Pre-analysis plan (10.02.2025)

1. Trial title

Unlocking agricultural potential through bundled services? Experimental evidence from a digital platform in Ghana

2. Country

Ghana

3. Status

On-going

4. Keywords

Digital platform, agriculture, weather shocks, climate change adaptation, extension services, bundled services, credit, soil testing, Ghana

5. Abstract

Agricultural productivity in Sub-Saharan Africa remains low compared to other areas of the world due to the various constraints that smallholder farmers face. Since addressing one constraint is typically insufficient, we evaluate the effects of bundled agricultural services that simultaneously tackle multiple constraints for farmers. We will implement a randomized controlled trial to investigate the impacts of four different service bundles for predominantly female soybean farmers in Northern Ghana. Initially, the implementation was planned for the 2024 season, however a severe drought in Northern Ghana and a flood in some communities prevented our implementing partner WamiAgro Ltd., an agricultural platform, from offering the intended service bundles in most of the communities in our sample. Therefore, we decided to conduct a second baseline survey in 2025 to elicit farmers' perceptions of changing weather patterns and their adaptation strategies, which may vary depending on the severity of droughts and floods. We proceed with the original experimental design in the 2025 season. In control communities, WamiAgro will provide farmers with weather information and training on farming practices, as well as the option to sell their soybeans via the platform, termed “Info + Market”. Our three treatment arms will add a soil test (“Soil test”), credit for harvesting and threshing (“Credit”), or a soil test and credit for harvesting and threshing (“Soil test + Credit”). We will analyze the effects of each treatment on agricultural production, land use, labor demand, and household welfare. The hypothesized impact of the soil test is that it will enhance farmers' understanding of their soil's health, enabling them to adjust their input selections and optimize the timing of their applications accordingly. We also expect improvements in self-reported soil quality, changes in production costs, and higher yields. For the credit treatment, we hypothesize that it increases the efficiency of harvesting and threshing. Mechanized threshing preserves the output quality, potentially leading to higher output

prices, sales, and incomes. We predict changes in labor demand and costs, as well as changes in household labor allocation due to increased mechanization. Households may also have lower levels of financial distress. One of the main research questions is whether soil tests and credit services are complements. Once farmers have optimized their inputs and practices based on soil test results, they are more likely to achieve higher yields. However, without sufficient resources for harvesting and threshing, these gains could be lost. Access to credit ensures that farmers can fully capitalize on the improved productivity by enabling timely and efficient post-harvest operations, thus preserving the quality and quantity of the produce. These synergies can improve outcomes in the short term and in future planting seasons. Lastly, we will assess whether the provision of agricultural extension services affects farmers strategies to cope with adverse weather events.

6. Trial start date

June 12, 2024

7. Intervention start date

June/July 2025

8. Intervention end date

November 2025

9. Trial end date

January 2026

10. Outcomes

Our first baseline has gathered data on the 2023 agricultural season and the second baseline will gather data on the 2024 agricultural season, while the follow-up survey in December 2025 will focus on the 2025 season and hence the short-term effects of the randomized service bundles.

We will assess the impact of the interventions on several intermediate outcomes, focusing on the uptake and use of credit and soil testing. In addition, we will collect data on risk coping measures and perceptions about climate change. These will be measured during the follow-up survey. Additionally, we are interested in evaluating the interventions' effects on ultimate outcomes, which reflect cumulative changes in the intermediate outcomes and relate to long-term socio-economic objectives.

<u>1. Intermediate outcomes</u>

<u>1.1 Agricultural production of soybeans</u>

- Yield and sales:
    - Number of acres cultivated with soybean
    - Total soybean harvest

    - Quantity sold
    - Types of buyers of soybean harvest
    - Price per unit of soybeans sold (by type of buyer)
    - Post-harvest losses
    - Quantity consumed
- Self-reported revenues and profits

<u>1.2 Timely and effective harvesting and threshing/employment and labor demand</u>

- Demand for agricultural labor for land preparation/planting and harvesting/threshing (separately):
    - Number of workers hired
    - Number of female workers hired
    - Number of youth workers (under age 35) hired
    - Daily wage/payment arrangement
    - Days each laborer worked on average
- Family labor on farm
    - Average days per week farming soybean
    - Average days per week farming other crops
    - Average hours per day farming soybean
    - Average hours per day farming other crops
    - Number of household members helping on farm during harvesting/threshing
    - Number of days per week household members helped during harvesting/threshing
    - Family labor was freed and pursued other activities during harvesting/threshing

<u>1.3 Input usage and changes in soil management</u>

- Input usage:
    - Fertilizer
        - Has the respondent applied fertilizer on his/her soybean acres?
        - Quantity used on soybean acres
        - Costs (for soybean acres)
        - Source of the fertilizer (self-made/manure or purchased)
        - Timing of fertilizer purchase
    - Pesticides (insecticides, weedicides, herbicides, molukticides, fungicides, akarisides, nematicides)
        - Has the respondent applied any pesticides on his/her soybean acres?
        - Costs (for soybean acres)

<u>1.4 Soil test take-up</u>

- Take-up:
    - Has the respondent ever heard about soil tests?
    - Has the respondent ever tested his/her soil?
    - Has the respondent ever gotten information about soil tests conducted on other farmer's plots?
    - Did the respondent receive any consultation from the field officer based on the soil test?

