Impacts of Africa RISING in Malawi

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

Author affiliations [1]World Bank, [2]International Food Policy Research Institute, [3]Michigan State University, [4]University of Bologna

Published by International Food Policy Research Institute, International Livestock Research Institute, International Institute of Tropical Agriculture

December 2023

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.

ILRI INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE   IFPRI   IITA Transforming African Agriculture

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 large-scale sustainable intensification (SI) program that has been implemented in Central Malawi's Dedza and Ntcheu districts beginning in 2012. Using a participatory action research framework, the program validated and promoted alternative SI options including fertilized maize, maize-legume intercropping, intercropping of two compatible legumes, cereal-legume rotation, and double-row planting of legumes. Impact is estimated on several SI indicators and domains using two rounds of panel data and difference-in-differences techniques. The unique study design allowed us to estimate impact 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). We also conduct a placebo test comparing non-beneficiaries in the two counterfactual groups. The within-village comparison shows positive impact on several agricultural and economic indicators including access to agricultural information, value of harvest, on-farm diversity, labor productivity, annual net household income, per capita household consumption expenditure, household wealth, commercial orientation, and household dietary diversity score. We do not find a statistically significant impact on human indicators such as child and maternal nutrition. 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 important lessons learned to inform future program design and impact assessments.

1. Introduction

Despite improvements over the last several decades, sub-Saharan Africa (SSA) continues to lag behind other developing regions on several fronts including productivity gains, poverty reduction, and food security. Average cereal yields in SSA have stagnated at 1.6 ton/ha since 2020, about a third of Southeast Asias's average yields (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, 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 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. 2016; 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

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

2012—Ethiopia, Malawi, Tanzania, Zambia, Mali, and Ghana. Its goal is to "to provide pathways out 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 Malawi and the interventions implemented in Dedza and Ntcheu districts in its Central region, primarily through an on-farm participatory approach known as *mother-and-baby* trials (Snapp 2002; Bezner Kerr, Berti, and Shumba 2011). The approach involves establishing researcher-designed, scientifically replicable agricultural demonstration (*mother*) trials centered around farmer action groups to test alternative SI innovations. Subsequently, group members would set up adaptive (*baby*) trials to test a subset of innovations from the mother trial on their plots. In the initial years of the program, the innovations promoted included fertilized maize, maize-legume intercropping, and intercropping of two compatible legumes, known as doubled-up legumes. Baby farmers received improved seeds from the program for the crop of their choice.

Groundnut, beans, pigeon peas, and soybeans were among the pulses targeted by the program. According to unpublished documents, baby farmers selected no more than four mother trial-tested technologies. The focus of the program expanded over the years to include cereal-legume rotation, double-row planting of legumes, supplementary feeding of goats with *Faidherbia* pods and *Gliricidia* leaves, construction of housing for goats, local processing of soybean and other grain legumes into nutritious products, and introduction of nutrient-dense common bean varieties (IITA 2019).

2. Materials and methods

2.1 Setting

In 2016, an estimated 69 percent of Malawi's population lived on less than $1.90 per-capita/day (2015 international prices) and, in 2019, 17 percent of the population's dietary caloric intake was found to be inadequate for maintaining a normal active and healthy life (defined as undernourishment) (World Bank

2022). Maize, which accounts for about half of the total-plant based caloric intake, is the main staple food, with pulses and groundnuts accounting for 7 percent (FAO 2013). While Malawi has made some progress toward achieving global stunting targets among children under five years old, it still faces a very high stunting prevalence of about 37 percent (USAID 2021).

Heavy reliance on rainfed agriculture exposes the country to cyclical rainfall patterns and their adverse effects on the levels and stability of yields and food security (Matita et al. 2021). For example, the large-scale floods of 2015 that were followed by a prolonged dry spell between 2015 and 2016 prompted the government to declare a state of disaster and had a significant impact on its economic growth (World Bank Group, United Nations, and European Union 2016; IMF 2017). Also, soils in vast areas of the country are low in essential nutrients due to inadequate resource inputs and continuous maize cultivation (Mungai, Messina, and Snapp 2020). Recognizing this, the government has been providing subsidies for inorganic fertilizers and improved cultivars for maize (since 2004/05) and legumes (since 2008/09) through its Farm Input Subsidy Program (FISP) to improve soil quality and food security.

2.2 Data

Panel data were collected as part of the monitoring, evaluation, and impact assessment of the Africa RISING program. Baseline data were collected in 2013 covering 26 program target and 28 program non-target (control) communities and three groups of households (IFPRI 2015). The first (group 1) consisted of 405 households identified as *baby* trial farm households based on project monitoring data shared by program implementers just before the baseline survey. The second (group 2) included a random sample of households in program target communities that did not participate in the program at baseline. As we show later, some of the households in group 2 reported participating in the program during the follow-up survey in 2019, and this group has benefited from the program in subsequent years. The third group of control households (group 3) consisted of a random sample of households from control communities. There was no expectation that the program would expand to control communities over time.

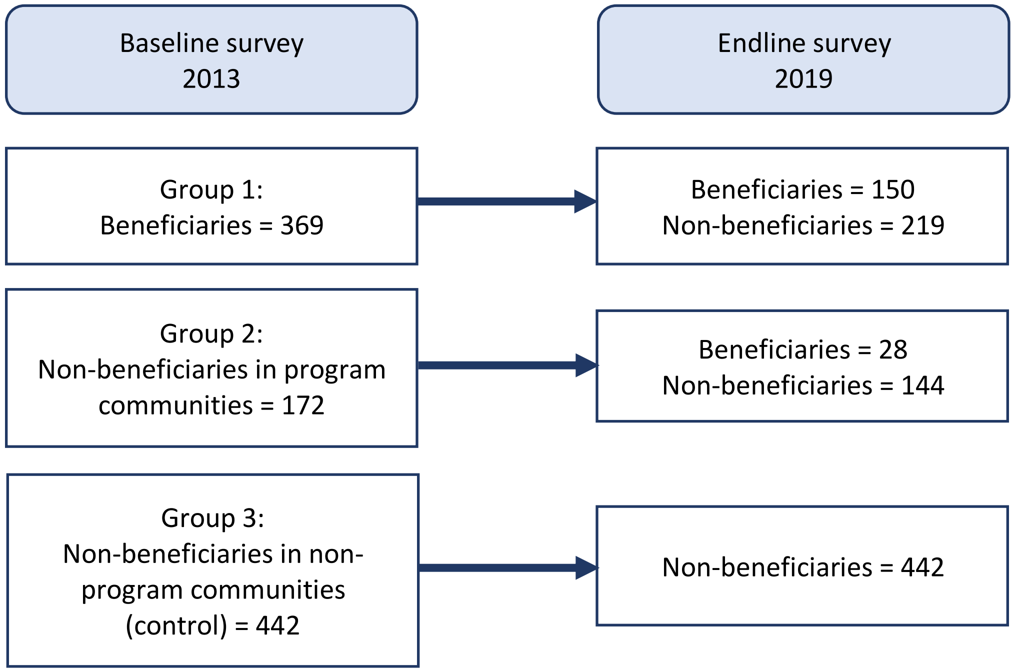
Before the implementation of program interventions, target districts (Dedza and Ntcheu) were characterized based on various GIS variables including length of growing period (LGP), elevation, temperature, rainfall, and market access (HarvestChoice 2013). Elevation and temperature-adjusted rainfall were found to adequately capture the heterogeneity in target districts and were used to stratify target areas. Program implementers then selected four Extension Planning Areas (EPAs) (Golomoti and Linthipe in Dedza district; Kandeu and Nsipe in Ntcheu district) and four extension Sections (Golomoti Centre, Mposa, Kampanje, and Mpamadzi) to represent three target strata: low rainfall and low elevation, medium rainfall and medium elevation, and high rainfall and high elevation. A sample of 200 non-beneficiary households (group 2) was randomly selected from program villages from a household list from which households in group 1 were excluded. This sample was distributed in each of the four program sections proportionate to the share of the total population in each section. This allocation was followed by a random sample selection of a fixed number of households per section.

