EVALUATION: NOVEMBER 2013 PUBLICATION: NOVEMBER 2014

WOMEN'S EMPOWERMENT IN ETHIOPIA

Evaluation of women's beekeeping and access to financial services

Effectiveness Review Series 2013/14

Photo: Tom Pietrasik/Oxfam

ROB FULLER

OXFAM GB

www.oxfam.org.uk/effectiveness

OXFAM

CONTENTS

ACKNOWLEDGEMENTS

We would like to thank the staff of Zembaba Union, the beekeepers' cooperatives in the four woredas of Bahir Dar Zuria, Gozamen, Gondar Zuria and Lebokemkem, and the Oxfam team in Ethiopia for all their support during the Effectiveness Review. Particular thanks are due to Sintayehu Mengiste, Mulu Tesfaye, Dawit Derbew, Meseret Cherie, Tigist Minlekalew and Nebyu Mehary, as well as to Bisrate Marcos and Abel Desta of Momentum Professionals for their excellent leadership of the field work.

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 '*Facilitating access to financial services for women beekeepers in Ethiopia*' was one of those selected for an Effectiveness Review in the 2013/14 financial year.

The project under review was carried out by Oxfam together with the Zembaba Bees Product Development and Marketing Cooperatives Union in seven *woredas* (districts) of Amhara Region between May 2010 and April 2013. The project aimed at strengthening women's livelihoods and promoting empowerment, especially by facilitating the formation of 32 women's 'self-help groups' across the seven woredas, by supporting their members in taking up or strengthening their beekeeping activities, and by facilitating links to financial services and other service providers.

Evaluation approach

This Effectiveness Review used a quasi-experimental evaluation design to assess the impact of the project activities approximately six months after implementation ended. The review was restricted to four of the seven woredas where the project was implemented, excluding the three that are included in the subsequent, larger-scale honey value-chain development project. It should be noted that the four woredas in which the Effectiveness Review was carried out are generally seen to have lower potential for honey production than the other three – so the results of this review may understate the effects of the project across the seven woredas.

The review sought to evaluate the impact only among women who directly participated in the self-help groups (SHGs) established under this project. All 230 women listed as members of the 13 SHGs across the four woredas were targeted for interview as part of this review. In the event, only 189 of these women were interviewed. In particular, those whose homes were located large distances on foot from their communities were not included in the survey. In addition, 438 women were selected at random from nearby communities where the project had not been implemented, for comparison purposes. At the analysis stage, the statistical tools of propensity-score matching and multivariate regression were used to control for apparent baseline differences between the SHG members and comparison respondents, to increase confidence when making estimates of the project's impact. In addition, 11 focus group discussions were held with SHG members and ten with male members of their households, and in-depth interviews were carried out with local leaders and extension agents, to provide deeper insights on the impact of the project than could be captured in the quantitative survey.

Results

The survey results provide good evidence that the project under review was successful in encouraging households to take up – or to continue to engage in – beekeeping (particularly among female-headed households), and in encouraging them to experiment with new technologies. Efforts to improve links to cooperatives and marketing channels also appear to have been successful, with a third of the households of SHG members having sold some honey to their local cooperative over the 12 months prior to the survey.

Key results of this Effectiveness Review

Outcome area	Evidence of positive impact	Comments
Engagement in beekeeping	Yes	More of the households of SHG members are engaged in beekeeping. Women have increased their participation, and many have experimented with new technologies.
Sales of honey	Yes	Revenue generated from sales of honey among households of SHG members is more than double that of the comparison households – though it is still small in magnitude.
Savings	Yes	Most self-help group members have some personal savings, and half had saved during the month prior to the survey
Access to and use of credit	Yes	SHG members are more likely to have borrowed both from the group and from microfinance institutions.
Dietary diversity	Yes	Self help group members and their households have a significantly more diverse diet than comparison households
Women's empowerment	Yes	There is a clear impact on women's personal savings and on participation in community groups. There is little or no indication of positive changes in other areas of women's empowerment.

Results apply to all members of the 13 self-help groups established under this project in the woredas of Bahir Dar Zuria, Gozamen, Gondar Zuria and Lebokemkem who were available and willing to be interviewed at the time of the survey, with the exception of the approximately 10 per cent of SHG members who live at particularly large distances from their kebele centre.

The prices gained from sales to the cooperatives were significantly higher than those realised through other channels, though it cannot be determined from the survey data whether this is because of differences in the quality of honey being sold, or purely because the cooperatives are more effective at marketing honey and can realise economies of scale. As a result of the difference in prices, households of SHG members generated considerably higher revenue from sales of honey than did comparison households. The level of revenue being generated is quite modest, and it is clear that for the vast majority of SHG members, honey production is not central to their livelihoods. However, this at least provides reasons for optimism that a project that could provide more intensive training and support to producers and overcome the difficulties in the supply of inputs has potential to generate significant returns.

The SHGs appear to have been successful in providing members with a means of saving and accessing credit. The personal savings of SHG members were considerably higher than those reported by comparison respondents, and approximately half believed that they would be able to obtain a loan of 1000 birr (approximately US$ 50) from their SHG if necessary. More of the SHG members (or other members of their households) had taken out loans during the 12 months prior to the survey, though most of the additional loans seem to have been quite small. Most loans were apparently invested in livestock or agriculture, with a smaller number going towards household businesses. It is notable that very few SHG members reported using their loans to invest in beekeeping.

SHG members and their households were apparently consuming a more diverse diet than comparison women, and were less likely to have suffered extreme food insecurity during the summer season previous to the survey. However, there was no indication of an effect from the project on an index of indicators of material wealth, such as housing conditions and ownership of assets.

The impact of the project was also assessed on 19 different indicators of women's empowerment that were thought to be appropriate for the context in Ethiopia. There was clear evidence of a change in terms of two characteristics directly linked to the

project activities – women's personal savings, and their involvement in community groups. There are some indications of a difference between SHG members and comparison respondents in other characteristics, including women's involvement in household-management decisions, control over assets, having the freedom to visit relatives in other communities and to participate in community activities, and taking up leadership positions in community groups. However, in each of these cases the evidence for an effect from the project is not strong. There are also indications that attending group activities has increased the burden on women, with SHG members being more likely than comparison respondents to say that they are sleeping less than they used to in 2009. They also say that their children and (less clearly) male members of their households have less leisure time than before. There was no indication that the project had had an effect on other areas of women's empowerment, such as involvement in decisions on the household's livelihoods activities or expenditure, influence in community decision-making, self-confidence and opinions on women's rights and roles, or experience of gender-based violence. While the majority of SHG members said that they were now involved in honey production, many also commented during focus group discussions that they still lack ownership of these activities or control over the income generated.

Programme learning considerations

Ensure that the appropriate level of support is in place when introducing new technologies.

The Effectiveness Review provides evidence that the new beekeeping technologies and practices introduced under this project can enable women to take on significant roles in honey production and marketing, overcoming traditional attitudes that beekeeping is suitable only for men. However, it is clear that prerequisites to success are that learning about the technologies is fully embedded, that the required equipment and inputs are available, and that extension agents are fully involved and able to provide effective advice and support. Even more importantly, it appears that a good level of support and follow up is required so that group members feel confident that they are able to apply the new technologies effectively, and that they will not be deterred by initial setbacks.

A related consideration is the appropriate level of intensity required for a project such as this. The project under review was originally intended to have a larger budget and to allow for a more intense package of advice and support to be provided to SHG members. Once it was clear that the planned budget was not available, it would perhaps have been more effective to have reduced the scale of the project in order to provide concentrated support to a smaller number of self-help groups, rather than a less comprehensive package for a greater number.

Continue to design more holistic projects and programmes, with specific strategies to contribute to women's empowerment and minimise risks of participation.

Another implication of this Effectiveness Review is that, even if a project can successfully encourage women's engagement in a new livelihood activity, this is not by itself sufficient to achieve empowerment in a broader sense. It is important to note that the project under review initially focused on beekeeping as a livelihoods activity and on promoting business development, and did not (at least at the design stage) have a clear strategy by which the project was expected to lead to social empowerment or to manage the risks of women's participation. The weakness of this approach has already been recognised, as evidenced by the design of the current honey value-chain development project, which incorporates a specific gender strategy, including community forums, literacy training, and the promotion of labour-saving devices. There

will always be a tension between social justice and the extent to which a programme should focus on business development, but it at least seems to be clear that social empowerment will not necessarily result from economic empowerment.

Carefully monitor the requirements that project activities place on participants' time, in order to mitigate any negative consequences.

The fact that SHG members were more likely than comparison respondents to report that they have experienced a decrease in the time they spend sleeping over the past few years, serves as a reminder of the importance of being aware of the additional burden that project activities place on participants. Clearly it is important to ensure that the value being realised through any intervention outweighs the potential costs, either in terms of increased stresses on participants or their family members, or in terms of reduced engagement in other livelihoods activities. This should be monitored carefully in current and future projects through the use of regular discussions with group members, so that action can be taken to reduce those burdens if necessary. It is also possible that the greater use of labour-saving technologies may be able to support women in reducing household responsibilities and encouraging men to take on some of these duties.

Review the monitoring, evaluation and learning approaches for current and future projects, particularly in order to monitor changes in the enabling environment and how project participants are responding to them, as well as how this is translating into increased empowerment.

The results of this Effectiveness Review reinforce the need for investment in a comprehensive system for monitoring implementation and outcomes. In particular, these findings suggest that emphasis should be put on monitoring of changes in the context (such as the availability of inputs or the involvement of extension services) and on regular discussions with project participants on the effects of the project activities, while implementation is proceeding. Just as important is to monitor how the effects of projects are distributed across different groups of participants, so as to improve the effectiveness with which interventions can be targeted.

The issues raised in this report can also be used to inform the questions to be investigated in more formal evaluation processes. In particular, the inclusion of indicators of access to and use of credit, of time use, and of contact with extension services and other service providers will strengthen the midline and endline surveys to be carried out for the current honey value-chain development project, and for future similar initiatives. At the same time, the refinement and testing of indicators of social empowerment should continue, to ensure that they provide as full and accurate a picture as possible of women's experience of empowerment in Ethiopia.

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 they have promoted change in relation to relevant Oxfam GB global outcome indicators.

The Effectiveness Review that took place in Amhara Region of Ethiopia in November and December 2013 aimed at evaluating the success of the project '*Facilitating access to financial services for women beekeepers in Ethiopia*' in strengthening women's livelihoods and promoting empowerment. The project was carried out by Oxfam, together with the Zembaba Bees Product Development and Marketing Cooperatives Union in seven *woredas* (districts) of Amhara Region between May 2010 and April 2013. In particular, the project facilitated the formation of 32 women's 'self-help groups' across the seven woredas, and supported their members in taking up or strengthening their beekeeping activities, as well as in facilitating links to financial services and other service providers.

In three of the woredas where this project was implemented, Oxfam, Zembaba Union and other partners have since launched a new project, which is continuing to strengthen the self-help groups established under the earlier project, while also scaling up these activities to other communities. In two of those woredas, Oxfam carried out research on the effects of women's collective action in 2012.[1] In order not to duplicate the work carried out under the women's collective action research, and so that results of the Effectiveness Review would reflect only the earlier project and not the more recent scale-up, the Effectiveness Review was carried out only in the four woredas where the newer project is *not* being implemented. It should be noted that the four woredas in which the Effectiveness Review was carried out were generally seen to have lower potential for honey production than the other three – so the results of this review may understate the effects of the project across the seven woredas.

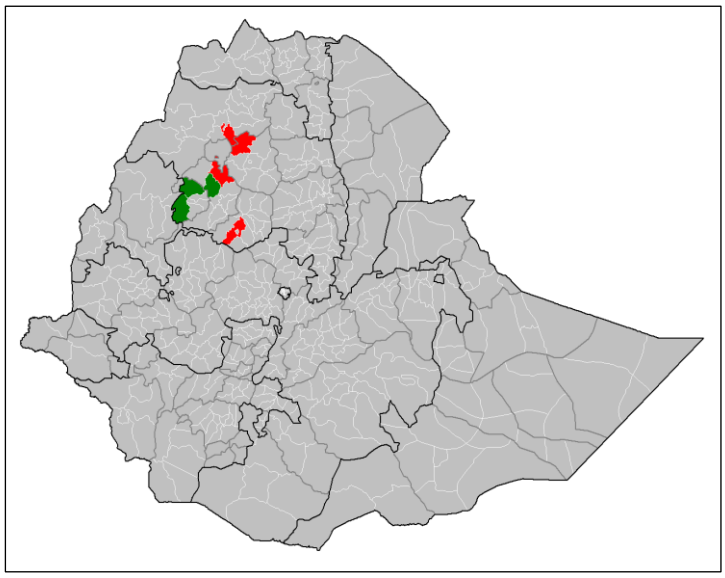


Figure 1.1: Map of Ethiopia showing project woredas
Woredas included in the Effectiveness Review are coloured red; those excluded from the review are coloured green.

This report presents the findings of the Effectiveness Review. Section 2 briefly reviews the activities and the intervention logic of the project. Sections 3 and 4 describe the evaluation design adopted and how this design was implemented. Section 5 presents the results of the data analysis, based on the comparison of outcome measures between the intervention and comparison groups. Section 6 concludes with a summary of the findings and some considerations for future learning.

2 PROJECT DESCRIPTION

2.1 PROJECT ACTIVITIES

The overall objective of the project under review was to strengthen the livelihoods of rural people – particularly women – by improving access to financial services in order to support production and marketing of agricultural products. Specifically, the project aimed to:

- support rural households to increase the productivity and quality of honey production, through training, technology transfer, and the distribution of inputs;
- build the capacity of producers' organisations, the woreda-level cooperatives and regional-level Zembaba Union, in particular by strengthening linkages to private-sector purchasers;
- improve access to finance, so as to allow producers to invest and increase their competitiveness;
- promote women's economic leadership in the honey value chain; and
- capture and communicate learning to support other projects in the future.

One of the initial stages of the project involved identifying and training 24 community facilitators who formed the women's self-help groups (SHGs) in their communities. The SHGs each consist of approximately 20 women who meet regularly to share experience and learning and to operate a revolving savings and loan fund. The groups are also intended to act as a vehicle to promote women's involvement in other community-level institutions and in cooperatives. A total of 32 groups were established between 2010 and 2012 across the seven woredas, with a total membership of just over 600 women.

The majority of members of the self-help groups received training and technical support in beekeeping. Most of the SHG members were not directly engaged in beekeeping before the project's launch, although all had some prior awareness of or exposure to beekeeping. In Ethiopia beekeeping is traditionally a male-dominated activity, and in many cases, the members' husbands or other male relatives were already engaged in the activity. The modern beehives encouraged under the project are intended to make beekeeping more appropriate for women's engagement. One hundred of the SHG members also received modern beehives and other inputs for honey production under this project, on a long-term loan.

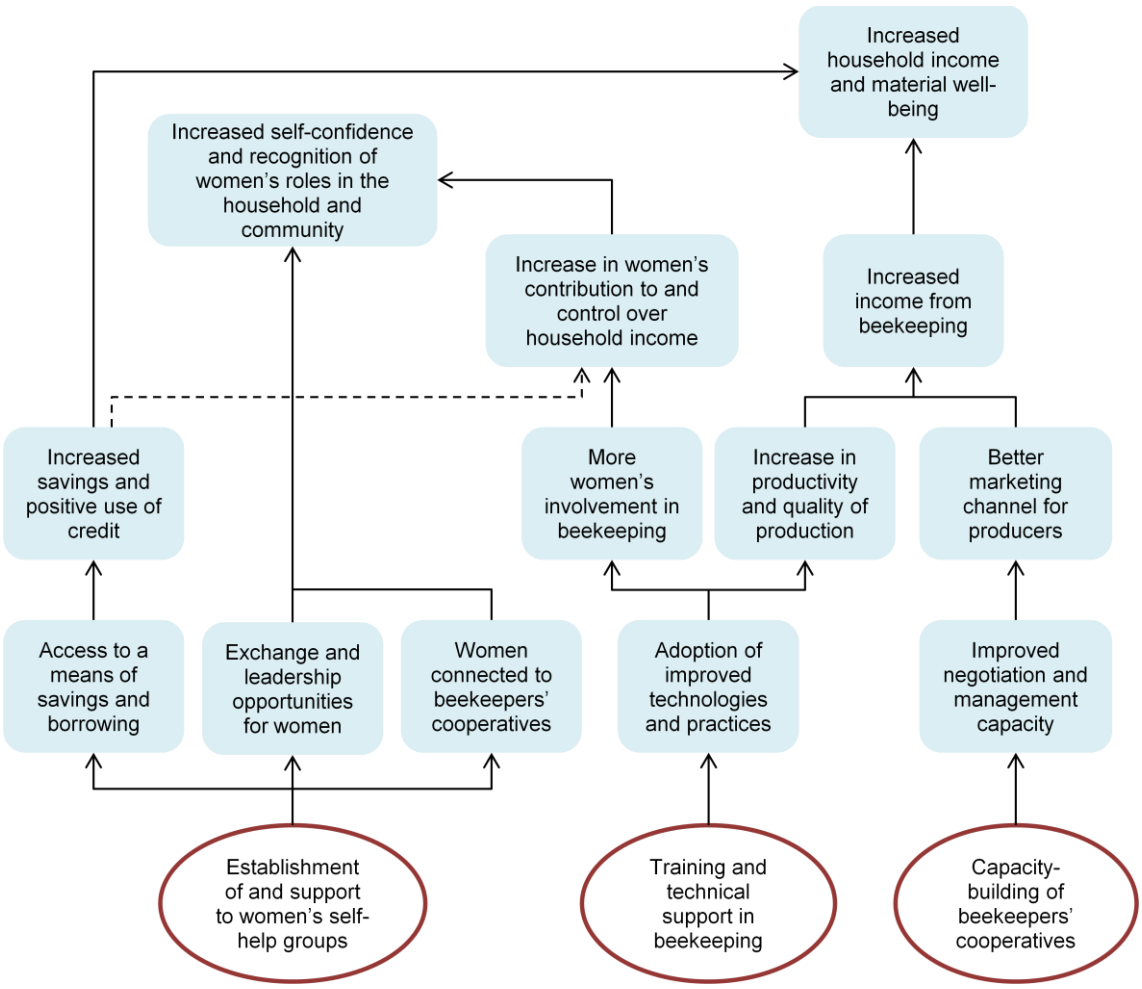
An additional, a programme of training for SHG members focused on household financial management – on the importance of making savings, and on how to use savings and credit to invest in productive activities. This training included encouragement to use improved agricultural practices and kitchen gardening.

