EVALUATION: MARCH 2015 PUBLICATION: OCTOBER 2015

WOMEN'S EMPOWERMENT IN PAKISTAN

Impact evaluation of the empowering small scale producers in the dairy sector project

Effectiveness Review Series 2014/15

Photo credit: Salman Shaukat/Oxfam

SIMONE LOMBARDINI & KIMBERLY BOWMAN

OXFAM GB

www.oxfam.org.uk/effectiveness

OXFAM

CONTENTS

ACKNOWLEDGEMENTS

We would like to thank Oxfam in Pakistan and local partner teams for being so supportive during the exercise. Particular thanks to Sohail Aziz, Irnum Malik and Rosa Garwood from Oxfam; Shazia Sarwar, Abbas Kahlid and Shams ud Din Chishti from Doaba Foundation; and Ahmed Nauman, Dr Tanveer Ahmad and Rizwana Waraich, the consultants who managed the fieldwork.

EXECUTIVE SUMMARY

Oxfam GB's Global Performance Framework is part of the organisation's effort to better understand and communicate its effectiveness, as well as enhance learning across the organisation. Under this Framework, a small number of completed or mature projects are selected at random each year for an evaluation of their impact, known as an Effectiveness Review. The project 'Empowering Small Producers, especially Women, in the Dairy Sector' (PKNC06) was one of those selected for an Effectiveness Review in the 2014/15 financial year.

The project's overall objective was to improve livelihoods opportunities, increasing income and employment, as well as raising women's empowerment by improving their economic leadership in the dairy sector. Oxfam GB implemented the project activities in conjunction with Doaba Foundation, a local partner organisation. The project started in 2011, supporting four cooperatives in four *tehsils* in Muzaffargarh district. In 2012 the project was interrupted due to a restructuring process and did not resume until January 2013, when it continued with only one of the four original cooperatives. This study will focus only on the impact on project participants involved from 2011 to 2014.

The project activities included the formation of one enterprise in the dairy sector and the establishment of ten collection centres where farmers could sell their milk production. The expectation was that the enterprise would buy milk from local farmers at higher prices than other competitors in the area, and still generate profits for the enterprise. The project activities also included the formation of ten community groups that provided training on milk production, animal health and the dairy market in order to improve the quality and quantity of the milk produced by the farmers. Half of the direct project participants and half of the members in the enterprise board were women.

The evaluation questions were:

- What has been the impact of the project in promoting women's empowerment among the project participants? Women's empowerment is defined by the project theory and stakeholder perception and measured by a composite index relevant to the socio-economic context of the project.
- What was the impact of the project on income and wealth for households involved in small dairy business who participated into the project?
- What was the impact of the project in changing knowledge, practices, and quality and quantity of milk production for small dairy farmers involved in the project?

Evaluation design

The Effectiveness Review took place in December 2014 in Muzaffargarh district, South Punjab – Pakistan. It intended to evaluate the success of the 'Empowering small producers, especially women, in the dairy sector' project in achieving its objectives: increasing income and employment among members of milk cooperative groups, and promoting women empowerment.

The review adopted a quasi-experimental impact evaluation design, complemented by a qualitative component. The quantitative impact evaluation aimed at measuring change that is causally attributable and representative of the project intervention. It involved comparing women that had been supported by the project with women in neighbouring communities that had similar characteristics in 2010. A household survey was carried out with 300 women participating into the project (either being directly involved in the project activities or being the spouse of a project participant), and 500 comparison women who had never been involved in any Oxfam project. At the analysis stage, the statistical tools of propensity-score matching and multivariate regression

were used to control for demographic and baseline differences between the households surveyed in project and comparison areas to provide additional confidence when making estimates of the project's impact.

The qualitative component informed decisions taken when developing the quantitative survey instrument and in interpreting data analysis. Qualitative work consisted of a literature review, field-based focus group discussions and individual interviews. Conducted in the project intervention area, these served to gather additional information, understanding and learning of the project implementation, as well as trying to answer the question of what women's empowerment means in South Punjab. In order to measure women's empowerment, this evaluation identified 18 indicators associated with empowerment of women in the dairy sector in South Punjab and employed a multidimensional measure to aggregate them into one composite index.

Results

The first important component of the project was to set up an enterprise with 10 milk refrigerators to buy milk from farmers for a higher price than other buyers in the area. The evaluation found that the median price per litre paid by the enterprise was higher than the price paid by local milkmen and middle-men, but in line with the price paid by other collection companies operating in the area. However the evaluation suggests that households involved in the project sold their milk production, on average, for lower prices relative to households in the comparison group. Evidence from focus group discussions suggests that this could be linked to problems in the implementation of the enterprise, with milk not being regularly collected in some areas, or with the enterprise failing to make payments in advance, unlike other competitors in the area.

Summary results of this Effectiveness Review

Outcome	Linked to project logic	Evidence of positive impact	Comments
Improved knowledge on milk production	Yes	Yes	The evaluation found evidence that the project led to higher levels of knowledge related to the dairy market, milk production and improved adoption of animal husbandry techniques, such as vaccination and de-worming.
Improved quality and quantity of milk produced	Yes	No	The evaluation did not find evidence of increased quantity or quality of milk produced attributable to the project.
Improved market conditions in dairy sector	Yes	No	The evaluation found that the median price paid by the enterprise was in line with the median price paid by other channels. However the average price per litre paid to project participants was lower than the average price paid to the comparison group. The evaluation also identified problems with the reliability of the project enterprise in regularly collecting milk and ensuring payments to the farmer, particularly when compared with other competitors operating in the same area.
Improved income and wealth	Yes	No	The evaluation failed to find evidence of higher income or wealth attributable to the project intervention.
Women's empowerment	Yes	Yes	There was evidence that the project led to an increase in the overall women's empowerment index. There was evidence of improved empowerment indicators in: self-confidence, personal autonomy, group participation, independent income, power in markets, control over time, and safety of movements outside the house. However, there is no evidence of change in: opinions on women's economic role, acceptability of violence against women, household decision-making power, and control of assets within the household.

Another important component of the project was providing training on livestock management and organisational management to members of milk cooperative groups. The evaluation found evidence that the project led to higher levels of knowledge related to dairy market and milk production, and to some improved adoption of animal husbandry techniques, such as vaccination and de-worming. These activities were intended to improve the quality as well as the quantity of milk produced. However, the evaluation did not find evidence of higher levels of production or higher quality of milk produced attributable to the project intervention.

Improvements in the market dairy sector, as well as improvements in the quantity and quality of milk produced, were expected to increase household wealth and incomes of the households involved into the project. However, this evaluation did not find evidence of higher income or material wealth associated with project participants.

The project also aimed at improving women's empowerment by strengthening their economic leadership in the dairy sector. This evaluation found that the project was successful in improving the overall women's empowerment, as measured by the composite index. Notably, project participants were associated with higher levels of self-confidence, personal autonomy, and perception of safety of movements outside the house. Project participants were more likely to influence community groups and reported a higher share, on average, of contribution to household income. However, there was no evidence of changes attributable to the project in other important women's empowerment indicators, such as: women's attitudes and beliefs regarding women's economic role, acceptability of violence against women, household decision making power, and control of household assets. More research should be undertaken to understand the impact on gender-based violence as results were not clearly defined.

Characteristics of women's empowerment examined in this Effectiveness Review

Level	Dimension	Characteristic	Connected to project logic	Evidence of positive impact
Personal	**Power from within**	Self-esteem/self-confidence	Yes	Yes
		Individual capability (knowledge)	Yes	Yes
		Women's opinions (attitude/beliefs):		
		• Women's economic role	No	No
		• Acceptability of GBV	No	No
		• Recognition of care	No	No
	Power to	Individual capability (apply knowledge)	Yes	No
		Personal autonomy	No	Yes
Relational	**Power with**	Social capital	No	No
		Participation in community groups	Yes	Yes
		Degree of influence in governing of community groups	Yes	Yes
		Attitudes and beliefs of the persons close to the woman	Yes	No
	Power over	Involvement in household decision-making	No	No
		Control over household assets	No	No
		Independent income	Yes	Yes
		Power in markets	Yes	Yes
		Experience of GBV	No	No
		Control over time	No	Yes
Environmental		Safety of movements outside the house	No	Yes

Programme learning considerations

Some important lessons that can be applied to other projects of this type in Pakistan and elsewhere have emerged from this evaluation. The Pakistan country team, and the programme team in particular, are encouraged to consider the following:

- **Improve project targeting and delivery quality**

The evaluation found that the project was implemented in an area where a number of other actors, both multilateral agencies and private sector companies, were already operating and working in similar thematic areas. Moreover, qualitative data suggested that some project participants perceived the enterprise as not being professionally run when compared with other businesses operating in the same market. Focus group discussions identified concerns and issues raised by project participants concerning the management of the enterprise, lamenting problems with collection of the milk, and having the milk spoiled as a consequence. Project participants also expressed dissatisfaction with the timing of the payments made by the enterprise, compared with other actors in the area.

This raised some questions around the value added of projects establishing a private enterprise in a context where there are already other private companies equipped with better skills and business means, although not necessarily committed to pay better prices for milk. A more careful targeting process of the intervention area should be put in place when selecting the market and location for project implementation.

- **Increase clarity over women's empowerment outcomes and pathways to change**

The evaluation identified no impact on income and wealth, but it did identify a positive impact attributable to the project on women's empowerment.

Survey results, as well as qualitative data, suggest that activity engagement with women at community level had a positive impact on a number of women's empowerment indicators. This, however, was achieved with a theory of change that lacked a clear definition of women's empowerment outcomes or pathways of change.

The project assumed that a higher contribution to household income and improved leadership for four women in the enterprise board would be the main levers for strengthening women's empowerment. The evaluation identified that the project had a positive impact on increasing the proportion of household income earned independently from other household members, but not on attitudes and beliefs about women's economic role, household decision-making, and control over assets.

The programme team has been encouraged to consider scaling up the women's empowerment components of this project, defining if an increase in household income is sufficient for women's empowerment or if other issues should be explicitly targeted and addressed as well. What the project means by empowerment and how change is expected to take place, including an explicit theory of change that differentiates between how change happens for women involved in the project and for women whose husbands participate, should be clearly defined.

- **Define what constraints are facing small milk-producer farmers other than knowledge**

One of the assumptions of the project was that providing training and increasing knowledge would also increase the quality and quantity of milk production. The evaluation identified that project activities had a positive impact on improving knowledge of milk quality and improved vaccination practices. However, there was no evidence of a higher quantity or quality of milk being produced as a consequence of the project.

The programme team has been encouraged to explore if there are other constrains that milk-farm producers are facing, other than limited knowledge.

- **Consider evaluation questions during programme design**

In future projects, it has been advised to consider including an evaluation framework in the project design.

Evaluation is a key tool for learning, to help projects and programmes succeed and generate evidence of success. When designing a project, the programme team is encouraged to consider and define key evaluative questions that they would like addressed; which components and characteristics of the intervention that should be evaluated; and finally, what are the reasons for conducting the evaluation (e.g. influencing, accountability, learning), and plan sufficient budget, time and resources. Different evaluation designs and methodologies provide different types of evidence, with different levels of confidence. For large-scale development interventions, a counterfactual evaluation design will allow the team to consider whether changes can be attributed to the project intervention.

1 INTRODUCTION

Oxfam GB's Global Performance Framework is part of the organisation's effort to better understand and communicate its effectiveness, as well as enhance learning across the organisation. Under this framework, a small number of completed or mature projects are selected at random each year for an evaluation of their impact, known as an Effectiveness Review. One key focus is on the extent to which they have promoted change in relation to relevant OGB global outcome indicators.

This Effectiveness Review took place in December 2014 in Muzaffargarh district, south Punjab – Pakistan, and it was intended to evaluate the success of the project 'Empowering Small Producers, especially Women, in the Dairy Sector' in supporting women to achieve a greater empowerment.

The OGB global outcome indicator under which this project has been selected is Women's Empowerment, which is defined as change in empowerment of supported women – measured by a composite index assessing indicators of empowerment that are relevant to the socio-economic context of the project under analysis. The index is explained in more details in Section 5.

The evaluation questions were:

- What has been the impact of the project in promoting women's empowerment among the project participants? Women's empowerment is defined by the project theory and stakeholder perception and measured by a composite index relevant to the socio-economic context of the project.
- What was the impact of the project on income and wealth for households involved in small dairy business who participated into the project?
- What was the impact of the project in changing knowledge, practices, and quality and quantity of milk production for small dairy farmers involved in the project?

The project, implemented by the Oxfam partner Doaba Foundation, started in January 2011 supporting four cooperatives in four tehsils in Muzaffargarh district. The project was interrupted in 2012, not resuming until January 2013 when it continued with only one of the four original cooperatives. This study will focus only on the impact on the project participants that were involved in the project from 2011 to 2014.

Figure 1.1: Map of Pakistan with Muzaffargarh highlighted.

Source: BBC

This report presents the findings of the Effectiveness Review. Section 2 briefly reviews the activities and the intervention logic of the project. Section 3 describes the evaluation design used, and Section 4 explains how this design was implemented. Section 5 thereafter presents the results of the data analysis, including the descriptive statistics of the population surveyed and the differences in outcome measures between the intervention and comparison groups. Section 6 concludes with a summary of the

findings and some considerations for future learning.

2 PROJECT DESCRIPTION

2.1 PROJECT ACTIVITIES

The overall objective of the 'Empowering Small Producers, Especially Women, in the Dairy Sector' project was to gain equitable access and control over increased income and economic resources for women in the dairy sector. In particular the project intended to:

- improve livelihoods opportunities by increasing income and employment
- improve women's empowerment by improving women's economic leadership in the dairy sector.

The project implemented the following activities:

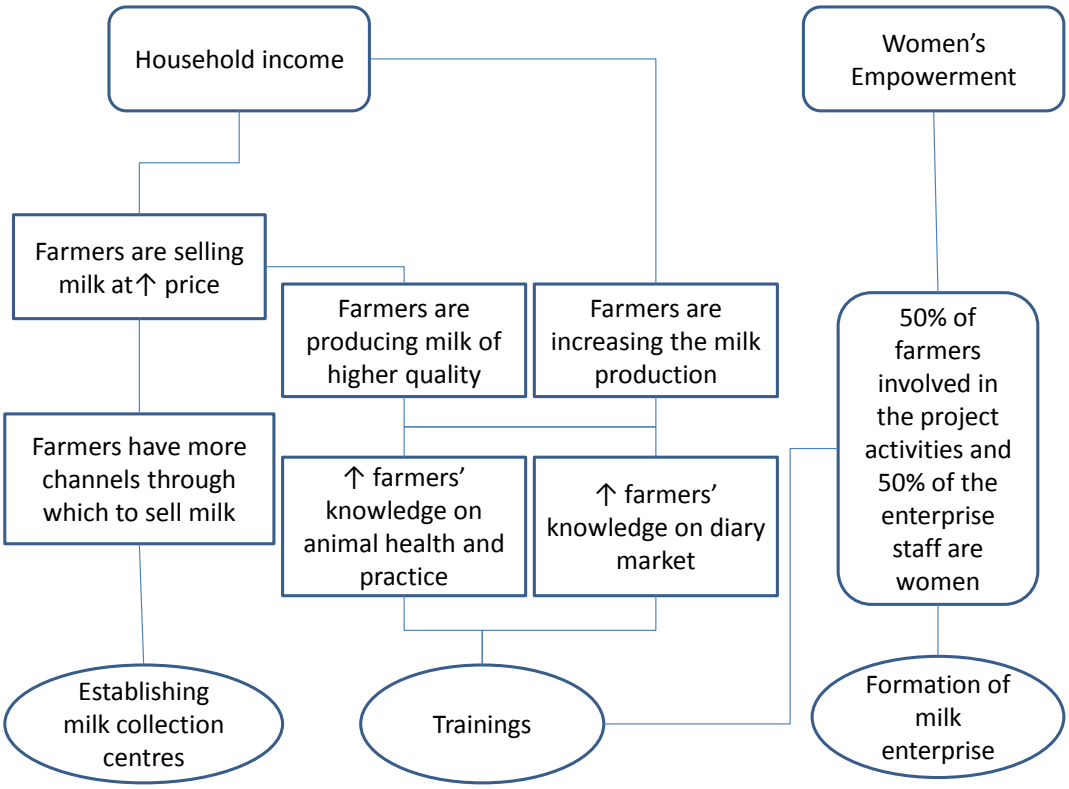
1- Formed ten community groups who received training on the dairy sector. Groups were formed in 24 villages and totalled 661 members, of whom 50 per cent were women.
2- Established ten collection centres for collecting milk from small farm producers participating into the community group. However, it has to be noted that eight out of ten milk collection centres were highly affected by the flood in 2014, and at the time of the survey only seven were working.
3- Created one enterprise with eight board members (four women, four men) and with three paid staff. The enterprise established a formal agreement with ENGRO, a major company in the dairy sector, in order to sell the milk collected from the collection centres.

