EVALUATION: NOVEMBER 2016 PUBLICATION: JULY 2019

LIVELIHOODS IN NIGER

Impact evaluation of the 'Community based integrated water resource management' project

Effectiveness Review Series 2016/17

A woman pours water that she drew from a tap connected to a 50,000 litre solar powered water tank, which was installed by Oxfam in north eastern Niger in 2016. Credit: Sam Tarling/Oxfam

ANDREW ANGUKO

OXFAM GB

www.oxfam.org.uk/effectiveness

OXFAM

CONTENTS

Acknowledgements **3**
Summary **4**
1 Introduction **9**
2 Project description **10**
3 Evaluation design **13**
4 Data collection **15**
5 Results **19**
6 Conclusions **32**
Appendix 1: baseline statistics before matching **34**
Appendix 2: Methodology used for propensity score matching **36**
Appendix 3: Robustness checks **39**
Notes **45**

ACKNOWLEDGEMENTS

We would like to thank Oxfam in Niger, particularly, Bruno Kokou Fugah, the Programme Director; Idrissi Hamidou, the Food Security and Livelihoods Officer; Katoumi Abdu Sirfi, the Food Security and Livelihoods Manager; and Aminata Ba, the Regional Programme Information Coordinator. Particular thanks are due to Moussa Harouna, the consultant who managed the fieldwork.

SUMMARY

Oxfam GB's Global Performance Framework is part of the organization's effort to better understand and communicate its effectiveness, as well as to enhance learning for staff and partners. Under this Framework, a small number of completed or mature projects are selected at random each year for an evaluation of their impact; this exercise is known as an 'Effectiveness Review'. One key focus is on the extent to which the projects have promoted change in relation to relevant Oxfam GB global outcome indicators. The global outcome indicator for the livelihoods thematic area is defined as 'total household consumption per adult equivalent per day'. This indicator is explained in more detail in section 5 of this report.

Niger's 'Community-Based Integrated Water Resource Management' project was one of those selected for an Effectiveness Review in the 2016/17 financial year. The project activities were implemented by Oxfam GB in conjunction with the partner organization Karkara and the Department of Agriculture of the Republic of Niger. The project was started in April 2013 and was completed in March 2015. It was evaluated one year after closure.

The project was implemented in the two villages of Banibangou and Soumatt in Banibangou commune. The project's overall objective was to increase agricultural production and income for farmers – in particular, women farmers - through integrated water resource management. The crops targeted for improvements in agricultural production included cabbages, tomatoes, onions, carrots, potatoes and sweet peppers. The choice of these particular interventions was derived in response to specific problems experienced by farmers in the two villages. The area has very low levels of rainfall, and local farmers reported low capacity in producing crops and were lacking the necessary inputs.

The partner organization played a key role in trying to solve these issues – particularly in the development of irrigation structures to allow for the cultivation of local vegetable crops. With funding support from Oxfam and its donor, wells and boreholes were dug, while water tanks were bought and installed with solar pumps. In addition, pipelines were connected to water basins in order to improve irrigation opportunities in the community. In the area of crop production, farmers were provided with inputs, including seeds and agricultural tools.

The partner also carried out capacity-building trainings in improved agronomic practices in conjunction with the Department of Agriculture. Exchange visits were organized and farmers were encouraged to organize themselves into groups in order to have better bargaining power in local markets. Oxfam provided the funds for project implementation and was in charge of the coordination of project activities. Regular monitoring visits were also carried out by Oxfam to ensure smooth implementation.

The project was intended to benefit up to 1,200 households in Banibangou and Soumatt through these interventions. With support from the programme, the beneficiaries were expected to increase their agricultural output, produce higher-value goods and reach more markets for their produce.

EVALUATION APPROACH

The review adopted a quasi-experimental impact evaluation design, which involved comparing households that had been supported by the project with households in neighbouring communities that had not been supported, but which had similar livelihoods characteristics in 2012 before the project was implemented.

The Effectiveness Review was carried out in four villages (two project villages and two comparison villages) in the commune where the project activities had been implemented. Households that had participated in the project were selected at random to be interviewed. For comparison purposes, interviews were carried out with farmer households from the two villages that had not participated in the project, but who had been eligible and had expressed an interest in doing so. These villages did not participate because the project did not have enough funds to cover all villages.

The comparison villages were selected purposively because they were deemed to have had similar characteristics to the implementation villages at baseline. Households in these villages were randomly selected and interviews were conducted. In total, 300 project participants and 404 non-participants were interviewed. At the analysis stage, the statistical tools of propensity score matching (PSM) and multivariate regression were used to control for demographic and baseline differences between the households surveyed in the project and comparison areas, in order to increase confidence when making estimates of the project's impact.

RESULTS

The data suggest that the project interventions made a positive contribution to the livelihoods of the target population.

One of the key questions for this review was to determine whether the project had an impact on household income and food security. In this study, 'total household consumption per adult equivalent per day' has been used as a proxy measure of net household income. Project participants had a higher overall household income compared with the matched comparison group, and this difference was statistically significant. Moreover, there is some evidence to suggest that the project had a positive effect on food security, with project beneficiaries spending more on food (food consumption per adult equivalent per day) compared with the comparison households.

In contrast with the results for food consumption and household income, there was no evidence to suggest that the project had a significant effect on household wealth (measured by normalized wealth index). It could be reasoned from the food consumption data that the project beneficiaries spent some of their increased income on purchasing food, rather than investing in assets. It is also possible that the project participants were investing in inputs for their vegetable business, and hence the increased income did not translate into assets. While interpreting the results, it should also be kept in mind that savings and subsequent asset creation can also depend on cultural, social and political contexts. Changes in wealth status may require a much higher income or longer time horizon in this particular context to become apparent.

Another key aspect of this Effectiveness Review was to determine whether the project had any effect on production, sales or revenue from the key vegetable crops it targeted. While there is evidence to show that project households produced more vegetables than their comparators, the review team was limited in the analysis and conclusions it could draw regarding sales and revenue by some shortcomings in the data collected.

An important finding was that, of the households who sold any produce, almost all reported selling some of it to local markets or middlemen. Among project households, none reported selling their produce through farmers' associations or cooperatives. This suggests that there are some questions regarding the effectiveness of the associations – at least in regards to collective marketing and selling.

The review also considered whether farmers adopted improved agricultural practices acquired through the capacity-building training offered by the project. These agricultural practices included seed nurseries, production of organic compost, organic farming, use of improved certified seeds/seedlings, integrated diversified farming systems and farm planning based on weather forecasts (e.g. rain gauges). The results indicated that almost all project households had implemented at least one of the improved practices in the 12 months prior to the survey, compared with 32 percent of comparison households. Indeed, project households implemented on average more than three of the six practices considered in the survey, compared with an average of less than one practice implemented by comparison households.

Although this Effectiveness Review was focused on livelihoods, it included some indicators to evaluate the project's impact on women's participation in group activities and their influence in household decision making. While there was no apparent evidence of the project positively affecting group participation, there was evidence of it effecting a positive change in women's influence in the household decision-making process.

The key results of the Effectiveness Review are summarized in Table 1.

Table 1: Key results of the Effectiveness Review

Outcome	Evidence of positive impact	Comments
Adoption of improved agricultural practices and technology	Yes	On average, farmers in the project areas adopted more improved agricultural practices/technologies compared with non-project areas.
Increased access to markets	Unclear	As so few comparison households sold the targeted crops, there are no reliable comparison data to draw robust conclusions. However, the data for the project households show no households selling their crops through farmer associations or cooperatives.
Households engaging in vegetable production	Yes	There is evidence that farmers in the project areas cultivated more of the vegetable crops targeted by the project.
Increased production of vegetables	Yes	There is evidence that project households harvested a significantly greater amount of vegetable crops than their comparators.
Increased revenue from vegetables	Unclear	As so few comparison households sold the targeted crops, there are no reliable comparison data to draw robust conclusions.
Overall crop diversity	Unclear	While project households were more likely to cultivate a greater number of vegetable crops, due to the limitations of the data collected (no information collected on crops not targeted by the project) it is not possible to draw conclusions as to the overall effect this had on crop portfolios among the project households.
Wealth index	No	No evidence of impact was found for changes in wealth status among project households.
Women's empowerment	Mixed	There is mixed evidence of the project having a positive effect on participation in groups, but there is stronger

		evidence that the project effected a positive change in women's influence in the household decision-making process.
Overall household income (global indicator)	Yes	The Effectiveness Review measured income indirectly through total household consumption per adult equivalent per day expressed as a logarithm of the local currency (CFA franc in this case). Households in project areas had an overall household income that was approximately 22 percent higher compared with households in the comparison communities.

PROGRAMME LEARNING CONSIDERATIONS

Continue to work on efforts to improve collaboration of production and marketing of vegetables in and around the project areas

While the review indicates success in encouraging households to diversify into cultivating a range of vegetables, it is also apparent that the majority of project households sell their goods to middlemen or local markets. The project team should consider whether there is opportunity for farmers to better collaborate on production of certain cash crops, and in turn explore options for marketing these goods in a coordinated way to achieve better returns for the farmers. Part of the strategy could include strengthening the organisation of local farmer groups and encouraging local leadership to represent their groups in potential markets.

Evaluate options for how to add value in the value chains targeted by the project

Clearly the project has been successful in encouraging greater production in the value chain targeted by the project, i.e. vegetable production. This should offer a key opportunity for the project team to consider how to maximise this increase in production by evaluating options for adding value to the produce. This may include improving opportunities for farmers to better preserve their crops, whether to take advantage of market fluctuations, or in producing dried end product, for example, dried tomatoes. There may also be opportunities to further process some of the crops produced to add value. An example given by the team was the production of onion jam, but there are likely other options, and as farmers organise collectively there may be options to invest in appropriate value-addition machines.

Review the key successes of the project and explore opportunities to replicate elements

The results from the review indicate the project has been largely successful in its key objectives of encouraging implementation of better vegetable production with the aim of increasing income among the target households. The steps taken to train farmers and implement irrigation and other initiatives to improve water availability seem to have borne good results. One of the more striking results from the review is how women from project households exhibit greater decision-making power at a household level. The project targeted women from the outset, including them in decisions in how the project should operate and which crops should be targeted. In the local culture, women often have the main responsibility for garden/vegetable production, and by purposely targeting such crops the project appears to have brought benefits to the women participants beyond improvements to household income and food security. There seems to be opportunity then here to review and further understand the reasons for this

indicator of success in women's empowerment, and evaluate whether there are opportunities to replicate the success of this elsewhere.

