Impacts of Africa RISING in Mali

Beliyou Haile[1], Carlo Azzarri[2], Ivan Tzintzun[3], Sedi-Anne Boukaka[2], and Sveva Vitellozzi[4]

Author affiliations [1]World Bank, [2]International Food Policy Research Institute, [3]Paris School of Economics, [4]University of Bologna

Published by International Food Policy Research Institute

May 2024

www.africa-rising.net

The Africa Research In Sustainable Intensification for the Next Generation (Africa RISING) program comprises three research-in-development projects supported by the United States Agency for International Development (USAID) as part of the U.S. Government's Feed the Future initiative.

Through action research and development partnerships, Africa RISING is creating opportunities for smallholder farm households to move out of hunger and poverty through sustainably intensified farming systems that improve food, nutrition, and income security, particularly for women and children, and conserve or enhance the natural resource base.

The three regional projects are led by the International Institute of Tropical Agriculture (in West Africa and East and Southern Africa) and the International Livestock Research Institute (in the Ethiopian Highlands). The International Food Policy Research Institute leads the program's monitoring, evaluation and impact assessment.

Africa RISING appreciates support from the American people delivered through the USAID Feed the Future initiative. We also thank farmers and local partners at all sites for their contributions to the program.

Executive summary

This study evaluates the impact of Africa RISING, a sustainable intensification (SI) program, implemented in Bougouni, Yanfolila, and Koutiala *cercles* in southern Mali beginning in 2012. Using a participatory action research framework, the program validated and promoted alternative SI options including fertilized groundnut and sorghum, crop-legume intercropping, intercropping of two compatible legumes, access to extension services, and fertilizer microdosing, while preserving ecosystem services in the face of projected population growth and climatic changes. Impact is estimated on several SI indicators and domains using two rounds of quasi-experimental panel data (surveys conducted in 2014 and 2022) and difference-in-differences techniques. The unique study design allows us to estimate the impact of Africa RISING by comparing outcomes among program beneficiaries with two different counterfactual groups—one located inside program villages (within-village comparison) and another in non-program (control) villages (out-of-village comparison) on several indicators across five SI domains—environment, productivity, economic, human, and social. We also conduct a *placebo* test comparing non-beneficiaries in the two counterfactual groups. We find few statistically significant differences in the averages of the characteristics in the environmental and productivity domain among households in the within-village and out-of-village comparisons, most likely because of misreporting of program participation. Overall comparisons between households in target and non-target villages show a positive impact of AR on environmental variables such as access to extension services, and adoption of improved crops; on productivity variables such as green bean, cotton and okra yield; and on economic variables such as an increase in the non-agricultural wealth index; but no statistically significant effect on human and social indicators, namely household dietary diversity, food consumption scores, and nutritional indicators for children 0–59 months old and women 15–49 years old. Estimates based on within-village, out-of-village, and *placebo* comparisons suggest important insights about the challenges in assessing the impact of agricultural programs in general and, specifically, participatory multi-intervention programs in the presence of sample (self-)selection and spillovers. Our study highlights useful empirical lessons learned to inform future program design and impact assessments.

1. Introduction

Despite improvements over the past several decades, sub-Saharan Africa (SSA) lags behind other developing regions on several fronts, including productivity gains, poverty reduction, and food security. Average cereal yields have stagnated at 1.6 ton/ha for SSA since 2020, about a third of the yield average in Southeast Asia (FAO 2022). SSA is also the only region where the number of chronically undernourished (stunted) children rose between 2000 and 2019, as the high rate of population growth outpaced the slower rate of reduction in stunting prevalence (UNICEF, WHO, and World Bank 2020). Demographic and Health Surveys data show that 32 percent of children under five years old in the region are chronically undernourished (ICF 2022). The extreme poverty headcount rate, defined as the share of population living below the $1.90 (per capita/day) international poverty line, declined from 55 percent to 40 percent between 1990 and 2018 in SSA, while other developing regions such as South Asia experienced a higher rate of reduction—from 50 percent to 15 percent (Schoch and Lakner 2020). In addition, temperatures are projected to increase faster as a result of climate change in SSA than the global average, posing a significant threat to agricultural production and food.

One approach pursued to enhance agricultural productivity, food security, and ecosystem services of the smallholder sector in SSA is sustainable intensification (SI). SI aims to improve resource-use efficiency, while increasing food supply from the same resources and enhancing beneficial environmental and social services (Garnett et al. 2013; Pretty et al. 2011; Smith et al. 2016). It provides a conceptual framework to achieve balanced outcomes across different dimensions of sustainability (Smith et al. 2017; Conway, Wilson and Wilson 2013). SI encompasses a wide range of innovations, including prudent use of chemical fertilizers, improved cultivars, integration of legumes and livestock into cereal-dominated farming systems, integrated soil and water conservation, crop rotation, agroforestry, and incorporation of manure and crop residues (D'Souza, Cyphers and Phipps 1993; Lee 2005). The approach departs from the historically narrow focus on enhancing productivity of a few staple crops and overall profitability, with little consideration of their linkages with nutrition and health (Bouis and Welch 2010).

One of the largest SI programs globally is *Africa Research In Sustainable Intensification for the Next Generation–Africa RISING* (AR).[1] This program has been implemented in six African countries since 2012—Ethiopia, Malawi, Tanzania, Zambia, Mali, and Ghana. Its goal is to "to provide pathways out

[1] More details about the program can be found here https://africa-rising.net/.

of hunger and poverty for smallholder families through sustainably intensified farming systems that sufficiently improve food, nutrition, and income security, particularly for women and children, and conserve or enhance the natural resource base” (IITA, IFPRI and ILRI 2012: p. 5). Phase I (2012–2016) focused on the validation of alternative demand-driven SI options with the potential to alleviate poverty and improve nutrition, equity, and ecosystem stability. Phase II (2016–2021) focused on scaling up validated SI innovations alongside development partners (IFPRI, IITA and ILRI 2016).

This impact evaluation study focuses on the Mali case, implemented in Bougouni, Yanfolila, and Koutiala *cercles* (administrative districts) in the Sikasso region of Mali. The innovations promoted included a cropping systems dimension focused on the adoption of improved crop varieties (sorghum, groundnut, okra, eggplant, and tomato) and farming methods (cereal-legume intercropping, cereal-vegetable intercropping, and fertilizer microdosing). Livestock systems interventions aimed at improving small ruminant production through improved feeding and vaccination. Natural resource management activities focused on the reduction of soil erosion through contour bunding. Finally, program interventions included a series of nutrition-oriented trainings directed at extension workers and women as well as the establishment of nutrition support groups. In addition, AR project in Mali aimed to foster collaboration between CGIAR centers, NARS, farmers, the private sector, universities, extension services, and government agencies, and to build human resources through short-term and post-graduate trainings (MSc and PhD).

2. Materials and Methods

2.1 Setting

Located in West Africa, Mali is one of the poorest countries in the world, marked by a high vulnerability to political instability and climate change. Since the outbreak of armed conflict in 2012, Mali has been dealing with economic and humanitarian crises, including food shortages, lack of basic services in conflict areas, and a significant increase in internally displaced people (AfDB, 2020). In addition, over the past years climate-related issues such as droughts, floods, and crop pests have posed major challenges to economic development and food security all over the country (World Bank, 2023).

Mali relies heavily on agriculture, with 80 percent of the working-age individuals employed in the sector (World Bank, 2023). Key subsistence crops include millet, sorghum, and maize, while cotton and rice are largely commercial crops (World Bank, 2011). Because of a lack of modernization, the

Malian agriculture sector faces structural challenges: value chains are underdeveloped, and the agricultural production is highly dependent on weather conditions and vulnerable to shocks (AfDB, 2020). Over the past years, the country has experienced a drop in annual rainfall and precipitation, as well as an increase in extreme weather events such as flooding and droughts (Nagarajan, et al., 2022). As a result of climatic shocks and armed conflict, food insecurity and poverty rates have risen dramatically in recent years: the proportion of the population that is undernourished increased from just over 3 percent in 2017 to nearly 10 percent in 2020, while the national poverty rate increased from 42.1 percent in 2018 to 46.9 percent in 2022 (World Bank, 2023).

A wide variety of risks, including climate change and high population growth rates, threatens the sustainability of Mali's agriculture sector. With an undiversified, low-income economy, food security and poverty are particularly dependent on agricultural market fluctuations. Between 2011 and 2015, extreme poverty has risen to 48 percent of the population, and reached more alarming levels in the southern part of the country. In addition to exogenous shocks, Malian farmers suffer from inadequate and inefficient market linkages that significantly constrain the adoption of productivity-increasing technologies. Among the policy instruments pursued to promote pro-poor economic growth through agriculture, improving agricultural market integration has become a priority. Low integration of input markets may result from thin markets (e.g., for labor and credit) or markets that are unavailable or difficult to access (e.g., for improved cultivars). Inadequate infrastructure, high costs of storage, and non-competitive output markets may also constrain the production of a marketable surplus by discouraging producers from striving to meet demand, both in terms of quantity and quality (Arias et al., 2013). Other knowledge-related factors about basic and more complex agronomic practices as well as marketing options can also deter adoption and limit agricultural revenue.

This study examines the impact of AR activities on a set of environmental, productivity, economic, social, and human indicators drawn from the Sustainable Intensification Assessment Framework (SIAF) (Musumba, 2017)-including cereal yield, agricultural production diversity, net crop and livestock income, household dietary diversity, household food security, and poverty status.

2.2 Data and summary

Panel data for the empirical analysis were collected as part of the monitoring, evaluation, and impact assessment of the Africa RISING program. Baseline and follow-up quasi-experimental data were collected in 2014 and 2022, respectively, covering 10 program target and 10 program non-target (control) communities and two groups of households (Azzarri, et al., 2016). The first (group 1) consisted of 319 households.[2] However, 112 households in group 1 reported not participating in the program during the follow-up survey in 2022. The second group of control households (group 2) consisted of a random sample of 292 households from control communities. There was no expectation or plan that the program would expand to control communities over time.