Figure 2.1 provides a visual representation of simplified project logic.

The first objective of the project was increasing farmers' income. The project provided training on dairy market and milk production practices as well as on animal health to women and men members of community groups. The producers' increased knowledge was expected to translate into higher quality of milk produced, as well as increased milk production. Higher milk quality (measured as fat content) should allow farmers to obtain higher prices. Increased milk production, combined with a greater selling price, was expected to increase income for households with members participating in community groups.

The project also aimed at increasing milk sales by establishing ten channels where small producers could sell their milk. It was expected that these channels could deliver higher prices for milk, compared with what normally paid by a 'middle man'.

Figure 2.1: Simplified logic model



The second objective of the project was improving women's empowerment by promoting women's group membership and economic leadership in the dairy sector, obtaining equal participation and benefit for women – and ensuring 50 per cent of participants were female. It also sought to include women in leadership positions in structures created by the project. This evaluation investigated more broadly whether involving women in development intervention might affect empowerment. In order to do this section 5.3.4 will discuss and present a measure for women's empowerment.

2.2 CONTEXTUAL BACKGROUND

This section aims to provide background information on the context of Pakistan, particularly in Muzaffargarh district, South Punjab, where the project was implemented.

Pakistan is a country of approximately 188 million people, with 40 per cent of the population living below the US$2 a day poverty line. Literacy rates in Pakistan are highly influenced by socio-economic factors, with a great disparity in the literacy rate between men and women. According to the Bureau of Statistics 2012–13, while the overall literacy rate for men in Pakistan reached 71.1 per cent, it was only 48.1 per cent for women. In rural areas literacy levels are even lower with the female literacy rate around 37.4 per cent and the male literacy rate around 65.1 per cent.

Employment figures from 2012–13 reveal that the total labour force in Pakistan has reached almost 60 million, with overall labour force participation rate of 53.1 per cent but only a 36.4 per cent female labour participation rate.

The livestock sector constitutes almost 12 per cent of the country's GDP, and employs directly or indirectly 30–35 million people in rural areas. Traditionally, livestock management has been dominated by women, particular in rural areas. Consequently, the female labour force participation in the agricultural sector reaches 74 per cent, while male labour force participation is only 34.5 per cent.

Although Punjab province in Pakistan is one of the most developed in the country, South Punjab is the least developed and most agricultural area in Punjab. Muzaffargarh is a district in southern Punjab with an area of 8,249Km2. Muzaffargarh district is also called Doaba in the local language, meaning 'a piece of land that lies between two rivers, because it lies between the Chenab river on its east and the Indus river on its west. This piece of land is often prone to floods; particularly violent was the flood that affected the area in August 2010. The district is divided into four tehsils, which are then divided into 93 union councils.

3 EVALUATION DESIGN

This Effectiveness Review employed a mixed method approach for impact evaluation, combining a quantitative quasi-experimental design, which provided representative and generalisable results, with qualitative information which provided a deeper understanding of the project's context and mechanisms. This section presents the two approaches.

3.1 QUASI-EXPERIMENTAL DESIGN

The central problem presented in designing an impact evaluation of any social programme is how to compare the outcomes that result from that programme with *what would have been the case* without that programme having been carried out. In the case of this Effectiveness Review, the situation of households in the villages where the project was implemented was examined 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 this 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), common practice is to make a comparison between units that were subject to the programme and units 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 programme, observing the situation of units where the programme was not implemented can provide a good estimate of the counterfactual.

An ideal approach to an evaluation such as this is to select the units in which the programme will be implemented at random. Random selection minimises the probability of there being systematic differences between the programme and non-programme units, and so maximises the confidence that any differences in outcome are due to the effects of the programme.

In the case of the project examined in this Effectiveness Review, the unit at which the programme was implemented was the village: within each of the project areas, specific villages were selected for a women's group to be established and for the other activities to be implemented, while other villages were not selected. The selection of villages was not made at random; in fact, activities were initiated based on distance from the rivers crossing the regions, how villages were affected by a big flood that affected the region in 2010, and finally by distance to communication roads. However, discussions with the implementation staff revealed that there were in fact more villages that were considered suitable for project implementation than could be covered by the programme. This allowed a 'quasi-experimental' evaluation design to be adopted, in which the situation of households in those non-implementation villages was assumed

to provide a reasonable counterfactual for the situation of households in the implementation villages.

Women in the project villages were 'matched' with women with similar characteristics in non-project (or 'comparison') villages. Matching was performed on the basis of a variety of characteristics – including household size, education level of the women and the head of the household, distance from the local market, distance from the river, a binary variable indicating whether the household has been affected by the flood in 2010, land size cultivated, number of groups in which the women were engaged, productive activities in which the households engaged, 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 productive activities and wealth indicators), matching was performed on the basis of these indicators *before* the implementation of the project. Since baseline data were not available, survey respondents were asked to recall some basic information about their household's situation from 2009, before the project was implemented. Although this recall data is unlikely to be completely accurate, it should not have led to significant bias in the estimates as long as measurement errors due to the recall data were not significantly different for respondents in the intervention and comparison groups.

The survey data provided a large number of individual and household characteristics on which matching could be carried out. Matching was based on a 'propensity score', which represented the conditional probability of the household being in an intervention village, given particular background variables or observable characteristics. Women in the project and comparison villages were matched based on their having propensity scores within certain ranges. Tests were carried out after matching to assess whether the distributions of each characteristic were similar between the two groups. Details on the validity of the propensity score matching procedure are reported in Appendix 2.

As additional check on the validity of the results derived from the propensity-score matching procedure, results were also estimated using multivariate regression models. Like propensity-score matching, multivariate regression also controlled for measured differences between intervention and comparison groups, but it did so by isolating the variation in the outcome variable explained by being in the intervention group after the effects of other explanatory variables have been accounted for. Appendix 3 provides estimates for the robustness checks.

It should be noted that both propensity-score matching 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, 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 weather (such as recurring climatic shocks) or other contextual conditions faced by households. The choice of which intervention and comparison villages to survey for this Effectiveness Review was made principally to minimise the potential for any such unobservable differences to bias the results.

3.2 QUALITATIVE COMPONENT

Past Effectiveness Reviews have shown that relying on a purely quasi-experimental design can provide limited understanding of the project's context and mechanisms that explain how the project works. For this reason this evaluation integrated the quantitative analysis with a qualitative component with the purpose of:

- providing additional understanding of the culture and socio-economic characteristics of the context under analysis, particularly in relation to women's empowerment
- providing additional understanding of the mechanism taking place as an effect of the project in order to aid the analysis and interpretation of the results coming from the quantitative component.

The qualitative component included a literature review of women's empowerment issues in Pakistan; six key informant interviews with active women leaders; and ten focus group discussions with female farmers group members involved in the project and four with male group members or husbands of female group members. Qualitative data was reviewed as part of data analysis, and evidence or findings from qualitative work were presented jointly with the results of the quantitative component, aiding in the interpretation of the results and understanding of the project.

4 DATA

4.1 SAMPLING OF INTERVENTION AND COMPARISON VILLAGES

The first stage in identifying an appropriate comparison group for a quasi-experimental evaluation is to understand the process by which participants were selected. The project started in January 2011 supporting four cooperatives in four tehsils in Muzaffargarh district. However, due to a restructuring process, the project was interrupted in 2012 and restarted again in January 2013 but implemented project activities in only one of the original four tehsils, Alipur. Since then, the project has only worked in four union councils, establishing ten community groups divided between 24 villages and reaching 660 members, of which 330 were women.

For this review, a decision was made to focus only on project participants that had been involved from 2011 to 2014, in the belief that project activities taking place between 2011 and 2012 were too reduced and too long ago to expect to be able to measure any kind of positive impact.

The focus of this study was to investigate the impact of the project on women's empowerment. It was decided to select the intervention group by randomly sampling 300 women from the 660 project participants. If the individual randomly selected for the interview was a man, the enumerators were instructed to interview his spouse or the closest female next of kin if he was unmarried. This choice was likely to underestimate the estimates on the project's impact on women's empowerment. However, it was not expected to bias estimates on income, wealth and milk production.

In order to identify a suitable comparison group, long and detailed discussions were held with staff from the partner organisation and the consultant team in order to locate comparable villages within the same geographical area that were not covered by the project. Villages identified as potential comparators needed to share similar socio-economic characteristics within the same tehsil, Alipur, and when that was not possible, neighbouring villages in Jatoi tehsil were identified. Particular attention was paid to the distance from the rivers (which is considered a proxy for prevalence to climatic shocks)[1] and from communication roads (which is considered an indicator of proximity to markets), as well as to the areas affected by the 2010 flood. Finally, ten union councils and 25 revenues villages were identified.

Table 4.1: Summary sample intervention group.

Union council	Number of intervention villages	Number of project participants	Number of project participant respondents
Bait Mullan Wali	8	240	118
Baz Wala	5	120	50
Mud Wala	8	240	106
Yaki Wala	3	60	26
Total	**24**	**660**	**300**

Within the selected villages, respondents were randomly selected starting from the largest mosque in the village, spinning a pen to determine a random direction and then contacting every fourth household. In order to be interviewed, the respondent had to meet the following criteria:

- Be a women aged between 18 and 60
- Live in the household in 2010
- Own between 1 and 6 milk animals
- Be involved in dairy activities

The final sample included 308 women randomly selected from the project participants (called intervention group) – of which 187 were group members themselves and 119 were spouses of the project participants – and 488 women randomly selected from the comparison villages (also called the comparison group).

4.2 ANALYSIS OF BASELINE CHARACTERISTICS

In order to control for the validity of the comparison group, women in project and comparison villages were compared in terms of their socio-economic characteristics in 2009. These data were based on information recalled during the questionnaire or reconstructed from the household composition at the time of the survey.

The full comparison is shown in Table 4.2. While distances from the river and the proportion of households affected by the flood showed no statistically differences between the intervention and comparison sample, some important differences were found between the households in project and comparison villages. For example, it appeared that on average the sample of women interviewed in the comparison group cultivated more land in 2009 than the sample of women in the intervention group. On average, women sampled in the intervention group were participating in a greater number of groups than the comparison sample. Conversely, households in the intervention group were less likely to be involved in the dairy sector than the intervention group. Finally, households in the intervention group appeared on average to be less wealthy than households in the comparison group.

These differences, which existed before the project, had the potential to bias any comparison of the project's outcomes between the project and comparison villages. It was therefore important to control for these baseline differences when making such comparisons. As described in Section 3, the main approach used in this Effectiveness Review was propensity-score matching (PSM). The full details of the matching procedure applied are described in Appendix 2. After matching, women in the project and comparison villages were reasonably well-balanced in terms of the recalled baseline data, with few significant differences between them. However, not all of the women interviewed in the project villages could be matched, and accordingly 75 of the

300 women surveyed in the intervention group had to be dropped from the analysis. The reasons for and consequence of these decisions are described in greater detail in Appendix 2.

Table 4.2: Baseline characteristics before matching

	Comparison mean	Intervention mean	Difference
1[HHH has no formal education]	0.730	0.760	-0.030
1[Respondent has no formal education]	0.935	0.940	-0.005
Household size 2009	6.216	6.153	0.063
1[Head of HH is female]	0.905	0.937	-0.032
1[Household was affected by the flood in 2010]	0.655	0.623	0.031
Distance of the house from the river in 2009 (in km)	5.609	5.537	0.072
Total area cultivated in 2009	2.907	1.791	1.117*
Number of groups involved in 2009	0.010	0.710	-0.700***
1[Household involved in dairy sector in 2009]	0.817	0.750	0.067**
1[Household farming in 2009]	0.956	0.940	0.016
1[Household involved in labour sector in 2009]	0.661	0.663	-0.003
1[Household involved in private business in 2009]	0.103	0.120	-0.017
1[HH is in the second wealth quintile]	0.185	0.227	-0.042
1[HH is in the third wealth quintile]	0.236	0.140	0.096***
1[HH is in the fourth wealth quintile]	0.179	0.237	-0.058**
1[HH is in the fifth wealth quintile]	0.214	0.173	0.041
Number of observations	300	504	804

4.3 QUALITATIVE SAMPLING

Qualitative data was collected via Focus Group Discussions (FGDs) and In Depth Interviews (IDIs) in ten different villages and four union councils in Tehsil Alipur, Muzaffargarh district, where the project under analysis has been implemented since 2011. Participants in qualitative fieldwork were purposefully selected by the field team, in consultation with the local partner. Given the time of year and busy schedule of rural woman, the evaluation team sometimes noted limitations in accessing participants for longer periods of time.

The selection criteria for village sites for the qualitative work were:

- The existence of active mixed-gender village groups and cooperatives.
- That the project had established a chiller/refrigerator, which was operational.
- Male and female group cooperatives members received training as a part of the project activities.

5 RESULTS

5.1 INTRODUCTION

This report is intended to be free from excessive technical jargon, with more detailed technical information being restricted to the appendices and footnotes. However, there are some statistical concepts that cannot be avoided in discussing the results. In this report, results will usually be stated as the average difference between women living in villages where the project was implemented (that is referred to as the 'intervention group') and the matched women in villages where the project was not implemented (named the 'comparison group').

In the tables of results on the following pages, statistical significance will be indicated by asterisks, with three asterisks (***) indicating a p-value of less than 10 per cent, two asterisks (**) indicating a p-value of less than 5 per cent and one asterisk (*) indicating a p-value of less than 1 per cent. The higher the p-value, the less confident we are that the measured estimate reflects the true impact. Results with a p-value of more than 10 per cent are not considered to be statistically significant.

The results are shown after correcting for observable baseline differences between the women interviewed in the project villages (the 'intervention group') and in the households in comparison villages using a propensity-score matching (PSM) procedure. The details of this procedure are discussed in Appendix 1. All outcomes have also been tested for robustness to alternative statistical models 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 footnotes.

5.2 INVOLVEMENT IN PROJECT ACTIVITIES

Before considering the project's effect on outcomes, it is important to examine whether the respondents reported having participated in the activities implemented under this project.

As presented in Section 2, one aspect of the project intervention was the provision of training. Figure 5.1 shows that more than 55 per cent of the respondents in the intervention group reported having a household member who has attended a livestock management training session since 2009. The proportion of women reporting having a household member having attended other training courses, such as organisational management, marketing, or water management, is slightly lower. Very few households in the comparison group reported having a household member attending any training sessions since 2009.