1 INTRODUCTION

Oxfam GB has put in place a Global Performance Framework (GPF) as part of its effort to better understand and communicate its effectiveness, as well as to enhance learning across the organization. This Framework requires project and programme teams to report output data annually across six thematic indicator areas. In addition, every year, for each thematic indicator area a small sample of mature projects is randomly selected to be evaluated through rigorous Effectiveness Reviews.[1] One key focus of the reviews is on the extent to which the projects have promoted change in relation to relevant Oxfam GB global outcome indicators. The global outcome indicator for the livelihoods thematic area is defined as total household consumption per adult equivalent per day. This indicator is explained in more detail in section 5 of this report.

This Effectiveness Review, which was conducted in November 2016, was intended to evaluate the success of the 'Community-Based Integrated Water Resource Management' project in Banibangou commune in the Republic of Niger in promoting food security and improving agricultural production and marketing among project participants. This was to be made possible through integrated water management in the project areas.

This project was implemented in two villages in Banibangou commune – Banibangou and Soumatt – between April 2013 and March 2015 by Oxfam in conjunction with a local organization, Karkara[1], and the Department of Agriculture of Niger. The project was intended to benefit up to 1,200 households in the two villages through interventions designed to increase agricultural production and in turn generate increased sales and revenues. With support from the programme, beneficiaries were expected to increase their output, produce higher-value goods and reach more markets for their produce. In general the project aimed to ensure improved quality and quantity of agricultural produce, through increasing the availability of water for agricultural production. The goal was the development of sustainable livelihood options for households in the project areas.

This report presents the findings of the Effectiveness Review. Section 2 briefly reviews the project's aims and objectives. Section 3 describes the evaluation design used, and section 4 describes how this design was implemented. Section 5 presents the results of the data analysis, based on the comparison of outcome measures between the intervention and comparison groups. Section 6 concludes with a summary of the findings and some considerations for programme learning. Finally, baseline statistics before matching are provided in Appendix 1, technical and methodological considerations on the propensity score matching (PSM) approach are given in Appendix 2, and tests of the robustness of the results are examined in Appendix 3.

[1] Karkara is a Niger-registered NGO established in 1992, who have extensive experience in sustainable agriculture, food security, integrated water-resource management and advocacy.

2 PROJECT DESCRIPTION

2.1 PROJECT ACTIVITIES

Oxfam embarked on an agricultural programme to contribute to poverty reduction efforts in two villages in the commune of Banibangou, intending to reach smallholder farmers by using an integrated approach to water management. Agriculture and small businesses provide livelihoods for the majority of the population in the commune. It is estimated that at least 90 percent of the population in this region are farmers and small traders.

One of the challenges to improved agricultural production in the area is limited productive land, low level of skills and lack of knowledge of modern agricultural practices and markets in indigenous communities, together with insufficient water sources. Some cash crops, such as vegetables, are grown in the area, but lack of water is a significant limitation to greater commercial production of vegetables. Oxfam and Karkara therefore planned initiatives aimed at increasing the production of vegetables, so that people could better attain self-sufficiency in food production and increase their revenues in order to boost household incomes.

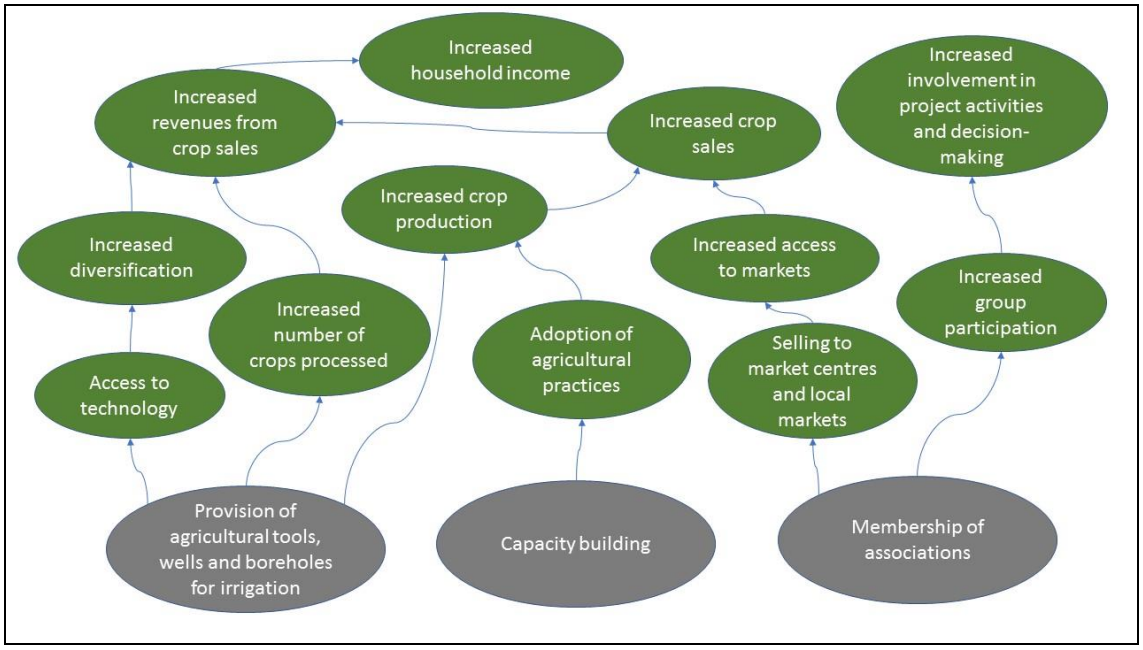
The programme covered the two villages of Banibangou and Soumatt in Banibangou commune and aimed to improve the livelihoods of almost 1,200 households in total. It was implemented from April 2013 to March 2015. The vegetables and other crops selected as the focus of improved production efforts were cabbages, tomatoes, onions, carrots, potatoes, peppers, garlic, anise, okra, lettuce and squash. The role of Karkara was to provide certified vegetable seeds for planting, as well as agricultural tools and equipment. It also mobilized the community and provided training in conjunction with the Department of Agriculture. Oxfam provided the funds and coordinated implementation of the project activities.

The activities implemented included the provision of cultivation, harvesting and processing tools such as slashers, hoes, jembes (a kind of hoe), pangas (a kind of machete), rakes, watering cans and sacks. Certified seeds were provided, together with technical assistance on farming techniques and market information for farmers. In the area of water management, wells and boreholes were dug. Water tanks were bought and installed with solar pumps. Water basins were connected with pipelines to the communities in order to provide water for irrigation. Integrated water resource management plans were also developed and the communities were trained on how to construct toilets and latrines. Oxfam and its partners also worked with farmers' groups or associations for effective and efficient management of the project.

2.2 PROJECT LOGIC AND INTENDED OUTCOMES

This section describes how the project was intended to achieve its goals. Using existing documentation about the project as well as through discussions with the implementing team, the intended hypothetical causal links can be mapped out from project activities, via outputs and intermediate outcomes, to overall changes in household income. Figure 2.1 presents a simplified version of the project's logic or theory of change.

Figure 2.1: The project's simplified logic model and impact



Project logic plays an important role in the design of quantitative evaluations: it is an explicit theory or model of how the project is meant to cause the intended or observed outcomes. It identifies project resources, project activities and intended project outcomes, and specifies a chain of hypothetical causal assumptions linking programme resources, activities and intermediate outcomes and ultimate project goals.

Project logic is important due to the recognition that a project's success or failure can be assessed only with a clear understanding of the problem it was intended to address, the rationale for choosing a particular approach and how the project was expected to operate.

A central element of many quantitative methodologies is the specification of hypotheses that can be tested through experimental or quasi-experimental designs. Sometimes an evaluation will test a single hypothesis. However, in this case, the evaluation is based on specifying and testing a causal chain of interlinked hypotheses.

The project was expected to increase household income and food security by improving vegetable production and marketing through its interventions. Increases in production were expected to be delivered through the training of farmers in modern agricultural practices, the provision of inputs such as seeds and farming tools, and more coordinated marketing of produce through farmers' associations.

The construction of water points such as wells, boreholes and irrigation systems was intended to provide more reliable and accessible sources of water which in turn could boost the production of crops. This could also lead to an increase in the number of crops being cultivated, potentially strengthening food security and increasing revenue. At the same time, membership of associations aimed to ensure that farmers had better and more productive access to markets through collective marketing power.

Attempts were also made to increase the participation of community members in the project, especially women. With increased membership and participation in groups, women were expected to influence decisions affecting the agricultural activities in which they were involved.

Based on this project logic, the Effectiveness Review sought to answer the following key evaluation questions:

- Did the farmers adopt modern methods of crop production and agricultural technology ?
- Was there increased access to markets for farmers involved in the project?
- Was there a difference between participants and non-participants in the quantities of the various kinds of vegetables produced and sold and the revenues obtained from sales?
- What was the impact of the project on total crop production, sales and revenues?
- What was the effect of the project on the overall household income of the participants?
- Were there any affects on women's empowerment, either at a community or household level?

3 EVALUATION DESIGN

The central problem in evaluating the impact of any project is how to compare the outcomes that result from that project with what would have been the case without that project having been carried out. In the case of this Effectiveness Review, information about the lives and livelihoods of project participants was collected through a household questionnaire – but clearly it was not possible to observe what their situation would have been had they not had the opportunity to participate in the project. In any evaluation, that 'counterfactual' situation cannot be directly observed: it can only be estimated.

In the evaluation of programmes that involve a large number of units (whether individuals, households or communities), it is common practice to make a comparison between units that were subject to the programme and those that were not. As long as the two groups can be assumed to be similar in all respects except for the implementation of the specific project, observing the situation of those where the project was not implemented can provide a good estimate of the counterfactual.

An ideal approach to an evaluation such as this is to select at random the areas in which the project is to be implemented. Random selection minimizes the probability of there being systematic differences between the project participants and non-participants, and so maximizes the confidence that any differences in outcomes are due to the effects of the project.

In the case of the project examined in this Effectiveness Review, the selection of the villages involved in the project was not made at random; in fact, villages were deliberately chosen based on them being particularly vulnerable in terms of the low quantity and quality of agricultural production among community members, lack of agricultural inputs, poor use of modern methods of crop production, low revenues and lack of access to markets and water sources.

However, discussions with the implementation staff revealed that there were neighbouring villages with farmers who would have been eligible and had the potential to benefit from participation in the project, but who were not given the opportunity to participate. These farmers, therefore, did not benefit from the project even though they had livelihoods characteristics similar to those of the project participants at baseline. This happened because the project did not have enough funds to cover all the villages. This allowed a 'quasi-experimental' evaluation approach to be adopted, in which the situation of farmers in non-project villages was assumed to provide a reasonable counterfactual for the situation of farmers who had participated in the project activities.

