Universidad de San Andrés

Departamento de Economía

Maestría en Economía

What Works to Strengthen Beekeeper Productivity and Climate Resilience? Evidence from México

Victoria LUCA

38.927.549

Mentor: Gabriela APARICIO

Buenos Aires

3 de febrero, 2024

Tesis de Maestría en Economía de

Victoria LUCA

"¿Qué funciona para reforzar la productividad y la resiliencia climática de los apicultores? Evidencia de México"

<u>Resumen</u>

Este trabajo estudia la efectividad de un programa de asistencia técnica destinado a mejorar la productividad y la resiliencia climática de apicultores en zonas rurales de México. El programa se ofreció de forma gratuita a los apicultores que perdieron colmenas durante la tormenta tropical Cristóbal en 2020. Los participantes del programa en 14 localidades de Yucatán tuvieron acceso a cuatro componentes principales: (i) insumos básicos, (ii) una capacitación de un año sobre manejo de colmenas, (iii) abejas reinas y (iv) una capacitación específica sobre cría de reinas. Debido a las dificultades para implementar el programa durante la pandemia de COVID-19, los participantes en otras 3 localidades cercanas sólo recibieron insumos básicos y sirven como grupo de control. Utilizamos un enfoque de diferencia en diferencias combinado con entropy balancing *para controlar por las diferencias preexistentes entre los grupos. Aprovechamos una encuesta detallada de línea base y de seguimiento de 356 apicultores que participaron en el programa. Los apicultores tratados aumentaron significativamente el número de colmenas, la producción de miel, los rendimientos y la adopción de mejores prácticas en comparación con el grupo de control, que solo recibió insumos básicos. Si bien la provisión de abejas reinas tuvo el mayor impacto sobre la producción y el rendimiento, las capacitaciones fueron especialmente importantes para la adopción de mejores prácticas, lo que es un buen augurio para la resiliencia climática a largo plazo.*

<u>Palabras clave</u>: agricultura, asistencia técnica, evaluación de impacto, capacitación, rendimiento, apicultores, abeja reina

"What Works to Strengthen Beekeeper Productivity and Climate Resilience? Evidence from Mexico"

Abstract

This paper studies the effectiveness of a technical assistance program aimed at improving the productivity and the climate resilience of beekeepers in rural Mexico. The program was provided free-of-charge to beekeepers who lost hives during tropical storm Cristobal in 2020. Program participants in 14 localities of Yucatan were offered four main program components: (i) basic inputs, (ii) a year-long training on beehive management, (iii) queen bees, and (iv) a specific training on queen breeding. Due to the difficulties of implementing the program during the COVID-19 Pandemic, participants in 3 other nearby localities were offered only basic inputs and serve as a control group. We use a difference-in-differences approach combined with entropy balance to control for pre-existing differences across groups. We take advantage of a detailed baseline and follow-up survey of 356 beekeepers who participated in the program. Treated beekeepers increased their number of hives, honey production, yields, and adoption of best practices substantially more than the control group, who only received basic inputs. While the provision of the queen bee asset had the largest impact on honey production and yield, the trainings were particularly important for the adoption of best practices, which bodes well for longer-term climate resilience.

Keywords: agriculture, technical assistance, impact evaluation, training, yield, beekeepers, queen bee

Códigos JEL: O12, O13, Q12, Q14, Q16, Q54

1. Introduction

Beekeeping is of high economic and social importance for Latin American and Caribbean (LAC) countries. As of 2021, LAC countries accounted for approximately 14% of the global honey production and 8% of the global bee livestock (Food and Agriculture Organization of the United Nations, 2023). Seven countries in the region were amongst the 20 leading countries in the world in terms of honey production and beehive stocks. Moreover, Argentina, México, and Brazil were amongst the top honey producers and exporters in the world. Honey production also provides an important livelihood opportunity for low-income populations across LAC. For example, in Mexico, honey production is concentrated in the Southeast of the country, which has lagged in terms of development and where alternative employment opportunities are limited. Paradoxically, LAC countries have been leaders in honey production despite poor colony management (Galetto et al., 2022). However, with growing threats to bee colonies, such as climate change and increased pesticide use, the adoption of improved beehive management practices is increasingly important.

Although many resources are invested in technical assistance for agriculture each year, several questions remain. There is little knowledge about how to design programs that are effective to diffuse knowledge, boost adoption of improved practices, and ultimately improve the long-term outcomes of agricultural producers in a financially sustainable manner. This paper studies the effectiveness of different components of a privately provided technical assistance program aimed at improving the productivity and climate resilience of beekeepers in rural Mexico. The year-long program was provided free-of-charge to beekeepers who lost hives during the tropical storm "Cristobal" that hit Southeastern Mexico in 2020, to help them to accelerate their productive recovery and improve their climate resilience. Program design considered best practices according to the available literature, such as defining the content bottom-up, providing training in the local language, and timing the training topics and input provision to coincide with beekeepers' needs during the agricultural calendar.

We took advantage of a detailed baseline and follow-up survey of 356 beekeepers who participated in the program. The program was highly comprehensive and substantial, allowing us to estimate its effects with a small sample. Program participants in 14 localities of Yucatan were offered four main components: (i) basic inputs, (ii) a year-long training on beehive management, (iii) queen bees for the genetic improvement of their bee colony, and (iv) a specific training on queen breeding. Instead, due to the difficulties of implementing the program during the COVID-19 Pandemic, participants in 3 other nearby localities were offered only basic inputs (such as sugar for feeding and inputs for plague control) and serve as a control group. For our analysis, we explore the difference in outcomes between treated producers (receiving two or more program components) relative to the control group (producers receiving only basic inputs). To control for pre-existing differences across groups, we combine a difference-in-differences approach with entropy balance. We also assess the robustness of our results using different matchings techniques, such as inverse probability weighting and propensity score matching.

Results show that the program was effective. Treated beekeepers improved their productive performance —number of hives, honey production, and yields— and adoption of best practices substantially more than the control group. These results are statistically significant at conventional levels and remain qualitatively unchanged regardless of the approach used to balance covariates between groups. While the different project components appear to have contributed to the effectiveness of the program, the provision of the queen bee asset appears to be an important driver of the impact on productive performance outcomes. Queen bees not only improved the genetic quality of the bee colony, which is important for honey production, but also allowed beekeepers to adopt good practices related to queen bee management, which were further improved by the provision of the queen breeding training. The breeding training contributed to the adoption of good hive management practices, which are particularly important for long-term resilience and producers' adaptation to climate change.

This paper contributes to various strands of literature. First, there is a growing, but still limited, body of evidence suggesting that bottom-up trainings for agricultural producers spur increased knowledge among program participants; however, it remains important to understand the external validity of these interventions in different countries and settings. Evidence has been mixed regarding whether increased knowledge can translate into adoption of best practices and ultimately into improved long-term outcomes. Moreover, there has been little attention concerning the joint adoption of more complex agricultural management systems, as opposed to the uptake of individual practices in an isolated manner. Our paper suggests that a comprehensive technical assistance program (including training on multiple key practices combined with the provision of key assets) can improve long-term outcomes. Lastly, despite the importance of bees —both for honey production and for the provision of ecosystem services though pollination— ours is one of the few evaluations to study the impact of training programs in this context.

2. Literature review

Given that information failures have been identified as important barriers affecting agricultural producers, it is important to understand whether training programs can succeed at disseminating information and improving knowledge (Kondylis et al., 2017). While traditional top-down public extension services have been subject to criticism, there is now growing evidence that more modern agricultural trainings, such as bottom-up or decentralized models, can spur knowledge diffusion and technology adoption (see Hörner et al., 2021; and the meta-evaluation of interventions in agribusiness Nankhuni & Paniagua, 2013). However, given the limited number of studies available, it remains important to evaluate additional bottom-up programs, such as the one presented in this paper, and learn whether the specific context of the intervention may affect the results (external validity).

Even when training can improve more immediate outcomes such as knowledge, several factors may hamper converting knowledge into changes in behavior, and ultimately into improved long-term outcomes like increased agricultural yield, production, farm income, assets, and/or profits (see Baul et al., 2023; DellaVigna, 2009; and Ravallion et al., 2015 for an example outside of agriculture). Several papers do find that more modern training can have positive impacts, not only on knowledge, but also on the adoption of practices and technologies (Hörner et al., 2021; Nankhuni & Paniagua, 2013). For example, a three-day hands-on training on organic farming increased the use of organic inputs (Grimm & Luck, 2020). Yet, evidence of positive impacts of training on long-term outcomes is still limited: the few papers available are in contexts other than agriculture (Blattman et al., 2014; Ruprah & Marcano, 2009); they do not account for self-selection into treatment (Danida, 2004); or are still scarce with few examples available (Baul et al., 2023; Wonde et al., 2022). Instead, it is more common for the impacts of training on long-term outcomes to be mixed, with positive findings only for some groups of beneficiaries (Kondylis et al., 2017; Nankhuni & Paniagua, 2013); or to be imprecisely estimated and not statistically significant, both in the context of agribusiness (Torres Franco et al., 2021) as well as in other contexts (McKenzie & Woodruff, 2014).

The lack of robust positive findings regarding the impacts of training on long-term outcomes may be due to the specific characteristics of the programs evaluated. Different program characteristics can influence the outcomes, such as the framing and/or timing of the information delivered, the credibility of whether a new technique will actually work, as well as farmers' belief in their capacity to successfully implement the practices promoted (Baul et al., 2023; Gignoux et al., 2023). For example, farmers are more likely to purchase inputs that are advertised at the time of harvest, when farmers are cash-rich, as opposed to other periods of the year (Duflo et al., 2011). Other aspects also matter, such as training duration and content (depth and breadth of topics covered). Few papers evaluate comprehensive training programs for the adoption of a set of interrelated practices or a more complex management system, such as ours, as opposed to promoting the uptake of individual practices in an isolated manner (one exception is Hörner et al., 2021).

Mixed results are also common when reviewing interventions providing access to inputs or capital assets. For example, while some studies find a positive effect of inputs or investment assets on productivity and welfare outcomes (Abman & Carney, 2020; M. Carter et al., 2021; M. R. Carter et al., 2013; Hemming et al., 2018; Wossen et al., 2017); other papers find smaller-than-expected effects on income and no effect on poverty incidence or severity (Jayne et al., 2018; Mason & Tembo, 2015; Ricker-Gilbert & Jayne, 2017). Gignoux et al. (2023) find that the provision of inputs or assets may even result in negative effects on input use and yield due to incorrect expectations of future transfers. There may be different explanations for these mixed results, such as the importance of the assets provided. For instance, key capital assets are likely to have larger impacts than basic inputs, as appears to be the case in our program.

Another important question is whether, rather than providing either inputs, assets, or training in an isolated manner, they should be provided jointly. Although evidence is quite limited, there is some indication that these "packages" are more effective than isolated interventions. For example, a meta-evaluation of private-sector interventions in agribusiness from Nankhuni & Paniagua (2013) finds that the more successful programs generally target all stages of the value chain, from training in good farming practices to training in post-harvest techniques, plus providing inputs such as credit and facilitating farmers' organization to help them obtain better prices from suppliers. Similarly, interventions facilitating access to credit for agricultural producers were found to be successful mainly when accompanied with training, technical advice, or other kinds of help. In a different context, Banerjee et al. (2022) find that a multi-faceted anti-poverty program, comprising the provision of productive assets in addition to training/coaching, leads to improved welfare effects relative to only a grant of productive assets. This may be the case because training enables households to become more successful in building businesses that persistently generate income, allowing them to accumulate more assets. Yet, we are not aware of any papers in the context of agriculture that can estimate the marginal impact of the different components in a complete "package" of support. Our paper aims to fill this gap.

Another important question is whether the types of interventions discussed above can be financially sustainable. One way to finance technical assistance may be for extension services to be provided by the private sector through anchor companies, taking advantage of arrangements where the anchor company and the agricultural suppliers in the company's value chain share the benefits (Nankhuni & Paniagua, 2013). Such arrangements could be used in cases where the training is linked to market transactions (like the provision of inputs or integration into a value-chain), as is the case for the program evaluated in this paper. Working with anchor companies also makes it more likely that training will be relevant to producers' needs and that there will be a market for producers' increased agricultural production. However, while the private sector can have a critical role in the provision of agricultural training, more research is needed to help design private interventions that are not only effective but also financially sustainable. Our paper aims to assess not only the effectiveness of different components of the programs, but also their cost-effectiveness.

Lastly, few papers study the effectiveness of interventions specifically in the context of beekeeping and most of them fail to control for self-selection into treatment (see for example Ahmad et al., 2017; Schouten, 2020; Schouten & John Lloyd, 2019; Woldewahid et al., 2012). This lack of evidence is surprising given the importance of apiculture worldwide. According to the Food and Agriculture Organization of the United Nations (2018), bees and other pollinators contribute to 35% of the world's total crop production, pollinating 87 of 115 leading food crops worldwide. Lack of improved management practices may be preventing beekeepers from realizing the full value of their enterprises, and providing technical assistance may be of great importance for increasing bee colony resilience and productivity, as well as to increase the ecosystem services that they provide, in an environmentally sustainable manner (Seagle, 2008). While improving management practices amongst beekeepers shares many of the challenges of other agricultural contexts (and training more broadly), there may be aspects specific to this activity, suggesting the need for more specific research. Our paper fills this gap by providing more rigorous evidence on a specific intervention to improve beekeepers' performance.

3. *Program description*

We study a pilot program financed by IDB Invest jointly with a leading private-sector Mexican exporter of honey (referred in this paper as "the Company"). The Company is a Mexican family-owned business engaged in the development, production, and sale of food (including not only honey but also nuts, grains, and spreads). At the time the program was designed, the Company had a large supply-chain with over 16,000 honey suppliers, 95% of which were micro, small, and medium sized agricultural producers. The Company had over 60 bulking centers throughout Mexico where honey was collected. The bulking centers determined which honey suppliers were eligible to participate in the program based on their location and whether they had supplied honey to the Company in the past. The Program was implemented by a consulting firm specialized in beehive management. Implementation was expected to take place between July 2021 and July 2022 to cover a full production cycle; however, due to delays in the provision of assets, it lasted until the beginning of 2023 (see Figure 1 for a full timeline of activities).

The program was implemented in 17 localities of the State of Yucatán (Southeastern Mexico). Yucatán was targeted for the pilot program due to its importance in Mexico's honey production (it produces approximately 15% of Mexico's honey on average) and because it was affected by recent weather shocks leading to substantial losses to beekeepers. Specifically, in June 2020, Yucatán's honey producers suffered substantial flooding derived from the tropical storm "Cristóbal". The floods were followed by rains and unusual temperatures that affected the flowering season, substantially contracting the state's honey production during that year (see Figure 2). Yucatán had also experienced severe climate shocks in the past, such as droughts during the 2017 season. This situation affected Yucatán's entire honey supply chain, but mainly small beekeepers, who faced serious economic difficulties due to the damage to their crops and scarce resources to rebuild their apiaries and recover lost hives.

The vertical logic of the program is presented in Figure 3. The program was highly comprehensive compared to other agricultural interventions (for comparison, see Table 1 for a summary of the magnitude and effects of other agricultural interventions). As a large majority of producers had not received previous technical assistance or training, the program was designed to include training components to increase beekeepers' knowledge and adoption of best practices in hive management. In addition, as many beekeepers had equipment with severe damage or deterioration and high loss of hives due to the climatic events, the program included components to help them restore some of the assets lost and to improve their production processes. Overall, the program had four main components, two of which focused on training, while the remaining two focused on the provision of assets or inputs. The four main program components are discussed below:

i. The first component was a grant of **basic inputs** such as inputs for feeding the bee colonies and plague control, as well as some field equipment (supers, breeding chambers, pollen and propolis traps). Some of the field equipment was delivered after the follow-up survey, and hence is not captured by this evaluation.

ii. The second component was the provision of **general training** on apiculture and beehive management. The training comprised the year-long production cycle and included group theoretical-practical workshops with 6 to 10 producers in each; as well as individual follow-up visits in apiaries, focused on increasing beekeepers' adaptation capacity, and developing knowledge and skills to improve beehive management. The content was defined bottom-up rather than top-down (based on the needs identified from a baseline survey) and focused on practical applications. Given that local languages are still common where the program was implemented, the training was provided through a "train the trainer" model, where 6 local Mayan-speaking beekeepers were trained as technicians. These technicians then trained the beekeepers who were selected as program beneficiaries.

iii. The third component was a grant of **queen bees** with European genetics, for the genetic improvement of the bee colonies. European bees are generally less aggressive than Africanized bees, which are prevalent in Mexico, have a lower tendency to swarm,[1] and also generate lower management costs and higher honey production (Gobierno de México, n.d.; Guzmán-Novoa et al., 2011). In total, as many as 1,894 fertilized bees were provided prior to the follow-up survey, and an additional grant of 1,787 queen bees was delivered after the follow-up survey, and hence are not captured by this evaluation.

iv. Finally, the fourth component was a **breeding training** to teach producers how to breed and raise queen bees adapted to the conditions of the area, and to select them for their behavioral characteristics. This training was provided only to a subgroup of the producers that received the queen bees (third component). Training participants were required to have good eyesight to be able to breed the queen bees.[2] Producers were expected to participate in two workshops; the first was carried out between the months of December 2021-January 2022 and the second one between May 2022-June 2022. A total of 30 theoretical-practical breeding workshops were carried out in the different localities, with 6 to 8 producers each. As a result of the workshops, producers bred 613 new queen bees with European genetics.

The combination of inputs and a comprehensive training program was expected not only to increase beekeepers' productive capacity but to also improve their climate resilience, understood as their ability to recover from, or reduce vulnerability to, climate shocks. While the provision of inputs was envisioned to help producers recover from past climatic events, the trainings for the adoption of best practices in hive management and queen bee breeding were expected to build their long-term resilience to future shocks, helping them improve their adaptive capacity by correctly identify flooding risk areas, improving knowledge of the potential effects of climate change, and preparing action protocols in the presence of adverse natural phenomena.

