RIDIE Study Registration
Pre-Analysis Plan

General

Study Overview

Title
Assessing the Downstream Socioeconomic and Land Health Impacts of Agroforestry in Kenya

Abstract

Agroforestry is widely purported to improve the livelihoods of smallholder farmers, rehabilitate degraded landscapes, and enhance the provisioning of critical ecosystem services, including carbon sequestration. Yet, the evidence base supporting these claims is weak. Using a quasi-experimental evaluation design based on a theory-based and mixed methods framework, this study investigates the longer-term and intermediate effects of an eight year effort led by Vi Agroforestry, a Swedish non-governmental organization (NGO), to promote agroforestry in large sections of Bungoma and Kakamega counties in western Kenya. Key longer-term outcomes of interest include consumption expenditure, asset accumulation, food security, shock resilience, soil organic carbon and erosion prevalence, and educational attainment. The causal effects of Vi Agroforestry's efforts will be estimated by comparing outcomes between two sets of households: those belonging to (a) 226 targeted and pre-existing farmer groups operating in 60 villages; and (b) 206 non-targeted pre-existing farmer groups operating in 61 villages located outside the intervention area that were matched to the intervention villages on selected geophysical and demographic variables. To further counter selection bias, doubly robust inverse probability weighted regression and other complementary econometric techniques will be combined with difference-in-differences estimation. Qualitative methods will also be used to triangulate key findings and interrogate impact pathways, unforeseen outcomes, and unexpected quantitative results.

Principal Investigator(s)

Name of First PI: Dr. Karl Hughes

Affiliation: The World Agroforestry Centre (ICRAF)

Email: k.hughes@cgiar.org

Name of Second PI: Dr. Katherine Baylis

Affiliation: The University of Illinois at Urbana-Champaign

Email: Baylis@illinois.edu

Intervention

Intervention Overview

Intervention

Within the program area, Vi promoted several interrelated, complementary agroforestry practices, including trees for (a) domestically consumable and marketable products, e.g. timber, fuel wood, and fruit; (b) soil health improvement practices, e.g. soil fertility enhancement, erosion control and increased water infiltration; and (c) livestock fodder. These promotional efforts were undertaken by targeting pre-existing farmer groups and providing their members with tree seeds, support to establish tree nurseries, and training and extension support on how to integrate trees with field crops on their farms.

The implementation of Vi's interventions in the program area (comprising of large sections of Bungoma and Kakamega counties in western Kenya) started in 2008 through two different projects: the Kenya Agricultural Carbon Project (KACP) and the Farmer Organizations and Agroforestry (FOA) project. The two projects have their own staff and funding structures but share similar approaches on the promotion of agroforestry and land management.

The KACP project, which is active in Bumula and Bungoma West sub-counties, focuses on increasing carbon sequestration in small-holder farming systems by encouraging tree planting and sustainable land management techniques. Tree planting in particular is incentivized by a small payment (equal to about $3.00 per person on average) disbursed to the farmer groups upon confirmation that trees have been planted and preserved on their farms.

The FOA project focuses more significantly on capacity building for farmer organizations, in addition to tree seed provision and land management training. FOA is active in Kimilili, Webuye East and Bungoma North sub-counties in Bungoma County, as well as Likuyani sub-county in Kakamega County. Vi staff describe the extension training and related support provided by FOA as being very similar to those of KACP. However, there are two key programmatic differences: FOA does not provide carbon payments and it does not monitor tree planning and management activities with the same level of rigor as KACP. Moreover, since FOA is focused on empowering farmer organizations, Vi made a decision in 2014 to hand over its activities in Kimilili and Webuye East to partner Savings and Credit Cooperatives (SACCOs). This means that Vi staff interaction with farmer groups was mediated by the SACCOs for two years out of the study period, i.e. from 2014 to 2016.

Given the above, there are three variations of Vi's program that will be factored into our analysis:

1. KACP project areas: farmer groups that participated in the KACP carbon credit scheme and have therefore received modest carbon payments and been subjected to monitoring activities to verify tree survival.
2. FOA project area with Vi staff: farmer groups in Bungoma North and Likuyani sub-counties are a part of the FOA program and continue to receive services directly from Vi Agroforestry staff.
3. FOA program areas handed over to SACCOs: farmer groups in Kimilili and Webuye East sub-counties continue to receive trainings with priorities set by Vi, but with actual activities performed by SACCO staff.

