Promoting a Bundle of Biofortified Seeds in Gombe, Nigeria: A Pre-Analysis Plan

September 2024

1. Introduction

Smallholder farmers in northeastern Nigeria face the triple threat of food insecurity, climate volatility, and conflict. Northeastern Nigeria has some of the highest levels of food insecurity in Nigeria (UNICEF, 2023). Additionally, delayed, shorter, and more volatile rainy seasons have led to massive floods, depleted soil quality, and disrupted agricultural growing seasons which ultimately constrain agricultural productivity. Moreover, conflict between pastoralist livestock herders and settled agricultural communities over land use, as well as extremism, has led millions of people to flee from their homes (International Crisis Group, 2023).

These threats are closely interrelated. As in much of sub-Saharan Africa, climate change has extended dry seasons in Nigeria. This change in seasonal rainfall patterns both limits local agricultural production and disrupts long-standing symbiotic relationships between pastoralist livestock herders and settled agricultural households, leading to conflict between these groups (McGuirk and Nunn, 2023). Exposure to conflict itself contributes to reduced agricultural production (Adelaja and George, 2019) and increased food insecurity (George et al., 2020). With limited social safety nets available, many households exposed to conflict either turn to informal self-employment while reducing agricultural labor (Bloem et al., 2023) or migrate domestically in search of safety, peace, and security.

Desertification and depletion of soil quality results in a cycle of yield and nutrition losses and adverse health effects associated with poor food micronutrient quality. As a result, this area suffers from high levels of anemia and zinc deficiency which can together adversely affect both cognitive, physical development and immunity (NPC and ICF, 2019). A key hurdle to productivity and risk management is that farmers do not simultaneously adopt an appropriate mix of improved agriculture technologies, choosing instead to perhaps expand area under cultivation or invest in discrete inputs that are more affordable—such as improved seed or specific pesticides—and rarely applying these inputs together on the same plot (Christiaensen, 2017).

In this pre-analysis plan, we document pre-registered analysis plans for data collected during the implementation of a randomized control trial designed to address these interrelated challenges by promoting the adoption of a bundle of agricultural inputs, including: biofortified seeds, fertilizer, information, crop protection products and insurance among smallholder farmers in Gombe state, Nigeria. The goal of bundling agricultural inputs together is to leverage agro-ecological complementarities that are critical for optimal crop production. These bundles address recent research that finds differences in estimated yield gains between agronomic trials and farmers using the same inputs under real-life conditions (Laajaj et al., 2020). One explanation for these differential yield effects is the limited adoption of complementary agricultural inputs and practices among smallholder farmers (Liverpool-Tasie et al., 2022). Thus, promoting complementary agricultural inputs together as a bundle holds the potential to close the gap between yield effects measured in agronomic trials and those measured under real-life conditions when adopted by smallholder farmers. Additionally, across northern Nigeria, availability of micronutrient enriched biofortified varieties has increased significantly since 2019 with support from the Agriculture Food Security Nutrition Strategy, Government of Nigeria and the Nigeria Agriculture Technology Innovation Policy (Birol et al., 2023). The availability of these seeds provides a real opportunity to address nutrition challenges while promoting the uptake of complementary agricultural inputs.

Our objective is twofold: (i) to assess smallholder farmer demand for these agricultural inputs marketed as a bundle, and (ii) to evaluate the effect of receiving this agricultural input bundle on agricultural productivity and indicators of potential nutritional benefits among participating households. The implementation of this study, designed to estimate smallholder farmer demand for a bundle of biofortified seed and other agricultural inputs, will provide policy-relevant insights about how to effectively promote the adoption of agricultural technology, and possibly in-turn improve farm productivity and nutrition outcomes, among vulnerable populations in a fragile setting.

2. Intervention

In a partnership between IFPRI, Harvest Plus, and Premier Seeds LTD Nigeria, we will implement a randomized control trial with three treatment groups and one control group. Causal identification comes from our randomization of treatment status across communities in Gombe State, Nigeria. Additional details are discussed below.