4. Data and Treatment Groups

The data used in this paper was obtained from two main sources: first, a baseline and a follow-up survey to program participants; and second, administrative data from program implementation.

From 16,000 honey suppliers of the Company in Mexico, a total of 431 beekeepers participated in the program, receiving at least some basic inputs. However, out of the 431 beneficiaries, only 356 producers had both a baseline and a follow-up survey. This is because some beneficiaries that started the program were then removed at the request of the bulking centers, as it was verified that they had not sold honey to the Company in the past. Other beneficiaries were included after the program had already started, to replace those that were removed, and they lacked a baseline survey.

The baseline survey was carried out before the start of program implementation, between the months of March and April of 2021. Overall, 424 producers were included in this survey. The follow-up survey was carried out during July 2022, after most program components were delivered,[3] and included only

[1] The swarming process consists of the abandonment of a part of the bee colony with a new queen, which results in the weakening of the remaining partial colony.
[2] According to the consulting firm who implemented the project, older producers with poor eyesight sometimes brought along younger family members to the breeding trainings to be able to participate.
[3] A final basic inputs delivery was carried out in April 2023, which included capital goods for 247 producers and queen bees for 295 producers. These inputs were not considered for our study since they were delivered after the follow-up survey.

the 356 producers that remained in the program since the start. Our analysis focuses on this group of 356 beekeepers. While our sample size may seem small, it is in line with that of some other related papers (see Table 1); and the program evaluated here is more substantial, requiring a smaller sample size to detect its effects. There was no attrition in data collection amongst those that met the requirements for program participation (according to the criteria set by the bulking centers). It is likely that producers were willing to participate in the surveys because they had received inputs and training for free.

The survey questionnaires provide very detailed information on producer characteristics and performance, as well as retrospective questions to measure hive stock and honey production as far back as 2016 for some producers. They also include comprehensive questions on hive management practices. See Annex 2 for the full questionnaire.

We also had access to administrative data from program implementation, such as the duration and topics of the workshops offered in each locality, number of workshops attended by each beneficiary, number of follow-ups received, etc. During the visits to individual apiaries, the technicians also recorded their observations about the number of hives available. This administrative data was used to confirm that the self-reported values of the number of hives were accurate at the time of the baseline and the follow-up survey. This helps us to corroborate that our results are likely not driven solely by improvements in reporting (such better data due to more use of beekeeping logs).

4.1. Descriptive statistics

Table 2 presents producer-level descriptive statistics for the full sample. The variables reported include producer characteristics, program interventions received, and the key outcome variables that will be studied in the main sections of the paper: number of hives, amount of honey produced and yield (amount of honey per hive), as well as hive management practices (a variable ranging from 0 to 9 that indicates how many key practices producers adopt, and individual dummies for each practice: annual change of queen bees, breeding of queen bees, keeping a record log of activities, supplemented feeding, sustainable pest control, condition of brood chambers, condition of honey suppers, condition of outer covers, condition of brood frames).

Out of the 356 producers in our sample, 71 (20%) are women, 117 (33%) have post-primary education and the median age is 43 years old. These demographic variables are roughly in line with data presented by other studies on apiculture, such as Contreras-Uc et al. (2018) for Mexico or Schouten (2020) for 19 low or middle-income countries, describing beekeepers as predominantly male and middle-aged, while levels of education vary considerably between countries, but seems to be limited in Mexico. Moreover, producers in our sample had an average of 22 hives before program implementation, and the amount of honey produced was approximately 22.6 kg per hive, below the 36.9 kg per hive presented in Contreras-Uc et al. (2018) for Mexican producers, which is reasonable given that producers in our sample had recently been affected by floods.

Table 2 also presents descriptive statistics for key outcome variables in 2020 (baseline survey) and 2022 (follow up survey). As expected, the comparison of the key outcome variables shows a marked improvement over time, both in terms of performance and adoption of key practices, which will be explored in more detail in later sections.

4.2. Treatment groups

The 356 beekeepers in our study sample can be divided into a control group, who were offered only basic inputs, and a treatment group, who were offered the four main program components discussed earlier. We further divide the treatment group into sub-groups depending on the take-up of different program components.[4] We describe each of these groups below, and Table 3 shows the mean and total amount of each program component received by group.

Control Group (C): Includes 21 producers that were offered basic inputs only. The beekeepers in the control group come from 3 localities in the western part of Yucatán, whereas the Treatment Group comes from 14 different localities in the southeast of Yucatán (Figure 4 shows the geographic location of the treatment and control groups). Initially all program participants were expected to be offered all program components; however, those in the West were excluded from receiving some components due to their distance from the other producers and the logistical difficulties of reaching them during the COVID-19 pandemic, as well as the fact that, according to the baseline survey, they had a slightly better level of hive management practices. Although not as good as random assignment, the Control Group is considered adequate because all the 17 localities were initially targeted for receiving the full program, they are not that far from each other, and the differences in characteristics among them is not that large (and can be controlled for); moreover, there was no self-selection into treatment and control (and consequently there is less concern about unobservable differences across groups).

Treatment Group (T): Includes 335 producers that were offered not only basic inputs, but also other program components. On average, each producer in the sample received inputs worth USD 395; however, the treatment group received slightly more inputs (USD 401/producer) relative to the control group (USD 312/producer). On average, producers in the treatment group participated in 12 group workshops on hive management, received 6 queen bees per producer, and participated in 1 breeding workshop. However, there was some variability within the Treatment Group. For example, some producers received as many as 14 European queen bees, while others did not receive any, and only 152 producers participated of a breeding training. Hence, the Treatment Group can be subdivided in the following sub-groups:

- **Partial General Training Group (T1):** Includes 61 producers that were offered the full package but dropped out of the program before completing it (but completed the follow-up survey and are included in our sample). On average, these producers participated in 5 group workshops per producer and received some basic inputs (USD 110/producer) but were no longer provided with inputs after they stopped attending the trainings, and they did not receive any queen bees nor the queen breeding training. 58 of the 61 producers in this group dropped out in the first 6 months of program implementation, citing as main reasons either prioritizing their other production activities (such as crops), getting a job in a nearby city or migrating to the United States. Because the training was not completed, it is likely that these producers did not obtain the full benefits from training. Moreover, those who chose to drop out may have substantially different unobservable characteristics relative to the producers that decided to remain in the program. While we recognize that this sub-group is not an ideal comparison to other groups due to the self-selection out of the program, we consider it can still provide some useful information about the benefits from partial training participation (if any), and to assess the robustness of our main findings.
- **Queen Bee Group (T2):** Includes 122 producers that received basic inputs, general training, and queen bees as assets, but did not receive the queen-bee breeding training. All producers in this group completed the general training, participating in 13 group workshops on hive management on average.
- **Full Package Group (T3):** Includes 152 producers that received basic inputs, general training, queen bees, and it was the only group receiving the queen bee breeding training. Producers are included in this group if they received at least one breeding training, even if they did not attend both

[4] Given that the program was provided free of charge and that it included asset transfers, the initial take-up of the program was 100%, for both the control and the treatment group (though some producers then abandoned the trainings and/or did not take-up all of the components).

sessions offered. All producers in this group completed the general training, participating in 13 group workshops on hive management on average (like the previous group). However, the Full package Group received slightly more queen bees on average relative to the Queen bee Group.

4.3. Preliminary Evidence

Table 5 presents the change in key outcome variables over time for the overall sample and for the different groups defined above (in Section 4.2). Specifically, we conduct a before-and-after analysis, where we compare the value of key outcomes after the program was implemented (2022 follow-up survey) relative to their value before the program (2020 baseline survey). The first row of the table presents the result of a regression for the entire study sample (356 producers), while the remaining rows consider separate regressions for each of the groups described earlier (Control, Treatment, Partial General Training, Queen Bee, and Full Package). As would be expected, performance in terms of number of hives, production, and yield improves over time for the treatment group, but not for the control, and it is highest for those beekeepers receiving the full package including queen bees and the breeding training. Given that the estimates for the full sample are a weighted average of the estimates for each of the other sub-samples, and that most producers were treated, the average change for the full sample resembles that of the treatment group. Table A 1 in Annex 1 shows that the differences observed across groups are statistically significant.[5]

5. Identification strategy and estimation method

While the previous estimates (in Section 4.3) are interesting and intuitive, they cannot be considered reliable estimates of the impacts of the program due to several reasons. For example, in 2020, beekeepers were affected by a tropical storm, and consequently after this adverse shock, it is to be expected that performance would improve over time, reversing to the mean prior to the shock. Additionally, there may be many other variables changing over time that are not controlled for in this analysis, such as other weather factors, COVID, etc. For example, pollen availability may have been better in 2022 than in other years. A preferable approach would be to compare the change in key outcome variables for those participating in the program to similar beekeepers that did not participate in the program.

5.1. Identification Strategy

In this section, we briefly describe the different group comparisons that are carried out to estimate the impacts of interest. To better explain our approach, Figure 5 presents the different sub-groups discussed above, and it marks in dark blue the different program components received by each sub-group. Based on these categorizations, it is feasible to identify the following impacts:

Overall impact (T-C): The main analysis in the paper studies the average treatment effect (ATE) of the overall program by comparing the outcomes of producers in the Treatment Group to those in the Control Group. Although a completely "clean" control group is not available for our study, since all producers included in the surveys received at least some basic inputs, our estimates of the impact of the program are conservative, as they do not include the entire contribution of the basic inputs provision (which were also received by the Control Group). As an extension, in Section 7 we attempt to estimate

[5] Table A 1 in Annex 1 presents difference-in-differences estimates that compare the changes over time in key outcome variables between the different sub-samples, while also using year and producer fixed effects. Each row presents the marginal effects for a different comparison. The first row compares the performance of the Full package group versus the Control group (which would give us the effect of the full package of general training, queen bee asset, and breeding training), while the remaining rows compare additional groups (in line with Figure 1 above). As one would expect, the first row indicates that the increases in hives, production and yield for the Full package group are large and significant when compared to the Control group. However, contrary to what would be expected, partial participation in the general training also appears to have substantial impacts, suggesting that there may be important differences across groups that need to be controlled for. Tables A 2-5 in Annex 1 present the before-and-after and difference-in-differences results for the adoption of key practices.

the impacts of receiving different levels of basic inputs.

Full Package impact (T3-C): We also estimate the average treatment effect on the treated (ATT) of specific sub-treatments. We estimate the impact of receiving full package, for those that receive all components, by comparing the outcomes of producers in the Full Package Group to those in the Control Group. In addition, we split the full-package impact, by estimating the marginal impact of sequentially adding each program component:

1. **Partial General Training impact (T1-C):** First, we estimate the impact of *partial* participation in the general training, by comparing producers in the Partial General Training Group (receiving some training sessions + basic inputs) to those in the Control Group (receiving basic inputs only).
2. **Queen Bee impact (T2-T1):** Second, we estimate the impact of completing the training and adding the provision of queen bees, by comparing the outcomes of those in the Queen Bee Group to those in the Partial General Training Group.
3. **Breeding Training impact (T3-T2):** Lastly, we estimate the impact of adding the breeding training, by comparing the outcomes of the Full Package Group to those in the Queen Bee Group.

While there are some limitations regarding how the different groups and impacts are defined, and it may not be possible to fully address these limitations and eliminate all possible bias in our setting, we attempt to address the different limitations as much as possible, as discussed below.

5.2. Estimation method

In Section 4.3, we presented some indications that the program resulted in improved outcomes for beekeepers. However, given that the treatments were not randomly assigned, there are ex-ante differences across groups that may bias the results. To address these concerns, our analysis combines difference-in-difference (implemented using a fixed-effects estimation) with entropy balancing.

A major advantage of having survey data before and after program implementation is that it allows us to account for unobservable factors – both at the producer level and time-varying – that may affect the outcomes of interest and the participation in the program. Specifically, the effects of the program will be estimated using the following fixed-effects linear regression model:

$$Y_{i,t} = \alpha_i + \alpha_t + \beta D_i Post_t + v_{i,t} \tag{1}$$

Where $Y_{i,t}$ represents the sets of outcomes to be considered for producer i at year t. Producer fixed effects α_i fully absorb any permanent heterogeneity at the producer level, and α_t represents yearly shocks that affect all producers. D_i is a binary variable that takes the value of one if the producer belongs to the treated group (T, T1, T2 or T3 depending on the effect estimated), and $Post_t$ is a binary variable that takes the value of one for the post-treatment period. Therefore, β represents the parameter of interest which captures the effect of the treatment on the outcome under consideration. When comparing the Treatment (T) and Control (C) groups, β will be the average treatment effect (ATE) of the program, while in the intermediate comparisons between groups (T3-T2, T2-T1, and T1-C), β will represent the average treatment effect on the treated (ATT) for each program component.

The validity of the fixed-effects estimator rests on the identification assumption that trends in the outcome variables for the Treatment and Control groups would have been equal in the absence of treatment. However, this assumption may be problematic given that producers in the control group can be very different from treated producers and may likely follow different trends as well.

To strengthen the validity of our identification strategy, we combine the fixed-effects (difference-in-difference) methodology with entropy balancing, a multivariate reweighting method described in Hainmueller (2012). The reweighting scheme assigns a weight to each sample unit such that the reweighted groups satisfy a set of balance constraints that are imposed on the sample moments of the covariate distributions. Since the weights are directly adjusted to the known sample moments, the scheme always improves on the covariate balance, while keeping the weights as close as possible to the base (unit) weights to prevent loss of information (Figal Garone et al., 2015; Hainmueller & Xu, 2013). This method helps eliminate a potential source of bias since weighted producers in the control group are expected to be more similar to producers in the treatment group.

In our case, we will reweight the control group for each comparison to match the sample mean of the treatment group in the pre-treatment period in order to then estimate equation (1) using the corresponding treatment group and the reweighted control group. As described in Hainmueller (2012), the weights ω_i are chosen by the following scheme:

$$\min_{\omega_i} H(\omega) = \sum_{\{i \backslash D_s = 0\}} h(\omega_i) \tag{2}$$

Subject to balance and normalizing constrains:

$\sum_{\{i \backslash D_s=0\}} \omega_i k_{ri} = m_r$ with $r \in 1, \dots, R$ and

$\sum_{\{i \backslash D_s=0\}} \omega_i = 1$ and

$\omega_i \geq 0$ for all i such that $D_s = 0$,

Where D_s is the treatment status, $h(.)$ is an entropy metric, and $k_{ri}(X_i) = m_r$ describes a set of R balance constraints imposed on the covariate mean of the reweighted control group in order to equal the covariate mean of the treatment group.[6]

For each set of comparisons (T-C, T3-T2, T2-T1, and T1-C) weights are created based on pre-treatment values of selected outcomes (such as hives, production, yield, and adoption of key practices) and on observable characteristics of the producers in 2020 (such as age, gender, education, and hives lost during floods). The obtained weights from this process are then passed on to the fixed-effects model (1) through sampling weights that denote the inverse of the probability that the observation is included as a result of the sampling design.[7]

After controlling for these sources of heterogeneity that affect both the set of outcomes and the participation in the program, the identifying assumption implies that the fixed-effects (difference-in-difference) method applied to the reweighted sample leads to a consistent estimator for β. Finally, we cluster standard errors at the locality level for inference robust to correlation across firms and account for multiple hypothesis testing by computing sharpened False Discovery Rate (FDR) q-values (Anderson, n.d., 2008; McKenzie, 2020).

5.3. Validity of the Identification Strategy

The different panels of Table 6 compare the mean values of key covariates for each of the treatment groups relative to a group receiving less project components. As expected, since the assignment to the different treatments was not random but instead considered beekeepers' baseline characteristics and performance, with the original sample (prior to reweighting the data), there are large and significant differences in ex-ante covariates and in lagged-values of the outcomes, especially when comparing the Treated, Partial General Training and Full Package groups with the Control group, since Control group producers were located in different localities and were generally better off in terms of production and best practices adoption, which was also part of the reason they only received the basic inputs component

[6] We use the Stata package called *ebalance* introduced by Hainmueller and Xu (2013).
[7] Entropy balancing is doubly-robust in the sense that applying a regression adjustment model to the reweighted data does not change the estimate of the treatment effect (as long as the same covariates are used in the regression adjustment), and using a regression adjustment model means no further correction is needed for the standard errors (Jann, 2020; Zhao & Percival, 2017). This is implemented in Stata using the *teffects ra* command to run the regression adjustment model.

of the program.

However, after using entropy balancing, those differences become non-significant in most cases when comparing each treatment group to their reweighted control group. Notably, differences in lagged production and hives become non-significant, supporting the parallel trends assumption underlying the estimation of our difference-in-differences model. Some significant differences remain, however, especially when comparing the Partial General Training and Control groups, since the relatively small number of observations makes it difficult to perfectly balance the groups. As a robustness check, in Section 8 we present the results using Propensity Score Matching (PSM) and Inverse Probability Weighting (IPW) as alternative balancing techniques.

6. Main results

Below we present our main results. Section 6.1 discusses the ATE of the overall program on key outcomes. Section 6.2 discusses the ATT of receiving the full package, as well as the impact of adding each of the individual program components. We study impacts on productive performance outcomes (Table 7 for ATE and Table 10 for ATT), and on practice adoption (Table 8 and Table 9 for ATE and Table 11 and Table 12 for ATT).