These three programmatic variations divide the study area into roughly three equal sub-areas, with adjacent sub-counties remaining as sources of comparison farmer groups that have never received services from Vi in any form. Due to funding restraints, Bungoma Central, Bungoma South, Webuye West, parts of Bungoma North and Lugari sub-county of Kakamega, remained outside of Vi's program area. Adjacent locations in these sub-counties were chosen as candidates for comparison to villages inside the Vi program zone.

Implementing Agency

Name of Organization

Vi Agroforestry

Type of Organization:

NGO and provider of agricultural extension services.

Program Funder

Name of Organization

World Bank BioCarbon Fund provides carbon credit funding for the KACP project. The FOA project is funding by private donations to Vi Agroforestry's Swedish parent organization and grant funding from the Swedish International Development Cooperation Agency (SIDA).

Type of Organization

The BioCarbon Fund is a World Bank program housed within the Carbon Finance Unit of the World Bank. SIDA is a Swedish government agency for the provision of official development assistance. Vi Agroforestry Sweden (Vi Skogen) is a private charitable organization.

Intervention Timing

Start Date

2008

End Date

Ongoing

Evaluation Method

Evaluation Method Overview

Primary (or First) Evaluation Method

Difference in differences

Other Method

Matching on observables (both at village and household levels)

Additional Evaluation Method (If Any)

Other Method

Theory-based evaluation methods—both quantitative and qualitative—will be used to investigate the mechanisms through which any of the hypothesized outcomes differences between the treatment groups and sub-group effects did or did not came about.

Method Details

Details of Evaluation Approach

The causal effects of Vi's program will be estimated by comparing outcomes between two sets of households: those belonging to (a) 226 targeted and pre-existing farmer groups operating in 60 villages; and (b) 61 non-targeted pre-existing farmer groups operating in 60 villages located outside the program area but within the same two Kenyan counties—Bungoma and Kakamega. The locations in which Vi implemented the program in general and farmer groups it targeted in particular were selected non-randomly. Consequently, it is possible that any differences in outcome identified between these two groups may be driven—either in whole or in part—by non-programmatic factors, e.g. the intervention groups may have already been better off in relation to the outcome variable in question to begin with and/or systematically subjected to differing external events/shocks during the implementation period.

Given the study's non-experimental nature, coupled by the fact that an appropriate baseline survey was never undertaken, several measures have and will be undertaken to counter both programme placement and self-selection bias, as well as other internal validity threats, i.e. program spillover effects and contagion. These are as follows:

1. **Village matching based on selected geophysical and demographic variables**. During the baseline period, Vi targeted specific geographical areas (i.e. Locations) and then pre-existing and active farmer groups within these Locations. Almost all of these groups operated in specific villages or clusters of neighboring villages. To counter program placement bias, 61 other villages with similar geophysical and demographic characteristics—i.e. population size and density, baseline soil conditions and tree cover, elevation, rainfall, and distance from major road networks—and with existing farmer groups that had similarly been active since the baseline period were identified and surveyed. These villages fall outside of Vi's program area but are located in various adjacent areas to the program villages within the two program counties, thereby increasing the likelihood that they were subjected to similar external events and shocks during the implementation period.

2. **Sampling from all farmer groups that existed in both the intervention and comparison villages at baseline.** We adopted this strategy to explicitly counter self-selection bias. As mentioned above, Vi initially targeted pre-existing farmer groups, as opposed to mobilizing new groups in the areas it targeted. There was likely significant self-selection at play in the formation of these groups. This would potentially bias a comparison of these groups with random samples of

farmers in the non-program villages. Consequently, only households belonging to active farmers groups that had existed since the baseline period were sampled in these villages. The assumption here is that Vi would have offered these groups the opportunity to participate in its program had there been sufficient resources to expand into matched non-program villages. Moreover, in the event, not all existing farmer groups in the program villages took up the offer to participate in Vi's program. Thus, those groups that did are actually a sub-set of these targeted famer groups, which may be unique in terms of both their observable and unobservable characteristics relevant to the outcomes of interest. Bias could therefore creep in if we were to directly compare them with the sample of pre-existing farmer groups in the intervention villages. We therefore sampled households from *all* farmer groups that existed in both the intervention and comparison sites at baseline, while seeking to identify two types of causal effect estimates, as outlined below.

