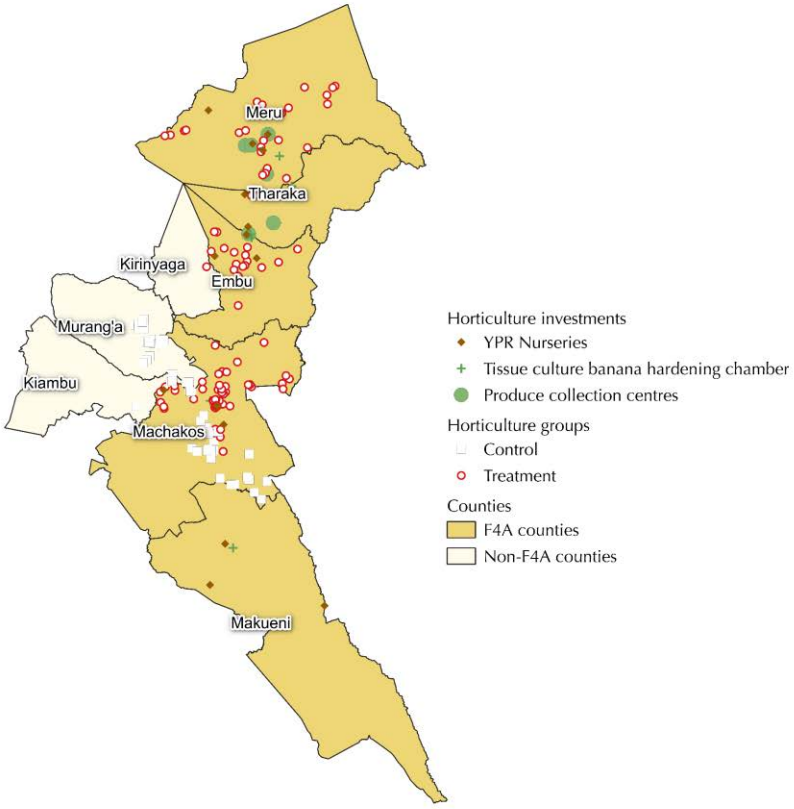


Figure 9: Location of horticulture groups in sample [29]



[29] There are 102 Horticulture treatment groups; 4 groups have missing GPS data.

4. Effectiveness (RQ 3-4)

4.1 Context: baseline sample characteristics

4.1.1 General characteristics

Table 11 gives average values for key farm household characteristics at *baseline*, by horticulture and dairy producer groups. Here and in all following tables we denote significance of differences using the following significance star symbols: * p<0.10 ; ** p<0.05 ; *** p < 0.01.

Table 11: Household characteristics at Baseline

	Horticulture			Dairy		
	Control	Treatment	T-C	Control	Treatment	T-C
Numbers						
Number of household members	4.74	4.98	0.24	4.35	4.21	-0.15
Number of children in household	1.67	1.91	0.24**	1.53	1.33	-0.19
Land owned (acres)	1.93	2.16	0.23	1.48	1.21	-0.26
Land available (acres)	2.15	2.49	0.34*	1.78	1.54	-0.24
Land used (acres)	3.53	3.47	-0.07	1.79	1.97	0.18
Years as a member of cooperative	4.58	2.38	-2.19***	5.97	13.00	7.04***
Fraction of households where:						
Household head is female	0.18	0.22	0.04	0.12	0.15	0.03
Household head completed O-level	0.36	0.27	-0.09**	0.33	0.36	0.03
Uses electricity for lighting	0.82	0.76	-0.06*	0.83	0.89	0.06
Has a private tap	0.12	0.33	0.21***	0.83	0.90	0.07
Owns fridge	0.06	0.02	-0.04**	0.04	0.03	-0.01
Owns smartphone	0.33	0.40	0.07*	0.49	0.60	0.12**
In the last year, at times the household did not have enough food	0.28	0.35	0.06*	0.22	0.18	-0.04
Experience drought in the past 12 months	0.18	0.23	0.05*	0.03	0.08	0.04*
Current food consumption is not enough	0.30	0.33	0.03	0.15	0.21	0.06
Life satisfaction	0.18	0.17	-0.01	0.27	0.27	0
Does irrigation	0.43	0.38	-0.06	0.11	0.59	0.47***
N	313	535	848	150	239	389

The average household has between four and five members, close to two children and owns about 1 to 2 acres of land. One fifth of households has a female household head: the share of female headed households is slightly higher in F4A POs but the difference is not significant. One-third of household heads have high school (O-levels) as highest education level (only 5 percent have no education at all).

The data show that dairy farmers on the whole have higher socio-economic indicators: on average they have fewer children, are slightly more likely to use electricity for lighting, much more likely to use a private water tap, more likely to own. That said, they own and use smaller land areas.

There are a few significant differences between treatment and control means, in both value chains. This is not unexpected given the quasi-experimental evaluation and sample design. For dairy these differences "favour" the treatment group, which has higher private tap use and smart phone ownership. For horticulture, the differences are more mixed, e.g. higher private tap use but lower electricity use for the treatment group. There are no treatment-control differences in acreage. In our regression analysis we control for a range of covariates, including those that are significantly different between treatment and control groups in Table 11.

One characteristic that deserves attention is the duration of producer group (cooperative) membership (row 6 of the table). Overall, dairy farmers have been member of their cooperative longer than horticulture farmers. Moreover, within the sectors it is clear that membership duration differs significantly between treatment and control groups. Of the four subgroups in Table 11, horticulture treatment farmers have the lowest membership duration. This suggests that these producer organizations are younger and have had less time to develop than the POs in the other three groups. In contrast, the dairy treatment farmers have the highest mean membership duration across these subgroups.

4.1.2 Importance of F4APK value chains: targeting and relevance

The F4APK program targets producer groups in two value chains, horticulture and dairy. In dairy, the key income generating output is milk. Within horticulture, the program targets seven marketable crops: French beans, banana, watermelon, butternut, onion, tomato, and capsicum.

Table 12 presents value chain engagement outcomes: it gives the baseline and endline means for the control and treatment groups. For each sector (horticulture and dairy) and for each variable, ten numbers are provided: the *means* for control and treatment, for baseline and endline (4 numbers); below each mean is the *number of observations* (again 4 numbers); and the *difference-in-differences (DiD) estimate*, with the standard error estimate below the DiD mean (2 numbers). All following tables have this general structure. In many cases fractions are presented: for example, the number in the upper-left cell in Table 12 implies that, at baseline, 61 percent of the 313 horticulture PO farmers in the control group cultivate any of the seven F4A target crops.

In Table 12, we provide two panels: the upper panel has outcomes that are especially relevant for horticulture farmers, the bottom panel has dairy outcomes. Estimates for horticulture farmers (POs) are in columns 1-3, estimates for dairy are in columns 4-6. Based on the program targeting, we would expect most of the "action" in the upper panel to be in the Horticulture (left hand side) columns, and in the lower panel in the Dairy (right hand side) columns.

Table 12: Value chain engagement

		Horticulture			Dairy		
		Baseline	**Endline**	**Diff-in-diff**	**Baseline**	**Endline**	**Diff-in-diff**
Panel A. Horticulture							
Cultivates any F4A crops (fraction)	Control	0.61	0.59		0.71	0.76	
	N	313	273		150	142	
	Treatment	0.54	0.46	-0.07**	0.72	0.71	-0.06

	N	535	480	(0.03)	239	220	(0.05)
F4A crops cultivated (number)	Control	5.16	2.41		1.41	2.76	
	N	190	160		107	108	
	Treatment	4.15	2.10	0.67**	1.49	2.78	-0.01
	N	290	219	(0.3)	172	156	(0.11)
F4A crop is among 2 most important crops in terms of value (fraction)	Control	0.26	0.26				
	N	313	273				
	Treatment	0.24	0.19	-0.05			
	N	535	480	(0.03)			
Acreage for F4A crops (fraction of total acreage)	Control	0.31	0.16		0.07	0.12	
	N	308	273		150	141	
	Treatment	0.22	0.11	0.04	0.08	0.13	0
	N	527	478	(0.03)	238	220	(0.02)
Panel B. Dairy							
Owns cows (fraction)	Control	0.76	0.75		1.00	0.92	
	N	313	273		150	142	
	Treatment	0.72	0.70	0.02	0.99	0.99	0.09***
	N	535	480	(0.03)	239	220	(0.03)
Does dairy (fraction)	Control	0.66	0.66		0.93	0.95	
	N	313	273		150	142	
	Treatment	0.53	0.55	0.03	0.95	0.97	0.01
	N	535	478	(0.04)	239	220	(0.03)
Acreage used for dairy (fraction)	Control	0.00	0.00		0.24	0.23	
	N	308	273		150	141	
	Treatment	0.01	0.01	0.00	0.22	0.35	0.14***
	N	527	478	(0)	238	220	(0.03)

F4APK target crop engagement decreases in the horticulture treatment group. For three of the four horticulture engagement variables, we see a decrease in the treatment group: the share growing any F4A crop, importance in terms of sales value and acreage devoted. F4A crop cultivation decreases among horticulture farmers in the treatment group,[30] from 54 to 46 percent of farmer growing any of these crops. The engagement share remains constant at about 60 percent in the control group, and thus the DiD estimate is negative. Second, the acreage share used for F4A crops is reduced over time for both control and treatment farmers. Third, the share of treatment farmers for whom an F4A crop is among their top two most important in terms of sales decreases. Finally, the average number of F4A crops grown (by farmers that engage in any F4A crops) decreases, both in the treatment and in the control group. In the treatment group the decrease is smaller and therefore we find a positive DiD estimate.

As observed at baseline, the share of farmers cultivating any of the seven F4APK target crops is significantly lower in the F4APK horticulture POs than in dairy POs (a finding that is consistent with the lower share using irrigation, see Table 11) and this difference grows larger over time. There is

[30] Among horticulture farmers that grow any F4APK focus crop at baseline, the F4APK crop grown by most farmers (more than 70 percent) is bananas. French beans are reported as being cultivated by 9 percent of these farmers in F4APK horticulture treatment POs (against 34 percent in control).

relative "despecialization" among horticulture farmers in the sense that the share of their land devoted to F4A horticulture crops was slightly larger than for dairy farmers at baseline, but has become smaller at endline.