6.1. Overall impact of the program (ATE)

Table 7 presents the average treatment effect **(T-C)** of the overall program on key productive performance outcomes such as the number of hives, honey production, and yield. Our results show a marked improvement in production variables for treated beekeepers relative to the control: Producers in the Treatment Group increased their hives by 20.3%, their honey production by 30.3% and their yield by 10.8% relative to the Control Group. The increase in the number of hives (4.95 hives) and honey production (246 kg) relative to the control, while substantial, is not surprising. It captures to some extent that most of the 6 queen bees received per producer were used to increase the number of hives and not only to replace old queens. Instead, more interesting is the impact that the program had on productivity, resulting in additional 2.3 kg. of honey per hive. The new queen bees, with improved genetics, appear to be an important driver for this improvement, as the more queen bees received the larger the impact on yields (see results in Section 7).

Table 8 and Table 9 present the average treatment effect of the overall program on the adoption of key practices. The first column of Table 8 presents the results for an aggregated practices variable, which takes the value of one if the producers adopt at least one key practice (that is, the effect on the **extensive margin**), while the second column is an index variable which ranges from 0-9 indicating how many key practices producers adopt (that is, the effect on the **intensive margin**). The rest of the columns in Table 8 are dummy variables for queen-bee related key practices that make up part of the index variable: (i) annual change of queen bees, (ii) breeding of queen bees, and (iii) keeping a record log of activities. Table 9, on the other hand, presents the rest of the individual key practices: (iv) use of supplemented feeding, (v) use of sustainable pest control, (vi) good condition of brood chambers, (vii) honey suppers, (viii) outer covers, (ix) and brood frames.

Although the extensive margin effect is relatively low (Treated producers are only 5.9% more likely to adopt any of the 9 key practices than producers in the Control group), our results show a marked improvement in the intensity of practice adoption for the Treatment Group relative to the Control Group:[8] Treated producers increased their best practice adoption by 2.6 practices on average relative to the control. The improvement is notable in several dimensions, with treated producers being 53.3%

[8] The larger effect found for the intensive margin relative to the extensive margin may be explained by the fact that relatively few producers had adopted zero key practices at the start of the program (3 in the Control group and 35 in the Treated group), and so there was a larger room for improvement in the number of producers adopting more key practices, which is reflected by the intensive margin.

more likely to keep a record log, 28.1% more likely to do an annual change of queen bees, 21.3% more likely to breed their own queen bees, as well as more likely to use supplemented feeding and sustainable pest control, and keep some beehive components in good conditions relative to the control.

The observed improvement in the adoption of key practices, especially those related to queen bee management (changing, and breeding) and log of apiary activities, are consistent with what is generally considered as good beehive management (Blog Apicultura y Miel, 2022). Changing bees frequently is considered very important to improve productive outcomes. While queen bees can live up to five years, their breeding potential wears out long before, and although they may be eventually replaced by worker bees, relying on natural replacement is not the best strategy as a queen's decline can be prolonged and negatively impact the colony. Additionally, natural breeding is somewhat uncertain and there could be cases of swarming, unsuccessful mating, and defective queens. For this reason, beekeepers are generally advised to do a systematic queen bee replacement every two years in temperate-cold climates, and annually in sub-tropical and tropical climates (Dini & Bedascarrasbure, 2011; Gobierno de México, 2019), as is the case in Yucatan. To achieve this, it is crucial to know when each queen is born, through the implementation of record logs or other tracking systems to record queen birth dates, their qualities and genetic lineage.

Finally, the adoption of key practices in beehive management is important not only to improve productivity but also for the long-term climate resilience of producers. While the provision of basic inputs to rebuild their lost hives was key to recover productive capacity in the short term, the adaptation of beekeepers to climate change and future climate shocks requires the adoption of best practices. Climate change can lead to shifts in weather patterns, affecting the thermal regulation of beehives and the availability of nectar and pollen, as well as contribute to the spread of diseases and pests affecting bees (Landaverde et al., 2023; Van Espen et al., 2023; Vercelli et al., 2021). In this context, practices such as providing supplemented feeding in times of low flowering or during climate events, implementing sustainable pest control that does not harm bees, and breeding genetically improved queen bees are key for colony health and resilience (Dequenne et al., 2022; Kovačić et al., 2020; Neumann & Straub, 2023).

6.2. Impact on performance (ATT)

Next, we estimate the ATT for producers receiving all the components in the full package, as well as the contribution of the different components. In Table 10, Table 11 and Table 12, each row presents the results of a separate regression, comparing the performance of a sub-group receiving more program components versus another sub-group receiving less program components, where the sub-group receiving less components was reweighted using entropy balancing, calculated as described in the Section 5 above.[9]

Full Package impact (T3-C): The first row of Table 10 shows a marked improvement in production variables for beekeepers in the Full Package Group relative to the Control Group. Package group producers increased their hives by 24.3% and their honey production by 33.3% relative to the control, although the 10.5% effect on yield is not significant. These ATT estimates are higher than the ATE because the Full Package Group includes only beekeepers that take-up the full program, whereas the Treatment Group also included beekeepers that abandon the program or do not receive some components. The impact of the full package (in the first row of Table 10) is approximately equal to the sum of the three sub-impacts: partial general training; queen bees, and breeding training, shown in the other rows of the table.

[9] This approach of comparing each pair of groups at a time, to estimate the impacts of each program component, provides estimates of the ATT for each treatment. However, the approach has the limitation that then it is not possible to test statistically which of the program components has the largest impact.

1. ***Partial General Training impact (T1-C):*** The second row of Table 10 shows that the relatively large differences between the Partial General Training Group and the Control Group, which were found in in the before and after analysis (Table 5), shrink and loose statistical significance when adjusting for pre-existing differences. It is possible that partial participation in the general training may not be effective, or that dropouts may have unobservable differences that cannot be addressed with our methodology (that controls only for observables). Nevertheless, the general training (when completed) may still contribute to the success of the program. For instance, the general training may provide beekeepers with the basic knowledge needed to succeed in the breeding trainings (for instance, the practice of recording information in a log taught during the general training is needed for queen breeding).

2. ***Queen Bee impact (T2-T1):*** The third row of Table 10 shows that completing the general training and receiving the queen bees was the main driver of the improvements in performance outcomes. Producers in the Queen Bee Group increased their hives by 17.5%, honey production by 26.5%, and yields by 7.1% relative to the Partial General Training Group. These coefficients are almost as large as those obtained for the full package. Hence, queen bee provision likely helped to increase productivity, in line with the perception that having high quality queen bees is one of the most crucial factors for productive beekeeping (Blog Apicultura y Miel, 2022). Note, however, that the log transformation makes coefficients non-significant, which may be due to the relatively smaller number of observations used in this comparison relative to the estimation of the ATE.

3. ***Breeding Training impact (T3-T2):*** The fourth row of Table 10 shows that the breeding training does not seem to have an additional effect on performance variables in our results, with all coefficients being non-significant. However, as we will discuss in Section 6.3 below, the breeding training appears to be important to improve the adoption of best practices, which may help beekeepers to improve their performance over the long term, especially considering the importance of good quality queen bees.

6.3. Impact on the adoption of best practices (ATT)

To further explore the forces driving the results on performance, Table 11 and Table 12 present the effects of the program in the adoption of key hive management practices, following the same approach as in Section 6.2.

Full Package impact (T3-C). The first row of Table 11 and Table 12 show a marked improvement in practice adoption for beekeepers in the Full Package Group relative to the Control Group. While the effect on the extensive margin is not significant, Package group producers increased the intensity of their best practice adoption in 3.9 practices on average relative to the control. Moreover, the improvement is clear in almost all the dimensions, with producers receiving the full package being 77.9% more likely to keep a record log, 48.9% more likely to do an annual change of queen bees, 46.2% more likely to breed their own queen bees, as well as more likely to use supplemented feeding and sustainable pest control, and keep some beehive components in good conditions relative to the control.

1. ***Partial General Training impact (T1-C):*** The second row of Table 11 and Table 12 show that, as in the case for production outcomes, the coefficients for most individual key practices are much smaller than in the Full Package vs Control comparison and not significant. Surprisingly, however, the effect on the both the extensive margin and intensive margin are significant, with Partial General Training Group producers 19% more likely to adopt at least one key practice relative to the control, and adopting 1.5 more key practices. These results suggest that even partial participation in the general training may lead to improvements in the adoption of some practices (such as keeping beehive components in good conditions).

2. **<u>Queen Bee impact (T2-T1):</u>** The third row of Table 11 and Table 12 show that completing the training and adding the queen bee asset provision appears to be an important driver for the adoption of best practices. Producers in the Queen bee Group adopt approximately 1.7 more practices relative to the Partial General Training Group, being 21.4% more likely to do an annual change of queen bees, 17.1% more likely to breed hem, and 43.4% more likely to keep a record log of activities in the apiary. Based on qualitative evidence from the field, the introduction of the fertilized queens motivated the tracking of activities in the apiary in a log. During the first few months of implementation, it was perceived that keeping a record log represented a burden to the producers, due to a lack of habit, but their reluctance began to change with the arrival of the fertilized queen bees, which was when producers realized the importance of having their hives numbered and keep records of each one, mainly in terms of the queen's posture and behavior of the colonies.[10] Instead, the coefficient for the improvement in queen bee breeding practices, while significant, is much smaller than observed in the Full Package impact, suggesting that most of the improvement seen in the Full package vs Control comparison may be due to the breeding training component, as discussed below.

3. ***<u>Breeding Training impact (T3-T2):</u>*** The fourth row of Table 11 and Table 12 show that the breeding training seems to have an important effect on the adoption of best practices, improving in 1.5 practices relative to the Queen Bee Group, especially those related to queen bee management and log of apiary activities, with producers receiving the breeding training 21.8% more likely to do an annual change of queen bees, 34.1% more likely to breed them, and 42.5% more likely to keep a record log. The observed increase in the likelihood of queen breeding is particularly important due to the difficulties of obtaining quality queen bees for purchase. In general, newly introduced queen bees can come from the beekeeper's own queen breeding efforts or be purchased from certified hatcheries who select the highest quality strains[11]. In Mexico, however, there are approximately 40 certified hatcheries with a production of around 200,0000 queens annually, with does not satisfy the 2.2 million queens a year that are needed given the country's number of hives (Food and Agriculture Organization of the United Nations, 2023; Gobierno de México, n.d.; Guzmán-Novoa et al., 2011), thus developing the abilities of producers to breed their own queen bees is key for them to access the amount and quality of queens needed for an adequate hive management.

7. Extensions

7.1. Heterogeneous effects

Gender. Table 13 presents the effect of the overall program (ATE) for women on key performance outcomes. For this, we interact the treatment variable with a dummy indicator for each gender (male and female).[12] Then, we evaluate if the difference in the coefficients for treated women and treated men are statistically significant. A priori, it was expected that the program would have larger impacts for women than for men as women had often been excluded from previous interventions. Moreover, it was expected that the introduction of queen bees with European genetics, less aggressive than Africanized bees, would improve the participation of women in apiaries and facilitate their work, thus improving their performance more than for men. However, our results in Table 13 suggest that the program did

[10] Producers that received the fertilized queen bees were advised and accompanied during the whole process to prepare the hives where the queens would be introduced, during their introduction, and after the delivery of the queens, to monitor the behavior of their hives, and observe the level of defensiveness, cleanliness, and productivity of the new queens.

[11] In Mexico, selective breeding practices are especially important to control the presence of Africanized bees, which are generally more defensive, have a higher tendency to swarm, and generate higher management costs and lower honey production compared to European bees (Gobierno de México, n.d.; Guzmán-Novoa et al., 2011).

[12] Since most producers in the control group are men (20 men and 1 woman), we do not interact the constant with a dummy indicator for each gender (to avoid splitting the control group between men and women).

not have a conclusive positive effect for women relative to men.[13] Although the increase in the number of hives seems to be higher for women (an increase of 12% relative to men), the results for production and yield suggest that women's performance was worse or non-distinguishable from men's (while the coefficient for kilos of production is negative and significant, the log specification is not significant, and vice versa for yield). The lack of conclusive heterogenous effects by gender is supported with several robustness tests. It is worth noting that the lack of significant effects may be because we have a relatively small sample and women only make up 20% of producers included in the program.[14]

Smaller vs. larger producers. Panel A in Table 14 presents the effect of the overall program for producers who were larger at the beginning of the program on key performance outcomes, defined as those beekeepers that had a level of production above the median in 2020 (315 kg). As shown in the table, the effect of the program is not statistically different for larger versus smaller producers.

Producers more vs. less affected by floods. Panel B in Table 14 presents the effect of the overall program for producers who were more affected by floods on key performance outcomes, defined as those beekeepers that had a level of hives lost due to floods above the median in 2020 (13 hives). As shown in the table, the effect of the program is not statistically different for more versus less affected producers.

7.2. Incremental effects of basic inputs and queen bees

In this subsection we attempt to estimate the impact on key performance outcomes using continuous treatment variables for the (log) amount of basic inputs, basic inputs per hive and queen bees received. The results are presented in Table 15, where each row presents the result of a separate regression, comparing different dosages of the program components using entropy balancing weights, calculated using an extension of the entropy method described in Section 5 for continuous treatment variables (EBCT), as presented in Tübbicke (2022).[15] The corresponding balance tests are presented in Table A 7 in Annex 1, which show that pre-balancing differences in observables become non-significant in all cases after using EBCT weights.

As shown in Table 15, marginally increasing the provision of basic inputs or queen bees results in increases mainly on honey production and yield. Panel A shows that a 1% increase in the amount of basic inputs received generates a 1.6kg. increase in production (although the log specification is not significant), while measuring the amount of basic inputs relative to the number of hives (shown in Panel B) suggests that a 1% increase in the amount of basic inputs per hive generates a 12.5% increase in yield. In terms of queen bees, results shown in Panel C of Table 15 suggest some effect on yield, although results are not significant in the log specification. The significance found for some of the log transformation of program components may suggest non-linear effects, where receiving at least some basic inputs or queen bees has an impact on performance, but with decreasing returns once producers already have some of these inputs. Table A 8 in Annex shows results using continuous basic inputs and queen bees (not log transformed), and the effect for basic inputs and queen bees is now not significant.

7.3. Cost-effectiveness

In this section we assess the cost-effectiveness of the program. Total budget for the program was

[13] Note the standard errors in these results do not consider the two-step procedure, as the *teffects ra* command in Stata does not allow for interaction terms. No effects are found for best practices either (not shown, but available from the authors upon request).

[14] These results for gender heterogeneity of the program are robust to different specifications. Table A 6 in the Annex shows no effects on number of hives and yield when comparing women in the Queen bee plus Full package groups relative to the Control plus Partial general Training groups, which allows us to split the comparison group between men and women. For production, the coefficient for kilos is negative and significant at the 10%, but the log specification is not).

[15] Note the standard errors in these results do not consider the two-step procedure, as the *teffects ra* command in Stata does not allow for continuous treatment variables.

US$299,281, which includes the provision of all components (basic inputs, general training, queen bees, and breeding trainings) as well as budget for evaluation purposes. From this total budget, we estimate that US$200,148 was spent on providing the program components to the 356 producers included in our evaluation before the follow-up survey (see Table A 9 in Annex 1 for details), of which: US$141,076 correspond to the basic inputs component, US$36,030 to the provision of general trainings, US$17,758 to the queen bee asset, and US$5,284 to the provision of breeding trainings.

Considering this budget and the effects on production obtained from our results on the ATT of the different program components presented in Table 10, we approximate the cost-effectiveness of the program in Table 16. Assuming a flat cost per producer, we estimate that US$1 invested in providing the full package results in an increase of 0.5 kg. in honey production or US$1.24 in honey sales. This effect is mostly driven by the provision of the queen bee asset, with a cost-effectiveness of US$1.37 per dollar invested. Note, however, that this calculation only includes the initial, short-term, increase in production and sales derived from the program, but it would be reasonable to expect that the training components of the program would also have an effect over the long-term, especially with the improvement in the adoption of best practices discussed in Section 6.

8. Robustness checks

8.1. Alternative methodologies for balancing pre-treatment differences: PSM and IPW

As a robustness check, we estimate our main tables using different matching techniques (PSM and IPW) as opposed to entropy balance. As in the case of entropy balance, for each set of comparisons (T-C, T3-T2, T2-T1, and T1-C), the groups are balanced based on pre-treatment values of selected outcomes (hives and production) and on observable characteristics of the producers in 2020 (such as age, gender, and education). Then, for each set of comparisons, we keep only the observations on the common support or overlap region and estimate equation (1) using propensity score or inverse probability weights in each case. A GMM estimator is used to implement all steps simultaneously, to ensure that standard errors are correct, by considering that the propensity score and the inverse probability are estimated and have some variability associated to it.[16]

The estimation of the ATE on performance outcomes is shown in Table 17, where Panel A presents our main results discussed in Section 6 using entropy weights, and Panels B and C show the effects using PSM and IPW, respectively.[17] Regardless of the methodology used, results are qualitatively similar, suggesting that our findings are robust and that the program was effective at improving producers' performance (number of hives, honey production and yields) as well as adoption of best practices.

8.2. Alternative comparisons between groups

As an additional robustness check, Table 18 presents alternative comparisons between groups in terms of their performance outcomes. Our main results from Section 6 suggest that each treatment group performed better or at least as well as a group receiving one fewer program component (see Table 10, in most cases, between group comparisons are either positive and significative or positive but not significant). In line with this, we would expect that a group receiving more than one additional program components relative to another group would also perform better.

In Table 18, the first line compares the Queen Bee group to the Control group (T2-C, where the Queen Bee group received the complete general training and queen bees as additional components relative to the Control group), and the second line compared the Full Package group to the Partial General Training

[16] This is implemented in Stata with the *teffects psmatch* command for PSM and *teffects ipw* for IPW.
[17] In addition, Tables A 8 – A 15 in the Annex present the estimation of the ATT for the different program components on performance outcomes and the adoption of best practices.

group (T3-T1, where the Full Package group received the complete general training, queen bees and the breeding training compared to the Partial General Training group). In both cases, the coefficients on performance outcomes are positive and significative, as expected.[18]

9. Next steps

The available data allows for additional analysis and extensions that may be completed in the future. First, producers were given fertilized queen bees from three different certified hatcheries (Elite, in the state of Nayarit; MAGOC, in the state of Veracruz; and Vallaloid, in the state of Michoacán), and it may be important to assess if there are heterogenous effects for different sources of queen bees. In addition, given that the topics covered by the general training varied slightly across locations, it may be interesting to assess if some training topics were more important than others and whether those receiving more training topics benefitted more. Finally, robustness tests using the year of treatment or the provision of queen bees in April 2023 as placebos can be added to strengthen our results.