Households used to construct the counterfactual group were identified using the following steps. First, four sections with comparable agroecologies were identified (Thete in Lobi EPA, Mtakataka Center in Mtakataka EPA, Mwalaoyera in Nsipe EPA, and Sitolo in Kandeu EPA) as program target sections while at the same time distanced 7–12 kilometers away from each other to minimize contamination. Subsequently, control villages were chosen such that they were physically isolated from program target villages (Haile et al. 2016). After 28 control villages were selected using probability proportional to size, 20 households were sampled from each village to form the control group (group 3).

A follow-up survey was conducted in 2019 in which 96 percent of baseline households were interviewed. The impact assessment is based on a balanced panel of 985 households interviewed in each wave. Comparable structured questionnaires were used to collect data on a wide range of topics including household demography, agricultural production, food consumption, non-food expenditure, and anthropometry for children under the age of five years old and women 15 to 49 years old. The reference period for agricultural production data was the main cropping season—October 2012 to May 2013—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 2013 and 2019. Households that reported participating in the program for at least three years were defined as program beneficiaries.

Figure 1 summarizes the study design with additional details about the evaluation design available in Haile et al. (2016). A significant percentage of households in group 1 identified as baby trial farm households in the project monitoring data at baseline reported never participating in the program during follow-up.

Figure 1. Study design



Note: Numbers in the figure refer to households.

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 SI is best measured as a multidimensional outcome focused on enhancing agricultural productivity while preserving social, economic, and environmental sustainability where each domain encompasses multiple elements (Haggar et al. 2021). Following previous research, we adopt the 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 indicator, metrics, and means to explore five domains of sustainability (agricultural production and productivity, economic, environment, human, and social). This framework is particularly appealing for our study because of its flexibility for indicator selection and its ability to examine potential trade-offs among domains and indicators. The SIAF 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 effects of mechanization on 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, leaving crop residues on the field, use of manure and inorganic fertilizers, maize-legume intercropping, and 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 legume yields (tons per hectare), the number of unique crops planted, livestock types owned (cattle, equines, sheep, goats, pigs, chickens, beehives, and other), tropical livestock units (TLU), agricultural labor used (person-days per hectare), value of harvest (MWK), and labor productivity measured as total value of harvest per unit of labor (MWK/person-days). Unit values computed by dividing total sales revenue by quantities sold are used to monetize harvest, in line with Deaton (1988).

Under the *economic* domain, we examine annual net household income, which encompasses crop, livestock, wage, transfers, self-employment, and other income. We also consider per capita daily consumption expenditure and poverty defined based on $1.90 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 expenditures on purchased food and imputed values of food consumed from own production or gifts. The latter two are monetized using unit values computed by dividing total expenditure on purchased foods with total quantity of purchased foods. To capture nonmonetary 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 that fall in the top tercile of asset-based indices are defined as non-poor.

Unlike consumption-based poverty that 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 been shown that the two do not always track well against each other (Foreit and Schreiner 2011). Additionally, we examine commercial orientation of farmers based on the value of agricultural products that were sold. Strong commercial orientation can both enhance productivity by increasing farmers' incentives to meet market demand and promote better household diets and nutrition by enhancing purchasing power. All monetary values from the follow-up survey are converted to real values based on the World Bank's Consumer Price Index (World Bank 2020).

Under the *human* domain, we consider two indicators based on food consumed inside the household over the last seven days. The first is the household dietary diversity score (HDDS) based on 12 food groups (Kennedy, Ballard, and Dop 2010), and the second is food security status based on whether household dietary caloric intake is above the minimum dietary caloric requirement (MDER) (FAO 2008). We complement these indicators with the number of (self-reported) months of food shortages during the postharvest (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 second captures food security over a longer period.

To complement household-level food security indicators, we use anthropometric data to construct individual-level nutritional indicators for children aged 0–59 months and women 15–49 years old. We follow WHO child growth standards (WHO 2006) to construct child indicators of 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 conditions and future predictions of individuals' growth and likelihood 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 an indicator of 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 females, either solely or jointly with males. We control for several other indicators in our multivariate framework including household socio-demographic characteristics and an index for market access based on self-reported travel time to various basic services. Households in the top tercile of the market access index are identified as living in remote areas. 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).[2] A descriptive summary of key indicators by survey round is provided in Appendix Table A1.

[2] NDVI data has a temporal resolution of 16 days, a spatial resolution of 250 meters, and covers the time period 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 (October to May).

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 for 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 previous companion study showed that households in group 1 (beneficiaries) are systematically different from those in group 2 (non-beneficiaries living in program villages) and group 3 (non-beneficiaries living in control villages) (Haile et al. 2016). In our case, households in group 1 are statistically larger, more educated, less likely to be female-headed, enjoy larger farms, report higher shares of livestock managed by females, are more likely to be non-poor based on durable agricultural and non-agricultural assets, and historically face less variable weather compared to households in group 3 (Table 1).

Table 1. Descriptive summary of selected variables at baseline (2013)