3. Research Design and Data

An initial survey with local government extension officers, a qualitative study with farmers and discussions with ICRISAT informed the design of the intervention. Three design considerations influence our choice of technologies in the biofortified bundle. First, farmers in the selected LGAs grow a mix of pearl millets and cowpea (intercropped) and maize in rotation. The bundle thus includes biofortified high iron pearl millets, cowpea, and Vit A maize seed varieties which are well tested for the agro-ecology of the area.

Second, fertilizer costs have increased considerably in Nigeria following the Russia-Ukraine war. For maize in Nigeria, the marginal physical output and profitability from fertilizer use are low, which can deter farmer investment (Liverpool-Tasie et. al, 2015). Additionally, fertilizer and improved seed tend to not often be applied intensively on the same plot. Moreover, microdosing can support an increase in maize yields and limit environmental damages associated with nitrogen runoff. The bundle therefore includes half the typical NPK quantities that are generally applied with broadcasting approaches.

Finally, crop protection products (weedicide, pesticide and herbicides) and a cost-effective insurance product are provided as part of the bundle. The insurance product is a weather index insurance offer (provided in partnership with Pula insurance) designed to protect farmers against loss of seed due to delayed rainfall. The bundle is applicable for a quarter hectare plot (i.e., the typical plot size for the region) and pictorial and verbal advisories will provide detailed information on crop management.

The bundle is offered by Premier Seeds LTD Nigeria, a commercial agricultural input company, which has a multinational presence in Africa and can directly distribute its product through their agro-dealer network in the study area. At the initiation of our intervention, the total cost of the bundle was approximately 74USD without discount. The specific quantities of each of the components of the bundle are as follows:

- 5kg of Sammaz 52 – Maize
- 2kg of IPM – Chakti
- 6kg of Sampea 14 cowpea
- 50kg of NPK
- 50kg of urea
- 1 liter of cypermethrin
- 1 liter of pendimethalin
- 3 sachets of seed dressing
- Insurance coverage for one growing season

3.1 Experimental Design

Our experiment includes the following three treatment groups and control group.

- Treatment 1 (T1): provides farmers with extension services and advisories relating to the kit but does not offer any discount on the purchase of the bundle. Extension will be offered at the community level. Additionally, farmers receive an informational brochure advertising the benefits of the bundle. Although farmers in this group do not receive any discount, they do receive a "marketing flyer" with unique IDs and the address of relevant agro-dealers. The marketing flyer is distributed both via SMS and in paper form during door-to-door visits.

- Treatment 2 (T2): provides farmers with a 50 percent nominal subsidy on the purchase of the bundle, in addition to the same community-level marketing activities provided in T1. Nominal because the actual price of the kit may vary due to volatility in the rate of imported inputs such as urea, however the seed company pre-purchased inputs for the bundle and fixed the price at this level.

- Treatment 3 (T3): provides farmers with a 75 percent nominal discount, in addition to the same community-level marketing activities in T1. All other modalities stayed the same as for T2 farmers.

- The control group (C) received no intervention. Premier Seeds is allowed to use regular market channels to promote improved seed but not marketed and sold as a bundle.

We hypothesize that information and prices are key constraints on the purchase of biofortified seeds and agricultural inputs among smallholder farmers in Gombe State, Nigeria. By comparing differences in purchase rates of the bundle between T1 and C communities, we test the effect of providing information, via community extension marketing campaigns and door-to-door visits, on take-up of the bundle. By comparing purchase rates between T2 and T3 communities relative to C communities, we can test the effect of price discounts on take-up of the bundle. By comparing differences in purchase rates of the bundle between T2 and T3 communities relative to T1 communities, we can estimate the slope of the demand curve for the bundle. As noted in more detail below, we also examine treatment effects on other downstream outcomes, such as measures of agricultural production and indicators of biofortified crop consumption.