10. Conclusion

Our results show that a program combining training with asset provision can be effective to improve the productivity and resilience of smallholder beekeepers. Treated beekeepers improved their productive performance (i.e., number of hives, honey production, and yields) and the adoption of best practices in hive management substantially more than the control group, who only received basic inputs. These results are statistically significant at conventional levels and remain qualitatively unchanged regardless of the approach used to balance differences between groups.

While the different project components appear to have contributed to the overall effectiveness of the program, the provision of the queen bee asset had the largest impact on productive performance outcomes. Queen bees not only improved the genetic quality of the bee colony, which is important for honey production and productivity, but also allowed beekeepers to adopt good practices related to queen bee management, which were further improved by the provision of the queen breeding training.

Although the general beehive management training and the queen bee breeding training did not significantly contribute to the improvement of productive outcomes in the short term, they did improve the adoption of good hive management practices, which are particularly important for long-term climate resilience. Climate change can lead to shifts in weather patterns, affecting the thermal regulation of beehives and the availability of nectar and pollen, as well as contribute to the spread of diseases and pests affecting bees, meaning practices such as supplemented feeding, sustainable pest control, and breeding of genetically improved queen bees are key for colony health and adaptation to climate change.

Our findings are important to guide the design of agricultural technical assistance programs, particularly those focused on apiculture. Investing in comprehensive and cost-effective interventions, such as the one studied in this paper, should be at the center of broader development strategies by governments, multilateral development banks and other institutions aimed at supporting sustainable agriculture and rural livelihoods. As climate change continues to pose challenges to beekeeping, supporting adaptive practices will be essential for ensuring the health and sustainability of bee colonies. This way, we can prolong the beneficial relationship between bees and their keepers for years to come.

[18] Practice adoption results are also positive and significant (not shown, but available from the authors upon request).

Tables

Table 1: Summary of studies evaluating agricultural interventions

Study	Location	Duration of program evaluated	Program components	Effects	Methodology
Training programs					
Hörner et al. (2021)	Ethiopia	1 cropping season (April-October)	Bottom-up training alone or combined with video	Farmers treated adopt an additional ~0.5-0.6 practices (~8-11%) from a complex technology package relative to the control group; no effects on income or yield 1 year after training	RCT (n = 2,382)
Baul et al. (2023)	India	2 cultivation and harvest seasons (June-August)	Bottom-up training alone or combined with video	~5-12% increase in output and ~9-24% increase in profits for treated producers, although not robust to different specifications	RCT (n = 2,520)
Grimm & Luck (2020)	Indonesia	3 days	Hands-on training	~8-13% increase in organic input adoption for treated producers; improvements in knowledge and perceptions	RCT (n = 1,149)
Kondylis et al. (2017)	Mozambique	Two 3-day trainings	Augmentation of on-going training with direct training of new technology	18% increase in adoption of new technology for treated farmers	RCT (n = 347)
Wonde et al. (2022)	Ethiopia	1 cropping season	Training and certification	27% increase in wheat yield, 10% increase in maize yield, 20% increase in maize income for treated producers	Cross-sectional data using PSM (n = 401)
Input provision programs					
Gignoux et al. (2022)	Haiti	One-time package of vouchers	Input subsidies	36% reduction in yield, 28% decrease in fertilizer use for treated producers	RCT (n = 515)
Carter et al. (2021)	Mozambique	One-time package of vouchers	Input subsidies	23% increase in yield, persistent after program ends	RCT (n = 514)
Hemming et al. (2018)	India, Malawi, Mali, Mozambique, Nigeria, Tanzania, Zambia	Varied	Input subsidies	Positive effects on adoption of inputs (0.23 SD), productivity (0.11 SD), and farm income (0.17 SD)	Meta-evaluation of 15 studies using experimental and quasi experimental methods
Jayne et al. (2018)	Ghana, Nigeria, Kenya, Tanzania, Malawi, Zambia, Ethiopia	Varied	Input subsidies	Improvements in yields and production levels, but overall welfare effects are limited	Review of 72 studies using experimental and

					quasi experimental methods
Training and input provision alone or combined					
Nankhuni & Paniagua (2013)	Africa, Asia, Latin America	Varied	Varied: training (top-down and bottom-up), credit and grants of inputs, or combination of both	Positive effects are generally found in knowledge, and technology adoption; impacts on production and productivity tend to be mixed; no effects are generally found in terms of poverty reduction; most successful programs provide training combined with provision of credit or in-kind inputs	Meta-evaluation of 65 studies using experimental and quasi experimental methods

Table 2: Descriptive statistics

Variable	Mean	SD	Min	Max	Median
Age	44.9	14.6	18	90	43.0
Post primary education (= 1)	0.3	0.5	0	1	0.0
Women (= 1)	0.2	0.4	0	1	0.0
Participation in general trainings (#)	10.8	4.2	0	14	13.0
Participation in queen bee trainings (#)	0.8	1.1	0	3	0.0
Basic inputs delivered (USD)	395.6	172.1	47	807	420.9
Basic inputs delivered per hive (USD/hives)	32.0	35.7	1	458	23.7
Queen bees received (#)	5.3	3.3	0	14	7.0
Hives 2020 (#)	22.1	26.4	1	250	16.0
Hives 2022 (#)	31.8	29.6	3	279	25.0
Production 2020 (kg)	490.0	654.5	16	5250	315.0
Production 2022 (kg)	886.8	989.1	0	7755	634.0
Yield 2020 (kg/hive, capped at 100)	22.6	13.4	0	100	21.0
Yield 2022 (kg/hive, capped at 100)	27.2	14.3	0	100	26.0
Practices 2020 (0-9)	2.8	2.1	0	8	2.0
Practices 2022 (0-9)	6.7	2.0	0	9	7.0
Annual change of queen bees 2020 (= 1)	0.4	0.5	0	1	0.0
Annual change of queen bees 2022 (= 1)	0.7	0.4	0	1	1.0
Breeds queen bees 2020 (= 1)	0.2	0.4	0	1	0.0
Breeds queen bees 2022 (= 1)	0.5	0.5	0	1	0.0
Keeps record log 2020 (= 1)	0.0	0.1	0	1	0.0
Keeps record log 2022 (= 1)	0.5	0.5	0	1	1.0
Supplemented feeding 2020 (= 1)	0.3	0.4	0	1	0.0
Supplemented feeding 2022 (= 1)	0.4	0.5	0	1	0.0
Sustainable pest control 2020 (= 1)	0.6	0.5	0	1	1.0
Sustainable pest control 2022 (= 1)	0.8	0.4	0	1	1.0
Good condition of brood chambers 2020 (= 1)	0.2	0.4	0	1	0.0
Good condition of brood chambers 2022 (= 1)	0.9	0.3	0	1	1.0
Good condition of honey suppers 2020 (= 1)	0.3	0.4	0	1	0.0
Good condition of honey suppers 2022 (= 1)	0.9	0.3	0	1	1.0
Good condition of outer covers 2020 (= 1)	0.5	0.5	0	1	0.0
Good condition of outer covers 2022 (= 1)	0.9	0.2	0	1	1.0
Good condition of brood frames 2020 (= 1)	0.3	0.4	0	1	0.0
Good condition of brood frames 2022 (= 1)	0.9	0.3	0	1	1.0

Table 3: Program components received, by group

	Basic inputs delivered (USD)	Participation in general trainings (#)	Queen bees received (#)	Participation in queen bee trainings (#)
Control group (C): 21 producers				
Mean	312	0	0	0
Total	6,557	0	0	0
Treated group (T): 335 producers				
Mean	401	12	6	1
Total	134,289	3,854	1,894	289
Partial General Training group (T1): 61 producers				
Mean	110	5	0	0
Total	6,739	296	0	0
Queen Bee group (T2): 122 producers				
Mean	428	13	6	0
Total	52,260	1,552	760	0
Full Package group (T3): 152 producers				
Mean	495	13	7	2
Total	75,290	2,006	1,134	289

Table 4: Mean of characteristics, by group

Variable	(C) Control	(T) Treated	(T1) Partial General Training	(T2) Queen Bee	(T3) Full Package
Age	55.3	44.3	46.9	43.6	43.8
Post primary education (= 1)	0.5	0.3	0.2	0.3	0.4
Women (= 1)	0.0	0.2	0.2	0.3	0.1
Hives 2020 (#)	21.5	22.1	18.8	25.4	20.8
Hives 2022 (#)	24.0	32.2	24.4	36.5	32.0
Production 2020 (kg)	497.4	489.5	275.1	570.5	510.6
Production 2022 (kg)	625.8	903.1	431.1	1034.7	987.0
Yield 2020 (kg/hive, capped at 100)	22.7	22.5	15.5	22.7	25.2
Yield 2022 (kg/hive, capped at 100)	25.6	27.3	18.4	27.5	30.6
Practices 2020 (0-9)	3.4	2.7	2.3	2.9	2.7
Practices 2022 (0-9)	4.8	6.8	4.2	6.8	7.8
Annual change of queen bees 2020 (= 1)	0.5	0.4	0.3	0.5	0.4
Annual change of queen bees 2022 (= 1)	0.4	0.8	0.3	0.8	0.9
Breeds queen bees 2020 (= 1)	0.2	0.2	0.1	0.3	0.2
Breeds queen bees 2022 (= 1)	0.2	0.5	0.1	0.4	0.7
Keeps record log 2020 (= 1)	0.1	0.0	0.0	0.0	0.0
Keeps record log 2022 (= 1)	0.1	0.6	0.1	0.6	0.8
Supplemented feeding 2020 (= 1)	0.4	0.3	0.1	0.3	0.3
Supplemented feeding 2022 (= 1)	0.4	0.4	0.1	0.4	0.6
Sustainable pest control 2020 (= 1)	0.2	0.6	0.6	0.5	0.6
Sustainable pest control 2022 (= 1)	0.3	0.8	0.7	0.8	0.9
Good condition of brood chambers 2020 (= 1)	0.3	0.2	0.2	0.2	0.2
Good condition of brood chambers 2022 (= 1)	0.8	0.9	0.7	1.0	1.0
Good condition of honey suppers 2020 (= 1)	0.4	0.2	0.3	0.2	0.2
Good condition of honey suppers 2022 (= 1)	0.8	0.9	0.8	1.0	1.0
Good condition of outer covers 2020 (= 1)	0.6	0.5	0.4	0.5	0.5
Good condition of outer covers 2022 (= 1)	0.8	0.9	0.8	1.0	1.0
Good condition of brood frames 2020 (= 1)	0.6	0.3	0.2	0.3	0.3
Good condition of brood frames 2022 (= 1)	0.9	0.9	0.8	0.9	1.0

Table 5: Before and After on performance

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
All producers						
Post=1 x All=1	9.699***	0.499***	396.791***	0.699***	4.609***	0.208***
	(2.100)	(0.055)	(62.861)	(0.071)	(1.038)	(0.046)
Observations	712	712	712	711	712	711
Control group (C)						
Post=1 x Control=1	2.476	0.124	128.333	0.244	2.905	0.120
	(2.571)	(0.118)	(98.328)	(0.155)	(2.287)	(0.094)
	42	42	42	42	42	42
Treated group (T)						
Post=1 x Treated=1	10.152***	0.523***	413.620***	0.728***	4.715***	0.213***
	(2.224)	(0.058)	(66.433)	(0.075)	(1.094)	(0.049)
	670	670	670	669	670	669
Partial General Training group (T1)						
Post=1 x Partial General Training=1	5.623	0.364**	156.008	0.598***	2.941	0.246*
	(4.626)	(0.145)	(106.198)	(0.182)	(2.501)	(0.140)
	122	122	122	121	122	121
Queen Bee group (T2)						
Post=1 x Queen bee =1	11.107**	0.542***	464.130***	0.741***	4.720***	0.197**
	(4.830)	(0.100)	(141.353)	(0.126)	(1.686)	(0.083)
	244	244	244	244	244	244
Full Package group (T3)						
Post=1 x Full Package=1	11.204***	0.571***	476.463***	0.764***	5.423***	0.210***
	(2.308)	(0.077)	(77.297)	(0.089)	(1.596)	(0.052)
	304	304	304	304	304	304

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on before-and-after comparison; robust SE in parentheses.

Table 6: Balance test of ex-ante covariates

Panel A: Treated vs Control (T-C)

Variable	Before balancing		Entropy balancing	
	Dif.	P-value	Dif.	P-value
Producer characteristics				
Age	-11.06***	(0.001)	-10.76	(0.370)
Post-primary education (= 1)	-0.21**	(0.050)	-0.19	(0.378)
Women (= 1)	0.16*	(0.073)	0.12	(0.202)
Hives lost due to floods (#)	4.58	(0.429)	4.76	(0.326)
Lagged and baseline productive outcomes				
2016 hives (#)	-2.78	(0.787)	5.50	(0.632)
2017 hives (#)	-0.54	(0.956)	4.28	(0.670)
2018 hives (#)	-12.56	(0.177)	2.78	(0.795)
2019 hives (#)	-11.88	(0.406)	1.45	(0.911)
2020 hives (#)	0.56	(0.925)	5.07	(0.144)
Ln(2020 hives)	-0.28	(0.129)	-0.06	(0.735)
2016 production (Kg.)	-497.51*	(0.062)	-381.59	(0.482)
2017 production (Kg.)	-498.40**	(0.045)	-469.53	(0.385)
2018 production (Kg.)	-929.57***	(0.000)	-502.70	(0.261)
2019 production (Kg.)	-1016.57***	(0.000)	-691.85	(0.152)
2020 production (Kg.)	-7.91	(0.957)	106.93	(0.346)
Ln(2020 production)	-0.39*	(0.079)	-0.11	(0.597)
2020 yield (Kg./hive, capped at 100)	-0.17	(0.956)	1.08	(0.639)
Ln(2020 yield)	-0.13	(0.373)	-0.07	(0.496)
Baseline best practices outcomes				
Practices (0-9)	-0.70	(0.130)	-0.20	(0.729)
Annual change of queen bees (= 1)	-0.08	(0.465)	-0.07	(0.747)
Queen bee breeding (= 1)	-0.01	(0.956)	0.12	(0.246)
Keeps record log (= 1)	-0.09***	(0.000)	-0.11	(0.267)
Supplemented feeding (= 1)	-0.16	(0.106)	-0.27	(0.206)
Sustainable pest control (= 1)	0.35***	(0.002)	0.42***	(0.000)
Good condition of brood chambers (= 1)	-0.12	(0.193)	0.05	(0.585)
Good condition of honey suppers (= 1)	-0.19*	(0.056)	0.04	(0.748)
Good condition of outer covers (= 1)	-0.08	(0.468)	-0.07	(0.758)
Good condition of brood frames (= 1)	-0.32***	(0.001)	-0.31	(0.153)
Producers	356		261	

Panel B: Full Package vs Control (T3-C)

Variable	Before balancing		Entropy balancing	
	Dif.	P-value	Dif.	P-value
Producer characteristics				
Age	-11.57***	(0.001)	-8.30	(0.417)
Post-primary education (= 1)	-0.12	(0.316)	-0.14	(0.468)
Women (= 1)	0.09	(0.246)	0.07	(0.414)
Hives lost due to floods (#)	2.77	(0.437)	4.10	(0.363)
Lagged and baseline productive outcomes				
2016 hives (#)	-3.97	(0.676)	3.02	(0.778)
2017 hives (#)	-1.88	(0.830)	1.74	(0.859)
2018 hives (#)	-13.04	(0.157)	0.70	(0.946)
2019 hives (#)	-7.51	(0.697)	4.92	(0.735)
2020 hives (#)	-0.73	(0.860)	3.11	(0.406)
Ln(2020 hives)	-0.25	(0.150)	-0.05	(0.799)
2016 production (Kg.)	-475.92*	(0.057)	-301.83	(0.455)
2017 production (Kg.)	-479.00*	(0.052)	-466.28	(0.285)
2018 production (Kg.)	-897.86***	(0.000)	-487.67	(0.197)
2019 production (Kg.)	-981.27***	(0.000)	-705.26	(0.119)
2020 production (Kg.)	13.12	(0.909)	78.01	(0.534)
Ln(2020 production)	-0.18	(0.347)	0.00	(0.986)
2020 yield (Kg./hive, capped at 100)	2.51	(0.424)	2.04	(0.413)
Ln(2020 yield)	0.04	(0.690)	0.02	(0.863)
Baseline best practices outcomes				
Practices (0-9)	-0.69	(0.140)	-0.19	(0.739)
Annual change of queen bees (= 1)	-0.11	(0.345)	-0.08	(0.689)
Queen bee breeding (= 1)	-0.01	(0.883)	0.11	(0.257)
Keeps record log (= 1)	-0.10***	(0.000)	-0.11	(0.197)
Supplemented feeding (= 1)	-0.15	(0.174)	-0.25	(0.184)
Sustainable pest control (= 1)	0.38***	(0.001)	0.46***	(0.000)
Good condition of brood chambers (= 1)	-0.11	(0.271)	0.04	(0.677)
Good condition of honey suppers (= 1)	-0.20**	(0.042)	0.01	(0.892)
Good condition of outer covers (= 1)	-0.08	(0.470)	-0.09	(0.635)
Good condition of brood frames (= 1)	-0.31***	(0.004)	-0.28	(0.151)
Producers	173		121	

Panel C: Partial General Training vs Control (T1-C)

Variable	Before balancing		Entropy balancing	
	Dif.	P-value	Dif.	P-value
Producer characteristics				
Age	-8.48**	(0.031)	-9.00	(0.543)
Post-primary education (= 1)	-0.33***	(0.004)	-0.29	(0.279)
Women (= 1)	0.12	(0.182)	0.08	(0.341)
Hives lost due to floods (#)	8.84	(0.397)	4.08	(0.520)
Lagged and baseline productive outcomes				
2016 hives (#)	-13.09	(0.154)	-0.89	(0.941)
2017 hives (#)	-10.24	(0.242)	-1.36	(0.897)
2018 hives (#)	-23.04**	(0.010)	-3.82	(0.734)
2019 hives (#)	-23.67**	(0.011)	-6.07	(0.657)
2020 hives (#)	-2.75	(0.619)	1.53	(0.699)
Ln(2020 hives)	-0.47**	(0.018)	-0.22	(0.273)
2016 production (Kg.)	-631.96**	(0.032)	-404.88	(0.481)
2017 production (Kg.)	-645.73**	(0.017)	-425.57	(0.437)
2018 production (Kg.)	-1068.04***	(0.000)	-450.41	(0.321)
2019 production (Kg.)	-1145.60***	(0.000)	-566.23	(0.236)
2020 production (Kg.)	-222.32*	(0.075)	-4.85	(0.967)
Ln(2020 production)	-1.07***	(0.000)	-0.53**	(0.015)
2020 yield (Kg./hive, capped at 100)	-7.24**	(0.022)	-2.59	(0.385)
Ln(2020 yield)	-0.60***	(0.002)	-0.31**	(0.020)
Baseline best practices outcomes				
Practices (0-9)	-1.13*	(0.055)	-0.67	(0.340)
Annual change of queen bees (= 1)	-0.21*	(0.083)	-0.23	(0.385)
Queen bee breeding (= 1)	-0.09	(0.348)	0.04	(0.740)
Keeps record log (= 1)	-0.10**	(0.014)	-0.08	(0.267)
Supplemented feeding (= 1)	-0.28***	(0.007)	-0.37	(0.160)
Sustainable pest control (= 1)	0.35***	(0.005)	0.41***	(0.001)
Good condition of brood chambers (= 1)	-0.15	(0.148)	0.04	(0.734)
Good condition of honey suppers (= 1)	-0.15	(0.208)	0.03	(0.811)
Good condition of outer covers (= 1)	-0.15	(0.255)	-0.16	(0.562)
Good condition of brood frames (= 1)	-0.36***	(0.002)	-0.35	(0.185)
Producers	82		60	

.