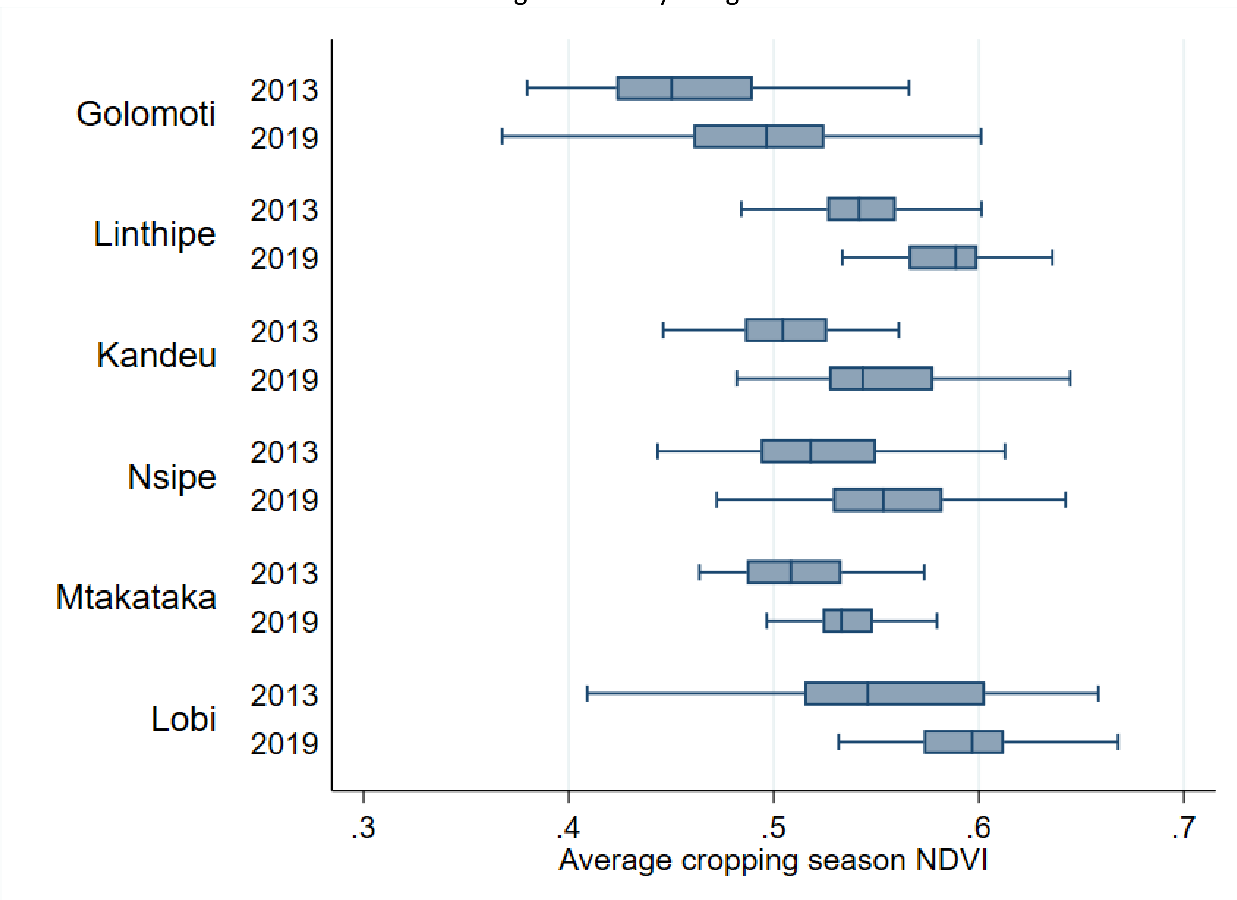
	(1) Group 1	(2) Group 2	(3) Group 3	(4) 1 vs 2	(5) 1 vs 3
Household demographic variables					
Household size	5.01	4.62	4.67	**	**
Age of the household head (years)	46.0	47.2	46.3		
Female-headed (%)	27	38	35	**	**
Household socioeconomic variables					
Average adult education (years)	5.16	4.30	4.54	***	***
Area operated (ha)	1.20	0.94	0.85	***	***
Plots managed by women (%)	86.2	87.7	83.9		
Livestock managed by women (%)	5.71	4.70	3.95	**	***
Tropical livestock units	0.45	0.3	0.21	***	***
Non-poor (agr. durable assets) (%)	48.1	31.6	29.3	***	***
Non-poor (non-agr. durable assets) (%)	38.3	28.2	30.7	**	**
Landscape level variables					
Mean monthly NDVI (historical)	0.54	0.55	0.54	***	
CV of NDVI (historical)	15.1	14.7	17.1		***
Remote households (%)	33.4	33.3	28.4		
Observations	369	174	442	543	811

Note: Columns 1–3 report average values (of continuous or dummy variables) while columns 4 and 5 show statistical significance levels from pair-wise 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

Some of these differences—not shown in Table 1 for reasons of space—persist when comparing non-beneficiaries in group 2 with non-beneficiaries in group 3, suggesting potential systematic differences between program target and control villages, given that households in both group 2 and group 3 were sampled randomly following a complete listing exercise.

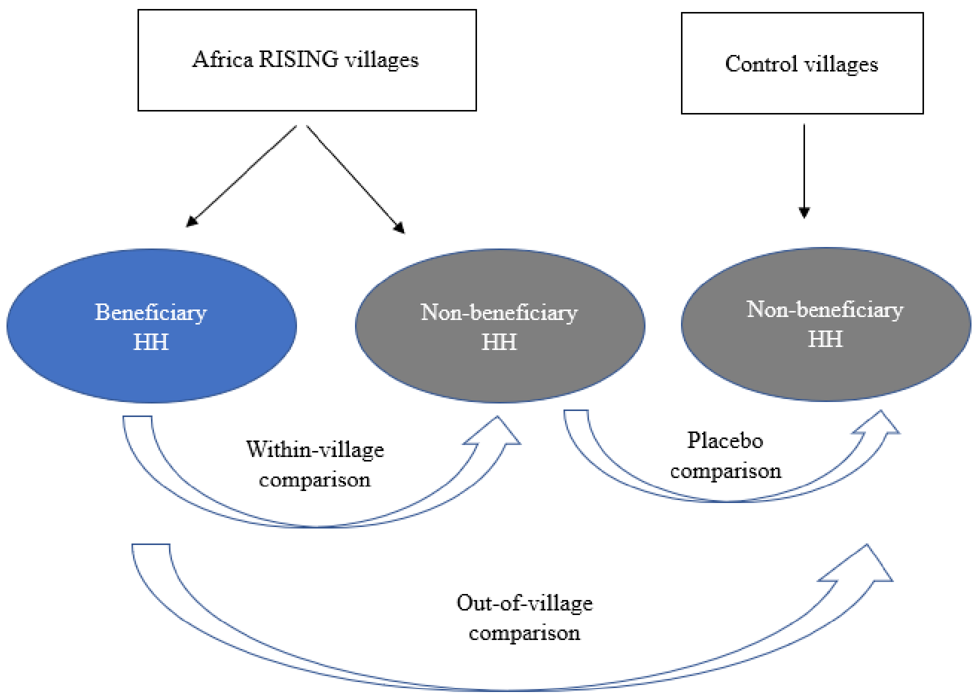
Indeed, geophysical variations within Dedza and Ntcheu have previously been documented with implications for agricultural production and productivity (Mungai et al. 2016). For example, among the four program target EPAs, Linthipe is associated to the highest average NDVI for the cropping seasons, corresponding with the baseline and follow-up data collection, while Golomoti shows the lowest average NDVI (Figure 2). Variation is also observed among control EPAs, with Lobi and Mtakataka EPAs reporting the highest and lowest average NDVI, respectively. For all sites, average NDVI at follow-up was higher than that at baseline with some variation in the percentage change by site.

Figure 2. Study design



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 3). These include comparisons between beneficiaries with non-beneficiaries in program target villages (*within-village* comparison), beneficiaries with non-beneficiaries in control villages (*out-of-village* comparison), beneficiaries with non-beneficiaries in program target and control villages (*overall* comparison), and non-beneficiaries in program villages with non-beneficiaries in control villages (*placebo* comparison).

Figure 3. Alternative pairwise comparisons



Note: HH stands for households

By design 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 to us. 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., social networks).

While spillovers are desirable to multiply 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 from attending field demonstration days. For example, in cases where non-beneficiaries in program villages benefit by significant spillovers but are not classified as participants when asked during follow-up, within-village impact estimates would be underestimated. Out-of-village impact estimates, on the other hand, 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 this latter effect, *placebo* comparison will capture potential spillover effect.

Our identification is based on propensity score matching (PSM) and difference-in-differences (DiD) strategies that are widely used in the evaluation literature to estimate the impact of policies and programs (for a 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 to a non-treated unit with similar characteristics. More specifically, PSM expresses the probability that a unit will enroll in 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).