Dairy farmers have increased their already high level of dairy specialization. Table 12 shows that, at baseline, farmers in the Dairy component of F4APK were already more specialized in milk production, the program target output. 99 percent of dairy treatment PO farmers own cows and 97 percent are engaged in dairy production at endline.[31] Interestingly, 53 percent of farmers in the horticulture program engage in dairy production at baseline, and 55 percent at endline. In other words, the horticulture farmers became slightly more engaged in dairy over time, while on average becoming less engaged in horticulture.

At baseline, F4APK dairy treatment farmers use 22 percent of their production acreage for dairy production on average, and this share increased to 35 percent at endline. Since this outcome remained practically constant in the control group, we see a strong positive and significant DiD estimate of +14 percentage points. This is a first indication of the success of the dairy component of F4APK. This result is consistent with the significant positive DiD estimate for cow ownership.

The respective value chain trends in engagement reported here are confirmed in discussions with programme management and implementers. The discussion highlighted the difference between the two sectors, including the difference in implementation. Selection of and outreach to horticulture farmers was done by the private sector partner (Meru Greens), and farmers became demotivated when payment commitments were not met because of financial difficulties. While the project was allowed to bring in other private partners, this turned out to be difficult in practice. The incentives to keep farmers loyal to the PO are stronger in dairy: there is a stable market price and the PO provides milk quality control and cooling, services that are helpful for marketing.

4.2 Outputs

4.2.1 Hardware/Organizations

Table 13 shows the percentage of farmers that live within 500 meters of the collection centre where they normally deliver their milk or crops. These data show that on average, produce aggregation and collection for dairy farmers were at baseline closer to home than for horticulture farmers. This is consistent with the fact that local collection centres are well established for dairy farmers.[32]

Table 13 also shows that, at baseline, slightly fewer treatment farmers than control farmers lived in close proximity (within 500 meter) of a collection centre, for both dairy and horticulture.

For horticulture farmers, the share with close collection centres decreased for both control and treatment farmers, but much more so for control farmers, leading to a positive DiD estimate. The annual reports show that none of the newly installed F4APK collection centres were functional in 2020, which may be reflected in the treatment mean for endline, but not in the control mean. Another possible explanation for the overall reduction, mentioned in stakeholder interviews, are COVID-19 restrictions. Note, however, that we do not see this reduction for the dairy control farmers.

[31] Defined as growing fodder, selling milk, selling cows or a combination of these activities.

[32] In 4 of the dairy cooperatives, the farmers bring milk to a local collection center from where a porter collects and transports the milk to the chilling plant (hub). It is only in Makueni that dairy cooperative where farmers deliver their milk directly to the chilling plant, in addition to other satellite collection centers.

Table 13: Distance to collection center

		Horticulture			Dairy		
		Baseline	Endline	Diff-in-diff	Baseline	Endline	Diff-in-diff
Collection / Distribution center within 500 meters of farmer	Control	0.19	0.03		0.63	0.65	
	N	313	273		150	142	
	Treatment	0.15	0.10	0.12***	0.58	0.50	-0.09
	N	535	480	(0.04)	239	220	(0.07)

4.2.2 Trainings

Table 14 provides a summary of survey data on training participation, by value chain and comparing treatment and control.

Intensity of F4APK training for horticulture has reduced, and remained at a high level for dairy. For horticulture treatment farmers in our sample, we observe a drop in the share that received training in the last year, from a 59 percent baseline training participation to 14 percent at endline. We observe a similar drop in the percentage aware of F4A training; and the percentage having received F4A training. These data are consistent with training dates recorded at baseline showing a majority of the horticulture treatment sample was trained by F4APK in 2018-19; and with much lower training intensity for this group at endline, at a level equal to the (horticulture) control group. The table further shows that other (non-F4APK) training for horticulturalists remained constant for the treatment group but, as training participation decreased for the control group, the relative importance of other trainings grew over the study period.

The dairy (right hand) side of the table shows that a) dairy training intensity is at a much higher level; and b) dairy training intensity does not decrease over the evaluation period; and c) dairy treatment farmers are more likely to receive training than the dairy control farmers.

Table 14: F4A training

		Horticulture			Dairy		
		Baseline	Endline	Diff-in-diff	Baseline	Endline	Diff-in-diff
Panel A. Training							
Received training	Control	0.31	0.14		0.65	0.65	
	N	313	273		150	142	
	Treatment	0.59	0.14	-0.29***	0.93	0.93	0.01
	N	535	480	(0.05)	240	221	(0.01)
Is aware of F4A program	Control	0.01	0.00		0.09	0.11	
	N	313	273		150	142	
	Treatment	0.60	0.42	-0.18***	0.95	0.93	-0.03
	N	535	480	(0.04)	240	221	(0.02)
Received F4A Training	Control	0.01	0.00		0.09	0.11	
	N	313	273		150	142	
	Treatment	0.47	0.04	-0.43***	0.88	0.87	-0.02
	N	535	480	(0.04)	240	221	(0.02)
Received other training	Control	0.29	0.14		0.55	0.54	
	N	313	273		150	142	
	Treatment	0.10	0.10	0.15***	0.06	0.07	0.03
	N	535	480	(0.04)	240	221	(0.02)

Female head: received training	Control	0.00	0.00		0.06	0.06	
	N	57	49		18	17	
	Treatment	0.51	0.55	0.08**	0.89	0.90	0
	N	120	108	(0.03)	36	31	(0.04)

Finally, we calculated the "Received training" variable again for the subgroup of households with a female head; see the last four rows of the table. We note that there are more female headed households in the treatment samples in both sectors (higher N); and that the training participation share in this subgroup is far higher in the treatment groups. Moreover, the participation in this subgroup does not decrease during the study period. This suggests that F4APK was successful in recruiting participants in female headed households and retaining these until the end of the project. These findings offer further support for our conclusions on female participation as an aspect of development additionality (see Section 2.7).

These are important results. They confirm – for a randomly selected sample in both value chains – that training outputs (results area 3) were delivered and, at baseline, dramatically improved the "training status" of the beneficiary group, compared to the controls. They also show that trainings are sustained at a high level in the dairy POs. Overall, training intensity was lower in 2020, possibly due to COVID restrictions, but certainly not zero. When farmers were asked why they did not participate in F4APK trainings at all, the most frequent answer is "Not offered, not aware of the program" and this answer is given by more farmers at endline (31 percent) than at baseline (23 percent).

On average, farmers have a high opinion of the **quality of the trainings** they participated in. Quality is measured using Likert scales, ranging from 1 (indicating "poor" or "very unhelpful") to 5 ("Excellent" or "very helpful"). Most of the average ratings are 4 ("good") or higher (data not shown in the table). In particular, overall quality of the training and trainer quality score higher than 4 on average, for both the F4APK and "other" training. Overall, ratings are slightly higher for dairy. For example, on the question "How helpful was the (Food 4 All) training to your farming practices (so far)?" the mean baseline score is 3.85 (between "somewhat helpful" and "very helpful") for horticulture; and 4.23 (between "very helpful" and "extremely helpful") for dairy.

As a measure of the direct impact of the training on all eligible farmers we use the share that are able to reproduce the training topics at baseline (when most farmers were trained). This measure is determined by the quantity of farmers trained and the how well the farmers are able to recall specific topics. For horticulture POs, we find that the share of farmers mentioning particular training topics is largely equal between treatment and controls. Two topics, crop agronomy and post-harvest management, are mentioned by a significantly larger share of farmers in the F4APK treatment group, suggesting a larger group trained and/or a more intense knowledge transfer in the project training activities. We see this difference more strongly for dairy POs, where nearly all training topics are mentioned by a significantly larger share of F4APK treatment farmers. Table 15 confirms the picture from the previous table: overall, F4APK farmers were trained more intensively than control group farmers; in addition, there was more intense training engagement among F4APK dairy farmers, compared with horticulture farmers.

Table 15: Training topics

	Horticulture	Dairy

	Control	Treatment	T-C	Control	Treatment	T-C
Share who report receiving training on…						
Soil nutrition	0.93	0.85	-0.08	0.06	0.16	0.1***
Crop agronomy	0.36	0.49	0.13*	0.25	0.42	0.17***
Dairy husbandry	0.08	0.09	0.01	0.98	0.97	-0.01
Group dynamics	0.08	0.10	0.01	0.03	0.26	0.23***
Human nutrition	0.09	0.08	-0.01	0.07	0.20	0.13***
Record keeping	0.11	0.08	-0.03	0.13	0.42	0.29***
Financial literacy	0.05	0.03	-0.02	0.06	0.26	0.2***
Post-harvest management	0.03	0.08	0.05**	0.03	0.11	0.08***
Marketing	0.08	0.07	-0.01	0.08	0.16	0.08**
N	98	326	424	100	231	331

4.2.3 Information

Table 16 shows information on the soil testing component for the farmers in horticulture POs. We find that in our sample, 4 percent of farmers report having had their soil tested in the last 12 months before the interview. At endline, there is no difference between treatment and control farmers. In the treatment group, this share was 5 percent at baseline and so there is no significant diff-in-diff estimate for this component.

Table 16: Soil tests

		Horticulture		
		Baseline	Endline	Diff-in-diff
Test received	Control	0.04	0.04	
	N	313	273	
	Treatment	0.05	0.04	-0.01
	N	535	480	(0.02)
Test received for free	Control	0.45	0.45	
	N	11	11	
	Treatment	0.56	0.79	0.29
	N	25	19	(0.29)
Changed farming practices after test	Control	0.64	0.73	
	N	11	11	
	Treatment	0.48	0.53	0.05
	N	25	19	(0.28)

How helpful did you find the test?	Control	3.18	3.64	
	N	11	11	
	Treatment	3.04	3.21	0.2
	N	25	19	(0.72)
How likely is it that you would recommend another farmer to do a soil test?	Control	4.00	3.64	
	N	11	11	
	Treatment	3.76	3.37	0.28
	N	25	19	(0.7)

79 percent of F4APK treatment farmers report that the test was free at endline, against 45 percent in the control group. The rating of the information that was received is mixed. About half of the horticulture farmers says they changed their production practice as a result of the soil test information, against 73 percent in the control group. Treatment farmers ratings of the soil test are close to 3 on average (information is "somewhat helpful"); and are between "somewhat likely" and "very likely" to recommend the soil test to other farmers. The ratings are slightly higher in the control group (but note that the respondent base is very small).