Panel D: Queen Bee vs Partial General Training (T2-T1)

Variable	Before balancing		Entropy balancing	
	Dif.	P-value	Dif.	P-value
Producer characteristics				
Age	-3.22	(0.138)	-1.07	(0.670)
Post-primary education (= 1)	0.07	(0.331)	0.00	(0.980)
Women (= 1)	0.16**	(0.025)	0.07	(0.506)
Hives lost due to floods (#)	-4.10	(0.430)	1.23	(0.790)
Lagged and baseline productive outcomes				
2016 hives (#)	16.09*	(0.050)	5.97	(0.491)
2017 hives (#)	15.41**	(0.047)	6.25	(0.430)
2018 hives (#)	15.95**	(0.020)	9.93	(0.211)
2019 hives (#)	12.33**	(0.050)	6.73	(0.401)
2020 hives (#)	6.59	(0.197)	4.14	(0.621)
Ln(2020 hives)	0.26*	(0.063)	0.15	(0.485)
2016 production (Kg.)	169.53	(0.431)	-0.04	(1.000)
2017 production (Kg.)	190.58	(0.322)	75.22	(0.680)
2018 production (Kg.)	166.71	(0.121)	6.40	(0.969)
2019 production (Kg.)	148.10	(0.178)	-48.30	(0.787)
2020 production (Kg.)	295.41**	(0.015)	199.46	(0.108)
Ln(2020 production)	0.76***	(0.000)	0.21	(0.322)
2020 yield (Kg./hive, capped at 100)	7.28***	(0.000)	1.77	(0.501)
Ln(2020 yield)	0.50***	(0.000)	0.05	(0.695)
Baseline best practices outcomes				
Practices (0-9)	0.64*	(0.052)	0.53	(0.334)
Annual change of queen bees (= 1)	0.23***	(0.003)	0.20	(0.127)
Queen bee breeding (= 1)	0.14**	(0.038)	0.15*	(0.093)
Keeps record log (= 1)	0.02	(0.317)	0.02	(0.158)
Supplemented feeding (= 1)	0.16**	(0.022)	0.24***	(0.004)
Sustainable pest control (= 1)	-0.05	(0.530)	0.01	(0.963)
Good condition of brood chambers (= 1)	0.03	(0.605)	-0.03	(0.772)
Good condition of honey suppers (= 1)	-0.03	(0.634)	-0.06	(0.634)
Good condition of outer covers (= 1)	0.10	(0.212)	0.03	(0.822)
Good condition of brood frames (= 1)	0.05	(0.469)	-0.02	(0.867)
Producers	183		140	

Panel E: Full Package vs Queen Bee (T3-T2)

Variable	Before balancing		Entropy balancing	
	Dif.	P-value	Dif.	P-value
Producer characteristics				
Age	0.13	(0.939)	0.31	(0.898)
Post-primary education (= 1)	0.15**	(0.012)	0.00	(0.983)
Women (= 1)	-0.18***	(0.000)	-0.00	(0.949)
Hives lost due to floods (#)	-1.96	(0.401)	-0.05	(0.984)
Lagged and baseline productive outcomes				
2016 hives (#)	-6.97	(0.253)	-0.16	(0.975)
2017 hives (#)	-7.05	(0.216)	-1.58	(0.770)
2018 hives (#)	-5.95	(0.264)	0.16	(0.977)
2019 hives (#)	3.83	(0.652)	7.94	(0.430)
2020 hives (#)	-4.57	(0.174)	-3.41	(0.415)
Ln(2020 hives)	-0.04	(0.726)	-0.04	(0.787)
2016 production (Kg.)	-13.49	(0.930)	96.05	(0.390)
2017 production (Kg.)	-23.84	(0.868)	59.31	(0.575)
2018 production (Kg.)	3.47	(0.969)	99.90	(0.253)
2019 production (Kg.)	16.24	(0.861)	91.80	(0.348)
2020 production (Kg.)	-59.97	(0.475)	-200.45	(0.239)
Ln(2020 production)	0.14	(0.230)	0.00	(0.989)
2020 yield (Kg./hive, capped at 100)	2.47	(0.127)	0.03	(0.988)
Ln(2020 yield)	0.15**	(0.034)	0.01	(0.836)
Baseline best practices outcomes				
Practices (0-9)	-0.20	(0.408)	0.10	(0.777)
Annual change of queen bees (= 1)	-0.13**	(0.037)	-0.05	(0.556)
Queen bee breeding (= 1)	-0.06	(0.233)	-0.01	(0.919)
Keeps record log (= 1)	-0.02	(0.114)	-0.04	(0.251)
Supplemented feeding (= 1)	-0.02	(0.713)	-0.04	(0.646)
Sustainable pest control (= 1)	0.08	(0.198)	0.14*	(0.100)
Good condition of brood chambers (= 1)	0.01	(0.834)	0.02	(0.707)
Good condition of honey suppers (= 1)	-0.02	(0.667)	-0.01	(0.877)
Good condition of outer covers (= 1)	-0.04	(0.536)	0.00	(1.000)
Good condition of brood frames (= 1)	0.00	(0.987)	0.08	(0.217)
Producers	274		201	

Notes: * p<0.10, ** p<0.05, *** p<0.01. Differences of treated minus control group means for each comparison, P-values in parentheses.

Table 7: Main results on performance – ATE

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
ATE (T-C)						
Treated=1 vs.	4.954***	0.203***	245.754***	0.303***	2.250*	0.108**
Control=1	(0.655)	(0.078)	(62.406)	(0.047)	(1.555)	(0.057)
Observations	261	261	261	260	261	260

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 8: Main results on adoption of practices index and queen-bee related individual practices – ATE

	(1) Practices (dummy)	(2) Practices (0-9)	(3) Change queen	(4) Breed queen	(5) Record log
ATE (T-C)					
Treated=1 vs.	0.059*	3.051***	0.281***	0.213***	0.533***
Control=1	(0.052)	(0.380)	(0.069)	(0.081)	(0.081)
Observations	261	261	261	261	261

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 9: Main results on adoption of other individual practices – ATE

	(1) Supplemented feeding	(2) Sustainable pest control	(3) Brood chambers	(4) Honey suppers	(5) Outer covers	(6) Brood frames
ATE (T-C)						
Treated=1 vs.	0.230**	0.188***	0.254***	0.275***	0.456***	0.621***
Control=1	(0.121)	(0.081)	(0.066)	(0.087)	(0.078)	(0.154)
Observations	261	261	261	261	261	261

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 10: Main results on performance – ATT

	(1)	(2)	(3)	(4)	(5)	(6)
	Hives	Ln(Hives)	Kg.	Ln(Kg.)	Yield	Ln(Yield)
Full Package impact (T3-C)						
Full Package=1 vs. Control=1	6.116***	0.243***	299.986***	0.333***	2.641	0.105
	(1.018)	(0.070)	(70.364)	(0.079)	(2.067)	(0.079)
Observations	121	121	121	121	121	121
Partial General Training Impact (T1-C)						
Partial General Training=1 vs. Control=1	0.279	0.053	23.975	0.182	1.362	0.143
	(1.468)	(0.061)	(58.242)	(0.174)	(1.904)	(0.142)
	60	60	60	59	60	59
Queen Bee impact (T2-T1)						
Queen Bee=1 vs. Partial General Training=1	5.551***	0.175	310.899**	0.265	2.824	0.071
	(1.543)	(0.084)	(123.278)	(0.138)	(2.048)	(0.150)
	140	140	140	139	140	139
Breeding Training impact (T3-T2)						
Full Package=1 vs. Queen Bee=1	0.842	0.034	-10.493	0.043	0.684	0.028
	(1.637)	(0.083)	(145.235)	(0.103)	(1.370)	(0.049)
	201	201	201	201	201	201

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 11: Main results on adoption of practices index and queen-bee related individual practices – ATT

	(1) Practices (dummy)	(2) Practices (0-9)	(3) Change queen	(4) Breed queen	(5) Record log
Full Package impact (T3-C)					
Full Package=1 vs. Control=1	0.012	3.923***	0.489***	0.462***	0.779***
	(0.056)	(0.504)	(0.087)	(0.106)	(0.088)
Observations	121	121	121	121	121
Partial General Training impact (T1-C)					
Partial General Training=1 vs. Control=1	0.190**	1.501***	-0.007	-0.093**	0.070*
	(0.081)	(0.410)	(0.069)	(0.042)	(0.049)
	60	60	60	60	60
Queen Bee impact (T2-T1)					
Queen Bee=1 vs. Partial General Training=1	-0.079	1.706**	0.214***	0.171**	0.434***
	(0.074)	(0.699)	(0.078)	(0.081)	(0.135)
	140	140	140	140	140
Breeding Training impact (T3-T2)					
Full Package=1 vs. Queen Bee=1	-0.021	1.513***	0.218**	0.341***	0.425***
	(0.061)	(0.556)	(0.120)	(0.105)	(0.108)
	201	201	201	201	201

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 12: Main results on adoption of other individual practices – ATT

	(1) Supplemented feeding	(2) Sustainable pest control	(3) Brood chambers	(4) Honey suppers	(5) Outer covers	(6) Brood frames
Full Package impact (T3-C)						
Full Package=1 vs. Control=1	0.415**	0.192**	0.278**	0.303**	0.415***	0.591***
	(0.151)	(0.091)	(0.095)	(0.106)	(0.107)	(0.141)
Observations	121	121	121	121	121	121
Partial General Training impact (T1-C)						
Partial General Training=1 vs. Control=1	-0.055	0.075	0.193***	0.222**	0.491***	0.606***
	(0.075)	(0.095)	(0.045)	(0.083)	(0.121)	(0.101)
	60	60	60	60	60	60
Queen Bee impact (T2-T1)						
Queen Bee=1 vs. Partial General Training=1	0.104	0.137	0.180	0.197	0.111	0.159
	(0.085)	(0.215)	(0.146)	(0.170)	(0.113)	(0.145)
	140	140	140	140	140	140
Breeding Training impact (T3-T2)						
Full Package=1 vs. Queen Bee=1	0.285**	0.018	0.060	0.092	0.072	0.004
	(0.112)	(0.090)	(0.084)	(0.066)	(0.105)	(0.084)
	201	201	201	201	201	201

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 13: Heterogeneous effects by gender

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
Women = 1	1.715**	0.122*	-64.699*	0.030	-1.226	-0.090*
	(0.580)	(0.056)	(36.437)	(0.053)	(0.903)	(0.041)
Observations	261	261	261	260	261	260

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 14: Heterogeneous effects by producer size and exposition to floods

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
Panel A: High vs low initial production						
High initial production=1	3.607	-0.035	184.478	-0.181	-2.301	-0.140
	(1.770)	(0.082)	(121.832)	(0.140)	(1.922)	(0.081)
Observations	261	261	261	260	261	260
Panel B: High vs low initial loss of hives						
High initial loss of hives=1	1.935	0.044	227.857	0.108	1.870	0.076
	(1.464)	(0.150)	(80.680)	(0.134)	(1.515)	(0.059)
	261	261	261	260	261	260

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 15: Incremental effects of basic inputs and queen bees delivered

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
Panel A: Basic inputs						
Ln(Basic inputs delivered – USD)	2.719	0.091	155.585*	0.204	2.483	0.099
	(1.372)	(0.051)	(51.528)	(0.123)	(1.103)	(0.098)
Observations	261	261	261	260	261	260
Panel B: Basic inputs per hive						
Ln(Basic inputs delivered per hive– USD/hive)	-1.719	0.013	7.540	0.138	3.265**	0.125*
	(3.274)	(0.116)	(90.046)	(0.132)	(0.718)	(0.042)
	261	261	261	260	261	260
Panel C: Queen bees						
Ln(Queen bees delivered - # + 1)	1.833	0.054	120.345	0.139	2.040*	0.079
	(1.339)	(0.038)	(49.922)	(0.081)	(0.660)	(0.059)
	261	261	261	260	261	260

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 16: Cost-effectiveness calculations

Program component	(1) Budget (US$)	(2) Producers receiving each (#)	(3) Cost per producer (US$)*	(4) Effect on production per producer (kg.)	(5) Effect on sales per producer (US$2.5 per kg.)	**(6)=1*(4) /(3) Increase in production per US$1 (kg)**	**(7)=1*(5) /(3) Increase in honey sales per US$1 (US$)**
Basic Inputs	$141,076	356	$396	-	-	**-**	**-**
Partial general training	$36,030	335	$504	24.0	$59.9	**0.05**	**$0.12**
Queen bee	$17,758	274	$569	310.9	$777.2	**0.55**	**$1.37**
Breeding training	$5,284	152	$603	0.0	-	**-**	**-**
Full package	$200,148		$603	300.0	$750.0	**0.5**	**$1.24**

Notes: *The cost per producer for each program component considers that the producer has also received the previous program components, e.g., a producer receiving the full package received $396 in basic inputs (total budget for that component, $141,076 divided by producers receiving that component, 356), $108 in general training, $65 in queen bees, and $35 in breeding training, adding up to $603. Effects on production are obtained from our entropy balancing results. No effect is assumed for the breeding training component (coefficient was negative but not significant). Effects on sales are calculated using the average price received by beekeepers in the 2022 campaign (US$2.5 per kg. of honey). Basic inputs budget also includes US$230 in inputs delivered directly to the bulking centers.