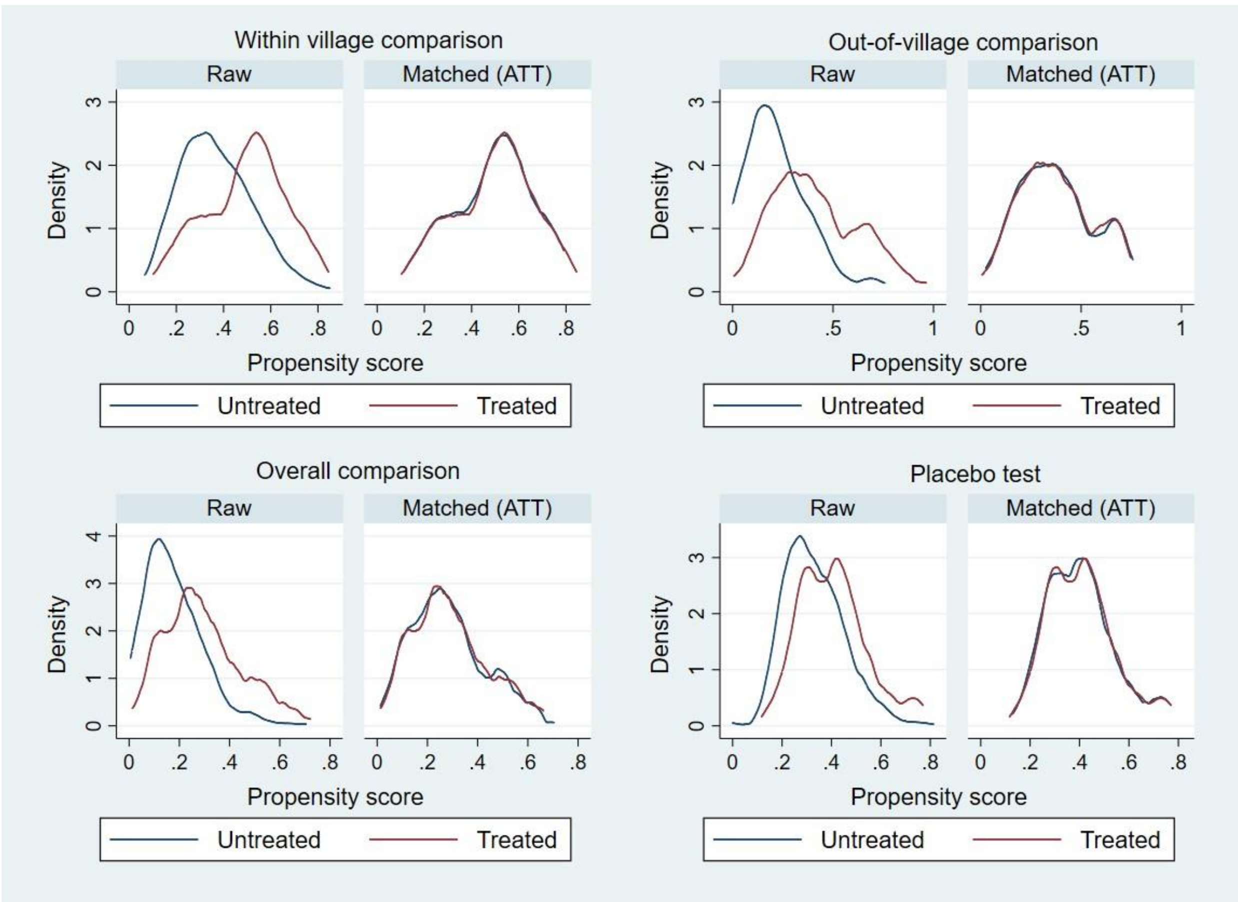
In our analysis, we computed the propensity score based on the following characteristics at baseline: age and gender of the household head, educational level, household's assets (agricultural and non-agricultural), elevation and size of the land, maize production, use of inputs for maize production, the total number of person-days needed for maize production, an index measuring the distance of the homestead from basic public services, and historical monthly average NDVI from 2000 to 2014 (Haile et al., 2016). More specifically, we estimated the following logistic regression:

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

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

Figure 4 shows the distribution of the propensity score between treatment and control groups before and after matching for all the pairwise comparisons was conducted. After matching, the distribution of the propensity score for the untreated group overlaps with that of the treated groups for all the four comparisons, suggesting that the matching was successful in balancing the covariates between treatment and control groups.

Figure 4. Distribution of propensity score before and after matching for four comparisons



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.

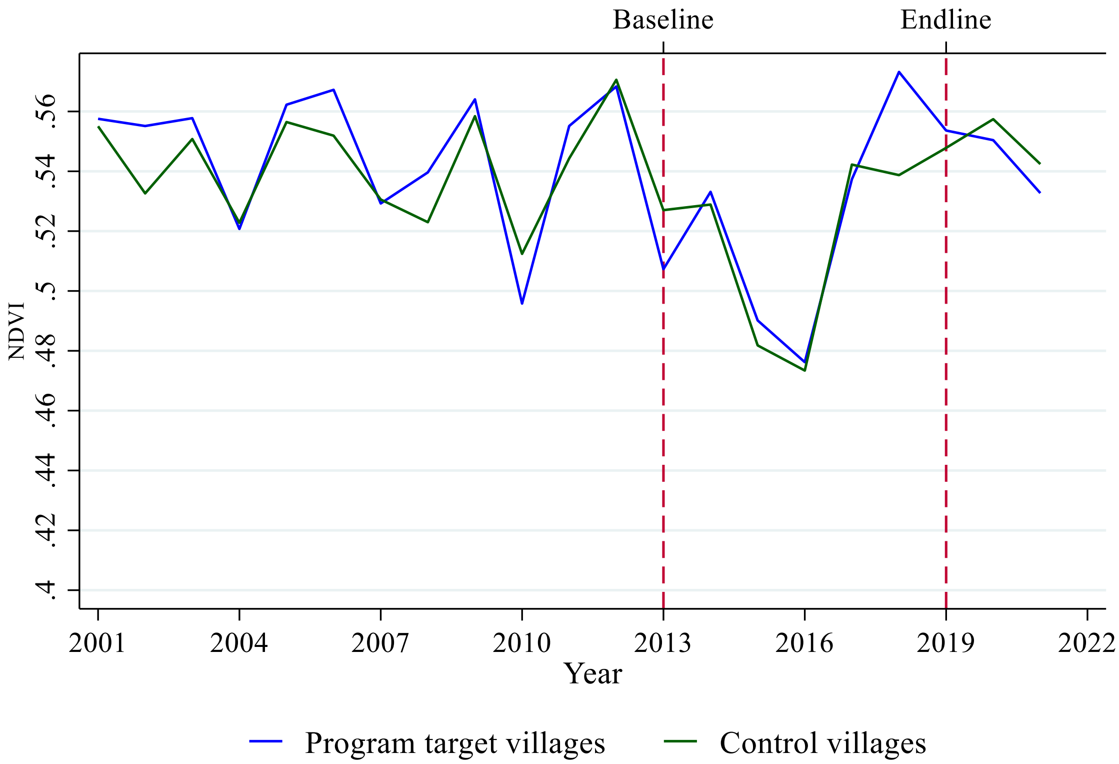
$$\boldsymbol{Y_{ht} = \alpha + \beta T_{ht} + \gamma P_{ht} + \delta(T_{ht} * P_{ht}) + \theta NDVI_{ht} + \varepsilon_{ht}} \quad 1$$

where *h* and *t* represent household and time, Y expresses the outcome indicator, T an indicator that takes value one for program beneficiaries (treated group) and zero for non-beneficiaries (control group), P an indicator that takes value one for follow-up and zero for baseline—capturing aggregate factors affecting Y the same way between the two groups, $NDVI_{ht}$ represents the NDVI index included as control variable in the regression adjustment (RA) specification of the DiD, and ε is the 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 the 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 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, this type of test is not possible because there are only two data points available for the outcome indicators and no time series data 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 program villages in the years preceding the 2019 follow-up survey (Figure 5). The dip in NDVI between baseline and follow-up corresponds to the prolonged dry spell that occurred in the 2015–2016 rainfall season (World Bank Group et al. 2016).