Figure 5.1: Proportion of households having received training since 2009

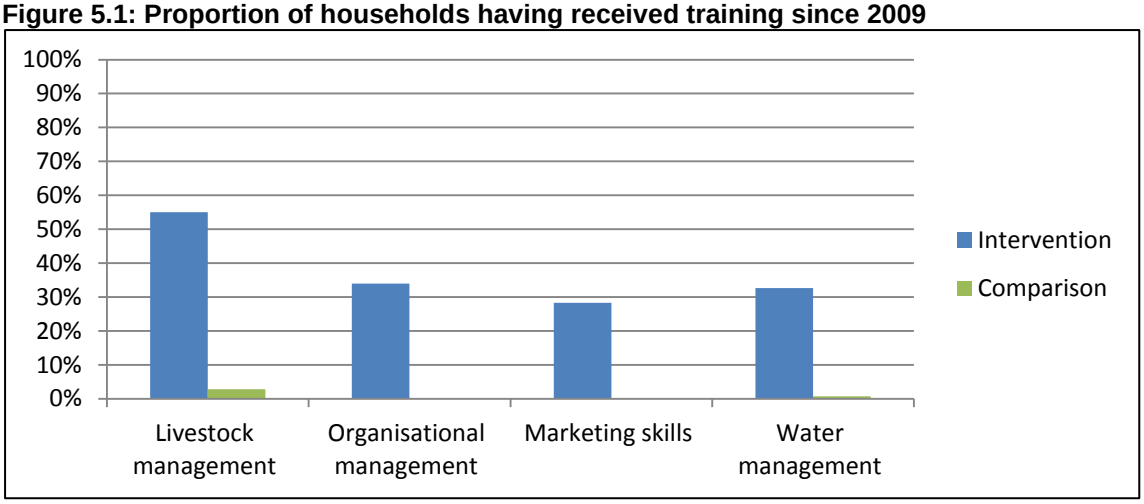


Figure 5.1 increases the confidence that the women indentified in the comparison group belong to households that were not exposed to training on livestock, organisational management, marketing, and water management.

The second set of project activities consisted of establishing a milk collection centre in order to enable project participants to sell their milk directly to the enterprise collection centre. Figure 5.2 shows that almost 80 per cent of the respondents in the intervention group reported being able to sell their milk production directly to the enterprise collection centre, while the equivalent proportion in the comparison group is lower that 15 per cent.

Figure 5.2: Markets for *potentially* selling milk production

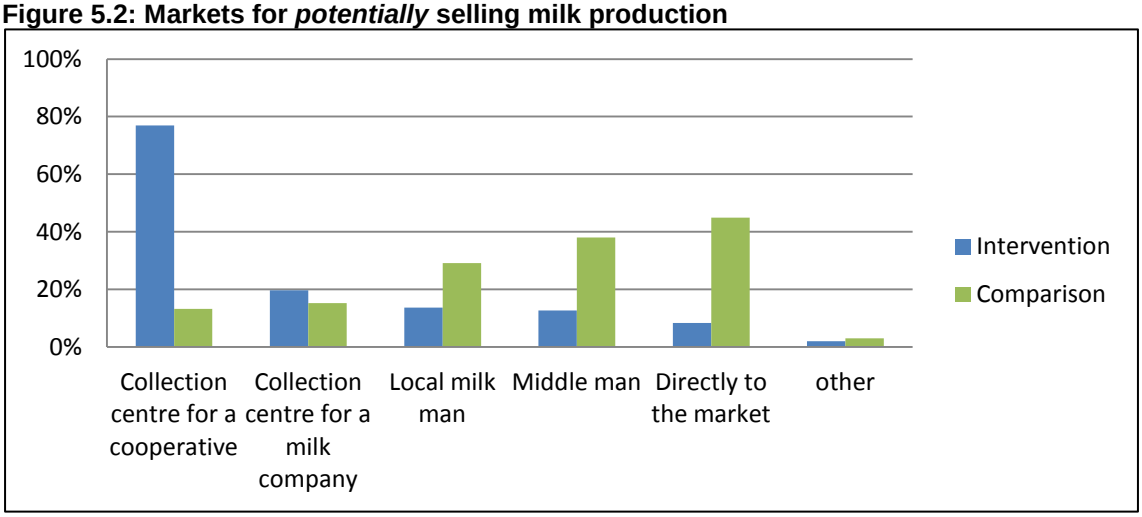
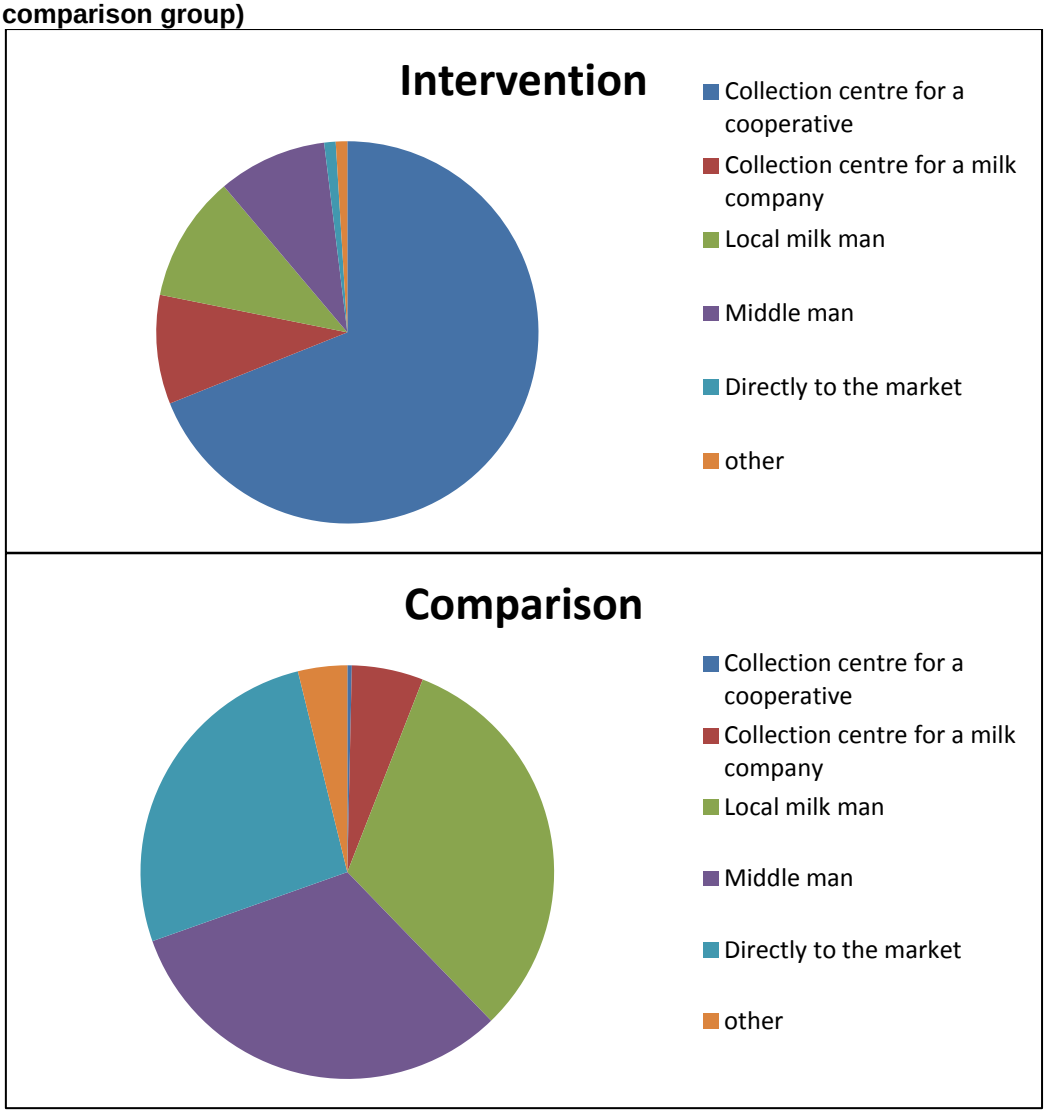


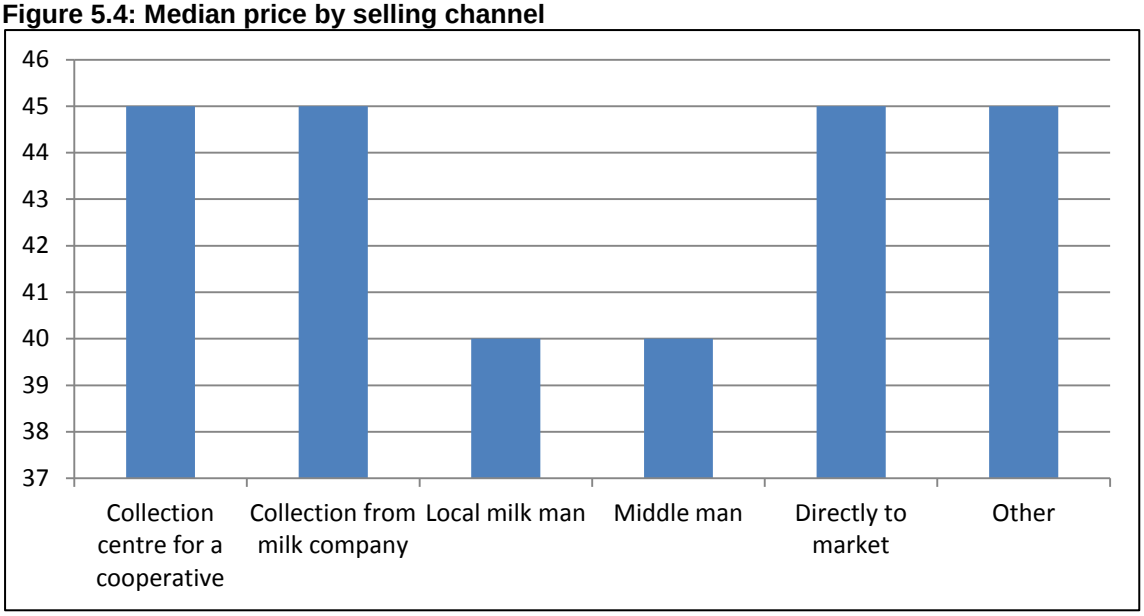
Figure 5.3 shows the marketing channels through which respondents in the intervention and comparison groups reported having sold their milk production in the last week. Almost 70 per cent of the sample in the intervention group sold their milk production to the collection centre. While in the comparison group than 60 per cent sold their milk production to the local milk man or to the middle man, and 26 per cent sold directly to the local market.

Figure 5.3: Where the milk production was actually sold (intervention group and comparison group)



The assumption of the project was that the price paid by the enterprise was going to be higher than the price that farmers were obtaining from other sources. Figure 5.4 shows the median price for the entire sample disaggregated by market channel. Figure 5.4 suggests that the price that farmers are obtaining from the project enterprise is the same as the price they would obtain selling directly to other collection points established by other milk companies (such as Endro or Nestlé) or if they were selling directly to the market. However this is higher than the price obtained by selling milk to the local milkman or the middle man.

Figure 5.4: Median price by selling channel



5.3 ANALYSIS OF OUTCOMES

This section will examine the differences between a sample of women involved in the project and matched women in comparison communities on an outcome measure capturing the project's theory of change as discussed in Section 2. The indicators on milk production, as well as consumption and wealth, will be presented at household level as the project intended to improve conditions for both men and women.

The outcomes measures examined in this section are:

- Milk production and markets
- Household consumption
- Household wealth
- Indicators of women's empowerment.

5.3.1 Milk production and markets

According to the theory of change presented in Section 2, there is the expectation that the project increased the quantity and quality of the milk production, as well as the price at which the producers sold their product, which was then expected to translate into an increased income. This section presents the estimates concerning milk quality, quantity and price.

Table 5.1: Milk production

	Milk production Litres	Quality of milk % fat content	Price milk PKR/litre
Intervention group mean:	14.658	4.276	43.007
Comparison group mean:	15.132	4.142	44.689
Difference:	-0.474	0.133	-1.683***
	(1.347)	(0.230)	(0.582)
Observations intervention:	223	191	150
Observations:	720	353	440

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

The first column of Table 5.1 suggests that there is no statistically significant difference between the weekly production of milk for project participants and the comparison group. Farmers in the intervention group reported on average production of 14.6 litres of milk per week.

The second column of Table 5.1 suggests that the quality of milk produced by farmers involved in the project was higher, but not significantly different from the milk produced by the comparison group. Quality of milk was measured by fat content of the milk produced. The reader should be aware that this measure is based on self-reported data, and therefore subject to possible bias as some farmers might not know the exact fat content of their milk. Estimates suggest that on average milk in the intervention group had 4.2 per cent fat, while milk produced in the comparison group had 4.1 per cent fat.

The final column in Table 5.1 suggests that on average the price received for selling milk is lower for the project participants than for the comparison group. While the intervention group reported to be paid on average more than 43 PKR per litre (equivalent to 0.42USD), the comparison group reported less than 44 PKR per litre (0.41USD).[2]

These estimates should be treated with caution, as it is possible that other actors within the area adapted their prices in response to the implementation of the cooperative. The project operated in a complex market environment where other private sector actors were also competing to buy milk. The project theory aimed to increase the 'channels' through which farmers could sell their milk. It is possible that the project could successfully contribute to generating upwards pressure on prices paid by companies and middlemen. For example, during one focus group, it was mentioned that the price paid by the middle-men used to be 25 PKR per litre, while now they pay 40 PKR per litre. In the absence of historical price data, and of a deeper understanding of the market dairy sector in Muzaffargarh, it is difficult to imply that this change is causally linked to the project alone.

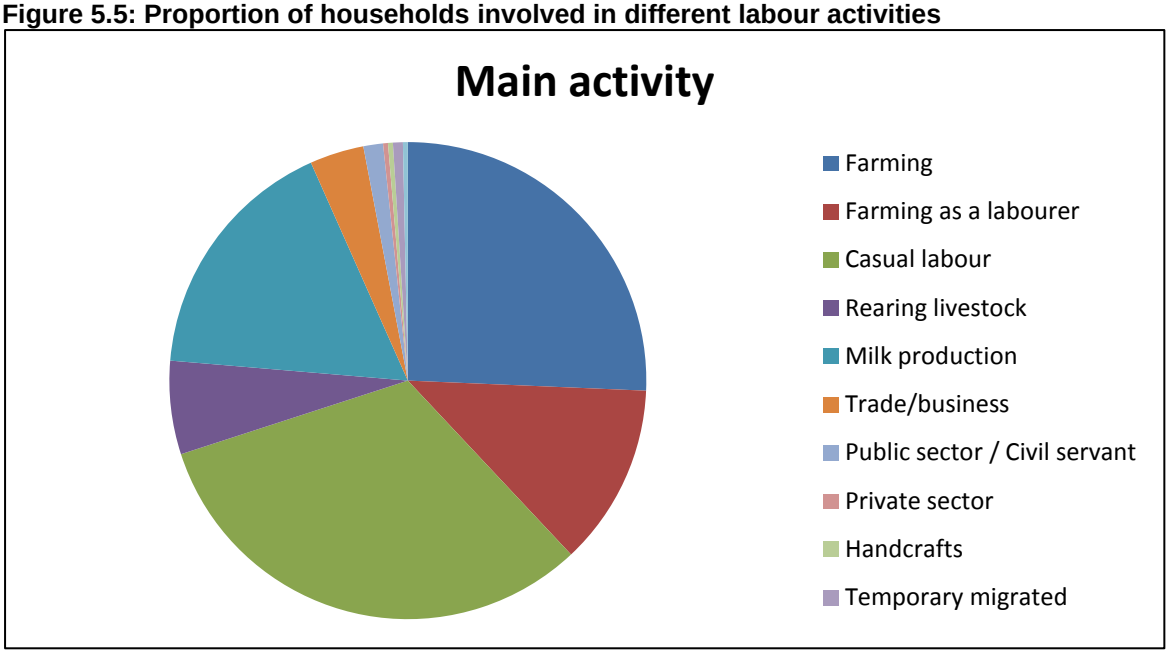
In some focus group discussions respondents expressed concerns about the efficiency of collecting centres, which allegedly paid after milk delivery rather in advance, and were less reliable in collecting milk than other channels and competitors in the formal market. Qualitative researchers also identified that Nestlé and other competitors may have at times provide loans in order to encourage farmers to sell their milk to them rather than to the enterprise. Unfortunately, the study did not investigate further the type of loans and interest rates that these companies were applying.

5.3.2 Household consumption

Measuring household income directly was problematic: self-reported measures of total income were generally regarded as unreliable, given the wide variety of endeavours such populations engaged in to generate income. Focus group discussions and survey data agreed in finding that selling milk was not a full-time livelihood activity for project participants, but rather one of many income-earning activities.