- Recall of soil test consultation:
    - Did the respondent receive any information about the nitrogen level of soil?
    - Did the respondent receive any information about the potassium level of soil?
    - Did the respondent receive any information about the phosphorus level of soil?
    - Did the respondent receive any information about micronutrients in the soil?
- Soil fertility:
    - How does the respondent rate the fertility of his/her soil? Self-reported soil fertility is a subjective rating based on a 5-point scale ranging from very unfertile to very fertile.
    - Does the respondent believe that soil tests would help farmers to increase their soil health/fertility?
    - How confident is the respondent that his/her input mix for his/her soybean acres is optimal for the soil?
- Knowledge of soil health and nutrient requirements:
    - Which criteria does the respondent use to assess soil fertility?
    - Does the respondent know which input to use when pH is too low?

<u>1.5 Credit take-up, repayment, and access</u>

- Take-up/Participation: Did the respondent take-up the mechanization credit?
- Use: Did the respondent use the credit for harvesting, or threshing, or both?
- Repayment:
    - Was the loan repaid on time (i.e. 20 weeks after the end of the production season)?
    - If the loan was not fully repaid, what was the amount that was repaid?
    - If the loan was repaid late, when was the loan fully repaid?
    - Did any other farmer group member help out with repayment?
    - Was the loan repaid in soybean produce?
    - Was the loan repaid in other produce?

<u>2. Ultimate outcomes</u>

2.1 <u>Shocks and strategies</u>

- Shock occurrences
    - Ex-post strategies used to cope with any household shock
    - Changes in asset holdings (household, agricultural and livestock assets)
- Perceptions and ex-ante adaptation strategies
    - Perceived risk of another drought or flood in the next twelve months
    - Perceptions about changes in rainfall patterns
    - Perceptions about climate change threats within the next ten years
    - Ex-ante adaptation strategies to cope with future drought or flood in the next twelve months
    - Ex-ante adaptation strategies to cope with future drought or flood in the next ten to 15 years
- Weather insurance
    - Respondent has weather insurance
    - Respondent received pay-out from the weather insurance
    - Willingness-to-pay for a hypothetical rainfall index insurance

- Perceptions about weather forecasts and trust in the forecast

<u>2.2 Household income, welfare and financial inclusion</u>

- Economic resilience
  - How much was the income from other crops the respondent sold?
  - Did the respondent engage in other non-farm activities?
  - What has been the most important income source for the household during the past six months?
  - Did the household receive any remittances in the last six months on a regular basis?
  - HH was in financial distress anytime during the last six months? (financial distress: unable to fulfil usual daily expenditures)
  - During the last six months, was there a time when, because of lack of money or other resources the household ran out of food?
  - Did the respondent take out any loan from another lender in the last six months?

- Subjective well-being: where do they locate their household on the 5 steps going from 1=poor to 5=rich
- Subjective relative well-being: where do they locate their neighbors on the 5 steps going from 1=poor to 5=rich

<u>Heterogeneity analysis</u>

We analyze heterogenous impacts for the following characteristics:

- Farm size
- Gender
- Educational level
- Age
- Experience (years cultivating soybeans)
- Risk attitudes
- Having taken out a loan in the previous six months

11. Experimental design

Sampling

Our sample consists of smallholder farmers who either previously collaborated with the platform or became new clients in 2024, prior to the start of the trial. We included only those communities that had reported to WamiAgro that they would be farming soybeans in 2024. The platform selected 17 communities based on these eligibility criteria. Each community comprises between one and 14 farmer groups, and a random subset of farmers was selected from each group.

Due to the planting phase already commencing in five communities in the Northern region before the baseline survey could be conducted in June, these communities were replaced with five others in the North East region, ensuring an equivalent number of farmers per group. In total, 1,363 farmers were surveyed for the baseline.

We will also conduct a short survey among the field officers of WamiAgro to collect data on the effects of weather shocks on the community level. These data will provide additional information how the field officers interact with farmers in the face of weather shocks and how services are delivered.