3. **Reconstruction of baseline data for difference-in-differences estimation**
A key limitation of the study is that no appropriate baseline survey was undertaken. Consequently, differences in outcome identified between the two groups could simply be reflective of their baseline differences. The absence such data also prevents difference-in-differences estimation, an appropriate identification strategy when it is reasonable to assume that the rate of outcome change experienced between the two groups would have been the same in the counterfactual state (i.e. the parallel trend assumption). However, during data collection, recalled baseline data were obtained from the respondents, particularly for items for which recall bias is expected to be minimal. This included ownership of assets, other wealth indicators (e.g. house characteristics), household livelihood pursuits, and tree planting and land management practices. We took advantage of the fact that a major historical event took place one year prior to the baseline, i.e. Kenya's post-election violence, and used this as a historical marker. The respondents were specifically requested to recall conditions prior to the events that took place in December 2007. The geographical area where the survey was carried out did not witness significant violence, as had taken place in other parts of Kenya, so we felt it was appropriate to use this particular historical marker.

   The recalled baseline data will therefore be used to generate difference-in-difference estimates for several important outcome measures, e.g. changes in asset wealth. For obvious reasons, consumption and expenditure data were only obtained for the endline period. However, the rich basket of assets and other wealth indicators obtained for both periods will be used together with the endline data to derive a differenced predicted consumption expenditure measure. In addition, three differenced biophysical measures—soil organic carbon, soil erosion, and fractional vegetation cover—will be used. As further elaborated below, these measures are based on the analysis of satellite imagery using models developed from data from the World Agroforestry Centre's Land Degradation Surveillance Framework (LDSF) field sites.

4. **Doubly robust estimation and other appropriate econometric techniques to control for other observable differences between the two groups.** The Achilles heel of difference-in-differences estimation is the parallel trend assumption; that is, such estimates will be biased if the outcome of interest would have changed at a different rate among either treatment group in the counterfactual state. This could be through either groups' systematic differential exposure to external events or shocks and/or due to the influence of their unique characteristics, whether these be observable or unobservable. An example of the latter is related to education. If one of the groups happens to be more educated overall, this may place them in a better position to respond to new market opportunities available to both groups, thereby resulting in more significant increases in income over time. The study will, therefore, strengthen the difference-in-differences identification strategy by combining it with econometric methods that control for observable differences (such as education and baseline wealth status) that may have caused a violation of the parallel trend assumption. This will not, however, address bias resulting from differential external event/shock experience and/or the influence of unobservable characteristics (e.g. entrepreneurial mindset) in affecting the evolution of the outcome of interest over time.

   In general and particularly in cases where only single difference estimation is possible—which is particularly the case for the study's food security and shock coping measures—attempts to mitigate

bias will be undertaken by controlling for baseline asset wealth status and other recalled baseline and time invariant household and participant characteristics. Given that it offers protection against the misspecification of either the participation or outcome model (but not both), doubly robust inverse probability weighted regression will be the first estimation technique of choice. However, the nature of the outcome data in question may favor other methods, e.g. robust regression in the presence of extreme values. In any event, several estimation techniques will be employed as a robustness check. A standard analysis will include: (a) the raw outcome difference between the two groups; (b) doubly robust estimation with fixed effects at project site level; (c) PSM kernel and nearest neighbor with no replacement estimation and exact matching within project sites; (d) multivariable regression with project-level fixed effects; and (e) robust multivariable regression for continuous outcome variables, particularly in the presence of extreme values.

5. **Use of Intention-to-treat (ITT) and local average treatment effect (LATE) estimation.** Particularly given what is stated under Point 2 above, we will generate two types of treatment effect estimates. ITT effect estimates will be derived by comparing all sampled households in the villages targeted by Vi with those in the comparison villages using the above estimation techniques, regardless of whether they happen to belong to a Vi group. However, given that the sample of households from the Vi groups includes a significant number (25%) that are not members of Vi groups (and, hence were not directly treated by the program), the ITT estimates will likely underestimate the impacts of VI's program (assuming that this program was effective).