A stratified random sample (in the control villages) and a random selection of households among beneficiaries (whose participation in the project was non-random) were used in the sampling design. Stratification for the control group was based first on the *cercle* and then on the village. The stratified random sample for the control villages was generated by aggregating subsamples drawn independently from each stratum of the population to generate the control village sample, and the full final sample was then obtained by aggregating the subsamples of each stratum. Table A1 in the Appendix reports the distribution of treatment and control households in the three targeted regions. The largest number of households in the sample are located within the Koutiala *cercle*, which is the most populous *cercle* in the region (Azzarri, et al., 2016).

Households in the control group were identified using similar criteria in the Koutiala and Bougouni cercles. The main criteria for selecting program target areas in Koutiala included road accessibility (during the dry and rainy seasons), presence of important implementing partners (such as AMEDD and AMASSA), and presence of agricultural cooperatives and community unions. Control villages in Bougouni *cercle* were selected based on their agricultural potential and market access, presence of key project partners, and accessibility of the sites. In both Koutiala and Bougouni, additional information was used to ensure that the sites fit the requirements for carrying out activities in the three target intervention domains (seed systems, livestock, land, fodder, and nutrition). After selecting 10 control villages with a probability proportional to size, 35 households from each village were sampled to form the control group (group 2). In Yanfolila, where only one treatment village was targeted by AR, the selection of the control village followed the same criteria adopted in Bougouni and Koutiala (Azzarri, et al., 2016).

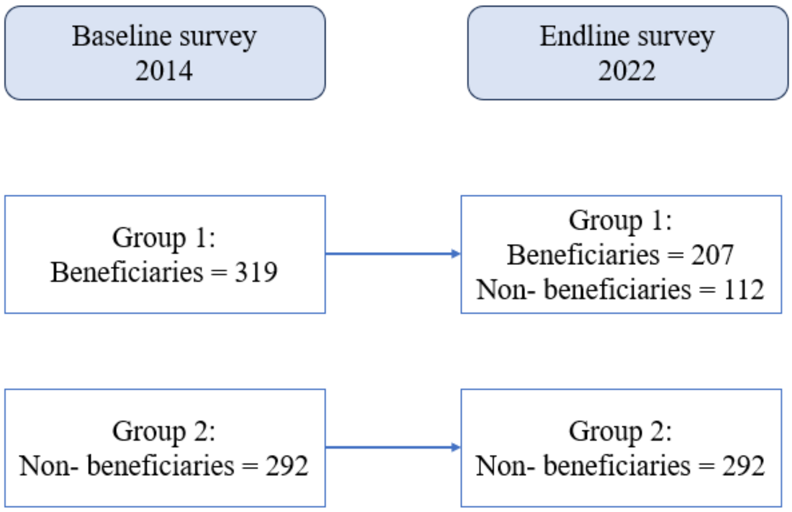
The biggest challenge in selecting eligible households for the baseline survey was related to the definition of the household unit. The Compagnie Malienne pour le Développement du Textile (CMDT) released the list of eligible households, based on the household definition as a "productive unit", which typically consists of an extended family of households farming together. The CMDT's list only included the names of the "concession/compound" heads, excluding the names of all other members who worked there. Consequently, in order to match beneficiaries with the households

[2] The number of participant households at baseline was 351. However, 32 households were lost due to attrition in the follow-up survey.

they belonged to, it was necessary to match individual beneficiaries listed within the "concession/compound" to its head. This exercise required a clear definition of the household unit.[3]

A follow-up survey was conducted in January 2022, and the impact assessment is based on a balanced panel of 611 households interviewed in each wave. Comparable, highly consistent, structured questionnaires were used to collect data on a wide range of topics including household demography, agricultural production, food consumption, non-food expenditure, and anthropometric measurements among children under the age of five years and women aged 15 to 49 years. The reference period for agricultural production data was the main cropping season while food consumption data were based on 7-day recall. One of the follow-up questions asked was about participation in the Africa RISING program between 2014 and 2022; households that reported participating in the program for at least three years were defined as program beneficiaries. Figure 1 summarizes the study design: around 35 percent of households in group 1 identified as agricultural trial farm households in the project monitoring data at baseline reported never participating in the program at follow-up data collection.

Figure 1. Survey design



Note: Numbers in the figure refer to households.

[3] The definition of household was phrased this way: "*The household is defined as a group of people who share expenses, live, and eat together most of the time (that is, at least 3 months in the past 12 months and at least three days in a typical week). The group may also share expenses and income with other groups (of individuals) living in the concession/compound. In addition to meals shared only by group members, the group may share communal meals with others in the concession/compound. A newborn less than 3 months must be considered a household member.*" (Azzarri, et al., 2016)

For an SI program such as Africa RISING with diverse interventions expected to improve agricultural production and economic outcomes, the identification of relevant indicators can be difficult. Indeed, the literature notes that the degree of SI is best assessed through a multidimensional outcome based on enhancing agricultural productivity while preserving social, economic, and environmental sustainability whereby each domain encompasses multiple elements (Haggar et al. 2021). Following previous research, we adopt a Sustainable Intensification Assessment Framework (SIAF) approach developed by Musumba et al. (2017). While various indicators have previously been suggested to measure different aspects of SI (Smith et al. 2017; Xie et al. 2019), the SIAF is meant to provide guidance on indicators, metrics, and means to explore five domains of sustainability (agricultural production and productivity, economic, environment, human, and social). Its flexibility for choosing indicators and its ability to examine potential trade-offs among domains and indicators are among the factors that make the SIAF appealing for our study. This framework has been used to assess patterns of SI in Ethiopia (Hammond et al. 2021), evaluate implications of groundnut production for SI in Ghana (Rahman et al. 2020), examine the implications of mechanization for equity and sustainability in Tanzania (Fischer et al. 2018), and evaluate farmers' preferences for different aspects of SI in Ghana (Kotu et al. 2022).

Given the diversity of program interventions and possible behavioral changes in the adoption of innovations that may be complementary to those promoted by the program, we consider several indicators of adoption under the *environment* domain, including crop rotation, use of improved seeds (sorghum and groundnut), and use of manure and inorganic fertilizers, as well as access to agricultural advisory services as a catch-all indicator for access to improved sustainable practices (Piñeiro et al. 2020; Nakano, Tanaka and Otsuka 2018).

Under the agricultural *productivity* domain, we consider cereals and legumes yields (tons per hectare), the number of unique crops produced, tropical livestock units (TLU), agricultural labor used (person-days per hectare), value of harvest (CFA), and labor profitability measured by total value of harvest per unit of labor (CFA/person-days). Unit values computed by dividing total sales revenue by quantities sold are used to impute a monetary value to harvest, in line with Deaton (1988).

Under the *economic* domain, we examine annual net household income, which encompasses crop, livestock, wage, self-employment, and other income. We also consider per capita daily consumption expenditure and poverty defined based on a $1.9 per capita/day poverty line (based on 2015 purchasing power parity (PPP) conversion data from the World Bank). Total food consumption expenditure is the sum of self-reported expenditure on purchased food and imputed values of food consumed from own production and gifts. The latter two are expressed in monetary value using unit values computed by dividing total expenditure by total quantity on purchased foods. To capture non-monetary dimensions of poverty, we construct asset-based poverty indices based on ownership of durable agricultural and non-agricultural assets using factor analysis principal-component factor method, following Filmer and Pritchett (2001). Households n the top tercile of asset-based indices are defined as non-poor. Unlike consumption-based poverty, which is prone to transient changes in expenditure, asset-based poverty captures long-term and persistent structural poverty (Carter and Barrett 2006; Brandolini, Magri and Smeeding 2010). It has also been shown that the two do not always track well against each other (Foreit and Schreiner 2011).

Under the *human* domain, we consider two indicators based on food consumed inside the household over the last seven days. These are household dietary diversity score (HDDS) based on 12 food groups (Kennedy, Ballard and Dop 2010) and food security status based on whether household dietary caloric intake is above the minimum requirement (MDER) (FAO 2008). We complement these indicators based on self-reported months of food shortages during the post-harvest (June–September) and peak months of the lean season (January–March). While the first set of indicators captures the state of food security around the time of data collection, the latter captures food security over a longer period. To complement household-level food security indicators, we use anthropometry data to construct individual-level nutritional indicators for children 0–59 months old and women 15–49 years old. We follow WHO child growth standards (WHO 2006) to construct indicators of child undernutrition: chronic undernutrition or stunting, severe undernutrition or wasting, and underweight. Stunting captures linear growth retardation and cumulative growth deficit due to a chronic state of deficiency resulting from undernutrition and illnesses. It is largely an irreversible outcome capturing past and future predictions, and revealing the growth environment of individuals' potential to fulfill their development and economic potential (Leroy and Frongillo 2019). Wasting occurs when a child is too thin for their height and results from a recent and severe weight loss due to acute starvation and/or severe disease. Underweight is a manifestation and combination of both chronic and acute factors. For women, we use body mass index (BMI) as indicator for undernutrition (BMI < 18.5).

Under the *social* domain, for which relatively limited data are available, we examine the share of plots and livestock owned by women, either solely or jointly with men.

Table A2 in the Appendix reports descriptive statistics by survey round on our outcomes of interest. To control for spatial and time-varying weather differences, we use the Normalized Difference

Vegetation Index (NDVI) data from the Terra Moderate Resolution Imaging Spectroradiometer Vegetation Indices (MOD13Q1) Version 6 (Didan 2015).[4]

2.3 Identification strategy

One of the biggest challenges in evaluating participatory programs is sample (self-)selection. This is because these types of programs often employ targeting criteria based on expected effectiveness and returns by recruiting farmers who are (perceived to be) better endowed to adopt and disseminate technologies (Winters, Salazar and Maffioli 2010; Phillips, Waddington and White 2014). Indeed, our results show that households are systematically different depending on treatment participation and intensity (Table 1).