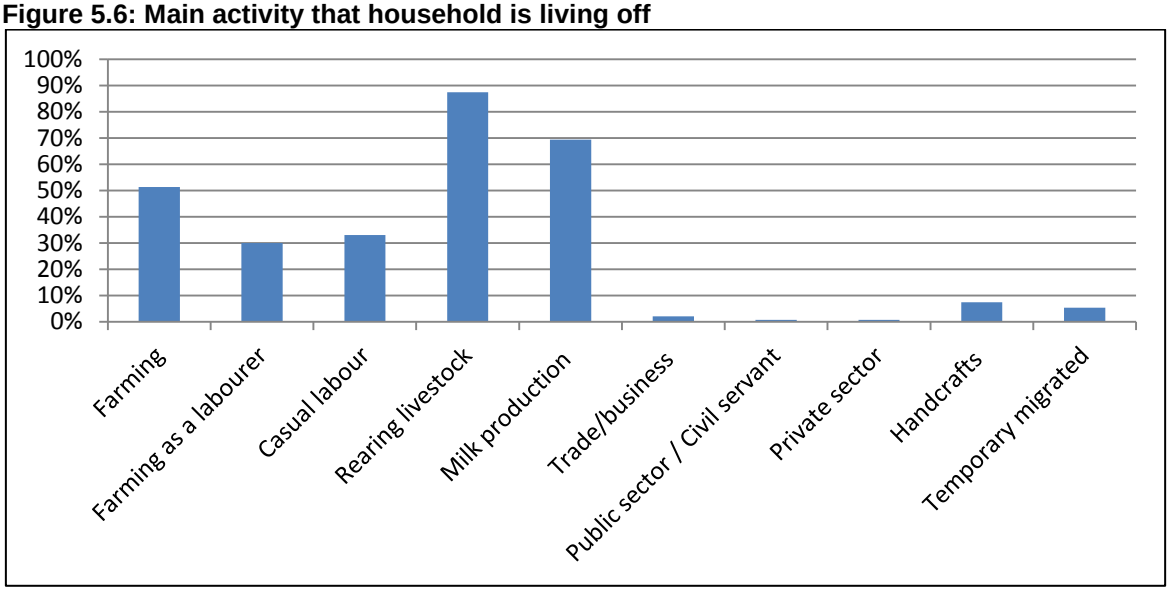
Figures 5.5 and 5.6 illustrate the variety of activities that project participants reported being involved in, as well as their main income-generating activity. The vast majority of project households reported being involved in rearing livestock (87 per cent of the sample) and in milk production (69 per cent of the sample). Moreover, farming and casual labour represent the main sources of income for 25 per cent and 32 per cent of the households involved in the project, respectively. A direct income measure would have to collect detailed information on the contribution of each of these activities to household income.

Figure 5.5: Proportion of households involved in different labour activities



Source: project participants only

Figure 5.6: Main activity that household is living off



Source: project participants only

For these reasons, the survey did not attempt to collect data on total household income directly. However, there is a widely recognised and strong association between household income and consumption.[3] The Effectiveness Review therefore followed the common practice of collecting data on 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 to select from a list of 17 products what types and quantities of food they had consumed over the previous seven-day period. 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 had been 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 13 items, such as transportation, education expenses, health expenses, wood or charcoal for

cooking/heating, clothes, etc. 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 10 items, which included equipment for the household, physical improvement of the house, jewellery, social festivals, livestock purchase and health management. The household expenditure measure was calculated by converting each of the expenditure types into a per-day per-capita[4] figure and adding them together. The expenditure variable had also been expressed on a logarithmic scale, to improve the model fit in regression analysis and reduce the influence of outliers.

Table 5.2 shows the comparison of total expenditure between project households and comparison households after the logarithmic transformation. There is no evidence of any statistically significant difference in this variable between project and comparison households. These estimates are consistent also when disaggregating yearly, monthly and by food consumption.

The final column in Table 5.2 shows the number of items the consumed in the household during the previous week from the list of 17 items of the questionnaire. This is a measure of food diversity ranging from zero to 17. On average, there are no statistically significant differences in diet diversity between intervention and comparison households.

Table 5.2: Consumption

	Log (Total consumption - daily per capita)	Log (Yearly expenditure per capita)	Log (Monthly expenditure per capita)	Log (Value food consumed in last 7 days - daily per capita)	Number of items consumed in the last week
Intervention group mean:	5.858	4.326	4.869	4.518	6.787
Comparison group mean:	5.862	4.283	4.817	4.583	6.605
Difference:	-0.003	0.043	0.053	-0.065	0.182
	(0.058)	(0.108)	(0.071)	(0.044)	(0.164)
Observations intervention:	225	225	225	225	225
Observations:	729	728	729	728	729

Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated *after* matching.

5.3.3 Household wealth

An alternative way to consider income was to investigate asset ownership. For this reason, respondents were asked about their ownership of various types of household goods and assets, as well as about the condition of their housing. These data were used to create a wealth index using Cronbach's alpha.[5] A total of 27 assets and other wealth indicators were used to construct the household wealth index with their inter-item correlations. The wealth indices were then created through applying Principal Component Analysis (PCA) to the selected indicators. PCA is a data reduction technique that narrows in on the variation in household asset ownership, which is assumed to represent wealth status: the more an asset type is correlates with this variation, the more weight it is given to it.

Table 5.3: Wealth index

	Wealth Index
Intervention group mean:	-0.020
Comparison group mean:	0.018
Difference:	-0.039
	(0.175)
Observations intervention:	225
Observations:	729

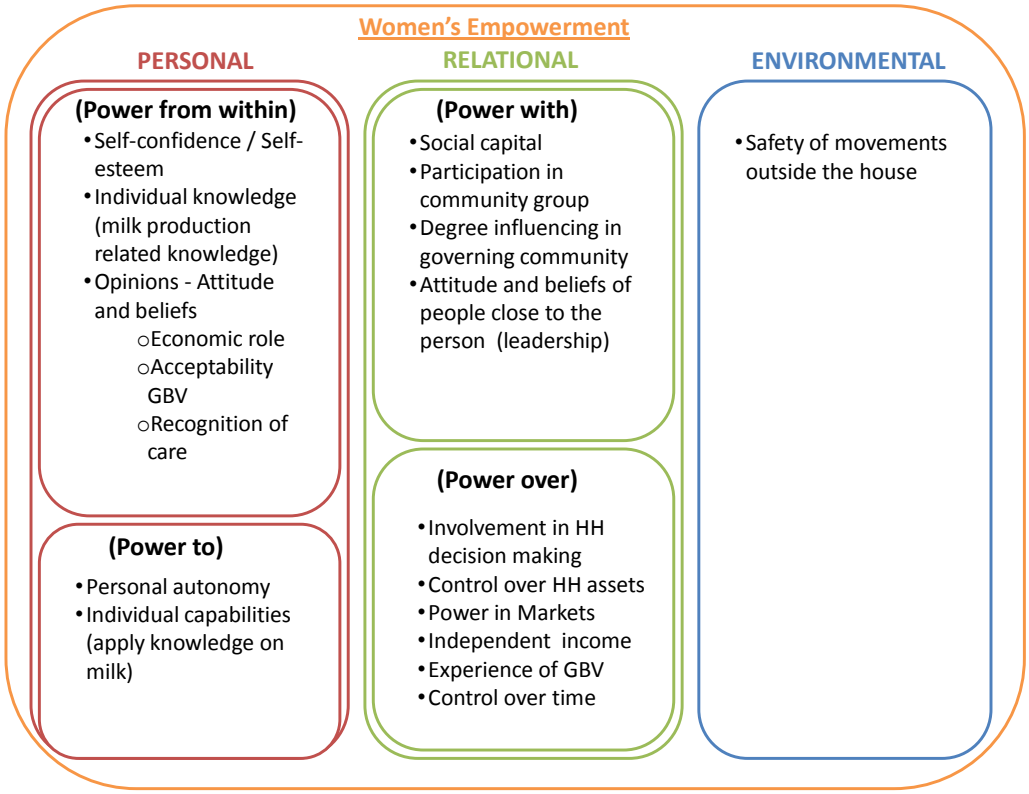
*Bootstrap Standard Errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped 1,000 repetitions*

Table 5.3 shows the difference in the wealth index measured between the intervention and comparison groups. Estimates suggest that on average there are no statistically significant differences between the intervention and comparison groups. These results are in line with results in Table 5.2 that failed to identify statistically significant differences in the levels of consumption for project participants.

5.3.4 Overall measure for women's empowerment

The project under review was specifically aimed at increasing women's empowerment. Oxfam GB has adopted and adapted an approach that assesses several dimensions of women's empowerment. This approach builds on the Women's Empowerment in Agriculture Index (WEAI) developed by the Oxford Poverty and Human Development Initiative (OPHI) with support from the United States Agency for International Development (USAID) and the International Food Policy Research Institute (IFPRI). Oxfam combined this multidimensional approach to women's empowerment with a theoretical framework on empowerment found in the literature (VeneKlasen and Miller (2002), Rowlands (1997), and CARE (2009)). The index used for this Effectiveness Review uses indicators associated with empowerment divided into three levels where change can take place (personal, relational, and environmental) and four dimensions of change in power.

Figure 5.7: Key dimensions of women's empowerment



There is no single generic set of 'women's empowerment' characteristics that are applicable to all contexts. The choice of indicators used to define and measure women's empowerment was constructed during a four-day workshop with programme staff and a partner organisation, with the explicit intention of defining what empowerment means for a woman in South Punjab. In addition to this, during the quantitative data collection survey, focus group discussions took place in order to better understand the concept of empowerment. At the end of the four-day workshop a list of 18 indicators listed in Table 5.4 was identified in order to measure women's empowerment.[6] It is important to note at this stage that while not all characteristics considered in this Effectiveness Review may be directly linked to the project activities, all were deemed to be important to women's empowerment in this particular context.

A questionnaire was designed and tested in order to include questions capturing each of the characteristics listed in Table 5.4. For each characteristic, a benchmark was defined, based on what it means for a woman to be faring reasonably well in relation to the characteristic in question. The particular benchmarks used for each characteristic are described in Appendix 1. Recognising that there is inevitably a degree of arbitrariness in defining such cut-off points, the sections which follow will present estimates of the same indicators without cut-off points, explaining in more detail the indicators and dimensions under analysis.

Table 5.4: Characteristics of women's empowerment examined in this Effectiveness Review

Level	Dimension	Characteristic
Personal	**Power from within**	Self-esteem/Self-confidence
		Individual knowledge
		Women's opinions (attitude beliefs): • Women's economic role • Acceptability GBV • Recognition of care
	Power to	Individual capability (apply knowledge)
		Personal autonomy
Relational	**Power with**	Social capital
		Participation in community groups
		Degree of influencing in governing of community groups
		Attitudes and beliefs of the persons close to the woman
	Power over	Involvement in household decision-making
		Control over household assets
		Independent income
		Power in markets
		Experience of GBV
		Control over time
Environmental		Safety of movements outside the house

This report will now consider how project participants differ from comparison women in each of the empowerment characteristics listed in Table 5.4. First, however, we examine how all of the characteristics combine to provide an overall measure of women's empowerment. This is the proportion of characteristics in which women scored positively, which we define as the *empowerment index*.[7] Table 5.5 presents the differences between the women surveyed in the project and comparison communities in terms of this overall empowerment index. On average, women participating in the

project appear to have higher levels of women's empowerment expressed using the empowerment index explained above.

Table 5.5: Overall women's empowerment index

	Women's empowerment Index
Intervention group mean:	0.574
Comparison group mean:	0.477
Difference:	0.097***
	(0.013)
Observations intervention:	225
Observations:	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

Figure 5.8: Results for dimensions of women's empowerment

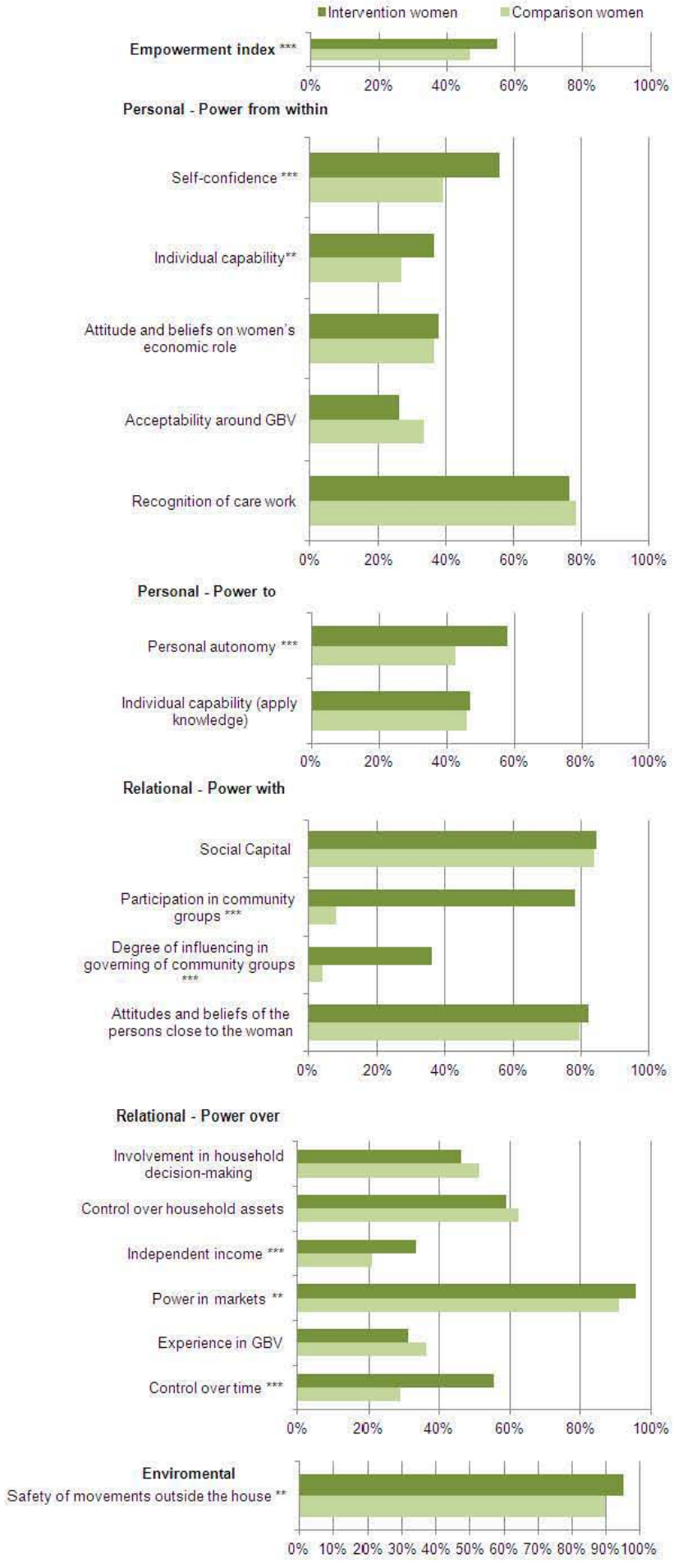


Table 5.6 presents the indicators in four levels of power: power within, to, with, and over. It suggests that the project was successful in changing power relations in all four levels of power.

Table 5.6: Women's empowerment – power dimensions

	Power within	Power to	Power with	Power over
Intervention group mean:	0.465	0.522	0.702	0.535
Comparison group mean:	0.428	0.440	0.440	0.485
Difference:	0.036*	0.082**	0.262***	0.050**
	(0.019)	(0.032)	(0.022)	(0.020)
Observations intervention:	225	225	225	225
Observations:	729	729	729	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

The following sections present in detail indicators and dimensions included in the index presented above. Figure 5.8 provides a graphical representation of the difference between intervention and comparison groups for each indicator employed in the construction of the women's empowerment index.

Personal

The first level of change refers to changes taking place within the person. Within this level of change power can express itself either from the *from within* or *to*.