3.2 Power Calculations

We use the following assumptions for the power calculations. According to the World Bank's LSMS (4^{th} round) data from 2020 in Nigeria, roughly 11 percent of the farmers use improved maize seed on at least one plot in the 2023 wet season. We also assume the following parameters, set at baseline: we have 10 farmers within each cluster (i.e., community) and 230 communities in our sample. Assuming an inter cluster correlation of 0.2 and that 11 percent of the comparison group correctly uses "improved varieties" of seeds, when comparisons are made between the control group and any of the treatment groups, we will be able to detect a change of 30 percent (a 3.4 percentage point change from 11 percent to 14 percent). An 11 percent rate of "correct" use of improved seed, however, might represent an overestimate. So, in the case that this rate is lower, for example at one percent, then we will be able to detect a change of 1.5 percentage points (a change from 1 percent to 2.5 percentage points).

So far, these power calculations assume no attrition. If we experience 20 percent attrition, equally distributed across each of the treatment and control groups, these power calculations will change slightly. Assuming an inter cluster correlation of 0.2 and that 11 percent of the comparison group correctly uses "improved varieties" of seeds, when comparisons are made between the control group and any of the treatment groups, we will be able to detect a change of 30 3.6 percentage points change from 11 percent to 15 percent. In the case that only one percent of the comparison group correctly uses improved seeds, for example, we will be able to detect a change of 1.6 percentage points from 1 percent to 2.6 percentage points.

3.3 Sample Selection and Data

We implement a cluster randomized control trial in Gombe, Nigeria. At the community level we randomly assigned 230 communities into one of either three treatment groups or the control group, as defined above. Within each community, we randomly select 10 farmers to generate our sample from a preliminary listing survey.

Timeline

- November 2023: Project begins, baseline survey
- May 2024: Intervention begins
- August 2024: Intervention ends
- October 2024: Endline survey

4. Outcomes and Economic Specifications

Our analysis includes two parts. First, we estimate the effect of the intervention on the purchase of the bundle and its use by each of the farmers in our sample. This "take-up" analysis represents an important first step in the sense that if the intervention does not inspire the purchase and effective use of the bundle, and its components, then we will have no reason to expect effects to materialize on other "downstream" outcomes. Second, we estimate the effect of the intervention on "downstream" outcomes such as agricultural production and biofortified crop consumption (i.e., an indicator of possible nutritional benefits).

4.1 Regression Specification

As this is a randomized controlled trial, our core regression specification is straightforward and defined as follows:

$$Y_{ci} = \alpha + \beta_1 T1_{ci} + \beta_2 T2_{ci} + \beta_3 T3_{ci} + \boldsymbol{X}'\delta + \epsilon_{ci}$$

In this regression Y_{ci} is one of our outcomes of interest, as discussed in more detail below, for farmer i in community c. The variables Tj_i for all $j = 1, 2, 3$ represents one of each of our three treatment groups, with the indicator for control group farmers omitted. These treatment group indicators are mutually exclusive. The vector $\boldsymbol{X}$ represent control variables, specifically including baseline values of Y_i, our outcome of interest, if the outcome was measured at baseline. Standard errors will be clustered at the community level.

This regression specification estimates the intent-to-treat effect of the random assignment of the intervention on each outcome of interest. This regression specification does not consider compliance with the intervention and, therefore, does not estimate the effect of participating in the intervention. To calculate the effect of participating in the intervention, we will discuss in the write-up of our analysis how these intent-to-treat effect estimates can be scaled by the intervention participation rate among the farmers in the treatment groups.

Missing Values: In all our specifications, if more than 30 percent of a pre-specified key outcome variable is missing, we will use available co-variants to predict the value of this outcome variable. If less than 30 percent of a key outcome variable is missing, we will restrict our analytical sample to the respondents for whom we have all outcome variables.

4.2 Outcomes of Interest

Our outcomes of interest fall into two categories. First, we use outcomes that measure the purchase of the bundle, the use of each of the components of the bundle, and the implementation of the recommended agricultural practices. Some of these outcomes will come directly from monitoring data collected by the agro-dealers affiliated with Premier Seeds LTD Nigeria and the other outcomes will be collected with our endline survey. Second, we will use outcomes that measure "downstream" outcomes measuring agricultural production and biofortified crop consumption. Each of these outcomes will be collected with our endline survey.