We consider three different groups: (1) Households in target villages that ever participated in the program: (2) Households in target villages that never participate in AR activities; and (3) Households in non-target villages.

[4] The NDVI has a temporal resolution of 16 days, a spatial resolution of 250 meters, and covers the time 2000 to 2021. It is a measure of greenness and vigor of vegetation with values ranging between 0 and 1 where NDVI <=0.1 represents bare rock, sand, or snow; 0.1 <NDVI<=0.5 captures sparse vegetation; and NDVI > 0.5 indicates dense, green vegetation. For both waves, we computed NDVI average value and coefficient for the completed cropping season prior to the data collection (Oct. to May).

Table 1. Descriptive summary of selected variables at baseline (2014)

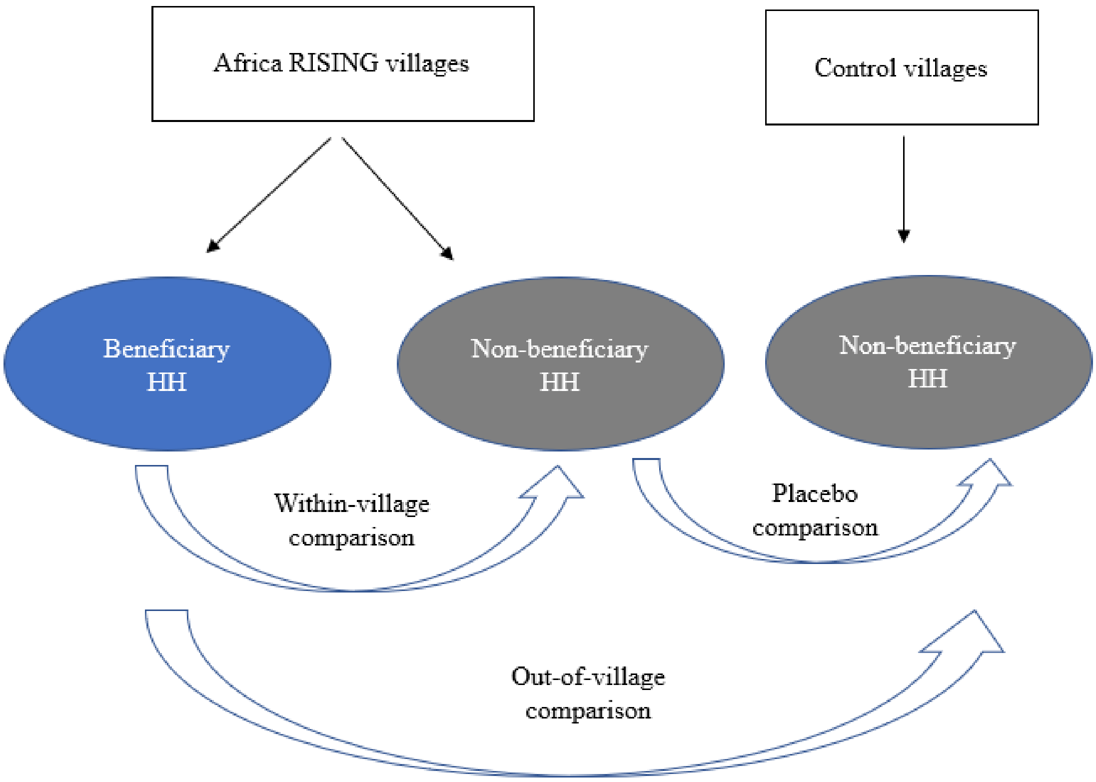
	(1) Group 1	(2) Group 2	(3) Group 3	(4) 1 vs 2	(5) 1 vs 3
Household demographic variables					
Household size	7.93	7.88	7.82	***	***
Age of the household head (years)	38.3	39.8	39.1		*
Female-headed (%)	1.45	0	2.74		
Household socioeconomic variables					
Years of education of the head	3.98	2.29	1.99		***
Area operated (ha)	10.2	8.72	8.81	***	***
Plots managed by women (%)	3.99	7.65	7.24	*	***
Livestock managed by women (%)	1.35	3.57	3.60		
Tropical livestock units	6.12	3.89	5.53	***	
Non-poor (agr. durable assets) (%)	68.6	58.9	56.8	**	***
Non-poor (non-agr. durable assets) (%)	31.4	33.9	47.3		***
Landscape level variables					
Mean monthly NDVI (historical)	0.43	0.46	0.47	**	***
CV of NDVI (historical)	13.9	13.9	13.7		
Observations	207	112	292	526	706

Note: Columns 1-3 report average values (of continuous or dummy variables) while columns 4 and 5 show statistical significance levels from pairwise tests of equality of means. NDVI is computed across the cropping season; CV stands for coefficient of variation. Plot and livestock management by females includes both sole and joint management with males.
* p<0.1, ** p<0.05, *** p<0.1

On average, households in group 1 are statistically larger, more educated, less likely to be female-headed, cultivate larger farms, raise more livestock, with fewer women managing—fully or partially—crop and livestock plots, are more likely to be non-poor based on durable agricultural assets, and historically face less variable weather compared to households in group 3. Some of these differences persist when comparing non-beneficiaries in group 1 with non-beneficiaries in group 2.

Exploiting the unique study design discussed in Section 2.2, we conduct multiple pairwise comparisons to check the sensitivity of impact estimates and assumed impact pathways (Figure 2). These include comparisons of beneficiaries vs. non-beneficiaries in program target villages (*within-village* comparison), beneficiaries vs. non-beneficiaries in control villages (*out-of-village* comparison), beneficiaries vs. non-beneficiaries in program target and control villages (*overall* comparison), and non-beneficiaries in program villages vs. non-beneficiaries in control villages (*placebo* comparison).

Figure 2. Alternative pairwise comparisons



By construction, within-village impact estimates are not prone to bias due to potential time-varying systematic village-level differences between program target and control villages that are correlated with the outcomes of interest. Nonetheless, these estimates may be prone to contamination bias if, for example, adoption behavior of households in program villages that have not directly participated in the program is affected by learning from beneficiaries who live in the same villages, i.e., contamination via spillover effects. Spillovers are often expected in agricultural programs that promote adoption among direct beneficiaries for subsequent dissemination of innovations either as part of a scaling exercise or more informally (e.g., through social networks).

While spillovers are desirable to magnify the impact of investments, they pose a challenge for impact evaluation by contaminating the counterfactual (Angelucci and Di Maro 2010; Winters et al. 2010). Spillovers can be especially important for a demand-driven and participatory program such as Africa RISING where residents in program target villages can be exposed to SI innovations by attending field demonstration days. For example, in a case where non-beneficiaries in program villages benefit from significant spillovers but do not report being participants during follow-up, within-village impact estimates would be underestimated. On the other hand, out-of-village impact estimates can be biased in the case of any systematic time varying village-level differences correlated with the outcomes of interest. Finally, impact estimates based on the *placebo* comparison can be prone to both spillovers and omitted time- and location-varying confounding factors. Absent the latter effect, *placebo* comparison will capture potential spillover effect.

Because participation in the program is voluntary and non-random, socio-demographic characteristics at baseline are not perfectly balanced between treatment and control group (as reported in Table 1). To address this imbalance between the two groups, our identification relies on a propensity score matching (PSM) and difference-in-differences (DiD) strategies (for general discussion, see Wooldridge 2002; Zeldow and Hatfield 2021).

PSM is used to estimate the causal impact of a program/policy when participation in the program is non-random. This method allows the identification of a valid control group by matching each treated unit with a non-treated unit with similar characteristics. More specifically, PSM expresses the probability that a unit will enroll into a program based on observed characteristics (the propensity score), which is then used to match treated and untreated units. This process balances differences in observable characteristics between the two groups and reduces bias, making it easier to compare outcomes between the two groups and draw more reliable conclusions about the impact of the treatment (Stuart et al., 2014).

In our analysis, we computed the propensity score based on the following characteristics at baseline: household size, age and gender of the household head, educational level, household's assets (agricultural and non-agricultural), size of the land, use of inputs for agricultural production, the total number of person-days needed for agricultural production, an index measuring the distance of the homestead from basic public services, and the historical monthly average and coefficient of variation of NDVI from 2000 to 2016 (Haile et al., 2016). More specifically, we estimated the following logistic regression:

$$P(Y_h \neq 0 \mid X_{hj}) = \frac{exp(X_{hj}\beta_j)}{[1 + exp(X_{hj}\beta_j)]}$$

with Y_h being the dummy variable equals to 1 if the household is in the treatment group; and X_{hj} the covariates at the household and environmental level at baseline.