Reach of the ICT platform has grown since baseline but quality rating of the messages is medium. Table 17 provides information about the ICT platform from the perspective of farmers. There is clear evidence that the quantitative reach of the platform increased considerably over the study period. For all F4APK treatment farmers, in both horticulture and dairy POs, the share of farmer that receive production related information on their mobile phone almost doubled between baseline and endline, reaching 37 percent of treatment farmers at endline. Over the same period, the share of control famers receiving messages increased by similar percentage points so that the DiD estimate for this output is close to zero. Dairy farmers under F4APK are more likely to have received the phone messages than in the control group, suggesting that in this value chain the program was more active compared with similar initiatives rolled out in the control group.

The share of farmers saying they changed their farming practices because of the messages is 50 percent or higher, depending on the survey and subsector. Taken at face value, this makes the ICT platform a potentially powerful source of information and behavioral change. However, the mean quality rating of the messages is lukewarm. When asked how helpful the messages are and how likely it is they would recommend the service to another farmer, scores are close to 3 ("Somewhat helpful" and "Somewhat likely"), across survey rounds and subsectors. We do not find any significant DiD estimates.

Table 17: Phone messages

		Horticulture			Dairy		
		Baseline	Endline	Diff-in-diff	Baseline	Endline	Diff-in-diff
Received phone messages	Control	0.18	0.41		0.08	0.25	
	N	313	273		150	142	
	Treatment	0.18	0.37	-0.03	0.20	0.37	0.01
	N	535	480	(0.04)	239	220	(0.05)
Changed farming practices because of messages	Control	0.57	0.73		0.50	0.50	
	N	56	111		12	36	
	Treatment	0.53	0.63	-0.05	0.60	0.51	-0.09
	N	98	178	(0.12)	48	82	(0.19)

How helpful do you find the messages?	Control	3.18	3.36		3.08	3.08	
	N	56	111		12	36	
	Treatment	3.16	3.17	-0.14	3.21	3.16	-0.16
	N	98	178	(0.24)	48	82	(0.51)
How likely is it that you would recommend another farmer to use this information service?	Control	3.54	3.46		3.42	3.39	
	N	56	111		12	36	
	Treatment	3.46	3.38	0	3.42	3.21	-0.22
	N	98	178	(0.24)	48	82	(0.5)

Summing up, the ICT platforms have developed quite strongly since the baseline survey, as measured by their outreach to farmers. The survey evidence for this component is consistent with the annual reporting (see Section 2.5.3). The farmer responses show that farm soil testing in the POs in our sample is not widespread and only a small percentage of the target population has had their farm soil tested.

4.3 Short-term outcomes: input use

A key question for the impact analysis of an agricultural program is to establish whether the investments and trainings translate into increased use of improved inputs and techniques by beneficiary farmers (compared with non-beneficiary farmers). In the tables below, we analyze this using the standard ITT estimator, that compares farmers in treatment POs with farmers in control POs, irrespective of whether they were trained or not. In effect, this estimator picks up all changes within the PO, including spill-over effects on farmers that did not participate in the trainings but may have heard about it from their neighbors.[33] To do a more direct analysis of the effects of training on input use, we also used an alternative treatment indicator that puts only trained farmers in the treatment group. However, this alternative definition does not change the overall results.

4.3.1 Horticulture

Table 18 shows the percentage of crop planting decisions for which the indicated production input or technique (over the last 12 months) was selected, conditional on the farmer cultivating that crop. The inputs are relevant in the context of the F4APK training curriculum. The table has two parts: the left-hand side (columns 1-3) describes decisions used for the seven F4APK target crops; the right-hand side describes decisions for other crops.

Use of inputs and techniques on F4APK crops among horticulture treatment farmers does not increase relative to control farmers. The survey data show that use of inputs and techniques among horticulture farmers is common, particularly fertilizer which is used by 84 percent of treatment and 88 percent of control farmers for F4APK focus crops at baseline. The data also suggest that the F4APK crops require more intense use of certain inputs. For example, fertilizer is used more frequently for F4APK crops (84 percent versus 69 percent, among treated farmers). Irrigation is particularly important for fruits and vegetables and is used much more often for F4APK crops than for other crops. However, it is not used more by F4APK treatment farmers than by the control group at baseline.

We also do not see a stronger increase in the use of these inputs for F4APK crops among treatment farmers: in all cases the DiD estimate is not significant, meaning the input use trend between baseline and endline is not different from the control group. For example, we see an increase in fertilizer use for these crops from 84 percent to 88 percent of crop decisions in the treatment group, but this is

[33] Note that the literature typically does not find such training spillovers (e.g. Waddington et al., 2014).

balanced by an increase from 88 to 92 percent in the control group. The use of pesticides is much lower for F4APK vegetable crops and it decreases over time, whereas it increases for other crops. Pesticide use for these vegetables is lower among F4APK treatment farmers than among control farmers and it decreases over the study period. This is interesting as lower pesticide use is consistent with the aim to satisfy certification standards and judicious pesticide use is part of the crop management trainings.[34] However, the control farmers also decrease their pesticide use, albeit from higher starting levels.

It is interesting to see significant and positive DiD effects for the non-F4A target crops for three inputs or techniques. These are intercropping, fertilizer use and use of hybrid seeds. For example, we see that fertilizer use for other crops is at 69 percent among treatment farmers at baseline and this increase to 84 percent at endline. This shift almost bridges the gap between treatment and control farmers in the application of fertilizer. These results suggest that lessons learned by the treatment group were not applied to the target crops but to other crops. This is consistent with the reduced engagement in terms of acreage devoted to F4A crops (see Table 12) noted earlier and could signify a positive spill-over of the program into other crops. We note that with the alternative "farmer training" treatment definition that directly compares trained and untrained farmers, we no longer find a significant positive effect on the use of hybrid seeds for other crops. This suggests that it is not the training itself but rather being part of the treatment PO that leads to the increase.

Table 18: Input use (fraction of crop decisions where input was used by horticulture farmers)

		F4A crops			Other crops		
		Baseline	Endline	Diff-in-diff	Baseline	Endline	Diff-in-diff
Intercropping	Control	0.16	0.06		0.73	0.62	
	N	972	535		1016	1572	
	Treatment	0.21	0.16	0.05	0.62	0.61	0.08**
	N	1180	629	(0.04)	1722	2806	(0.04)
Fertilizer	Control	0.88	0.92		0.81	0.87	
	N	972	535		1020	1572	
	Treatment	0.84	0.88	0	0.69	0.84	0.09***
	N	1180	629	(0.04)	1726	2806	(0.03)
Pesticides	Control	0.48	0.37		0.40	0.53	
	N	972	535		1016	1572	
	Treatment	0.38	0.29	0.01	0.45	0.60	0.04
	N	1180	629	(0.04)	1722	2806	(0.04)
Irrigation	Control	0.57	0.50		0.20	0.17	
	N	972	535		1020	1572	
	Treatment	0.52	0.49	0.01	0.14	0.11	0
	N	1180	629	(0.04)	1726	2806	(0.02)
Hybrid seeds	Control	0.50	0.39		0.44	0.51	
	N	972	535		1020	1572	
	Treatment	0.41	0.36	0.04	0.31	0.43	0.06*

[34] The use of pesticides in not discouraged but the advice provided is the use of integrated pest management (IPM) which advocates for the use of chemical pesticides as a last resort after the nonchemical pest/disease control methods are unable to keep the pests below the economic threshold. The standards do not prohibit the use of pesticides but rather encourage use of pesticides in a judicious manner that will not result in exceeding of minimum residue levels MRL).

	N	1180	629	(0.05)	1726	2806	(0.03)

4.3.2 Dairy

Table 19 shows that dairy specific input use is relatively common at baseline. For example, 79 percent of dairy PO farmers in the treatment group report spending money on artificial insemination; 71 percent on acaricide (a pesticide that kills, a.o., ticks); 67 percent on veterinary services; 97 percent on deworming; and 96 percent on feed supplements. In other words, a (large) majority of farmers are aware of these inputs and shows a willingness to use them already at baseline. In other words, for these inputs it looks harder to achieve productivity increases through additional input use.

The main category where F4APK dairy treatment farmers at baseline spent significantly less relative to controls is fodder. This is the input where we see the only positive and significant DiD estimate. The increase is such that the treatment group has caught up with the control group by endline in this respect. The increase is consistent with the dairy training content (see Appendix B, training topic labeled "Feeds and Feeding strategies"). We note that with the alternative "farmer training" treatment definition that directly compares trained and untrained farmers, we find a significant positive effect on the use of one more input: acaricide.

Table 19: Dairy Inputs (fraction of farmers using the input)

		Baseline	Endline	Diff
Acaricide	Control	0.66	0.63	
	N	149	135	
	Treatment	0.71	0.73	0.09
	N	234	220	(0.08)
Artificial insemination	Control	0.70	0.67	
	N	149	135	
	Treatment	0.79	0.73	0.02
	N	234	220	(0.06)
Vet services	Control	0.65	0.67	
	N	149	135	
	Treatment	0.67	0.71	0.04
	N	234	220	(0.07)
Deworming	Control	0.97	0.95	
	N	149	135	
	Treatment	0.97	0.95	0
	N	234	220	(0.03)
Feed supplements	Control	0.93	0.95	
	N	149	135	
	Treatment	0.96	0.95	-0.02
	N	234	220	(0.03)
Fodder	Control	0.58	0.50	
	N	149	135	
	Treatment	0.39	0.49	0.21***
	N	234	220	(0.06)

Bull service	Control	0.07	0.03	
	N	149	135	
	Treatment	0.04	0.05	0.04
	N	234	220	(0.03)
Animals	Control	0.17	0.19	
	N	149	135	
	Treatment	0.24	0.20	-0.05
	N	234	220	(0.06)
Irrigation	Control	0.07	0.16	
	N	151	135	
	Treatment	0.17	0.25	0.03
	N	234	220	(0.04)

4.4 Medium-term outcomes

4.4.1 Productivity

A key medium-term outcome along the F4APK results chain is productivity, or output per reference unit of the production factors land and livestock. We choose as our main outcome indicators weight (kg) produced per acre for the F4APK horticulture crops (combined an individually); and litres produced per cow for milk.

For the main F4APK target crops we do not find positive productivity effects, except when focusing on the pre-COVID period. In Table 20 we present productivity estimates for the three **horticulture** crops grown by a relatively large number of farmers: French beans, tomatoes and bananas.[35] We present two sets of estimates: columns 1-3 have the standard baseline-endline comparison, which includes three growing seasons before each survey (2018-19 at baseline, 2020-21 at endline). Columns 4-6 compare the main season of 2018 with the main season of 2019, which ended before the start of the pandemic; we call these the "pre-COVID" estimates.