To improve the confidence in making this comparison, households in the project villages were 'matched' with households with similar characteristics in the non-project (or 'comparison') villages. Matching was performed on the basis of a variety of characteristics – including household size, level of education and indicators of material well-being, such as housing conditions and ownership of assets. Since some of these characteristics may have been affected by the project itself (particularly those relating to wealth indicators), matching was performed on the basis of these indicators *before* the implementation of the project. Baseline data were not available, and so survey respondents were asked to recall some basic information about their household's situation from 2012, before the project was implemented. While these recall data were unlikely to be completely accurate, this should not have led to significant bias in the

estimates as long as the measurement errors due to the recall data were not significantly different for the project participants and the comparison group.

Recall survey data provided a variety of baseline household characteristics on which matching could be carried out. These characteristics were used to calculate a 'propensity score', which is the conditional probability of the household being a participant, given the set of observable characteristics at the baseline. Project households and comparison households were then matched based on their having propensity scores within certain ranges. Appendix 2 provides a more extensive explanation of the matching procedure and tests carried out after matching to assess whether baseline characteristics were similar between the two groups.

As a check on the results derived from the propensity score matching (PSM) process, results were also estimated using multivariate regression models. Like PSM, multivariate regression controls for measured differences between intervention and comparison groups, but it does so by isolating the variation in the outcome variable explained by being a project participant, after the effects of other explanatory variables have been accounted for. The regression models tested are described in Appendix 3.

It should be noted that both PSM and multivariate regression rely on the assumption that the 'observed' characteristics (those that are collected in the survey and controlled for in the analysis) capture all of the relevant differences between the two groups. If there are 'unobserved' differences between the groups that matter for project participation, then estimates of outcomes derived from them may be misleading. Unobserved differences between the groups could potentially include differences in attitudes or motivation (particularly important when individuals have taken the initiative to participate in a project), differences in community leadership or local-level differences in wealth or other contextual conditions faced by households.

4 DATA COLLECTION

4.1 SAMPLING APPROACH

The project was implemented in the commune of Banibangou in the two villages of Banibangou and Soumatt. The intervention group consisted of farmer households cultivating vegetables who participated in the project from inception to closure. The intervention group was selected based on its vulnerability, as explained earlier. Since it was not possible to include all villages due to resource constraints, some villages within the Banibangou commune did not take part in the project activities, even though they were also vulnerable. This group therefore formed the comparison group.

Lists of participating households in the sampled villages of Banibangou and Soumatt were obtained from the partners. The number of farmers to be selected from each of the villages was determined by dividing the number of members in each of the villages by the total participants in the selected villages and multiplying by the sample size that had been determined for the intervention group. The result provided the number of farmers to be interviewed from each of the two intervention villages, as shown in Table 4.1. The sample frame was constructed by identifying those households that received at least one project intervention. Among the households that benefited directly from the project, 300 were randomly selected to be interviewed.

There were villages in Banibangou commune that did not participate in the project but could have participated, since they were similar to those that did take part. The project staff and partners identified two villages in Banibangou – Garbey and Gossou – that did not participate in the project and had livelihoods characteristics similar to those of the project participants. These villages therefore provided a potentially good comparison group, to the best knowledge of the project staff. A total of 404 households from these villages (also shown in Table 4.1) were randomly selected to be interviewed.

The comparison villages were selected purposively. Households were, however, randomly selected from the comparison villages using the random walk sampling method. This involves first selecting a starting point where an interviewer would begin the interviews, observing the following rules: starting points were chosen using a landmark, such as a church, school or road junction or any other identifiable mark; starting points were spread all over the sampling area and were not close to each other; no road was ‘walked on’ for a second time until all the roads of the village had been ‘walked on’; if a selected starting point did not fulfil one of the above requirements, the next road had to be selected, and if this was not suitable, the next had to be chosen, and so on, until the right point was reached. The interviewer would begin by going to a starting point provided by the field supervisor. As the enumerator walked along the road, they sampled at systematic intervals defined by the supervisor (say, after every five households depending on the distribution of households in the village). This method was employed because there were no household lists in comparison villages.

A household questionnaire was developed by Oxfam staff, in collaboration with partners, to capture data on various outcome measures associated with the project’s activities. Demographic data and recalled baseline data were also collected to statistically control for differences between the supported and comparison households that could not plausibly be affected by the project. The questionnaire was pre-tested by

local Oxfam staff and then by the enumerators during a practice exercise, and revised accordingly. Data collection involved the use of mobile devices using SurveyCTO software. This software uses an Excel platform where the survey is developed, uploaded onto the server and then downloaded onto the mobile devices. The advantage of this method is that data collection can be done offline, and at the end of the day the completed records can be sent to the server online. The use of mobile devices reduces the cost and time of data collection and, more importantly, reduces the errors associated with data entry.

The enumerators participated in a three-day training workshop, which was led by Oxfam staff. The first and second days of the workshop involved training using the paper questionnaire and mobile phone devices respectively. The third day of the workshop involved a piloting exercise, where a community in Banibangou commune was identified. Following this exercise, the performance of each of the enumerators was reviewed individually before their appointments were confirmed.

The consultant and staff were also trained on focus group discussion (FGD) methodologies. An FGD guide was developed and the consultant and two supervisors were taken through the guide. This allowed them to conduct an FGD among the project beneficiaries. The findings from the FGD have been used to explain some of the quantitative findings in this report.

The field supervisor had the overall responsibility for sending completed records to the server at the end of each day. The full list of villages with numbers of households/farmers interviewed in intervention and comparison villages is shown in Table 4.1.

Table 4.1: Intervention and comparison groups sample sizes

	Project participants			**Sample comparison group**		
Commune	Villages/ farmer associations randomly selected from intervention communities	Households/ farmers participating in the project	Households /farmers interviewed	**Commune**	Villages/farmer associations selected in comparison communities	Households/ farmers interviewed in comparison communities
Banibangou	Banibangou	320	147	Banibangou	Garbey	200
	Soumatt	392	153		Gossou	204
Total		**712**	**300**			**404**

4.2 ANALYSIS

Households of project participants and non-participants were compared in terms of their demographic characteristics, livelihoods activities and economic situations in 2012. These data were based on information either recalled during the questionnaire or reconstructed from the household composition at the time of the survey.

The full comparison is shown in Appendix 1. Some important differences were found between the project participants and non-participants. For example, on average, household sizes of participant households were larger than those of non-participants. The proportion of wealthier households (those in the wealthiest 25 percent in terms of an index of wealth indicators) was much higher in the project villages than in the non-project villages. Conversely, comparison households were, on average, located closer to the nearest market than the average household in the project villages. In addition, household heads from comparison households were more likely to have completed secondary education, while project households had a greater proportion of members with primary education than their comparators.

Any differences between project and comparison households that existed before the project commenced had the potential to bias comparisons of its outcomes between project and comparison respondents. Therefore, the review team tried to control for these baseline and demographic differences when making such comparisons. This was especially important for wealth and women's membership of groups, which could be regarded as potential outcomes of the project: it was hoped to establish whether the project affected such outcomes, rather than there simply being differences between the project participants and the comparison group.

Some of the differences between the project participant households and the comparison group identified above may be down to recall error. However, this would require the project participants to have systematically overstated their wealth and group membership and/or the comparison households to have systematically understated. There are two potential reasons why this might have happened in the survey. In particular, it may have been difficult for project respondents to remember back to a time before any project activities, so their recall answers may include some mix of their baseline status and the effects of the project.

However, in the absence of these types of systematic bias for the project participant households and comparison households, any measurement error that arises due to recall would actually lead to any differences between project and non-project households being *underestimated*.[2] Thus, it is unlikely that there are truly no differences between project households and comparison group households in terms of wealth or group participation (in 2012). As such, it is important to control for these differences in the analysis.

As described in section 3, the main approach used in this Effectiveness Review to control for baseline differences was propensity score matching (PSM). The variables on which respondents were matched were selected from among the full list detailed in Appendix 2, based on two key factors. Firstly, those variables were selected that were thought to be the most significant in influencing respondents' participation in the project. Secondly, the review aimed to include variables that could affect potential project outcomes as well as the likelihood of participating in the project.

After matching, project participant households and comparison households appeared to be reasonably well balanced in terms of each of the selected variables. One caveat is that 15 of the 300 project households and four of the 404 comparison households in

the sample had to be dropped from the analysis. The exclusion of 19 observations at this stage of the analysis is unlikely to make a substantial difference to the main findings, however.

All the results described in section 5 of this report were tested for robustness by estimating them with several alternative statistical models, including alternative PSM models and linear or probit regression models. These robustness checks are shown in Appendix 3. However, the results of the alternative PSM and regression models generally produced estimates of outcomes that were similar in magnitude and statistical significance to those derived from the original PSM model. The few cases where the models produced divergent results are discussed in section 5, in the text or in endnotes.

As mentioned in section 3, PSM and regression models can only control for the baseline differences between project and comparison households for which data were collected in the survey. If there are any 'unobserved' differences between the two groups – such as individuals' attitudes or motivation, differences in local leadership, or weather or other contextual conditions – then these may bias the estimates of outcomes described in section 5. The evaluation design and the selection of respondents were intended to minimize any potential for unobserved differences, but this possibility cannot be excluded and must be borne in mind when interpreting the results.

5 RESULTS

This report is intended to be free from excessive technical jargon, with more detailed technical information being reserved for the appendices and endnotes. However, there are some statistical concepts that cannot be avoided when discussing the results. In this report, results are usually stated as the average difference between the project households (referred to as the 'intervention' group) and the matched non-project households (the 'comparison' group). In the tables of results on the following pages, statistical significance is indicated with asterisks, with three asterisks (***) indicating a *p*-value of less than 1 percent, two asterisks (**) indicating a *p*-value of less than 5 percent and one asterisk (*) indicating a *p*-value of less than 10 percent. The higher the *p*-value, the less confident the review team could be that the measured estimate reflects the true impact, as opposed to simply random variation in the data. Results with a *p*-value of more than 10 percent are not considered to be statistically significant.

5.1 INTRODUCTION

This section presents a comparison of the households interviewed in project and comparison communities in terms of various outcome measures relating to the project under review. As described above, asterisks are used in the results tables to indicate where the differences are statistically significant, to at least the 10 percent significance level.

The results are shown after correcting for observed baseline differences between the households interviewed in the project communities and those in the comparison communities, using a propensity score matching (PSM) procedure. This means that when differences are reported in the means for the intervention group outcomes and comparison group outcomes, this is for the matched sample. More information about the procedure applied is found in Appendix 2. All outcomes discussed here have also been tested for robustness with alternative statistical models, as described in Appendix 3; where those alternative models produce markedly different results from those shown in the tables in this section, this is discussed in the text or in endnotes.

There is a key limitation to this analysis, which has been described above, but is repeated here because it affects the interpretation of the results. There may be 'non-observable' differences between the project participants and comparison households – such as individuals' attitudes or motivation or differences in local leadership, weather or other contextual conditions. If these unobserved differences also influence the potential outcomes considered in this section, then the estimates of the project's effects will be biased. This possibility must be borne in mind when interpreting the results.