Another strand of the project was to strengthen the woreda-level beekeepers' cooperatives (and in the case of Dangila woreda, to establish a cooperative where one did not exist previously) and Zembaba Union. In particular, this project (in partnership with Oxfam's Enterprise Development Programme) supported Zembaba Union in procuring loans of working capital from two microcredit institutions to enable it to purchase honey from its members, and in establishing partnerships to supply honey to private-sector companies.

2.2 PROJECT LOGIC AND INTENDED OUTCOMES

This Effectiveness Review sought to discover what changes could be observed at the level of the individual women and their households as a result of the activities described above. Figure 2.2 shows a simplified version of the outcomes that the project was intended to affect.

Figure 2.2: Simplified logic model for the project



As discussed above, the training, technical support and beekeeping inputs that were provided to SHG members were mainly intended to encourage the adoption of modern practices and technologies. The use of modern hives in particular was thought to provide opportunities for women to actively engage in beekeeping. These improvements should also lead to increased efficiency (such as through harvesting multiple times per year), and to improvements in the quality of the honey produced. This would allow producers to realise greater prices when selling their honey, and so increase their household income and overall well-being. At the same time, women and men were encouraged to engage with, and become members of, their woreda-level beekeepers' cooperatives, providing them with a new channel through which they could market their honey. The project also sought to build the capacity of those cooperatives and Zembaba Union, particularly in marketing and negotiation, so that they could negotiate contracts with private-sector buyers and therefore provide a dependable marketing channel for their members.

Apart from acting as a forum for training and sharing experience in beekeeping, the SHGs established under this project were also intended to function as savings and credit groups. Encouraging regular saving among members and giving them access to small loans should enable women to better manage their household finances and also to generate useful amounts to invest in beekeeping and other productive activities, such as agriculture. Another intervention, not shown explicitly in Figure 2.2, was to facilitate links between SHG members and microfinance institutions, again with the aim of enabling them to use credit to make productive investments.

All of these various activities were intended to lead to greater empowerment for women. Providing a means for women to engage in beekeeping and therefore to contribute to household income was intended to increase their control over household resources, and to lead to greater recognition of and respect for women's engagement in economic activities. Encouraging saving among women and providing them with greater access to credit should also result in their having greater control over resources within the household. The SHGs were also intended to act as a forum for sharing experience and taking collective action, with the aim of building self-confidence and having women's roles more recognised at the community level. All these changes should, over time, lead to a rebalancing of power relationships within the household and community, with women being able to make and influence decisions and to have the confidence, freedom and support necessary to make positive changes in their lives.

The project's impact on each of the outcome areas discussed here – at least insofar as success can be observed at the individual or household level – is examined in Section 5 of this report.

3 EVALUATION DESIGN

The central problem in evaluating the impact of any project is how to compare the outcomes that result from that project with *what would have been the case* without that project having been carried out. In the case of this Effectiveness Review, information about the situation of those who participate in the self-help groups was collected through a household questionnaire – but clearly it was not possible to know 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 those that were not. As long as the two groups can be assumed to be similar in all respects except for the implementation of the specific project, observing the situation of those where the project was not implemented can provide a good estimate of the counterfactual.

An ideal approach to an evaluation such as this is to select the sites in which the programme will be implemented at random. Random selection minimises the probability of there being systematic differences between the project participants and non-participants, and so maximises the confidence that any differences in outcomes are due to the effects of the project.[2]

In the case of project examined in this Effectiveness Review, the implementation sites were not selected at random. Instead, women with the capacity to act as community facilitators were identified by the cooperatives in each woreda, and those facilitators then encouraged other community members to join a self-help group. However,
It was clear from discussions with the implementation staff that there were many more communities in the implementation areas where groups could have been established – that is, where there were numbers of women who would be eligible to participate, had the community facilitators mobilised in those communities instead. During visits to project sites, it appeared that people in communities located even a short distance from the implementation communities did not have any awareness of the project activities. This allowed a 'quasi-experimental' evaluation design to be adopted, in which the situation of women in nearby non-project communities was assumed to provide a reasonable counterfactual for the situation of women who participated in the project. Since the project involved support to the woreda-level cooperatives, it was also assumed that any individual members could have benefited from the project activities. Non-project communities with significant numbers of individual cooperative members were therefore not considered suitable for comparison purposes, and were excluded from the survey.

It is important to note that, within the implementation communities, the self-help group members were not selected at random. Instead women came to participate in the self-help groups, firstly through social connections to the community facilitators, and secondly, through their having made an active decision to participate. However, in the comparison communities it could not be known who *would have* joined a self-help group had they had the opportunity. For that reason, the women interviewed in the comparison communities were selected at random from among the households in those communities. At the data analysis stage, project participants were then 'matched' with women with similar characteristics in the comparison communities. Matching was performed on the basis of a variety of characteristics – including household size, ethnicity, education level, productive activities, and indicators of material well-being, such as housing conditions, ownership of assets, and participation in other community

groups and activities. Since some of these characteristics may have been affected by the project itself (particularly those relating to productive activities and wealth indicators), matching should normally be 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 lead to significant bias in the estimates as long as measurement errors due to the recall data are not significantly different for respondents in the intervention and comparison groups.

The survey data provided a large number of baseline household characteristics on which matching could be carried out. One practical problem is that it would be very difficult to find households in the comparison communities that correspond exactly in all these characteristics to households in the project communities. Instead, these characteristics were used to calculate a 'propensity score', the conditional probability of the household being in an intervention community, given particular background variables or observable characteristics. Households in the project and comparison communities were then matched based on their having propensity scores within certain ranges. Tests were carried out after matching to assess whether the distributions of each baseline characteristic were similar between the two groups. Technical details on this approach are described in Appendix 3.

As a check on the results derived from the propensity-score matching process, results were also estimated using multivariate regression models. Like propensity-score matching, multivariate regression also controls for measured differences between intervention and comparison groups, but it does 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.

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. This is a cause for particular caution when interpreting the results of an evaluation for a project in which participants were to some extent self-selected. This point is further discussed in the context of the results in Section 5.

4 DATA

4.1 SAMPLING APPROACH

As discussed in Section 1, it was decided to restrict the Effectiveness Review to the four woredas in which the project was implemented that have not been included in the women's collective action research or the more recent project. In order to have an adequate number of observations for quantitative analysis, it was decided to target for interview all the 230 women listed as members of the 13 self-help groups across those four woredas. In the event, only 189 of these women were interviewed. Twenty-three members were excluded from the survey because they were reported to live very far from the centre of their *kebele* (sub-district), and reaching the houses was not feasible within the time and budget constraints available. An important consequence of this is that the results of the Effectiveness Review do not apply to the 10 per cent of SHG members who live far from the kebele centres. Five of the SHG members were found to be deceased by the time of the survey, and another 13 were not available or did not wish to participate in the survey.

Focus group discussions were held with SHG members in each of the 11 kebeles, immediately after completion of the individual survey. In 10 of the kebeles, focus group discussions were also held with the spouses of SHG members (or other male household members). Participants in these discussions were asked about changes in women's situations in the community, as well as their experience in beekeeping and for feedback on the project. In-depth interviews were also carried out in several communities with local leaders and extension agents.

The selection of communities suitable for comparison was made in two stages. Firstly, based on their knowledge of the area, staff from Zembaba Union and the relevant woreda cooperatives suggested communities that were potentially suitable. The manager of the survey team then visited each of the suggested comparison communities to ascertain that there was little awareness of the idea of self-help groups, and that there had been little or no recent transfer of knowledge on beekeeping with the project communities. Respondent households were selected at random in the comparison communities by distributing enumerators at different points in the community, having them spin a pen to choose a random direction, and selecting the nearest household physically lying in that direction. After interviewing a respondent from that household (or ascertaining that there was no potential female respondent living in the house and willing to participate in the survey), enumerators proceeded to the next house in the original direction, and so on until they reached the edge of the community. In cases where more than one woman was living in a household and was present and willing to take part in the survey, enumerators interviewed the most senior (usually the oldest) woman household member.

Table 4.1: Numbers of respondents interviewed

Woreda (district)	Kebele (sub-district)	Number of SHG members in kebele	Number of SHG members interviewed	Number of comparison respondents interviewed
Bahr Dar Zuria	Kenbaba	20	14	39[b]
	Feres Woga	19	16	38
	Zege	20	16	40
	Meshenti	19	18	38[b]
Gozamen	Enerata	22	18	46
	Addis Ena Gulit	13	10	26
	Lekelekita Mariam	15	15	30[b]
Gondar Zuria	Das Dinzaz	20	17	40
	Degola Chenchaye	39[a]	35	80
Libokemkem	Genaza Selkisa	25[a]	15	30
	Yifag Ena Bura	18	15	32[b]
		230	**189**	**439**

[a] Two self-help groups were established in each of these two kebeles.
[b] In these cases, comparison communities could not be identified within the kebele, so comparison respondents were interviewed from communities in neighbouring kebeles.

4.2 ANALYSIS

The households of SHG members and those of comparison respondents were compared in terms of their demographic characteristics, livelihoods activities and economic situation 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 Appendix 2. Some important differences were found between the SHG members and comparison respondents. For example, the SHG members generally had larger households (with 5.6 household members on average, against 4.6 members among the comparison respondents), and members of their households were more likely to have some education. Most notably, 81 per cent of SHG members said that they or other household members were engaged in beekeeping in 2009, against only 13 per cent of the comparison respondents. The households of SHG members also tended to be wealthier, more likely to own livestock, and more likely to have an electricity connection in 2009 than were the comparison respondents.

These differences, which existed before the project, have the potential to bias any comparison of the project's outcomes between the SHG members and comparison respondents. 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 to control for the baseline differences was propensity-score matching (PSM). The full details of the matching procedure applied are described in Appendix 3. After matching, SHG members and comparison respondents were reasonably well-balanced in terms of the recalled baseline data. However, even after matching, the SHG members tend to be slightly wealthier than the comparison respondents: they are more likely to be in the top two quintiles of the wealth index[3] (i.e. in the wealthiest 40 per cent of the sample), and less likely to have been receiving

income from rent in 2009. Another important caveat is that 17 of the 189 members interviewed could not be matched and had to be dropped from the analysis. The consequence of this is that the estimates of the project's impact presented in Section 5 are not based on the whole population of SHG members interviewed, but exclude a non-random minority.

All the results described in Section 5 of the report were tested for robustness by estimating them with several alternative statistical models, including alternative PSM models and linear or probit regression models. Some of these alternative models were constructed using the same subset of households as were used to construct the primary PSM model, while others were constructed using the full set of SHG members interviewed. Where the alternative statistical models produce markedly different results from those shown in the tables in this section, this is discussed in Section 5, in the text or in the footnotes.

It is important to recall, as highlighted in Section 3, that PSM and regression models can control only for the baseline differences between the SHG members and comparison respondents for which data was collected in the survey. If there are any 'unobserved' pre-existing differences between the two groups – such as individuals' attitudes, motivation, skills or confidence – then these may bias the estimates of outcomes described in Section 5. Given that the project participants are a self-selected group, this possibility cannot be excluded and must be borne in mind when interpreting the results.

5 RESULTS

Statistics primer

This report is intended to be free from excessive technical jargon, with more detailed technical information being restricted to the footnotes and appendices. However, there are some statistical concepts that cannot be avoided in discussing the results.

Effect size

The size or magnitude of an effect when evaluating outcomes refers to the size of the difference between groups. In this report, results will usually be stated as the average difference between the SHG members (that is, the 'intervention group') and the matched women interviewed in the communities where the project was not implemented (the 'comparison group').

Statistical significance

When we refer to 'impact' in this report, we mean differences between the SHG members and comparison respondents that are 'statistically significant'. For example, imagine that we find that the average SHG member is producing honey from one more beehive than the average woman sampled in the comparison communities. This seems to be a large difference between the two groups, given that the average comparison respondent owns only two beehives. However, it is important to remember that this estimated average impact is derived from data on a *sample* of comparison respondents, rather than data on the whole population. It is possible that, by chance, we happen to have interviewed comparison respondents who own relatively small numbers of beehives, but that the ownership in the overall pool of women in the comparison communities is similar to that found among the project participants.

For this reason, it is necessary to take into consideration the statistical probability of finding a difference of one additional beehive if there were in reality no difference in beehive ownership between the SHG members and comparison respondents. This probability is usually referred to as the *p*-value. *p*-values help to evaluate study hypotheses. The default hypothesis is always that there are no differences between the intervention and comparison groups. When a difference is detected, the *p*-value is used to evaluate whether the default hypothesis (that there is no difference between the intervention and comparison groups) should be rejected – that is, to conclude that the project had an impact. If the *p*-value is small, for instance one per cent, this means that the probability that our sample would show SHG members owning an average of one additional beehive compared to comparison respondents when the true difference were zero is only one per cent. This is a small probability, and so we would have confidence in rejecting the default hypothesis that the project had no impact on this outcome. We would then say that the result is 'statistically significant'. Note that the smaller the sample size and the greater the variation in the outcome measures among the sampled households, the larger the *p*-value will be, and hence the less likely we are to be able to conclude that a result is statistically significant.

In the tables of results on the following pages, statistical significance will be indicated with asterisks, with one asterisk (*) indicating a *p*-value of less than one per cent, two asterisks (**) indicating a *p*-value of less than five per cent and three asterisks (***) indicating a *p*-value of less than 10 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 usually not considered to be statistically significant.

5.1 INTRODUCTION

This section presents a comparison of the self-help group (SHG) members and comparison respondents in terms of various outcome measures relating to the project under review. In the tables of results, asterisks are used to indicate where the differences are statistically significant at least at the 10 per cent significance level.

The results are shown after correcting for apparent baseline differences between the SHG members or their households (the 'intervention group') and the comparison respondents or their households (the 'comparison group') using a propensity-score matching (PSM) procedure. The details of this procedure are discussed in Appendix 3. All outcomes have also been tested for robustness to alternative statistical models. 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 the footnotes.

It is important to stress that the results presented in this section are average results across the SHG members who were interviewed in the four woredas. Clearly it would be of interest to investigate the effects of the project at a more local level and for specific subgroups, such as female-headed households – but the small sample sizes available limit the potential for detecting any differences between these various subgroups.[4] Nevertheless, all outcome variables were examined for differential impacts by woreda and for female-headed and male-headed households: where differences in outcomes have been found, this has been noted in the text.[5]

Three further points that were discussed in Section 4.2 should be recalled when interpreting the results presented in this section. Firstly, 17 of the 189 SHG members surveyed were excluded from the analysis during the matching process. This means that the results shown in the tables in this section are not fully representative even of those who were interviewed. However, some of the alternative statistical models tested (and discussed in the text or in footnotes where appropriate) do take account of data from all 189 SHG members surveyed. Secondly, even in the matched sample, the SHG members appear to have been slightly wealthier before the project began, according to the information recalled from 2009. Finally, the statistical estimation procedures used to derive estimates of outcomes are based only on 'observable' baseline characteristics. If there are any 'unobserved' pre-existing differences between the project participants and comparison respondents – such as individuals' attitudes, motivation, skills or confidence – then these may bias the estimates of outcomes described in this section. Given that the project participants are a self-selected group, this possibility cannot be excluded and must be borne in mind when interpreting the results.

5.2 INVOLVEMENT IN PROJECT ACTIVITIES AND SUPPORT RECEIVED

The first step in understanding what impact this project has had is to examine the extent to which respondents participated in the activities implemented under the project.

Table 5.1 shows the differences between the proportions of SHG members and comparison respondents who reported participating in key project activities. It should be noted that these figures make a comparison between outcomes after correcting (as far as possible) for the baseline and demographic differences between the SHG

members and comparison respondents, using the propensity-score matching process described in Appendix 3. Asterisks are used to indicate where the differences are statistically significant at least at the 10 per cent significance level.