In the standard results (cols 1-3) we observe slightly decreasing productivity levels in the treatment group, constant levels in the control group and DiD estimates that are not or marginally significant. In the right-hand side of Table 20 (pre-COVID) we observe positive and significant DiD productivity estimates for French beans and bananas, and a negative one for tomatoes. This suggests that the COVID pandemic does affect the productivity DiD estimates, and reduces them for French beans and bananas. A caveat for the pre-COVID estimates is that the numbers of observations are in some cases small and much reduced between baseline and endline, suggesting the possibility of (strong) selection effects. This is especially true for French beans (and tomatoes): we interpret the positive significant effect for bean growing farmers as a selection effect, rather than a program effect. That is, the very small sample observed at the pre-COVID main season is not randomly selected: the four remaining French beans farmers are likely observed because they are specialized and produce throughout the year.

This selection effect seems less problematic for bananas. For this crop we observe quite strong productivity growth among the treatment farmers, while productivity among control farmers declines.

[35] The other crops (watermelon, capsicum, onion and butternut) are grown by very few farmers in the treatment and control groups.

The combined effect reverses the order, making F4APK the more productive group. This effect can be more plausibly attributed to the program. The comparison with the pre-COVID seasons provides some evidence of a negative effect of COVID on productivity growth, especially for bananas.

Table 20: Productivity (Ton per acre)

		All seasons			Pre-COVID, Main seasons		
		Baseline	Endline	Diff	Baseline	Endline	Diff
French beans	Control	2.27	2.46		2.19	2.08	
	N	63.00	38.00		44.00	24.00	
	Treatment	2.70	1.81	-1.1*	1.85	5.83	3.67**
	N	55.00	37.00	(0.65)	21.00	4.00	(1.75)
Tomatoes	Control	6.47	5.73		5.65	9.16	
	N	54.00	51.00		27.00	10.00	
	Treatment	5.33	4.42	-0.11	4.76	2.88	-7.01**
	N	80.00	41.00	(2.1)	35.00	14.00	(3.14)
Bananas	Control	4.86	3.91		5.28	3.88	
	N	243.00	230.00		136.00	114.00	
	Treatment	6.48	5.17	-0.20	3.83	4.77	2.45**
	N	345.00	307.00	(1.04)	193.00	158.00	(0.97)

Productivity results for **dairy farmers** are presented in Table 21, which shows monthly production per cow in liters. We observe an increase in dairy productivity of about 20 liters per cow per month or about 9 percent relative to the baseline level. This causes the F4APK treatment farmers on average to catch up with control farmers at endline. The implied productivity level at baseline in the treatment group is 8.5 liters per cow per day and the productivity increase in this group is three quarters of a liter per cow per day. The pre-COVID comparison for dairy shows very similar results.

Table 21: Monthly Productivity Dairy (Liters)

		Baseline	Endline	Diff
Production per cow per month	Control	248.31	242.44	
	N	231.00	227.00	
	Treatment	232.30	255.52	19.77
	N	393.00	381.00	(18.21)

4.4.2 Sales (Use of produce)

How do **horticulture farmers** use their F4APK crops? Table 22 answers this question for French beans and bananas. From a PSD perspective, the third outcome listed in the table, "fraction of produce sold", is especially important. Clearly, French beans are a key cash crop: they play almost no role in the household diet and have much higher sales fractions than bananas. We observe an increase in share sold for both products, both in the F4APK treatment group and the control group. However, the numbers are small for French beans and the increase could be a function of sample selection, with

only specialists remaining in the sample at endline. Bananas are produced by far more farmers in our sample; and producers consume about half of their production, which is important with an eye to the long-term food security goals of the program.

The three bottom rows in the table present the share of sales going to three marketing channels: sales to a distribution (collection) center, sales to a local market, and other sales channels. Here we find that, in the treatment group, French beans had a substantive share (62 percent) of sales through the collection/distribution center at baseline. This share is reduced at endline to 44 percent, while the share sold by control farmers through this channel increases. We note that this finding is not consistent with the survey evidence on distance to collection centers (Table 13), which showed a relative improvement in access for horticulture treatment farmers. But it is consistent with the annual reports showing that none of the newly installed F4APK collection centres were functional in 2020. Treatment farmers have replaced this channel partially with the "local market", but overall the centers remain the main channel for sales of French beans. For bananas a major shift took place from "other channels" to the "local market" (possibly under the influence of COVID-19 travel restrictions).

Table 22: Utilization of harvest

		French beans			Bananas		
		Baseline	Endline	Diff	Baseline	Endline	Diff
Consumed (fraction)	Control	0.09	0.03		0.49	0.45	
	N	65.00	37.00		148.00	128.00	
	Treatment	0.08	0.07	0.04	0.53	0.46	-0.03
	N	25.00	18.00	(0.04)	207.00	175.00	(0.05)
Sold (fraction)	Control	0.83	0.97		0.49	0.53	
	N	65.00	37.00		148.00	128.00	
	Treatment	0.83	0.93	-0.06	0.45	0.51	0.02
	N	25.00	18.00	(0.06)	207.00	175.00	(0.05)
Fraction of sales to							
Distribution center	Control	0.88	0.95		0.01	0.01	
	N	61.00	37.00		105.00	101.00	
	Treatment	0.62	0.44	-0.18	0.00	0.00	0.00
	N	22.00	18.00	(0.22)	125.00	129.00	(0)
Local market	Control	0.01	0.05		0.31	0.82	
	N	61.00	37.00		105.00	101.00	
	Treatment	0.10	0.28	0.09	0.51	0.60	-0.41***
	N	22.00	18.00	(0.1)	125.00	129.00	(0.08)
Other channels	Control	0.10	0.00		0.68	0.17	
	N	61.00	37.00		105.00	101.00	
	Treatment	0.28	0.28	0.09	0.49	0.40	0.41***
	N	22.00	18.00	(0.15)	125.00	129.00	(0.08)

Table 23 shows for **dairy farmers** the share (of production quantity) used for a particular consumption or marketing destination. For all farmers, the two main destinations for their produced milk are a) sold to cooperatives and b) consumed by household members. At baseline cooperative sales are at 50 percent for treatment farmers and 59 percent for control farmers. The most significant development reported in Table 23 is the substantial increase in the cooperative sales share, while the control sales remain flat. The regression DiD is 20 percentage point (significant at the 10 percent level) and translates into an impressive relative increase of 40 percent of the sales share (relative to the baseline level of 50 percent). The change means that treatment farmers caught up and overtook the control

farmers in this respect during the study period. This result is a success for the F4APK theory of change and links the supply side (productivity) pathway to the private sector development pathway.

Table 23: Use of milk production (share of production)

		Baseline	Endline	Diff
Consumed by household	Control	0.33	0.28	
	N	214.00	184.00	
	Treatment	0.29	0.26	0.87
	N	351.00	318.00	(6.7)
Sold to cooperatives	Control	0.59	0.59	
	N	214.00	184.00	
	Treatment	0.50	0.67	0.20*
	N	351.00	318.00	(0.10)
Sold to intermediaries	Control	0.06	0.07	
	N	214.00	184.00	
	Treatment	0.16	0.14	-0.05
	N	351.00	318.00	(0.06)
Sold to consumers	Control	0.08	0.11	
	N	214.00	184.00	
	Treatment	0.04	0.04	-0.04
	N	351.00	318.00	(0.03)

4.4.3 Prices

We find that prices for all seven F4APK crops have decreased for control group farmers, while prices decreased less or increased for treatment farmers for all crops except bananas. Since the number of observations is small for most F4A crops, especially at endline, we analyse the weighted average price obtained by farmers for the seven targeted F4A crops in Table 24. Note that there is a process of self-selection that probably influences prices and the price effects. We have seen that many farmers have disengaged from horticulture over the study period, especially in the treatment group.

The conclusion from Table 24 is twofold: (1) horticulture prices decreased over the study group, but the decline was stronger for the control group farmers so that a positive DiD effect is obtained; (2) the positive DiD is driven by price changes in local markets and "other" channels (e.g. traders), not in distribution centers. The latter are the key horticulture marketing channel targeted by the programme, and therefore we do not interpret these price results as a convincing positive program effect.

The bottom rows of the table provide tests of equality of the price *levels* between the different channels. The null-hypothesis is that the price levels are the same and we cannot reject this for any comparison. However, the largest difference is observed at baseline (p=0.13), with higher prices obtained at distribution centers. This difference is much less pronounced at endline, meaning that distribution centers have become a less attractive proposition over time.

Table 24: Average horticulture prices, by channel

		Baseline	Endline	Diff
All	Control	32.09	21.76	

	N	160	135	
	Treatment	28.13	23.56	5.95**
	N	209	167	(2.68)
Distribution centers	Control	40.03	32.19	
	N	61	40	
	Treatment	33.03	25.52	3.83
	N	37	28	(5.26)
Markets	Control	36.32	22.75	
	N	89	116	
	Treatment	29.23	23.34	7.48**
	N	151	121	(3.51)
Other	Control	31.15	17.81	
	N	132	32	
	Treatment	28.05	23.68	8.12**
	N	133	76	(3.6)
P-values for tests of equality for different hypotheses	All=Distrib	0.13	0.62	
	All=Markets	0.55	0.93	
	All=Other	1.00	1.00	
	Distrib=Markets	0.27	0.60	
	Distrib=Other	0.15	0.66	
	Market=Other	0.57	0.91	

At the disaggregated level, we find that French beans prices reported by treatment farmers are slightly higher than for control farmers at baseline; and prices for control farmers decrease while they remain nearly constant for treatment farmers. This suggests that the treatment farmers' strategy of moving to alternative channels pays off through. Banana sales prices are higher for treatment than for control farmers, and decrease for both groups between baseline and endline; here we find a negative significant DiD. The decrease in banana prices can be linked to increased competition in local markets. As shown in Table 22, 2020 saw a large supply increase in local markets by existing sellers, especially in the control group. In addition, new entry into local produce markets by non-traditional sellers such as office employees and taxi drivers, as a result of COVID-19, was mentioned in several interviews.

Milk prices of sales to cooperatives increase for both treatment and control farmers between baseline and endline. For treatment farmers the mean reported increase is substantial, from 34 to 40 KES per liter (18 percent), but the DiD (increase relative to the control group increase) is not significant.