5.2 INVOLVEMENT IN PROJECT ACTIVITIES

The first step in understanding the project's impact is to examine the extent to which respondents reported having received the types of support and having participated in the various activities implemented by the project.

Table 5.1 shows the differences between intervention and comparison households in terms of the training in which they participated during the life of the project. There are positive and significant differences in favour of the project across all of the different trainings considered in the survey.

Some of the largest differences are for training on vegetable production, where 76 percent of project households reported receiving such training compared with 20 percent of non-project households, as well as training on use of quality seeds (73 percent compared with 19 percent) and fertilizers (76 percent compared with 19 percent). This is perhaps not surprising given that improved agricultural production was a key priority for the project. However, positive and large differences can also be seen in favour of the project households in terms of training on hygiene and nutrition, and some smaller – but still statistically significant – differences in receipt of training on water resource monitoring and agricultural processing/storage techniques.

Table 5.1: Training received by farmer households

	1	2	3	4
	Training on vegetable production (%)	**Training on hygiene and malnutrition prevention (%)**	**Training on water resource monitoring through water and rainfall monitoring systems (%)**	**Training on use of quality seeds (%)**
Intervention group mean	0.76	0.80	0.27	0.73
Comparison group mean	0.20	0.40	0.10	0.19
Difference:	0.56***	0.40***	0.17***	0.54***
	(0.04)	(0.05)	(0.04)	(0.05)
Observations (intervention group)	285	285	285	285
Observations (total)	685	685	685	685

	5	6	7
	Training on use of fertilizers (%)	**Training on use of pesticides (%)**	**Training on food processing and storage techniques (%)**
Intervention group mean	0.76	0.59	0.32
Comparison group mean	0.19	0.15	0.12
Difference:	0.57***	0.44***	0.19***
	(0.05)	(0.05)	(0.04)
Observations (intervention group)	285	285	285
Observations (total)	685	685	685

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

Table 5.2 reports the differences between project and comparison communities in terms of particular activities carried out in the community since 2012 (based on households' responses). Similar to the results presented for households' receipt of training, there are large and statistically significant differences between project and comparison households in each of the activities considered in the questionnaire.

Firstly, project households were far more likely to receive agricultural inputs, such as seeds (89 percent of project households) and tools (84 percent). In each of these cases only around a quarter of comparison households reported receiving such inputs. There is further evidence that project households received greater technical support on modern methods of farming (47 percent of project households, compared with just 7 percent among comparison households), and were more likely to attend farmer field days and exhibitions.

Table 5.2: Activities carried out in the community

	1	2	3	4
	Household members provided with seeds (%)	**Household members provided with agricultural tools (%)**	**Household members provided with cash for work (%)**	**Household members provided with tech. support on modern methods of farming (%)**
Intervention group mean	0.89	0.84	0.52	0.47
Comparison group mean	0.27	0.25	0.26	0.07
Difference:	0.62***	0.59***	0.27***	0.39***
	(0.05)	(0.05)	(0.05)	(0.04)
Observations (intervention group)	285	285	285	285
Observations (total)	685	685	685	685

	5	6	7	8
	Household members had access to sand dams (%)	**Household members had access to water wells (%)**	**Household members had access to irrigation services (%)**	**Household members attended farmer field days (%)**
Intervention group mean	0.79	0.23	0.54	0.22
Comparison group mean	0.22	0.04	0.06	0.03
Difference:	0.58***	0.19***	0.48***	0.19***
	(0.05)	(0.03)	(0.03)	(0.03)
Observations (intervention group)	285	285	285	285
Observations (total)	685	685	685	685

	9
	Household members exposed to agricultural exhibitions (%)
Intervention group mean	0.21
Comparison group mean	0.03
Difference:	0.18***
	(0.03)
Observations (intervention group)	285
Observations (total)	685

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

Taken together with the results on training, these results provide compelling evidence that project households were well served with agricultural-related training and inputs from the project.

There is also evidence that project households had greater access to water improvements implemented by the project. Almost 80 percent of project households had access to sand dams (22 percent of comparison households) and 23 percent had access to water wells (4 percent of comparison households). Particularly striking are the results for access to irrigation services, where more than 50 percent of project households reported having such access, compared with just 6 percent of comparison households. Column 3 of Table 5.2 also shows that members of project households were more likely to benefit from a 'cash-for-work' scheme implemented by the project.

Therefore, taking the results from this section together, the intervention group – as expected – clearly participated in more activities implemented by the project than the comparison group.

5.3 ADOPTION OF AGRICULTURAL PRACTICES

This section considers one of the next 'links' in the project's theory of change, or logic model. It examines whether, having received the training and inputs detailed in section 5.2, project households were more likely to implement some of the 'modern' or improved agricultural practices.

Column 1 of Table 5.3 shows that almost all project households (93 percent) had implemented at least one of the improved practices in the 12 months prior to the survey, compared with 32 percent of comparison households. Column 2 shows that, on average, project households implemented more than three of the six practices considered in the survey, compared with an average of less than one practice implemented by comparison households. With again such a large difference between project and comparison households, this provides evidence that the project effected a positive change in household agricultural behaviour.

In terms of the specific practices reported in columns 3–8, it can be seen that approximately three-quarters of project households reported implementing a seed nursery and producing organic materials for cultivation, compared with just one-quarter of comparison households. Large and positive differences between project and comparison households were also detected for whether the household practised organic farming, used improved seeds or implemented a diversified farming system. Column 8 shows that, while there are still statistically significant differences between project and comparison households for whether farm planning was based on weather forecasts, only 6 percent of project households reported doing so.

However, overall the results in Table 5.3 provide evidence that the project had a positive impact on the implementation of improved agricultural practices at a household level.

Table 5.3: Adoption of improved agricultural practices

	1	2	3	4
	Household adopted at least one practice (%)	**Number of improved practices adopted**	**Household practised use of seed nurseries (%)**	**Household practised production of organic materials (%)**
Intervention group mean	0.93	3.24	0.77	0.73
Comparison group mean	0.32	0.82	0.27	0.28
Difference:	0.60***	2.42***	0.50***	0.45***
	(0.05)	(0.16)	(0.05)	(0.05)
Observations (intervention group)	285	285	285	285
Observations (total)	685	685	685	685

	5	6	7	8
	Household practised organic farming (%)	**Household practised use of improved seeds (%)**	**Household practised use of integrated diversified farming system (%)**	**Household practised farm planning based on weather forecasts (%)**
Intervention group mean	0.51	0.66	0.51	0.06
Comparison group mean	0.08	0.04	0.14	0.01
Difference:	0.42***	0.62***	0.37***	0.05***
	(0.04)	(0.03)	(0.04)	(0.02)
Observations (intervention group)	285	285	285	285
Observations (total)	685	685	685	685

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

5.4 CROPS

This section considers the project's impact on crop cultivation – or more specifically, vegetable cultivation. As this was a key intended outcome of the project, it is important to evaluate any evidence of impact on this measure.

One of the weaknesses of the survey design was that only vegetable cultivation was considered in the questionnaire. Unfortunately, no questions were asked regarding cultivation of key staple crops – which was perhaps due to the project's focus on increasing production of vegetables. This meant, however, that there was great difficulty in finding appropriate 'matches' from the pool of comparison households, due to so few comparison households cultivating vegetables at the time of the survey. This resulted in a very small sample size from the comparison households – particularly for the analysis of overall sales of vegetable crops. With such few comparators (just 30 comparison households), it is not possible to be confident in presenting results connected to sales or revenues from the selected crops.

In hindsight, not collecting harvest and sales information for staple crops was an error. Having such information – across all crops – would have allowed greater analysis of sales and revenue, such as whether households had generated more income from agriculture, and the degree to which the project encouraged a shift away from staple crop cultivation into vegetable production.

For the reasons set out above, the conclusions that can be drawn from the production and sales data collected in the survey are unfortunately limited.

Table 5.4 presents the summary information on crop cultivation. As already mentioned, no staple crops were included, but the following 14 crops were considered in the survey: cabbage, tomato, onion, carrot, potato, aubergine, lettuce, courgette, garlic, anise, okra, moringa, pepper, and chilli pepper.

Table 5.4: Selected crop cultivation analysis

	1	2	3	4
	Number of selected crops farmed in last 12 months	**Number of selected crops sold in last 12 months**	**Total kilograms (kg) harvested in last 12 months**	**Total kg harvested in last 12 months – log**
Intervention group mean	6.34	2.37	2,653.56	6.10
Comparison group mean	1.93	0.26	164.32	4.54
Difference:	4.41***	2.11***	2,489.24***	1.60***
	(0.35)	(0.18)	(585.72)	(0.18)
Observations (intervention group)	285	285	285	275
Observations (total)	685	685	685	402

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

Column 1 of Table 5.4 shows that project households cultivated on average six of the crops considered in the survey. This compares with an average of just under two crops among comparison households. This difference is statistically significant, indicating that the project had a positive impact on the cultivation of the vegetable crops targeted by the project. Indeed, with such high average figures for the number of crops, it seems that project households have very much diversified into cultivating a range of crops. Column 2 shows that project households sold an average of just over two of the targeted crops.

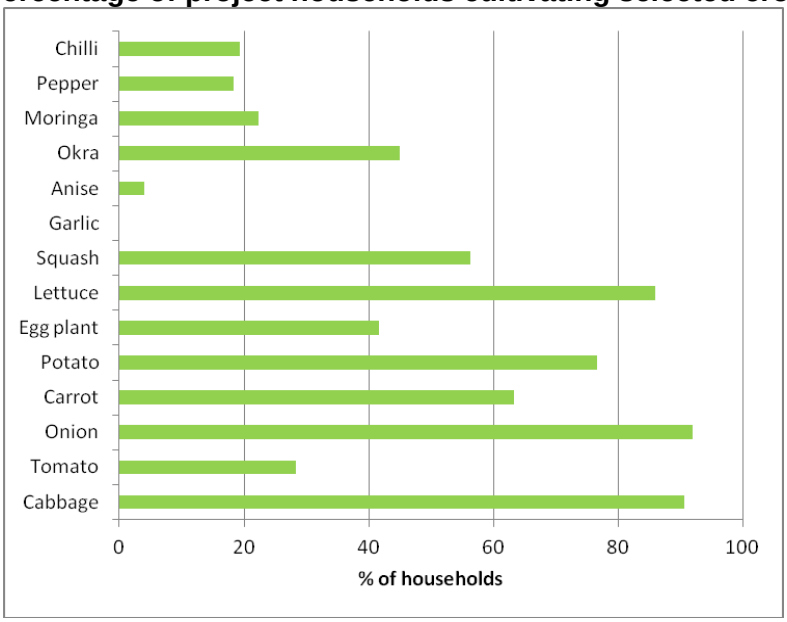
Table 5.4 also presents both the raw average (column 3) and the log analysis (column 4) of the amount of target crops harvested in the 12 months prior to the survey. The raw average is presented for information only – no assumptions can be drawn from these results. The log analysis is more reliable as it reduces the effects of extreme outliers in the data, which are likely to be due to the data being estimates from the respondents. Column 4 shows that project households were more likely to harvest a significantly larger amount of the selected crops in the year prior to the survey.