The first column of Table 5.1 confirms that most of those (approximately two-thirds) who were registered as members of the SHGs and interviewed for the Effectiveness Review said that they were regularly attending SHG meetings.[6] Approximately 20 per cent of members said that they had attended one meeting in the past month, approximately the same proportion had attended two or three meetings, and another 22 per cent had attended a meeting every week during the past month. Fifteen per cent of members reported that they had held a leadership position in the group at some time.

It will also be noted from column 1 that 20 per cent of the women interviewed in the comparison communities said that they were participating in a self-help group. Since it was confirmed by the survey team on arrival at each comparison community that no such SHGs existed, it is probable that these respondents were in fact referring to their participation in some other type of community group or potentially an initiative from another organisation.

Survey respondents were also asked about whether they and other household members belong to the primary (woreda-level) beekeepers' cooperative. As shown in columns 2 and 3 of Table 5.1, most of the SHG members said that they are members of the cooperative – in fact, most said that they are the sole members among their household. More surprisingly, 40 per cent of the comparison respondents also said that they were members of the cooperative. It is possible that there is confusion here with the former practice in Ethiopia, whereby everybody was considered to be a member of the general woreda cooperative by default. However, it can be seen in column 4 that few women in the comparison communities reported having attended a meeting of the cooperative during the past month, whereas most SHG members did so.[7] A substantial number of SHG members (13 per cent) reported having held a leadership position in the cooperative in the recent past.

Table 5.1: Involvement in project activities

	1	2	3	4
	Respondent regularly attends self-help group meetings %	Some household member is a member of the primary beekeepers' cooperative %	Respondent herself is a member of the primary beekeepers' cooperative %	Respondent regularly attends meetings of the primary beekeepers' cooperative %
Intervention group mean:	64.0	87.2	83.7	83.7
Comparison group mean:	19.9	46.0	39.9	15.1
Difference:	44.1*** (6.8)	41.2*** (7.0)	43.8*** (7.0)	68.6*** (6.5)
Observations (intervention group):	172	172	172	172
Observations (total):	581	581	581	581

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

Respondents were also asked whether they or other household members had received various forms of training and support since 2009.[8] As can be seen in the first column of Table 5.2, most SHG members confirmed that they had received training in beekeeping, compared to only a small minority of comparison respondents. It was clear from the responses that this support is seen to be ongoing: 40 per cent said that they had received some training on beekeeping in the current year, and another third said that they received training last year. Columns 2 and 3 of the table show the figures relating to the second dimension of training provided under the project, which involved training in financial management (savings, credit and investment) and use of improved agricultural practices, such as kitchen gardening. Interestingly, few of the SHG members recognised that they had had training in business or finance skills at any time since 2009. It is possible that the way the question was presented in the survey was unfamiliar to them. The majority of both SHG members and comparison respondents reported that they had received training in agricultural production, but the proportion was significantly higher among the SHG members.

The promotion of women's leadership and empowerment was integrated into all the training provided under this project. It can be seen from column 4 of Table 5.2 that, while only a third of the SHG members than the comparison respondents recognised that they had received training on women's leadership, this proportion was significantly higher than among the comparison respondents.

Although figures are not shown in the table, respondents were asked about donations of productive equipment they had received during the 12 months prior to the survey. Nearly one in five (19 per cent) of SHG members reported having received modern beehives, three per cent said that they had received transitional hives, and eight per cent other inputs for beekeeping. In contrast, very few comparison households (less than two per cent) reported receiving any beekeeping inputs at all.[9]

Table 5.2: Types of training received by respondents' households since 2009

	1	2	3	4
	Training on beekeeping %	Training on business or finance skills %	Training on crop production %	Women's leadership training %
Intervention group mean:	88.9	7.0	68.4	30.4
Comparison group mean:	20.0	4.2	54.6	17.4
Difference:	68.9*** (6.4)	2.85 (2.8)	13.9* (7.5)	13.1** (6.1)
Observations (intervention group):	171	171	171	171
Observations (total):	579	579	579	579

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

One aim of the project was to improve links between participants and service providers, including the beekeepers' cooperatives, but also government extension services and microfinance institutions. Table 5.3 shows the proportions of households that had contact with various officials during the 12 months prior to the survey. As would be expected from the higher participation of the SHG members in the beekeepers' cooperatives, they were much more likely to say that they had had some contact with cooperative representatives during that 12-month period. Also as expected, in most cases it was said to be the respondent herself who had had that contact, whereas in the comparison households the contacts were much more frequently with male household members. In column 2 of Table 5.3 it can be seen that SHG members (and other members of their households) were also considerably more likely to have had

contact with representatives of other woreda-level producers' cooperatives, not only with those of the beekeepers' cooperative.

Column 3 of Table 5.3 shows that the overall proportion of households who came into contact with an agent from the government's livestock extension service was no higher among SHG members than among comparison respondents. There is some indication of a difference between the four woredas in this regard: SHG members in Bahir Dar Zuria woreda may have been slightly more likely to have had contact with the livestock extension service than comparison respondents, but those in other woredas were not.[10] This is surprising, since the livestock extension service is responsible for support to beekeeping. Qualitative interviews with extension agents in the project sites conducted concurrently with the survey found that most had little or no knowledge of this project, and one that had been aware was critical that implementers had not tried to involve him. However, project participants highlighted during focus group discussions that extension agents had changed often, that there are not sufficient numbers of them, and that some of them are absent from the sites where they work for long periods, which limits the potential for them to provide meaningful support.

On the other hand, the majority of survey respondents interviewed had had some contact with an agricultural extension agent during the 12 months, and this proportion was significantly higher among the households of SHG members. However, subgroup analysis suggests that this positive effect is confined to the SHG members in Bahir Dar Zuria woreda, and that those in other woredas were no more likely to have had contact with the agricultural extension service than were those in the comparison group.[11] Finally, in column 5 it can be seen that it is not clear whether SHG members were more likely to have contact with microfinance institutions than were comparison respondents: there is a positive difference, but this is not statistically significant.[12]

Table 5.3: Contact with officials by respondents' households in the 12 months prior to the survey

	1	2	3	4	5
	Contact with representative of beekeepers' primary cooperative %	Contact with representative of other primary cooperative %	Contact with a livestock extension agent %	Contact with an agricultural extension agent %	Contact with representative of a microfinance institution %
Intervention	62.4	31.4	25.9	77.3	19.3
Comparison	15.5	6.0	27.6	60.1	11.8
Difference:	46.9*** (6.8)	25.5*** (7.4)	-1.7 (7.0)	17.2** (7.4)	7.4 (7.6)
Observations (intervention group):	170	172	170	172	171
Observations (total):	576	577	576	580	576

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

5.3 BEEKEEPING AND HONEY PRODUCTION

We now turn to examining the evidence for changes brought by this project in the livelihoods of participants and their households. The first outcome area to be considered is the project's impact on households' engagement in beekeeping.

Table 5.4 shows the difference between households of SHG members and those of comparison respondents in terms of indicators of their engagement in beekeeping in the 12 months prior to the survey. It can be seem from column 1 that a large majority of the SHG members reported that their households were engaged in beekeeping during the 12 months prior to the survey, and that this figure was significantly higher among the SHG members than among comparison households.[13] This suggests that the project has succeeded in encouraging participants to take up and/or remain engaged in beekeeping. The difference is particularly large among female-headed households: most (80 per cent) of the female-headed households participating in the SHGs were engaged in beekeeping, against only 34 per cent of female-headed households in the comparison group.[14]

In almost all cases where the household was said to be engaged in beekeeping, the respondent reported that she herself was involved in the work. As can be seen in column 2 of the table, SHG members were more likely than comparison respondents to say that the time they personally spend on beekeeping has increased since 2009. Despite this increased participation in beekeeping, it should be noted that during focus group discussions conducted at the same time as the survey SHG members highlighted that they still generally lack ownership and decision-making control from income gained from sales of honey.

Table 5.4: Engagement in beekeeping

	1	2	3	4	5	6
	Household engaged in beekeeping in past 12 months %	Respondent increased her time spent on beekeeping since 2009[a] %	Number of beehives owned by household at time of survey	Change in number of beehives owned by household since 2009	Household owned any modern beehives at time of survey %	Household owned any transitional beehives at time of survey %
Intervention	87.2	45.9	4.4	-0.8	34.5	8.8
Comparison	70.2	26.0	2.1	-0.7	7.1	0.9
Difference:	17.1*** (4.7)	19.7** (7.9)	2.3*** (0.6)	-0.1 (0.7)	27.4*** (4.6)	7.9*** (2.4)
Observation	172	135	172	172	171	171
Observation	581	187	578	578	577	577

[a] Among those households that were engaged in beekeeping at the time of the survey.
Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

Column 3 of Table 5.4 compares the households of SHG members and comparison respondents in terms of the number of beehives owned at the time of the survey. There is clearly a positive difference here – though it should also be recalled from Section 4.2 that SHG members reported that they also owned a large number of beehives before the project started in 2009. When examining the *change* in the number of beehives since 2009 (in column 4), there is no sign of an impact from the project.[15]

Where an effect of the project is clearly visible is in the use of newer technologies for beekeeping. As shown in columns 5 and 6 of the table, much larger proportions of households of SHG members had modern or transitional hives at the time of the survey than did households of comparison respondents.

Table 5.5 shows the comparison between SHG members and comparison women in terms of honey production. As can be seen in column 1, just over three quarters of the SHG members reported that their household produced some honey during the 12 months prior to the survey, against only half of the matched comparison respondents. On average, they also reported harvesting honey more often – but even so, it should be noted that two thirds of the SHG members whose households were producing honey said that they only harvested once during the 12-month period.

Column 3 of Table 5.5 shows that the average honey production among the households of the SHG members was only a little higher on average (and not statistically significantly so) than among the comparison households. Among those households that were actually engaged in honey production, the average quantity produced over the year was approximately 22 kg, and this figure was very similar among the SHG members and comparison households.

It may be surprising that the adoption of modern and transitional hives apparently had not led to an increase in production. Some context for this finding is provided by the focus group discussions conducted with SHG members and with male members of their households at the same time as the questionnaire. In these focus groups, many participants described difficulties in applying the new technologies effectively, and some reported having lost bee colonies as a result. Some participants felt that the level of technical support provided (particularly to those who joined the SHGs later in the project's lifetime[16]) was insufficient for people to be able to use the new technologies effectively.

Table 5.5: Honey production

	1	2	3	4
	Household produced any honey during the past 12 months %	Number of harvests of honey during the past 12 months	Total quantity of honey produced during the past 12 months (kg)	Total quantity of honey consumed in household during the past 12 months (kg)
Intervention group mean:	77.3	1.1	18.5	2.8
Comparison group mean:	52.5	0.8	15.7	2.3
Difference:	24.8*** (6.8)	0.3** (0.1)	2.8 (6.9)	0.5 (0.7)
Observations (intervention group):	172	172	172	172
Observations (total):	581	581	581	581

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

Survey respondents were also asked to estimate how much of the honey they harvested was kept for consumption within their household. As shown in column 4 of the table, the quantity consumed did not differ significantly between households of SHG members or comparison respondents. Households that had produced some honey during the year kept 4 kg on average for their own consumption.

Table 5.6 presents results from the survey data on sales of honey. Most of the households (approximately 80 per cent) that produced some honey during the 12-month period went on to sell some of their production. It is already known that the

households of SHG members were more likely to be producing honey than comparison households, so – not surprisingly – they were also more likely to have sold some honey. This difference can be clearly seen in the first column of the table.

Column 2 of Table 5.6 shows that the *quantity* of honey sold was, if anything, only slightly higher among the households of SHG members than among comparison households.[17] However, where there is a clear difference between the two groups is in the channel through which they made their sales: just over a third of the SHG members said that they sold some honey to the primary cooperative, whereas none of the comparison households did so. The price received from sales to the cooperative seems to have been higher than that realised from sales through other channels (an average of 48 Ethiopian birr per kg across all the sales to the cooperative recorded in the survey, against 36 birr per kg for sales in local markets, and 41 birr per kg for sales in woreda markets). This is reflected in column 3 of the table, where the average price realised for all sales of honey was higher among the households of SHG members than comparison households.[18] These data do not reveal whether the price difference is due to producers having greater negotiating power when selling to the cooperative, or simply whether honey sold to the cooperative was of higher quality than that sold through other channels. Of course, it is possible that a combination of both factors is at work.

The effect of the slightly higher quantity of honey sold and higher prices realised by project participants is shown in columns 4 and 5 of Table 5.6, which show the total revenue generated from sales of honey during the 12 months prior to the survey. In column 4, the comparison is restricted to households that had sold some honey. Despite the apparently large estimated difference in revenue generated between households of SHG members and those of comparison respondents, it is not clear that this represents a statistically significant difference.[19] However, the full impact of the project can only be understood by comparing the revenue generated from honey production for all the SHG members and the corresponding comparison households. This comparison is shown in column 5 of the table, where we find that households of SHG members on average were generating more than twice the revenue from sales of honey than were the comparison households. This difference is strongly statistically significant, and remains so even when comparing only those households that were actively engaged in beekeeping.

No differences could be detected between the four woredas in the project's effect on the quantity of honey produced or sold, nor in the effect on price or revenue generated. Nor could any differences be observed between female-headed and male-headed households for these outcomes. However, as discussed in Section 5.1, the sample sizes for this subgroup analysis are small, so this failure to find any differences should not be taken to imply that the project's effects were uniformly distributed around the woredas and between household types.

Table 5.6: Sales of honey during the 12 months prior to the survey

	1	2	3	4	5
	Household sold any honey during past 12 months %	Total quantity of honey sold (kg)	Average price for which honey was sold[a] (birr per kg)	Revenue generated from sales of honey[a] (birr)	Revenue generated from sales of honey – average across all households (birr)
Intervention group	62.8	15.5	43.3	1075	675
Comparison group	41.8	13.5	36.7	706	295
Difference:	20.9*** (7.3)	1.9 (6.9)	5.3* (3.3)	399** (169)	380*** (121)
Observations (intervention group):	172	171	107	108	172
Observations (total):	581	580	143	144	581

[a] Among those households that sold some honey during the past 12 months.
Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

One point to note is that, even though the project apparently had a positive impact on revenue generated from sales of honey, the size of this effect is quite modest: an increase of 380 Ethiopian birr (approximately than US$ 20) on average over a 12-month period. Importantly, these figures represent only gross revenue – it is possible that the net effect on household income could be smaller if the project participants have had higher costs of production (such as through investment in new beehives) or have diverted time from other livelihoods activities to engage in beekeeping.

In summary, then, the survey data provides good evidence that the project has been successful in encouraging SHG members to take up or remain engaged in honey production, and in promoting the use of modern and transitional hives at least among a proportion of project participants. The quantity of honey produced on average was no different between the households of SHG members and households of comparison respondents. However, the SHG members seem to have been able to realise higher prices when making sales of honey (partly because they were able to sell to the cooperative), and therefore generated greater revenue overall.

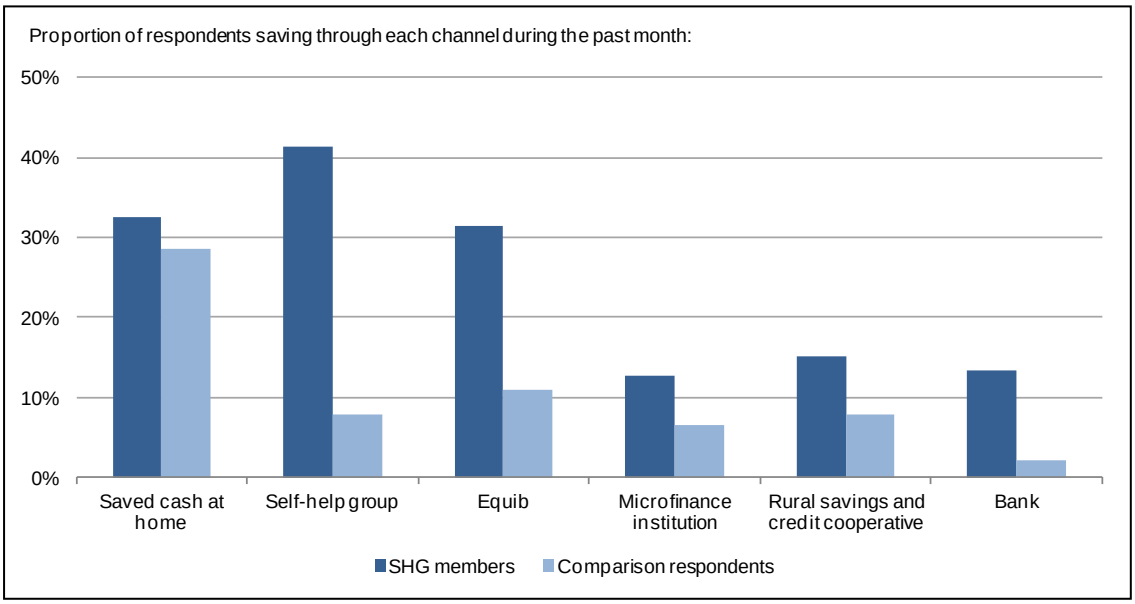
5.3 SAVINGS AND ACCESS TO CREDIT

As discussed in Section 2, one way in which the SHGs established under this project were intended to support their members was through acting as a revolving savings and credit group. This section examines the evidence that the project made a significant impact on the savings and borrowing of SHG members.