Power from within

This dimension measures changes in personal self-confidence and self-esteem, personal opinions, attitudes and beliefs. These dimensions establish a change in which a woman sees and perceives herself and other women in society. However, this change in perception does not require any change in behaviour as a consequence.

In the context under analysis the following indicators have been identified:

1. Self confidence/self-esteem
2. Individual knowledge
3. Attitude and beliefs about women's economic role
4. Acceptability around GBV
5. Recognition of care work.

Table 5.7: Power from within

	Self-confidence – number	Knowledge – number	Opinion Women's economic role – number	Attitude to GBV – number	Recognition of care – number
Intervention group mean:	2.271	1.907	1.267	2.573	1.658
Comparison group mean:	2.042	1.663	1.211	2.493	1.685
Difference:	0.229***	0.243**	0.056	0.081	-0.027
	(0.081)	(0.095)	(0.071)	(0.259)	(0.056)
Observations intervention:	225	225	225	225	225
Observations:	729	729	729	729	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

Self confidence/self-esteem measures the attitude the respondent has towards herself. Respondents were asked to which extent they agreed or disagreed with the following statements:

- I feel that I'm a person of worth, at least on an equal plane with others
- I feel that I have a number of good qualities
- I feel I do not have much to be proud of
- I am equal to my peers (e.g. sisters, friends, colleagues, etc.)

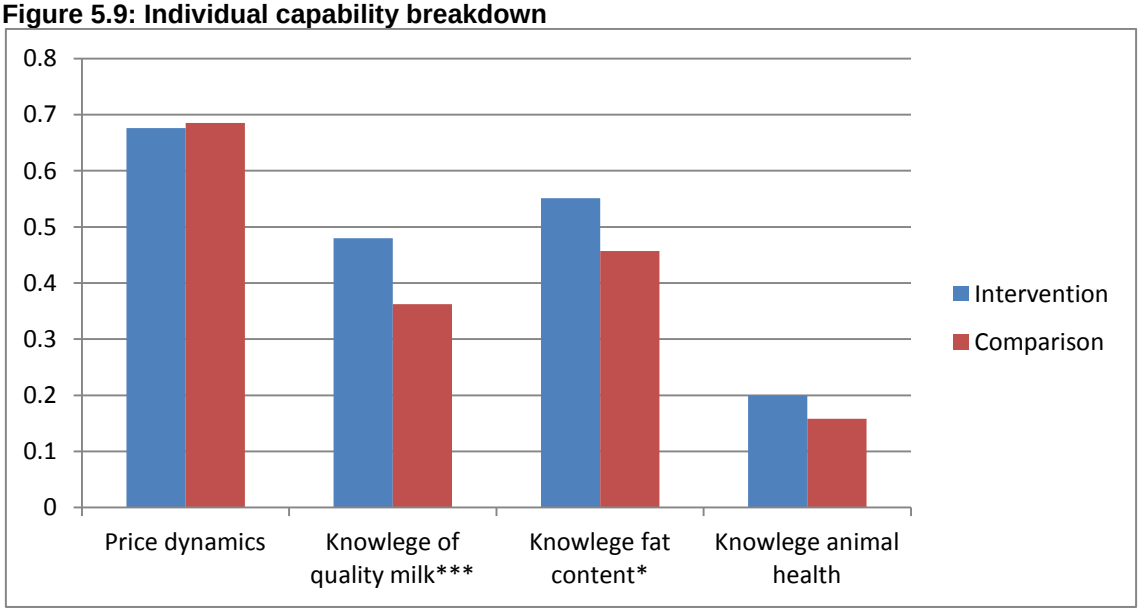
The first column in Table 5.7 provides estimates on the number of responses indicating self-confidence in the intervention and comparison group. Out of the four statements, women in the intervention group reported 2.2 answers of self-esteem, compared to an average of 2.04 in the comparison group. This difference is significantly different from zero.

Individual capability measures knowledge on the dairy sector expected to be gained from training provided by the project. In order to test their increased knowledge respondents were asked the following questions:

- A milk producer group collects and sells milk to one processor. They are paid the standard rate 35 rupees per litre. Another firm enters the market and approaches the milk producer group, asking if they can buy the milk instead. What is likely to happen to the price? Answers: go up, down, stay the same, I don't know. (Correct answer: go up)
- Did you know that some buyers pay different amounts depending on the quality of the milk? What is the fat content of the best-valued quality milk from a cow? (Correct answer: 3–4%)
- Do you know the fat content of your milk?
- In which period of the year is it more likely for cows to suffer from black quarter? (Correct answer: March to May)

The second column in Table 5.7 provides estimates on the average number of correct responses in the intervention and comparison groups. On average it appears that women in the participating into the project answered correctly 1.9 out of 4 questions, compared with 1.6 correct answers in the comparison group. Figure 5.9 provides a visual representation of the responses rate for each question. These estimates suggest that the training provided by the project was successful in improving knowledge about milk quality and fat content.

Figure 5.9: Individual capability breakdown



The third column in Table 5.7 provides estimates on attitudes and beliefs regarding **women's economic role.** Respondents were given the following pairs of statements and asked with which they most agreed:

- A woman can be a leader, just like a man can/Men are better leaders than women
- It is a waste of time to train a woman to keep financial records when you could train a man and he will do the job better/It is good to train a woman to keep financial records because she can do the job as well as a man
- A good marriage is more important for a girl than a good education/A good education is more important for a girl than a good marriage

On average, there appears to be no difference between the intervention and comparison groups in the opinion of respondents on women's economic role.

Acceptability towards GBV measures the extent to which a woman considers violence against women acceptable. Respondents were asked whether they believed it is acceptable for a man to hit his wife in the following instances:

- If she disobeys her husband or other family members
- If he suspects that she has been unfaithful
- If she neglects the children
- If she spends money without permission
- If she is not supporting her husband in livestock and agricultural activities
- If she goes to see her family without the permission of her husband

The fourth column in Table 5.7 provide estimates of the number of questions in which the respondent thought it was never acceptable, with an indicator ranging from zero to six. On average women in both the intervention and comparison groups reported that is acceptable for a man to hit his wife in approximately 2.5 cases out of 6. There is no statistically significant difference between the intervention and comparison groups, which suggests that the project was not successful in changing attitudes towards gender-based violence. The qualitative fieldwork underlined a high level of discomfort felt by respondents in discussing issues of gender-based violence. While gender-based violence may be a significant problem, its discussion appeared to be taboo.

Recognition of care work measures the extent to which women recognise that unpaid care work should be distributed equally in the household. Respondents were asked the extent to which they agreed or disagreed with the following statements:

- Certain care tasks (housework and care of persons) should also be done by other members of my household, including my husband
- I can convince my husband to complete care work (housework, care of persons)

The final column in Table 5.7 suggests that there is no statistically significant difference in the way project participants recognise care work, compared with similar women in the comparison group.

Power to

This second dimension of power measures changes in individual agency, meaning the capability to decide actions and carry them out. While in the previous dimension it was an internal process in how a woman perceives herself, in this dimension it is required to exercise the agency and carry out the action.

In the context under analysis the following indicators have been identified:

1. Personal autonomy
2. Individual capability (apply knowledge)

Table 5.8: Power to

	Personal autonomy – proportion	Individual capability (apply knowledge)
Intervention group mean:	0.598	0.467
Comparison group mean:	0.451	0.457
Difference:	0.147***	0.010
	(0.050)	(0.049)
Observations intervention:	219	225
Observations:	696	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

Personal autonomy is the capacity to decide actions for oneself and pursue a course of action in one's life, often regardless of any particular moral content. For each of the following decision-making areas, the respondent was first asked who normally takes the decisions about that area (if it was at all applicable to the household) and then, if the woman reported not being the one responsible or the only one responsible, to what extent she thought she could influence the decision on a scale from 'not at all' to 'a large extent'.[8]

- Whether you can personally travel to visit relatives outside the community
- Whether you can personally participate in community group activities or meetings

The first column of Table 5.8 provides estimates of the proportion of answers in which the respondent can make the decision herself or influence it to a great extent. On average, women in the intervention group reported being able to take decisions on travel and community-group participation in almost 60 per cent of cases. This represents a statistically significant difference compared with the comparison group, where women reported being able to take decisions autonomously on average in only 45 per cent of the cases. This indicator suggests that the project was successful in increasing personal autonomy.

Individual capability (apply knowledge to milk production) measures to what extent the knowledge acquired is then put into practice. Respondents were asked:

- Did you vaccinate your cows in the last year?
- Did you de-worm your cows in the last year?
- Did you ever use artificial insemination on your cows or buffalos?
- If you did wash the container and utensils for the milk yesterday. What was the main material used for washing it? (Soap)

The second column of Table 5.8 provides estimates for a variable equal to one if the respondent answered positively to all the previous questions, zero otherwise. On average, no statistically significant difference between groups was found in individual capability. Figure 5.10 provides a visual representation of the breakdown of the individual capability measure.

Figure 5.10: Individual capability breakdown

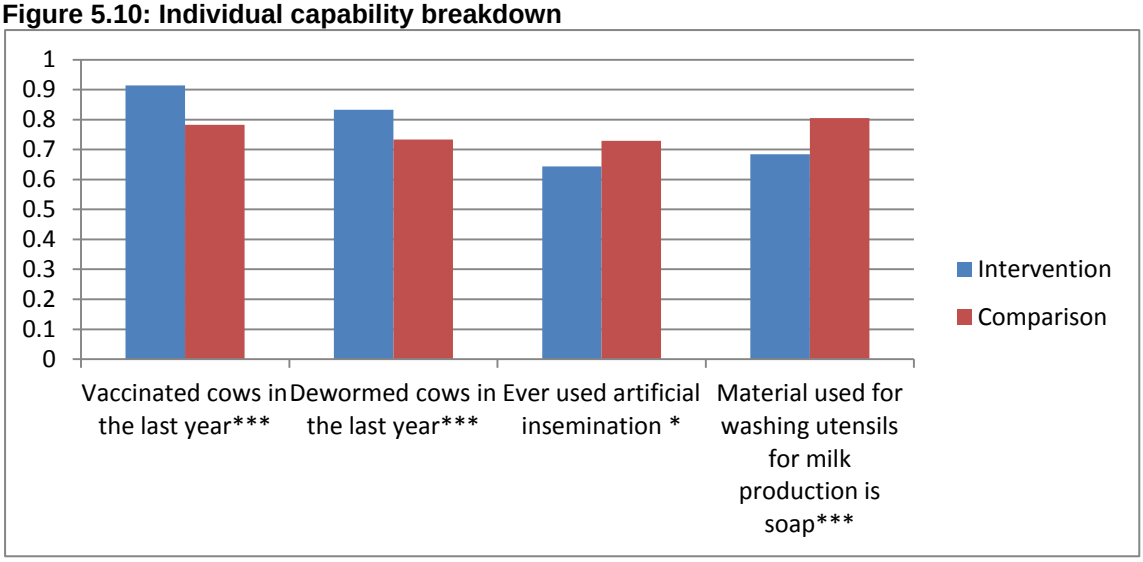


Figure 5.10 suggests that the project was successful in promoting vaccination and de-worming practices among the project participants. On the other hand, there is evidence that project participants were less likely to use more sophisticated practices, such as artificial insemination, as well as basic hygiene practices, such as cleaning utensils used for milk production process with soap. It should be noted that in some cases, project training benefited the husbands rather than the women. For example, some women who participated in the qualitative study reported did not know how to vaccinate an animal because it was their husbands who were sent to Islamabad to receive training on feed preparation and injections.

Relational

This second level of change measures changes taking place in power relations within the woman's surrounding network. The dimensions *power with* and *power over* both require changes in power relations in the interactions with other actors.

Power with

This dimension reflects the recognition that empowerment is a *collective process*, which requires the support and interaction of other peers and organisations. In the context of the project this evaluation identified the following indicators.

1. Social capital
2. Participation in community groups
3. Degree of influencing in governing community groups
4. Attitudes and beliefs of people close to the person

Table 5.9: Power with

	Social capital – number	Participation in community groups	Degree of influencing in governing	Attitudes and beliefs
Intervention group mean:	2.800	1.871	0.662	1.804
Comparison group mean:	2.761	0.132	0.093	1.695
Difference:	0.039	1.739***	0.569***	0.109**
	(0.046)	(0.106)	(0.092)	(0.055)
Observations intervention:	225	225	225	225
Observations:	729	729	729	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

Social capital measures to what extent a woman has access to a network of people or groups to support her, and to what extent she makes use of that network. Respondents were asked the extent to which they agreed or disagreed with the following statements:

- I can contact my parents and siblings if I need their support
- I feel that I have respect and friendship in my husband's family
- If I needed I could ask a favour of my neighbours

The first column in Table 5.9 provides estimates the number of answers where the respondent agreed with the above statements. On average, women in the intervention and comparison groups reported a similar number of answers indicating they benefited from social capital.

Participation in community group measures the extent to which a woman participates in groups – assuming that such participation brings benefits, such as the ability to interact with other people outside the family. Opportunities for formal participation in associations are limited for many women in rural Pakistan, with restrictions, for example, in participating in village-level governance groups or political parties. Respondents were asked if they regularly attended meetings of the following groups: farmer's groups/cooperatives, board of enterprise, milk collection centre, early warning, religious or political group, local NGOs or community organisations, or in any other formal or informal group.

The second column in Table 5.9 provides estimates on the number of groups in which the respondent was involved. This is a number, which can vary from zero to seven. On average women in the intervention group reported participation in 1.8 groups, compared to only 0.13 in the comparison group. This is a large and statistically significant difference of, on average, almost two groups.

Upon further consideration, we believe this to be a positive, but logical and somewhat unsurprising finding – women in intervention areas (where the intervention involves forming community groups) are more likely to be group members. Therefore, this should not be considered an outcome-based measure, but rather an output-based measure. The indicator was nonetheless retained in the index because the qualitative component suggested that in becoming members of village group women had the opportunity to interact more frequently with their fellow female villagers, sitting together and sharing their problems beyond milk production and finding solutions together to overcome these problems.

Degree of influencing in governing community groups measures the extent to which a woman has decision-making power in the groups she is involved in. The respondents were asked to what extent they were involved in making important decisions in the groups they regularly attended. The third column in Table 5.9 provides estimates of the number of groups women were members of in which they were involved in taking important decisions. On average, women in the intervention group report taking important decisions in 0.6 groups, compared with less than 0.1 groups for women in the comparison group.

This is considered to be a mark of success for the project, as it is evidence that women were *active* participants in local groups.

This finding is in line with what was identified during the qualitative component, suggesting that women can gain confidence from meaningful participation in project activities. For example, should they become members and attend board meetings, they are able to make suggestions and raise voices to the benefit of milk-producing women.

Attitudes and beliefs of people close to the person measures to what extent women's families would invest in the education of daughters and recognise the possibility of women having leadership positions. Respondents were asked the extent to which they agreed or disagreed with the following statements:

- If there was a school in our village, I think my family would send our daughter
- My family and neighbours are supportive of women having a leadership position in community groups.

The last column in Table 5.9 provides estimates on the number of questions where the respondent agreed or strongly agreed with the previous statements. The variable ranged from zero to two, and suggests that on average women in the intervention group reported positive attitudes in 1.8 out of 2 answers relative to women in the comparison group who reported on average positive responses in 1.6 of the 2 cases. This difference was positive and statistically significant, suggesting that the project improved the attitudes and beliefs of people close to the women involved in the project.[9]

Power over

This dimension measures changes in the power of the strong over the weak. Also this dimension represents changes taking place in the power relationship between individuals. In the context of the project, this evaluation identified the following indicators.