Figure 3: Distribution of propensity score before and after matching for four comparisons

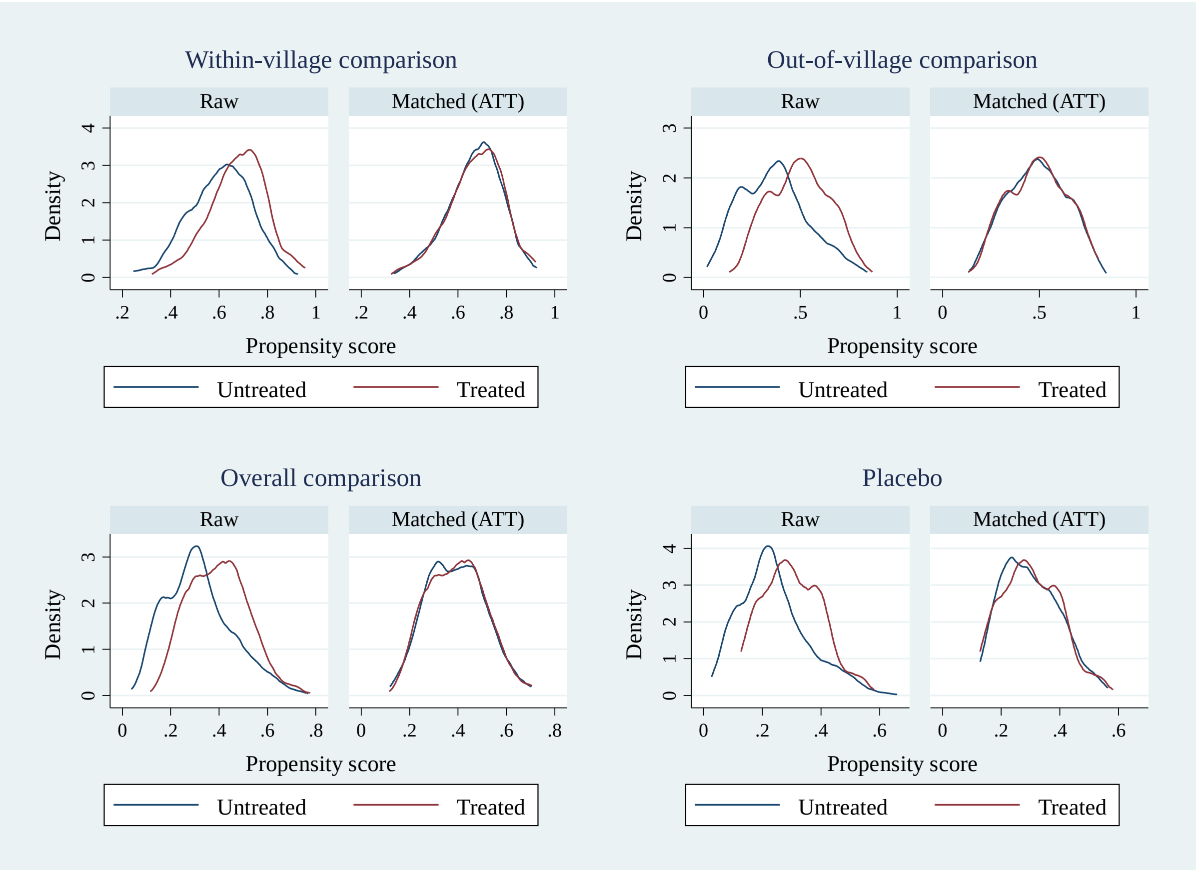


Figure 3 shows the distribution of the propensity score between treatment and control groups before and after the matching for all the pairwise comparisons conducted. After matching the distribution of the propensity score for the untreated group overlaps with the one of the treated groups for all the four comparisons, suggesting that the matching was successful in balancing the covariates between treatment and control groups.

Then, we include the estimated propensity score in the main DiD equation as a weight to obtain more balanced estimates (Stuart et al. 2014). The DiD strategy can be formalized as shown in Equation 2.

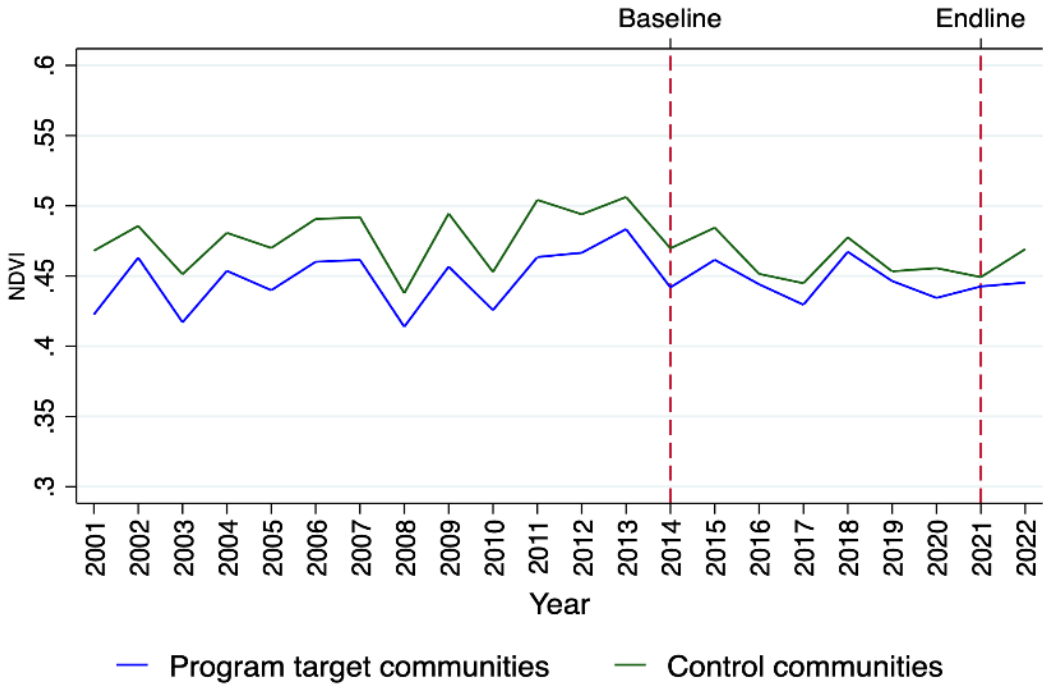
$$Y_{ht} = \alpha + \beta T_{ht} + \gamma P_{ht} + \delta(T_{ht} * P_{ht}) + \theta NDVI_{ht} + \varepsilon_{ht} \quad 2$$

where *h* and *t* are indices for household and time, Y is an outcome indicator, T is indicator that takes value one for program beneficiaries (treated group) and zero for non-beneficiaries (control group), P is an indicator that takes value of one for follow-up and zero for baseline and captures aggregate factors affecting Y the same way between the two groups, $NDVI_{ht}$ the NDVI average and coefficient or variation in the 12 months prior to the interview included as control variable in the regression adjustment (RA) specification of the DiD, and ε is model error term assumed to be i.i.d.. Ordinary least squares (OLS) estimate of δ measures the average effect of treatment on the treated (ATT) on Y. Absent additional covariates, $\hat{\delta}$ is given by $(\bar{Y}_1^B - \bar{Y}_0^B) - (\bar{Y}_1^C - \bar{Y}_0^C)$, where superscript B and C represent beneficiary and control group, respectively; subscripts 1 and 0 represent post- and pre-treatment periods, respectively, and $\bar{Y}$ represents sample average. For ease of interpretation, all dummy outcome indicators are multiplied by 100 so that $\hat{\delta}$ captures percentage changes in Y.

DiD relies on one key assumption—outcomes of the two groups would have evolved similarly in the absence of the treatment—known as the parallel trend assumption. This strategy produces unbiased estimates if the treatment is not systematically related to other confounding factors that affect Y. Unlike cross-section-based impact estimates, DiD estimates will be unbiased even when baseline covariates are unbalanced between the two groups as long as temporal trends in outcomes do not vary by treatment status (Zeldow and Hatfield 2021).

One approach to test the validity of the identifying assumption for DiD is to compare trends in outcome indicators over time. In our case, conducting such a test is not possible since only two data points are available for the outcome indicators and no time series data are available on the various other indicators. On the other hand, analysis of temporal trends of indicators such as NDVI can serve as an indirect test of the identifying assumption given the reliance of the target population on rainfed agriculture. NDVI trends were generally comparable for treated and control villages over time, with cropping season NDVI being marginally higher for non-program villages (Figure 4).

Figure 4. Average NDVI during the cropping season by community treatment status



As robustness checks, we conduct three additional analyses. First, we re-estimate a version of the model in Equation 1 controlling for NDVI (historical average as well as average and coefficient of variation for the cropping seasons for which agricultural production data were collected).[5] One strategy for dealing with bias in DiD estimates due to systematic time varying confounding factors such as NDVI is to control for their time varying effects on Y through RA. In a DiD with RA model, OLS estimate of $\hat{\delta}$ no longer captures the double mean differences as shown for the simple DiD model but still measures the effect of program participation. Controlling for covariates will change $\hat{\delta}$ if one or more covariates are correlated with the interaction term between survey round and treatment status conditional on the main effects of survey round and treatment (Angrist and Krueger 1999).

Second, both the simple DiD and DiD with RA models are estimated using different counterfactuals as shown in Figure 2. By construction, $\hat{\delta}$ from the *overall* comparison will be an average of the impact estimate between the within-village and the out-of-village comparisons. Third, Equation 1 is estimated based on a *placebo* comparison where T equals one for non-beneficiaries in program villages and zero for non-beneficiaries in control villages. Given that both groups were randomly sampled and did not participate in the program, $\hat{\delta}$ should not be significant unless in the case of systematic time varying village-level omitted confounding factors correlated with Y, spillovers, or both.

[5] We do not run a specification including socio-demographic characteristics, as we included these dimensions in the equation to compute the propensity score for household matching.

3. Results and Discussion

In this section, we present the results for all the SI domains (environmental, productive, economic, social, and human) for the four comparisons listed in the previous section[6]. We find that participation in the program led to statistically significant impacts when considering all the four comparisons, with more robust effects when comparing all households in target villages with households in control villages (i.e., *overall comparison*), and when comparing non-beneficiary households in target villages with households in control villages (i.e., *placebo test*).

These findings are most likely affected by the fact that households in target villages that reported not being enrolled in the program may have benefited, since they know the program by a different name. When we include these households in the analysis, the increased sample size and statistical power increase the significance of the estimates, the magnitude of which falls between the magnitudes of the within-village and out-of-village estimates. The *placebo* test results also suggest that there was a misreporting of AR participation: because non-beneficiary households in target villages have been exposed to and engaged in AR interventions, the program's spillover effects appear to be particularly pronounced.

Table 2 shows estimates of indicators under the environmental domain. In the within-village comparison (Table 2, columns 1 and 2), being a beneficiary household is associated to a 10 percentage points higher likelihood to utilize fertilizers, as well as an increase in seeds purchased per hectare by 1,400 FCFA, compared to non- beneficiaries. In the out-of-village comparison (Table 2 columns 3 and 4), the Africa RISING program has an impact on the real value of improved seeds purchased per hectare by 600 FCFA and the likelihood of using improved seeds for groundnut by 12 percentage points. In all comparison groups, the program increased the use of improved seeds for sorghum between 11 and 18% percentage points depending on the comparison.