Firstly, respondents were asked about whether they had made any savings during the one month prior to the survey in any one of several locations. It can be seen in column 1 of Table 5.7 that more of the SHG members reported having made savings during the past month than did comparison respondents – though the difference is not statistically significant. Where there is a very clear difference between the groups is in the number of channels through which they reported having made savings: an average of 1.5 channels for the SHG members, and only 0.6 channels for the comparison respondents. Clearly a large proportion of the SHG members had saved with the SHG itself. But as seen in Figure 5.1, there were also differences between the SHG members and comparison respondents in the proportions of who had saved with an

equib (a traditional savings group), who with a rural savings and credit cooperative, and who perhaps through other channels as well. This implies that members were not simply saving in the SHGs as an alternative to existing savings channels. Instead, saving through those existing channels appears to have become more common as well.

Figure 5.1: Savings channels used by respondents



As a follow-up question, respondents were asked to provide an indication of the total value of their savings. Rather than being asked for the monetary value directly, they were asked to estimate for how many days they could support their household in an emergency from the money they had saved. Column 4 of Table 5.7 shows the proportion of respondents who reported that they had any savings at all, and column 5 shows the average number of days that they estimated that their households could live from those savings. In both cases there is a very clear difference between the SHG members and comparison respondents, representing what appears to be a significant effect from the SHG's activities on women's savings.[20]

As with engagement in beekeeping, the project's effects on savings are particularly clear among female-headed households: 90 per cent of SHG members living in female-headed households said that they had some savings at the time of the survey, against only 17 per cent of comparison respondents in female-headed households.

Table 5.7: Respondent's personal savings

	1	2	3	4	5
	Respondent saved any money during the past month %	Number of channels through which respondent saved money during the past month	Respondent saved money with an SHG during the past month %	Respondent currently has any savings %	Number of days household could live from respondent's current savings
Intervention group	50.0	1.5	41.3	88.4	60.2
Comparison group	44.1	0.6	7.9	54.1	25.6
Difference:	5.9 (7.3)	0.8*** (0.1)	33.3*** (5.6)	34.3*** (7.0)	34.6*** (8.5)
Observations (intervention group):	172	172	172	172	172
Observations (total):	581	581	581	581	581

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

Another key intended result of the project was to improve women's (and their households') access to credit. The project's success in this respect was assessed by asking women whether they would be able to borrow approximately 1000 Ethiopian birr (approximately US$ 50) from any of various sources, if they needed it to invest in a household business. As can be seen in column 1 of Table 5.8, most women said that they would have at least some source for such a sum – though the proportion was higher among the SHG members, of whom more than 98 per cent responded positively. There is an even clearer difference between the SHG members and the comparison respondents in terms of the *number* of potential sources of credit: the SHG members mentioned an average of 2.5 different sources from which they could borrow the 1000 birr, compared to only 1.9 sources among the comparison respondents. As can be seen in column 3 of the table, a major contributing factor to this difference is the self-help group itself, from which approximately half of the SHG members said they would be able to borrow. (SHG members in Gonder Zuria woreda were particularly likely to cite the SHG itself as a potential source of credit.[21]) On the other hand, as shown in column 4, there was no significant difference between SHG members and comparison respondents in the numbers who said they could borrow from a microfinance institution or a bank. This is consistent with the finding mentioned earlier (Section 5.2) that there was no difference between the two groups in the numbers having had contact with representatives of microfinance institutions.

Table 5.8: Respondents' access to credit

	1	2	3	4
	Respondent is able to borrow 1000 birr from any source %	Number of sources from which respondent would be able to borrow 1000 birr	Respondent is able to borrow 1000 birr from an SHG or other community group %	Respondent is able to borrow 1000 birr from a microfinance institution or bank %
Intervention group mean:	98.3	2.5	49.4	64.5
Comparison group mean:	94.2	1.9	9.3	70.9
Difference:	4.1* (2.1)	0.7*** (0.2)	40.1*** (6.1)	-6.4 (6.2)
Observations (intervention group):	172	172	172	172
Observations (total):	581	581	581	581

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

Respondents were further asked for some information about loans they and other household members had received during the 12 months prior to the survey. The results from these questions are shown in Table 5.9. It can be seen in the first column that SHG members were much more likely to say that their households had taken out at least one loan during the year, and in column 2 that the number of loans they reported borrowing was higher. Column 3 shows that nearly one in five SHG members reported having borrowed from the group (or from another community group) during the year.[22] There are also indications of a significant difference between the SHG members and comparison respondents in the proportions of households that had borrowed from a microfinance institution or bank.[23] This is interesting, given that (as discussed above) SHG members and comparison women were equally likely to report that they had *access to* credit from those more formal providers.

Respondents were also asked about the size of the largest loan they (or other members of the household) had borrowed during the 12 months prior to the survey. Column 5 of Table 5.9 compares the households of SHG members and comparison respondents in this respect. Even though a larger proportion of the comparison households had not borrowed at all during the year (that is, their largest loan was zero), the difference in average loan size between the two groups is not statistically significant. In fact, among those households who had borrowed at all, the largest loan borrowed by SHG members was 2100 birr on average, while among the comparison households it was nearly 3000 birr on average.[24] Among the households of SHG members that had taken out a loan during the year, 60 per cent of them had borrowed no more than 2000 birr. In contrast, the comparison households who had borrowed tended to take out larger loans, with only 30 per cent of them borrowing no more than 2000 birr. This suggests that many of the loans borrowed from the self-help groups were for small amounts, and were made to households that otherwise would not have borrowed at all.

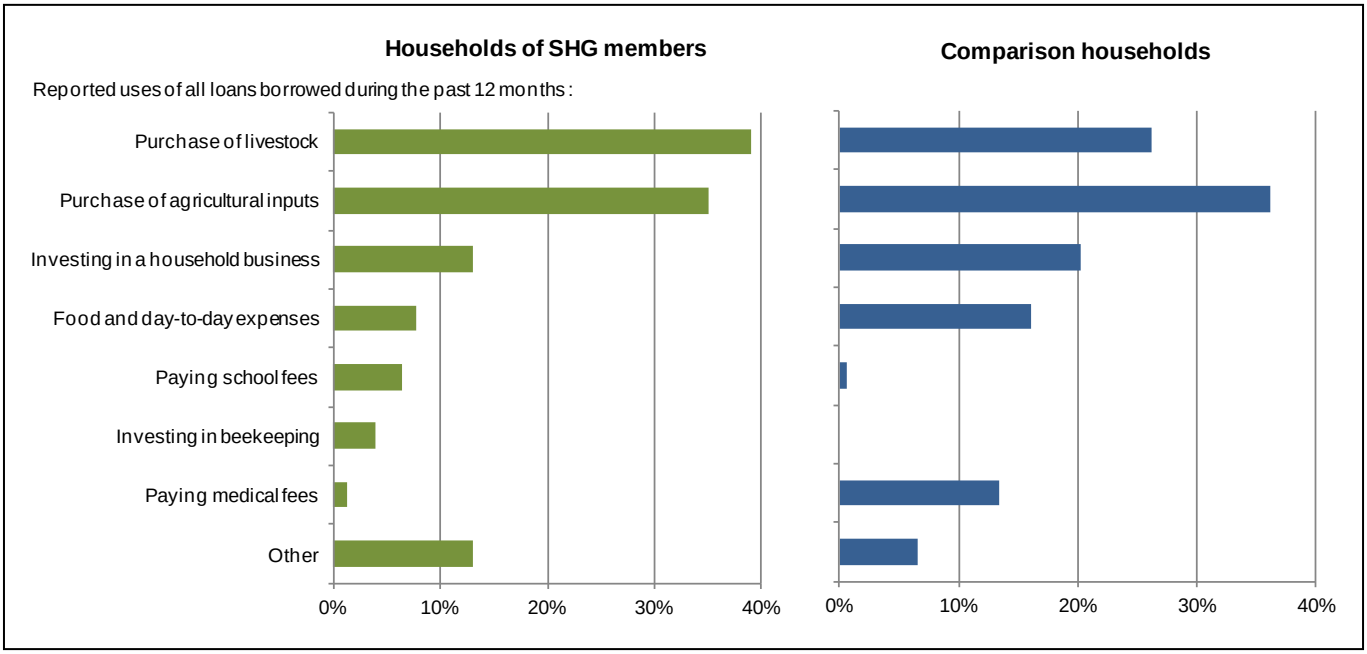
Table 5.9: Households' use of credit

	1	2	3	4	5
	Household took out any loan during past 12 months %	Number of sources of credit used by household during past 12 months	Household borrowed from a self-help group or other community group during past 12 months %	Household borrowed from a microfinance institution or bank during past 12 months %	Largest loan borrowed by household during past 12 months (birr)
Intervention group	44.8	0.49	18.6	23.3	928.2
Comparison group	28.6	0.29	5.1	15.1	846.3
Difference:	16.2** (7.1)	0.20*** (0.07)	13.5*** (4.7)	8.1* (4.8)	81.9 (245.3)
Observations	172	172	172	172	172
Observations (total):	581	581	581	581	581

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

Survey respondents were also asked for some information about the reasons for which they had borrowed. As can be seen in Figure 5.2, the most common use of loans reported among both the SHG members and comparison households was for investment in livestock or agricultural inputs.[25] Among the comparison households, loans were widely used for paying day-to-day expenses and medical fees – something that was reported less often among the SHG members. A small but significant minority of SHG members reported using loans to pay school fees.

Figure 5.2: Reported loan uses



It is notable in Figure 5.2 that only three SHG members (that is, approximately two per cent of the SHG members overall) reported having used credit to invest in beekeeping. This is intriguing in the light of feedback from focus group discussions that lack of availability of finance had prevent participants from making greater investments in beekeeping. It is possible that the loan sizes or loan terms available to respondents were not considered suitable for use in beekeeping – although such limitations would

be expected to apply to the use of credit for investing in agriculture and livestock as well. Alternatively, it may be that the supply of inputs for beekeeping (such as the modern hives) was so limited that borrowers were not able to find opportunities to invest additional funds in beekeeping. Another plausible explanation is that people may simply have felt that agriculture and livestock have the potential to generate larger, or less risky, returns.

5.4 FOOD SECURITY AND DIETARY DIVERSITY

We have seen that the households of SHG members generated greater revenue from sales of honey than did comparison households. In addition, SHG members were saving more regularly and more of them were taking out loans (particularly small-value loans) than those in comparison households, mostly for use in productive investment. Each of these effects may be expected to result in an increase in income for the household, at least in the long term. The questionnaire used for this Effectiveness Review did not include questions about overall household income, but two measures that can be examined with the survey data may be reasonably good indicators of a household's overall well-being. In this section we will look measures of food security and dietary diversity, and in Section 5.5 at asset ownership and other indicators of material wealth.

Survey respondents were asked a series of questions intended to reflect whether their household has secure access to food throughout the year. The analysis of this characteristic was based on seven common indicators of food security, adapted from the Household Food Insecurity Access Scale.[26] Respondents were asked how frequently they and other household members had experienced the following:

- Having to reduce the size of meals because there was not enough food.
- Having to eat fewer meals in a day than normal because there was not enough food.
- Having to eat poorer quality or less preferred types of food than normal.
- Having to reduce the amount eaten by adults, so that children could eat.
- Having to go to sleep at night hungry because there was not enough food.
- Spending a whole day and night without eating because there was not enough food.

Unusually in this case, these questions were asked not about the household's experiences during the weeks prior to the survey, but instead about their experience during the previous summer (approximately June to August). Since this is the time of year when food shortages are typically most severe, it was thought that this would provide the best indication of the household's ability to secure year-round food security.

Further, respondents were asked to answer separately how often each of these situations had occurred to them personally, and how often to their husband or other adult male members of the household. Responses were ranked on a scale from zero (meaning that the problem was never encountered during the previous winter) to three (meaning that the problem was encountered most days during the previous winter).

Table 5.10: Indicators of food security

	1	2	3	4
	Food security score[a] (female respondent)	Food security score[a] (male household members)	Severe food insecurity (female respondent) %	Severe food insecurity (male household members) %
Intervention group mean:	15.7	16.2	10.1	6.9
Comparison group mean:	15.4	15.8	17.3	12.4
Difference:	0.3 (0.3)	0.3 (0.35)	-7.3 (5.4)	-5.2 (4.6)
Observations (intervention group):	168	144	168	144
Observations (total):	571	475	571	475

[a] On a scale from zero to 18. Higher values represent fewer food security problems.
Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

A food security score was created separately for women and for men by adding together the rankings for incidence of each of the six food security problems. The resulting scores range from zero to 18, with higher scores representing lower incidence of food security difficulties.

The results are shown in columns 1 and 2 of Table 5.10. There is no evidence of a difference between the SHG members and comparison women (and between men in their respective households) in terms of food security problems experienced.
By comparing the figures in those two columns, it can be seen that fewer food security problems were reported among men than among women. For example, 6.4 per cent of respondents said that they personally had to reduce the size of their meals most days during the past summer, but only 5.3 per cent said that the same applied to male household members. However, most of these gender differences are small: the majority of respondents reported that male members of their households experienced the same food security difficulties as they did themselves.

The survey questions were also used to generate an indicator of severe food insecurity – defined as having to go to for a whole day and night without eating during the past summer, or suffering any of the other five food security problems frequently during that season. The results of this indicator are shown in columns 3 and 4 of Table 5.10. On this measure smaller proportions of SHG members (and male members of their households) seem to have experienced food insecurity than comparison respondents. Although these differences are large, it is not clear that they are statistically significant.[27] However, there is some evidence of a difference specifically among female-headed households: 43 per cent of comparison respondents in female-headed households suffered from severe food insecurity by this definition, whereas only 20 per cent of SHG members in female-headed households did so.[28]

The second approach adopted to assess households' food consumption was to ask directly about the range of food types eaten by the respondent and by male household members. Respondents were presented with a list of 13 food types, and asked on how many days of the past seven days the food type had been eaten by household members.[29] Note that this time respondents were asked about their current dietary situation, rather than recalling the previous summer.[30]

The results were used to create two different indicators of dietary diversity, the first being a simple score created by adding together the number of days that each food types was consumed. The maximum score possible score was 91. In fact, the maximum score encountered among the respondents interviewed was 55, and the

mean and median scores were both 30. Columns 1 and 2 of Table 5.11 show that there is a clear positive difference between the diets of SHG members (and of male members of their households) and that of comparison women.

Secondly, the food consumption data was also used to construct an indicator of positive dietary diversity, defined to be positive for households meeting all of these standards:

- Household members consumed a protein source (pulses, eggs, dairy products, meat or fish) on at least four of the previous seven days.
- Household members consumed fruit or vegetables on at least three of the previous seven days.

Nearly a quarter (23 to 24 per cent) of the SHG members and male members of their households met this standard for dietary diversity, compared to only nine per cent of comparison respondents.

One interesting factor to consider is the extent to which households' dietary diversity is linked to the use of kitchen gardens. A higher (though not statistically significant) proportion of the SHG members were farming from a kitchen garden at the time of the survey than were the comparison respondents (73 per cent, against 67 per cent). These households generally had higher dietary diversity scores than did those who were not producing from kitchen gardens. However, there is a significant difference between SHG members and comparison respondents in dietary diversity even if fruit and vegetables are excluded – showing that the use of kitchen gardens cannot be the full explanation for this difference.

The measures of dietary diversity, then, show that SHG members and their households are generally eating a more diverse diet than comparison respondents. This gives added weight to the indication from Table 5.10 that they were also less likely to suffer from extreme food insecurity during the previous summer.[31] We should be cautious in claiming that this difference is an effect of the project. This is because (as noted in Section 4.2 and in Appendix 3) even under the matching process used to derive these results, more of the households of SHG members were in the top quintile according to wealth indicators in 2009 than were the households of comparison respondents. Since diet is likely to be linked to material wealth,[32] it is possible that there were differences between the two groups in the diversity of their diet even before the project started. However, it seems unlikely that the small baseline differences in wealth status would be sufficient to fully account for the large difference in the proportions experiencing positive dietary diversity that are shown in columns 3 and 4 of Table 5.11.

It will also be noted from Table 5.11 that, in contrast to the food security indicators shown in Table 5.10, there were few differences in the variety of food reported to be consumed by the female respondents and by male household members.

Table 5.11: Indicators of dietary diversity

	1	2	3	4
	Dietary diversity score[a] (female respondent)	Dietary diversity score[a] (male household members)	Positive dietary diversity (female respondent) %	Positive dietary diversity (male household members) %
Intervention group mean:	33.3	33.7	23.3	23.8
Comparison group mean:	30.6	30.3	8.6	8.6
Difference:	2.6*** (0.9)	3.6*** (1.1)	14.7*** (4.9)	15.1*** (5.5)
Observations (intervention group):	172	147	172	147
Observations (total):	572	469	572	470

[a] On a scale from zero to 91. Higher values represent improved dietary diversity.
Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01; PSM estimates bootstrapped with 1000 repetitions, with standard errors clustered by community.

5.5 INDICATORS OF MATERIAL WEALTH

To provide an indicator of each household's overall material wealth, a household wealth index was created. Respondents were asked to provide information about their household's ownership of various assets (including land, livestock, productive equipment and household goods), as well as about the conditions of the family's house, both in 2009 and at the time of the survey.