4.4.4 Certification

Table 25 shows the fraction of farmers reporting having received some type of certification. The upper panel reports about any type of certification, the lower panel only counts if farmers mention KenyaGAP or GlobalGAP certification.

The table provides evidence that the overall level of certification reported by farmers is low at baseline, but especially for French beans there is a clear drive towards increased levels of certification. However, this drive is present in both the control and the treatment group and there is no significant

DiD effect, except for a lower generic certification among treatment farmers. This makes sense because of the program emphasis on the KenyaGap/GlobalGap certifications. For these salient forms of certification, there is an increase for French beans from 4 percent to 28 percent among treatment farmers. There is a similar increase for control farmers.

Table 25: Certification

		Baseline	Endline	Diff
Any F4A crops	Control	0.02	0.08	
	N	190.00	161.00	
	Treatment	0.09	0.04	-0.11**
	N	290.00	223.00	(0.04)
French beans	Control	0.05	0.28	
	N	65.00	46.00	
	Treatment	0.16	0.28	-0.14
	N	25.00	18.00	(0.18)
KENYAGAP/GLOBALGAP				
Any F4A crops	Control	0.02	0.07	
	N	190.00	161.00	
	Treatment	0.02	0.04	-0.04
	N	290.00	223.00	(0.04)
French beans	Control	0.03	0.26	
	N	65.00	46.00	
	Treatment	0.04	0.28	0.02
	N	25.00	18.00	(0.16)

4.5 Long-term outcomes

The key long-term outcomes of interest are the value of sales and the resulting farmer surplus (profit). The farmer surplus in the F4APK target value chains arguably determines the long-term attractiveness of the programs. In this section we present estimates of sales, costs and surplus for the target horticulture crops and milk, by value chain.

Profit is defined as the KES value of produce sales minus production costs. In the results that follow, we provide the per capita value of absolute sales, cost and value levels (the values reported at household level have been divided by the number of adults in the household). The values are reported in thousands of Kenyan Shillings (KES) per year. The sales and home consumption measures are based on reported quantities (sold or consumed) and prices of the relevant products. Production costs consist of all direct costs of production for each farmer, summed over her production related expenditures including agriculture and dairy activities, and irrigation.[36]

4.5.1 Sales and profit

Summary results of the calculations are provided in Table 26, which consists of three panels. Panel A (Horticulture) presents profits for the F4APK horticulture crops (by PO type), panel B does this for

[36] Production costs have been allocated to the milk and crop production, based on the amount of land devoted to a particular value chain.

Dairy products (mainly milk). Each of these two panels provides four results: (1) the per capita sales; (2) per capita costs; (3) per capita business profit, that (1) minus (2); (4) the per capita value of home consumption for F4A crops (in A, only for horticulture farmers) and milk (in B, only for dairy farmers), at market prices. Panel C provides the per capita sales of maize and beans, for comparison.[37]

A number of results stand out. First, per capita profit (sales minus costs) are positive for dairy production and mostly negative for the horticulture target crops. The losses on horticulture crops are not very large, but are present for both the horticulture POs and Dairy POs, both at baseline and endline. Positive profits are only calculated at baseline for the horticulture treatment group. These results do suggest that overall, the experience with these crops is mixed and profit is not guaranteed.

Second, the value of home consumption does not change the conclusion on surplus. For F4A crops, the home consumption value is very small relative to sales and costs. Nevertheless, there is a significant positive DiD for home consumption value, as the negative trend in home consumption was much less pronounced in the treatment POs. For dairy, the value of home consumption is substantial, both at base- and endline, and exceeds profits for the treatment group.

Third, horticulture sales and (investment) costs decrease between baseline and endline for horticulture PO farmers, especially for the control group. This is consistent with the declining engagement (number of crops, acreage devoted) noted in Table 12. The opposite is true for dairy PO farmers: especially for dairy control farmers horticulture costs increase substantially, leading to a larger loss. The profit DiD is positive because treatment farmers manage to reduce their losses to approximately zero.

Fourth and conform expectations, in the Dairy panel we find that sales and costs are at higher levels for the dairy POs. For these POs, we also see strong growth in sales per capita (consistent with earlier findings) and also in profits. The positive trend for the treatment group is matched by the control group, therefore we do not see a positive (significant) DiD. An unexpected finding is the strong growth in dairy sales and profit for the horticulture POs. This suggests that the observed decrease in horticulture engagement was matched by a successful engagement in dairy, that more than compensated the lower profit income for horticulture crops. Profitability (profit over sales) of dairy production is between 25 and 45 percent for dairy farmers.[38]

The table confirms the respective specialization levels, with F4A crop sales higher in the horticulture POs. Dairy POs have much higher dairy sales than horticulture POs. We do not find large changes in costs per capita in the treatment group. This suggests that, relative to the control farmers, the program did not change the aggregate level of investment or production expenditure in the treatment group. This is consistent with the already high level of adoption of techniques at baseline (see Section 4.3).

Fifth, there is a general reduction in the value of home consumption between baseline and endline, especially for treatment farmers. The value of crops and dairy products consumed at home is between 10 and 18 thousand KES (depending on the subgroup and survey round) and therefore typically somewhat larger than farm profits from dairy.

Sixth, we observe strong sales growth for the staple crops maize and beans, across PO types. For horticulture POs, at endline, these staple crop sales on average represent a larger value than F4A

[37] Note that these means are calculated over all surveyed farmers: farmers that do not produce a product group are assigned zero sales (profit) for that group.

[38] Remarkably, profitability of dairy is higher for horticulture farmers, at lower overall dairy sales levels.

horticulture crops, and also than dairy sales, for both control and treatment farmers. For dairy POs, staple crop sales are smaller but growing; at endline they are equal in size to F4A horticulture crop sales, but much smaller than dairy sales.

How high are these per person values in relation to other earning opportunities? As a benchmark for comparison, we use the baseline survey data to calculate a mean agricultural wage for casual labour of about KES 16,000 (annualized). This amount is equal to the per person home consumption in the horticulture control group, both at baseline and endline. If we add up the mean sales values (F4A crops, dairy and staples) for the horticulture treatment farmers at endline, the total sales amount to 39 thousand KES or more than twice the casual wage rate.

Table 26: Sales and profit

(A) Horticulture crops		PO type					
		Horticulture			Dairy		
		Baseline	Endline	Diff-in-diff	Baseline	Endline	Diff-in-diff
Sales per capita	Control	14.81	10.33		1.21	4.08	
	N	313.00	273.00		149.00	135.00	
	Treatment	8.56	6.01	1.39	3.87	5.94	-0.67
	N	534.00	480.00	(1.61)	234.00	220.00	(1.08)
Cost per capita	Control	15.44	11.47		2.94	9.69	
	N	313.00	273.00		149.00	135.00	
	Treatment	5.74	6.36	2.17	4.43	5.96	-5.11**
	N	534.00	480.00	(1.83)	234.00	220.00	(1.95)
Profit per capita	Control	-0.63	-1.15		-1.72	-5.61	
	N	313.00	273.00		149.00	135.00	
	Treatment	2.82	-0.34	-0.78	-0.56	-0.03	4.44**
	N	534.00	480.00	(1.85)	234.00	220.00	(2.16)
Home consumption per capita	Control	1.02	0.24				
	N	313.00	273.00				
	Treatment	0.47	0.22	0.56*			
	N	533.00	478.00	(0.32)			

(B) Dairy		PO type					
		Horticulture			Dairy		
		Baseline	Endline	Diff-in-diff	Baseline	Endline	Diff-in-diff
Sales per capita	Control	15.71	14.75		25.74	31.85	
	N	313.00	273.00		149.00	135.00	
	Treatment	8.90	13.29	5.51***	21.24	27.92	0.53
	N	534.00	480.00	(1.78)	234.00	220.00	(4.06)
Cost per capita	Control	4.51	5.18		18.10	17.56	
	N	313.00	273.00		149.00	135.00	

	Treatment	2.99	4.02	0.3	15.99	15.96	0.77
	N	534.00	480.00	(0.62)	234.00	220.00	(1.75)
Profit per capita	Control	11.21	9.56		7.65	14.29	
	N	313.00	273.00		149.00	135.00	
	Treatment	5.91	9.27	5.21***	5.25	11.95	-0.24
	N	534.00	480.00	(1.75)	234.00	220.00	(3.81)
Home consumption per capita	Control				8.98	6.50	
	N				149.00	135.00	
	Treatment				19.75	16.72	-0.33
	N				234.00	220.00	(2.4)

(C) Other		PO type					
		Horticulture			Dairy		
		Baseline	Endline	Diff-in-diff	Baseline	Endline	Diff-in-diff
Sales per capita: maize and beans	Control	12.94	22.80		2.09	8.12	
	N	313.00	273.00		149.00	135.00	
	Treatment	9.03	17.09	-1.72	0.48	4.25	-2.02
	N	534.00	480.00	(3.37)	234.00	220.00	(2.52)

4.5.2 Food access and well-being

A second group of long-term outcome indicators represents food access status and well-being. Table 27 reports results on three of these. Overall, these indicators are moving in the "right direction" for horticulture farmers and in the "wrong direction" for dairy farmers. As a result, at endline these indicators have moved closer together for the two types of farmers.

We find that horticulture treatment farmers have higher reported food shortages than dairy farmers. At baseline, 35 percent of horticulture treatment farmers report there were times over the past 12 months when the household did not have sufficient food to meet the needs of the family; at endline this is nearly the same, at 33 percent. 18 percent of dairy treatment farmers report this at baseline, and 24 percent at endline.

The share of dairy farmers saying their household did not have enough food last year has increased. For dairy treatment farmers this share increases to 24 percent, which represents a lower increase than in the dairy control group. We see similar trends for both treatment groups for the second, more acute food access outcome ("Current food consumption is not enough"). There are no significant DiD effects. For dairy farmers, these welfare indicators are consistent with the downward trend in their home consumption values.

On the question "Taking all things together, how would you say things are going for you these days?", 27 percent of dairy farmers respond "Good" or "Very Good" at baseline, against 17-18 percent for horticulture farmers. Interestingly, the share goes down (to 21 percent) for dairy farmers and up to 19 percent for horticulture farmers.