Table 17: Comparison of entropy balancing results with PSM and IPW – ATE

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
Panel A: Entropy						
Treated=1 vs. Control=1	4.954***	0.203***	245.754***	0.303***	2.250*	0.108**
	(0.655)	(0.078)	(62.406)	(0.047)	(1.555)	(0.057)
Observations	261	261	261	260	261	260
Panel B: PSM						
Treated=1 vs. Control=1	13.057**	0.699**	377.736**	0.778**	2.157	0.089*
	(4.556)	(0.249)	(184.030)	(0.299)	(2.130)	(0.054)
	286	286	286	286	286	286
Panel C: IPW						
Treated=1 vs. Control=1	9.884***	0.535***	267.549***	0.614***	1.224*	0.089**
	(1.418)	(0.091)	(60.840)	(0.125)	(1.321)	(0.047)
	316	316	316	315	316	315

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing, PSM and IPW; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table 18: Alternative comparisons between groups

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
Queen Bee vs Control (T2-C)						
Queen Bee=1 vs. Control=1	5.902***	0.239**	279.387***	0.330***	1.981*	0.089*
	(1.058)	(0.116)	(116.989)	(0.079)	(1.850)	(0.094)
Observations	114	114	114	114	114	114
Full Package vs Partial General Training (T3-T1)						
Full Package=1 vs. Partial General Training=1	8.037***	0.337***	411.962***	0.522***	5.510***	0.199***
	(1.259)	(0.049)	(65.571)	(0.081)	(1.286)	(0.063)
	147	147	147	146	147	146

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Figures

Figure 1: Program Timeline

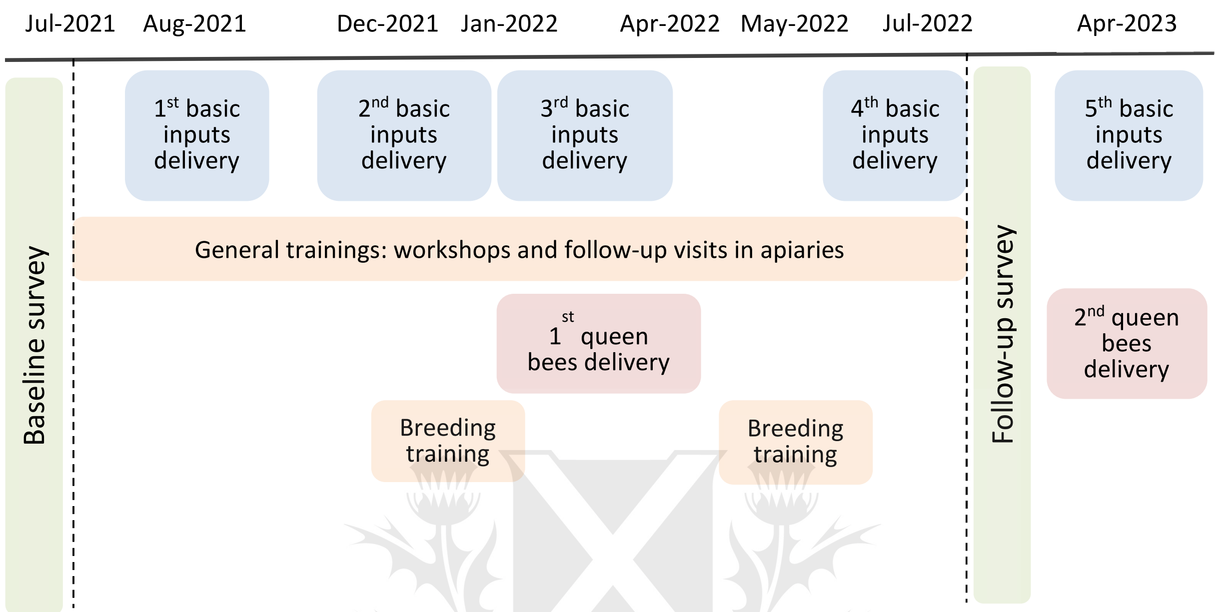

Figure 2: Mexico's Honey Production

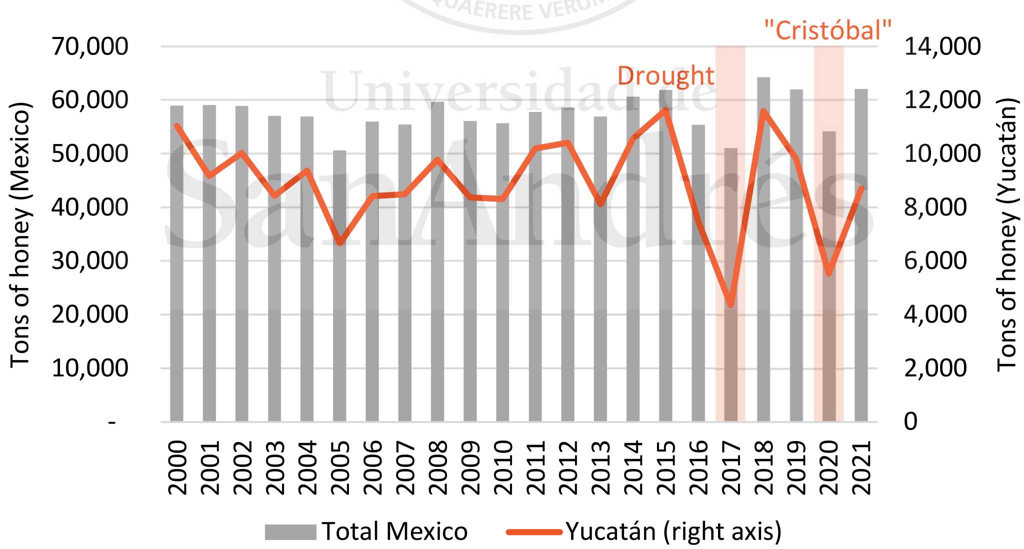

Figure 3: Vertical Logic

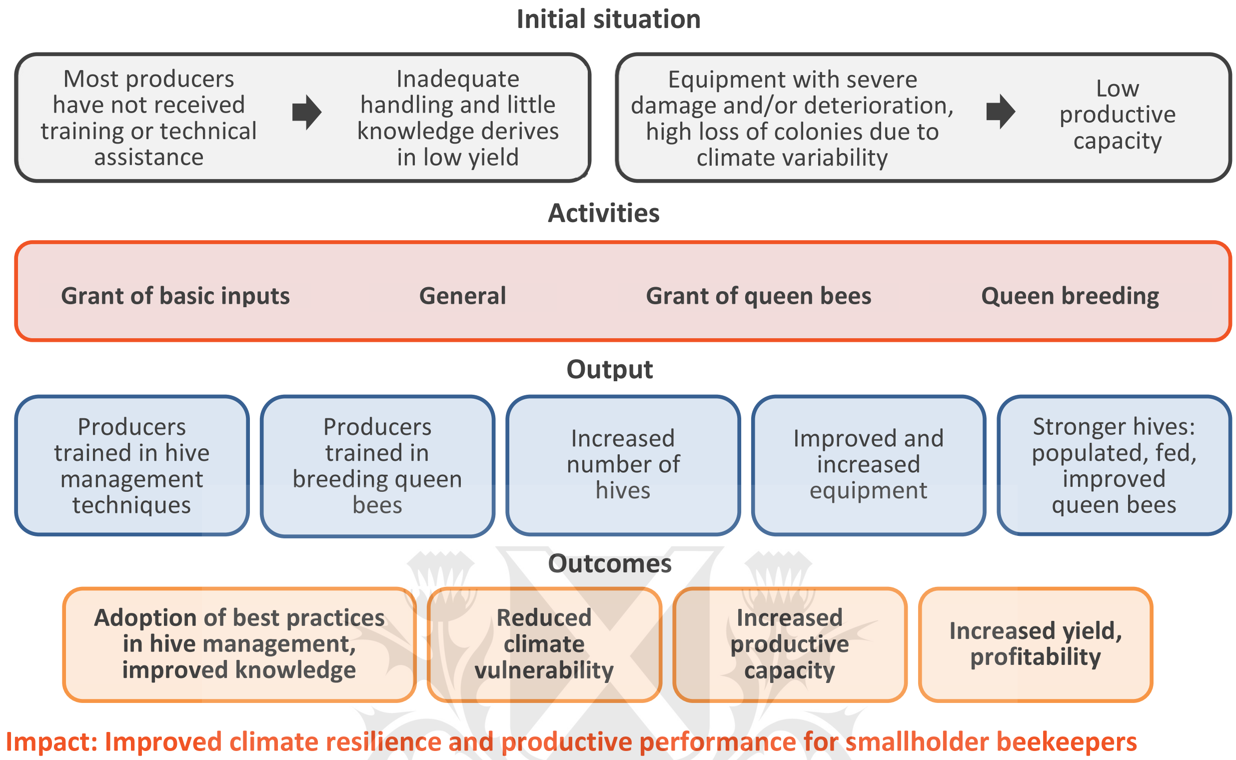

Figure 4: Program Beneficiaries

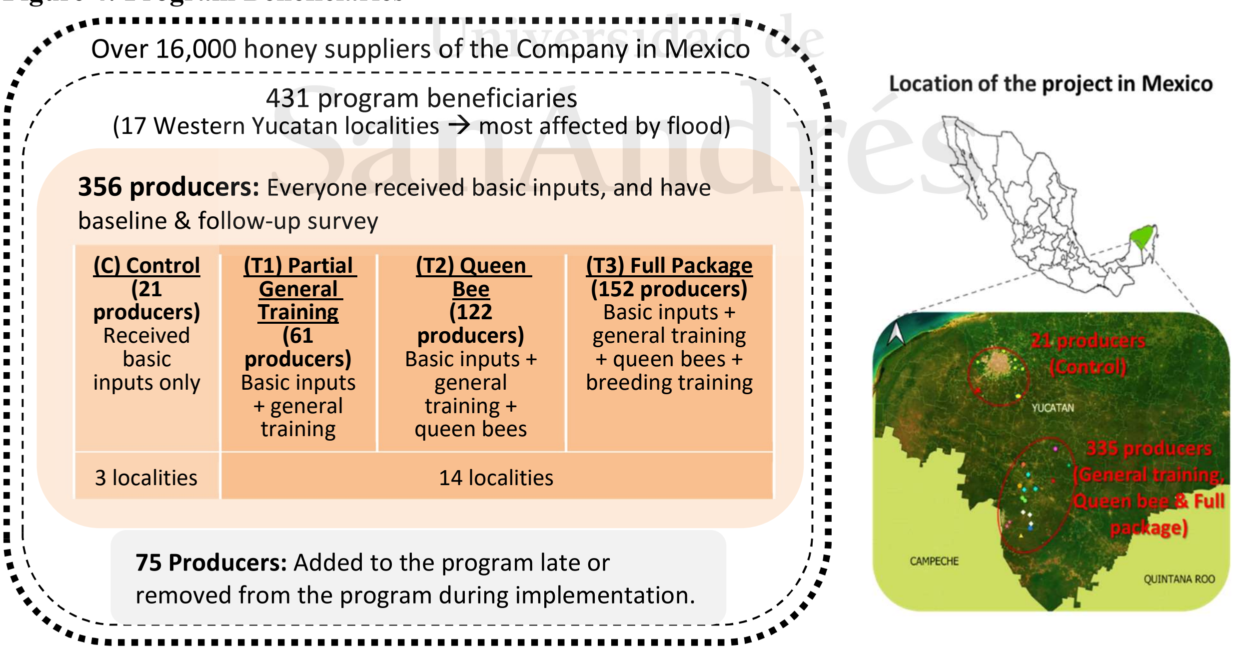

Figure 5: Treatment Groups and Program Components

		Treatment Groups (N=356)			
		C: Control (N=21)	T: Treated (335)		
			T1: Partial General Training (N=61)	T2: Queen Bee (N=122)	T3: Full Package (N=152)
Program Components	**Basic Inputs**				
	Partial general training		**T1 - C:** Impact of partial general training		
	Complete general training and queen-bee asset			**T2 - T1:** Impact of complete general training + queen-bee	
	Queen breeding training				**T3 - T2:** Impact of queen breeding training

References

Abman, R., & Carney, C. (2020). Agricultural productivity and deforestation: Evidence from input subsidies and ethnic favoritism in Malawi. *Journal of Environmental Economics and Management*, *103*, 102342. https://doi.org/10.1016/j.jeem.2020.102342

Ahmad, T., Shah, G.-M., Ahmad, F., Partap, U., & Ahmad, S. (2017). Impact of Apiculture on the Household Income of Rural Poor in Mountains of Chitral District in Pakistan. *Journal of Social Sciences (COES&RJ-JSS)*, *6*(3), 518–531.

Anderson, M. (n.d.). *Stata code to compute "sharpened" False Discovery Rate (FDR) adjusted q-values*. Retrieved September 20, 2023, from https://are.berkeley.edu/~mlanderson/downloads/fdr_sharpened_qvalues.do.zip

Anderson, M. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495. https://doi.org/10.1198/016214508000000841

Banerjee, A., Karlan, D., Osei, R. D., Trachtman, H., & Udry, C. (2022). *Unpacking a Multi-Faceted Program to Build Sustainable Income for the Very Poor*.

Baul, T., Karlan, D., Toyama, K., & Vasilaky, K. (2023). Improving Smallholder Agriculture via Video-Based Group Extension. *IPR Working Paper Series*, *23*(06). https://www.ipr.northwestern.edu/documents/working-papers/2023/wp-23-06.pdf

Blattman, C., Fiala, N., & Martinez, S. (2014). Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda*. *The Quarterly Journal of Economics*, *129*(2), 697–752. https://doi.org/10.1093/qje/qjt057

Blog Apicultura y Miel. (2022, February 20). Cambiar la reina de las colmenas: Cuándo y cómo debe hacerse. *Blog Apicultura y Miel*. https://apiculturaymiel.com/abejas/cuando-como-cambiar-la-reina-de-las-colmenas-para-lograr-productividad/

Carter, M., Laajaj, R., & Yang, D. (2021). Subsidies and the African Green Revolution: Direct Effects and Social Network Spillovers of Randomized Input Subsidies in Mozambique. *American Economic Journal: Applied Economics*, *13*(2), 206–229. https://doi.org/10.1257/app.20190396

Carter, M. R., Laajaj, R., & Yang, D. (2013). The Impact of Voucher Coupons on the Uptake of Fertilizer and Improved Seeds: Evidence from a Randomized Trial in Mozambique. *American Journal of Agricultural Economics Papers and Proceedings*.

Contreras-Uc, L. C., Magaña-Magaña, M. A., Sanginés-García, J. R., Contreras-Uc, L. C., Magaña-Magaña, M. A., & Sanginés-García, J. R. (2018). Características técnicas y socioeconómicas de la apicultura en comunidades mayas del Litoral Centro de Yucatán. *Acta universitaria*, *28*(1), 77–86. https://doi.org/10.15174/au.2018.1390

Danida. (2004). *Farm Women in Development. Impact Study of Four Training Projects in India*.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, *47*(2), 315–372. https://doi.org/10.1257/jel.47.2.315

Dequenne, I., Philippart de Foy, J.-M., & Cani, P. D. (2022). Developing Strategies to Help Bee Colony Resilience in Changing Environments. *Animals*, *12*(23), Article 23. https://doi.org/10.3390/ani12233396

Dini, C., & Bedascarrasbure, E. (2011). *Manual de Apicultura para ambientes subtropicales* (1a. ed.). Ediciones INTA.

Duflo, E., Kremer, M., & Robinson, J. (2011). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *American Economic Review*, *101*(6), 2350–2390.

Food and Agriculture Organization of the United Nations. (2018). *Why bees matter: The importance of bees and other pollinators for food and agriculture*. FAO. https://www.fao.org/documents/card/en/c/i9527en

Food and Agriculture Organization of the United Nations. (2023). *FAOSTAT statistical database* [dataset]. https://www.fao.org/faostat/en/#data

Galetto, L., Aizen, M. A., Arizmendi, M. del C., Freitas, B. M., & et al. (2022). Risks and opportunities associated with pollinators' conservation and management of pollination services in Latin America. *Ecología Austral*, *32*, 55–76.

Garone, L. F., Maffioli, A., de Negri, J. A., Rodriguez, C. M., & Vázquez-Baré, G. (2015). Cluster development policy, SME's performance, and spillovers: Evidence from Brazil. *Small Business Economics*, *44*(4), 925–948.

Gignoux, J., Macours, K., Stein, D., & Wright, K. (2023). Input subsidies, credit constraints, and expectations of future transfers: Evidence from Haiti. *American Journal of Agricultural Economics*, *105*(3), 809–835. https://doi.org/10.1111/ajae.12337

Gobierno de México. (n.d.). *Abejas y apicultura*. Retrieved August 10, 2023, from https://atlas-abejas.agricultura.gob.mx/cap2.html#26_Acciones_y_Programas_Gubernamentales_en_Pro_de_la_Apicultura

Gobierno de México. (2019). *Manual de Buenas Prácticas Pecuarias en la Producción de Miel* (4ta ed.).

Grimm, M., & Luck, N. (2020). Can Training Enhance Adoption, Knowledge and Perception of Organic Farming Practices? Evidence from a Randomized Experiment in Indonesia. *SSRN*. https://developmentevidence.3ieimpact.org/search-result-details/impact-evaluationrepository/can-training-enhance-adoption-knowledge-and-perception-of-organic-farming-practices-evidence-from-a-randomized-experiment-in-indonesia/14225

Guzmán-Novoa, E., Benítez, A. C., Montaño, L. G. E., & Novoa, G. G. (2011). Colonization, impact and control of Africanized honey bees in Mexico. *Veterinaria México*, *42*(2), 149–178.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, *20*(1), 25–46. https://doi.org/10.1093/pan/mpr025

Hainmueller, J., & Xu, Y. (2013). ebalance: A Stata Package for Entropy Balancing. *Journal of Statistical Software*, *54*, 1–18. https://doi.org/10.18637/jss.v054.i07

Hemming, D. J., Chirwa, E. W., Dorward, A., Ruffhead, H. J., Hill, R., Osborn, J., Langer, L., Harman, L., Asaoka, H., Coffey, C., & Phillips, D. (2018). Agricultural input subsidies for improving productivity, farm income, consumer welfare and wider growth in low- and lower-middle-income countries: A systematic review. *Campbell Systematic Reviews*, *14*(1), 1–153. https://doi.org/10.4073/csr.2018.4

Hörner, D., Bouguen, A., Frölich, M., & Wollni, M. (2021). Knowledge and Adoption of Complex Agricultural Technologies: Evidence from an Extension Experiment. *The World Bank Economic Review*, *36*(1), 68–90. https://doi.org/10.1093/wber/lhab025

Jann, B. (2020). Influence functions continued. A framework for estimating standard errors in reweighting, matching, and regression adjustment. *University of Bern Social Sciences Working Papers*, Article 35. https://ideas.repec.org//p/bss/wpaper/35.html

Jayne, T. S., Mason, N. M., Burke, W. J., & Ariga, J. (2018). Review: Taking stock of Africa's second-generation agricultural input subsidy programs. *Food Policy*, *75*, 1–14. https://doi.org/10.1016/j.foodpol.2018.01.003

Kondylis, F., Mueller, V., & Zhu, J. (2017). Seeing is believing? Evidence from an extension network experiment. *Journal of Development Economics*, *125*, 1–20.