   One possible method of obtaining a more refined estimation of the impacts of Vi's program on those households that actually participated would be to estimate the average treatment effect on the treated (ATT). This would involve comparing households that belonged to the Vi affiliated groups with those households in the intervention villages that are observationally similar (statistically speaking). However, such treatment effect estimates would really strongly on the "selection on observables" assumption and undermine the bias mitigation strategies presented under Point 1 and Point 2 above. In other words, it would not rule out the possibility of there being non-program related unobserved differences between the two groups driving any identified differences in outcome.

   Alternatively, given that the program and non-program villages were matched on key geophysical and demographic characteristics, we will assume that they are as good as randomly assigned, i.e. the matching process eliminated program placement bias, leaving self-selection bias as the primary bias that needs to be tackled. We will also assume that the opportunity provided by Vi to the pre-existing farmer groups in the program sites to participate made it more likely and never less likely for them to have actually participated (i.e. the monotonicity assumption). With these two assumptions, we will use two-stage least square regression (2SLS) to derive LATE estimates. Given that there are very few households in the non-program areas that are members of Vi groups, these effect estimates will essentially pertain to those households that were members of the farmer groups that actually participated in Vi's program.

6. **Use of mixed methods to interrogate mechanisms**
   We have adopted a theory-based approach and will use both quantitative and qualitative methods to explore how any observed treatment effects came or did not come about. Quantitative data were therefore captured through the household survey to enable us to trace the expected ways in which program participation was expected to lead to positive effects on consumption expenditure, food security, and resilience. Data were captured, for example, to assess the extent to which the Vi affiliated farmer groups were better exposed to the types of interventions associated with this program. If these efforts were successful, we would then expect to see greater uptake of the types of trees and agroforestry and related practices Vi promoted and, in turn, effects on intermediary outcomes, such as increased access to and sales of various tree products like firewood, timber, fruit, and medicine. Women's time collecting firewood would further decrease significantly, as would household expenditure on firewood and timber. Dairy farmers would further be expected to have either reduced production costs or increased milk yields from the fodder introduced by the program. Reduced soil erosion and possibly increased soil fertility would also be expected with integration of

the right trees in the right places in farmers' fields, thereby positively impacting crop production. Positive effects on the final outcome variables presented below would then follow suit.

In the unlikely event that the quantitative data evidences such as consistent story, qualitative methods will be used to embellish it and add more depth. However, there will likely be inconsistencies in general and for particularly categories of participating households in particular. Our sub-team of qualitative researchers will iteratively formulate and test hypotheses on why this is the case. A particular focus will be on understanding how the participating households actually interpreted the program, as well as put in place and adapted what was being promoted.

Outcomes (Endpoints)

Given our adoption of a theory-based evaluation framework, we will measure various exposure, adoption, intermediate and final outcome variables. These are listed below but explained in greater detail in the measurement section.

Exposure Variables:
1. Agroforestry group participation
2. Receipt and Implementation of agroforestry training
3. Receipt and Implementation of agroforestry extension
4. Other agricultural support

Uptake of Promoted Practices:
1. Agroforestry Practice Index
2. Differenced fractional vegetative cover

Intermediary Outcome Variables:
1. Cash earned from sale of agroforestry products
2. Estimated cash value of firewood harvested from farm
3. Estimated hours collecting firewood per month
4. Average % change in milk yields among dairy producers
5. Self-reported increase in income from dairy production
6. Estimated soil erosion prevalence and soil organic carbon (SOC)

Primary Outcome Variable:
Differenced predicted consumption expenditure

Other Outcome Variables:
1. Differenced asset wealth
2. Single-difference consumption expenditure
3. Adapted Coping Strategies Index
4. Minimum Dietary Diversity – Women (MDD-W)
5. Months of Adequate Food Provisioning
6. Education Progression
7. Education Spending
8. Perceived Change on Economic Ladder

Measurement

Exposure Variables:

1. **Agroforestry group participation:** Respondent reports being a member of group with a focus on tree planting and management, with group participation in at least 4-5 times in last 12 months and a role in decision-making to a medium extent
2. **Receipt and Implementation of agroforestry training:** Respondent reports having received training in tree planting and management at least 1 time during last 3 years and reports having implemented the training at least to a medium extent. A comparison of the specific agroforestry practices the respondents report having been trained in will also be compared.
3. **Receipt and Implementation of agroforestry extension:** Respondent reports having received extension support related to tree planting and management during the last 3 years and reports having implemented the extension support at least to a medium extent. A comparison of the specific AF practices the respondents report having had received extension support on will also be compared.
4. **Other agricultural support:** Similar binary variables will be constructed for group participation in other agricultural and non-agricultural topics, as well as exposure to other types of training and agricultural extension support.