Figure 5. Average NDVI of the cropping season by treatment status



To check for robustness, we conduct three additional regressions. First, we re-estimate a version of the model in Equation 2 controlling for NDVI both in terms of the historical average and average as well as the coefficient of variation for the cropping seasons for which agricultural production data were collected.[3] 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 regression adjustment (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 the different counterfactuals shown in Figure 3. By construction, $\hat{\delta}$ from the *overall* comparison will be an average of the impact

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

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.

3. Results and discussion

Table 2 shows DiD estimates on indicators under the environmental domain. Beneficiaries are more likely to receive agricultural advisories than the control group, irrespective of the type of counterfactual group (Table 2, columns 1–6). In the within-village comparison, impact is equivalent to an increase of 50 percent from the baseline average compared to the control group (as reported in Appendix Table A1). The program also fostered adoption of improved groundnut seeds irrespective of the type of counterfactual group (Table 2, columns 1–6). However, we also observe that beneficiaries are less likely to use manure but use higher intensity of inorganic fertilizer relative to the control group in non-program villages (Table 2, columns 3–6). The causal mechanism of this finding is unclear, given that the program emphasized complementarity—not substitution—between manure and fertilizers.

Inorganic fertilizers complement organic nutrient sources from manure and nitrogen fixation by leguminous crops through rotation or intercropping with cereals as well as agroforestry (Nalivata et al. 2017). A previous study in Malawi found a positive association between use of manure and inorganic fertilizers (Holden and Lunduka 2012). We do not find a statistically significant impact on intercropping of maize with legumes, while we find a negative impact—at a low significance level and only for some of the comparisons conducted—of the program on intercropping of short-term and long-term legumes, despite the strong focus of the program on legumes and nuts. According to program implementers, maize-pigeon pea and maize-common bean intercropping systems are quite common in the study area and, because this information is unavailable in our data, it can be difficult to estimate impact unless the intercropping indicator considers whether cultivars used were improved or not. We also do not find statistically significant impact on the farm-level rotation indicator. Groundnut or soybean and maize rotational systems were among the innovations promoted by the program. Unfortunately, only rotational data at the farm level, not the crop level, are available.

Placebo test results (Table 2, columns 7 and 8) show a marginally significant and positive impact on the likelihood of receiving agricultural advice among non-beneficiaries in program target villages relative to non-beneficiaries in non-program villages. Given that the source of agricultural information is both formal (from agricultural development/extension agents and farmers' groups) and informal (from neighbors, friends, and other farmers), *placebo* test results may suggest either systematic village-level differences in access to agricultural information by treatment status or presence of spillovers. In participatory programs, agricultural information spreads in program villages through, for example, field demonstration days, which are often held to introduce residents to program innovations. The impact on the other environmental indicators is not statistically significant, which suggests that even if spillovers occurred, they did not impact adoption behavior along the dimensions measured.

Table 2. Impact on environmental domain (2013–2019)

	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Received agricultural advice (%)	51.1***	51.9***	19***	17.5***	33.1***	33.9***	14.5*	17.1**
Used improved beans (%)	1.2	1.9	-4.3	-8.7	-2.3	-4.4	-2	-6.2
Used improved groundnut (%)	18.6***	18.1***	25.8***	26.1***	23.2***	23.3***	-7.3	-6.1
Used improved pigeon pea (%)	1.4	1.1	2.3	3.6*	2	2.6	-1	-0.6
Practiced crop rotation (%)	5.5	5.9	2.9	-0.6	3.4	2.9	5.2	2.1
Practiced fallowing (%)	-1.9	-2.2	-1.5	0.7	-1.9	-1.2	0.2	1.1
Applied manure (%)	-1.4	-0.8	-14.8**	-17.3***	-9.5*	-10.3**	0	0.1
Practiced cereal-legume intercropping (%)	4.6	4.7	9.8**	9.7*	6.3	7.3	4.9	0.9
Practiced legume-legume intercropping (%)	-1.8	-1.8	-6.8	-11.1**	-4.4	-6.5*	-4.6	-6.2*
Inorganic fertilizer used (kg/ha)	24.5*	25*	41.6**	40.1**	35.2***	34.2**	-6.4	-5.3

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 measures. 0.01 - ***; .05 - **; .1 - *

Impact on agricultural production and productivity is reported in Table 3. Results from the within-village comparison (Table 3, columns 1 and 2) show positive impact on the total value of harvest (79 percent increase from baseline), unique livestock types owned (42 percent increase from baseline), and total livestock wealth in TLU (17 percent increase from baseline). Positive impact on livestock may be due to supplementary feeding of livestock with *Faidherbia* pods and *Gliricidia* leaves and the introduction of improved livestock housing. Impact estimates from the out-of-village comparison (Table 3, columns 3 and 4) or that of the overall comparison (Table 3, columns 5 and 6) show significant impact on legume yield driven mostly by yield gains in cowpeas, beans, and legumes. Cowpea was one of the main crops targeted by the program: it also introduced a new cowpea cultivar called *IITA18*.

These results, coupled with insignificant impact on legume yield in within-village and placebo comparisons, may suggest potential spillovers and/or systematic village level differences. As we showed above, access to advisory services improved not only among beneficiaries but also among non-beneficiaries in program villages. Information dissemination within program villages might have affected adoption patterns along dimensions not measured here that subsequently affected agronomic outcomes of non-beneficiaries in program villages. For example, if the latter were associated to higher biophysical suitability for legumes than control villages and, in turn, this higher suitability was correlated with other systematic time-varying factors, out-of-village impact estimates on legume yields would be biased.

Impact on crop diversity—proxied by the number of different crops produced—from within-village comparison is not statistically significant, and, when the counterfactual group includes non-beneficiaries in non-program villages, we find a negative impact. To further validate the impact pathway, we estimate impact on the likelihood of growing six food groups: cereals, starch roots and tubers, legumes and nuts, vegetables, fruits, and cash crops (e.g., cotton, tobacco, tea, and coffee). Results shown in Appendix Table A2 show a negative impact on the likelihood of growing roots and tubers but no impact on the likelihood of growing legumes. Impact estimates show that beneficiaries own more diverse and more livestock relative to the control group, which may be due to the introduction of improved livestock feed and housing.

Table 3. Impact on productivity domain (2013–2019)

	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Yield (ton/ha)								
Maize	0.03	-0.02	-0.02	-0.22	-0.01	-0.1	-0.01	-0.02
Bean	0.04	0.03	0.23***	0.19***	0.16***	0.13**	0.07	0.05
Soybean	-0.87	-0.82	-0.27	-0.2	-0.56	-0.55	0.43*	0.42*
Pigeon pea	0.65	0.64	1.52*	1.63*	1.12	1.07	0.36	0.38
Groundnut	0.07	0.06	5.44	4.99	2.99	2.59	3.41	2.69
Cowpea	0.25	0.24	0.72***	0.71***	0.43*	0.39*	0.07	0.1
Cereal	-0.04	-0.08	-0.02	-0.2	-0.04	-0.12	0.05	0.02
Legume	0.07	0.06	0.41***	0.35***	0.27***	0.23***	0.1	0.13
Real value of total harvest ('000 MWK)	79.76***	79.39**	-4.77	-10.61	28.28	29.82	-7.87	-12.41
Unique crops produced (#)	-0.14	-0.13	-1.07***	-1.17***	-0.68**	-0.62**	-0.32	-0.36
Unique livestock types owned (#)	0.42***	0.41***	0.15	0.13	0.27***	0.29***	-0.14	-0.12
Tropical livestock units	0.17***	0.17***	0	0.01	0.06	0.08	-0.11	-0.09
Agricultural labor (person-days/ha)	26.93	24.77	15.61	17.24	24.14	20.74	26.26	38.6
Labor productivity (MWK/person-day)	644.7***	604.9***	423.9**	378.4*	529.4***	499.7**	-154.9	-206.6