Table 27: Food access and well-being (% of households)

	Horticulture	Dairy

		Baseline	Endline	Diff-in-diff	Baseline	Endline	Diff-in-diff
Fraction with							
In the last year, at times the household did not have enough food	Control	0.28	0.27		0.22	0.30	
	N	313	273		150	142	
	Treatment	0.35	0.33	-0.02	0.18	0.24	-0.06
	N	535	480	(0.04)	239	220	(0.06)
Current food consumption is not enough	Control	0.30	0.19		0.15	0.24	
	N	313	273		150	142	
	Treatment	0.33	0.30	0.06	0.21	0.25	-0.08
	N	535	480	(0.04)	239	220	(0.06)
Life satisfaction ("good/very good")	Control	0.18	0.15		0.27	0.15	
	N	313	273		150	142	
	Treatment	0.17	0.19	0.05	0.27	0.21	0.07
	N	535	480	(0.04)	239	220	(0.06)

4.6 Robustness checks

We have performed two sets of robustness checks on our regression analyses. The first robustness analysis uses a different, more "active" definition of treatment: here a farmer is in the treatment group only if she has indeed participated in a training after baseline. This definition provides an impact estimate that is, in contrast to the Intention-to-Treat (ITT) definition we used so far, not subject to possible dilution because of non-participation of individual farmers. We have used this new definition to re-estimate our regressions but do not find any major qualitative differences in the impact estimates.[39]

A second set of checks asks for a number of outcomes whether the double difference estimates are different for female headed households (in effect doing a triple difference analysis). For most of these outcomes we do not find that the relative changes (DiD estimates) are different for female headed households. In a few cases where we do find a (marginally) significant difference, for example for cow ownership, the sign is negative. This implies that, relative to male headed households, cow ownership decreased for female headed households in the treatment group; whereas for the group as a whole (on average) there was no change in this respect. One explanation, supported by the literature (see Rosenzweig and Wolpin, 1993; Upton, 2004), is that cows embody savings and provide a form of self-insurance. As such, they provide a savings buffer against risks and may well have been used to bridge a difficult time for female headed households (COVID-19).

Apart from the robustness checks, we can obtain information about female headed households from the regressions that provide the DiD estimates in the tables presented in this chapter. When we look at the regression coefficient for "female headed household" in the regressions used to create Table 26, we find that these households do not do worse than other households in the horticulture POs and better in the dairy POs, where they display slightly higher mean per capita profit for dairy and higher staple sales.

[39] As an example, using the new definition, for dairy farmers in Table 19 we find two significant positive DiD estimates for input use (for acaricide and fodder) instead of the one we reported; while we find two rather than three significant estimates for horticulture farmers (Table 18, non-F4A crops).

5. Sustainability and CSR

5.1 Sustainability (RQ 5-6)

RQ 5: To what extent do the benefits of the project (outcome & impact level) continue after FDOV-funding ceased and how was this influenced by the business case and/or revenue model?

The outcome and impact level benefits of the project have been described in the previous chapter, based on measurements that took place in February-March of 2021. This report was written in June-July 2021 and the evidence beyond the data already described is limited. We therefore use data from the implementation period to answer these questions.

5.1.1 Financial sustainability of the business cases

The project implemented three business cases: processed beans, seedling nurseries and dairy (hubs). For RQ 5, we first use the financial data for the business cases, as presented in the F4APK Annual Reports (For a full assessment of the business cases, see the F4APK BC assessment Report Final).

For the processed beans business case, from the financial overview 2016-2019 in the 2019 AR, and the results for 2020 in the AR 2020, we can observe an impressive growth in turnover after the initial investment in the 2016 book year.[40] Sales grew to a highpoint of 4.2 million EURO in 2018; then, influenced by drought and COVID-19, reduced to 3.3 and 3.0 million EURO in 2019 and 2020, respectively.

A potential sustainability concern is that the net cash-flow margins appear to be thin (1.5-2 percent) and were negative in two of the five book years. This is consistent with the financial difficulties of Meru Greens reported on earlier, but of course this period includes the start-up years. Overall, the operational costs are high and close to the turnover, leaving little room for capital expenditures. The production level exceeds what was originally estimated: in 2020 1458 MT canned beans were processed, whereas the project document targeted 800 MT. As shown in Chapter 2, substantial job creation took place. In 2018 and 2019, the farm gate value of the beans represents approximately 20-25 percent of the operational expenditure, so if other costs can be managed downwards there is potentially room for increased profitability without reducing the farmers' income from sales.

The nursery business cases appear to be very healthy financially and do not raise sustainability concerns. There is clearly demand for the products of the nurseries and costs appear to be low relative to sales. In 2019, the ratio of net cash-flow to turnover was about 75 percent. At the same time, 378 jobs (FTE) were created.

The dairy business case also appears to be financially healthy and stable: the ratio of net cash-flow to turnover was 6.3 percent in 2018; and 5.8 percent in the 2019 book year. (The 2020 AR shows that the profit and loss report for 2020 was not shared). The total turnover was lower in 2019 and 2020 compared with 2018, which may relate to the drought and COVID-19 disruption in these respective years.

[40] We note that the financial reporting has been revised and consolidated in the 2019 Annual Report and use the last (revised) reporting as per the 2019 and 2020 Annual Reports.

Finally, we add the "micro"-financial sustainability. Table 26 shows that farm level profit margins for the F4APK horticulture crops are negative at endline and (in most cases) at baseline, and mostly decreasing over the period studied. This means that these crops are not a financially sustainable choice for most farmers. This finding is supported by the decreasing share of farmers growing these crops. The picture is much more positive for dairy, where we find positive and increasing profits for most farmers.

5.1.2 Sustainability hardware (Results Area 2)

We next review some of the sustainability evidence by results area, based on the Annual Reports. In Result Area 2, the main investments are capital (hardware) and organisational investments. We have seen in Chapter 2 that these were largely delivered as planned.

The evidence on the condition of the hardware at the end of the project is mixed. The 2020 Annual Report finds that three out of six investments are in good condition and functional at the end of the project, and three are not. On the positive side we find 1) the dairy mini processing units; 2) the YPR nurseries; and 3) the beans processing factory. These investments are in good condition and functional. As also discussed in section 2.5.4, in 2020 the following hardware was reported as not being (fully) utilised: 1) the banana tissue hardening nurseries; 2) produce collection centres; and 3) the banana ripening chamber. In conclusion, on the horticulture side the sustainability of the hardware and infrastructure development is mixed, although the largest share (by far) of the project investment is located in the still successfully operating beans factory. The non-utilization of the collection centres is reflected in the farmer survey responses. This appears to represent a valuable opportunity that is lost to both farmers and the market uptaker.

The reports are especially positive about the dairy investments (two mini-milk processing plant each with a 5,000-litre chilling tank capacity). These are still operational at the end of the project and adding value to the chain. The milk cooling raises the price of milk sold and reduces spoilage: "Dairy farmers, for instance will no longer pour away surplus milk due to spoilage, they will rather process the milk into yoghurt and other by-products". Yoghurt production was, however, loss making and suspended, according to the 2020 AR.

The private-sector development pathway outputs included the establishment of long-term contracts within the value chain. Contracts were signed and renewed in several value chains, as reviewed in Chapter 2. However, the Solidaridad End of Project Evaluation (EPE) notes that: "As the project comes to an end, most of the groups, even though were trained on group governance, may lack advanced negotiation skills to enable sustained market penetration, linkages with industry players and emerging business opportunities." A second concern mentioned in the EPE is that "Some of the subsidised support includes input access and collective output marketing which were largely organized through the project." so that "Closure of project leaves a technical gap and despite the capacity already developed, this is likely to slow down the competitiveness of these hubs."

5.1.3 Sustainability training (Results Area 3)

The trainer and extension staff were, according to the 2020 AR ".. carefully and strategically selected from county government extension staff, private sector out grower companies' extension staff, farmer aggregators, and renowned knowledgeable lead farmers. This was to ensure continuity of farmer training and extension support after the end of the project." As reported in the EPE, the skills trained include "book keeping, grafting techniques, milk handling which not only strengthens the existing the rural knowledge culture and asset ownership but are also alternative sources of income."

It is reasonable to assume that the knowledge and experience invested by the project will remain with the individual trainers and can be applied in future training activities, provided these are budgeted and organised. The application of this knowledge by trainee farmers in production activities will depend on the production incentives, including the costs associated with some of the techniques. For milk production and for some of the horticulture products (beans) these incentives seem to be there, but training is often dependent on project budgets. As formulated by the EPE: "Some of the existing project components especially field trainings or soil sampling enjoyed financial support, facilitation or subsidies. Specific support provided by F4APK included facilitation of trainings, providing transport and lunch allowances for the extension agents to visit the groups and offering technical support or financial linkages. Without project support, it remains uncertain whether the hubs will meet some of these costs from their coffers."

The soil testing component faced a number of challenges as discussed, and adjusted its strategy accordingly. Possibly due to the the COVID-19 disruption, the share of farmers in the target producer organizations that say they have had their soil tested is only 4 percent (equal to the control group share). Since most of these tests were provided for free, at the end of the project it is uncertain that a sustainable business model has been created.

The Global GAP training and certification is a potentially sustainable component in this result area. The certification trend is positive, but is mainly restricted to French beans and is equal to the positive trend in the control group. Even though the certificates are valid for one year only, farmers have a clear incentive to keep pursuing certification in the presence of the processing factory market linkage.

5.1.4 Scalability and systemic change

RQ 6: Did the project/intervention lead to systemic change and/or was the intervention scalable? If yes, in what way?

The core intervention model is being used across the country and region. The contract farming model that links smallholder farmers to market uptakers through output purchases and extension work is not a wheel that needed to be invented. In fact, the basic model has been around for years and has been replicated many times over, across different supplier networks.

Against this basic contract farming background, F4APK added a number of innovations. When asked what were the most innovative aspects of F4APK, stakeholders highlighted a few aspects. First, processing facility and milk coolers were new investments (although not a new concept in the model). Second, the project added new techniques, including drip irrigation, which led to higher recruitment of farmers; and the start-up of successful seedling nurseries. Third, the selection of the trainers (of trainers) was changed so that they are part of the farm communities, not outsiders; this will help further enrolment according to the implementers. Fourth, the ICT component and use of information is regarded as a major innovation.

Horticulture export sales grew by 12 percent in 2016. As long as demand in the supported value chains grows at current rates, the adoption of these F4APK production innovations seems feasible.

5.2 Corporate Social Responsibility (RQ 7)

5.2.1 How relevant were the designed CSR plans?

The project CSR plans and mitigation measures were relevant. A range of relevant (International) Corporate Social Responsibility concerns were addressed in the project design document, and related risks were assessed (see section 2.6 of the design document, Netherlands Enterprise Agency, 2014). First, child labor is not used by any of the project partners, as per Kenyan laws and ILO requirements, and was therefore not a specific issue in project implementation. This is relevant for the established processing plant and for other project units that created employment, such as the Young Plant Raisers.