As so few comparison households cultivated the selected crops, reliable comparison data cannot be presented for the individual crops, but Figure 5.1 presents descriptive data showing the proportion of households cultivating each of the targeted crops. As can be seen from the figure, cabbages, tomatoes and lettuce were the most popular crops cultivated by the project households, but there is also clearly a whole range of crops being produced.

Again, due to the very small sample of comparison households who sold the selected crops, it is not possible to present comparative data, but the descriptive findings on where project households sold their crops are nevertheless important. Of the 228 project households who reported selling any crops in the 12 months prior to the survey, 62 percent sold to local traders or middlemen, 48 percent sold in local markets and 1 percent sold to institutional buyers, such as hotels or restaurants. None of the project households surveyed reported selling their crops to farmer associations or cooperatives. Strengthening farmer associations to be able to buy and sell produce is a

potential opportunity to develop the project further, and the learning considerations at the end of the report will return to this topic.

Figure 5.1: Percentage of project households cultivating selected crops



5.5 OVERALL HOUSEHOLD INCOME

Measuring household income directly is problematic: self-reported measures of total income are generally regarded as unreliable, given the wide variety of endeavours such populations engage in to generate income. Most households in this sample were engaged in a range of livelihood activities; a direct income measure would therefore have to collect detailed information about the contribution of each of these activities to household income.

For these reasons, the survey did not attempt to collect data on total household income directly. However, there is a widely recognized and strong association between household income and consumption. The Effectiveness Review therefore followed common practice in micro-level socio-economic analysis by considering household consumption and expenditure as an indicator of income.

To that end, respondents were asked to provide detailed information about their recent expenditure on both food and non-food items. Firstly, the respondents were asked, from a list of 30 products, what types of food they had consumed over the previous seven-day period, and the particular quantities. The quantities of each food item consumed were then converted into a monetary value. This was done by asking the respondent how much was paid for the food item in question or – if the food item was from the household's own production – how much it would be worth if it was purchased from the local market. The respondents were also asked how much they spent on particular regular non-food items and services from a list of 19 items, such as fuel, toothpaste and transport fares, over the previous four weeks. Finally, they were asked to estimate the value of other occasional types of expenditure that they had incurred over the previous 12 months from a list of 19 items, which included clothes, medical expenses and home repairs.

The household expenditure measure was calculated by converting each of the expenditure types into a per day per capita figure and adding them together. This figure was then divided by a factor representing household size, to generate a per day per person expenditure figure. As with the measures of agricultural sales, the expenditure

variable has been expressed on a logarithmic scale, to reduce the influence on the overall result of any households with extreme values for total consumption. The comparison of consumption per adult equivalent per day and total household consumption per adult equivalent per day between households supported by the project and comparison households, after logarithmic transformation, is shown in Table 5.5.

It can be seen in column 1 that the value of food consumed within the households of project participants was significantly higher than that of comparison households. The difference is highly significant, indicating confidence that there is a true difference between the project and comparison households that is not just due to random sampling error. The difference in the logarithmic values of 0.18 implies a difference in food consumption between project and comparison households of approximately 18 percent. Similarly for total household consumption, as shown in column 2, there is a strong and positive difference in favour of project households, with total consumption estimated to be 22 percent higher among project households.

These findings indicate that there is strong evidence that the project had a positive effect on household income, as measured by average consumption across the sampled households.

Table 5.5: Household consumption

	1	2
	Food consumption per adult equivalent per day (logarithm of CFA franc)	**Total household consumption per adult equivalent per day (logarithm of CFA franc)**
Intervention group mean	5.58	5.83
Comparison group mean	5.40	5.60
Difference:	0.18***	0.22***
	(0.05)	(0.05)
Observations (intervention group)	285	285
Observations (total)	685	685

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

5.6 HOUSEHOLD ASSET WEALTH

This section explores the project's impact on households' wealth. Wealth may be interpreted in two ways from the perspective of resilience – a very important concept in the context of this project. Firstly, wealth may be seen as a *driver* of resilience, insofar as households can sell off assets in times of crisis, but can also more easily finance the costly investments needed to adapt livelihood strategies and innovate. However, wealth may also be regarded as exactly the type of well-being indicator – a 'final' outcome – that would be improved in spite of shocks, stresses and uncertainty in more resilient households. Typically, these types of final well-being outcome take more time to change than immediate changes in income.

In the questionnaire, respondents were asked to provide information about their household's ownership of various assets (including livestock, productive equipment and household goods), as well as about the conditions of the family's house, both in 2012 and at the time of the survey. This information on asset ownership and housing conditions was used to generate an index of overall household wealth.

The wealth index was generated under the assumption that, if each of the assets and housing characteristics constituted suitable indicators of household wealth, they should be correlated with one another. That is, a household that scores favourably on one particular wealth indicator should be more likely to do so for other wealth indicators. A small number of items that had low or negative correlations with the others were therefore not considered to be good wealth indicators and so were excluded from the index.

A data reduction technique called principal component analysis (PCA) was used to produce two indices of overall wealth, one based on the recalled data from 2012 and one on the household's situation at the time of the survey. In particular, the wealth index is taken directly from the first principal component.[3] PCA enables weights to be assigned to the different assets, to capture as much information as possible from the data. Broadly, PCA assigns more weight to those assets that are *less* correlated with all the other assets, as these carry more information. By contrast, items with *more* intra-correlation are given less weight.

In order to ensure that the same weights were applied to assets for both the recalled wealth index and the wealth index for the time of the survey, data from these two time periods were pooled before undertaking the PCA procedure. This means that changes in wealth can be more easily compared over time. It should also be noted that the wealth index for 2012 is the measure that has been used throughout this analysis to control for baseline differences in wealth status between project and non-project households.

The analysis in this section starts by 'normalizing' the wealth index.[4] This means that the impacts of the project that are reported can be directly understood as the number of standard deviations by which the project improved wealth. This means that the results from this Effectiveness Review can be more easily compared with other similar evaluations.

Table 5.6 estimates the project's impact on wealth in two ways. Column 1 reports wealth for the project and non-project households at the time of the survey, using the regular matching procedure that was used throughout the other tables in this report. Column 2, however, takes a slightly different approach: the differences between wealth at the time of the survey and in 2012 are calculated, and these differences are compared between project and non-project households in the matched sample. For the results in column 2, it is necessary to omit recalled wealth from the matching process.[5]

Table 5.6: Household wealth

	1	2
	Normalized wealth index	**Difference in normalized wealth index**
Intervention group mean	0.25	0.21
Comparison group mean	0.05	0.24
Difference:	0.21**	-0.03
	(0.10)	(0.07)
Observations (intervention group)	290	290
Observations (total)	689	689

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

Although the current wealth index is higher among project households, there are no differences in wealth status between 2012 and the time of the survey. This indicates that there may have been significant differences in the wealth status of the project and comparison households which could not be properly controlled for in the analysis.

Appendix 1 shows that there were significant differences between the project and comparison households at baseline, which strengthens the case for looking at the results in column 2 for a more accurate picture of changes in wealth between 2012 and the time of the survey. Column 2 shows no significant difference in the change in wealth index between the project and comparison households. This finding is supported by the robustness checks in Appendix 3.

Therefore, the data do not provide evidence that the project had a positive impact on overall household wealth as measured by assets. It should, however, be remembered that changes in wealth status may require a longer time horizon in this context to become apparent. As such, it may be useful to follow up this evaluation with future efforts to measure wealth status after a number of years.

5.7 GROUP PARTICIPATION AND DECISION MAKING BY WOMEN

The final section of this review considers women's involvement in community groups and decision making at a household level. As mentioned in section 2, the project aimed to increase community members' participation in the project, especially that of women. With increased membership and participation in groups, women were expected to be able to better influence decisions affecting the agricultural activities in which they were involved.

In the questionnaire, women respondents were first asked questions regarding their participation in certain groups in their community. The groups in question were:

- women's associations
- farmers' associations/groups
- cooperatives
- savings or micro-finance groups
- disaster management groups
- social support groups
- other relevant community groups.

Table 5.7: Women's involvement in community groups

	1	2
	Number of groups women participate in	**Number of groups women are involved in decision making**
Intervention group mean	1.82	0.55
Comparison group mean	1.68	0.49
Difference:	0.14	0.05
	(0.11)	(0.11)
Observations (intervention group)	280	280
Observations (total)	664	664

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

Column 1 in Table 5.7 shows that, on average, women from the project communities were each involved in approximately two groups, which is not significantly higher than women from comparison communities. Column 2 presents the average number of groups in which women had some decision-making responsibility. Again, there is no significant difference in the results for women from the project and comparison communities. It should be mentioned that in the robustness checks in Appendix 3, some of the analytical models suggest that women from project households did attend more community groups, so there may be some evidence that the project affected this measure. However, there is no clear evidence that the project successfully led to an increase in women being involved in group-level decision making.

Table 5.8: Other indicators of women's empowerment

	1	2
	Women disagreeing that standing up for their concerns is intimidating (%)	**Women's self-reported contribution to household needs (%)**
Intervention group mean	0.71	44.7
Comparison group mean	0.63	34.5
Difference:	0.08*	10.2***
	(0.04)	(0.24)
Observations (intervention group)	278	280
Observations (total)	660	664

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

However, Table 5.8 does present some evidence of differences between project and comparison group women in terms of their responses to an opinion statement in the questionnaire. Respondents were asked to agree or disagree with the following statement:

> *Public forums held in your village can be intimidating – it is difficult for a woman like you to stand up and voice any concerns.*

The proportion of women *disagreeing* with this statement is reported in column 1 of Table 5.8. This shows some modest evidence that the proportion of women disagreeing with the statement was higher among the project communities than in comparison communities. It is obviously difficult to tell from the data whether this is a result of women from the project areas feeling more empowered, or whether their community forums are less intimidating. Regardless, this is a positive finding.

Column 2 presents the findings from the following question in the survey:

> *Here are 10 small beans. The beans together represent all the resources your household needs, such as food and money. From what you get, either crop or cash, how many beans represent your contribution?*

For presentation, the result was multiplied by 10 to show an estimated percentage contribution. There are obviously caveats with self-estimated data, but there is evidence to suggest that women from project households are making a greater contribution to overall household needs than their comparators. This may well be connected to the increases in overall household income seen among project households. Nevertheless, it is encouraging to see from these results that it appears that women's personal income – or at least their responsibility for household income – has also increased.