In the overall comparison (columns 5 and 6), results show that households in target villages are 20 percentage points more likely to have access to agricultural advisories, and we observe a positive and significant impact of the program on the adoption of improved crop varieties. Households in target villages have increased the purchase of improved seeds per hectare (+500 FCFA) compared to

[6] In Tables 2 through 5, columns (1) and (2) report DiD impact estimates based on within-village comparison. Columns (3) and (4) report impact estimates based on out-of-village comparison. Columns (5) and (6) report impact estimates where all non-beneficiaries inside and outside program villages are used as control groups. Columns (7) and (8) report results from the placebo test. Results in columns (1), (3), (5) and (7) are from simple DiD, while those in columns (2), (4), (6), and (8) are from DiD with regression adjustment where we control for NDVI average and coefficient of variation in the 12 months prior to the interview. Comments refer to parameters of the DiD with regression adjustment model.

households in non-target villages, and are more likely to use improved seeds for the cultivation of sorghum and groundnuts, two crops that were purposively targeted by the program (Table 2, columns 5 and 6). The *placebo* test results (Table 2, columns 7 and 8) confirm the results from the overall comparison.

Table 2. Difference-in-differences (DiD) on environmental indicators

	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inorganic fertilizers used (kg/ha)	1.1	1.7	8.3	7.1	8.2	8.2	12.2	13.7
Real value of pesticides used per ha ('000 CFA)	-0.8***	-0.8*	-1.8	-1.5	-1.5	-1.3	-1.7	-1.5
Real Value of fertilizers per ha ('000 CFA)	1.4	1.5	-0.4	-0.4	0.1	0.1	3.4	3.7
Real Value of purchased seeds per ha ('000 CFA)	1.4**	1.4*	0.9	1	1.1*	1.2**	1.3*	1.5**
Real Value of purchased improved seeds per ha ('000 CFA)	0.3	0.3	0.6**	0.6**	0.5**	0.5**	0.5**	0.5**
Accessed extension services (%)	18.7*	18.4*	19.8*	18.1*	20.5*	20.1**	19.1*	18.6*
Practiced intercropping (%)	-7.1	-7.2	-2.3	-2.7	-4.5	-4.8	-11.1*	-11.4*
Household uses fertilizer (%)	10.1*	10.3*	1.6	1.4	4.5	4.4	5.8	5.9
HH uses improved seeds: Sorghum	10.6**	10.6**	18.3*	18.2*	16.3*	16.1*	14.4**	14.3*
HH uses improved seeds: Groundnut	-2.1	-2.1	11.8*	11.9*	7.5**	7.3*	7.2	7

Note: Columns 1 and 2 report DiD impact estimates based on within-village comparison. Columns 3 and 4 report impact estimates based on out-of-village comparison. Columns 5 and 6 report impact estimates where all non-beneficiaries inside and outside program villages are used as counterfactual. Columns 7 and 8 report results from the placebo test. Results in columns (1), (3), (5) and (7) are from simple DiD, while those in columns (2), (4), (6) and (8) are from DiD with regression adjustment where we control for NDVI average and coefficient of variation in the 12 months prior to the interview. 0.01 - ***; .05 - **; .1 - *

Impact on agricultural productivity is reported in Table 3. Results from within-village and out-of-village comparisons show limited impact of participation in AR on most of the considered agriculture-related outputs (columns 1 through 4 in Table 3). Beneficiary households in the within-village comparison (Table 3 columns 1 and 2) experience a decrease in maize yield by 150Kg per hectare, while increasing okra yield by about a ton per hectare, compared to control households. Looking at the out-of-village comparison (Table 3 columns 3 and 4), by contrast, the program has experienced an increase in cotton yield (+310Kg per hectare) and green bean yield (+170 Kg per hectare), and a decrease in agricultural labor by 29 days per hectare. In the overall comparison (columns 5 and 6), program participation is associated with higher green bean yields, cotton yields, and okra yields. Beneficiary households grow also less cotton and millet crops, significant at the 10 percent level, while there is no positive impact on growing other crops (Table A3 in Appendix). Despite the positive impact of the program on the adoption of improved seeds for groundnut and sorghum, we do not find strong evidence of an increase in yields for the two crops, but we surprisingly observe a significant decrease in the productivity of sorghum. According to our local partners, the implementing agency—Fenabe—extensively encouraged the use of biological and organic agricultural inputs, particularly in Bougouni, which may have reduced traditional yields. Also, we find a decrease in agricultural labor, proxied by the number of person-days per hectare, suggesting that beneficiary households became more productive. The *placebo* comparison (columns 7 and 8) confirms results from the overall comparison, showing an increase in green bean and okra yields, and a decrease in sorghum yield.

Table 3. Difference-in-differences (DiD) on productivity indicators

	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Yield (ton/ha)								
Maize	-0.14	-0.15***	-0.02	-0.07	-0.07	-0.11	0.09	0.04
Sorghum	-0.13	-0.12	-0.15*	-0.17**	-0.14**	-0.15**	-0.13**	-0.14*
Groundnut	0.09	0.09	-0.34	-0.24	-0.17	-0.12	-0.36	-0.3
Green Bean	0.16	0.16	0.15*	0.17**	0.15**	0.15**	0.17***	0.18***
Cotton	0.13	0.13	0.27**	0.31**	0.24*	0.26*	0.19	0.21
Okra	0.93***	1.15***	0.22	0.17	0.54**	0.65**	0.49***	0.64**
Number of crop items produced	0.08	0.1	-0.19	-0.28	-0.08	-0.13	0.07	0.03
Tropical livestock units	0.56	0.56	-0.74	-0.73	-0.3	-0.29	0	-0.05
Agricultural labor (person-days/ha)	13.58	14.69	-24.73**	-28.9**	-12.54*	-14.93*	-6.77	-10.3
Labor profitability (CFA/person-day)	-0.83	-0.81	1.23	1.12	0.68	0.62	0.71	0.67

Note: Columns 1 and 2 report DiD impact estimates based on within-village comparison. Columns 3 and 4 report impact estimates based on out-of-village comparison. Columns 5 and 6 report impact estimates where all non-beneficiaries inside and outside program villages are used as counterfactual. Columns 7 and 8 report results from the placebo test. Results in columns (1), (3), (5) and (7) are from simple DiD, while those in columns (2), (4), (6) and (8) are from DiD with regression adjustment where we control for NDVI average and coefficient of variation in the 12 months prior to the interview. 0.01 - ***; .05 - **; .1 - *

Table 4 reports the results across the economic indicators. As is the case with the productivity outcomes, we observe only few statistically significant impacts in the within-village and out-of-village comparisons. In the overall comparison (columns 5 and 6), in line with the general lack of impact of the program on indicators of agricultural production and productivity, results show little evidence of the effect of the program on indicators of economic mobility, especially income. However, we find a positive impact on total consumption expenditure (+$1.3 per capita/day in PPP), with the latter result confirmed also when looking at the within- and out-of-village comparisons (columns 1 through 4). In the out-of-village and overall comparisons beneficiary households are also more likely to be non-poor based on non-agricultural durable assets (19 and 11 percentage points, respectively). Finally, AR initiatives had an impact on the prevalence of diversified livelihoods (+13 percentage points) for households in target villages compared to non-beneficiaries in non-target villages, with an even higher impact (+21 percentage points) among beneficiary *versus* non-beneficiary households in control villages (columns 3 and 4). These results hold also when we look at the placebo comparison, with a similar magnitude of the effects.

Table 5 summarizes the impacts on human and social indicators. We observe no statistically significant impact of participation in the program on food security and household dietary diversity scores. Despite nutritional improvement being one of the program objectives, we surprisingly do not find improvements in either child or maternal nutrition along the various nutritional indicators examined.

Table 4. Difference in difference (DiD) on economic indicators

	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income/Earnings ($ PPP)								
Annual net household income (PM) ($ PPP)	1218.9	1232.2	64.7	-31.6	547.7	538.6	-378.7	-380.5
Net crop income (PM) ($ PPP)	1080.1	1100.5	-543.5	-670.9	99.5	69.2	-818.8	-843.7
Net livestock income (PM) ($ PPP)	133.6	143.1	202.7	190.8	194.3	181.8	106.7	97.2
Other income ($ PPP)	1.5	1.2	19.6	20.5	12.7	12.6	15.8*	15*
Poverty								
Per capita daily expenditure ($ PPP)	1.1**	1.1**	1.3**	1.3**	1.3***	1.3***	-0.5	-0.5
Non-poor (>$1.9 PPP) (%)	-5.6	-6	5.8	5.7	1.3	1.3	-8.6	-9.1
Non-poor (agr. durable assets) (%)	-1.6	-1.3	5	5.6	2.4	2.5	4.4	4.5
Non-poor (non-agr. durable assets) (%)	-5.9	-6.4	17.4*	19.3**	9.6**	10.5**	13.8**	14.7***
Diversified livelihood								
Households with diversified livelihoods (%)	-3.8	-4	19.4***	21.4***	11.9**	12.7**	17.2**	17.9**