4.2.1 Take-up—We measure take-up in several ways. First, and most fundamentally, we consider the purchase of the bundle as the primary measure of take-up. Second, we also consider both the self-reported use of each of the components of the bundle and the self-reported implementation of the recommended agricultural practices.

1. We use sales data we collected from agro-dealers when monitoring our intervention. These sales data define take up as whether the farmer purchased the bundle within the study period.
2. We use the following information collected in our follow-up survey:
    a. Self-reported purchase of the bundle by farmers

b. Self-reported use of each of the components of the bundle (conditional on purchasing the full bundle from the agrodealer?)
    i. Planted 2kgs of pearl millet (IPM – Chakti)
    ii. Planted 6 kgs of cowpea seeds (Sampea 14 cowpea)
    iii. Planted 5 kgs of vitamin A maize seeds (Sammaz 42 maize)
    iv. Used 50 kgs of NPK fertilizer
    v. Used 50 kgs of urea fertilizer
    vi. Used one liter of cypermethrin
    vii. Used one liter of pendimethalin
    viii. Used three sachets of seed dressing
    ix. Self-reported knowledge of insurance coverage
c. Self-reported implementation of recommended agricultural practices
    i. Applied all bundle components on one quarter hectare plot
    ii. Intercropped maize and cowpea (i.e., two rows of maize and two rows of cowpea)
    iii. Intercropped millet and cowpea (i.e., two rows of millet and two rows of cowpea)
    iv. Weeded the plot every three to four weeks
    v. Planted seeds after rainfall is fully established
    vi. Applied pendimethalin shortly after planting
    vii. Applied the first batch of fertilizer (i.e., 50 kgs of NPK and 25 kg of urea) 10 days after sowing
    viii. Used microdosing technique when applying fertilizer
    ix. Applied the second batch of fertilizer (i.e., 25 kgs of urea) four-five weeks after sowing
    x. Applied cypermethrin shortly after the cowpeas begin to flower
    xi. Harvest crops as they mature and store in a dry place for two weeks

4.2.2 Agricultural productivity—A primary potential benefit of marketing these agricultural inputs as a bundle is the agricultural productivity effects associated with agronomic complementarities between improved seeds, fertilizer, and agro-chemicals. Our endline survey includes a detailed module that measures the amount of each crop the respondent harvested on each of their plots. This provides us with plot-crop level data on agricultural production (measured in terms of kg harvested). We can combine this information on plot area cultivated (measured in terms of hectares cultivated). To generate a measure of yield at the household level, we aggregate crop-plot level measures of kg harvested to the household and divide by the plot level measure of hectares cultivated aggregated to the household. We use this yield measure, along with the kg harvested and hectares cultivated components as outcome variables (to minimize the influence of measurement error), expressed in terms of levels, as outcomes variables measuring agricultural productivity.

Given that the promoted bundle is specifically formulated for use on a one quarter hectare plot, agricultural productivity effects might not materialize at the household level. Rather, we might only be able to observe these effects at the plot level. Specifically, on the plot where the bundle was planted (among those who purchased the bundle) and on a comparable plot where the farmer usually would plant improved maize, millet, and cowpea seeds (among those who did not purchase the bundle).

We ask farmers to indicate the usual plot on which they use improved seeds of maize, millet and cowpea. We then ask farmers who purchased the bundle which plot they planted the bundle. With these data, we will compare the overlap between these plot identification questions among farmers who purchased the bundle. The extent of the overlap will validate the following two comparisons, which we will make using the regression specification defined above. First, we will compare yield (i.e., kg harvested divided by hectares cultivated) for the plot specified as the "usual plot," as discussed above. As in the household-level analysis discussed above, we will also separate the kg harvested and hectares cultivated into two separate outcome variables to minimize the influence of measurement error. Second, we will compare yield for the plot where the bundle was planted for those who purchased the bundle with the "usual plot" for those who did not purchase the bundle. The measures of yield, kg harvested, and hectare cultivated will be defined as discussed above for each of the identified plots.