The final aspect of the review considers the involvement of women in household-level decision making. The seven decision-making areas included in the survey were as follows:

- keeping and managing household income
- buying and selling of productive assets (e.g. land and machinery)
- buying and selling livestock
- how much to invest in business activities
- what food to buy and consume
- how children should be educated
- housework and care of family or other community members

For each decision-making area, the respondent was first asked who normally makes most of the decisions about this area (i.e. the respondent herself, her husband, respondent and husband jointly, other household member, etc.). Where the respondent was not responsible for decision making, she was asked to what extent she felt that she could change the decision.

Table 5.9: Women's involvement in household decision making

	1	2	3	4
	Proportion of household decision-making areas where the woman has sole decision-making power/able to change decisions (%)	**Proportion of household decision-making areas where the woman has at least joint decision-making power (%)**	**Women who have sole decision-making power/able to change decision in more than half of household decisions (%)**	**Women who have at least joint decision-making power in more than half of household decisions (%)**
Intervention group mean	0.51	0.82	0.47	0.81
Comparison group mean	0.36	0.73	0.33	0.68
Difference:	0.15***	0.09**	0.14***	0.13***
	(0.05)	(0.04)	(0.05)	(0.05)
Observations (intervention group)	280	280	280	280
Observations (total)	664	664	664	664

Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01; PSM estimates are bootstrapped with 1,000 repetitions.

Column 1 presents the proportion of the household decision-making areas in which the woman respondent either has sole decision-making power or feels that she is able to effect a change in the decision. Overall, women from project households reported this to be the case for half of the decision-making areas covered in the survey. Conversely, women from comparison households reported that they had the same decision-making power for 36 percent of decisions. This difference is significant, indicating that there appears to be evidence that the project effected a change in this area. Looking at the proportion of decision-making areas where women have at least joint decision-making power (column 2), the figure rises to 82 percent for project respondents and 73 percent for comparison respondents. Again, this is a significant difference, meaning that we can be more confident that there is a true difference between project and comparison households in this regard.

Columns 3 and 4 present binary indicators of the same data. Column 3 shows the proportion of women who have sole decision-making power, or are able to change the decisions made, in more than half of the decision-making areas. In this case, 47 percent of women from project households have such power, compared with 33 percent of women from comparison households. Column 4 presents the proportion of women who have at least joint decision-making power in more than half of household decisions – here 81 percent of women from project households reported this to be the case, compared with 68 percent of women from comparison households. Both of the differences reported in columns 3 and 4 are significant, providing further evidence that the project appears to have positively impacted decision-making by women at a household level.

6 CONCLUSIONS

6.1 CONCLUSIONS

The results clearly show that the project households in our survey have significantly higher income and food security – as measured by household consumption - compared to the comparison group. Total consumption and food consumption was estimated to be approximately 22 percent and 18 percent higher respectively among project participants than non-participants, providing evidence that the project has had a corresponding positive effect on overall household income and food security. However, no evidence of changes in the wealth status of respondents – as measured by ownership of assets - was detected.

There is evidence that the project has effected a change in the diversity of crops cultivated by project households, and project households also used or adopted a significantly higher number of 'modern' farming practices. Unfortunately, due to limitations in how the survey was designed, we are unable to draw any conclusions in changes of volume of crops harvested and income generated from their sales.

There is evidence too that women in project households have experienced a greater measure of empowerment, in terms of the indicators used in this study. Firstly, there is evidence to suggest that women from project households are making a greater contribution to overall households needs than women from comparison households. In addition, there is consistent evidence across several indicators that women from project households have more decision-making power at a household level than their comparators.

6.2 PROGRAMME LEARNING CONSIDERATIONS

Continue to work on efforts to improve collaboration of production and marketing of vegetables in and around the project areas

While the review indicates success in encouraging households to diversify into cultivating a range of vegetables, it is also apparent that the majority of project households sell their goods to middlemen or local markets. The project team should consider whether there is opportunity for farmers to better collaborate on production of certain cash crops, and in turn explore options for marketing these goods in a coordinated way to achieve better returns for the farmers. Part of the strategy could include strengthening the organisation of local farmer groups and encouraging local leadership to represent their groups in potential markets.

Evaluate options for how to add value in the value chains targeted by the project

Clearly the project has been successful in encouraging greater production in the value chain targeted by the project, i.e. vegetable production. This should offer a key opportunity for the project team to consider how to maximise this increase in production by evaluating options for adding value to the produce. This may include improving opportunities for farmers to better preserve their crops, whether to take advantage of market fluctuations, or in producing dried end product, for example, dried tomatoes.

There may also be opportunities to further process some of the crops produced to add value. An example given by the team was the production of onion jam, but there are likely other options, and as farmers organise collectively there may be options to invest in appropriate value-addition machines.

Review the key successes of the project and explore opportunities to replicate elements

The results from the review indicate the project has been largely successful in its key objectives of encouraging implementation of better vegetable production with the aim of increasing income among the target households. The steps taken to train farmers and implement irrigation and other initiatives to improve water availability seem to have borne good results. One of the more striking results from the review is how women from project households exhibit greater decision-making power at a household level. The project targeted women from the outset, including them in decisions in how the project should operate and which crops should be targeted. In the local culture, women often have the main responsibility for garden/vegetable production, and by purposely targeting such crops the project appears to have brought benefits to the women participants beyond improvements to household income and food security. There seems to be opportunity then here to review and further understand the reasons for this indicator of success in women's empowerment, and evaluate whether there are opportunities to replicate the success of this elsewhere.

APPENDIX 1: BASELINE STATISTICS BEFORE MATCHING

Table A1.1: Descriptive statistics before matching

	Intervention mean	Comparison mean	Difference	Standard error
Household size	7.07	5.80	1.27***	0.20
Proportion of household members who are children (less than 15 years) (%)	49.07	46.11	2.96	1.72
Proportion of household members who are school age (7–18 years) (%)	31.00	23.23	7.78***	1.57
Proportion of household members who are elderly (more than 65 years) (%)	3.86	3.24	0.62	0.85
Proportion of household members who are male (%)	25.71	26.49	-0.78	1.24
Proportion of household members who have completed primary education (%)	46.17	37.17	9.00***	2.31
Proportion of household members who have completed secondary education (%)	16.27	21.10	-4.83*	2.08
Household head is male (%)	81.33	87.13	-5.80*	2.74
Age of household head (years)	49.88	44.44	5.44***	1.02
Household head completed primary education (%)	21.33	22.28	-0.94	3.15
Household head completed secondary education (%)	3.67	11.63	-7.97***	2.08
Number of minutes it took to walk to the nearest market in 2012	52.11	34.55	17.56***	4.53
Farming crops in 2012 (%)	89.67	88.86	0.81	2.37
Household was in the lowest 20% of the wealth distribution in 2012 (%)	33.33	41.83	-8.50*	3.69
Household was in the second 20% of the wealth distribution in 2012 (%)	1.00	3.96	-2.96*	1.23
Household was in the third 20% of the wealth distribution in 2012 (%)	16.33	22.03	-5.70	3.02
Household was in the fourth 20% of the wealth distribution in 2012 (%)	24.00	16.34	7.66*	3.02
Household was in the highest 20% of the wealth distribution in 2012 (%)	25.33	15.84	9.49**	3.03

Household was in the top 25% of the wealth distribution in 2012 (%)	33.33	18.81	14.52***	3.26
Household participated in any community groups in 2012 (%)	78.00	48.51	29.49***	3.55
Number of community groups in which household participated in 2012 (%)	1.66	0.76	0.90***	0.09

The construction of the wealth index is described in section 5.
Variables dated 2012 are estimates, based on recall data.
* p < 0.1, ** p < 0.05, *** p < 0.01.

APPENDIX 2: METHODOLOGY USED FOR PROPENSITY SCORE MATCHING

The results presented in section 5 of this report were estimated using propensity score matching (PSM). PSM is a statistical technique that allows the effect of an intervention to be estimated by accounting for the covariates that predict receiving the intervention, or 'treatment'. The idea behind PSM is to match similar individuals in the treatment or intervention group to those in the control or comparison group, based on observed characteristics at baseline. After each participant is matched with a non-participant, the average treatment effect on the treated (those who benefited from the intervention) is equal to the difference in average outcomes of the intervention and the comparison groups after project completion. This appendix describes and tests the specific matching procedure employed in this Effectiveness Review. A practical guide on the different approaches to matching may be found in Caliendo and Kopeinig (2008).[6]

Estimating propensity scores

Finding an exact match for treated individuals, based on various baseline characteristics, would be very hard to implement in practice. Rosenbaum and Rubin (1983) demonstrated that a 'propensity score' could summarize all of this information in one single variable.[7] The propensity score is defined as the conditional probability of receiving the intervention given background variables. Specifically, propensity scores are calculated using a statistical probability model (e.g. probit or logit[8]) to estimate the probability of participating in the project, conditional on a set of characteristics.

Table A2.1 shows the variables used to estimate the propensity scores. This reports the marginal effects at the mean and the corresponding standard errors. Following Caliendo and Kopeinig (2008), only variables that influence the participation decision, but are not affected by participation in the project, were included in the matching model. In the table, the dependent variable corresponds to whether or not an individual received the intervention – it is equal to 1 if the household belongs to one of the communities that benefited from the project activities, and 0 otherwise. The coefficients in the table correspond to the marginal effects, i.e. the change in the probability of receiving the intervention if the independent variable is increased by one.