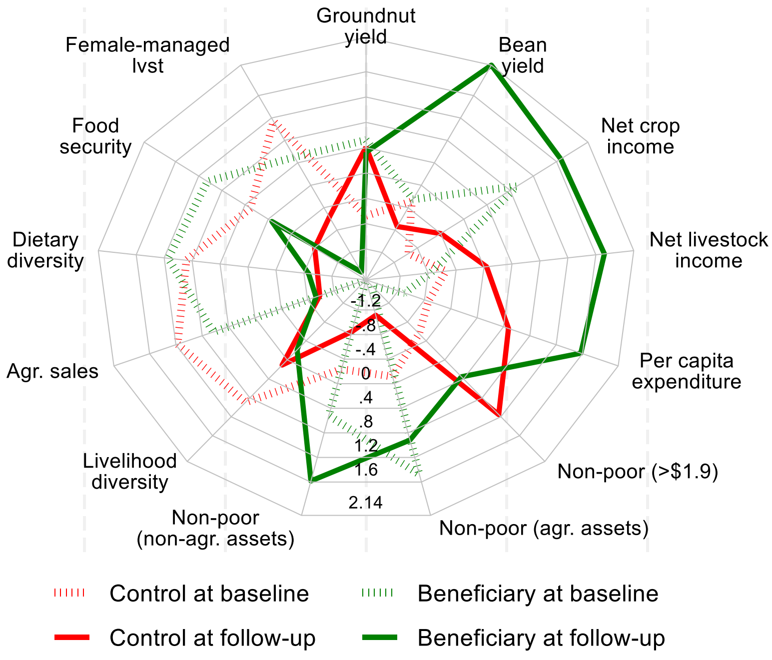
Note: Columns 1 and 2 report DiD impact estimates based on within-village comparison. Columns 3 and 4 report impact estimates based on out-of-village comparison. Columns 5 and 6 report impact estimates where all non-beneficiaries inside and outside program villages are used as counterfactual. Columns 7 and 8 report results from the placebo test. Results in columns (1), (3), (5) and (7) are from simple DiD, while those in columns (2), (4), (6) and (8) are from DiD with regression adjustment where we control for NDVI average and coefficient of variation in the 12 months prior to the interview. 0.01 - ***; .05 - **; .1 - *

Table 5. Difference-in-differences (DiD) on human indicators

	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Human								
Household								
Household dietary diversity score (out of 12)	-0.02	-0.04	-0.13	-0.11	-0.1	-0.08	-0.3	-0.28
Incidence of food secure households (%)	8.64	8.71	-3.36	-5.24	-0.16	-0.89	-7.48	-8.24
Months of food shortage (last year)	-0.39	-0.4	-0.55	-0.55	-0.5	-0.48	-0.75**	-0.74**
Child								
HAZ	0.13	0.12	0.43**	0.49*	0.29	0.33	0.38	0.42
WHZ	0.13	0.13	-0.38*	-0.35*	-0.19	-0.17	-0.09	-0.08
WAZ	0.33	0.33	0.29	0.4	0.28	0.34	0.21	0.27
Prevalence of wasting (%)	-0.92	-1.26	2.56	2.03	1.37	1.33	-1.06	-0.93
Prevalence of stunting (%)	-8.9	-9.02	-14.33*	-15.26*	-11.95	-12.21	-16.08	-16.22
Prevalence of underweight (%)	-1.68	-1.56	-4.33	-6.16	-2.48	-3.11	-1.71	-2.24
Women								
BMI	-0.39	-0.41	0.96	1.04	0.46	0.45	0.13	0.15
Panel B. Social								
Plots under female responsibility (%)	1.2	1	-0.1	-0.1	0.1	0.3	2.6	2.7
Livestock under female responsibility (%)	2.1	1.9	-2	-2.2	-0.7	-0.6	-0.3	-0.3

Note: Columns 1 and 2 report DiD impact estimates based on within-village comparison. Columns 3 and 4 report impact estimates based on out-of-village comparison. Columns 5 and 6 report impact estimates where all non-beneficiaries inside and outside program villages are used as counterfactual. Columns 7 and 8 report results from the placebo test. Results in columns (1), (3), (5) and (7) are from simple DiD, while those in columns (2), (4), (6) and (8) are from DiD with regression adjustment where we control for NDVI average and coefficient of variation in the 12 months prior to the interview. 0.01 - ***; .05 - **; .1 - *

Figure 5. Predicted impact values of the indicators from the overall comparison



The radar graph in Figure 5 represents the DiD estimates across a selection of relevant variables referred to the five dimensions of the SIAF indicators from the overall comparison. The graph allows to compare the characteristics associated to beneficiaries and non-beneficiaries before (dashed green and red lines, respectively) and after (solid green and red lines, respectively) program participation, as well as contrast the outcomes of interest between treatment and control groups based on standardized values of each variable.

Bean yield, net crop income, net livestock income, per capita expenditure, and the probability of being non-poor based on non-agricultural assets show the greatest variation between treated and control households, with outcomes among beneficiary households being significantly higher than among non-beneficiary households. Food security, livelihood diversification, groundnut production, and the percentage of plots under women's responsibility show smaller differences. We find no variation in the probability of being non-poor based on the poverty line ($1.9 per day) criterion or based on agricultural assets.

Beneficiary households before and after AR interventions (dashed and solid green lines, respectively) show significant differences in outcomes at baseline and follow-up: they were more food secure, had a higher dietary diversity score, and a higher value of agricultural sales at baseline than at follow-up. On the other hand, between baseline and follow-up they have experienced increases in bean yield, net crop and livestock revenue, per capita expenditure, and the probability of not being poor based on non-agricultural assets.

Even non-beneficiary households before and after AR interventions (dashed and solid red lines, respectively) show significant differences in outcomes between baseline and follow-up: as with beneficiary households, they were more food secure, had a higher dietary diversity score, greater livelihood diversity, agricultural sales, and percentage of plots owned by female at baseline than at follow-up. They have experienced an increase in net crop and livestock income, per-capita expenditure, and higher probability of being non-poor -based on the poverty line- between baseline and follow-up, while there are no significant differences in bean yields and the probability of being non-poor, based on non-agricultural assets, over time.

In all three comparisons, both beneficiaries and non-beneficiary households show a higher likelihood of being food insecure, as shown in Figure A1 in the Appendix, as well as a lower dietary diversity score and agricultural sales values, factors likely due to the long-term effects of COVID-19.

4. Conclusions and implications

Sustainable intensification (SI) of the smallholder sector has gained increased recognition as an effective strategy for improving food and nutrition security as well as ecosystem services in the face of projected population growth and climatic changes. The focus of SI is on improving resource-use efficiency while achieving higher production from the same resources and enhancing beneficial environmental and social services.

We evaluated the impact of an SI program in southern Mali that was implemented between 2012 and 2021. The main goal of the program was to provide smallholders with pathways out of hunger and poverty through SI. Toward that end, the program has validated and promoted several SI innovations, including improved sorghum and groundnut varieties, cereal-legume intercropping and rotation, and legume-legume intercropping, as well as improved livestock feeding and animal housing. Given the diversity of interventions and expected effects on several dimensions of sustainability, we use the Sustainable Intensification Assessment Framework (SIAF) to guide the selection of indicators covering environmental, productivity, economic, human, and social domains. To account for systematic targeting of geographic areas, self-selection of participants, and possible contamination of the counterfactual group through spillovers we took advantage of the unique evaluation design using panel data collected in 2014 and 2022 and applying propensity score matching (PSM) and a difference-in-differences (DiD) strategies to estimate impacts based on alternative counterfactual groups. In addition, we conduct a *placebo* test to shed light on potential spillover effects to confirm the causality mechanisms at play.

We document positive results on use of improved seeds, access to extension services, green bean, cotton and okra yield, per capita household consumption expenditure, likelihood of being non-poor based on durable assets, and likelihood of having diversified livelihoods. However, we do not find improvements in women and child nutrition despite the program's focus on nutritional status especially of women and children, which remains a persistent public health and development challenge in Mali. To cope with this challenge, integration of behavioral change communication and water, sanitation, and hygiene projects (WASH) into agricultural programs can help maximize nutritional gains of agriculture-focused programs.

The positive impacts of the program are lower than expected, and various interpretations can be put forward to explain this finding. As discussed earlier, the challenge of defining a household in Mali made it difficult to identify nuclear households—as opposed to families living in the same compound—in the treatment and control groups, resulting in data that are less homogeneous than those collected in other countries where Africa RISING was implemented. Misreporting on program participation might also have biased estimated parameters. To mitigate this potential misclassification bias, we attempted to cross-reference self-reported program participation with program monitoring data but then decided to rely on the former due to inconsistencies between the two measurements and the inability to verify the accuracy of monitoring data.

Another possible, non-testable explanation of these results is related to factors outside program implementation. The COVID-19 outbreak disrupted the timing of planned follow-up surveys in Mali, resulting in misalignment of the reference periods between baseline and follow-up data collection. This is especially problematic when conducting a before and after comparison of indicators based on 7-day recall food consumption data—e.g., dietary diversity and caloric-intake based food security—given the evidence on the strong association between agricultural season, food security, and nutritional status among smallholder farmers (Aweke et al. 2022; Bonuedi, Kornher and Gerber 2022). In addition, to the extent the outbreak might have differentially affected program participation and benefits (also due to differences in livelihoods), our impact estimates might be inaccurate. For instance, our results seem to suggest a worsening of severe food insecurity among AR beneficiaries following the COVID-19 outbreak.

Disentangling the contribution of different innovations to the overall impact of the program is necessary to inform targeted interventions. This objective would have required an adequate number of households to adopt the various SI innovations being offered and routine and frequent monitoring data collection on the type and timing of innovations tested and applied. However, given the diversity of program interventions rolled out over time, it was not possible to obtain accurate measures of impact of single, discrete interventions, but only aggregate impact estimates of the overall program. Despite the limitations of our study design, i.e., not designed to assess the impact of adoption of a specific (or package of) SI innovation(s), but rather to estimate the overall impact of participation in Africa RISING program, we learned valuable lessons to improve future evaluations of SI programs.

Given the diversity of SI innovations promoted and the fact that a specific mix of SI innovations were tested as part of the agronomic trials conducted, our impact estimates based on program participation are likely prone to dilution bias, providing a conservative, lower-bound estimate of the real effect. Also, the relatively small sample size did not allow us to examine potential heterogenous impacts by various dimensions of policy relevance, including intensity of program participation and location, with the latter especially important given the diversity in agroecologies covered within the country.

References

AfDB (African Development Bank). 2020. "Mali: 2015-2019 Country Strategy Paper Completion Report Combined with the 2019 Country Performance Review."

Angelucci, M., and V. Di Maro. 2010. "Program Evaluation and Spillover Effects: Impact-Evaluation Guidelines." IZA Discussion Paper No. 9033, IZA, Bonn, Germany.