Kovačić, M., Puškadija, Z., Dražić, M. M., Uzunov, A., Meixner, M. D., & Büchler, R. (2020). Effects of selection and local adaptation on resilience and economic suitability in Apis mellifera carnica. *Apidologie*, *51*(6), 1062–1073. https://doi.org/10.1007/s13592-020-00783-0

Landaverde, R., Rodriguez, M. T., & Parrella, J. A. (2023). Honey Production and Climate Change: Beekeepers' Perceptions, Farm Adaptation Strategies, and Information Needs. *Insects*, *14*(6), Article 6. https://doi.org/10.3390/insects14060493

Mason, N., & Tembo, S. (2015). Do Input Subsidies reduce Poverty among Smallholder Farm Households? Panel Survey Evidence from Zambia. *2015 Conference, August 9-14, 2015, Milan, Italy*, Article 212232. https://ideas.repec.org//p/ags/iaae15/212232.html

McKenzie, D. (2020, June 1). *An overview of multiple hypothesis testing commands in Stata*. https://blogs.worldbank.org/impactevaluations/overview-multiple-hypothesis-testing-commands-stata

McKenzie, D., & Woodruff, C. (2014). What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World? *The World Bank Research Observer*, *29*(1), 48–82. https://doi.org/10.1093/wbro/lkt007

Nankhuni, F., & Paniagua, G. (2013). *Meta-evaluation of private sector interventions in agribusiness: Finding out what worked in access to finance and farmer or business training*. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/353231500640835766/Meta-evaluation-of-private-sector-interventions-in-agribusiness-finding-out-what-worked-in-access-to-finance-and-farmer-or-business-training

Neumann, P., & Straub, L. (2023). Beekeeping under climate change. *Journal of Apicultural Research*, *62*(5), 963–968. https://doi.org/10.1080/00218839.2023.2247115

Ravallion, M., van deWalle, D., Dutta, P., & Murgai, R. (2015). Empowering poor people through public information? Lessons from a movie in rural india. *Journal of Public Economics*, *132*, 13–22.

Ricker-Gilbert, J., & Jayne, T. S. (2017). Estimating the Enduring Effects of Fertiliser Subsidies on Commercial Fertiliser Demand and Maize Production: Panel Data Evidence from Malawi. *Journal of Agricultural Economics*, *68*(1), 70–97. https://doi.org/10.1111/1477-9552.12161

Ruprah, I., & Marcano, L. (2009). Does technical assistance matter? An impact evaluation approach to estimate its value added. *Journal of Development Effectiveness*, *1*(4), 507–528. https://doi.org/10.1080/19439340903370451

Schouten, C. (2020). Factors influencing beekeepers income, productivity and welfare in developing countries: A scoping review. *Journal of Apicultural Research*. https://doi.org/10.1080/00218839.2020.1844464

Schouten, C., & John Lloyd, D. (2019). Considerations and Factors Influencing the Success of Beekeeping Programs in Developing Countries. *Bee World*, *96*(3), 75–80. https://doi.org/10.1080/0005772X.2019.1607805

Seagle, J. (2008). *Agricultural Certifications and Beekeeping: Lessons from an Apicultural Cooperative in Northeastern El Salvador, Central America*. https://scholarworks.umt.edu/etd/1247

Torres Franco, N. A., Dávalos, E., & Morales, L. F. (2021). Heterogeneous Effects of Agricultural Technical Assistance in Colombia. *Journal of Agricultural and Applied Economics*, *53*(4), 459–481. https://doi.org/10.1017/aae.2021.18

Tübbicke, S. (2022). Entropy Balancing for Continuous Treatments. *Journal of Econometric Methods*, *11*(1), 71–89. https://doi.org/10.1515/jem-2021-0002

Van Espen, M., Williams, J. H., Alves, F., Hung, Y., de Graaf, D. C., & Verbeke, W. (2023). Beekeeping in Europe facing climate change: A mixed methods study on perceived impacts and the need to adapt according to stakeholders and beekeepers. *Science of The Total Environment*, *888*, 164255. https://doi.org/10.1016/j.scitotenv.2023.164255

Vercelli, M., Novelli, S., Ferrazzi, P., Lentini, G., & Ferracini, C. (2021). A Qualitative Analysis of Beekeepers' Perceptions and Farm Management Adaptations to the Impact of Climate Change on Honey Bees. *Insects*, *12*(3). https://doi.org/10.3390/insects12030228

Woldewahid, G., Gebremedhin, B., Hoekstra, D., Tegegne, A., Berhe, K., & Weldemariam, D. (2012). *Market-oriented beekeeping development to improve smallholder income: Results of development experiences in Atsbi-Womberta District, northern Ethiopia*. https://cgspace.cgiar.org/handle/10568/21583

Wonde, K. M., Tsehay, A. S., & Lemma, S. E. (2022). Training at farmers training centers and its impact on crop productivity and households' income in Ethiopia: A propensity score matching (PSM) analysis. *Heliyon*, *8*(7), e09837. https://doi.org/10.1016/j.heliyon.2022.e09837

Wossen, T., Abdoulaye, T., Alene, A., Feleke, S., Ricker-Gilbert, J., Manyong, V., & Awotide, B. A. (2017). Productivity and Welfare Effects of Nigeria's e-Voucher-Based Input Subsidy Program. *World Development*, *97*, 251–265. https://doi.org/10.1016/j.worlddev.2017.04.021

Zhao, Q., & Percival, D. (2017). Entropy Balancing is Doubly Robust. *Journal of Causal Inference*, *5*(1). https://doi.org/10.1515/jci-2016-0010

Annex

Annex 1: Additional tables

Table A 1: Difference-in-Differences on performance

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
Full Package impact (T3-C)						
Post=1 x Full Package=1 vs. Post=1 x Control=1	8.728***	0.448***	348.129***	0.520***	2.519***	0.089**
	(1.362)	(0.066)	(48.017)	(0.079)	(0.974)	(0.035)
Observations	712	712	712	710	712	710
Partial General Training impact (T1-C)						
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	3.147**	0.241***	27.675	0.356***	0.037	0.126*
	(1.591)	(0.076)	(44.400)	(0.101)	(1.056)	(0.069)
	712	712	712	710	712	710
Queen Bee impact (T2-T1)						
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	5.484***	0.177***	308.122***	0.141	1.779	-0.049
	(1.259)	(0.066)	(53.575)	(0.096)	(1.142)	(0.078)
	712	712	712	710	712	710
Breeding Training impact (T3-T2)						
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.097	0.030	12.332	0.023	0.703	0.012
	(0.955)	(0.054)	(56.609)	(0.072)	(1.066)	(0.050)
	712	712	712	710	712	710

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences; robust SE in parentheses.

Table A 2: Before and After for adoption of practices index and queen bee-related individual practices

	(1) Practices (dummy)	(2) Practices (0-9)	(3) Change queen	(4) Breed queen	(5) Record log
All producers					
Post=1 x All=1	0.101***	3.890***	0.298***	0.236***	0.528***
	(0.017)	(0.153)	(0.035)	(0.035)	(0.027)
Observations	712	712	712	712	712
Control group (C)					
Post=1 x Control=1	0.143*	1.381**	-0.095	0.000	0.000
	(0.078)	(0.663)	(0.157)	(0.135)	(0.093)
	42	42	42	42	42
Treated group (T)					
Post=1 x Treated=1	0.099***	4.048***	0.322***	0.251***	0.561***
	(0.017)	(0.155)	(0.036)	(0.036)	(0.027)
	670	670	670	670	670
Partial General Training group (T1)					
Post=1 x Partial General Training=1	0.164***	1.934***	0.016	-0.082	0.066**
	(0.056)	(0.375)	(0.085)	(0.056)	(0.032)
	122	122	122	122	122
Queen Bee group (T2)					
Post=1 x Queen bee=1	0.074***	3.828***	0.238***	0.107*	0.549***
	(0.024)	(0.231)	(0.059)	(0.061)	(0.047)
	244	244	244	244	244
Full Package group (T3)					
Post=1 x Full Package=1	0.092***	5.072***	0.513***	0.500***	0.770***
	(0.024)	(0.189)	(0.045)	(0.050)	(0.034)
	304	304	304	304	304

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on before-and-after comparison; robust SE in parentheses.

Table A 3: Before and After for adoption of other individual practices

	(1) Supplemented feeding	(2) Sustainable pest control	(3) Brood chambers	(4) Honey suppers	(5) Outer covers	(6) Brood frames
All producers						
Post=1 x All=1	0.160***	0.213***	0.699***	0.669***	0.441***	0.646***
	(0.035)	(0.034)	(0.026)	(0.027)	(0.030)	(0.028)
Observations	712	712	712	712	712	712
Control group (C)						
Post=1 x Control=1	0.000	0.095	0.476***	0.381***	0.238*	0.286**
	(0.156)	(0.142)	(0.137)	(0.141)	(0.141)	(0.136)
	42	42	42	42	42	42
Treated group (T)						
Post=1 x Treated=1	0.170***	0.221***	0.713***	0.687***	0.454***	0.669***
	(0.036)	(0.035)	(0.027)	(0.027)	(0.030)	(0.028)
	670	670	670	670	670	670
Partial General Training group (T1)						
Post=1 x Partial General Training=1	-0.049	0.066	0.557***	0.475***	0.344***	0.541***
	(0.060)	(0.088)	(0.075)	(0.080)	(0.084)	(0.077)
	122	122	122	122	122	122
Queen Bee group (T2)						
Post=1 x Queen Bee=1	0.123**	0.246***	0.738***	0.705***	0.443***	0.680***
	(0.061)	(0.059)	(0.042)	(0.044)	(0.048)	(0.045)
	244	244	244	244	244	244
Full Package group (T3)						
Post=1 x Full Package=1	0.296***	0.263***	0.757***	0.757***	0.507***	0.711***
	(0.054)	(0.047)	(0.036)	(0.036)	(0.041)	(0.038)
	304	304	304	304	304	304

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on before-and-after comparison; robust SE in parentheses.

Table A 4: Difference-in-Differences for adoption of practices index and queen bee-related individual practices

	(1) Practices (dummy)	(2) Practices (0-9)	(3) Change queen	(4) Breed queen	(5) Record log
Full Package impact (T3-C)					
Post=1 x Full Package=1 vs. Post=1 x Control=1	-0.051	3.691***	0.608***	0.500***	0.770***
	(0.080)	(0.499)	(0.077)	(0.050)	(0.034)
Observations	712	712	712	712	712
Partial General Training impact (T1-C)					
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	0.021	0.553	0.112	-0.082*	0.066**
	(0.093)	(0.566)	(0.081)	(0.042)	(0.032)
	712	712	712	712	712
Queen Bee impact (T2-T1)					
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	-0.090	1.893***	0.221***	0.189***	0.484***
	(0.058)	(0.380)	(0.069)	(0.067)	(0.055)
	712	712	712	712	712
Breeding Training impact (T3-T2)					
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.018	1.244***	0.275***	0.393***	0.221***
	(0.034)	(0.270)	(0.064)	(0.072)	(0.057)
	712	712	712	712	712

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences; robust SE in parentheses.

Table A 5: Difference-in-Differences for adoption of other individual practices

	(1) Supplemented feeding	(2) Sustainable pest control	(3) Brood chambers	(4) Honey suppers	(5) Outer covers	(6) Brood frames
Full Package impact (T3-C)						
Post=1 x Full Package=1 vs. Post=1 x Control=1	0.296***	0.168**	0.280**	0.376***	0.268**	0.425***
	(0.108)	(0.079)	(0.115)	(0.131)	(0.122)	(0.143)
Observations	712	712	712	712	712	712
Partial General Training impact (T1-C)						
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	-0.049	-0.030	0.081	0.094	0.106	0.255
	(0.105)	(0.089)	(0.129)	(0.149)	(0.146)	(0.156)
	712	712	712	712	712	712
Queen Bee impact (T2-T1)						
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	0.172***	0.180**	0.180**	0.230**	0.098	0.139
	(0.066)	(0.079)	(0.080)	(0.090)	(0.101)	(0.085)
	712	712	712	712	712	712
Breeding Training impact (T3-T2)						
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.173**	0.017	0.019	0.052	0.064	0.030
	(0.070)	(0.068)	(0.055)	(0.056)	(0.062)	(0.059)
	712	712	712	712	712	712

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences; robust SE in parentheses.

Table A 6: Heterogeneous effects by gender – alternative specification

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
Women = 1	-9.969	-0.430	-163.830*	-0.356	1.213	0.021
	(8.980)	(0.377)	(53.906)	(0.263)	(3.956)	(0.146)
Observations	261	261	261	260	261	260

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table A 7: Balance test of ex-ante covariates for continuous treatment variables

Panel A: Ln(Basic inputs delivered - USD)

Variable	Before balancing		EBCT	
	Dif.	P-value	Dif.	P-value
Age	-0.86	(0.450)	0.49	(0.731)
Post-primary education (= 1)	0.07**	(0.044)	0.00	(1.000)
Women (= 1)	0.04	(0.153)	0.00	(1.000)
Hives lost due to floods (#)	0.23	(0.911)	0.00	(1.000)
2016 hives (#)	11.64***	(0.002)	0.00	(1.000)
2017 hives (#)	11.59***	(0.001)	1.59	(0.666)
2018 hives (#)	11.77***	(0.000)	3.91	(0.285)
2019 hives (#)	14.03***	(0.005)	5.08	(0.227)
2020 hives (#)	7.56***	(0.000)	0.95	(0.809)
Ln(2020 hives)	0.36***	(0.000)	0.01	(0.911)
2016 production (Kg.)	206.35**	(0.039)	27.18	(0.727)
2017 production (Kg.)	190.04**	(0.042)	72.75	(0.299)
2018 production (Kg.)	189.03***	(0.006)	90.68	(0.168)
2019 production (Kg.)	193.41***	(0.006)	66.77	(0.409)
2020 production (Kg.)	209.12***	(0.000)	3.77	(0.957)
Ln(2020 production)	0.65***	(0.000)	0.00	(1.000)
2020 yield (Kg./hive, capped at 100)	4.96***	(0.000)	0.00	(1.000)
Ln(2020 yield)	0.28***	(0.000)	-0.02	(0.850)
Producers	356		261	

Panel B: Ln(Basic inputs delivered per hive - USD/hive)

Variable	Before balancing		EBCT	
	Dif.	P-value	Dif.	P-value
Age	-1.96**	(0.024)	0.30	(0.856)
Post-primary education (= 1)	0.04	(0.175)	0.00	(1.000)
Women (= 1)	0.05**	(0.035)	0.00	(1.000)
Hives lost due to floods (#)	-8.16***	(0.000)	-0.00	(1.000)
2016 hives (#)	-17.59***	(0.000)	-0.00	(1.000)
2017 hives (#)	-17.00***	(0.000)	-0.30	(0.950)
2018 hives (#)	-16.09***	(0.000)	-1.77	(0.686)
2019 hives (#)	-20.27***	(0.000)	-2.86	(0.530)
2020 hives (#)	-16.79***	(0.000)	-3.29	(0.125)
Ln(2020 hives)	-0.63***	(0.000)	-0.03	(0.753)
2016 production (Kg.)	-353.23***	(0.000)	-127.59	(0.277)
2017 production (Kg.)	-336.10***	(0.000)	-102.17	(0.352)
2018 production (Kg.)	-222.74***	(0.000)	-33.50	(0.732)
2019 production (Kg.)	-232.54***	(0.000)	-17.92	(0.869)
2020 production (Kg.)	-249.18***	(0.000)	-165.32	(0.125)
Ln(2020 production)	-0.36***	(0.000)	-0.00	(1.000)
2020 yield (Kg./hive, capped at 100)	3.29***	(0.000)	0.00	(1.000)
Ln(2020 yield)	0.23***	(0.000)	0.03	(0.677)

Producers	356		261	

Panel C: Ln(Queen bees delivered - # + 1)

Variable	Before balancing		EBCT	
	Dif.	P-value	Dif.	P-value
Age	-2.57***	(0.003)	0.15	(0.876)
Post-primary education (= 1)	0.05	(0.107)	0.00	(1.000)
Women (= 1)	0.03	(0.146)	-0.00	(1.000)
Hives lost due to floods (#)	-1.71	(0.269)	-0.00	(1.000)
2016 hives (#)	3.59	(0.207)	0.00	(1.000)
2017 hives (#)	3.96	(0.138)	-0.49	(0.900)
2018 hives (#)	3.17	(0.208)	-0.43	(0.911)
2019 hives (#)	5.36	(0.162)	0.23	(0.956)
2020 hives (#)	1.81	(0.252)	0.54	(0.805)
Ln(2020 hives)	0.08	(0.123)	0.01	(0.842)
2016 production (Kg.)	9.34	(0.901)	-88.91	(0.386)
2017 production (Kg.)	12.90	(0.853)	-85.64	(0.398)
2018 production (Kg.)	-36.62	(0.486)	-94.26	(0.320)
2019 production (Kg.)	-48.41	(0.371)	-125.92	(0.209)
2020 production (Kg.)	84.88**	(0.030)	-14.84	(0.718)
Ln(2020 production)	0.28***	(0.000)	0.00	(1.000)
2020 yield (Kg./hive, capped at 100)	3.04***	(0.000)	0.00	(1.000)
Ln(2020 yield)	0.20***	(0.000)	-0.02	(0.726)
Producers	356		261	

Notes: * p<0.10, ** p<0.05, *** p<0.01. P-values in parentheses.