Regions (Districts) and communities

- Northern Region (Zabzugu)
    - Zabzugu (including Nakohigu)
    - Tindang
    - Gor-Tanei
    - Kukpalgu
- Savannah Region (Sawla-Tuna-Kalba, West Gonja)
    - Tuna
    - Larabanga
    - Alhassan Kura Damongo
- Upper West Region (Sissala West, Wa)
    - Jeffisi
    - Busa
    - Goripie
    - Bihee
- North East Region (Chereponi)
    - Jakpa
    - Kpenchi
    - Nawqari
    - Ando Ngamanu
    - Cherekpegri
    - Famisha

Timeline

Activity	**Timeline**
Baseline Survey	*Jun 24*
Planting	Jul – Oct 24
Interventions: Info Soil test Credit Market	**2024** Jul – Oct (not implemented) (implemented only in 1 out of 78 farmer groups) (implemented only in 3 out of 17 communities)
Harvest	Oct – Nov 24
Sales	Oct – Dec 24
2nd Baseline Survey & Field officer survey	*Jan/Feb 25*
Interventions: Info Soil test Credit Market	2025 agricultural season
Follow-up Survey	*Dec 25*

Randomization procedure

The original 17 communities were divided into five strata based on the number of registered farmers in each community. The random assignment into the three treatment groups and the control group was then done within each stratum. Within each treatment group and the control group, the number of farmers to be interviewed was determined so that the number of farmers drawn per group was inversely proportional to the number of farmer groups in each treatment and control group.

The randomization had been done before the five communities had to be replaced. The random assignment was then transferred from the "old" to the "new communities" with only very minor adjustment to the number of sampled farmers per group. Overall, more farmers were sampled to prevent a loss in power in case of attrition.

Details on services

This study evaluates the effectiveness of different bundles of agricultural extension services on smallholder soybean farmers. We implement a randomized controlled trial (RCT) using a 2x2 design. Farming communities are randomized into one of three treatment groups and a control group. The control group receives the standard services offered by WamiAgro Ltd., which consist of information provision, trainings, and granting smallholders access to international markets. In the first treatment group ("Soil test"), each smallholder receives information on their soil's nutrient composition and crop-specific input recommendations, in addition to the general information provision and market access services provided to the control group. The second treatment group ("Credit") receives harvesting and/or threshing services on credit along with the general information and market access services. The third treatment group ("Soil test + Credit") receives both the soil test and the credit service, in addition to the general information and market access service. This design allows to estimate the combined impact of the soil test and credit services as compared to their individual impacts.

2x2 Design

		Credit	
		No	Yes
Soil test	No	Control "Info + Market"	Treatment 2 "Credit" (includes Info + Market)
	Yes	Treatment 1 "Soil test" (includes Info + Market)	Treatment 3 "Soil test + Credit" (includes Info + Market)

Spillovers

Given the design of our study, we do not expect significant spillover effects. The treatment groups are geographically dispersed, which minimizes the likelihood of interaction between treated and untreated units. Furthermore, the treatment is highly specific and tailored to individual participants, further reducing the possibility of spillovers. These factors together suggest that any unintended influence of the treatment on control groups is likely to be minimal.

However, we cannot be certain to rule out spillover effects for communities that were assigned a different treatment status, but are located within the same region and geographically relatively close to one another. Therefore, we will test whether knowledge and demand for the use of soil tests and credit services is higher in the untreated communities in our sample that are closer in distance to the respective treatment groups. We can estimate this by incorporating the number of treated soybean farmers within a meaningful radius. This variable will be included both linearly and interacted with the treatment to examine whether spillover effects differ between treated and untreated farmers.

Another way of looking at spillovers is to check for impacts directly at the market level. In the case that the treatments lead to significant improvements in productivity for treated farmers, it could potentially affect local labor markets. Increased demand for labor to manage higher yields or expanded cultivation could drive up wages in treated areas. This effect might spill over into neighboring areas, particularly if there is labor mobility between treated and untreated regions. Similarly, if the treatment results in a substantial increase in soybean production among treated farmers, this could affect local soybean prices. An oversupply in treated regions might lead to a decrease in prices, which could spill over into adjacent markets, especially if these markets are closely linked. Conversely, if the treatment enhances the quality of soybeans, it could lead to price premiums that might influence market prices more broadly. While these impacts are plausible, their magnitude would depend on the scale of the treatment effects and the degree of market integration across regions. It is important to monitor these potential spillovers as they could have broader economic implications beyond the immediate scope of the study. In order to better understand any potential spillover effects, we will collect qualitative data via focus group discussions, interviewing famers, farmer group leaders and field agents in the control and treatment groups.