Angrist, J.D., and A.B. Krueger. 1999. "Emprical Stratagies in Labor Economics." In *Handbook of Labor Economics Vol. 3C*, eds. O. Ashenfelter and D. Card. 1277–1357. North-Holland.

Awotide, B.A., A.A. Karimov, and A. Diagne. 2016. "Agricultural Technology Adoption, Commercialization and Smallholder Rice Farmers' Welfare in Rural Nigeria." *Agricultural and Food Economics* 4 (1).

Azzarri, C., B. Haile, P. Howard, C. Roberts, I.F. Africa, S. Enkhbayar, and S. Signorelli. 2016. "A User Guide to Mali Africa Research in Sustainable Intensification for the Next Generation (Africa RISING) Baseline Evaluation Survey Data." IFPRI.

Bezner Kerr, R., P.R. Berti, and L. Shumba. 2011. "Effects of a Participatory Agriculture and Nutrition Education Project on Child Growth in Northern Malawi." *Public Health Nutrition* 14 (8): 1466–1472.

Birhanu, F.Z., A.S. Tsehay, and D.A. Bimerew. 2021. "The Effects of Commercialization of Cereal Crops on Multidimensional Poverty and Vulnerability to Multidimensional Poverty among Farm Households in Ethiopia." *Development Studies Research* 8 (1): 378–395.

Bouis, H.E., and R.M. Welch. 2010. "Biofortification: A Sustainable Agricultural Strategy for Reducing Micronutrient Malnutrition in the Global South." *Crop Science* 50 (Supplement 1): S-20-S-32.

Brandolini, A., S. Magri, and T. Smeeding. 2010. "Asset-Based Measurement of Poverty." *Journal of Policy Analysis and Management* 29 (2): 267–284.

Carter, M.R., and C.B. Barrett. 2006. "The Economics of Poverty Traps and Persistent Poverty: An Asset-Based Approach." *Journal of Development Studies* 42 (2): 178–199.

Conway, G., K. Wilson, and E. Wilson. 2013. "Sustainable Intensification: a New Paradigm for African Agriculture: A 2013 Montpellier Panel Report." London: Agriculture for Impact, Imperial College.

D'Souza, G., D. Cyphers, and T. Phipps. 1993. "Factors Affecting the Adoption of Sustainable Agricultural Practices." *Agricultural and Resource Economics Review* 22 (2): 159–165.

Deaton, A. 1988. "Quality, Quantity, and Spatial Variation of Price." *The American Economic Review* 78 (3): 418–430.

Didan, K. 2015. "MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]."

FAO. 2022. “Crop and livestock products.” Available at: https://www.fao.org/faostat/en/#data/QCL.

FAO. 2008. “FAO Methodology for the Measurement of Food Deprivation.” *FAO Methodology for the Measurement of Food Deprivation* (Oct.): 2–16.

Filmer, D., and L.H. Pritchett. 2001. “Estimating Wealth Effects without Expenditure Data - or Tears: An Application to Educational Enrollments in States of India.” *Demography* 38 (1): 115–132. Available at: https://doi.org/10.2307/3088292.

Fischer, G., S. Wittich, G. Malima, G. Sikumba, B. Lukuyu, D. Ngunga, and J. Rugalabam. 2018. “Gender and Mechanization: Exploring the Sustainability of Mechanized Forage Chopping in Tanzania.” *Journal of Rural Studies* 64 (September):112–122.

Foreit, K., and M. Schreiner. 2011. “Comparing Alternative Measures of Poverty: Assets-Based Wealth Index vs. Expenditures-Based Poverty Score.” *Working Paper Series: MEASURE Evaluation PRH*:1–21.

Garnett, T., M.C. Appleby, A. Balmford, I.J. Bateman, T.G. Benton, P. Bloomer, B. Burlingame, M. Dawkins, L. Dolan, D. Fraser, M. Herrero, I. Hoffmann, P. Smith, P.K. Thornton, C. Toulmin, S.J. Vermeulen, and H.C.J. Godfray. 2013. “Sustainable Intensification in Agriculture: Premises and Policies.” *Science* 341 (July): 33–34.

Haggar, J., V. Nelson, R. Lamboll, and J. Rodenburg. 2021. “Understanding and Informing Decisions on Sustainable Agricultural Intensification in Sub-Saharan Africa.” *International Journal of Agricultural Sustainability* 19 (5–6): 349–358.

Haile, B., C. Azzarri, C. Roberts, and D.J. Speilman. 2016. “Targeting, Bias, and Expected Impact of Complex Innovations on Developing-Country Agriculture: Evidence from Malawi.” *Agricultural Economics* 48 (2016): 1–10.

Hammond, J., M. van Wijk, N. Teufel, K. Mekonnen, and P. Thorne. 2021. “Assessing Smallholder Sustainable Intensification in the Ethiopian Highlands.” Elsevier Ltd.

HarvestChoice. 2013. “Africa RISING USAID’s Sustainable Intensification Program in Africa East and Southern Africa: Malawi Site Selection Report.” IFPRI.

Holden, S., and R. Lunduka. 2012. “Do Fertilizer Subsidies Crowd Out Organic Manures? The Case of Malawi.” *Agricultural Economics* 43 (3): 303–314.

ICF. 2022. “The DHS Program STATcompiler.”

IFPRI, IITA, and ILRI. 2016. “Africa Research in Sustainable Intensification for the Next Generation Program Pproposal for a Second Phase, 2016–2021.”

IITA, IFPRI, and ILRI. 2012. “Africa Research in Sustainable Intensification for the Next Generation (Africa RISING): Program Framework 2012–2016.”

Jayne, T.S., D. Marther, and E. Mghenyi. 2005. “Smallholder Farming in Difficult Circumstances: Policy Issues for Africa in IFPRI.” IFPRI.

Kennedy, G., T. Ballard, and M. Dop. 2010. “Guidelines for Measuring Household and Individual

Dietary Diversity." Rome: Nutrition and Consumer Protection Division, Food and Agriculture Organizaiton of the United Nations.

Koppmair, S., M. Kassie, and M. Qaim. 2017. "Farm Production, Market access and Dietary Diversity in Malawi." *Public Health Nutrition* 20 (2): 325–335.

Kotu, B.H., O. Oyinbo, I. Hoeschle-Zeledon, A.R. Nurudeen, F. Kizito, and B. Boyubie. 2022. "Smallholder Farmers' Preferences for Sustainable Intensification Attributes in Maize Production: Evidence from Ghana." *World Development* 152: 105789.

Lee, D.R. 2005. "Agricultural Sustainability and Technology Adoption: Issues and Policies for Developing Countries." *American Journal of Agricultural Economics* 87 (5): 1325–1334.

Leroy, J.L., and E.A. Frongillo. 2019. "Perspective: What Does Stunting Really Mean? A Critical Review of the Evidence." *Advances in Nutrition* 10 (2): 196–204.

Mungai, L.M., S. Snapp, J.P. Messina, R. Chikowo, A. Smith, E. Anders, R.B. Richardson, and G. Li. 2016. "Smallholder Farms and the Potential for Sustainable Intensification." *Frontiers in Plant Science* 7 (Nov.).

Musumba, M., C. Palm, P. Grabowski, and S.S. Snapp. 2017. "A Framework for Selecting and Analyzing Indicators of Sustainable Intensification." Google Scholar.

Nagarajan, C., Binder, L., Destrijcker, L., Michelini, S., Rüttinger, L., Sangaré, B., Zaatour, J. V. (2022). "Climate, Peace and Security Assessment: Mali." Adelhpi.

Nakano, Y., Y. Tanaka, and K. Otsuka. 2018. "Impact of Training on the Intensification of Rice Farming: Evidence from Rainfed Areas in Tanzania." *Agricultural Economics (United Kingdom)* 49 (2): 193–202.

Nalivata, P.C., M.W. Munthali, W.I. Makumba, E. Mbewe, and H. Manase. 2017. "Optimizing Fertilizer use within the Context of Integrated Soil Fertility Management in Malawi." *Fertilizer Use Optimization in sub-Saharan Africa*: 113–124.

Phillips, D., H. Waddington, and H. White. 2014. "Better Targeting of Farmers as a Channel for Poverty Reduction: A Systematic Review of Farmer Field Schools Targeting." *Development Studies Research* 1 (1): 113–136.

Piñeiro, V., J. Arias, J. Dürr, P. Elverdin, A.M. Ibáñez, A. Kinengyere, C.M. Opazo, N. Owoo, J.R. Page, S.D. Prager, and M. Torero. 2020. "A Scoping Review on Incentives for Adoption of Sustainable Agricultural Practices and their Outcomes." *Nature Sustainability* 3 (10): 809–820.

Pretty, J., C. Toulmin, and S. Williams. 2011. "Sustainable Intensification in African Agriculture." *International Journal of Agricultural Sustainability* 9 (1): 5–24.

Rahman, N.A., A. Larbi, B. Kotu, F. Kizito, and I. Hoeschle-Zeledon. 2020. "Evaluating Sustainable Intensification of Groundnut Production in Northern Ghana Using the Sustainable Intensification Assessment Framework Approach." *Sustainability (Switzerland)* 12 (15).

Schoch, M., and C. Lakner. 2020. "Global Poverty Reduction is Slowing, Regional Trends Help Understanding Why." World Bank Data Blog, Nov. 5.

Smith, A., S. Snapp, R. Chikowo, P. Thorne, M. Bekunda, and J. Glover. 2017. "Measuring Sustainable Intensification in Smallholder Agroecosystems: A Review." *Global Food Security* 12 (Nov. 2016): 127–138.

Snapp, S. 2002. "Quantifying Farmer Evaluation of Technologies: The Mother and Baby Trial Design." n *Quantitative Analysis of Data from Participatory Methods in Plant Breeding*, eds. M.R. Bellon and J. Reeves. Mexico, DF: CIMMYT. https://cgspace.cgiar.org/handle/10568/76948

UNICEF, WHO, and World Bank Group. 2020. "Levels and Trends in Child Malnutrition: Key Findings of the 2020 Edition of the Joint Child Malnutrition Estimates."