Table A 8: Incremental effects of basic inputs and queen bees delivered – alternative specification

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
Panel A: Basic inputs						
Basic inputs delivered – USD	0.008	0.000	0.407	0.001	0.008	0.000
	(0.005)	(0.000)	(0.289)	(0.001)	(0.004)	(0.000)
Observations	261	261	261	260	261	260
Panel B: Basic inputs per hive						
Basic inputs delivered per hive – USD/hive	-0.054	0.002	-0.168	0.003	0.091	0.001
	(0.075)	(0.001)	(3.100)	(0.004)	(0.045)	(0.004)
	261	261	261	260	261	260
Panel C: Queen bees						
Queen bees delivered - #	0.539	0.014	25.847	0.038	0.534	0.021
	(0.287)	(0.007)	(21.688)	(0.019)	(0.230)	(0.017)
	261	261	261	260	261	260

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Differences-in-Differences combined with Entropy Balancing; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table A 9: Budget detail

	Total budget (431 producers, US$)			Budget in evaluation (356 producers, US$)		
Program component	**Total**	Before follow-up survey	After follow-up survey	**Total**	Before follow-up survey	After follow-up survey
Basic inputs	**$189,373**	$164,081	$25,292	**$163,895**	$141,076	$22,818
Partial general training*	**$43,620**	$43,620	-	**$36,030**	$36,030	-
Queen bee	**$42,038**	$20,594	$21,444	**$36,370**	$17,758	$18,612
Breeding training*	**$6,250**	$6,250	-	**$5,284**	$5,284	-
Total**	**$281,281**	$234,545	$46,736	**$241,579**	$200,148	$41,430

* Approximate. Total technical assistance cost for 431 producers was $47,317. Since there were 354 general trainings and 30 breeding trainings, cost was proportionately split between the two components. To calculate the budget for 356 producers, the total budget was multiplied by 356/431. The breeding training also includes the cost of materials needed ($2,553 for 431 producers, of which $2,230 was delivered to the 356 producers included in the evaluation).

**Total program budget was $299,281, also including $18,000 for evaluation purposes.

Table A 10: Results on performance – PSM

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
(T3-C) Full Package impact						
Post=1 x Full Package=1 vs. Post=1 x Control=1	13.567***	0.729***	433.375**	0.824**	2.861	0.114
	(4.045)	(0.223)	(163.166)	(0.267)	(2.039)	(0.054)
Observations	155	155	155	155	155	155
(T1-C) Partial General Training impact						
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	12.667	0.726	286.812	0.821	1.121	0.095***
	(6.521)	(0.394)	(145.924)	(0.468)	(0.574)	(0.021)
	54	54	54	54	54	54
(T2-T1) Queen Bee impact						
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	-0.796	0.057	151.330	0.202	3.966	0.145
	(7.247)	(0.275)	(81.435)	(0.168)	(3.160)	(0.140)
	131	131	131	131	131	131
(T3-T2) Breeding Training impact						
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.515	0.022	90.405	0.102	2.121	0.100
	(1.004)	(0.046)	(47.902)	(0.075)	(1.622)	(0.065)
	232	232	232	232	232	232

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences combined with Propensity Score Matching; robust SE in parentheses; P-values corrected for multiple hypothesis testing.

Table A 11: Results on adoption of practices index and queen-bee related individual practices – PSM

	(1) Practices (dummy)	(2) Practices (0-9)	(3) Change queen	(4) Breed queen	(5) Record log
(T3-C) Full Package impact					
Post=1 x Full Package=1 vs. Post=1 x Control=1	-0.410	2.231***	0.560***	0.522***	0.739***
	(0.362)	(0.464)	(0.052)	(0.051)	(0.038)
Observations	155	155	155	155	155
(T1-C) Partial General Training impact					
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	-0.636	-1.030	0.061	0.000	0.091*
	(0.577)	(1.760)	(0.075)	(0.044)	(0.049)
	54	54	54	54	54
(T2-T1) Queen Bee impact					
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	-0.031	2.592***	0.439**	0.296*	0.510***
	(0.097)	(0.692)	(0.166)	(0.166)	(0.053)
	131	131	131	131	131
(T3-T2) Breeding Training impact					
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.000	0.873**	0.246***	0.284**	0.291***
	(0.055)	(0.383)	(0.081)	(0.130)	(0.079)
	232	232	232	232	232

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences combined with Propensity Score Matching; robust SE in parentheses; P-values corrected for multiple hypothesis testing.

Table A 12: Results on adoption of other individual practices – PSM

	(1) Supplemented feeding	(2) Sustainable pest control	(3) Brood chambers	(4) Honey suppers	(5) Outer covers	(6) Brood frames
(T3-C) Full Package impact						
Post=1 x Full Package=1 vs. Post=1 x Control=1	0.209*	0.194***	-0.052	0.052	-0.060	0.067
	(0.085)	(0.045)	(0.059)	(0.066)	(0.367)	(0.064)
Observations	155	155	155	155	155	155
(T1-C) Partial General Training impact						
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	-0.121	0.030	-0.273*	-0.273	-0.455	-0.091
	(0.081)	(0.106)	(0.103)	(0.578)	(0.581)	(0.579)
	54	54	54	54	54	54
(T2-T1) Queen Bee impact						
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	0.173	0.020	0.327*	0.449	0.265	0.112
	(0.076)	(0.176)	(0.121)	(0.208)	(0.348)	(0.199)
	131	131	131	131	131	131
(T3-T2) Breeding Training impact						
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.090	-0.104	-0.037	0.060	0.030	0.015
	(0.113)	(0.121)	(0.064)	(0.068)	(0.106)	(0.078)
	232	232	232	232	232	232

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences combined with Propensity Score Matching; robust SE in parentheses; P-values corrected for multiple hypothesis testing.

Table A 13: Results on performance - IPW

	(1) Hives	(2) Ln(Hives)	(3) Kg.	(4) Ln(Kg.)	(5) Yield	(6) Ln(Yield)
(T3-C) Full Package impact						
Post=1 x Full Package=1 vs. Post=1 x Control=1	11.180***	0.654***	312.105***	0.760***	2.213	0.118*
	(1.733)	(0.099)	(74.801)	(0.144)	(1.601)	(0.059)
Observations	316	316	316	315	316	315
(T1-C) Partial General Training impact						
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	7.549**	0.390***	71.912	0.365**	-2.067*	-0.030
	(3.059)	(0.105)	(67.127)	(0.137)	(1.131)	(0.046)
	316	316	316	315	316	315
(T2-T1) Queen Bee impact						
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	3.141	0.156*	212.990***	0.383**	4.144***	0.232*
	(2.260)	(0.083)	(61.087)	(0.160)	(1.268)	(0.132)
	316	316	316	315	316	315
(T3-T2) Breeding Training impact						
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.490	0.108*	27.203	0.012	0.136	-0.085
	(1.492)	(0.064)	(54.742)	(0.159)	(1.560)	(0.134)
	316	316	316	315	316	315

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences combined with Inverse Probability Weighting; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table A 14: Results on adoption of practices index and queen-bee related individual practices – IPW

	(1) Practices (dummy)	(2) Practices (0-9)	(3) Change queen	(4) Breed queen	(5) Record log
(T3-C) Full Package impact					
Post=1 x Full Package=1 vs. Post=1 x Control=1	-0.186	2.614***	0.583***	0.520	0.753
	(0.144)	(0.545)	(0.097)	(.)	(.)
Observations	316	316	316	316	316
(T1-C) Partial General Training impact					
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	-0.108	-0.713	-0.059	-0.155	0.054
	(0.184)	(0.676)	(0.184)	(.)	(.)
	316	316	316	316	316
(T2-T1) Queen Bee impact					
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	-0.096	2.011***	0.323	0.346*	0.459***
	(0.111)	(0.720)	(0.214)	(0.207)	(0.122)
	316	316	316	316	316
(T3-T2) Breeding Training impact					
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.018	1.317***	0.320***	0.329***	0.240**
	(0.050)	(0.429)	(0.082)	(0.073)	(0.091)
	316	316	316	316	316

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences combined with Inverse Probability Weighting; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Table A 15: Results on adoption of other individual practices – IPW

	(1) Supplemented feeding	(2) Sustainable pest control	(3) Brood chambers	(4) Honey suppers	(5) Outer covers	(6) Brood frames
(T3-C) Full Package impact						
Post=1 x Full Package=1 vs. Post=1 x Control=1	0.204	0.197	-0.027	0.084	0.149	0.150
	(0.126)	(0.069)	(0.090)	(0.112)	(0.114)	(0.133)
Observations	316	316	316	316	316	316
(T1-C) Partial General Training impact						
Post=1 x Partial General Training=1 vs. Post=1 x Control=1	-0.087	0.152	-0.348	-0.307	-0.008	0.045
	(0.061)	(0.183)	(0.152)	(0.237)	(0.125)	(0.128)
	316	316	316	316	316	316
(T2-T1) Queen Bee impact						
Post=1 x Queen Bee=1 vs. Post=1 x Partial General Training=1	0.176	-0.068	0.279	0.309	0.105	0.082
	(0.078)	(0.136)	(0.155)	(0.215)	(0.124)	(0.092)
	316	316	316	316	316	316
(T3-T2) Breeding Training impact						
Post=1 x Full Package=1 vs. Post=1 x Queen Bee=1	0.115	0.113	0.042	0.082	0.052	0.023
	(0.073)	(0.087)	(0.074)	(0.056)	(0.094)	(0.065)
	316	316	316	316	316	316

Notes: * p<0.10, ** p<0.05, *** p<0.01. Estimates based on Difference-in-Differences combined with Inverse Probability Weighting; SE clustered at the locality level in parentheses; P-values corrected for multiple hypothesis testing.

Annex 2: Questionnaires

Producer identification

Questionnaire N°: Technician Key:	Date:

Name of producer:	Gender of producer:	1. Male		2. Female	
Beekeeping registration N°:	Geographic coordinates:	Latitude		Longitude	
Municipality:	Name of the apiary:				
Locality:	N° of hives in the apiary:				
Main occupation:	1. Agriculture	2. Beekeeping	3. Livestock	4. Other	
Education level:	1. Primary	2. Secondary	3. High school	4. Technical	5. Higher
Time living in the region:	1. 1-4 years	2. 5 years	3. 6-10 years	4. 11-20 years	5. +20 years

1. Diagnostic review of the apiary

1. Apiary condition

1.1. Cleanliness inside the apiary	1. Clean	2. With weeds	3. Garbage
1.2. Is the apiary fenced?	1. Yes	2. No	
1.3. Cleaning method	1. Manual	2. Herbicides	3. None
1.4. Water sources	1. Natural	2. Drinking fountains	3. Does not have
1.5. Adequate orientation	1. Yes	2. No	
1.6. Presence of shadow	1. Yes	2. No	
1.7. Type of hive	1. Langstron	2. Jumbo	3. Rustic
1.8. Hives with bases	1. Yes	2. No	

2. Condition of field materials

1.9. Material of the bases	1. Concrete	2. PVC box	3. Metal
1.10. Height of the bases (cm)	1. 20 cm	2. 30 cm	3. 40 cm
1.11. Condition of brood chambers	1. Optimal	2. Moderate deterioration	3. Severe deterioration
1.12. Hives have external cover?	1. Yes. Type:		2. No
1.13. Hives have hive entrance?	1. Yes	2. No	
1.14. Hives have inner cover?	1. Yes	2. No	
1.15. Condition of honey suppers	1. Optimal	2. Moderate deterioration	3. Severe deterioration
1.16. Condition of top covers	1. Optimal	2. Deterioration	3. Does not have
1.17. Top covers material	1. Metal sheet	2. Wood	3. Plastic
1.18. Number of beehives affected by climatic events:			
1.19. Condition of brood frames	1. Optimal	2. Moderate deterioration	3. Severe deterioration
1.20. Condition of honeycombs	1. With brood	2. With reserve	3. Empty
1.21. Hive activity	1. High	2. Medium	3. Reduced

3. Reproductive and genetic management of hives

1.22. Annual change of queen bees	1. Yes	2. No. Why?		
1.23. How are queens obtained?	1. Own production	2. Purchase from another producer	3. Registered hatchery	4. Hatchery
1.24. Clips the queen's wings?	1. Yes. Why?		2. No	
1.25. How are queens bought?	1. Fertilized	2. Virgin	3. Royal cell	
1.26. Last time you increased your hives	1. 3 months	2. 3 months – 1 year	3. More than 1 year	

4. Plagues and diseases

1.27. Pests present	1. Yes. Which?		2. No
1.28. Diseases present	1. Yes. Which?		2. No
1.29. Number of affected hives:			
1.30. Measures applied to control plagues and diseases:			
1.31. Date of last application:			
1.32. Measure(s) applied in last application:			

5. Feeding

1.33. Do you provide food for the hives?	1. Yes. In which months?	
	2. No. Why?	
1.34. Type of feeding provided	1. Support	2. Supplemented
1.35. Type of food provides	1. Syrup	2. Cake with pollen
1.36. Inputs used for the syrup	1. Sugar	2. Fructose

6. Climatic conditions

1.37. Level of flooding reached	1. Less than 30cm	2. 30cm-50cm	3. More than 50cm
1.38. Type of damage	1. Total damage	2. Loss of biological material	3. Decrease in populations
1.39. Number of beehives affected in last rainy season:			
1.40. Mention main damages in the apiary:			
1.41. Other type(s) of damages:			
1.42. Measures implemented to recover from the damage:			

7. Apiary environment

1.43. Approximate area of the apiary:				
1.44. Is it located within a parcel?	1. Yes	2. No		
1.45. Use of the land where the apiary is located	1. Crops	2. Livestock	3. Acahual	
1.46. Type of access	1. Paving	2. Dirt road	2. Dirt path	
1.47. Approximate distance from the apiary to:	1. Housing	2. Roads	3. Other apiaries	4. Foraging sources
1.48. Do you keep a record log of activities in the apiary?	1. Yes	2. No		

1.49. Have you received training or technical assistance in beekeeping?	1. Yes. Which subjects?		2. No

1.50. Observable nectapolyniferous species:

1.51. Observable degraded and/or deforested areas:

8. Historical hive and production data

Year	Hives	Kg of honey produced
2016		
2017		
2018		
2019		
2020		

2. Socioeconomic and livelihood aspects

1. Ethnic group and household composition

2.1. Ethnic group:

2.2. Mother tongue:

2.3. How many members make up your household? | N° of men: | N° of women:

2. About gender gaps

2.4. Who manages the financial administration of production?

2.5. What is the role of women in production?

2.6. Do you consider that women have the same capacity as men to be beekeepers? | 1. Yes | 2. No. Why? |

2.7. Why do you think there is little participation of women in beekeeping?

(if the interviewee is a woman) 2.8. What barriers have you had to face to work in beekeeping?

3. Income and credit access

2.9. Approximately how much is your monthly income?

2.10. What is the activity on which your income is most dependent?

2.11. How much does the income from the sale of honey contributes to the household income?

2.12. In your household, do other members contribute to the family economy? | 1. Yes | 2. No

2.13. Do you receive any government support? | 1. Yes | 2. No

2.14. Do you have access to any type of credit for your activity? | 1. Yes | 2. No. Why? |

2.15. Do you currently require any type of credit? | 1. Yes | 2. No

2.16. For which concepts do you require credit?

2.17. Do you have savings for an emergency? | 1. Yes | 2. No

4. Assets and resources

2.18. Do you have your own land? | 1. Yes, Ejido | 2. Yes, private property | 3. No

2.19. Do you own the house you live in? | 1. Yes | 2. No |

2.20. Type of vehicle used to reach the apiary: | |

2.21. Do you have an extraction room? | 1. Yes, fixed | 2. Yes, mobile | 3. No |

2.22. Extraction equipment available:

Extractor	Uncapping bench
Sedimentation vat	Strainers

5. Availability of inputs

2.23. Approximate amount of wax required annually? (kg) | |

2.24. Are you self-sufficient in your wax needs? | 1. Yes | 2. No. Why? | |

2.25. What difficulties have you had accessing wax? | |

2.26. Do you have difficulties to feed the hives in critical season? | |

2.27. Do you obtain pollen for commercialization? | 1. Yes | 2. No. Why? | |

2.28. Would you be interested in producing pollen for commercialization? | 1. Yes | 2. No. Why? | |

2.29. Where do you get inputs for pest and disease control? | |

2.30. Do you do nucleus division to increase hives? | 1. Yes | 2. No. Why? | |

2.31. In addition to apis mellifera beehives, do you have meliponines bees? | 1. Yes | 2. No |

2.32. Do you elaborate products based on beehive derivatives (pollen, propolis, royal jelly)? | 1. Yes | 2. No |

3. Perceptions of climate change and use of ecosystem resources

1. Knowledge on climate change

3.1. Have you heard about climate change? | 1. Yes | 2. No |

3.2. Do think the climate has been changing?

1. Yes. Why?	
2. No. Why?	

3.3. If yes, how long have you been noticing these changes? | 1. <5 years | 2. 5-20 years | 3. +20 years |

3.4. What changes have you noticed in weather behavior?

Excessive rainfall	Floods
Droughts	High temperatures
Other:	

3.5. To what do you think these weather changes are due? | |

3.6. Approximately how often do they occur? | |

3.7. Do you think people can do anything to slow down these changes?

1. Yes. How?	
2. No. Why?	

3.8. How do you think you could contribute to reduce these changes? | |

2. Past affectations

3.9. Have you been affected by any weather-related event?

1. Yes. What kind?	2. No
Flooding	
Drought	
Hurricanes	

3.10. How did this natural disaster affect you? | |

3.11. Did you receive any support during or after the incident?	1. Yes. What kind?		2. No
3.12. During the period of the incident, was it difficult for you to obtain food?	1. Yes	2. No	
3.13. How did the climate phenomenon affect the availability of work?			
3.14. Do you think women are more affected than men in the event of a climate disaster?	1. Yes. Why?		
	2. No. Why?		

3. Use of ecosystem resources

3.15. What are the main activities in your locality?			
3.16. Where does water for human consumption come from?			
3.17. Is there a year-round supply of water?	1. Yes.	2. No. In which months there is shortage?	
3.18. How do you deal with this shortage?			
3.19. What fuel do you use for cooking?	1. Gas	2. Wood. Where do you obtain it from?	
3.20. Is there a decrease in the number of plants useful for bees and for humans?	1. Yes	2. No	
3.21. Do you remember the names of any of these species?			