

Pre-analysis plan of the impacts of the project “Promoting coherence between disaster risk reduction, climate action and social protection in Malawi”

Silvio Daidone and Nicholas Sitko

Food and Agriculture Organization of the United Nations (FAO)

March 21, 2023

1. Background

The project “Promoting coherence between disaster risk reduction, climate action and social protection in Sub-Saharan Africa (Malawi)” aims to support poor and vulnerable households to strengthen their resilience to climate change and climate variability through social protection (SP) and the adoption of proven Climate Smart Agriculture (CSA) practices blended with Disaster Risk Reduction (DRR). FAO Malawi leads the implementation of the project in two targeted districts of Mwanza and Neno, targeting 2,400 farmers, some of them being beneficiaries of existing SP programs. At community level, the project is implemented through the Farmer Field School (FFS) approach and delivered through 80 FFS groups located in 74 villages.¹

The standard package of support provided by the project includes the following elements: CSA training, farming as a business training, and an enterprise grant disbursed collectively to each of the 80 FFS group. To this standard packet, the project randomly provided some combination of the following forms of support to the beneficiaries:

- 1) In-kind or voucher for CSA related inputs comprising maize and pigeon peas seeds, basal and top-dressing fertilizers, and inoculants for pigeon pea seeds.
- 2) Cash voucher equivalent in value to CSA input packet.

2. Hypotheses tested

Hypothesis 1: Cash and inputs transfers will increase food security.

Social protection instruments such as cash and input transfers can have a positive effect on food and nutrition security, which may in turn enhance labor productivity. In the short term, beneficiaries have greater access to sufficient, safe, and nutritious food to meet dietary needs, which improves physical strength and stamina and reduces days of work lost.

Hypothesis 2: Cash transfers will increase income diversification.

Cash transfers induce diversification by enabling households to invest in income-generating activities that offer a higher return than subsistence agriculture.

¹ In 4 villages, multiple groups have been formed: 4 FFS in Donda (Section Neno Central, TA Ngozi), 2 FFS each in Chirombo (section Kalioni, TA Ngozi), Tulongkhondo (section Tulongkhondo, TA Kanduku) and Nyakoko (section Ligowe North, TA Dambe).

Hypothesis 3: Cash and input transfers will increase adoption of positive coping-strategies.

By helping to smooth consumption and/or income when shocks occur, transfers can mediate household sensitivity to adverse weather events, and reduce the likelihood that a household will turn to negative coping strategies, such as productive asset liquidation, and directly ease the credit and liquidity constraints faced by poor rural households, thus increasing their capacity to invest in productive farm assets

Hypothesis 4: Participation in FFS activities will increase: a) adoption of climate smart agricultural practices; b) access to climate information; c) adoption of positive coping-strategies.

FFS build farmers' capacity and promote adoption of better agroforestry and climate smart agricultural practices, and consequently improve farmers' lives in terms of agricultural outcomes, health, environment, and empowerment. The FFS process enables farmers to internalize the advantages of the improved agricultural practices through learning by doing and observation. FFS empower farmers cognitively by encouraging them to develop skills in problem-solving using "scientific" methods of analysis, and organizationally through group activities. Further, by providing information concerning climate change and weather shocks, FFS promote the adoption of positive disaster risk-reduction strategies.

3. Study design

To evaluate impacts of the project, we will use a crossover design to compare the relative merits of its different components, and combine various evaluation methods.

The standard package of support provided by the project includes the following elements: CSA training, farming as a business training, and an enterprise grant for the field school groups.² To this standard packet, the project could randomly provide some combination of the following forms of support to the beneficiaries:

- 1) In-kind transfer for CSA related inputs comprising at least 10 Kgs of maize seeds, 50Kgs of NPK (basal) fertilizers, 50Kgs of Urea (top dressing) fertilizers, 4Kgs of Pigeon pea seeds and Nyonga pack (inoculants for pigeon pea seeds). The complete package per farmer is valued at 127.75 US dollars (USD).³

² The enterprise grant consists of a one-off transfer valued at 1 250 US dollars, disbursed collectively to each of the 80 FFS group, not to the single farmer, with the aim of fostering group formation and consolidation.