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 measures. 0.01 - ***; .05 - **; .1 - *

We find a positive impact on the likelihood of being non-poor based on consumption expenditure (10 percentage points), durable agricultural assets (17.5 percentage points), and non-agricultural durable assets (10 percentage points) (Table 4, columns 1 and 2). Unlike the impact on consumption expenditure, impact on ownership of durable assets captures the cumulative effects of program interventions between 2013 and 2019. Overall, based on the within-village comparison, the program had a positive and statistically significant impact on per capita daily expenditure and household net income regardless of the type of counterfactual group. Total household net income gains are mainly due to livestock activities and, specifically, earnings from livestock, goat milk, and egg sales that contributed to a +$146 PPP increase in net livestock income for beneficiaries compared to the control group in the within-village comparison. Unlike environmental and productivity indicators, consumption-based poverty and commercial orientation can be more prone to systematic village level differences in commodity prices due to general equilibrium effects or other market-related factors, while durable asset ownership is less prone to such transient factors. According to project M&E data, the number of *baby* farmers increased from 400 at the time of the baseline survey to over 2,000 around the time of the follow-up survey.

In addition, although consumption-based poverty is highly correlated with agricultural income for the target population, it may also reflect spatial differences in access to and income from non-agricultural activities (e.g., *ganyu* labor) that may have varied by treatment status regardless of program presence.

Table 4. Impact on economic domain (2013–2019)

	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income/Earnings ($PPP)								
Annual net household income	496.5**	456.9***	408.2*	435.7**	431.5**	451.7**	-182.7	-186.7
Gross crop income	156.76	143.15	69.38	63.35	96.47	104.28	-133.25	-165.03
Net livestock income	146***	146.9***	129.7***	125.2**	137.6***	142.2***	-36.8	-42.5
Earnings from goat milk sales	6*	5.7**	5.7	6.7*	5.8*	6.5*	-2.7	-1.6
Earnings from eggs sales	3.5***	3.4***	2.8**	3**	3.1**	3.2**	-0.3	-0.3
Earnings from livestock sales	102.4**	107.4**	82.2**	70.4*	90**	89.3**	-16.4	-21.7
Annual income from self-employment	222.9**	206.1**	79.9	115.9	135.5	150.5*	-71.3	-40.8
Poverty								
Per capita daily expenditure ($PPP)	0.6***	0.6***	0.3*	0.4*	0.5***	0.6***	-0.4	-0.4
Non-poor (>$1.9 PPP) (%)	8.7*	9.9*	6.6	8.7	6.6	8.9*	-8.3	-10.3
Non-poor (agr. durable assets) (%)	17.4***	17.5***	9.5	9.7	13.5**	14.6**	6.2	5.5
Non-poor (non-agr. durable assets) (%)	10.1*	9.5*	4.9	6.8	6.6	8.9*	-0.8	3.8
Diversified livelihood								
Households with diversified livelihoods (%)	12	12.8	15.3**	14.7*	14.5*	14.8**	1.3	-0.4

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 measures. 0.01 - ***; .05 - **; .1 - *

Program participation has increased HDDS and household food security regardless of the type of counterfactual group (Table 5, columns 1–6), but impact on child and maternal nutrition is not statistically significant. Results also show a negative impact on the number of months of food shortage during the postharvest season while impact on food shortage during the lean season is not significant (Table 5, columns 1–6). A positive impact on the number of months of food shortage during the post-harvest season from the placebo test is found (Table 5, columns 7–8), suggesting potential worsening of food security in program villages between baseline and follow-up relative to the trends in the control villages. This result also suggests potential underestimation of the impact of food shortages in the within-village comparison. Despite nutritional improvement being one of the program objectives, we do not find improvements in either child or maternal nutrition along the various nutritional indicators examined. Finally, we find a positive effect on the share of plots and livestock managed by women either solely or jointly with men (Table 5, columns 1–2), and improvements in women's management of livestock are mirrored in the diversity and stock of livestock raised (Table 3).[4]

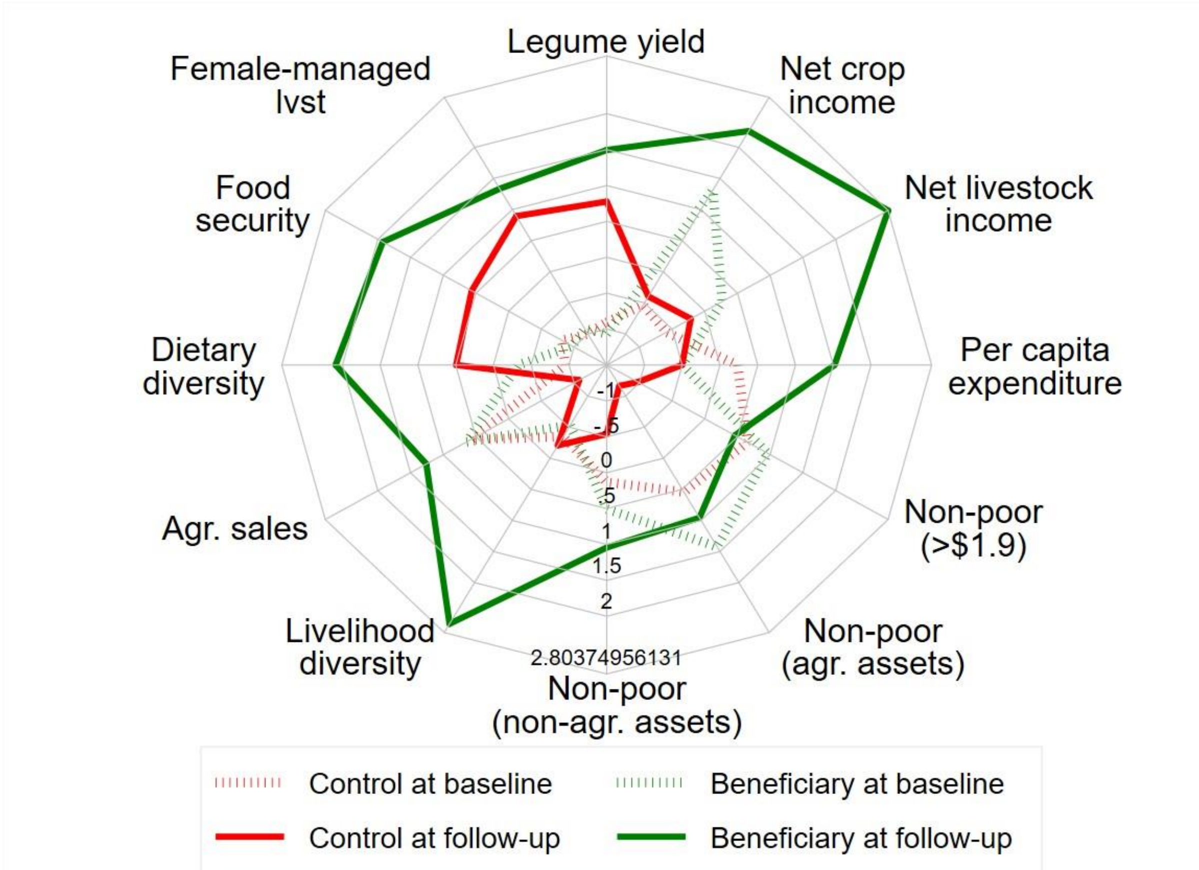
[4] These results do no change when we estimate a DiD model without first matching the observations using PSM, as shown in Tables A3–A6 in the Appendix.