Uptake of Promoted Practices:

Agroforestry Practice Index:

To compare groups in relation to the extent to which they actually took up the specific agroforestry practices Vi promoted under its program, we will construct a multi-dimensional practice uptake index. It will comprise of three dimensions with several binary indicators under each. The index will be computed for both the baseline and end line periods, thereby enabling the construction of a differenced index. The dimensions and binary indicators that underpin each, including how each are specified are as follows:

Practice Uptake:

1. *2 or more tree products from largest farm plot*. Specification: Trees (non-Eucalyptus) along boundary or within largest food crop farming plot which are: generating two or more products for either sale or domestic use (timber, fuel wood, fruits, medicine, fodder, green manure), at least some of which were planted or natural occurring species that were protected and nurtured..
2. *Tree-based NRM within largest farm plot*. Specification: Trees (non-Eucalyptus) within largest food crop farming plot which are: planted on grass strips, contours, alleys, as windbreaks, as improved fallows, at least some of which were planted or natural occurring species that were protected and nurtured; or over 20 interspersed trees all of which were planted or natural occurring species that were managed (i.e. FMNR).
3. *1 or more complementary agroforestry practices*. Specification: One or more of the following: fruit orchard (or over 10 fruit trees at homestead); woodlot; tree-based fodder bank (or fodder shrubs on plot or homestead with reported use as fodder) or dairy farmers reporting using shrub fodder as feed; or tree-based fallow.

Intensity of Practice:

1. *Density of tree cover on farm plots*. Specification: At least 50 trees on at least one food crop farm plot, with all other food crop farm plots having at least 20 trees, some of which were planted or naturally occurring species that were managed.
2. *Sales from tree products from farm plot*. Specification: At least one tree product from food crop farming plot produced for sale.

3. *Intensity of complementary practices*. Specification: Fruit tree orchard with over 20 trees (or fruit trees at homestead with over 10 trees); woodlot with over 100 trees; and/or tree-based fodder bank with over 50 shrubs.
4. *Sales from complementary AF practices*. Specification: At least one tree product from complementary AF practices produced for sale (fruit from orchard or homestead, timber/firewood from woodlot, fodder from fodder bank).

Tree Species:

1. *1 or more leguminous shrubs*. Specification: Presence of at least 1 of the Vi-introduced leguminous shrub species (*Sesbania sesban*, *Calliandra calothyrsus*, *Leucaena leucocephala*).
2. *3 or more long-term exotic tree species*. Specification: Presence of at least 3 long-term exotic species promoted by Vi (*Grevillea robusta, Casuarina equisetifolia FRUIT TREE SPECIES Persea americana, Cassimiroa edulis, Annona spp., Syzygium cuminii, Annona senegalensis*)
3. *3 or more native species*. Specification: Presence of at least 3 native species promoted by Vi (*Cordia africana, Markhamia lutea, Croton megalocarpus, Croton macrostachus, Prunus africana, Olea capensis, Vitex keniensis, Podocarpus falcatus*)

Differenced fractional vegetative cover:

This is a relative index that indicates the extent to which the sampled fields are covered by green vegetation. It is constructed from the analysis of satellite imagery and, in particular, the spectral signatures derived from them pertaining the phenomena in question. These spectral signatures are derived from predictive models for mapping land health, which are based on systematically collected field and lab measurements from about 150 Land Degradation Surveillance Framework (LDSF) sites across the global tropics. The LDSF framework[1] consists of a set of methods for capturing data on various ecosystem health metrics, including soil condition, vegetation structure and cover, landform, land use, and land degradation (e.g. soil erosion). The LDSF is developed for integrated landscape analysis, using a spatially balanced hierarchical sampling design based on 10 by 10 kilometre (100 km^2) sites (or landscapes), with each site consisting of 16 clusters, each 1 km^2, with 10 randomly generated 1,000 m^2 plots per cluster.[2]

Intermediary Outcome Variables:

1. **Cash earned from sale of agroforestry products**. For each farm plot profiled, respondents were asked whether any agroforestry product had been produced from the plot in the last 12 months and, if so, sold, including the approximate income earned through these sales. The reported values will be added together and compared between the treatment groups.