1. Involvement in household decision-making
2. Control over household assets
3. Power in markets
4. Independent income
5. Experience of violence
6. Control over personal time

Table 5.10: Power over

	Involvement in HH decision-making	Control over household assets	Power in markets	Independent income	Experience of violence	Control over personal time
Intervention group mean:	0.520	0.591	0.956	3.444	0.311	0.556
Comparison group mean:	0.547	0.625	0.909	2.766	0.363	0.290
Difference:	-0.027	-0.034	0.047**	0.678***	-0.052	0.265***
	(0.037)	(0.044)	(0.020)	(0.170)	(0.045)	(0.047)
Observations intervention:	225	225	225	225	225	225
Observations:	729	729	729	729	729	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

Involvement in household decision-making measures the extent to which the respondent is involved in decisions taking place in her own household. Each respondent was asked who normally takes decisions in the following areas:

- Keeping and managing household income
- How to spend the money made from milk sales
- Buying and selling livestock
- How much money to invest in business activities
- Deciding how much to borrow and save
- What food to buy and consume

- How children should be educated
- Approve marriages
- Transfer of property to a relative or any other person
- How many children to have
- Purchase of furniture for the house
- Housework and care of persons

If the woman reported not to be the one responsible or not to be the only one responsible, than she was asked to what extent she thought she could influence the decision, on a scale from 'not at all' to 'a large extent'. Estimates in the first column in Table 5.10 report the proportion of activities in which the respondent takes the decision by herself or in which she is able to influence others to a large extent. On average, it seems women are involved in roughly half the household decisions, and there is no statistically significant difference between the women in the intervention and comparison groups.

Qualitative fieldwork suggests that this evaluation may have failed to measure an important dimension of women's empowerment in the family. Respondents indicated that some men took a second wife. This would represent a significant 'decision' for the household – and potentially a hugely *dis*empowering event for the women involved. We recognise that the relative absence of questions relating to marriage practices (aside from marriage of children) is possibly a considerable limitation of the empowerment framework in this context.

Control over household assets measures women's ownership and control over strategic assets, such as land, livestock or household equipment (time saving equipment). Respondents were asked about their household's ownership of various types of asset. As a follow-up to these questions, they were then asked to specify which household member could make decisions about whether to sell, trade or give away an item if need be. Estimates in the second column in Table 5.10 report the proportion of assets over which a woman has joint control. On average, women in both the intervention and comparison groups reported having control over roughly 60 per cent of the household assets, with no statistically significant difference between the two groups.

The qualitative research component addressed the role played by women in conducting livestock activities. The quantitative survey component suggested that women in the intervention group were on average more in control over livestock, such as cows, buffalos, sheep, goats and poultry, compared with the women in the comparison group.

This suggests that the project may be successful in improving women's control over livestock assets, an area that traditionally belongs to women.

For the indicator **power in markets** respondents were asked the extent to which they agreed or disagreed with the following statement:

- It is important for farmers to work together in order to get a better price for their products

The third column in Table 5.10 provides estimates of the variable equal to one if the respondent agreed or strongly agreed to the statement above, zero otherwise. On average, 95 per cent of women in the intervention group agreed to the statement above, compared with 90 per cent of women in the comparison group, with a statistically significant difference of five percentage points.

Independent income attempts to measure the proportion of income women earn independently from their spouse or other household members.

Given the low level of literacy among the survey participants, this indicator was assessed in the following way. Respondents were asked:

- Here are ten small beans. The beans together represent all the resources your household needs, such as food and money. From what you receive, either crops or cash, how many beans represent your contribution?

Estimates in the fourth column in Table 5.10 provide the average number of beans women in the intervention and comparison group reported to contribute to household need. On average women in the intervention group reported contributing almost 35 per cent of the total household income, compared with 28 per cent in the comparison group. This difference is statistically significant, and it represented a crucial indicator of success in the project's theory of change, suggesting that the project increased the proportional contribution of women to their households' incomes.

Qualitative observations suggest that income from the sale of milk is spent on the household's needs. It is unclear if women themselves are able to fully control/influence how this additional income is spent. It is possible that woman have control and influence over this additional income that they have earned. It is also possible that formal or informal pressures mean they are not able to fully choose how to use the additional income.

The indicator on **experience of violence** aims to measure the extent to which women are exposed to violence, asking if in the previous month someone had done any of the following to a woman close her:[10]

- Said something to humiliate her in front of others
- Threatened to hurt or harm her or someone she cares about
- Insulted her or made her feel bad about herself
- Pushed her, shaken her, slapped or punched her or thrown something at her
- Threatened or attacked her with a knife, gun or other weapon

The fifth column provides estimates on the probability that respondents reported that a woman close to them had been exposed to violence. This variable takes value equal to one if the respondent reported not being exposed to all of the five violent events listed above, zero otherwise. Estimates suggest that on average 31 per cent of women in the intervention group never reported any event of violence in the past month, compared with 36 per cent of women in the comparison group, with differences not being significantly different from zero.

These estimates are very different if looking at the probability that respondents reported that a woman close to them had been exposed to violence. On average, 18 per cent of women in the intervention group reported at least one case of violence that had happened to a woman close to her, compared with 10 per cent of the women in the comparison group. It is difficult to determine if this difference is due to an increased level of violence or rather to an increased confidence in reporting these events. Unfortunately the qualitative component was not able to explore this component.

Finally **control over time** is estimated with a variable equal to one if the woman reported that time devoted to personal activities, meaning leisure time (e.g. socialising with neighbours), sleeping at night, and personal care and rest, had increased since the project activities begun, zero otherwise. On average, more than 55 per cent of the women in the intervention group reported an increase in personal time, compared with almost 30 per cent in the comparison group. This difference of 26 percentage points is statistically different from zero. This is notable, as one might expect that participation in project activities would have a negative effect on women's control over their time – but this appears to not have been the case.

Environmental

The final level of empowerment measures changes in the broader environment. Only one indicator was indentified for this evaluation:

1. Safety of movements outside the house

In order to measure **safety of movements outside the house** respondents were asked the extent to which they agreed or disagreed with the following statement:

- I feel safe to walk alone in my village

Table 5.11 estimates the proportion of women in the intervention and comparison groups who reported feeling safe walking alone in their village. On average, more than 95 per cent of the women in the intervention group agreed to feeling safe to walk alone in their village, compared with almost 90 per cent in the comparison group, with a difference of 5 percentage points that is statistically different from zero.

The qualitative data do not clearly suggest any potential explanations for this difference.

Table 5.11: Environment

	Safety of movements outside the house
Intervention group mean:	0.951
Comparison group mean:	0.897
Difference:	0.054**
	(0.021)
Observations intervention:	225
Observations:	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

6 CONCLUSIONS

6.1 CONCLUSIONS

This evaluation provides evidence on a project that has no results on income and markets, but which is making an impact on some women's empowerment indicators.

Both qualitative and quantitative data provide evidence of improved knowledge about the milk and dairy sector and milk production among project participants as a result of project intervention. However, the evaluation did not find evidence that this improved knowledge translated into a higher quantity and quality of milk production.

There might be some indications suggesting that the project may have contributed to shaping the local dairy market as focus group discussions mentioned an increase in milk price. However, this evaluation is not able to provide evidence that this change is attributable to the project intervention alone. Moreover, survey data provide evidence that current prices of milk paid by the enterprise are higher than the price paid by middle-men, but are in line with the other prices paid by other milk competitors operating in the area. On average, it appears that price per litre paid to project participants is lower than the average price paid to the comparison group, which is in line with the implementation problems identified with the qualitative component. In some focus group discussions, respondents expressed concerns about the efficiency of some collecting centres, which allegedly paid after milk delivery rather in advance, and were less reliable in collecting milk than other channels and competitors in the formal market.

Overall there is no evidence that the project had an impact on income and on wealth levels of households involved in the project.

Despite these modest results on income, markets and livelihood practices, the project was found to be positively associated with overall higher levels of women's empowerment for women involved into the project activities. In particular, the evaluation identified positive effects on the following indicators: self-confidence, personal autonomy, group participation, independent income, power in markets, control over time, and safety of movements outside the house. However, there is no evidence of changes in important women's empowerment indicators, such as opinions on women's economic role, household decision-making power, and control over assets within the household.

6.2 PROGRAMME LEARNING CONSIDERATIONS

From the evaluation are emerging some important lessons that can be applied to other projects of this type in Pakistan and elsewhere. The Pakistan country team and the programme team in particular are encouraged to consider the following:

- **Improve project targeting and delivery quality**

The evaluation found that the project was implemented in an area where a number of other actors, both multilateral agencies and private sector companies, were already operating and working in similar thematic areas. Moreover, qualitative data suggest that some project participants perceived the enterprise as not being professionally run when compared with other businesses operating in the same market. Focus group discussions identified concerns and issues raised by project participants concerning

the management of the enterprise, lamenting problems with collection of the milk and having the milk spoiled as a consequence. Project participants also expressed dissatisfaction with the timing of the payments made by the enterprise, compared with other actors in the area.

This raises some questions about the value added of projects establishing a private enterprise in a context where there are already other private companies equipped with better skills and business means. A more careful targeting process of the intervention area should be put in place when selecting the market and location for project implementation.

- **Increase clarity over women's empowerment outcomes and pathways to change**

The evaluation identified no impact on income and wealth, but it did identify a positive impact attributable to the project on women's empowerment.

Survey results, as well as qualitative data, suggest that activity engagement with women at community level had a positive impact on a number of women's empowerment indicators. This, however, was achieved with a theory of change that lacked a clear definition of women's empowerment outcomes or pathways of change.

The focus of the project for empowering women was only through higher contribution to household income, and improved leadership for four women on the enterprise board. The evaluation identified that the project had a positive impact on increasing the proportion of household income earned independently from other household members, but not on attitudes and beliefs about women's economic role, household decision-making, and control over assets.

The programme team has been encouraged to consider scaling up the women's empowerment components of this project, defining whether an increase in household income is sufficient for women's empowerment or if other issues should be explicitly targeted and addressed. What the project means by empowerment and how change is expected to take place, including defining a theory of change that differentiates how change happens for women involved into the project and for women whose husbands participate, should be clearly defined.

- **Define what constraints are facing small milk producer farmers other than knowledge**

One of the assumptions of the project was that providing training and increasing knowledge would also increase the quality and quantity of milk production. The evaluation identified that project activities had a positive impact on improving knowledge of milk quality and improved vaccination practices. However, there is no evidence of higher quantity or quality of milk being produced as a consequence of the project.

The programme team has been encouraged to explore if there are other constrains that milk-farm producers are facing, other than limited knowledge.

- **Consider evaluation questions during programme design**

In future projects, if there is an interest in exploring impact questions, it has been advised to consider including impact evaluation frameworks in the project design.

Impact evaluation is a key tool for learning, to help projects and programmes succeed and generate evidence of success. When designing a project, the programme team is encouraged to consider and define what the questions within the evaluation should address; what are the components and characteristics of the intervention that should be evaluated; and finally, what are the reasons for conducting the evaluation (e.g. influencing, accountability, and learning). Depending on the answers to these points, the programme team has been encouraged to plan budget, time and resources

adequately. Different evaluation designs and methodologies provide evidence of impact with different levels of confidence. For large-scale development intervention, a counterfactual allows the attribution of change to the project intervention.

APPENDIX 1: THRESHOLDS FOR CHARACTERISTICS OF WOMEN'S EMPOWERMENT

Level	Dimension	Characteristic	Threshold: a women scores positively if she...	Coherent with continues estimates	Estimate difference
Personal	**Power from within**	Self-confidence	… 'agrees' or 'strongly agrees' with • I feel that I'm a person of worth, at least on an equal plane with others • I feel that I have a number of good qualities • I am equal to my peers (e.g. sisters, friends, colleagues, etc.) OR 'disagrees' or 'strongly disagrees' with at least half of the statements: • I feel I do not have much to be proud of	Yes	0.165*** (0.041)
		Individual capability (gain knowledge)	… answers at least two correct answer that tests knowledge on the milk sector: • A milk producer group collects and sells milk to one processor. They are paid the standard rate of 35 rupees per litre. Another firm enters the market and approaches the milk producer group, asking if they can buy the milk instead. What is likely to happen to the price? • Did you know that some buyers pay different amounts depending on the quality of the milk? What is the fat content of best valued quality milk from cow? (Correct answer is 3–4%) • Do you know the fat content of your milk? • In which period of the year is more likely for cows to suffer from black quarter? (Correct answer: March to May)	Yes	0.095** (0.041)
		Attitude and beliefs on women's economic role	… 'agrees' or 'strongly agrees' with the statement • A woman can be a leader, just like a man can/Men are better leaders than women Or 'disagrees' or 'strongly disagrees' with at least two of the three statements: • It is a waste of time to train a woman to keep financial records, when you could train a man and he will do the job better/It is good to train a woman to keep financial records, because she can do the job as good as the man • A good marriage is more important for a girl than a good education / A good	Yes	0.012 (0.043)

Level	Dimension	Characteristic	Threshold: a women scores positively if she...	Coherent with continues estimates	Estimate difference
			education is more important for a girl than a good marriage		
		Acceptability around GBV	…reports being unacceptable for a man to hit his wife in *all* of the following cases: • She disobey her husband or other family members • He suspects that she has been unfaithful • She neglects the children • She spends money without permission • She is not supporting her husband in livestock and agricultural activities • She goes to see her family without the permission of her husband	Yes	-0.072 (0.045)
		Recognition of care work	…'agrees' or 'strongly agrees' with both of the following statements: • Certain care tasks (housework and care of persons) should also be done by other members of my household, including my husband • I can convince my husband to complete care work (housework, care of persons)	Yes	-0.018 (0.038)
	Power to	Personal autonomy	…can take the decision herself or is can influence to a great extent both of the following actions: • Whether you can personally travel to visit relatives outside the community • Whether you can personally participate in community group activities or meetings	Yes	0.154*** (0.047)
		Individual capability (apply knowledge)	…answered positively to all the following questions: • Did you vaccinate your cows in the last year? • Did you de-worm your cows in the last year? • Did you ever use artificial insemination on your cows or buffalos? • If you did wash the container and utensils for the milk yesterday. What was the main material used for washing it? (Multiple choices. Correct answer: soap)	Yes	0.010 (0.049)
Relational	**Power with**	Social capital	… 'agreed' or 'strongly agreed' to all the following statement: • I can contact my parents and siblings if I need their support • I feel that I have respect and friendship in my husband's family • If I needed I could ask a favour of my neighbours	Yes	0.004 (0.031)
		Participation in community groups	…participates at least in one group	Yes	0.700*** (0.035)

Level	Dimension	Characteristic	Threshold: a women scores positively if she...	Coherent with continues estimates	Estimate difference
		Degree of influencing in governing of community groups	…is involved to a large extent in taking important decisions in at least one group	Yes	0.317*** (0.038)
		Attitudes and beliefs of the persons close to the woman	If she agrees or strongly agrees on both these statements • If there was a school in our village, I think my family would send our daughter • My family and neighbours are supportive of women having a leadership position in community groups.	No	0.027 (0.036)
	Power over	Involvement in household decision making	...is involved in or could influence to large extent at least 50% of the decisions taken within the household	Yes	-0.050 (0.046)
		Control over household assets	…has decision-making power over at least 75% of the 24 assets listed in the questionnaire	Yes	-0.034 (0.044)
		Independent income	…reports contributing at least half of the household income.	Yes	0.124*** (0.038)
		Power in markets	… 'agreed' or 'strongly agreed' with the following statement: • it is important for farmers to work together in order to get a better price for their products.	Yes	0.047** (0.020)
		Experience in GBV	…didn't report any experience of violence to women close to her	Yes	-0.052 (0.046)
		Control over time	…reported increasing time devoted to leisure, sleeping at night, and personal care and rest	Yes	0.265*** (0.047)
Environmental		Safety of movements outside the house	…she reported feeling safe to walk alone in her village.	Yes	0.054** (0.021)

APPENDIX 2: METHODOLOGY USED FOR PROPENSITY-SCORE MATCHING

The analysis of outcome variables, presented in Section 5 of this report, involved group mean comparisons using propensity-score matching (PSM). The basic principle of PSM is to match each participant with a non-participant that was observationally similar at baseline and to obtain the treatment effect by averaging the differences in outcomes across the two groups after project completion. Unsurprisingly, there are different approaches to matching, i.e. to determining whether or not a household is observationally 'similar' to another household. For an overview, we refer to Caliendo and Kopeinig (2008).[11] This appendix describes and tests the specific matching procedure followed in this Effectiveness Review.