³ At an average exchange rate in November 2021 of 802 Malawi Kwacha (MWK) per USD, the complete package is equivalent to a subsidy of approximately 102 500 Kwacha.

- 2) Cash transfer equivalent in value to CSA input packet.

The FFS groups are selected randomly by the Evaluation Team into one of three groups described below.⁴ The groups are:

- T1: receives the standard packet of CSA training, farming as business training, and an enterprise grant for the FFS
- T2: Standard packet + voucher restricted to CSA input packet
- T3: Standard packet + cash voucher equivalent to the CSA packet

A fourth group (C) of farmers receive the same CSA packet offered to group T2, without being involved in the FFS. This fourth group comes from neighboring communities within the same districts and beneficiary farmers would be selected in the same way that farmers in the field schools are selected.

This design allows calculating several types of impacts at the project level in relation to the outcomes and outputs of interest:

- 1) The differential impact of the CSA input package (T2 vs. T1)
- 2) The differential impact of cash (T3 vs. T1)
- 3) The relative effectiveness of CSA input package vis-à-vis the cash (T3 vs. T2)
- 4) The impact of the FFS standard packet (T2 vs. C)

3.1. Comparison group selection

The Evaluation Team carried out the sampling of the comparison households in 25 villages. These villages were selected among those in Mwanza and Neno districts that are located farther away from the FFS. The identification of these villages has been carried out with the following procedure:

1. From a national shapefile of Malawi published in 2015, we extracted 427 villages located in Mwanza and Neno districts.⁵
2. From a mapping exercise conducted in early 2021 by FAO Malawi in Mwanza and Neno for the purposes of project targeting, we extracted 212 villages not selected for the FFS.

⁴ Originally, it was foreseen to organize a public event and a lottery to randomly allocate the benefits provided by the project. However, given the existing difficulties related to travel because of the outbreak of covid-19, and especially to organize an open event with public authorities and the civil society, the randomization procedure was managed by the Evaluation Team with a statistical software.

⁵ <http://landscapeportal.org/layers/geonode:villagesgeo>

3. We match the two sources based on the following string variables: village names, district, traditional authority (TA), section (156 matched villages).
4. Based on available GPS coordinates, we constructed distances between each matched village and the location of the 74 villages where the 80 FFS are located.
5. We created a ranking of villages, based on the maximum minimum distances between the pot of potential comparison villages and the FFS.

In Figure 1 and Figure 2 we show respectively the project districts map and the location of the comparison villages and of those hosting FFS. In each village, we randomly select at least 22 farmers for at least 530 farmers. Each sample from each village should have similar characteristics to the treatment groups.

The sampling of farmers in the comparison villages tried to mimic the FMM project targeting. Hence, we selected a group of farmers in each village according to the criteria listed below.

1. Both women and men being equally eligible to be sampled. Women and men to be selected should be within the active age group of 18 to 69 and able to engage in agricultural production activities.
2. At least 33% of the sample being composed of young people engaged in farming activities.
3. Potential farmers should be smallholders, with some land available for crop production.
4. The farmers meeting the above requirements and targeted by different social protection programmes, such as the Affordable Inputs Programme (AIP), the Social Cash Transfer Programme (SCTP), the Malawi Social Action Fund (MASAF) Public Works Programme and/or micro-credit programmes, were prioritized in the sampling process.

These criteria were verified with the village chief or the elderly committees. Further, after identification of potential interviewees, enumerators approached them, asking the following questions:

1. “Are you interested in innovating your agricultural practices to improve your productivity?”
2. “Are you interested in agroforestry or soil and water conservation fertility enhancement practices?”