Second, the project focused on intensification of production on existing land (not on expansion into new land), thus avoiding deforestation or land conflicts. Third, environmental impact assessments were planned to be undertaken on all the projects that that, according the law of Kenya, required it, such as dairy cooling plants. Consultation with the National Environment Management Authority (NEMA) was planned to indicate whether the other hardware establishments required such assessments.

Fourth, the project emphasised organic production, quality upgrading and certifying production through KENYAGAP (Local) and GLOBALGAP (Global). The quality control related to these certificates makes it more likely that production is sustainable, takes care of the employee's welfare and consumer health.

The project annual reports provide an "Annex 3f: update ICSR risks and mitigation progress". This annex lists a number of ICSR concerns, including some that are not mentioned above: political risks such as election instability (2018), county level legislation affecting the project, and a breakdown of relations between the Kenya and Netherlands governments; bankruptcy of Meru Greens; water scarcity and other climate change related issues; outbreak of pests and disease; cultural issues affecting participation of women and youth. All of these concerns seem relevant and the mitigation measures proposed in the ICSR table appear to be sensible.

5.2.2 What effects can be observed of CSR plans of private partners in consortia?

This question has a counterfactual nature and is not easy to answer in general. We provide a number of observations, based on our different data sources.

A review of the "Annex 3f: update ICSR risks and mitigation progress" in the project annual reporting shows that the risk mitigation measures were implemented as planned. No special concerns were raised in the status updates. An example of specific evidence of the environmental mitigation measures comes in the form of the training topics (see Appendix B), that addressed concerns like pest management, GAP certification standards and improved environmental resource management.
In the survey data, we see that pesticide use among horticulture group members is lower in the F4APK treatment clusters. At baseline, we reported a positive treatment effect of the project in this regard: pesticide use was especially low among treatment farmers for those crops targeted by the project.

A major CSR concern emphasised by the project is the participation of women at all relevant levels. According to the project documentation, female (and youth) participation was a success. First, the share of female and youth reached by the project climbed steadily, from 45 percent in 2017 to around 70 percent in 2019-20. 64 percent of all trainees in 2020 were female. Second, a majority of the jobs created at the Meru Greens factory (73%) were for women, 60% of whom were on permanent contracts. Close to 50 percent of jobs at the YPR nurseries were occupied by women, and about 75 percent by youth. Third, the project emphasized female leadership development. According to the 2020 AR, "The number of women in the boards of directors of the five (5) dairy hubs has increased to

22% (11 out of 50 board members). Two (2) of the five (5) dairy hubs are managed by women while the rest are managed by men."

5.2.3 To what extent did the projects have a major positive or negative influence on their direct natural environment or contributed (combatting) global climate change?

From the previous subsections it is possible to conclude that environment related project CSR concerns were identified and addressed in project planning, reporting and execution. The reports do not include any signs of acute environmental problems, and there is some evidence that the interventions in fact lowered environmental pressure, based on training of good agricultural practices.

The project does interact with climate change. There is scientific evidence that the global livestock industry contributes to climate change. According to Grossi et al. (2019), the livestock sector ".. is responsible for about 14.5% of total anthropogenic greenhouse gas emissions" (see also Gerber et al., 2013).

The reverse problem is also present: agricultural production problems resulting from climate change have been identified in several places in the project documentation. The 2020 AR (cited earlier) warns that "Climate change has become the greatest challenge to sustainable agricultural production. Water for irrigation is increasingly becoming scarce in the project areas." To combat these resulting problems, "The project trained and introduced farmers to water efficient climate smart irrigation technologies and other farming methods geared towards climate change adaptation. Such technologies included drip irrigation, use of hybrid drought tolerant crop and fodder varieties, enterprise diversification and vegetable production under shade nets." (AR 2020, page 28).

References

Bamberger, M., Rugh, J., Mabry, L., 2011. RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints. SAGE.

Banerjee, A., 2015. Policies for a Better-fed World (Working Paper No. 21623). National Bureau of Economic Research. https://doi.org/10.3386/w21623

Bellemare, M.F., Bloem, J.R., 2018. Does contract farming improve welfare? A review. World Development 112, 259–271. https://doi.org/10.1016/j.worlddev.2018.08.018

Fabregas, Raissa; Michael Kremer, Matthew Lowes, Robert On, Giulia Zane, 2019. SMS-extension and Farmer Behavior: Lessons from Six RCTs in East Africa. Manuscript.

FAO, 2019. Kenya at a glance | FAO in Kenya | Food and Agriculture Organization of the United Nations [WWW Document]. URL http://www.fao.org/kenya/fao-in-kenya/kenya-at-a-glance/en/ (accessed 8.30.19).

Gerber, P. J., H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, and G. Tempio. 2013. Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities. Rome: FAO. Available from http://www.fao.org/3/a-i3437e.pdf

Giampiero Grossi, Pietro Goglio, Andrea Vitali, Adrian G Williams, Livestock and climate change: impact of livestock on climate and mitigation strategies, Animal Frontiers, Volume 9, Issue 1, January 2019, Pages 69–76, https://doi.org/10.1093/af/vfy034

Heinrich, Melina (2014), Demonstrating Additionality in Private Sector Development Initiatives: A Practical Exploration of Good Practice for Challenge Funds and other Cost-Sharing Mechanisms", Donor Committee For Enterprise Development (DCED).

KNBS and Agriculture and Food Authority, Horticultural Crop Directorate: Horticulture Validated Report 2015-2016.

Ministry of Agriculture, Livestock, Fisheries and Irrigation, 2019. Agricultural Sector Transformation and Growth Strategy 2019-2029 (Abridged Version).

Netherlands Enterprise Agency, 2014. Project Plan for Food for All in Kenya Project.

Ojwang, J.D; Odhiambo, H.G; Maraga, K., Mutisya, M.; Kavatha, A., Mwangi F. and Wambui E., 2020. Impact of COVID-19 on household dairy enterprise in Kenya. Heifer Project International Kenya, Nairobi.

Ragasa, C., Lambrecht, I., Kufoalor, D.S., 2018. Limitations of Contract Farming as a Pro-poor Strategy: The Case of Maize Outgrower Schemes in Upper West Ghana. World Development 102, 30–56. https://doi.org/10.1016/j.worlddev.2017.09.008

Rosenzweig, M.R. and Kenneth I. Wolpin., 1993. Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India. *Journal of Political Economy* 101, 223–244.

Ruth Stewart, Laurenz Langer, Natalie Rebelo Da Silva, Evans Muchiri, Hazel Zaranyika, Yvonne Erasmus, Nicola Randall, Shannon Rafferty, Marcel Korth, Nolizwe Madinga, Thea de Wet, 2015. The effects of training, innovation and new technology on African smallholder farmers' wealth and food security. The Campbell Collaboration. https://doi.org/10.4073/csr.2015.16

Strohm, Kathrin, Hoeffler, Heike, 2006. Contract Farming in Kenya: Theory, Evidence from selected Value Chains, and Implications for Development Cooperation.

United Nations Statistical Division, 2005. Household Surveys in Developing and Transition Countries. United Nations Publications.

Upton, M., 2004. The Role of Livestock in Economic Development and Poverty Reduction. FAO PPLPI Working Paper 10, Rome.

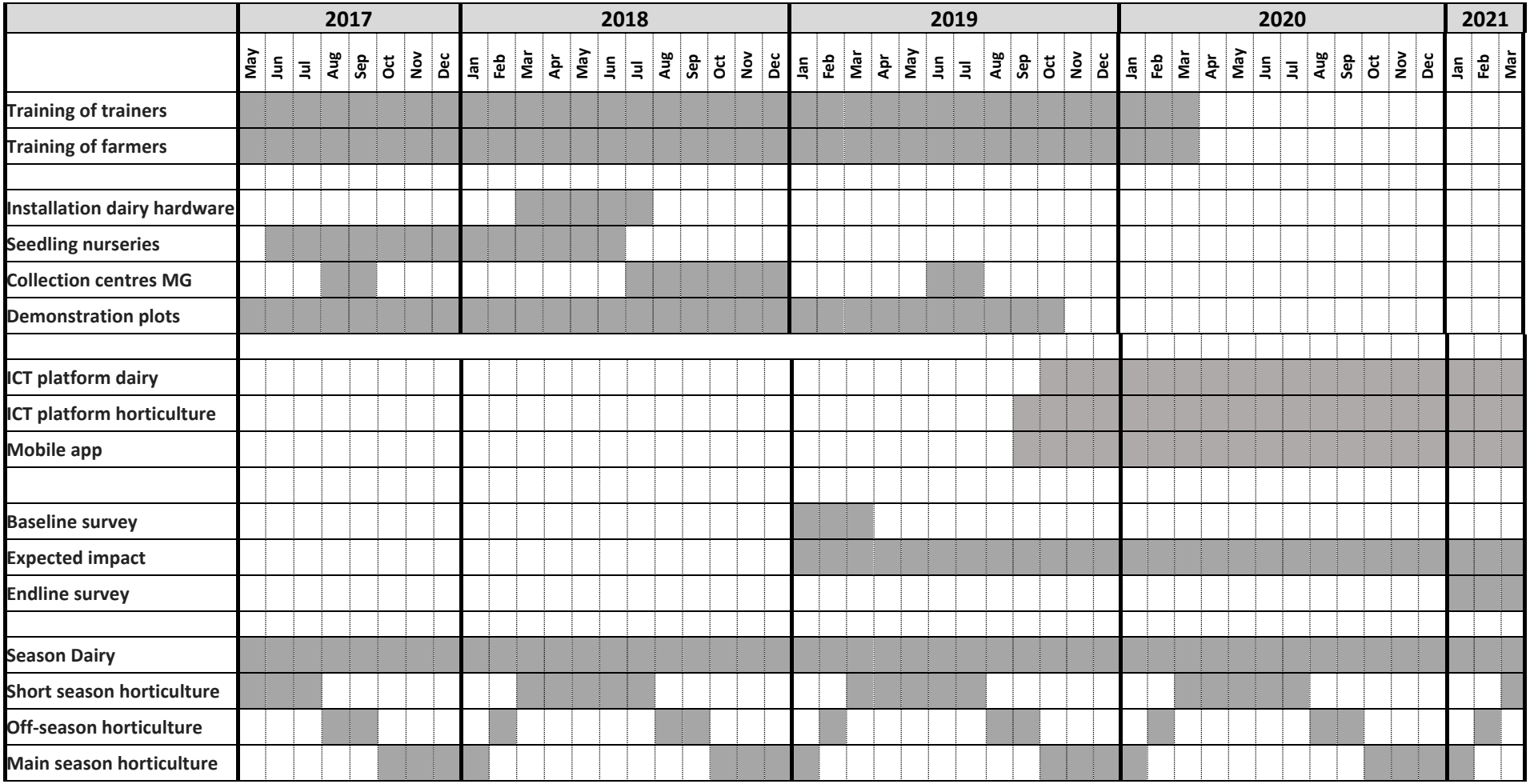
Waddington, H, Snilstveit, B, Hombrados, J, Vojtkova, M, Phillips, D, Davies, P and White, H. Farmer Field Schools for Improving Farming Practices and Farmer Outcomes: A Systematic Review. Campbell Systematic Reviews 2014:6. DOI: 10.4073/csr.2014.6.