2. **Estimated cash value of firewood harvested primarily from own farm.** During the survey, respondents were asked to report (a) whether their household had used firewood over the past month; (b) if so, the sources and primary source of this firewood; and (c) if collected during the average month, how much this would have costed if it was to be purchased at local market. Restricting the analysis to only those households reported using firewood over the past 12 months, we will compare the treatment groups in terms of the estimated cash value of firewood accessed from their own farms.

3. **Estimated hours collecting firewood per month.** Respondents were also asked how many times they collected firewood in the average month and how many hours is spent on each of these

[1] http://landscapeportal.org/blog/2015/03/25/the-land-degradation-surveillance-framework-ldsf/

[2] See, for example: Vågen, T.G. et al., 2013. Landsat-based approaches for mapping of land degradation prevalence and soil functional properties in Ethiopia. *Remote Sensing of Environment*, 134(August), pp.266–275.

occasions. From this, an estimated number of hours per month spent collecting firewood will be derived, and the treatment groups will be compared against this.

4. **Average % change in milk yields among dairy producers.** The respondents who reported having either improved or local cows or dairy goats were asked about the average milk yield per day per cow or goat during the milking period at present and back in 2007. The percentage of reported change will be computed between the two time periods and averaged per household if the household in question has more than one of the 3 dairy producing animals in question (i.e. improved cow, local cow, or dairy goat). The analysis will be restricted to those respondents reporting milk production in both the baseline and endline periods.

5. **Self-reported changes in income from dairy production.** This analysis will also be restricted to those respondents reporting milk production in both the baseline and endline periods. These respondents were asked the extent to which their milk production has increased on a 5 point scale – about the same; increased significantly; increased somewhat; decreased significantly; and decreased somewhat, and the treatment groups will be compared against this scale.

6. **Estimated soil erosion prevalence and soil organic carbon (SOC).** The complex measurement methodology via remote sensing underlying these two measures is the same as that for fractional vegetation over explained above. The soil erosion measure pertains to the estimated percentage of soil that is eroded in the satellite pixel in question, while the SOC measure pertains to the number of grams of SOC per kilogram of soil in this same pixel.

Primary Outcome Variable:

Differenced predicted consumption expenditure. Given the absence of baseline data—coupled by our assumption that people can recall assets they owned and particular characteristics of their homes in the past with relatively good precision (particularly when promoted with a good historical marker)—we will take advantage of the fact that we have consumption expenditure data for the endline period and a rich set of wealth indicators for both baseline and endline periods.

Following a similar approach to that of the World Banks Living Standards Measurement Survey (LSMS), the consumption expenditure data were collected during the survey using a one week recall food consumption sub-module; a four week regular non-food spending sub-module; and one year non-regular occasional spending module. Following cleaning, these data will be used to compute the daily per adult equivalent per day consumption expenditure, adjusted for purchase power parity (PPP).

Data on over 80 of assets and other household wealth indicators were also collected during the household survey for both the endline and baseline periods, most of which fall on varying ordinal scales. To both enable sensible aggregation and exploit maximum variation in the data, the median ordinal values will be used to generate binary scores, with 1 for being above the median and 0 otherwise. Stepwise regression will be used on the 2016 variables to determine the set that best predict household consumption expenditure. We will check to ensure that the resulting predications are sufficiently correlated (i.e. $r > 0.60$) with the actual consumption expenditure. We will then assign weights to each shortlisted asset/wealth indicator based on their respective coefficients. These same weights will also be applied to the same set of asset/wealth indicators for the baseline period. We will then difference the two to create the differenced predicted consumption expenditure measure.