Table 5. Impact on human and social domains (2013–2019)

	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)	0.67***	0.65***	0.84***	1***	0.71***	0.83***	-0.33	-0.26
Incidence of food secure households (%)	9.07	9.36	16.16**	17.79***	12.86**	14.2**	-0.3	-0.38
Months of food shortage (June-Sep)	-0.23***	-0.21***	-0.2**	-0.24***	-0.2***	-0.24***	0.22**	0.2**
Months of food shortage (Oct-Mar)	0.21	0.23	0.01	0.01	0.09	0.1	0.11	0.13
Child								
Height-for-age z-score (HAZ)	-0.12	-0.1	0.12	0.13	-0.03	-0.03	-0.07	0.01
Weight-for-height z-score (WHZ)	0	0.03	-0.16	-0.07	-0.11	-0.04	-0.15	-0.19
Weight-for-age z-score (WAZ)	-0.16	-0.15	-0.19	-0.14	-0.2	-0.19	-0.06	0
Prevalence of stunting (%)	15.6	13.93	11.78	10.77	14.79	15.14	-11.37	-11.75
Prevalence of wasting (%)	-0.79	-0.93	1.97	1.13	0.52	-0.5	3.88	5.72
Prevalence of underweight (%)	11.9	11.23	10.95*	9.77*	10.62*	10.13	-0.72	-0.69
Women								
Body Mass Index (BMI)	-0.67	-0.81	-0.46	-0.41	-0.5	-0.52	-0.14	0.17
Prevalence of undernutrition (BMI<18.5)	1.09	1.08	3.12	2.98	2.38	2.32	1.77	2.15
Prevalence of overweight (25 <BMI<30)	-5.91	-7.26	-3.64	-4.11	-4.67	-5.06	-2.03	2.08
Prevalence of obesity (BMI>30)	-0.08	-0.82	-2.21	-2.46	-1.08	-1.25	-1.72	0.7
Panel B. Social								
Plots under women’s responsibility (%)	12.5	12.5	0.4	1.8	4.7	5.1	-10	-7.8
Livestock under women’s responsibility (%)	13.9**	13.9**	1.2	1	5.8	5.4	-15.4***	-14.8***
Household belonging to farmers' groups (%)	18**	18.2**	-26.9***	-27.5***	-8.5	-5.5	16.7**	21.6***

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 measures. 0.01 - ***; .05 - **; .1 - *

Figure 6. Predicted DiD estimates of the indicators from the overall comparison



The radar graph in Figure 6 represents the DiD estimates across a selection of relevant variables selected from the five dimensions of the SIAF indicators in the overall comparison. The graph compares beneficiaries and non-beneficiaries before (dashed green and red lines, respectively) and after (solid green and red lines, respectively) program participation, as well as contrasting the outcomes of interest between treatment and control groups based on standardized values of each variable.

The differences in outcomes between treated and control households are striking. At baseline there are almost no differences between the two groups in the outcomes of interest, except for net crop income among beneficiary households higher than among non-beneficiary households, yet a clear pattern emerges at follow-up. Beneficiary households enjoy both higher livelihood diversity and net livestock income compared to households in the control group. Per capita expenditure has increased as well, yielding a higher share of households defined as non-poor (i.e., those falling above the poverty line of $1.90 per day). Beneficiary households also show higher levels of food security and a greater dietary diversity score.

When comparing beneficiary households before and after Africa RISING participation (green dashed and solid lines), most of the indicators under analysis have increased, except for the likelihood of being non-poor when measured based on the poverty line of $1.90 per day or based on agricultural assets. Values of the indicators among control households also differ between baseline and follow-up (red dashed and solid lines): while non-agricultural assets and agricultural sales are lower at follow-up than at baseline, legume yield, food security, net livestock income, and dietary diversity are all higher at follow-up, along with the proportion of plots owned by women.

Indicators among control households between baseline and follow-up vary as well. While at follow-up non-agricultural assets and agricultural sales exhibit lower values than at baseline, legume yield, food security, and dietary diversity show higher values. In addition, the share of plots owned by women and net livestock income have also increased.

4. Conclusion and implications

Sustainable intensification (SI) of the smallholder sector offers an integral approach for achieving several of the Sustainable Development Goals (SDGs) in Africa south of the Sahara (SSA)--from eradication of hunger and poverty to reversing land degradation and combating climate change. SI of agriculture 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. SI focuses on improving resource-use efficiency while achieving higher production from the same resources and enhancing beneficial environmental and social services, contrasting sharply with the historical focus of agricultural development interventions that primarily target yield and profitability of a few staple cereal crops such as maize and rice.

We evaluated the impact of an SI program in Central Malawi that was implemented between 2012 and 2021, the goal of which was to provide smallholders pathways out of hunger and poverty. To that end, the program has validated and promoted several SI innovations including fertilized maize, maize-legume intercropping and rotation, intercropping of two legumes, row planting of legumes, improved legume cultivars, and improved livestock feeding and housing. Given the diversity of interventions and expected effects on several dimensions of sustainability, we use the Sustainable Intensification Assessment Framework to inform the selection of indicators across 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 via spillovers, we took advantage of the unique evaluation design using panel data collected in 2013 and 2019 and a propensity score matching with a difference-in-differences technique 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 mechanism at play.

We document positive impact on access to agricultural information, value of harvest, number and diversity of livestock, labor productivity, annual net household income, per capita household consumption expenditure, likelihood of being non-poor based on monetary and asset-based wealth indices, commercial orientation, and household dietary diversity score, and on the share of plots and livestock managed by women. Beneficiaries also report fewer months of food shortages during the post-harvest season but not the lean season.

We do not find improvements in the nutrition of women and children despite the program's focus on nutritional status of these groups. However, undernutrition of women and children remains a persistent public health and development challenge in Malawi, and integration of behavioral change communication into agricultural programs can help maximize nutritional gains of agriculture-focused programs. As expected, we find evidence of spillover effects on access to agricultural extension and advisory services, given the participatory nature of the program and the demand-driven approach used to validate and promote innovations that include field demonstration days.

A positive association between agronomic indicators, poverty, household dietary diversity, and commercial orientation is also documented. This finding highlights the importance of interventions that ease output market constraints (e.g., remoteness, output price fluctuations, and postharvest losses due to pests and insects) that negatively impact the likelihood, timing, and volume of agricultural sales. Given that weather variability and climatic risks are among the main challenges facing smallholder farming households, there is an urgent need to prioritize investments to build climate-resilient agricultural systems.

We also learned some important lessons that point to areas of improvement for future impact assessments of sustainable intensification programs. First, while we do not find a positive impact of program participation on the incidence of intercropping, impact is positive on several agronomic and economic indicators. Given that intercropping practice (cereals-legumes and one short-term and one long-term legume) is the main target of the Malawi program, lack of a significant impact on intercropping suggests that the impact pathway from program participation to agronomic and economic gains is unclear. Further analysis is necessary to shed light on which aspects of the program are more impactful using other data collected from beneficiaries between baseline (2013) and follow-up (2019). Second, while the positive impacts of program participation presented here are encouraging, disentangling the contribution of different innovations to the overall impact of program participation is necessary to inform targeted interventions. This objective requires that an adequate number of households adopt the various SI innovations being offered and monitoring data about the type and timing of innovations received are routinely collected. This study relied on self-reported retrospective program participation data collected at follow-up that might be affected by recall bias. Indeed, data show some classification mismatches between program participation status—defined in project monitoring data at the time of the baseline (2013) and program participations status self-reported during follow-up (2019)—that should be addressed and resolved when trying to rigorously estimate program impact.