Defining the region of common support

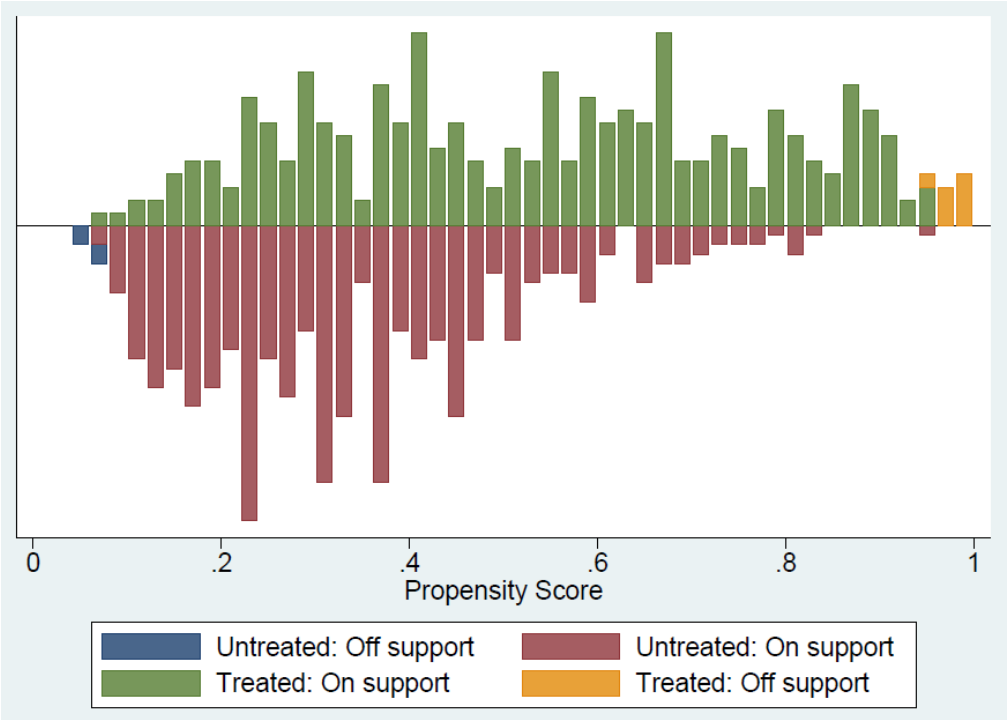
After estimating the propensity scores, it is necessary to verify that there is a potential match for the observations in the intervention group with those from the comparison group. This means checking that there is *common support*. The area of common support is the region where the propensity score distributions of the intervention and comparison groups overlap. The common support assumption ensures that each 'treatment [intervention] observation has a comparison observation "nearby" in the propensity score distribution' (Heckman, LaLonde and Smith, 1999).[9] Figure A2.1 shows the propensity score density plots for both groups. It can be observed that, although the distributions of propensity scores are clearly different between the intervention and comparison groups in each case, there is a reasonably good area of overlap between the groups. However, in constructing the model for household-level outcomes, 15 observations from the intervention group and four observations from the comparison group were dropped because there was not a suitable match for them.

Table A2.1: Estimating the propensity score

	Marginal effect	Standard error	*p*-value
Household head has primary education	-0.05	0.06	0.40
Household head is male	-0.17**	0.06	0.00
Proportion of household members with primary education	0.26**	0.08	0.00
Household in top 25% for wealth in 2012	0.14**	0.05	0.01
Number of household members	0.04***	0.01	0.00
Household's main economic activity was farming in 2012	0.12	0.06	0.05
Number of groups women attended in 2012	0.16***	0.02	0.00

The construction of the wealth index is described in section 5. Variables dated 2012 are estimates, based on recall data.
Dependent variable is binary, taking 1 for project participant households, and 0 otherwise.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Figure A2.1: Histogram of propensity scores in the intervention and comparison groups



Matching intervention households to comparison households

Following Rosenbaum and Rubin (1983), households were matched on the basis of their propensity scores. The literature has developed a variety of matching procedures. After a series of checks, it was decided to employ the kernel matching algorithm for the results presented in this Effectiveness Review. Kernel matching assigns more weight to the closest comparison group observations that are found within a selected 'bandwidth'. Thus 'good' matches are given greater weight than 'poor' matches. The review team used the *psmatch2* module in Stata and restricted the analysis to the area of common support. When using PSM, standard errors of the estimates were bootstrapped using 1,000 repetitions, to account for the additional variation caused by the estimation of the propensity scores.[10]

Checking balance

For PSM to be valid, the intervention group and the matched comparison group need to be balanced. In other words, the intervention and comparison groups need to be similar in terms of their observed characteristics. The most straightforward method of doing this is to test whether there are any statistically significant differences in baseline covariates between the groups in the matched sample. The balance of each of the matching variables after kernel matching is shown in Table A2.2. There are no statistically significant differences between the intervention and comparison groups for any of the matching variables used in the matched sample. For all of these variables, the *p*-values for the difference in means tests are larger than 0.2. It can therefore be concluded in each case that a satisfactory match has been found for the observable variable in the sample.

Table A2.2: Balancing test on matching variables

	Intervention group mean	**Comparison group mean**	***p*-value**
Household head has primary education	0.22	0.21	0.70
Household head is male	0.82	0.86	0.29
Proportion of household members with primary education	0.46	0.43	0.41
Household in top 25% for wealth in 2012	0.32	0.34	0.76
Number of household members	6.96	6.9	0.84
Household's main economic activity was farming in 2012	0.89	0.91	0.70
Number of groups women attended in 2012.	1.52	1.59	0.70

The construction of the wealth index is described in section 5. Variables dated 2012 are estimates, based on recall data.

* p < 0.1, ** p < 0.05, *** p < 0.01.

APPENDIX 3: ROBUSTNESS CHECKS

In order to assess the robustness of the results presented in section 5, a series of checks were carried out to determine whether the main findings of this report are sensitive to the estimation procedure – propensity score matching with the kernel method – that was used to control for observable differences between the intervention and comparison groups. This appendix presents five types of robustness check.

1 Multivariate regression

The first robustness test run is to estimate the impact of project participation using an ordinary least squares (OLS) regression. The main idea behind OLS is to isolate the variation in the outcome variable that is due to the intervention status – the project's impact – by controlling directly for the influence that observable differences between the intervention and comparison groups have on outcomes. To do this, Equation 1 is estimated.[11]

Equation 1

$$Y_i = \alpha + \beta_1 \tau_i + \boldsymbol{X}_i' \boldsymbol{\delta} + \varepsilon_i$$

In Equation 1, Y_i is the dependent variable (the outcome) and $\boldsymbol{X}_i$ is a column vector of the same matching variables listed in Tables A3.1 and A3.2. The intervention status is given by a dummy variable (τ_i), which takes the value 1 if the household participated in the project and 0 otherwise. The key difference between this OLS regression model and the PSM procedure used in the main report is that the OLS regression estimates a direct parametric relationship between the covariates in $\boldsymbol{X}_i$ and the dependent variable Y_i. This means that it is possible to include the observations that were excluded due to being off common support by extrapolating the relationship between $\boldsymbol{X}_i$ and Y_i. It should be borne in mind, however, that extrapolating in this way may bias the results if the covariates are distributed very differently between the intervention and comparison groups (Rubin, 2001).[12]

It is also important to note that, as with the PSM methods used in the main body of the report, OLS regressions can only account for observable differences between the intervention and comparison groups. Unobservable differences may still bias the results. In the tables that follow, only the estimate of $\beta_1$is reported.

2 Multivariate regression with alternative matching variables

Given the importance of controlling for recalled group participation for minimizing bias, the review team tested whether the results were sensitive to the way this information is captured by the matching variables. To do this, they altered $\boldsymbol{X}_i$, removing the dummy variable for whether or not the household participated in any community groups in 2012and adding instead: (1) the number of community groups in which the household participated in 2012 and (2) a dummy variable for whether or not the household participated in a women's community group. Once again, only the estimates of β_1 are reported.

3 Propensity score weighting

Following the example of Hirano and Imbens (2001),[13] OLS regressions were also estimated, using exactly the same model as in Equation 1, but weighting the observations according to the propensity score. Observations are assigned weights equal to 1 for the intervention households and $\hat{P}(\boldsymbol{X}_i)/(1-\hat{P}(\boldsymbol{X}_i))$ for the comparison households. The variable $\hat{P}(\boldsymbol{X}_i)$ represents the probability of a household being in the intervention group, given their observable characteristics, measured through the vector of matching variables $\boldsymbol{X}_i$ – this was estimated in the probit regressions in Appendix 2 The estimates of β_1 are reported in the same way as the regular OLS regressions.

4 Nearest neighbour matching

The nearest neighbour (NN) matching algorithm matches each observation from the intervention group with an observation from the comparison group that is closest in terms of their propensity score.[14] In this robustness check, the NN method is applied 'with replacement', meaning that comparison observations can be matched to intervention observations more than once.[15] The tables below report the estimated differences between the intervention and comparison groups. In the remainder of this appendix, these robustness checks are reported for the main results of the report.

Table A3.1: Training received by households

	1	2	3	4
	Training on vegetable production (%)	**Training on hygiene and malnutrition prevention (%)**	**Training on water resource monitoring through water and rainfall monitoring systems (%)**	**Training on use of quality seeds (%)**
OLS regression	0.60***	0.42***	0.15***	0.55***
	(0.03)	(0.04)	(0.03)	(0.03)
N	704	704	704	704
OLS regression with alternative matching variables	0.65***	0.46***	0.22***	0.62***
	(0.03)	(0.03)	(0.03)	(0.03)
N	704	704	704	704
OLS with PS weighting	0.57***	0.41***	0.17***	0.55***
	(0.04)	(0.05)	(0.04)	(0.04)
N	685	685	685	685
Nearest neighbour	0.61***	0.45***	0.19***	0.60***
	(0.08)	(0.08)	(0.07)	(0.07)
N	692	692	692	692

	5	6	7
	Training on use of fertilizers (%)	**Training on use of pesticides (%)**	**Training on food processing and storage techniques (%)**
OLS regression	0.59***	0.47***	0.18***
	(0.03)	(0.04)	(0.03)
N	704	704	704
OLS regression with alternative matching variables	0.63***	0.49***	0.27***
	(0.03)	(0.03)	(0.03)
N	704	704	704
OLS with PS weighting	0.57***	0.44***	0.20***
	(0.04)	(0.04)	(0.04)
N	685	685	685
Nearest neighbour	0.61***	0.48***	0.24***
	(0.07)	(0.07)	(0.07)

N	692	692	692

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3.2: Activities carried out in the community

	1	2	3	4
	Household members provided with seeds (%)	**Household members provided with agricultural tools (%)**	**Household members provided with cash for work (%)**	**Household members provided with technical support on modern methods of farming (%)**
OLS regression	0.69***	0.66***	0.23***	0.35***
	(0.03)	(0.03)	(0.04)	(0.03)
N	704	704	704	704
OLS regression with alternative matching variables	0.75***	0.71***	0.25***	0.44***
	(0.03)	(0.03)	(0.04)	(0.03)
N	704	704	704	704
OLS with PS weighting	0.63***	0.59***	0.27***	0.40***
	(0.04)	(0.05)	(0.04)	(0.04)
N	685	685	685	685
Nearest neighbour	0.66***	0.64***	0.27***	0.41***
	(0.07)	(0.07)	(0.07)	(0.06)
N	692	692	692	692

	5	6	7	8
	Household members had access to sand dams (%)	**Household members had access to water wells (%)**	**Household members had access to irrigation services (%)**	**Household members attended farmer field days (%)**
OLS regression	0.63***	0.16***	0.44***	0.15***
	(0.03)	(0.03)	(0.03)	(0.02)
N	704	704	704	704
OLS regression with alternative matching variables	0.69***	0.21***	0.50***	0.21***
	(0.03)	(0.03)	(0.03)	(0.03)
N	704	704	704	704
OLS with PS weighting	0.58***	0.19***	0.48***	0.20***
	(0.05)	(0.03)	(0.03)	(0.03)
N	685	685	685	685
Nearest neighbour	0.61***	0.18***	0.50***	0.20***
	(0.07)	(0.03)	(0.04)	(0.03)
N	692	692	692	692