If the answers were positive for both questions, the farmer was confirmed for the sample. If the farmer answered no to at least one of the above two questions or could not be reached by phone, another farmer was randomly selected. When 22 farmers were reached and answered positively

to the two questions, the sampling for the village was finished and enumerators proceeded with interviews. Similarly to the case of treatment villages, refusal/non-compliance with the survey led to farmers' replacement.

The decision to restrict the comparison villages within the Mwanza and Neno districts was driven by the pressure to conduct the fieldwork before the end of June, at a time of the year when the COVID-19 pandemic curve was increasing and travel restrictions within Malawi seemed a realistic threat to the rollout of the survey. In fact, while the permissions to conduct the fieldwork were already obtained from the Mwanza and Neno authorities, the necessary administrative steps were not undertaken in the neighbor districts, which could have also represented an equally suitable alternative for the counterfactual. We must remark that despite being located within the same target district and at most 80km distance from the FFS, comparison villages come from Lower Neno, in proximity to the Shire river and to a tarmac road leading to Blantyre. Geographic and agro-ecological conditions in this area are slightly different relative to upper Neno and Mwanza. It is therefore important to take into account these considerations while comparing the experimental groups with the comparison villages.

3.2. Data collection

Quantitative data are collected in two rounds: baseline and 24-months follow-up.

The baseline data collection occurred from 17 June to 14 July 2021. The fieldwork was carried out by 2 supervisors and 19 enumerators, all led by one national consultant specialized in survey data collection. The original goal set by the Evaluation Team was to collect 1,856 interviews, distributed as follows:

- Approximately 442 interviews for each of the three experimental groups (1,326 overall), divided in 26 equally sized clusters of 17 farmers each. However shortly after the establishment of the FFS, it appeared that multiple FFS were formed in two target villages. This led to a reduction from 78 to 74 clusters available for the experimental study and the decision to interview 18 farmers per cluster.
- For the non-experimental C group, it was planned to have at least 530 interviews, divided in 25 equally sized groups.

The baseline data collection was necessary to capture the baseline living conditions of the FFS participants and the comparison group before the start of the FFS activities and before any cash and CSA inputs transfers had been disbursed. In addition, these data provide a detailed description of beneficiaries and allow the evaluation team to assess whether any systematic

differences between the treatment and the comparison groups exist at baseline so that the differences can be controlled for during the impact evaluation analysis. It is crucial that the baseline data be collected before the three experimental groups received support payments and before the comparison group received the CSA inputs voucher. Because some short-term indicators, such as land management practices and inputs use will be impacted by the project soon after households start FFS activities and eventually receive cash transfers or CSA input transfers, we conducted the baseline survey before these impacts occurred to ensure that we accurately measure the full impacts of the programme.

Identification of households in FFS villages was a straightforward process, facilitated by the targeting activity carried out in the previous months to form the farmer field schools. However, the sampling of households had to balance gender and age composition of farmers participating in the FFS activities, following the existing registry data. For this reason, and combined with some tracking problems of FFS households, in a small share of FFS clusters it was not possible to reach the target of 18 households. This led to slightly unequal cluster size. Identification of farmers and their households in comparison villages was instead supported and monitored by village heads, councilors, and other key community leaders.

As shown in Table 1, the team successfully gathered the expected data from the two districts, collecting surveys from 1,886 households in 99 clusters, representing 8,961 individuals. Comparison households come exclusively from villages located in TA Symon and one residual cluster in TA Mluli. Because of the above-mentioned issue concerning the gender-age composition of participating FFS farmers, the clusters randomization procedure conducted after the data collection in the FFS villages boiled down in unequal cluster size and a slight oversampling of 457 households in the FFS+inputs group, followed by 440 and 436 households in the FFS+cash group and FFS only group, respectively.