If each of the assets and housing characteristics are indicators of household wealth, they should be correlated with each other. That is, a household that scores favourably on one particular wealth indicator should be more likely to do so for other wealth indicators. Items that had low correlations with the others were therefore not included in the index.[33] In particular, data relating to the type of water source used by the household and the type of cooking fuel used by the household were found not to be correlated with the other wealth indicators, and so were not used in the construction of the wealth index.

A data reduction technique called principal component analysis (PCA) was used to produce two indices of overall wealth, one based on the recalled data from 2009, and one based on the household's situation at the time of the survey. PCA produces a measure that maximises the variation in asset types by assigning more weight to those assets that are most highly correlated with the inter-item variation. Hence, each household's weighted index score is determined by both the number of assets it owns, and by the weight assigned to each asset type. The resulting index enables the relative wealth status of the households to be compared. The wealth index for 2009 is the measure that has been used throughout this analysis to control (to the greatest extent possible) for baseline differences in wealth status among the households of the various treatment groups. After calculating the wealth index for both 2009 and the date of the survey, households were categorised according to the quintile in which they lie – that is, the top 20 per cent of households according to wealth indicators were categorised together, as were those in the next 20 per cent, and so on. The measure reported in Table 5.12 is based on households moving between quintiles. For example, a household that changed from being among the bottom 20 per cent of the sample in 2009 to being in the 20–40 per cent quintile at the time of the survey would be given a score of +1. A household that moved from the middle quintile to the bottom quintile would have a score of –2.

Table 5.12: Change in index of wealth indicators

	Number of quintiles of wealth index in which household increased
Intervention group mean:	0.047
Comparison group mean:	0.061
Difference:	-0.015 (0.17)
Observations (intervention group):	172
Observations (total):	581

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; PSM estimates are bootstrapped with 1000 repetitions, with standard errors clustered by community.

Unfortunately the result in Table 5.12 does not provide any indication that households of SHG members saw more positive changes in terms of indicators of material wealth since 2009 than did comparison households.

5.6 WOMEN'S EMPOWERMENT

The project under review was specifically aimed at increasing women's empowerment. The approach used to assess progress on women's empowerment in this and other Effectiveness Reviews involved using responses from the survey to construct a multi-dimensional measure of empowerment. This approach builds on the 'Women's Empowerment in Agriculture Index'[34] (WEAI) developed by the Oxford Poverty and Human Development Initiative with support from the United States Agency for International Development (USAID) and the International Food Policy Research Institute (IFPRI).

This Effectiveness Review assesses the project's impact against five dimensions of women's empowerment:

1. Women's ability to make decisions and influence
2. Self-perception.
3. Personal freedom.
4. Access to and control over resources.
5. Support from social networks.

Several indicators, listed in Table 5.13, were specified for each of these five dimensions. 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 are deemed to be important to women's empowerment in this particular context.

Table 5.13: Specific characteristics of women's empowerment examined in this Effectiveness Review

Dimension	Characteristic
Ability to make and influence decisions	Involvement in productive decisions of the household
	Involvement in expenditure decisions of the household
	Involvement in household management decisions
	Influence in community decision-making
Self-perception	Self-confidence
	Opinions on women's economic roles
	Opinions on gender rights
	Opinions on power relations in the household
Personal freedom	Personal autonomy
	Able to adjust burden of care responsibilities
	Labour-saving equipment
	Attitude to gender-based violence
	Experience of violence
Access to and control over resources	Independent income
	Ownership of strategic assets
	Personal savings
	Literacy
Support from social networks	Social connections
	Participation in community groups

The questionnaire used in the Effectiveness Review included questions relating to each of the characteristics listed in Table 5.13. 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 the sections that follow, and are presented in summary form in Appendix 1. There is inevitably a degree of arbitrariness in defining such cut-offs. In some cases alternative definitions of the indicators and the cut-offs have been tried, as a check on the robustness of the results.

In the pages that follow, we will consider how project participants differ from comparison households in each of the women's empowerment characteristics listed in Table 5.13. First, however, we examine how all of the characteristics combine to provide an overall measure of women's empowerment.

Aggregate measures of women's empowerment are constructed a multidimensional measurement methodology known as the Alkire–Foster Method.[35] The first measure of overall women's empowerment considered is the proportion of characteristics in which the respondent scored positively, which we call the *base empowerment index*.[36] This is the measure for which the results are shown in column 1 of Table 5.14. A woman was then defined as having positive empowerment *overall* if she met the threshold for positive empowerment in at least two thirds of these characteristics. A second empowerment index was then created, which takes a value of 1 if the respondent reaches the threshold for overall empowerment, and otherwise is equal to the proportion of characteristics in which the household scored positively. This modified index is known as the *Alkire–Foster empowerment index*, and has the effect of concentrating the analysis on differences between those who do not meet the benchmark for overall empowerment. The results from applying this measure are shown in column 2 of Table 5.14.

The Oxfam GB global indicator for women's empowerment is based on whether women are doing better in terms of overall empowerment than a 'typical' household in the area. This is defined by comparing each household's women's 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. Column 3 of Table 5.14 shows the comparison between SHG members and comparison respondents in terms of this measure.

Table 5.14: Overall indices of women's empowerment

	1	2	3
	Base empowerment index	Alkire-Foster empowerment index	Respondents meeting threshold for global indicator for women's empowerment %
Intervention group mean:	0.55	0.79	54.7
Comparison group mean:	0.50	0.75	45.7
Difference:	0.05** (0.02)	0.05 (0.03)	8.9 (7.45)
Observations (intervention group):	172	172	172
Observations (total):	581	581	581

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; PSM estimates are bootstrapped with 1000 repetitions, with standard errors clustered by community.

As can be seen from the first column of the table, SHG members scored positively in 55 per cent of the characteristics, compared to 50 per cent of the comparison respondents, a difference which is statistically significant. In terms of the Alkire-Foster index, the difference between the groups is again around five percentage points, though it is not clear that the difference is statistically significant. Subgroup analysis shows that there is no evidence of a positive effect on the indices of women's empowerment among SHG members in Libokemkem woreda, but the result is present (and cannot be differentiated in size) across the other three woredas. There is no evidence of a difference in the effect on the indices of empowerment between respondents in female-headed and male-headed households.[37]

Column 3 of the table shows clearly that more of the SHG members met the global indicator for women's empowerment than did the comparison respondents – though again this difference is not clearly statistically significant.

These results (particularly in terms of the base empowerment index) imply that the project has resulted in some positive effects on empowerment among the SHG members. This raises the question of which specific dimensions and characteristics of empowerment have seen this increase. The split between SHG members and comparison respondents in each of the characteristics of empowerment considered in this review is shown in Figure 5.3. The following sub-sections describe these results in more detail.

Dimension 1: Women's ability to make and influence decisions

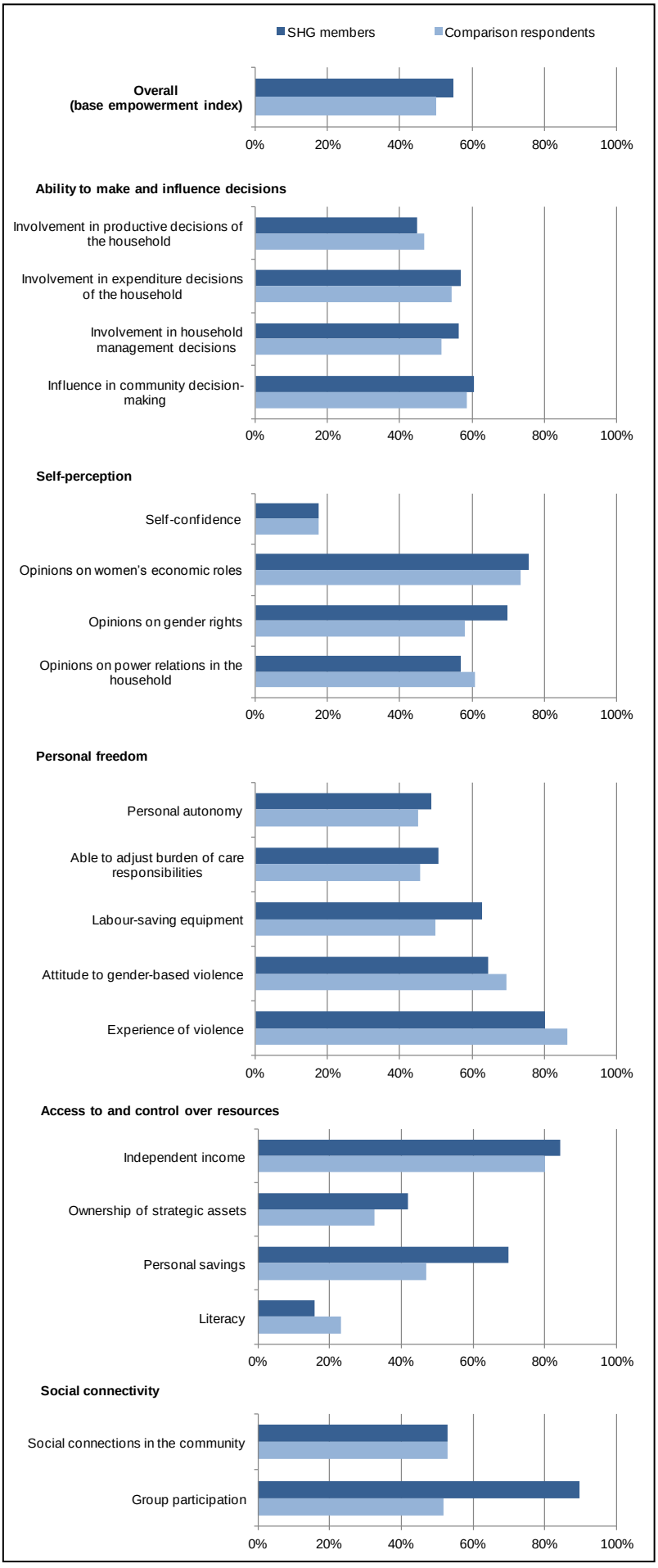
The first dimension of women's empowerment considered in the effectiveness review focused on women's influence in decision-making processes in their households and in their communities.

The results regarding women's decision-making power in the household are based on questions in the survey that addressed household decision-making in three different areas, specifically:

- **Decisions on productive activities**: Decisions relating to the conduct of a household's farming activities (e.g. type of crops household plants), to household businesses (e.g. how the business is managed, how many days to work, etc.) and to the sales or purchases of agricultural and non-agricultural produce/assets.
- **Decisions on household expenditures**: Decisions over how the money earned from various agricultural and non-agricultural activities is spent
- **Decisions on household management**: Decisions over such things as participation in or contributions to community events, decisions about the education of children and how to respond when a household member becomes ill.

For each of these decision-making areas, the respondent was first asked who in the household normally takes the decisions (if the household makes such decisions at all). If the respondent reported that she was not the sole decision-maker, she was also asked to what extent she thought she could influence such a decision, on a four-point scale ranging from 'not at all' to 'a large extent'. A woman scored positively on the measure of involvement in productive decisions if she reported being involved to at least a medium extent in at least half of the productive decision-making areas in which the household was active. The same applies to the indicators for involvement in expenditure and household-management decisions.

Figure 5.3: Results for indicators of women's empowerment



The results for these three measures of involvement in household decision-making are shown in columns 1 to 3 of Table 5.15. It can be seen that there are no statistically significant differences between SHG members and comparison respondents in any of these three areas. In terms of involvement in household-management decisions, the results from the various alternative statistical models used are consistently positive and some of them are statistically significant. This provides some indication that there is a positive effect from the project on this characteristic, but this cannot be stated with confidence.

Table 5.15: Women's ability to make and influence decisions

	1	2	3	4
	Involvement in productive decisions in the household %	Involvement in expenditure decisions in the household %	Involvement in household-management decisions %	Influence in community decision-making %
Intervention group mean:	44.8	57.0	56.4	60.5
Comparison group mean:	46.9	54.4	51.6	58.6
Difference:	-2.1 (7.5)	2.5 (7.7)	4.8 (7.9)	1.8 (7.4)
Observations (intervention group):	172	172	172	172
Observations (total):	581	581	581	581

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; PSM estimates are bootstrapped with 1000 repetitions, with standard errors clustered by community.

The final column of Table 5.15 shows the comparison of the SHG members and comparison respondents in terms of **influence in community decision-making**. Results for this characteristic were obtained by asking each survey respondent for the extent to which she agreed or disagreed with the following four statements:[38]

- Kebele leaders now take the opinions of women into account more than they used to.
- Women are able to influence the important decisions which are taken in this community.
- Women have become more active in decision making in the last few years.
- Women's participation in local groups is as important as men's participation.

Each respondent was deemed to have scored positively in terms of community influencing if she agreed with all four of these statements. As shown in column 4 of the table, approximately 60 per cent of respondents did so. Again there was little difference in the proportion between the SHG members and the comparison respondents in the proportion who met that threshold.[39]

Dimension 2: Women's self-perception

The second dimension of women's empowerment considered in this report includes four different elements of women's self-perception. The first of these is **self-confidence**. Respondents were asked for the extent to which they agreed or disagreed with the following statements:[40]

- I often do what others (partners or community group leaders) tell me to do, even if it is against my wishes.
- I often trust others' decisions over mine concerning my life.
- I often compare myself with others.

Respondents were scored positively in terms of self-confidence if they disagreed with at least two of these three statements. As can be seen in the first column of Table 5.16, this applied to only a minority of respondents, and there was no indication that this varied between the SHG members and comparison respondents overall.[41] SHG members in Bahir Dar Zuria responded especially positively, but those in Gozamen responded more negatively than the comparison respondents.[42]

Table 5.16: Characteristics of women's self-perception

	1	2	3	4
	Self-confidence %	Opinions on women's economic roles %	Opinions on gender rights %	Opinions on power relations in the household %
Intervention group mean:	17.4	75.6	69.8	57.0
Comparison group mean:	17.4	73.5	57.9	60.8
Difference:	0.0 (5.6)	2.05 (6.9)	11.9 (7.5)	-3.8 (7.5)
Observations (intervention	172	172	172	172
Observations (total):	581	581	581	581

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; PSM estimates are bootstrapped with 1000 repetitions, with standard errors clustered by community.

The remaining three indicators considered under this dimension examined respondents' opinions on women's and men's roles and rights, both in the home and outside the home. Each of these indicators is informed by female respondents' reactions to a subset of statements that were presented to them during the questionnaire. Again, women were asked to state the extent of their agreement or disagreement with each of the statements, on a four-point scale.

Opinions on women's economic roles were assessed by asking for the extent to which they agreed or disagreed with the following three statements:[43]

- These days women are contributing more income to the household's needs than they used to before.
- A woman needs to have some income/resources of her own to take care of herself and her children.
- Women are just as capable as men of contributing to household income.

Column 2 of Table 5.16 shows the proportions of respondents who agreed with all three of these statements. This applied to about three quarters of respondents, but again there was no indication of a difference between the responses of SHG members and comparison respondents.

Similarly, three statements were used to assess respondent's **opinions regarding gender rights**:[44]

- A good marriage is more important for a girl than a good education.
- Boys and girls should be given equal opportunities to learn how to read and write.
- It is important that a girl is old enough to make her own decision before she can be married.

Respondents were scored positively on this indicator if they disagreed with the first statement and agreed with the other two. Again this applied to the majority of respondents. Although the proportion scoring positively was estimated to be higher among the SHG members, this difference was not statistically significant, so again cannot be concluded that this represents an impact of the project.[45] However, there is a very clear difference specifically among female-headed households: most (87 per cent)

of the SHG members from female-headed households responded positively, against only 28 per cent of the corresponding comparison respondents.

Finally, three statements were used to assess respondents' **opinions on power relations within the household:**[46]

- If a woman does not agree with her husband, she should express her opinion directly.
- If a woman does not agree with her husband, she should show him some indication of her disagreement.
- A wife should never question the decisions made by her husband.

Women were scored positively if they agreed with either of the first two statements and disagreed with the third statement. Once again, although the majority of respondents scored positively, there was no indication of a difference between the SHG members and comparison respondents.[47]

Dimension 3: Personal freedom

The survey included questions relating to five characteristics of personal freedom. Table 5.17 shows the results of the comparison of women in project and comparison communities in terms of each of these characteristics.

The first characteristic considered under this dimension is the degree of **autonomy** that the respondent has in her movements and her participation in activities outside the home. This indicator is based on questions that followed the same format as those on household decision-making, discussed above. Respondents were asked who would take the decision about whether they could travel to visit relatives outside the community, and whether they could participate in community group activities or meetings. If the respondent stated that the decision was not made solely by her (i.e. that her husband or other household members had some involvement in the decision), she was asked to what extent she would be able to influence that decision. Respondents were scored positively if they stated that they made these decisions alone, or if they said that they have a great deal of influence over the decisions.[48] Just under half of respondents met this criterion, though again there was no indication of a difference between SHG members and comparison respondents.

An alternative measure of personal autonomy was constructed, using another three questions in which respondents were asked whether they agreed or disagreed with the following:

- A woman should always seek permission from her husband before participating in community meetings.
- You have to consult your husband before going to attend women group activities.
- You could travel alone to the woreda capital without asking for permission from anyone, if you wanted to.

Women were scored positively if they disagreed with the first two statements and agreed with the third. This applied to about six per cent of the SHG members, but to only about one per cent of the comparison respondents.[49] This would imply that the project may have had some positive effect on women's freedom to participate in community activities and to travel. However, the evidence for this is not strong.