USAID. 2021. "Malawi: Nutrition Profile."

WHO. 2006. "WHO Child Growth Standards Length/ Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age Methods and Development."

Winters, P., L. Salazar, and A. Maffioli. 2010. "Designing Impact Evaluations for Agricultural Projects: Impact-Evaluation Guidelines." *SSRN.*

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

World Bank. 2020. "Consumer price index (2010 = 100) - Mali."

World Bank. 2011. "Climate Risk and Adaptation Country Profile Mali."https://climateknowledgeportal.worldbank.org/country/mali.

World Bank. 2023. "Mali Economic Update – April 2023 Special Chapter: Strengthening Financial Resilience of Pastoralists to Drought."

Xie, H., Y. Huang, Q. Chen, Y. Zhang, and Q. Wu. 2019. "Prospects for Agricultural Sustainable Intensification: A Review of Research." *Land* 8(11): 1–27.

Zeldow, B., and L.A. Hatfield. 2021. "Confounding and Regression Adjustment in Difference-in-Differences Studies." *Health Services Research* 56 (5): 932–941.

Appendix

Table A1. Sample distribution

Cercle	Commune	Village	Type	Total HHs
Bougouni	Keleya	Dialakoro	Control	33
Bougouni	Sido	Siratogo	Control	33
Bougouni	Sido	Sakoro	Control	32
Bougouni	Syentoula	Dossola	Control	34
Koutiala	Gouadji-kao	N'Togonassa	Control	32
Koutiala	Konina	Konina	Control	32
Koutiala	Konseguela	Konseguela	Control	33
Koutiala	Zanfigue	Bobola-Zangasso	Control	29
Yanfolila	Wasselou-Balle	Goulala 1	Control	35
			Control total	293
Bougouni	Danou	Dieba	Treatment	25
Bougouni	Faradiele	Flola	Treatment	27
Bougouni	Faragouaran	Sibirila	Treatment	15
Bougouni	Kouroulamni	Madina	Treatment	19
Koutiala	N'golonianasso	N'golonianasso	Treatment	48
Koutiala	Songoua	Sirakele	Treatment	47
Koutiala	Sincina	Nampossela	Treatment	33
Koutiala	M'Pessoba	M'Pessoba	Treatment	46
Koutiala	Fakolo	Zansoni	Treatment	24
Yanfolila	Gouanan	Yorobougoula	Treatment	35
			Treatment total	319
			Total households	611

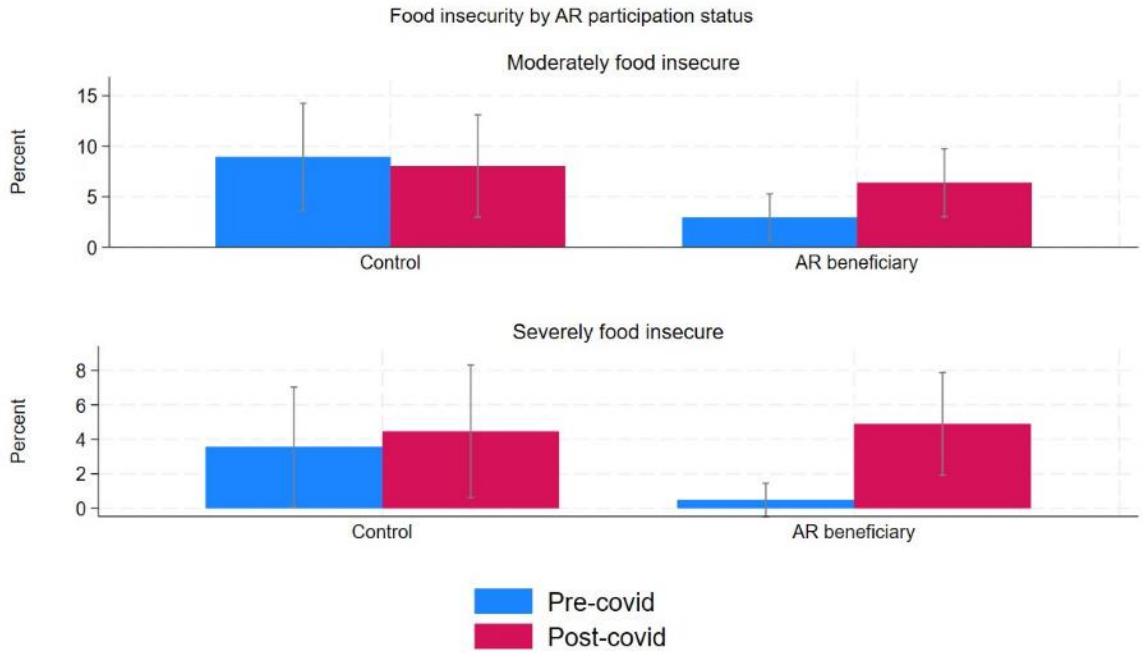
Table A2. Descriptive summary by survey round

	(1) 2014	**(2)** 2021	**(3)** **1 vs 2**
DOMAIN: ENVIRONMENT			
Inorganic fertilizers used (kg/ha)	150.9	103.5	***
Real value of pesticides used (CFA/ha)	2.59	4.39	***
Real value of fertilizers per ha ('000 CFA)	21.0	23.4	***
Real value of purchased seeds per ha ('000 CFA)	3.23	4.35	***
Real value of purchased improved seeds per ha ('000 CFA)	0.16	0.68	***
Accessed extension services (%)	79.2	63.6	***
Practiced intercropping (%)	4.41	4.58	
HH practices cereals/legumes intercropping (%)	1.14	1.96	
Household uses fertilizer (%)	93.5	93.3	
HH uses improved seeds: Sorghum	0.49	8.66	***
HH uses improved seeds: Groundnut	1.14	8.82	***
DOMAIN: PRODUCTIVITY			
Yield of maize(ton/ha)	1.74	1.22	***
Yield of sorghum(ton/ha)	1.21	1.27	
Yield of groundnut(ton/ha)	0.23	0.26	
Yield of green bean (ton/ha)	1.03	0.98	
Yield of cotton(ton/ha)	0.86	0.58	***
Yield of okra(ton/ha)	0.87	0.75	
Yield of legumes(ton/ha)	0.85	0.98	**
Number of crop items produced	1.06	0.90	***
Tropical livestock units	87.6	83.8	*
Agricultural labor (person-days/ha)	74.3	43.6	***
Labor profitability (CFA/person-day)	41.7	16.7	***

	(1) 2014	(2) 2021	(3) 1 vs 2
DOMAIN: ECONOMIC			
Annual net household income (PM) ($ PPP)	3565.8	3924.5	
Net crop income (PM) ($ PPP)	2481.6	3102.3	***
Net livestock income (PM) ($ PPP)	152.0	284.0	***
Other income ($ PPP)	31.5	4.49	***
Per capita daily expenditure (2011 PPP)	2.32	3.58	***
Non-poor (>$1.9 2011 PPP) (%)	48.7	58.3	***
Non-poor (agr. durable assets) (%)	62.9	61.1	
Non-poor (non-agr. durable assets) (%)	60.5	60.1	
Real value of agricultural sales ('000 CFA)	704.1	186.2	***
Households with diversified livelihoods (%)	41.0	37.9	
DOMAIN: HUMAN			
Household dietary diversity score (out of 12)	6.34	5.12	***
Incidence of food secure households (%)	48.4	44.4	
Months of food shortage (last year)	0.44	0.79	***
HAZ	-1.64	-2.18	***
Prevalence of stunting (%)	37.9	49.3	***
WHZ	-0.55	-0.40	
Prevalence of wasting (%)	13.9	12.8	
WAZ	-1.34	-1.49	
Prevalence of underweight (%)	28.6	37.9	***
BMI	21.7	22.3	**
Prevalence of underweight (%)	14.6	15.2	
DOMANI: SOCIAL			
Plots under female responsibility (%)	21.8	6.21	***
Livestock under female responsibility (%)	7.60	2.88	***
Observations	**612**	**612**	**1224**

*Note: T-test significance level: * p<0.1; ** p<0.05; *** p<0.01*

Figure A1. Trends in food insecurity before and after COVID-19



Note: Food insecurity defined based on current & retrospective Food Insecurity Experience Scale (FIES) data collected at follow-up.

Table A3. Incidence of crop-growing households

	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grows maize (%)	2.46	2.62	-4.87	-5.32	-2.29	-2.33	1.36	1.25
Grows sorghum (%)	4.25	4.85	-4.42	-6.87	-2.67	-4.63	1.17	-1.2
Grows bean (%)	0.89	0.61	-5.32	-4.22	-2.18	-1.47	-4.6	-3.61
Grows groundnut (%)	0.18	0.44	10.64	12.4	7.12	7.68	16.19	17.01
Grows okra (%)	0.53	0.45	-8.67	-9.32	-4.91	-4.81	-2.87	-2.75
Grows cotton (%)	-11.5**	-10.46**	-21.41	-23.77	-18.69*	-20.42*	-6.04	-7.69
Grows millet (%)	-6.38	-6.04	-21.23*	-23.67**	-16.17*	-17.55*	-14.72	-16.35
Grows legumes (%)	6.42	6.37	27.6	29.82	21.03*	21.87*	25.37*	26.65*

Note: Columns 1 and 2 report DiD impact estimates based on within-village comparison. Columns 3 and 4 report impact estimates based on out-of-village comparison. Columns 5 and 6 report impact estimates where all non-beneficiaries inside and outside program villages are used as counterfactual. Columns 7 and 8 report results from the placebo test. Results in columns (1), (3), (5) and (7) are from simple DiD, while those in columns (2), (4), (6) and (8) are from DiD with regression adjustment where we control for NDVI average and coefficient of variation in the 12 months prior to the interview. 0.01 - ***; .05 - **; .1 - *