It is important to note that none of these measures of agricultural productivity are without limitations. The household-level analysis might obscure effects by aggregating plots together for each household. The first approach for plot-level analysis risks the farmer not identifying the plot where the bundle was planted. Finally, the second plot-level approach risks making comparisons between incomparable plots. This analysis plan, however, aims to conduct three types of complementary analysis with the aim of "triangulating" any true effect of the intervention on agricultural productivity.

4.2.3 Biofortified crop consumption—An important possible benefit of promoting the purchase and cultivation of biofortified seeds is the nutrition benefits associated with consuming the crops grown from these seeds.

We ask the following set of questions to measure the consumption of biofortified crops. First, we ask if the respondent, or any member of the respondent's household consumed food that the household cultivated and processed. If the respondent answers affirmatively, we then ask them to list the crops that their household cultivated, processed, and ultimately consumed. If the respondent indicates either maize, millet, or cowpea, we ask if the maize was—in part or in whole—cultivated with the specific biofortified variety of seed included in the bundle. These questions generate three binary indicators of biofortified crop consumption that we use as indicators of possible nutrition benefits that could potentially materialize in the future.

4.3 Heterogeneity

Given the motivation of this study in a fragile setting with a relatively large number of IDPs, a primary source of heterogeneity based on IDP status. We hypothesize that we might find differential effects among IDPs than among the existing population. On the one hand, IDPs might be more resource constrained and price sensitive, leading the discounts to inspire the purchase of the bundle. On the other hand, IDPs might have constrained access to cultivatable land, which might limit their willingness to invest in agricultural inputs. Using data from our baseline survey, we identify farmers who have been internally displaced. With this data we categorize 17 percent of our sample as being internally displaced. We will, therefore, estimate an augmented version of the regression specification defined above that includes an indicator expressing the IDP status of the respondent and this indicator interacted with the treatment status variables. This regression will allow us to test differences in the effect of the intervention based on the IDP status of the respondent.

4.4 Additional Exploratory Analysis

This document does not contain a comprehensive description of all the analysis we will conduct. We plan to do additional exploratory analysis with our data to better understand our main results expressed above. This additional exploratory analysis will be framed as such in any write-up of our results.

For example, we will aim to understand what farmers are doing with their crops after production. While our endline survey will be conducted at a time before all farmers generally have completed all post-production sales and/or processing activities, we might be able to partially understand these activities based on the actual timing of our endline survey with the specific 2024 agricultural season on Gombe by asking about the amount of a crop stored for future consumption and stored for future sales.

4.5 Multiple Hypothesis Testing Adjustments

We will conduct multiple hypothesis testing adjustments across the main outcomes measuring take-up of the bundle noted in Section 4.2.1 above. Specifically, we consider measures of purchasing the bundle, both with the sales data provided by agro-dealers our self-reported data, as one “family” of outcomes. Next, we consider measures of the use of each of the components of the bundle as a second “family” of outcomes. Finally, we consider measures of the implementation of recommended practices as a third “family” of outcomes. Specifically, we will construct adjusted q-values using the Anderson (2008) Family-Wise Error Rate (FWER) adjustment. We will not conduct multiple hypothesis testing adjustments on the main outcomes measuring agricultural production and biofortified crop consumption, as these are all different representations of the same conceptual outcome.

4.6 Attrition

We define attrition as a respondent not responding to the endline survey. We will test for differences by study arm (e.g., T1, T2, T3, and C) in the attrition rate of respondents who were surveyed at baseline. Non-random attrition that is correlated with our treatment assignment could threaten the validity of our experiment. If we find evidence of non-random attrition correlated with our treatment assignment, then we will construct inverse probability weights to correct for non-random attrition based on observable characteristics included in our survey data. We will also use Lee bounds to address the attrition if there is any (Lee, 2009).

5. Data Analysis Completed Thus Far

We have conducted preliminary descriptive analysis on the baseline data collected at the end of 2023. We have also conducted preliminary data assessing purchases of the bundle using the monitoring data collected by the agro-dealers affiliated with Premier Seeds LTD Nigera. No additional data analysis has been initiated or completed thus far.