Model specification (ANCOVA)

The primary objective of this analysis is to estimate the effect of the intervention on the outcome variables measured after the intervention, controlling for baseline outcomes and other covariates.

The standard ANCOVA model will be specified as follows:

$$Y_{ict} = \beta_0 + \beta_1 SoilTest_c + \beta_2 Credit_c + \beta_3 (SoilTest_c \times Credit_c) + \lambda Y_{ic0} + X_{ict}{}' \gamma + \varepsilon_{ict}$$

Where Y_{ict} is the post-treatment outcome for farmer i in community c measured in survey wave t. $SoilTest_c$ is the treatment indicator equal to 1 if the community c received Treatment 1 "Soil test", and 0 otherwise. $Credit_c$ is the treatment indicator equal to 1 if community c received Treatment 2 "Credit". $(SoilTest_c \times Credit_c)$ represents the interaction term between the two treatments, which captures any combined effect of receiving both treatments (or, Treatment 3 "Soil test + Credit"). β_1, β_2, and β_3 are the effects of either treatment relative to the control group. Y_{ic0} are the baseline outcomes. X_{ict} is a vector of control variables including status at baseline (i.e. having cultivated soybeans in the past or being new soybean farmer and having received a service from WamiAgro in 2024) and ε_{ict} is the error term. The intervention was randomized at the community level. Therefore, we cluster the standard errors at the community level.

Covariates X_i:

- Age
- Gender

- Household head status
- Education
- Region
- Farm size
- Experience (years cultivating soya)
- Having received a service from WamiAgro in 2024

Multiple Hypothesis Testing

Given the large number of outcomes estimated in our study, there is an increased risk of falsely rejecting the null hypothesis due to multiple comparisons. To address this, we will control for the False Discovery Rate (FDR) using the Benjamini-Hochberg (1995) procedure. Specifically, sharpened q-values will be calculated within each domain for the groups of outcomes.

12. Was the treatment clustered?

Yes, by community.

13. Planned number of clusters
    78

14. Planned number of observations

At baseline, we interviewed 1,363 farmers. We had to drop four farmers from our sample who stated that they did not cultivate soybeans, nor were they planning to cultivate soybeans in the future and we had to drop one respondent who was the only one in their farmer group who cultivated soybeans. This reduces the number of farmers to 1,358. We plan to interview the same farmers for the second baseline and the follow-up survey. This will result in a total of 4,074 observations for outcomes measured in all three survey rounds and 2,716 observations for outcomes measured in two survey rounds. In addition, depending on available financial resources, we plan to conduct one or two short phone surveys. We anticipate an attrition rate of approximately 5% due to factors such as non-response, traveling, migration, and death. We will observe and document any kind of attrition carefully. We will analyze whether it is systematically related to the treatment status.

Power

We used our first baseline data to perform power calculations as it provides information on the means, standard deviations and intra-cluster correlation of key outcome variables. Our power calculations are for intention-to-treat effects. The MDEs are based on a 95% confidence interval and a power of 80%.

The baseline data shows a mean of harvest that was sold in 2023 of 784.05 kg of soybeans, with a standard deviation of 773.37 kg and an intra-cluster correlation (ICC) of 0.23. With a simple post-intervention cross-section and given our average cluster size of 17.41 individuals per cluster and 78 clusters in total, we are powered to detect a minimum effect of 280.6 kg, or 0.36 standard deviations. Conditioning on baseline outcomes and covariates, i.e. estimating ANCOVA specifications, will further absorb noise from the data and hence increase power. Further follow-up surveys are planned which will allow to estimate ANCOVA specifications over several waves of post-intervention data. In terms of power, this is especially helpful for those outcomes that are difficult to measure and typically show a relatively low autocorrelation over time, such as sales and profits (McKenzie, 2012).

The baseline data shows a mean of sales made from soybeans in 2023 of 3,954.58 GHS, with a standard deviation of 3,827.16 and an intra-cluster correlation (ICC) of 0.21. Given our average cluster size of 17.41 individuals per cluster and 78 clusters in total, we are powered to detect an MDE of 1,329 GHS, or 0.35 standard deviations. With regard to soybean yields (kg/acres) in 2023**,** the baseline data show a mean 327.65, with a standard deviation of 197.17 and an ICC of 0.32. The MDE is 83.99, or 0.43 standard deviations for a simply ex-post comparison. Again, in both cases adding a further wave of data, conditioning on baseline outcomes and covariates will further increase power.

15. Was IRB approval obtained?
    a. IRB Name: Ethics Committee for Humanities, University of Ghana
    b. IRB Approval Date: 11. June 2024
    c. IRB Approval Number: ECH 277 23-24