Table 5.17: Characteristics of women's personal freedom

	1	2	3	4	5
	Personal autonomy %	Able to adjust burden of care responsibilities %	Labour-saving equipment %	Attitude to gender-based violence %	Experience of violence %
Intervention group	48.8	50.6	62.8	64.5	80.2
Comparison group	45.1	45.5	49.8	69.4	86.4
Difference:	3.7 (7.8)	5.05 (7.5)	13.0* (7.35)	-4.9 (7.1)	-6.2 (5.1)
Observations (intervention group):	172	172	172	172	172
Observations (total):	578	581	581	581	581

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; PSM estimates are bootstrapped with 1000 repetitions, with standard errors clustered by community.

The second characteristic considered under this dimension was whether a woman is able to **adjust the burden of her care responsibilities** if necessary. This was assessed again by asking respondents about the extent to which they agreed or disagreed with these statements:

- If you want to participate in a group meeting or activities, then your husband or relatives would help by taking care of the children or housework.
- These days men are helping more with housework than they used to.

Respondents were scored positively if they agreed strongly with the first statement, and agreed at all with the second. (The requirement for *strong* agreement with the first statement was specified because more than 90 per cent agreed at least partially.) There was no indication of a difference in this respect between SHG members and comparison respondents.

In a separate section of the questionnaire, respondents were asked directly about how their time is allocated between various activities – including care responsibilities, livelihoods work and leisure or personal care – and also how this time allocation had changed since 2009. As would be expected, SHG members reported that the time they spend attending group meetings had increased since 2009 (80 per cent said that they now spend more time attending meetings than they did in 2009, against 65 per cent of the comparison respondents). It has already been noted in Section 5.3 that SHG members said that they personally had increased the time they spend in beekeeping since 2009 – and interestingly, the majority of them also reported increases (significantly larger than the comparison respondents) in the time they spend working on a kitchen garden or tending livestock. The result for kitchen gardening, at least, is consistent with the encouragement provided to SHG members to engage in this as a supplementary activity, and with the results on dietary diversity found in Section 5.4.

It would be important to understand what consequences this increase in the time SHG members were spending on livelihoods activities and attending meetings had on their other activities. Unfortunately the survey data do not provide a clear indication on this: there are few significant differences between the SHG members and comparison women in terms of changes in their time spent on care responsibilities or other work.[50] There is some evidence that SHG members were more likely than comparison women to say that they have less time for sleeping now than they did in 2009: 52 per cent of SHG members reported decreasing the time they spend sleeping, against only 38 per cent of comparison respondents. On the other hand, SHG members were more likely than comparison respondents to say that men and children in their households had

experienced a decrease in leisure time since 2009,[51] perhaps suggesting that there has in fact been some reallocation of care responsibilities from women towards men and children. More detailed data would be needed from a larger number of respondents to investigate these effects in more depth.

Access to labour-saving equipment was also thought to be an important contributor to women's empowerment. Respondents were deemed to score positively in this respect if either (a) their household owns a labour-saving stove, or if (b) they reported that the time it takes them to collect water is not more than 10 minutes. In this respect there is some indication that SHG members scored slightly higher than comparison respondents. However, this result is only marginally statistically significant,[52] so it does not provide clear evidence of an effect of the project activities.

Respondents were also asked for their opinion on the **acceptability of gender-based violence**. Specifically, women were asked whether they believe it is acceptable for a man to hit his wife if:

- She spends money on things he does not approve of.
- She goes outside of the home without his permission.
- He suspects that she has been unfaithful.
- She does not serve him as he expects to be served.
- For any reason at all, if he wants to.

Women scored positively in terms of their attitude towards domestic violence if they deemed it unacceptable for a husband to hit his wife in all of these situations. This applied to approximately two thirds of women – again, with no indication of a difference between the SHG members and comparison respondents.

The final characteristic of personal freedom considered in the survey was actual **experience of violence**. Rather than being asked whether they personally had experienced violence, respondents were asked whether any of the following incidents had happened to 'women close to them' during the past 12 months:[53]

- Having anything that belongs to her being stolen or taken away.
- Being insulted, humiliated, or made to feel bad about herself.
- Being threatened herself, or having somebody she cares about threatened.
- Being hit or attempted to being hit.
- Having serious injury inflicted to her.
- Being forced to have sex when she did not want to.

Fewer than 20 per cent of respondents reported that they knew of any of these types of incidents having happened during the past 12 months. Respondents were scored positively if they said that they did not know of any of these incidents occurring. Once again, there was no indication of any difference between the SHG members and comparison respondents in this respect.

Dimension 4: Access to and control over resources

Four indicators were examined in the survey to assess women's access to and control over resources. The results of this analysis are shown in Table 5.18.

The first indicator considered was whether a woman has some **income** independently from her spouse or other household members. To assess this, respondents were asked to estimate the proportion of income that she personally contributes to household income and resources, and was considered to score positively on this basis if she reported that she personally contributes at least 40 per cent. The majority of respondents reached that level, and again there was no difference between the SHG members and comparison respondents.[54]

The second characteristic examined under this dimension was women's **ownership of strategic assets** such as land, livestock, agricultural equipment and household goods. As already noted in Section 5.5, respondents were asked about their household's ownership of various types of assets. As a follow-up to these questions, they were then asked to specify which household member has decision-making control over these assets: that is, who could make decisions about whether to sell an item if necessary. Respondents were deemed to have scored positively if they reported having sole decision-making control over at least two types of asset.

Table 5.18: Characteristics of women's access to and control over resources

	1	2	3	4
	Independent income %	Ownership of strategic assets %	Access to savings %	Literacy %
Intervention group mean:	84.3	41.9	69.8	15.7
Comparison group mean:	80.1	32.6	46.9	23.2
Difference:	4.2 (5.7)	9.3 (7.1)	22.9*** (7.5)	-7.5 (6.7)
Observations (intervention group):	172	172	172	172
Observations (total):	581	581	581	581

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; PSM estimates are bootstrapped with 1000 repetitions, with standard errors clustered by community.

On this indicator, the proportion scoring positively was slightly higher among the SHG members, but again this result is not clearly statistically significant, so it cannot be stated with confidence that this is an effect of the project.[55]

Personal **savings** were also thought to be an important indicator of a woman's independence and empowerment. As discussed in Section 5.4, respondents were asked about the extent of their personal savings in terms of the number of days those savings would enable the household to live on in an emergency. Respondents were deemed to score positively in terms of savings if they reported that their household could live for 15 days or more in an emergency. Consistent with the results found earlier (in Table 5.7), there appears to be a clear effect of the project on the proportion of respondents with this level of personal savings, and this difference is even larger among respondents in female-headed households than in male-headed households.

Finally, each respondent was scored positively for **literacy** if she reported that she could read and write (which was presented in the questionnaire in contrast to simply being able to write her name). Not surprisingly, there was a strong correlation between respondents' literacy and their level of education. Since the project had not sought to address literacy among participants, it is also not surprising that there is no difference between SHG members and comparison respondents in this respect.

Dimension 5: Support from social networks

The final two characteristics included in the Effectiveness Review attempted to evaluate the strength of respondents' social networks. The results of the comparison for women in project and comparison communities in terms of these characteristics are shown in Table 5.19.

The first characteristic attempted to evaluate each woman's degree of social connectivity by presenting four further statements, and asking respondents the extent to which they agreed with them:[56]

- I feel that I am on the same wavelength as people in my social network.
- I am happy about the social network I have in my locality.
- Woman do not need to establish friendship with their neighbours in their locality as the husbands have such connections already.
- The social contacts I have with people in my social network feel superficial.

Each respondent scored positively on this indicator if she agreed with the first two and disagreed with the last two of these statements. This applied to around half of the respondents, with no indication of a difference between SHG members and comparison respondents.

Finally, respondents were asked which **community groups** and types of meetings they participate in, including *mahabar*, savings or microcredit groups, voluntary groups, kebele meetings and one-to-five groups, as well as women's self-help groups and honey cooperatives. Women scored positively if they reported that they regularly attended meetings of at least two of these types of groups, and that they had attended at least three meetings during the previous month.

Table 5.19: Characteristics of support from social networks

	1	2
	Social connections %	Participation in community groups %
Intervention group mean:	52.9	89.5
Comparison group mean:	52.8	51.8
Difference:	0.1 (7.3)	37.7*** (6.9)
Observations (intervention group):	172	172
Observations (total):	581	581

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; PSM estimates are bootstrapped with 1000 repetitions, with standard errors clustered by community.

In this respect, there is a very large and clear difference between the SHG members – of whom almost 90 per cent scored positively – and comparison respondents, of whom only 52 per cent did. Most of this difference results from respondents' participation in the SHGs themselves and in the beekeepers' cooperatives (as well as in local savings and microcredit groups, which may have been confused with self-help groups in the responses). However, this does not account fully for the difference: the SHG members also appear to be more likely than comparison respondents to participate in one-to-five groups and perhaps (though the evidence is less clear) also in kebele meetings and voluntary activities in the community.[57]

It is also interesting to note that there was a significant difference between SHG members and comparison respondents in the proportions who reported that they had

held a leadership position in some group, during the last few years. A third of SHG members said that they had held a leadership position, against only 20 per cent of comparison respondents. Even when considering excluding leadership positions in community groups directly connected with the project activities (that is, self-help groups, savings or credit groups, and beekeeping cooperatives), there is still a large difference between SHG members and comparison respondents that is verging on being statistically significant. This may imply that the project has had some impact on women's engagement in groups in the community at large. However, it is also possible that this difference reflects that the women who joined the SHGs were already more engaged in community groups than were the average comparison respondents (who were selected at random in their communities), and that this difference is not fully accounted for by the matching approach used in this evaluation. Unfortunately the data available cannot provide a conclusive judgement on this.

Summary of results on women's empowerment

From the breakdown of results discussed above, it can be seen that the two areas in which there is evidence of a clear impact of the project on women's empowerment are those directly related to the project activities: respondents' personal savings, and participation in community groups. There are some indications of an impact in other areas, including women's involvement in household-management decisions, attitudes towards gender rights (particularly among female-headed households), control over assets, having the freedom to travel and participate in community activities, and taking up leadership positions in community groups. However, in each of these cases the evidence for an effect from the project is not strong. In particular, the fact that there may have been baseline differences between the SHG members and comparison respondents that may not have been fully controlled for in the statistical analysis means that caution should be taken in attributing these differences to the impact of project activities. This conclusion is reinforced by the focus group discussions, in which the factors that were mentioned as having resulted in changes in women's position over the past few years were mostly government initiatives or broader changes in society in Ethiopia – including a new land titling law, the establishment of one-to-five groups and women's committees at kebele level, and decreasing acceptance of early marriage, female genital mutilation and gender-based violence – rather than being specifically linked to this project.

6 CONCLUSIONS

6.1 CONCLUSIONS

The survey results provide good evidence that the project under review was successful in encouraging households to take up – or to continue to engage in – beekeeping (particularly among female-headed households), and in encouraging them to experiment with new technologies. Efforts to improve links to the cooperatives and marketing channels also appear to have been successful, with a third of the households of SHG members having sold some honey to their local cooperative over the 12 months prior to the survey.

The prices gained from sales to the cooperatives were significantly higher than those realised through other channels, though it cannot be determined from the survey data whether this is because of differences in the quality of honey being sold, or purely because the cooperatives are more effective at marketing honey and can realise economies of scale. As a result of the difference in prices, households of SHG members generated considerably higher revenue from sales of honey than did comparison households. The level of revenue being generated is quite modest, and it appears that, for the vast majority of SHG members, honey production is not central to their livelihoods. However, this at least provides reasons for optimism that a project that could provide more intensive training and support to producers and overcome the difficulties in the supply of inputs has potential to generate significant returns.

The SHGs appear to have been successful in providing members with a means of saving and accessing credit. The personal savings of SHG members were considerably higher than those reported by comparison respondents, and approximately half believed that they would be able to obtain a loan of 1000 birr (approximately US$ 50) from their SHG if necessary. More of the SHG members (or other members of their households) had taken out loans during the 12 months prior to the survey, though most of the additional loans seem to have been quite small. Most loans were apparently invested in livestock or agriculture, with a smaller number going towards household businesses. It is notable that very few SHG members reported using their loans to invest in beekeeping.

SHG members and their households were apparently consuming a more diverse diet than comparison women, and were less likely to have suffered extreme food insecurity during the summer season previous to the survey. However, there was no indication of an effect from the project on an index of indicators of material wealth (asset ownership and housing conditions).

It is unclear to what extent this project has had a significant impact on women's positions within their households. While the majority of SHG members said that they were now involved in honey production, most also commented during focus groups that they lack ownership of these activities or control over the income generated. Among the various indicators of women's empowerment included in the survey, clear evidence was found of an impact only on those directly linked to the project activities – women's savings and involvement in community groups.

It should be noted that a follow-on project was launched in early 2013 in three of the seven woredas where the project under review was implemented (the three woredas that were *not* included in this Effectiveness Review, which are also thought to have greater potential for honey production). The newer project is being implemented by Oxfam and the Zembaba Union in partnership with Facilitator for Change, an Ethiopian

non-governmental organisation, and is also seeking to establish women's self-help groups and support them in honey production and in other income-generating activities. This project is being implemented at a larger scale and has employed an experimental approach to evaluation. Over time, this project should provide a stronger evidence base on the impacts that these interventions can have on livelihoods and on women's positions in their homes and communities.

6.2 PROGRAMME LEARNING CONSIDERATIONS

Ensure that the appropriate level of support is in place when introducing new technologies.

The Effectiveness Review provides evidence that the new beekeeping technologies and practices introduced under this project can enable women to take on significant roles in honey production and marketing, overcoming traditional attitudes that beekeeping is suitable only for men. However, it is clear that prerequisites to success are that learning about the technologies is fully embedded, that the required equipment and inputs are available, and that extension agents are fully involved and able to provide effective advice and support. Even more importantly, it appears that a good level of support and follow up is required in order that group members feel confident that they are able to apply the new technologies effectively, and that they will not be deterred by initial setbacks.

A related consideration is the appropriate level of intensity required for a project such as this. The project under review was originally intended to have a larger budget and to allow for a more intense package of advice and support to be provided to SHG members. Once it was clear that the planned budget was not available, it would perhaps have been more effective to have reduced the scale of the project in order to provide concentrated support to a smaller number of self-help groups, rather than a less comprehensive package for a greater number.

Continue to design more holistic projects and programmes, with specific strategies to contribute to women's empowerment and minimise risks of participation.

Another implication of this Effectiveness Review is that, even if a project can successfully encourage women's engagement in a new livelihood activity, that is not by itself sufficient to achieve empowerment in a broader sense. It is important to note that the project under review initially focused on beekeeping as a livelihoods activity and on promoting business development, and did not (at least at the design stage) have a clear strategy by which the project was expected to lead to social empowerment or to manage risks of women's participation. The weakness of this approach has already been recognised, as evidenced by the design of the current honey value-chain development project, which incorporates a specific gender strategy, including community forums, literacy training, and the promotion of labour-saving devices. There will always be a tension between the extent to which a programme should focus on business development and social justice, but it at least seems to be clear that social empowerment will not necessarily result from economic empowerment.

Carefully monitor the requirements that project activities place on participants' time, in order to mitigate any negative consequences.

The fact that SHG members were more likely than comparison respondents to report that they have experienced a decrease in the time they spend sleeping over the past few years serves as a reminder of the importance of being aware of the additional burden that project activities place on participants. Clearly it is important to ensure that the value being realised through any intervention outweighs the potential costs, either in terms of increased stresses on participants or their family members, or in terms of reduced engagement in other livelihoods activities. This should be monitored carefully in current and future projects through the use of regular discussions with group members, so that action can be taken to reduce those burdens if necessary. It is also possible that the greater use of labour-saving technologies may be able to support women in reducing household responsibilities and encouraging men to take on some of these duties.

Review the monitoring, evaluation and learning approaches for current and future projects, particularly in order to monitor changes in the enabling environment and how project participants are responding to them, as well as how this is translating into increased empowerment.

The results of this Effectiveness Review reinforce the need for investment in a comprehensive system for monitoring implementation and outcomes. In particular, these findings suggest that emphasis should be put on monitoring of changes in the context (such as the availability of inputs or the involvement of extension services) and on regular discussions with project participants on the effects of the project activities, while implementation is proceeding. Just as important is to monitor how the effects of projects are distributed across different groups of participants, so as to improve the effectiveness with which interventions can be targeted.

The issues raised in this report can also be used to inform the questions to be investigated in more formal evaluation processes. In particular, the inclusion of indicators of access to and use of credit, of time use, and of contact with extension services and other service providers will strengthen the midline and endline surveys to be carried out for the current honey value-chain development project, and for future similar initiatives. At the same time, the refinement and testing of indicators of social empowerment should continue, to ensure that they provide as full and accurate a picture as possible of women's experience of empowerment in Ethiopia.