Estimating propensity scores

Given that it is extremely hard to find two individuals with exactly the same characteristics, Rosenbaum and Rubin (1983) demonstrate that it is possible to match individuals using a prior probability for an individual to be in the intervention group, naming it *propensity score*. More specifically, propensity scores are obtained by pooling the units from both the intervention and comparison groups and using a statistical probability model (e.g. a probit regression) to estimate the probability of participating in the project, conditional on a set of observed characteristics.

Table A2.1 presents the probit regression results used to estimate the propensity scores in our context. To guarantee that none of the matching variables were affected by the intervention, we only considered variables related to baseline, and only those variables that were unlikely to have been influenced by anticipation of project participation (Caliendo and Kopeinig, 2008).

Table A2.1: Estimating the propensity score

	Intervention
1[HHH has no formal education]	0.134
	(0.124)
1[Respondent has no formal education]	-0.028
	(0.210)
Household size 2009	0.003
	(0.019)
1[Head of HH is female]	0.204
	(0.190)
1[Household was affected by the flood in 2010]	-0.335***
	(0.111)
Distance of the house from the river in 2009 (in km)	0.001
	(0.008)
Total area cultivated in 2009	-0.026**
	(0.012)
Number of groups involved in 2009	1.900***
	(0.275)
1[Household involved in dairy sector in 2009]	-0.359***
	(0.131)
1[Household was farming in 2009]	0.067
	(0.246)
1[Household involved in labour sector in 2009]	-0.120
	(0.111)
1[Household involved in private business in 2009]	-0.005
	(0.167)
1[HH is in the second wealth quintile]	0.014
	(0.163)
1[HH is in the third wealth quintile]	-0.413**
	(0.171)
1[HH is in the fourth wealth quintile]	0.175
	(0.163)
1[HH is in the fifth wealth quintile]	-0.017
	(0.178)
_cons	-0.268
	(0.377)
Number of observations	804

Standard errors in parentheses

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

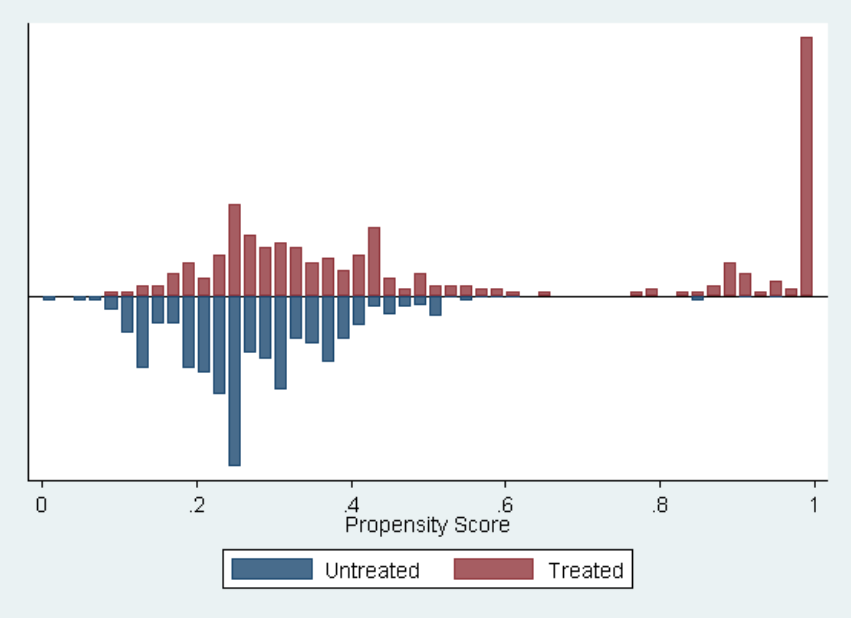
Defining the region of common support

After estimating the propensity scores, the presence of a good *common support area* needs to be checked. The area of common support is the region where the propensity score distributions of the treatment and comparison groups overlap. The common support assumption ensures that 'treatment observation have a comparison observation "nearby" in the propensity score distribution' (Heckman, LaLonde and Smith, 1999). Since some significant differences were found between the intervention and comparison groups in terms of their baseline characteristics (as detailed in Section 4.2), some of the women in the intervention group were too different from the comparison group to allow for meaningful comparison. We used a minima and maxima comparison, deleting all observations whose propensity score is smaller than the minimum and larger than the maximum in the opposite group (Caliendo and Kopeinig, 2008). Seventy-five of the 300 women interviewed in the project communities and were

dropped because they lay outside the area of common support. The consequence of dropping project participant households is that the estimates of differences in outcome characteristics between the various treatment groups only apply to those intervention households that were not dropped; that is, they do not represent the surveyed population as a whole.

Figure A2.1 illustrates the propensity scores and shows the proportion of women lying on and off the areas of common support, by treatment group.

Figure A2.1: Propensity score on and off area of common support



Matching intervention and comparison households

Following Rosenbaum and Rubin (1983), after estimating the propensity scores and defining the area of common support, individuals are matched on the basis of their propensity score. The literature has developed a variety of matching procedures. For the main results presented in this Effectiveness Review we chose to employ the method of kernel matching. Kernel matching weights the contribution of each comparison group member, attaching greater weight to those comparison observations that provide a better match with the treatment observations. One common approach is to use the normal distribution with mean zero as a kernel, and weights given by the distribution of the differences in propensity score. Thus 'good' matches are given greater weight than 'poor' matches.

The *psmatch2* module in Stata was used with a bandwidth of 0.06 and with the analysis restricted to the area of common support.

When using PSM, standard errors of the estimates were bootstrapped using 1,000 repetitions (clustered by village), to account for the additional variation caused by the estimation of the propensity scores and the determination of the common support.[12]

Check balancing

For PSM to be valid, the intervention group and the matched comparison group need to be balanced, in that they need to be similar in terms of their observed baseline characteristics. This should be checked. The most straightforward method to do this is to test whether there are any significant differences in baseline covariates between the intervention and comparison group in the matched sample, as reported in Table A2.2. None of the variables implemented for the matching are statistically significant in the matched sample.

Table A2.2: Balancing test

	Unmatched	Mean		t-test	
Variables	Matched	Treated	Control	t	p>|t|
1[HHH has no formal education]	U	0.76	0.73016	0.93	0.351
	M	0.74667	0.72348	0.56	0.578
1[Respondent has no formal education]	U	0.94	0.93452	0.31	0.758
	M	0.93333	0.88663	1.73	0.084
Household size 2009	U	6.1533	6.2163	-0.31	0.76
	M	6.2133	5.9493	0.99	0.322
1[Head of HH is female]	U	0.93667	0.90476	1.58	0.114
	M	0.92889	0.9161	0.51	0.613
1[Household was affected by the flood in 2010]	U	0.62333	0.65476	-0.9	0.369
	M	0.56	0.57449	-0.31	0.757
Distance of the house from the river in 2009 (in Km)	U	5.537	5.6087	-0.15	0.877
	M	5.7493	5.5321	0.36	0.715
Total area cultivated in 2009	U	1.7907	2.9074	-1.96	0.051
	M	1.8462	1.7548	0.31	0.756
Number of groups involved in 2009	U	0.71	0.00992	13.16	0
	M	0.10667	0.10228	0.15	0.879
1[Household involved in dairy sector in 2009]	U	0.75	0.81746	-2.29	0.023
	M	0.72889	0.73365	-0.11	0.91
1[Household was farming in 2009]	U	0.94	0.95635	-1.03	0.303

	M	0.94222	0.9461	-0.18	0.858
1[Household involved in labour sector in 2009]	U	0.66333	0.66071	0.08	0.94
	M	0.65778	0.61239	1	0.318
1[Household involved in private business in 2009]	U	0.12	0.10317	0.74	0.461
	M	0.12	0.13864	-0.59	0.557
1[HH is in the second wealth quintile]	U	0.22667	0.18452	1.44	0.149
	M	0.20889	0.19895	0.26	0.794
1[HH is in the third wealth quintile]	U	0.14	0.23611	-3.31	0.001
	M	0.12444	0.117	0.24	0.809
1[HH is in the fourth wealth quintile]	U	0.23667	0.17857	1.99	0.047
	M	0.25778	0.26981	-0.29	0.773
1[HH is in the fifth wealth quintile]	U	0.17333	0.21429	-1.41	0.16
	M	0.21333	0.21603	-0.07	0.945

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

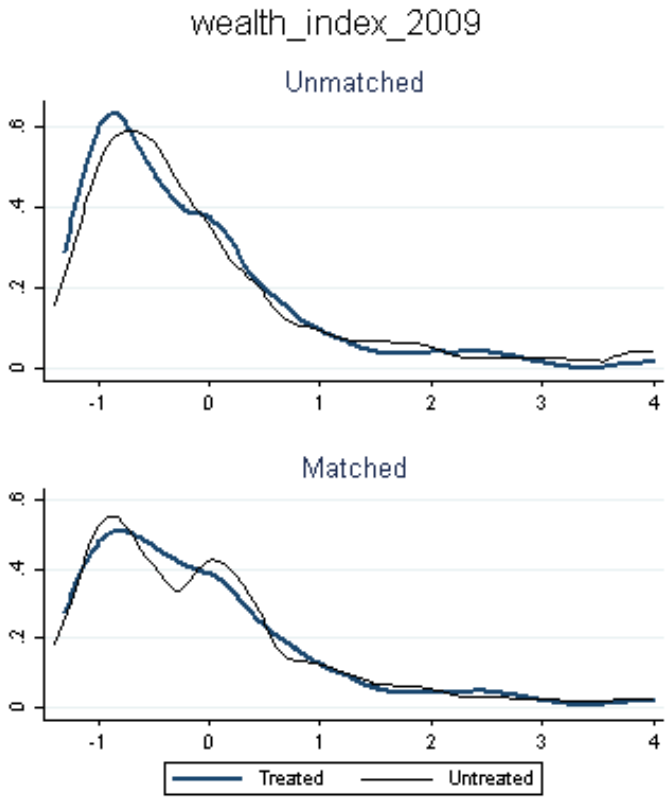
*** if 'bad', i.e. variance ratio <0.5 or >2*

Sample	Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%concern	%bad
Unmatched	0.224	237.9	0	13.7	8.9	91.9*	30.34*	38	13
Matched	0.01	6.09	0.987	4.6	3	23.2	0.91	6	0

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

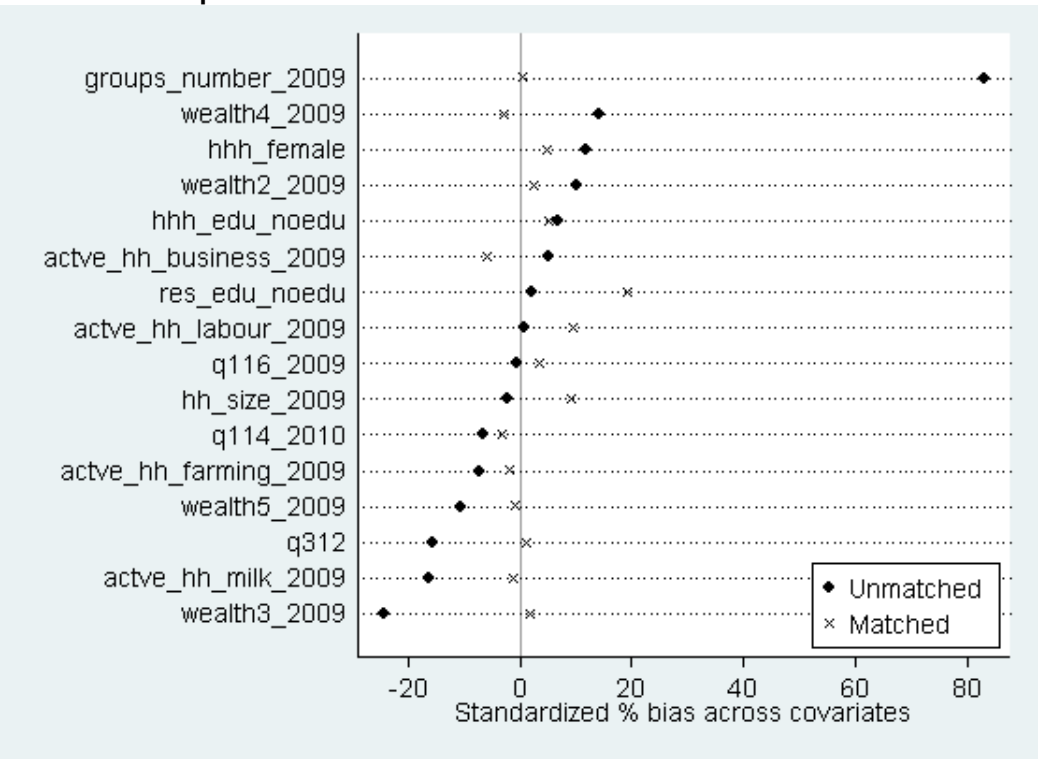
Figure A2.2 provides the distribution of the wealth index in 2009 for treated and untreated women before and after matching.

Figure A2.2: Wealth index distribution for matched and unmatched sample



Finally Figure A2.3 shows the standardised percentage of bias across matching variables for matched and unmatched samples. As already explained in Section 4, in the unmatched sample intervention and comparison differed for the average number of groups women were involved in 2009 before the project began and for wealth indicators.

Figure A2.3: Standardised % of bias across matching variables for matched and unmatched samples



APPENDIX 3: ROBUSTNESS CHECKS

In order to address the validity of the results presented in Section 5, a series of robustness checks were carried out to check if the preferred matching algorithm is the one that best performs the matching between intervention and comparison groups. This section presents a number of alternative matching algorithms used to test the robustness of the estimates presented in Section 5.

1- Multivariate regression

The first basic specification for estimating the impact of project participation is an OLS model (when the dependent is continuous) or probit model when the dependent is binary.

$$Y_i = \alpha + \beta_1 Project\ participation_i + \delta' X_i + \varepsilon_i$$

Where Y_i is the dependent variable; X_i is a vector of household covariates used in the model in table A2.1; finally the variable of interest is the dummy variable *Project Participation* that assumes value equal to one when the household is enrolled in the project, zero otherwise. When the dependent variable Y_i is binary variable, a probit model replaces the OLS specification. It is important to note that in the absence of randomised allocation of the project among the population in our sample, OLS and probit models fail to identify the causal effect of the programme, and can only be used as additional qualitative checks for the non-parametric estimates. Only the estimate of β_1 will be reported.

2 - Propensity Score Matching – Nearest Neighbour

The Nearest Neighbour (NN) matching algorithm finds an observation from the comparison group to be matched with an observation from a treated individual that is closest in terms of their propensity score. Several variants of NN matching are possible, e.g. NN matching 'with replacement' and 'without replacement'. In the former case, an untreated individual can be used more than once as a match, whereas in the

latter case it is considered only once. Matching with replacement involves a trade-off between bias and variance. If we allow replacement, the average quality of matching will increase and the bias will decrease. This is of particular interest with data where the propensity score distribution is very different in the treatment group and the control group (Caliendo and Kopeinig, 2008).

3- Propensity Score Matching – Caliper

NN matching faces the risk of bad matches, if the closest neighbour is far away. This can be avoided by imposing a tolerance level on the maximum propensity score distance (caliper). Imposing a caliper works in the same direction as allowing for replacement. Bad matches are avoided and hence the matching quality rises. However, if fewer matches can be performed, the variance of the estimates increases. Applying caliper matching means that an individual from the comparison group is chosen as a matching partner for a treated individual that lies within the caliper ('propensity range') and is closest in terms of propensity score. Estimates in this analysis will impose a caliper of 0.05.

4- Propensity Score Weighting

Following the example of Hirano and Imbens (2001)[13] we implemented a regression adjustment with weights based on the propensity score. The average treatment effect can be estimated in a parametric framework as follows:

$$Y_i = \alpha + \beta_1 Project\ participation_i + \delta_2' Z_i + \delta_1' X_i + \varepsilon_i$$

Where Y_i represents the outcome of interest; Project participation$_i$ is a dummy binary variable equal to one if an individual/household is enrolled into the programme and zero otherwise; X_i is a vector of matching covariates used to estimate the propensity score match; and Z_i is a vector of control variables which cannot be used for the matching as they are not supposed to influence project participation. The regression is estimated with weights equal to one for the treated units and $\hat{e}(x)/(1 - \hat{e}(x))$ for control units.