Winters, P., Salazar, L., Maffioli, A., 2010. Designing Impact Evaluations for Agricultural Projects (Impact-Evaluation Guidelines Technical Notes No. IDB-TN-198).
World Bank, 2018. Policy Options to Advance the Big 4: Unleashing Kenya's Private Sector to Drive Inclusive Growth and Accelerate Poverty Reduction. World Bank Economic Update, April 2018.

Appendix A Timeline

Figure 10 gives a timeline of the inception and completion of all program elements, the agricultural seasons for horticulture and dairy in Kenya, as well as the planned evaluation surveys. The implementation timing data are from project documents and discussions with the F4APK management.

Figure 10: Timeline of project activities



Note: monthly timing is approximate, based on project documents and discussions with F4APK management.

Appendix B Training topics

F4APK Training topics

Horticulture training topics

Topics	Details
Importance of GAP in the horticultural supply chain Horticultural Nursery management.	Elements of good agricultural practices (GAP) Nutritional value of fruits and vegetables Advantages of nursery seedling raising. Nursery Site selection Seedling management.
Horticultural crop nutrition	Farm soil management Maintenance of soil fertility Soil and Cultural practices
Soil fertility management, soil sampling, analysis and crop nutrition	
Crop pests and diseases	Crop infestation and damage Development of plant disease GAP in management of crop pest and diseases (IPM, Pesticide application, field hygiene, health and safety)
GAP specific for each crop	GAP for: Tomato Capsicum Butter nut Water melon French bean Onion Banana
Integrated Pest Management (IPM)	Cultural, Biological and chemical pest and disease control
Record Keeping Market and Marketing	Traceability and record keeping Horticulture Value chain Market Players Functions Gross margin analysis Break even Analysis Marketing Channels
Global GAP certification Scheme training	Introduction to standards Benefits of standards Introduction to Global GAP

Note: The trainings have a separate component on certification that specifically targets 20 producer groups that are undergoing training on Global GAP leading to certification. Credit was not covered in the trainings. Only the youth and women (nursery groups) are trained on financial literacy.

GlobalG.A.P IFA Implementation Action Plan
Terms of Reference: Solidaridad project number PLAZA1312 Specification: GlobalGAP IFA Ver 5.1 Quality Management System and IFA Implementation action Plan
Training Location: Kaguru Agricultural Training Centre- Nkubu Training Dates: 12th-13th February 2019
Trainers: Peter Kanyarati (qualityapproach@gmail.com) , James Kituri (kimkahortconsultants@gmail.com John Muteti (johnmuteti01@gmail.com) Tel: 0722838678 ,0721951353,0720808374 Email: qualityapproach@gmail.com

Meru Greens Horticulture (MGH) Contact: (Head Office) Mr Samuel Afanda (0725702831), Irene Wahome (0715080344), Gatembe Office: Francis Mutinda (0729822134 francis.mutinda@merugreens.com)

Time	Sessions Day 1 12th February 2019		Methods
0830 – 0900		Delegates Registration	
0900 – 1000		Introduction and Delegate Expectations:	Q&A Presentation
1000 – 1030		Introducing GlobalG.A.P IFA Fruits and Vegetables Standard	Q&A Presentation
1030 – 1100	Break		
11.00-12.30		Requirements for GLOBALG.A.P. QMS 1	Presentation Discussion
1230 – 1400	Lunch		
14.00- 1530		Requirements for GLOBALG.A.P. QMS 2	Group exercise
1530 – 1600	Break		
16.00- 1700		Farmer Group Requirements for GLOBALG.A.P. QMS	Group exercise reporting
Time	Sessions Day 2 13th February 2019		Methods
0800 – 0930		Assessment Process (Registration-Assessment-Certification) 1	Q&A Presentation
	Break		
1000- 1300		Assessment Process (Registration-Assessment-Certification) 2	Q&A Presentation
	Lunch		
1400-1530		Farmer Group Documentation/ records	Guided group work
	Break		
1530-1700		Launching the Quality Management System Manual	Group work Reporting
1700-1800	Delegates leave at their pleasure		

DAIRY FARMING TRAINING TOPICS, KEY MESSAGES, TRAINING APPROACHES AND TRAINING MATERIALS

Module	Topic	Key Messages
Breeding	Dairy cow breeding	Types of breeds Selection of suitable /best characteristics of a dairy cow Heat signs and heat detection Advantages of A.I over bull service Dairy cow fertility
Feeds and Feeding strategies	Animal Nutritional Requirement	Types of feeds (energy), proteins, minerals and water, concentrates. Feeding regime for heifers, calves, in calf cows and milking cows. Annual feed plan development
	Types of fodder and fodder establishment	(S. African Napier Variety, Ouma 2 Variety, Brachiaria, Boma Rhodes, Desmodium (silver leaf), Caliandria, leacaena, sesbania sesban, sweet potatoes vines
	Use of crop residues	Crop residue storage, urea treatment, key crop residues products.
	Hay making	Hay making and storage
	Silage making	Nappier grass, maize, Brachiaria etc

Calf rearing	calf management at birth Calf feeding Calf housing and management practices Weaning	Preparation for calving, Calf feeding, housing, calf diseases
Housing and security	Standard zero grazing units	Importance of housing, siting, components of the zero-grazing unit and milk hygiene.
Diseases	Causes, symptoms and control, vaccination and deworming schedules, spraying demos,	Mastitis, tick borne diseases, parasites, hygienic milking techniques, milk fever spraying demos

Group dynamics and social capital training topics

Cornerstones/Social capital	Nutrition, health and Income
Cornerstones/Social capital	Passing on the gift
Cornerstones/Social capital	Sharing and caring
Cornerstones/Social capital	Accountability
Cornerstones/Social capital	Improving the environment
Cornerstones/Social capital	Training, education and communication
Cornerstones/Social capital	Spirituality
Cornerstones/Social capital	Gender and family focus
Group formation	Group formation
Cornerstones/Social capital	Genuine need and Justice
Cornerstones/Social capital	Sustainability and self-reliance
Cornerstones/Social capital	Full participation
Cornerstones/Social capital	Improved animal and resource management
Cornerstones/Social capital	Genuine need and Justice
Cornerstones/Social capital	Accountability, Nutrition Health and Income
Cornerstones/Social capital	Gender and family focus

Appendix C Equations

This appendix provides the methodology that was used to calculate the different sales and profit figures.[41]

	All Crops	F4A Crops
Sales	$\sum_{i=S,O,M} Sales_i^{ALL}$	$\sum_{i=S,O,M} Sales_i^{F4A}$
Sales per acre	$\frac{\sum_{i=S,O,M} Sales_i^{ALL}}{\sum_{i=S,O,M} Acreage_i^{ALL}}$	$\frac{\sum_{i=S,O,M} Sales_i^{F4A}}{\sum_{i=S,O,M} Acreage_i^{F4a}}$
Cost per acre	$\frac{Total\ Costs}{\sum_{i=S,O,M} Acreage_i^{ALL}}$	
Total profit	$Sales^{ALL} - \frac{Costs}{acre} * \sum_{i=S,O,M} Acreage_i^{ALL}$	$Sales^{F4A} - \frac{Costs}{acre} * \sum_{i=S,O,M} Acreage_i^{F4A}$
Profit per acre	$\frac{Sales^{ALL}}{\text{Acreage}^{ALL}} - \frac{Costs}{Acre}$	$\frac{Sales^{F4A}}{\text{Acreage}^{F4A}} - \frac{Costs}{Acre}$
Profit per person	$\frac{Profit^{ALL}}{No.of\ Adults\ in\ household}$	$\frac{Profit^{F4A}}{No.of\ Adults\ in\ household}$

	Dairy
Sales	$Fodder\ sales + Milk\ sales + Other\ cowrelated\ sales\ (e.g. Cheese, animals, etc.)$
Sales per acre	$\frac{Sales^{Dairy}}{Acreage\ used\ for\ fodder}$
Total costs	$Sum\ of\ dairy\ related\ expenditures$[42]
Total profit	$Dairy\ Sales - Dairy\ costs$
Profit per acre	$\frac{Dairy\ profit}{Acreage\ used\ for\ fodder}$
Profit per person	$\frac{Dairy\ profit}{No.of\ adults\ in\ households}$

41 Notes: S = Short season, O = Off Season, and M = Main season. ALL stands for all crops, while F4A stands for all Food 4 All crops, which are banana, onions, water melon, French beans, capsicum, butternut, and tomatoes. Summing Acreage over three seasons, gives the size of land that is effectively used for production. This is different from physical land as land is reused throughout the year.

42 These expenditures include the money spent on commercial feeds, hay, grazing/pasture, veterinary drugs, veterinary chemicals, animal health care

This is a publication of:

Netherlands Enterprise Agency
Prinses Beatrixlaan 2
PO Box 93144 | 2509 AC The Hague
T +31 (0) 88 042 42 42
Contact us
www.rvo.nl

This publication was commissioned by the ministry of Economic Affairs and Climate Policy.

Publication number: RVO-169-2022/RP-DUZA

NL Enterprise Agency is a department of the Dutch ministry of Economic Affairs and Climate Policy that implements government policy for agricultural, sustainability, innovation, and international business and cooperation. NL Enterprise Agency is the contact point for businesses, educational institutions and government bodies for information and advice, financing, networking and regulatory matters.

Netherlands Enterprise Agency is part of the ministry of Economic Affairs and Climate Policy.