APPENDIX 1: THRESHOLDS FOR CHARACTERISTICS OF EMPOWERMENT

Dimension	Characteristic	Threshold: respondent scores positively if...
Ability to make and influence decisions	Involvement in productive decisions of the household	Respondent reports having influence in at least half of the productive decisions that the household engages in (out of 10 types of decision listed).
	Involvement in expenditure decisions of the household	Respondent reports having influence in at least half of the expenditure decisions that the household engages in (out of 11 types of decision listed).
	Involvement in household management decisions	Respondent reports having influence in at least half of the household management decisions that the household engages in (out of seven types of decision listed).
	Influence in community decision-making	Respondent agrees with all four of the following statements: • Kebele leaders now take the opinions of women into account more than they used to. • Women are able to influence the important decisions which are taken in this community. • Women have become more active in decision making in the last few years. • Women's participation in local groups is as important as men's participation.
Self-perception	Self-confidence	Respondent disagrees with at least two of the following three statements: • I often do what others (partners or community group leaders) tell me to do, even if it is against my wishes. • I often trust others' decisions over mine concerning my life. • I often compare myself with others.
	Opinion on women's economic roles	Respondent agrees with all three of the following statements: • These days women are contributing more income to the household's needs than they used to before. • A woman needs to have some income/resources of her own to take care of herself and her children. • Women are just as capable as men of contributing to household income.
	Opinion on gender rights	Respondent agrees with the following statement: • A good marriage is more important for a girl than a good education. and disagrees with both of the following statements: • Boys and girls should be given equal opportunities to learn how to read and write. • It is important that a girl is old enough to make her own decision before she can be married.
	Opinion on power within the household	Respondent agrees with at least one of the two following statements: • If a woman does not agree with her husband, she should express her opinion directly. • If a woman does not agree with her husband, she should show him some indication of her disagreement. and disagrees with the following statement: • A wife should never question the decisions made by her husband.

Dimension	Characteristic	Threshold: respondent scores positively if...
Personal freedom	Personal autonomy	Respondent reports that she alone is responsible for decisions about her travel to visit relatives outside the community and about her participation in community group activities, or that she has a great deal of influence over those decisions.
	Able to adjust burden of care responsibilities	Respondent agrees strongly with the following statement: • If you want to participate in a group meeting or activities, then your husband or relatives would help by taking care of the children or housework. and agrees with the following statement: • These days men are helping more with housework than they used to.
	Labour-saving equipment	Household owns a time-saving stove and/or collecting water takes no longer than 10 minutes.
	Attitude to gender-based violence	Respondent does not agree that it is acceptable for a man to hit his wife in any of the five situations listed.
	Experience of violence	Respondent does not report that any of the six types of violence listed have happened to women close to her during the past 12 months.
Access to and control over resources	Independent income	Respondent estimates that she personally contributes at least 40 per cent to the household's income.
	Ownership of strategic assets	Respondent reports that she has sole decision-making control over at least two types of livestock or household asset.
	Access to savings	Respondent reports that her personal savings are sufficient to enable the household to live on in an emergency for at least 15 days.
	Literacy	Respondent reports that she is able to read and write (more than simply her name).
Support from social networks	Social connections	Respondent agrees with both of the following statements: • I feel that I am on the same wavelength as people in my social network. • I am happy about the social network I have in my locality. and disagrees with both of the following statements: • Woman do not need to establish friendship with their neighbours in their locality as the husbands have such connections already. • The social contacts with people in my social network feel superficial.
	Participation in community groups	Respondent reports that she participates in at least two types of group in the community, and has attended group meetings at least three times during the past month.

APPENDIX 2: BASELINE STATISTICS BEFORE MATCHING

		Intervention mean	Comparison mean	Difference	Standard error of difference
Number of household members in 2009		5.65	4.6	1.1***	(0.2)
Proportion of household members who were children (less than 18 years old) in 2009	%	58.2	49.8	8.3***	(1.7)
Proportion of adult household members in 2009 who were fit and able to work	%	97.1	94.6	2.5*	(1.4)
Household had only one adult member in 2009	%	13.8	13.9	-0.2	(3.0)
Household had no male adult members in 2009	%	15.3	16.4	-1.1	(3.2)
All adult household members were elderly (over 60 years old) in 2009	%	0.5	2.3	-1.75	(1.1)
Respondent is the head of household	%	18.5	19.2	-0.7	(3.4)
Age of respondent in 2009	years	36.3	35.0	1.3	(1.0)
Respondent was elderly (over 60 years old) in 2009	%	1.1	4.6	-3.5**	(1.6)
Respondent is fit and able to work	%	100.0	97.0	3.0*	(1.7)
Respondent is widowed or divorced	%	15.9	18.5	-2.6	(3.3)
Respondent has any education	%	30.2	25.6	4.6	(3.9)
Respondent has any formal education	%	11.6	13.7	-2.1	(2.9)
Number of years respondent spent in formal education	years	0.8	0.7	0.1	(0.2)
Household head is female	%	19.6	21.2	-1.7	(3.5)
Age of household head in 2009	years	44.3	42.9	1.4	(1.1)
Household head is fit and able to work	%	96.3	93.6	2.7	(2.3)
Household head has any education	%	60.3	42.2	18.1***	(4.3)
Household head has any formal education	%	29.1	18.3	10.8***	(3.6)
Number of years household head spent in formal education	years	2.0	1.0	1.0***	(0.25)
Proportion of adult household members in 2009 with any education [a]	%	51.3	37.2	14.1***	(3.35)
Proportion of adult household members in 2009 with any formal education [a]	%	28.8	21.5	7.3***	(2.8)
Household was engaged in beekeeping in 2009	%	81.5	12.6	68.9***	(3.05)
Number of beehives owned by household in 2009[a]		5.2	0.4	4.8***	(0.4)
Number of beehives owned by household in 2009 (restricted to those who were engaged in beekeeping) [a]		6.4	3.15	3.3***	(1.1)
Household cultivated any land in 2009	%	87.3	88.1	-0.83	(2.8)
Land area cultivated by household in 2009	*qodda*	4.9	4.3	0.6*	(0.35)
Household owned any livestock in 2009	%	87.3	80.4	6.9**	(3.3)
Any household member was engaged in casual labour in 2009	%	10.1	10.3	-0.2	(2.6)
Any household member was engaged in a non-agricultural household business in 2009	%	32.8	25.3	7.5*	(3.9)
Any household member was engaged in formal employment in 2009	%	2.1	1.1	1.0	(1.0)
Any household member received income from renting out land or property in 2009	%	7.4	8.0	-0.6	(2.3)
Household received remittances during the 12 months prior to the survey	%	7.9	2.7	5.2***	(1.8)
Household was in the lowest 20% of the sample according to wealth indicators recalled from 2009[a,b]	%	12.2	23.5	-11.3***	(3.5)
Household was in the second 20% of the sample according to wealth indicators recalled from 2009[b]	%	10.6	24.0	-13.4***	(3.4)
Household was in the middle 20% of the sample according to wealth indicators recalled from 2009[b]	%	13.8	22.8	-9.1***	(3.5)

		Intervention mean	Comparison mean	Difference	Standard error of difference
Household was in the fourth 20% of the sample according to wealth indicators recalled from 2009[b]	%	27.5	16.7	10.8***	(3.5)
Household was in the upper 20% of the sample according to wealth indicators recalled from 2009[b]	%	36.0	13.0	23.0***	(3.4)
Household had an electricity connection in 2009	%	13.2	3.9	9.35***	(2.15)
Distance of house from nearest market place in 2009 minutes on foot (estimated by respondent)		99.2	89.3	9.8	(7.8)
Number of observations		**189**	**438**	**627**	

[a] These variables were not used for matching or as covariates in regression models, since they are highly correlated with other variables on the list.

[b] The construction of the wealth index is described in Section 5.5.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Variables dated 2009 are estimates, based on recall data or reconstructed from the composition of the household at the time of the survey.

APPENDIX 3: 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 woman is observationally 'similar' to another woman. For an overview, we refer to Caliendo and Kopeinig (2008).[58] 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)[59] 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 A3.1 and A3.2 present the probit regression results used to estimate the propensity scores in our context. **Error! Reference source not found.**Table A3.1 hows the probit results for the non-parsimonious model entering the full set of matching variables considered in this study. 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 (Caliende and Kopeinig, 2008).

Table A3.1: Estimating the propensity score: non-parsimonious model

	Coefficient	Standard error	*p*-value
Number of household members in 2009	0.176	(0.065)	0.007
Proportion of household members who were children (less than 18 years old) in 2009	-0.035	(0.598)	0.953
Proportion of adult household members in 2009 who were fit and able to work	0.848	(0.846)	0.316
Household had only one adult member in 2009 = 1	0.478	(0.328)	0.145
All adult household members were elderly (over 60 years old) in 2009 = 1	0.073	(1.193)	0.951
Respondent is the head of household = 1	0.022	(0.669)	0.974
Age of respondent in 2009 years	-0.002	(0.015)	0.902
Respondent was elderly (over 60 years old) in 2009 = 1	-0.626	(0.908)	0.490
Respondent is fit and able to work = 1	0.286	(0.446)	0.521
Respondent is widowed or divorced = 1	-0.401	(0.444)	0.367
Respondent has any education = 1	-0.141	(0.231)	0.542
Respondent has any formal education = 1	-0.811	(0.635)	0.202
Number of years respondent spent in formal education years	0.114	(0.098)	0.246
Household head is female = 1	0.716	(0.587)	0.223
Age of household head in 2009 years	0.004	(0.012)	0.714
Household head is fit and able to work = 1	-0.343	(0.442)	0.439
Household head has any education = 1	-0.031	(0.199)	0.877
Household head has any formal education = 1	-0.214	(0.372)	0.565
Number of years household head spent in formal education years	0.094	(0.053)	0.075
Household was engaged in beekeeping in 2009 = 1	1.967	(0.153)	0.000
Household cultivated any land in 2009 = 1	0.253	(0.293)	0.388
Land area cultivated by household in 2009 *qodda*	-0.011	(0.027)	0.672
Household owned any livestock in 2009 = 1	0.065	(0.228)	0.774
Any household member was engaged in casual labour in 2009 = 1	0.321	(0.253)	0.205
Any household member was engaged in a non-agricultural household business in 2009 = 1	0.101	(0.175)	0.563
Any household member was engaged in formal employment in 2009 = 1	0.049	(0.571)	0.931
Any household member received income from renting out land or property in 2009 = 1	0.152	(0.294)	0.604
Household received remittances during the 12 months prior to the survey = 1	0.483	(0.344)	0.161
Household was in the second 20% of the sample according to wealth indicators recalled from 2009 = 1	-0.149	(0.253)	0.555
Household was in the middle 20% of the sample according to wealth indicators recalled from 2009 = 1	-0.176	(0.254)	0.489
Household was in the fourth 20% of the sample according to wealth indicators recalled from 2009 = 1	0.285	(0.253)	0.260
Household was in the upper 20% of the sample according to wealth indicators recalled from 2009 = 1	0.165	(0.269)	0.540
Household had an electricity connection in 2009 = 1	0.529	(0.267)	0.048
Distance of house from nearest market place in 2009 minutes on foot (estimated by respondent)	0.001	(0.001)	0.090
Number of observations	**627**		

Notes: Probit regression. Variables dated 2009 are estimates, based on recall data or reconstructed from the composition of the household at the time of the survey. Explanatory variables expressed as $x = 1$ represent binary variables taking values of either 0 or 1. The dependent variable is 1 if the woman is a member of a self-help group supported by the project, and 0 otherwise. The coefficients represent the contribution of each explanatory variable/characteristic to the probability that a woman participates in the project.

The final set of variables used in the matching process were identified using a backwards stepwise regression to identify those variables correlated with being in an intervention group at *p*-values of 0.20 or less. Twelve such variables were identified. Table A3.2 shows the results of the probit model restricted to this final (restricted) set of matching variables.

Table A3.2: Estimating the propensity score: parsimonious model

	Coefficient	Standard error	*p*-value
Number of household members in 2009	0.163	(0.043)	0.000
Household had only one adult member in 2009 = 1	0.508	(0.267)	0.057
Respondent has any formal education = 1	-0.313	(0.230)	0.174
Household head is female = 1	0.365	(0.228)	0.110
Number of years household head spent in formal education years	0.074	(0.025)	0.003
Household was engaged in beekeeping in 2009 = 1	1.914	(0.143)	0.000
Any household member was engaged in casual labour in 2009 = 1	0.343	(0.226)	0.130
Household received remittances during the 12 months prior to the survey = 1	0.544	(0.320)	0.090
Household was in the fourth 20% of the sample according to wealth indicators recalled from 2009 = 1	0.398	(0.175)	0.023
Household was in the upper 20% of the sample according to wealth indicators recalled from 2009 = 1	0.327	(0.185)	0.078
Household had an electricity connection in 2009 = 1	0.483	(0.255)	0.058
Distance of house from nearest market place in 2009 minutes on foot (estimated by respondent)	0.001	(0.001)	0.137
Number of observations	**627**		

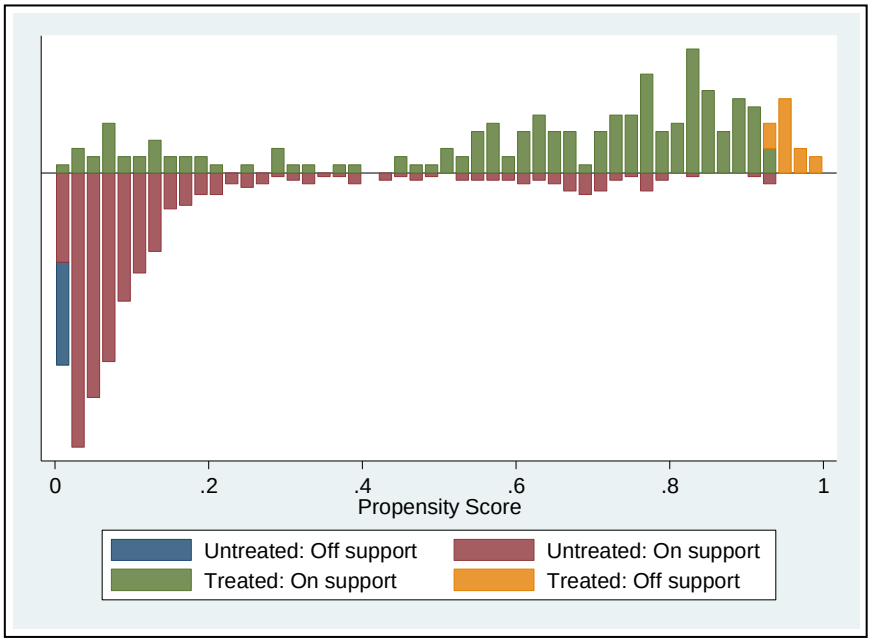
Notes: Probit regression. Variables dated 2009 are estimates, based on recall data or reconstructed from the composition of the household at the time of the survey. Explanatory variables expressed as $x = 1$ represent binary variables taking values of either 0 or 1. The dependent variable is 1 if the woman is a member of a self-help group supported by the project, and 0 otherwise. The coefficients represent the contribution of each explanatory variable/characteristic to the probability that a woman participates in the project.

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[60]). Since some significant differences were found between the intervention and comparison groups in terms of their baseline characteristics (as shown in Appendix 2), some of the women in the intervention group are 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 (Caliende and Kopeinig, 2008). In this case, 17 of the 189 self-help group members and 29 of the 438 comparison women surveyed were dropped because they lay outside the area of common support. This means 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 A3.1 illustrates the propensity scores and shows the proportion of women lying on and off the areas of common support, by treatment group. It is apparent that many of the control observations have very low propensity scores: this is not surprising given that comparison respondents were selected at random from among community members, while the treatment group were self-selected.

Figure A3.1: Propensity score on and off area of common support



Matching intervention women to comparison women

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.

We used the *psmatch2* module in Stata using a bandwidth of 0.12, and restricted the analysis on the area of common support. When using PSM, standard errors of the estimates were bootstrapped using 1000 repetitions (clustered by community), to account for the additional variation caused by the estimation of the propensity scores and the determination of the common support.[61]

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 statistically significant differences in baseline covariates between the intervention and comparison group in the matched sample. Efforts were made to ensure that the covariates were balanced across groups at *p*-values greater than 0.20. The balance of each of the matching variables after kernel matching is shown in Table A3.3. None of the variables implemented for the matching are statistically significant in the matched sample. However, an important caveat is that higher proportions of the treated group than the control group are in the top two quintiles of the baseline wealth index.

Table A3.3: Balancing test on the restricted set of matching variables

	Treated	Untreated	*p*-value
Number of household members in 2009	5.599	5.673	0.727
Household had only one adult member in 2009 = 1	0.122	0.115	0.844
Respondent has any formal education = 1	0.110	0.126	0.649
Household head is female = 1	0.174	0.167	0.862
Number of years household head spent in formal education years	1.692	1.534	0.617
Household was engaged in beekeeping in 2009 = 1	0.797	0.797	0.997
Any household member was engaged in casual labour in 2009 = 1	0.099	0.071	0.358
Household received remittances during the 12 months prior to the survey = 1	0.052	0.078	0.344
Household was in the fourth 20% of the sample according to wealth indicators recalled from 2009 = 1	0.262	0.216	0.319
Household was in the upper 20% of the sample according to wealth indicators recalled from 2009 = 1	0.343	0.293	0.318
Household had an electricity connection in 2009 = 1	0.116	0.136	0.586
Distance of house from nearest market place in 2009 minutes on foot (estimated by respondent)	89.157	81.058	0.366
Number of observations	**172**	**409**	

Notes: Variables dated 2009 are estimates, based on recall data or reconstructed from the composition of the household at the time of the survey. Explanatory variables expressed as *x* = 1 represent binary variables taking values of either 0 or 1.