This parametric regression analysis framework has the advantage to explore heterogeneity in the treatment effect. Moreover it allows controlling for variables that cannot be included in the propensity score equation. The robustness check tables will only report β_1.

Table A3.1: Milk production

	(1) OLS / probit	(2) PSM NN	(3) PSM Caliper	(4) Propensity Score Weighting
Milk production	-1.446	-0.954	-0.474	-0.446
	(1.303)	(1.709)	(2.654)	(1.268)
N	794	794	794	794
Quality of milk	-1.286**	0.229	0.133	0.376
	(0.554)	(0.312)	(0.284)	(0.246)
N	421	421	421	421
Price milk	-1.377**	-0.961	-1.683	-1.286**
	(0.560)	(0.976)	(1.190)	(0.554)
N	499	499	499	499

Robust standard errors in parentheses. PSM estimates bootstrapped 1000 repetitions.

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

Table A3.2: Household consumption and wealth

	(1) OLS / probit	(2) PSM NN	(3) PSM Caliper	(4) Propensity Score Weighting
Log(Total consumption - daily per capita)	-0.009	0.028	-0.003	-0.046
	(0.208)	(0.079)	(0.113)	(0.066)
N	804	804	804	804
Log(Yearly consumption - daily consumption)	-0.020	-0.046	0.043	0.048
	(0.089)	(0.141)	(0.237)	(0.104)
N	803	803	803	803
Log(Monthly consumption - daily consumption)	-0.046	0.128	0.053	-0.020
	(0.066)	(0.096)	(0.095)	(0.089)
N	804	804	804	804
Log(Value food consumed in last 7 days - daily per capita)	0.129***	-0.082	-0.065	-0.072*
	(0.015)	(0.065)	(0.087)	(0.042)
N	803	803	803	803
Number of items consumed in the last week	-0.072*	0.088	0.182	-0.009
	(0.042)	(0.220)	(0.369)	(0.208)
N	804	804	804	804
Wealth Index	0.048	-0.310	-0.039	-0.071
	(0.104)	(0.246)	(0.374)	(0.114)
N	804	804	804	804

Robust standard errors in parentheses. PSM estimates bootstrapped 1000 repetitions.

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

Table A3.3: Women's empowerment index

	(1) OLS / probit	(2) PSM NN	(3) PSM Caliper	(4) Propensity Score Weighting
Women's empowerment index	0.098***	0.099***	0.097***	0.094***
	(0.010)	(0.017)	(0.034)	(0.012)
N	808	804	804	804
Power within	0.094***	0.046*	0.036	0.054***
	(0.012)	(0.025)	(0.026)	(0.019)
N	804	804	804	804
Power to	0.054***	0.090**	0.082	0.043
	(0.019)	(0.040)	(0.077)	(0.037)
N	804	804	804	804
Power with	0.043	0.261***	0.262***	0.255***
	(0.037)	(0.026)	(0.060)	(0.019)
N	804	804	804	804
Power over	0.255***	0.048*	0.050	0.044**
	(0.019)	(0.029)	(0.037)	(0.020)
N	804	804	804	804

Robust standard errors in parentheses. PSM estimates bootstrapped 1000 repetitions.

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

Table A3.4: Power from within

	(1) OLS / probit	(2) PSM NN	(3) PSM Caliper	(4) Propensity Score Weighting
Self-confidence - number	0.234***	0.229	0.278**	0.254***
	(0.083)	(0.234)	(0.117)	(0.078)
N	804	804	804	
Knowledge - number	0.301***	0.243	0.198	0.378***
	(0.084)	(0.249)	(0.138)	(0.106)
N	804	804	804	804
Opinion Women's Economic Role-number	0.033	0.172*	0.056	0.117
	(0.060)	(0.089)	(0.116)	(0.073)
N	804	804	804	804
Attitude to GBV - number	0.237	0.000	0.081	0.146
	(0.219)	(0.341)	(0.403)	(0.238)
N	804	804	804	804
Recognition of care – number	-0.019	-0.123	-0.027	-0.053
	(0.054)	(0.077)	(0.073)	(0.058)
N	804	804	804	804

Robust standard errors in parentheses. PSM estimates bootstrapped 1000 repetitions.

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

Table A3.5: Power to

	(1) OLS / probit	(2) PSM NN	(3) PSM Caliper	(4) Propensity Score Weighting
Personal autonomy – proportion	0.167***	0.153**	0.147**	0.109**
	(0.042)	(0.065)	(0.069)	(0.054)
N	770	770	770	770
1[Individual capability (apply knowledge)]	0.154**	0.009	0.010	0.010
	(0.077)	(0.065)	(0.124)	(0.124)
N	804	804	804	804

Robust standard errors in parentheses. PSM estimates bootstrapped 1000 repetitions.

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

Table A3.6: Power from with

	(1) OLS / probit	(2) PSM NN	(3) PSM Caliper	(4) Propensity Score Weighting
Social capital- number	0.039	0.004	0.039	-0.011
	(0.048)	(0.069)	(0.085)	(0.049)
N	804	804	804	804
Participation in community groups	1.800***	1.736***	1.739***	1.837***
	(0.087)	(0.114)	(0.319)	(0.111)
N	804	804	804	804
Degree of influencing in governing	0.553***	0.564***	0.569***	0.570***
	(0.075)	(0.104)	(0.212)	(0.088)
N	804	804	804	804
Attitude and beliefs	0.063	0.093	0.109	0.084
	(0.043)	(0.075)	(0.071)	(0.057)
N	804	804	804	804

Robust standard errors in parentheses. PSM estimates bootstrapped 1000 repetitions.

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

Table A3.7: Power over

	(1) OLS / probit	(2) PSM NN	(3) PSM Caliper	(4) Propensity Score Weighting
Involvement in HH decision making	-0.040	-0.016	-0.027	-0.024
	(0.032)	(0.053)	(0.083)	(0.041)
N	804	804	804	804
Independent income	0.590***	0.744***	0.678	0.744***
	(0.161)	(0.254)	(0.420)	(0.172)
N	804	804	804	804
Experience of violence	0.214**	0.163	0.159	0.235**
	(0.090)	(0.138)	(0.299)	(0.104)
N	804	804	804	804

Robust standard errors in parentheses. PSM estimates bootstrapped 1000 repetitions.

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

APPENDIX 4: TIME ALLOCATION AND CARE

The survey also explored time allocation measurements within households. It was decided to report these estimates in the appendix because, even if relevant and important for women's empowerment, they are not directly linked with the theory of change of the project under analysis.

Table A4.1 provides estimates on the **number of hours devoted to different activities** the women reported spending in the previous 24 hours. Estimates in the first column suggest that women involved in the project activities spend an average of two hours per day on care work. The second column in Table 1 suggests that women in the intervention group reported spending on average 12 hours per day on household care, which includes: care of children and elderly, fetching water and wood, cooking, cleaning the house, washing clothes, and tending livestock. It has to be noted that these estimates are based on self-reported estimates of the time devoted to different activities, and the sum of the time allocated to these activities might be more than 24 hours. The third column suggests that women reported spending less than one hour per day on business activities. This is in line with the idea that women have limited access to business activities, and even if involved in the dairy sector this represents a secondary activity that does not require an extensive investment of time. None of these estimates are statistically different from the estimates in the comparison group. Finally, the fourth column in Table A4.1 suggests that on average women involved in the intervention group spent 12 hours per day in personal activities, such as leisure time (e.g. socialising with neighbours), sleeping at night, and personal care and rest. This appears to be significantly higher than the comparison group.

Table A4.1: Time allocation

	Time devoted to care	Time devoted to HH care	Time devoted to business activities	Time devoted to personal activities
Intervention group mean:	2.298	11.993	0.845	12.004
Comparison group mean:	2.145	12.253	0.882	11.372
Difference:	0.153	-0.260	-0.037	0.632**
	(0.123)	(0.357)	(0.131)	(0.266)
Observations intervention:	225	220	223	224
Observations:	726	676	693	711

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

Estimating the number of hours devoted to a certain activity can be difficult. In order to triangulate the responses provided in Table A4.1, the questionnaire also investigated the **self-reported perception of change in time devoted to** activities such as care, household care, business activities, and personal activities since the project started in 2009.

Estimates in Table A4.2 suggest that on average 82 per cent of the women in the intervention group reported having increased the time devoted to household care since 2009, compared with 71 per cent of women in the comparison group. Almost 50 per cent of the women in the intervention group reported having increased time devoted to business compared with 31 per cent in the comparison group. Finally, more than 55 per cent of the intervention group reported having increased time devoted to personal care, compared with almost 30 per cent in the comparison group. All these differences are statistically different from zero, but only estimates of personal time (i.e. sleeping at night, personal care, and leisure time) seem to be consistent between Table A4.1 and Table A4.2.

Table A4.2: Increased personal time

	1 [Increased time for care]	1 [Increased time for HH care]	1 [Increased time for business]	1 [Increased personal time]
Intervention group mean:	0.711	0.827	0.489	0.556
Comparison group mean:	0.646	0.712	0.310	0.290
Difference:	0.065	0.115***	0.179***	0.265***
	(0.045)	(0.041)	(0.045)	(0.045)
Observations intervention:	225	225	225	225
Observations:	729	729	729	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

It has been suggested that time devoted to care activities is only one component of care. In their everyday work, women have to conduct multiple tasks and not focus on one main task at the time. **Intensity of care work** can be estimated by using an indicator looking at multiple activities. In order to estimate multiple activities the questionnaire asked for each activity listed if the respondent was also responsible for care of children or other adult household member at the same time the activity was being carried out.

Table A4.3 provides estimates of the number of multiple activities conducted every day, with a variable ranging from zero (i.e. no multiple activities conducted) to nine. Estimates suggest that on average women in the intervention group are conducting more intense work than women in the comparison group.

Table A4.3: Multiple activities

	Multiple activities
Intervention group mean:	5.027
Comparison group mean:	4.040
Difference:	0.986***
	(0.325)
Observations intervention:	225
Observations:	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

It has been suggested that one possible mechanism that can lead to changes in time and intensity of the care activities in women's life is a more equal **redistribution of responsibilities within the house**. One possible method to investigate gender roles and domestic responsibilities for care is to determine if the amount of time that men and boys in the household spent on care activities had increased or decreased since the beginning of the project.

Estimates in Table A4.4 suggest that 17 per cent of women in the intervention group reported that since 2009 men in the household have increased the time they devote to care activities. This is compared with only 7 per cent of women in the comparison group. This might suggest that in households in the intervention group are experiencing a gradual and modest shift towards redistribution of care activities with men in the household.

The third column in Table A4.4 suggests that time devoted to business appears to have increased more for the comparison group than the intervention group. This is a puzzling result given that the project was also aiming to increase business opportunities for men.

Table A4.4: Increased time – men

	1 [Increased time for care – men in HH]	1 [Increased time for HH care – men in HH]	1 [Increased time for business – men in HH]	1 [Increased personal time – men in HH]
Intervention group mean:	0.178	0.342	0.147	0.160
Comparison group mean:	0.076	0.306	0.214	0.124
Difference:	0.102***	0.036	-0.067*	0.036
	(0.027)	(0.041)	(0.036)	(0.035)
Observations intervention:	225	225	225	225
Observations:	729	729	729	729

*Notes: Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates are bootstrapped with 1,000 repetitions. All means are calculated after matching.*

NOTES

1 The actual occurrence of the climatic shock affecting the household was not captured.

2 However, this difference is not statistically significant when different matching techniques are used.

3 See Gujarati, Damodar N. (2003) *Basic Econometrics*: Fourth Edition. New York: McGraw Hill.

4 Per capita figure refers to adult equivalent units. Daily total consumption was divided by a factor representing household size, to generate a per-day, per-person expenditure figure. To reflect that the existence of economies of scale within households, and the lower consumption needs of children, the formula used for calculating household size is $(A+\alpha K)^{\theta}$, where A is number of adults in the household; K is the number of children; α is the consumption of a child relative to an adult; and θ stands for the extent of economies of scale. This Effectiveness Review follows the common practice of setting α equal to 0.33 and θ equal to 0.9.

5 When items are used in a scale or index, they should all measure the same underlying latent construct (e.g. household wealth status). The items, then, must be significantly correlated with one another. Cronbach's alpha is a measure of this inter-item correlation. The more the variables are correlated, the greater is the sum of the common variation they share. If all items are perfectly correlated, alpha would be 1 and 0 if they all were independent from one another. For comparing groups, an alpha of 0.7 or 0. 8 is considered satisfactory. See: Bland, M. J. & Altman, D. G. 1997. Statistics notes: Cronbach's alpha. *BMJ*, 314, 572.

6 The qualitative component identified at least one additional dimension that was not identified during the workshop, this being voting freedom, which reflects the ability to cast a vote according to one's own conscience. Qualitative data suggests that this may have been a serious oversight, as there is evidence that women's voting freedoms are severely restricted, formally and informally. Firstly it appears that voter registration is unequal – with disproportionately more men registered as voters than women. Secondly, qualitative interview data suggests that of those women who are registered to vote, many do so at the 'will and wish' of their husbands or families. While this project did not seek to influence voting behaviour, this kind of basic self-determination is an important component of personal power. In future, we suggest that evaluations of women's empowerment in rural Punjab should look at the degree to which women are able to make informed decisions about casting their votes.

7 In the previous year Oxfam GB's global indicator for women's empowerment was based on whether women are doing better in terms of overall women's empowerment than a 'typical' woman in the area. This is defined by comparing each woman's women empowerment index with the median of the comparison group. In particular, the global indicator takes the value of 1 if the base empowerment index is greater than the median of the comparison group and zero otherwise. This measure is not reported any more, but results are consistent with the current empowerment index.

8 These questions were also triangulated with the attitude shown during the interview. Enumerators were asked to report which one of the two options better described the respondent's behaviour during the interview:

1 – Looks you in the eye during most of the interview / 2 – Looks away or on the ground during most of the interview

1 – Speaks audibly and clearly; I do not have to ask her to repeat what she says./ 2 – Speaks quietly or inaudibly. I have to listen very closely or ask her to repeat herself.

1 – Speaks freely and answers questions with long answers and opinions. / 2 – Is reluctant to give her opinion or gives very short answers.

Results show, however, that there are no statistically significant differences between intervention and comparison groups in the way enumerators marked respondents. It is entirely possible that this measure is bias due to the enumerators' perceptions.

9 It should be noted that this measure is not robust to alternative estimation techniques.

10 Given the high sensitivity of the question and the cultural context, it was decided to not ask directly about the respondent's experience but rather asking someone close to her and using it as a proxy for violence.

11 Caliendo, M. and Kopeinig, S. 2008. Some Practical Guidance for the Implementation of Propensity Score Matching, *Journal of Economic Surveys*, Wiley Blackwell, vol. 22(1), pages 31–72.

12 Bootstrapping is a statistical procedure where repeated samples are drawn from the original sample with replacement. This results in a statistical distribution of parameter estimates (the sampling distribution). The bootstrapped standard error is the standard deviation of this sampling distribution and it can be shown that as the number of repeated samples becomes large, provided certain technical conditions are met this is a good estimate for the standard error of the estimate.

13 Hirano, K. & Imbens G.W. (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.

Oxfam Effectiveness Reviews

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

The information in this publication is correct at the time of going to press.

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

OXFAM

Oxfam is an international confederation of 17 organisations 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:

Oxfam America (www.oxfamamerica.org)
Oxfam Australia (www.oxfam.org.au)
Oxfam-in-Belgium (www.oxfamsol.be)
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 India (www.oxfamindia.org)
Intermón Oxfam (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 (www.oxfamnovib.nl)
Oxfam Québec (www.oxfam.qc.ca)

Please write to any of the agencies for further information, or visit www.oxfam.org.