Other Outcome Variables:

1. Differenced asset wealth

This measure will complement the differenced predicted consumption expenditure, but just focus on using the data associated with the assets and other wealth indicators captured in both time periods. To construct

this measure, changes in household ownership of each wealth item/indicator will be differenced over the two time periods. In one of our analyses, we will simply add up the resulting binary scores to generate an asset change score and compare the treatment groups accordingly. However, in this analysis, all assets will be given the same weight. Consequently, in a complementary analysis, we will use principal component factor analysis to narrow in on the variation in the data assumed to be most reflective in changes in asset wealth over time. To execute this procedure, we will first verify if the 2007 basket of binary assets and wealth indicators each are significantly correlated, i.e. generate an alpha > 0.8. Those with negative signs will be dropped. A new variable will then be generated to indicate the assets and wealth indicators that saw positive change over the implementation period. The inter item correlation associated with this positive wealth change variable will also be checked. It will then be placed on a tetrachoric matrix followed by being subjected to PCA to generate a predicted variable pertaining to the first principle component, which we will assume significantly captures positive changes in asset wealth from the baseline period.

2. Single difference consumption expenditure estimates
We will also do two complementary analyses to compare the actual reported consumption expenditure between the treatment groups. The first will involve controlling for baseline asset wealth estimated via PCA and other baseline and time invariant covariates, while the second will control for predicted baseline consumption expenditure using the same method for the study's primary outcome variable.

3. Adapted Coping Strategies Index. This analysis will be restricted to respondents reporting that their local area had been subjected to one of four significant shocks during the previous three years: (a) drought or irregular rains; (b) flooding or landslides; (c) crop-related pest or disease outbreaks; and/or (d) livestock-related pest and disease outbreaks. If the respondent reports that their household was not affected by any of the shocks, this household will be given 0 points. Two-points will be awarded to those reporting that there household was affected to some extent but nothing serious for at least one of the four shocks but nothing greater. Households will be awarded 4 points if their respective respondents reported being moderately affected by at least one of the shocks but were able to recover after some time. Finally, households reporting that they were severely affected to at least one of the shocks will be given the maximum highest possible score of 8 points. These scores will not be added together for each shock; rather, the highest score associated for the four shocks will apply.

Households that were reported as being moderating and seriously affected will get additional points for the particular coping strategies that they used, using similar weights associated with the Coping Strategies Index (Maxwell and Caldwell 2008). The weights that will be used are as follows: borrowed money/credit (2 points); reduced meals (2 points); limited portion sizes at mealtimes (1 point); skipped an entire day without eating (4 points); restricted food consumption of adults so children can eat (2 points); restricted food consumption of children (4 points); substituted commonly bought food with cheaper foods (1 point); modified cooking method (1 points); mortgaged/sold assets (4 points); borrowed from neighbors/relatives (2 points); relied on remittances (2 points); sent household members out to beg (4 points); sent household members to eat elsewhere (2 points); and gathered unusual types/amounts of wild food or hunted (2 points). These scores will also not be added together for each shock; rather, the highest score associated coping strategy will be used.

4. Minimum Dietary Diversity – Women (MDD-W). The food items the respondents reported to have had consumed during the previous data will be grouped into the 10 MDD-W (FAO and FHI 360, 2016) food groups, giving each respondent a score out of 10. The groups will be compared in relation to this 10 point score but also the binary form of this indicator with the cut-off of 5 points.

5. Months of Adequate Food Provisioning. The number of months reported by the respondent where they had insufficient food to meet household needs over the past 12 months will be compared. Given that households recall this for the baseline year as well, differenced estimates will also be generated, but with the recognition that the recall error involved is likely to be substantial.

6. Education Progression. This analysis will be restricted to those households that have children between the ages of 14 to 17, i.e. households that presumably had school aged children at baseline. The average number of years of education for all children in each household (scaled for age) will be compared. This will provide an indication of whether the program households were in a better position to send their younger children to secondary school during the implementation period.

7. Educational Spending. These data were directly collected from the respondents and will be compared between the groups, while controlling for number of children and age of children.

8. Perceived Change on Economic Ladder. How the respondent rated where their household was on the six point scale at baseline will be differenced with how they rate their households now. These self-perceived scores will be differenced and compared between the groups.