Similarly, as shown in Table A3.4, we also pass the balance tests when using the full (unrestricted) set of matching variables. Only four variables in the complete set are unbalanced with *p*-values of less than 0.2. However, it is again important to note that two of those unbalanced variables (income from rent at baseline, and being in the middle quintile of the wealth index at baseline) probably imply that treated households at baseline had on average higher material well-being than did control households.

Table A3.4: Balancing test on the full set of baseline covariates

	Treated	Untreated	*p*-value
Number of household members in 2009	5.599	5.673	0.727
Proportion of household members who were children (less than 18 years old) in 2009	0.573	0.565	0.629
Proportion of adult household members in 2009 who were fit and able to work	0.969	0.956	0.430
Household had only one adult member in 2009 = 1	0.122	0.115	0.844
All adult household members were elderly (over 60 years old) in 2009 = 1	0.006	0.016	0.378
Respondent is the head of household = 1	0.169	0.163	0.884
Age of respondent in 2009 years	36.163	35.969	0.855
Respondent was elderly (over 60 years old) in 2009 = 1	0.012	0.022	0.458
Respondent is fit and able to work = 1	1.000	0.969	0.037
Respondent is widowed or divorced = 1	0.140	0.127	0.727
Respondent has any education = 1	0.291	0.367	0.132
Respondent has any formal education = 1	0.110	0.126	0.649
Number of years respondent spent in formal education years	0.750	0.729	0.927
Household head is female = 1	0.174	0.167	0.862
Age of household head in 2009 years	44.360	44.064	0.813
Household head is fit and able to work = 1	0.959	0.956	0.906
Household head has any education = 1	0.587	0.645	0.272
Household head has any formal education = 1	0.267	0.292	0.620
Number of years household head spent in formal education Years	1.692	1.534	0.617
Household was engaged in beekeeping in 2009 = 1	0.797	0.797	0.997
Household cultivated any land in 2009 = 1	0.895	0.849	0.201
Land area cultivated by household in 2009 *qodda*	4.907	4.506	0.269
Household owned any livestock in 2009 = 1	0.890	0.856	0.359
Any household member was engaged in casual labour in 2009 = 1	0.099	0.071	0.358
Any household member was engaged in a non-agricultural household business in 2009 = 1	0.308	0.256	0.283
Any household member was engaged in formal employment in 2009 = 1	0.017	0.013	0.717
Any household member received income from renting out land or property in 2009 = 1	0.070	0.023	0.039
Household received remittances during the 12 months prior to the survey = 1	0.052	0.078	0.344
Household was in the second 20% of the sample according to wealth indicators recalled from 2009 = 1	0.116	0.114	0.948
Household was in the middle 20% of the sample according to wealth indicators recalled from 2009 = 1	0.151	0.235	0.049
Household was in the fourth 20% of the sample according to wealth indicators recalled from 2009 = 1	0.262	0.216	0.319
Household was in the upper 20% of the sample according to wealth indicators recalled from 2009 = 1	0.343	0.293	0.318
Household had an electricity connection in 2009 = 1	0.116	0.136	0.586
Distance of house from nearest market place in 2009 minutes on foot (estimated by respondent)	89.157	81.058	0.366
Number of observations	**172**	**409**	

NOTES

[1] Results are summarised in 'Involving marginalized women in collective action – Making a difference through NGO interventions', Oxfam, February 2013: http://policy-practice.oxfam.org.uk/publications/womens-collective-action-in-the-honey-sector-in-ethiopia-275773. More details on the research design and results can be found at http://womenscollectiveaction.com/.

[2] This approach has been taken to the evaluation of the newer honey value-chain project that is being implemented by Oxfam, Zembaba Union and Facilitator for Change in the three woredas that were excluded from the Effectiveness Review.

[3] The construction of the wealth index is discussed in Section 5.5.

[4] For example, only 37 of the SHG members interviewed were living in female-headed households, of whom seven were dropped from the figures presented in this section because they were outside the area of common support for the PSM model.

[5] This analysis was done by adding terms for the interaction between being an SHG member and residing in each of the four woredas, or for being an SHG member and from a female-headed household, to the parametric PSM and linear regression models for each of the outcome measures examined in this report.

[6] It will be recalled from Section 4.1 that 10 per cent of those identified as members of the SHG groups were not interviewed because their houses were located far from the community centre and could not be reached within the logistic constraints of the survey. It would be natural to assume that those who live at such far remove from community centre tend to participate less in SHG meetings – so the figures for participation reported here are probably overestimate the level of participation among the membership as a whole.

[7] Indeed, a quarter of SHG members said that they had participated in a cooperative meeting every week during the past month – a figure which seems unrealistically high, and may reflect some confusion on the part of respondents.

[8] As a check on the robustness of these results, there was no difference found between the SHG members and comparison respondents in terms of receipt of training of types not carried out as part of the project, including literacy classes or training on family planning. There was some evidence (at the 10 per cent significance level) that comparison households had received more training in health issues than the households of SHG members.

[9] Respondents were also asked about whether they had received donations of agricultural inputs during the 12 months prior to the survey: as expected, only a very small proportion (two per cent) responded positively, with no indication of a difference between households of SHG members and those of comparison respondents.

[10] After adding an interaction term for being a member of a SHG in Bahir Dar Zuria woreda to the parametric PSM and linear regression models for this outcome, the estimated coefficient on the interaction term is positive and statistically significant at at least the 10 per cent level. This does not apply when adding interaction terms for membership of an SHG in each of the other three woredas.

[11] Again, this analysis was performed by adding an interaction term for being a member of a SHG in Bahir Dar Zuria woreda to the parametric PSM and linear regression models for the relevant outcome variable. The estimates of the coefficient on the interaction term are all positive and statistically significant at the one per cent level, while the coefficient on the dummy variable for being an SHG member is estimated to be small and not statistically significant.

It should be noted that there were no significant differences between the households of SHG members and comparison respondents in their contacts with natural resource conversation or health extension services. Again, since contacts with these other extension services would not be expected to be affected by the project, this adds confidence that the significant differences that were found do represent an effect of the project activities.

[12] A related measure, the proportion of respondents that said that they had *ever* met a representative of a microfinance institution, showed a difference between SHG members and comparison respondents that was positive and statistically significant at at least the 10 per cent level under most of the statistical models tested, though not under the standard PSM kernel model for which results are presented in the tables in this section.

[13] It will be recalled from Section 4.2 that a much larger proportion of SHG members than comparison respondents recalled that their households were engaged in beekeeping in 2009. However, this baseline engagement in beekeeping is one of the factors that has been controlled for in the propensity-score matching (PSM) models. It is shown in Appendix 1 that, after applying PSM, the two groups are very closely balanced in this respect, with 80 per cent of each group reporting that they were engaged in beekeeping in 2009. We can have some confidence, therefore, that the difference shown in column 1 of Table 5.2 does represent an effect of the project.

[14] This was tested by adding an interaction term for residing in a female-headed household and being a member of a SHG to the parametric PSM and linear regression models for this outcome variable. The coefficient on the interaction term was estimated to be positive and statistically significant at at least the 10 per cent level – though the coefficient on the dummy variable for participation in a SHG remained positive and statistically significant at the one per cent level, confirming that there is a difference in adoption of beekeeping among male-headed households as well. The same applies to the outcome

variables for the household having produced any honey and sold any honey in the 12 months prior to the survey (examined in Tables 5.5 and 5.6 respectively).

[15] The results shown in column 4 of Table 5.2 make use of data on the households' situations in 2009. This was collected during the same interview as the rest of the data, but respondents were asked to recall their situation in 2009, using locally-appropriate landmark events from around that time. Knowing that the accuracy of the recalled 2009 data may vary, outcome measures based on data recalled from 2009 are always presented in this report alongside alternative estimates that do not make use of this recalled information.

[16] Unfortunately the length of time each respondent had been an SHG member was not recorded in the survey, so differences in outcomes by length of participation could not be tested.

[17] The estimates of the difference derived from the various statistical models are all positive, but none of them are statistically significant at conventional levels.

[18] It should be noted that this result is only marginally statistically significant, and that the estimates derived from alternative models are not all positive in sign. However, these estimates are based on the small number of comparison households who reported having sold honey during the 12-month period. A more sensitive indicator is provided by the revenue generated from sales, in column 4.

[19] The estimated difference is negative under three of the eight statistical models tested, and only positive and statistically significant at the 10 per cent level under two of them.

[20] It should be recalled from Section 4.2 and Appendix 3 that the SHG members tended to be wealthier than the comparison respondents at baseline (according to the recalled data from 2009) even after matching. However, it seems very unlikely that the baseline imbalance is large enough to fully account for the large differences in savings observed in columns 4 and 5 of Table 5.7.

[21] The coefficients on an interaction term for being a member of a SHG in Gonder Zuria woreda added to the parametric PSM and linear regression models for this outcome are estimated to be positive and statistically significant at the 5 per cent level.

[22] Although, as noted above, SHG members in Gonder Zuria were particularly likely to cite the self-help group as a source of credit, they appear to have been just as likely to actually borrow from the SHG as those in other woredas. On the other hand, SHG members in Gozamen woreda were *less* likely than those in the other woredas to have borrowed from the group.

[23] The estimate is comparable in size and statistically significant at at least the 10 per cent level under each of the estimation methods applied except for the PSM nearest-neighbour no-replacement model and the probit regression models.

[24] Although this difference is large in size it is not statistically significant under most of the estimation methods used, even at the 10 per cent level.

[25] Detailed questions on agricultural production and livestock management were not included in the survey, so it cannot be assessed how productive these investments were. However, it can be confirmed from the data available survey data that there were no significant differences between the households of SHG members and comparison respondents in terms of the land area cultivated or in terms of the number of cattle, sheep, goats or poultry owned between 2009 and the time of the survey.

[26] Jennifer Coates, Anne Swindale and Paula Bilinksy, *Household Food Insecurity Access Scale (HFIAS) for Measurement of Food Access: Indicator Guide*, version 3, Food and Nutrition Technical Assistance Project (FANTA), August 2007: http://www.fantaproject.org/monitoring-and-evaluation/household-food-insecurity-access-scale-hfias

[27] The differences in proportions suffering from severe food insecurity for both the female respondent and male household members are consistently estimated to be negative, using each of the different estimation techniques. A parametric PSM model for the difference among respondents produces an estimate that is statistically significant at the 10 per cent confidence level – but those generated by probit regression are not statistically significant.

[28] The coefficient of the interaction term for being a SHG member and in a female-headed household, when added to the parametric PSM and linear regression models for this outcome, were all negative and statistically significant at at least the 5 per cent level.

[29] This approach is adapted from the Household Dietary Diversity Score approach, described in Anne Swindale and Paula Bilinsky, *Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide*, version 2, Food and Nutrition Technical Assistance Project (FANTA), September 2006: http://www.fantaproject.org/monitoring-and-evaluation/household-dietary-diversity-score. The difference with the HDDS approach was that recall was requested over a seven-day period, rather than over the past 24 hours.

[30] However, since various religious celebrations and festivals occurred during the survey period, on which people usually fast, respondents were asked about a typical seven-day period, rather than necessarily the seven days prior to the survey.

[31] On the other hand, there is no indication of a difference in dietary diversity between SHG members and comparison respondents specifically in female-headed households.

[32] Indeed, there is a strong correlation between dietary diversity score and the index of wealth indicators in the dataset for this Effectiveness Review.

[33] Cronbach's alpha was used to measure this inter-item correlation. The Cronbach's alpha obtained for all the indicators for the recalled 2009 data was 0.72. This alpha was increased to 0.77 by removing those items that were negatively correlated or had a low correlation with the others. The alpha derived for the index of change in wealth indicators was originally 0.73, and was increased to 0.78 by removing those items that were negatively correlated or had a low correlation with the others.

[34] Information is available at http://www.ifpri.org/publication/womens-empowerment-agriculture-index.

[35] Sabina Alkire and James Foster, Counting and multidimensional poverty measurement, *Journal of Public Economics* vol. 95 (2011), pp. 476–487: http://www.sciencedirect.com/science/article/pii/S0047272710001660

[36] It will be noted that in calculating these overall measures of women's empowerment, each of the individual characteristics presented in Figure 5.3 was weighted equally. This means that the index is weighted more towards characteristics of dimensions 'Women's ability to make decisions and influence' and 'Women's access to and control over resources', and less towards the other three dimensions. Alternative weights could be given to the various characteristics and dimensions, which would necessarily result in changes in the overall indices and potentially in the magnitude of differences between the intervention and comparison groups.

[37] The coefficient on an interaction term for being a SHG member in Libokemkem, when added to the parametric PSM and linear regression models for the base empowerment index, is consistently estimated to be negative and statistically significant at at least the 5 per cent level. The coefficients on the equivalent interaction terms for each of the other woredas and for being in a female-headed household are not found to be statistically significant. Again it should be noted that the number of project participants in each woreda, and the number of project participants in female-headed households, were small, so the potential for detecting significant differences between these subgroups is limited.

[38] A fifth statement on this subject was included in the questionnaire, but responses were found to be uncorrelated with the other four, so the results were not used in this analysis.

[39] A continuous measure of community influencing, constructed using principal-component analysis with the results of these four statements, also showed no significant difference between the SHG members and comparison respondents, but the difference was consistently estimated to be negative.

[40] The results derived from a fourth statement relating to self-confidence ('I am capable of overcoming challenges and achieving my goals') were found not to be correlated with the results from the other three statements, so were excluded from this analysis.

[41] An alternative measure, constructed from the results of the three statements using principal-component analysis, also did not show any significant difference between the SHG members and comparison respondents.

[42] In both cases the estimated coefficients on the interaction variables associated with residing in the relevant woreda and being a SHG member are statistically significant at at least the 10 per cent level, both for the binary variable described here and for an alternative continuous variable, generated through principle component analysis.

[43] The responses to a further two statements presented in the questionnaire were found not to be correlated with these three, and so were excluded from the analysis.

[44] The questionnaire also included one statement asking respondents whether they agreed or disagreed that 'If a man divorces his wife, it is her right to take half of his land'. It is not clear whether this statement was interpreted as a test of respondents' knowledge of Ethiopian law (where it is in fact a wife's right to half her husband's land on divorce), or whether if generally elicited respondents' opinions about the rightfulness of that law. In any case, 96 per cent of respondents agreed with this statement (and 82 per cent agreed strongly), so this was not used in the analysis.

[45] Estimates derived from a corresponding measure constructed by principal-component analysis from the responses of the three statements were not consistently positive.

[46] Again, the responses to a fourth statement that were not positively correlated with the others were excluded from the analysis.

[47] Again, this is corroborated by considering an alternative measure, constructed by principal-component analysis from the responses to the same three statements.

[48] In the small number (3 per cent) of cases where the respondent said that one of these two types of decision was not relevant to her, the determination of the indicator was based solely on her influence in the other decision-making area. Three respondents said that neither of these two types of decision was relevant to her: they were deemed to score negatively on this characteristic.

[49] The estimate is significant at the 5 per cent level under the PSM kernel model, and at at least the 10 per cent level under the various regression models tested (though not quite significant under the PSM nearest-neighbour model). An alternative measure, constructed from the same responses using principal component analysis, produces estimates that are positive, but not statistically significant under most of the models tested.

[50] There is evidence (statistically significant at the 10 per cent level under the PSM kernel model and some of the other estimation methods) that SHG members spend a small proportion of their day collecting water or firewood than do the comparison women. However, this appears to be related to the fact that SHG members were generally living closer to their water source (as discussed in the following paragraph), and is unlikely to be linked to the project activities.

[51] The difference for children's leisure time is statistically significant at the 10 or 5 per cent level under each of the estimation methods. The difference for men's leisure time is statistically significant at the 10 per cent level only under some of the estimation methods.

[52] The sign of the estimate is corroborated by the alternative statistical models used, but its statistical significance is not.

[53] No further details were asked about the specific situation, the frequency of its occurrence, or the identity of the victim or perpetrator.

[54] In fact, a majority of respondents stated that they contribute exactly 50 per cent to household income.

[55] When examining a continuous measure of the number of asset types over which a respondent has sole decision-making control, the differences between the SHG members and comparison respondents are consistently positive, but are not consistently statistically significant. There is no indication of a difference between the groups in the number of assets over which respondents have joint decision-making control.

[56] Once again, one statement on the same theme was excluded from the analysis because the responses were not correlated with these four.

[57] In the case of kebele meetings, it is not clear that this represents a statistically significant change in the proportion participating since before the project started.

[58] Marco Caliendo and Sabine Kopeinig 'Some Practical Guidance for the Implementation of Propensity Score Matching', *Journal of Economic Surveys*, vol. 22(1) (2008), pages 31–72.

[59] Paul R. Rosenbaum and Donald B. Rubin, 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika*, vol. 70(1) (1983), pages 41–55.

[60] James J. Heckman, Robert J. Lalonde and Jeffrey A. Smith, 'The Economics and Econometrics of Active Labor Market Programs', *Handbook of Labor Economics*, vol. 3, part A (1999), pages 1865–2097.

[61] 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.

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 (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.