References

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).

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 [rice.” *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* (October): 2–16.

FAO. 2013. “Malawi: BEFS Country Brief.” Available at: https://www.fao.org/3/aq168e/aq168e.pdf.

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. Spielman. 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. 2015. "Malawi Africa Research in Sustainable Intensification for the Next Generation (Africa RISING) Baseline Evaluation Survey."

IFPRI, IITA, and ILRI. 2016. "Africa Research in Sustainable Intensification for the Next Generation Program proposal for a second phase, 2016–2021."

IITA. 2019. "Africa RISING in Malawi: Creating Sustainable Systems for Agriculture."

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

IMF. 2017. "Malawi Economic Development Document."

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.

Matita, M., E.W. Chirwa, D. Johnston, J. Mazalale, R. Smith, and H. Walls. 2021. “Does Household Participation in Food Markets Increase Dietary Diversity? Evidence from Rural Malawi.” *Global Food Security* 28 (January 2020): 100486. Available at: https://doi.org/10.1016/j.gfs.2020.100486.

Mazunda, J., and K. Droppelmann. 2012. “Maize Consumption Estimation and Dietary Diversity Assessment Methods in Malawi.” Malawi Stratagy Support Program (MaSSP) Policy Note 11. IFPRI.

Mungai, L.M., J.P. Messina, and S. Snapp. 2020. “Spatial Pattern of Agricultural Productivity Trends in Malawi.” *Sustainability (Switzerland)* 12 (4): 1–22.

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.

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. 2016. “Measuring Sustainable Intensification in Smallholder Agroecosystems: A Review.” *Global Food Security*.

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 TrialDesign.” In *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

Stuart, E.A., H.A.Huskamp, K. Duckworth, J. Simmons, Z. Song, M.E. Chernew, and C.L. Barry. 2014. “Using Propensity Scores in Difference-in-Differences Models to Estimate the Effects of a Policy Change.” *Health Services and Outcomes Research Methodology* 14: 166–182.

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) - Malawi.”

World Bank. 2022. “Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population) - Malawi.”

World Bank Group, United Nations, and European Union. 2016. “Malawi Drought 2015-2016: Post-Disaster Needs Assessment.”

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.

Acknowledgments

This research was conducted as part of the Africa Research In Sustainable Intensification for the Next Generation (Africa RISING) Program, funded by United States Agency for International Development as part of the U.S. Government's Feed the Future Initiative [grant #AID-BFS-G-11-00002]. We thank Leulsegged Kasa for the contribution in data processing. The donor had no role in the study design, data analysis, or write-up. All errors and omissions remain ours.

Appendix

Table A1. Descriptive summary by survey round

	(1) 2013	(2) 2019	(3) 1 vs 2
DOMAIN: ENVIRONMENT			
Received agricultural advice (%)	54.1	48.2	***
Practiced crop rotation (%)	71.2	63.0	***
Left crop residues on the field (%)	84.3	43.9	***
Applied manure (%)	54.2	68.9	***
Maize-legume intercropping (%)	72.6	66.2	***
Inorganic fertilizer used (kg/ha)	109.3	130.3	***
DOMAIN: PRODUCTIVITY			
Area cultivated with maize (ha)	0.49	0.56	***
Yield of maize (ton/ha)	2.10	1.31	***
Cereal yield (ton/ha)	2.01	1.30	***
Area cultivated with legumes (ha)	0.38	0.23	***
Legume yield (ton/ha)	0.76	1.09	***
Yield of groundnut (ton/ha)	1.42	1.64	**
Yield of bean (ton/ha)	0.59	0.49	**
Yield of soybean (ton/ha)	0.87	0.94	
Yield of cowpea (ton/ha)	0.42	0.89	***
Grew cereals (%)	99.6	99.7	
Grew roots and tubers (%)	11.2	15.3	***
Grew legumes and nuts (%)	88.2	88.4	
Grew vegetables (%)	0.20	4.98	***
Grew fruits (%)	0.61	8.04	***
Grew cash crops (%)	13.2	7.22	***
Real value of total harvest ('000 MWK)	163.9	66.8	***
Unique crops produced (#)	4.51	4.45	
Unique livestock types owned (#)	1.33	1.36	
Tropical livestock units	0.32	0.31	
Agricultural labor (person-days/ha)	321.2	143.6	***
Labor productivity (MWK/person-day)	876.0	589.7	***

	(1) 2013	(2) 2019	(3) 1 vs 2
DOMAIN: ECONOMIC			
Per capita daily expenditure (PPP)	1.83	1.82	
Non-poor (>$1.9 PPP) (%)	35.4	28.2	***
Non-poor (agr. durable assets) (%)	36.8	30.3	***
Non-poor (non-agr. durable assets) (%)	33.2	34.3	
Real value of crop and livestock sold ('000 MWK)	37.3	33.1	
DOMAIN: HUMAN			
Household-level indicators			
Dietary diversity score	6.62	8.22	***
Food secure households (%)	24.4	39.7	***
Months of food shortage (June-Sep)	0.21	0.42	***
Months of food shortage (Jan-Mar)	0.87	1.08	***
Child-level indicators			
HAZ	-1.70	-1.63	
Prevalence of stunting (%)	41.0	37	
WHZ	0.24	0.34	
Prevalence of wasting (%)	3.19	1.01	**
WAZ	-0.81	-0.71	
Prevalence of underweight (%)	12.6	11.1	
Woman-level indicators			
Body Mass Index (BMI	22.6	23.3	***
Prevalence of undernutrition (BMI<18.5)	6.75	3.83	***
DOMAIN: SOCIAL			
Plots managed by females (%)	85.4	49.2	***
Livestock managed by females (%)	4.74	36.9	***
Observations	983	983	1966

Note: Columns 1 and 2 report means of indicators at baseline and follow-up, respectively. Column 3 shows statistical significance levels from pair-wise tests of equality of means. * p<0.1, ** p<0.05, *** p<0.1

Table A2. Impact on the likelihood of production of different food groups (2013–2019)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Within-village comparison		Out-of-village comparison		Overall comparison		Placebo test	
Grew cereals (%)	1.15	1.11	0.94	0.96	1.01	1.02	-0.66	-0.66
Grew roots and tubers (%)	-1.83	-2.35	-11.03**	-15.66**	-7.36*	-9.81**	-14.91***	-16.01***
Grew legumes and nuts (%)	3.84	4.12	4.25*	3.22	2.78	3.39	0.06	-2.06
Grew vegetables (%)	2.64	2.53	2.75	2.90	2.96	2.95*	-2.02	-1.82
Grew fruits (%)	-3.16	-2.64	-1.25	-0.89	-2.18	-2.14	5.68	5.52
Grew cash crops (%)	1.15	1.11	0.94	0.96	1.01	1.02	-0.66	-0.66

Note: Difference-in-differences (DiD) impact estimates reported are based on Equation 1 (δ). 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), and (5) are from simple DiD. while those in columns (2), (4), and (6) are from DiD with regression adjustment where we control for household socio-demographic characteristics (household size. age and gender of the household head. average education of adults). NDVI measures (historical average. as well as average and coefficient of variation for the cropping season corresponding to the baseline and follow-up surveys). and an index for travel time to different services. 0.01 - ***; .05 - **; .1 - *