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3.3: Adoption of improved agricultural practices

	1	2	3	4
	Household adopted at least one practice (%)	**Number of improved practices adopted**	**Household practised use of seed nursery (%)**	**Household practised production of organic materials (%)**
OLS regression	0.66***	2.40***	0.53***	0.50***
	(0.03)	(0.13)	(0.04)	(0.04)
N	704	704	704	704
OLS regression with alternative matching variables	0.72***	2.84***	0.61***	0.58***
	(0.03)	(0.12)	(0.03)	(0.03)
N	704	704	704	704

OLS with PS weighting	0.60***	2.42***	0.50***	0.46***
	(0.04)	(0.17)	(0.05)	(0.05)
N	685	685	685	685
Nearest neighbour	0.66***	2.55***	0.53***	0.50***
	(0.07)	(0.23)	(0.07)	(0.08)
N	692	692	692	692
	5	**6**	**7**	**8**
	Household practised organic farming (%)	**Household practised use of improved seeds (%)**	**Household practised use of integrated diversified farming system (%)**	**Household practised farm planning based on weather forecasts (%)**
OLS regression	0.37***	0.58***	0.37***	0.05**
	(0.03)	(0.03)	(0.04)	(0.02)
N	704	704	704	704
OLS regression with alternative matching variables	0.49***	0.63***	0.47***	0.05***
	(0.03)	(0.03)	(0.03)	(0.01)
N	704	704	704	704
OLS with PS weighting	0.42***	0.62***	0.37***	0.05***
	(0.04)	(0.03)	(0.04)	(0.02)
N	685	685	685	685
Nearest neighbour	0.43***	0.64***	0.40***	0.05***
	(0.05)	(0.03)	(0.07)	(0.02)
N	692	692	692	692

Rob

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3.4: Analysis of selected crop cultivation

	1	2	3
	Number of selected crops farmed in last 12 months	**Total kg harvested in last 12 months**	**Total kg harvested in last 12 months – log**
OLS regression	4.37***	2,217.04***	1.58***
	(0.23)	(660.61)	(0.18)
N	704	704	418
OLS regression with alternative matching variables	5.06***	2,647.53***	1.82***
	(0.21)	(578.04)	(0.18)
N	704	704	418
OLS with PS weighting	4.40***	2,415.60***	1.71***
	(0.32)	(541.89)	(0.18)
N	685	685	402
Nearest neighbour	4.91***	2,484.72***	1.51***
	(0.63)	(570.23)	(0.25)
N	692	692	409

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3.5: Household consumption

	1	2
	Food consumption per adult equivalent per day (logarithm of CFA franc)	**Total household consumption per adult equivalent per day (logarithm of CFA franc)**
OLS regression	0.18***	0.24***
	(0.05)	(0.05)
N	704	704

OLS regression with alternative matching variables	0.19***	0.27***
	(0.05)	(0.04)
N	704	704
OLS with PS weighting	0.19***	0.23***
	(0.05)	(0.05)
N	685	685
Nearest neighbour	0.14**	0.18***
	(0.06)	(0.06)
N	692	692

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3.6: Household wealth

	1	2
	Normalized wealth index	**Difference in normalized wealth index**
OLS regression	0.01	-0.01
	(0.07)	(0.06)
N	704	704
OLS regression with alternative matching variables	0.26***	-0.02
	(0.08)	(0.06)
N	704	704
OLS with PS weighting	-0.03	-0.01
	(0.09)	(0.07)
N	685	685
Nearest neighbour	-0.07	0.05
	(0.19)	(0.10)
N	692	692

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3.7: Women's involvement in community groups

	1	2
	Number of groups women participate in	**Number of groups women are involved in decision making**
OLS regression	0.21***	-0.01
	(0.06)	(0.06)
N	682	682
OLS regression with alternative matching variables	0.98***	0.36***
	(0.10)	(0.07)
N	682	682
OLS with PS weighting	0.21***	0.07
	(0.06)	(0.13)
N	664	664
Nearest neighbour	0.06	-0.07
	(0.16)	(0.19)
N	671	671

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3.8: Other women's empowerment indicators

	1	2
	Women disagreeing that standing up for their concerns is intimidating (%)	**Women's self-reported contribution to household needs (%)**
OLS regression	0.14***	0.80***

	(0.04)	(0.22)
N	678	682
OLS regression with alternative matching variables	0.16***	0.70***
	(0.04)	(0.22)
N	678	682
OLS with PS weighting	0.09*	0.90***
	(0.05)	(0.22)
N	660	664
Nearest neighbour	0.07	1.24***
	(0.06)	(0.27)
N	667	671

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3.9: Household decision making

	1	2	3	4
	Proportion of household decision-making areas where the woman has sole decision-making power/able to change decisions (%)	**Proportion of household decision-making areas where the woman has at least joint decision-making power (%)**	**Women who have sole decision-making power/able to change decision in more than half of household decisions (%)**	**Women who have at least joint decision-making power in more than half of household decisions (%)**
OLS regression	0.12***	0.04	0.12**	0.06
	(0.04)	(0.03)	(0.04)	(0.04)
N	682	682	682	682
OLS regression with alternative matching variables	0.13***	0.03	0.13***	0.04
	(0.03)	(0.03)	(0.04)	(0.03)
N	682	682	682	682
OLS with PS weighting	0.14***	0.08**	0.13***	0.12**
	(0.04)	(0.04)	(0.05)	(0.05)
N	664	664	664	664
Nearest neighbour	0.17**	0.07	0.15*	0.10
	(0.08)	(0.06)	(0.08)	(0.08)
N	671	671	671	671

Robust standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

In general, these robustness checks support the main results. Even if the coefficient is no longer statistically significant, there are only a few instances where the sign on the coefficient differs from what was presented in section 5, and these cases do not apply to the headline results. The nearest neighbour matching algorithm seems to cause the most change to the results.

NOTES

[1] A mature project has been running for long enough – typically at least 2.5 years – to have a reasonable expectation of impact, with either an expenditure rate of at least 70 percent or the completion of most project activities.

[2] This arises due to 'classical measurement error', which attenuates effect sizes – including for basic t-tests – towards zero.

[3] This follows the guidance in Filmer and Pritchett (2001). The first principal component captures sufficient variation in the data. Filmer, D. and Pritchett, L.H., 2001. *Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of* India. *Demography*, *38*(1), pp.115-132.

[4] To do this, the mean of the wealth index is subtracted, and it is then divided by its standard deviation.

[5] These results present something similar to a Difference-in-Difference specification. However, the baseline data are recalled rather than measured at baseline.

[6] M. Caliendo and S. Kopeinig (2008). *Some Practical Guidance for the Implementation of Propensity Score Matching*. *Journal of Economic Surveys*, Wiley Blackwell, vol. 22(1), pp.31-72.

[7] P.R. Rosenbaum and D.B. Rubin (1983). *The Central Role of the Propensity Score in Observational Studies for Causal Effects*. *Biometrika* 70(1), 41-55.

[8] Hagle, T.M. and Mitchell, G.E., 1992. *Goodness-of-fit measures for probit and logit*. *American Journal of Political Science*, pp.762-784.

[9] J.J. Heckman, R.J. LaLonde and J.A. Smith (1999). *The Economics and Econometrics of Active Labor Market Programs*, *Handbook of Labor Economics*, 3(A), 1865-2097.

[10] The review team elected not to cluster standard errors at the community level because this would result in a small number of clusters and would be likely to bias the standard errors downwards.

[11] It should be noted that, for all these regression techniques, robust standard errors are reported. However, the standard errors are not bootstrapped as in the main results in section 5.

[12] It is possible to test whether the covariates are distributed sufficiently similarly for the intervention and comparison groups using Rubin's (2001) tests. For the matching variables used in this report, with the kernel matching algorithm, Rubin's B = 24.0, and Rubin's R = 1.41. According to Rubin's recommendations, this suggests that the covariates are sufficiently balanced for OLS regression methods to be valid. Rubin, D.B., 2001. *Using propensity scores to help design observational studies: application to the tobacco litigation*. *Health Services and Outcomes Research Methodology*, *2*(3), pp.169-188.

[13] K. Hirano and G.W. Imbens (2001), *Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization*. *Health Services & Outcomes Research Methodology*, vol. 2, pp. 259-278.

[14] Choosing whether to match with and without replacement involves a trade-off between bias and variance. If replacement is allowed, the average quality of matching will increase and the bias will decrease, especially when the distribution of the propensity score is very different in the intervention and comparison groups. However, allowing for replacement increases the variance of the estimates because, in effect, the number of distinct comparison observations is reduced (Caliendo and Kopeinig, 2008).

[15] Following the guidance of Abadie and Imbens (2008), robust standard errors are calculated analytically using the *teffects* module in Stata. These standard errors are not bootstrapped. Abadie, A. and Imbens, G.W., 2008. *On the failure of the bootstrap for matching estimators*. *Econometrica*, *76*(6), pp.1537-1557.

For more information, or to comment on this report, email policyandpractice@oxfam.org.uk

ISBN: 978-1-78748-464-1

DOI: 10.21201/2019.4641

GB, Oxfam House, John Smith Drive, Cowley, Oxford, OX4 2JY, UK.

OXFAM

Oxfam is an international confederation of 20 organizations networked together in more than 90 countries, as part of a global movement for change, to build a future free from the injustice of poverty. Please write to any of the agencies for further information, or visit www.oxfam.org.

Oxfam America (www.oxfamamerica.org)
Oxfam Australia (www.oxfam.org.au)
Oxfam-in-Belgium (www.oxfamsol.be)
Oxfam Brasil (www.oxfam.org.br)
Oxfam Canada (www.oxfam.ca)
Oxfam France (www.oxfamfrance.org)
Oxfam Germany (www.oxfam.de)
Oxfam GB (www.oxfam.org.uk)
Oxfam Hong Kong (www.oxfam.org.hk)
Oxfam IBIS (Denmark) (http://oxfamibis.dk/)
Oxfam India (www.oxfamindia.org)
Oxfam Intermón (Spain) (www.intermonoxfam.org)
Oxfam Ireland (www.oxfamireland.org)
Oxfam Italy (www.oxfamitalia.org)
Oxfam Japan (www.oxfam.jp)
Oxfam Mexico (www.oxfammexico.org)
Oxfam New Zealand (www.oxfam.org.nz)
Oxfam Novib (Netherlands) (www.oxfamnovib.nl)
Oxfam Québec (www.oxfam.qc.ca)
Oxfam South Africa (http://www.oxfam.org.za/)

OXFAM