Unit of Analysis

Household

Hypotheses

Our **primary hypothesis** is that a participation in Vi's program led to increased household consumption expenditure and asset accumulation on average, albeit with considerable variation among the participating households. Given that data were obtained from households belonging to non-Vi farmers from the program villages, we also hypothesis that the LATE estimates on consumption expenditure and asset accumulation will be greater than ITT effect estimates. However, the differences will not be that significant, given that these households represent only about one-quarter of those making up the program village sample.

Secondary hypotheses include:

1. **Shock resilience**. We further hypothesize that that households belonging to the Vi groups were less affected by shocks that hit their local areas (again, on average and with significant heterogeneity), given that the trees on their farms made them more resilient to such shocks by providing alternative (and relatively climate insensitive) sources of income.

2. **Education progression and spending**. During the study's scoping exercise, Vi group members outside the study area reported that having trees on their farms (which can easily be harvested and sold as firewood or timber) help supports them to meet non-regular lump sum expenses, e.g. school fees and related educational expenses. We therefore expect more children from Vi group households to have progressed on to secondary school and even higher levels of education than their comparators. We also expect household spending on education for these households to be higher on average.

3. **Food Security and Dietary Diversity**. For the same reasons as explained under the two points above, we expect that Vi group households will be less likely to report having problems with food provisioning. However, given that it is significantly driven by socio-cultural factors, we do not expect to see significant difference between the treatment groups in relation to dietary diversity (i.e. the MDD-W measure).

4. **Soil Erosion Prevalence and Soil Organic Carbon**. We hypothesis that, because of the promotion of agroforestry and other sustainable land management practices by Vi's program, relative decreases over the implementation period in the estimated prevalence of soil erosion will be greater, on average, in the farming plots belonging to Vi group members. However, among this group, we expect differential effects depending on the extent to which they have taken up the various agroforestry and other sustainable land management practices promoted under Vi's program. Moreover, given the long duration of time required to build up soil organic carbon, coupled

by the relatively large number of agroforestry and other sustainable land management practices promoted by Vi (and hence likely heterogeneous take-up), we do not expect to evidence significant causal effects for this particular variable. However, we may see significant effects in the CASP sites, given its more intensive promotion of complementary sustainable land management practices, e.g. mulching.

5. **Uptake of Promoted Agroforestry Practices**. We hypothesis that the Vi groups will score higher, on average, on our agroforestry practice index. However, we also expect that there will be significant heterogeneity in the Vi program area in general and among the three project sites in particular. We expect, in particular, that uptake will be greater in the KACP project site, given its more intensive monitoring system and issuing of modest payments to the participating farmer groups. While it happened only two years prior to the endline data collection exercise, we also expect lower agroforestry practice uptake in the FOA project sites that were taken over by SACCOs. We hypothesize that this differential uptake of the promoted agroforestry practices will, in turn, translate into corresponding differential effects on the above higher-level outcomes.

6. **More Significant Effects on Dairy Farmers**. Given the established efficacy of *Calliandra calothyrsus* and *Sesbania sesban* as highly nutritious sources of animal feed (Frazel, Kiptot, and Kukuyu, 2014), coupled with the fact that its use as such was strongly promoted under Vi's program, we expect larger treatment effects to be exhibited among dairy farmers.

7. **Less Significant Effects for Female Farmer Group Members**. Given that farm investment decision-making remains predominately the domain of men in the study area, we hypothesize that the impacts of Vi's program will be less on households with married female groups members whose husbands do not belong to the same or some other Vi group.

8. **Moderating Effects of Education, Wealth Status, Community Influence, and Landholding Size**. The types of agroforestry and related practices promoted by Vi's program varied, with some being relatively more technical and challenging to implement. In addition, a main focus of the program involved the provision of technical and organizational capacity to the participating groups and their members, rather than material support, i.e. the households themselves were expected to significantly rely on their own resources and labor to put in place the promoted practices on their farmers. Poorer households may have, therefore, not been in a position to implement many of the promoted practices and/or been preoccupied with meeting more immediate needs. Those with more land may have been in a position to plant more trees on their farms and, hence, generate greater returns. For the above reasons, we expect there to be differential effects for more educated, wealthy, and influential Vi groups members, as well as those with households having relatively larger agricultural landholdings.

9. **Agro-ecological Zone.** We hypothesize that adoption rates will be higher in mountainous areas with steeper topography, given that such areas are more prone to erosion and, thereby, motivating farmers to take remedial action.