3.3. Survey instruments

Two survey instruments are used in the impact evaluation of project: household questionnaire and community questionnaire. The design of the household instrument was guided by three core principles:

- The instrument must contain the key list of indicators presented in the project's log frame that will allow the programme to be assessed against its stated objectives. These core indicators include land management and agricultural practices, agroforestry, crop production, climate shocks, climate information and coping strategies, food security, access

to markets and savings, although the final instrument contains many more relevant indicators.

- Where possible, indicators are measured using the questions and approaches that have already been field tested in similar surveys in the country, thus ensuring that they are appropriate for local conditions and that the resulting data can be compared with national data. We followed two main household surveys currently available to researchers: 1) The Fifth Integrated Household Survey (IHS) 2019-2020, carried out by the Malawi National Statistical Office in collaboration with the Living Standard Measurement Survey group at the World Bank. 2) The impact evaluation data of the Malawi National Social Cash Transfer Program 2013-2014, collected by the University of Malawi in collaboration with the University of North Carolina at Chapel Hill
- The survey instrument must be a manageable length to avoid interviewer or respondent fatigue. Table 2 provides a list of topics covered in each of the two instruments

4. Randomization

Randomization is a critical step for ensuring exogeneity in experimental methods, in which all eligible units in a sample are randomly assigned to the treatment arms. Randomization ensures that the treatment groups are equal in both observed and unobserved characteristics, thus ruling out selection bias. The only difference between the treatment groups, then, is their participation in the intervention itself, and the difference in their outcomes thus represents the intervention's impact.

While typically randomization is carried out through a public lottery, this option was not considered feasible for this study given the current situation with the covid-19 pandemic. The Evaluation Team has been therefore assigned to conduct the randomization procedure to support project implementation once the baseline data collection had finished. In this way, neither the enumerators nor the respondents knew which group project beneficiaries would be included. Researchers in the evaluation team decided to carry out the randomization with Stata, because of the transparency and reproducibility of the process. To randomize with replicability in Stata, we followed the simple guidelines provided by Development Impact Evaluation (DIME) at World Bank:⁶

1. Making sure the baseline dataset includes a unique cluster ID

⁶ See https://dimewiki.worldbank.org/Randomization_in_Stata

2. While writing a do-file, pay close attention to the following:
 - a. Setting version: this ensures that the randomization algorithm is the same, since the randomization algorithm sometimes changes between Stata versions
 - b. Setting the seed: this ensures that the same random number is generated for all observations, every time the code is run.
 - c. Sorting the data by the unique ID: the data should be sorted such that observations are in the same order every time the code is run.
3. Converting the random numbers into categorical variables for treatment status.

In Table 3 we report the Stata codes used to randomize the treatment allocation for the three experimental groups.

5. Main outcomes

In this section we describe a list of indicators that corresponds to the goal and outcomes included in the project's log frame. This will allow to assess the programme against its main stated objectives, namely: income diversification, crop diversification, climate-adaptive agricultural practices, food security, climate information and coping strategies. Many of these outcomes are multidimensional concepts and many indicators can be used to measure them. For this reason, we decided to consider for the main outcome variables a summary index approach, which facilitates generalized findings about the program's effectiveness.

We calculate summary indexes by adopting the standardized weighted mean approach of Anderson (2008), where we use the comparison group as the default reference group for the standardization.⁷ These standardized summary indexes à la Anderson do not have a specific meaning as they merely reflect deviations from the comparison group and can be thus interpreted as effect sizes.

To summarize, we compute the following index variables:

5.1. Income diversification index

The income diversification index is a standardized weighted average of the (positively coded) number of income sources, a Simpson index of income concentration (Simpson, 1949) and a Shannon index of income diversity (Shannon, 1948). To calculate income diversification index, we consider the following labor and non-labor income sources: crop production, vegetables

⁷ See Schwab et al. (2020) for a detailed step-by-step guide to construct such summary indexes à la Anderson.

production, fruits production, livestock production, livestock-by-products, non-farm businesses, sales of forest products, private transfers, which includes both remittances from abroad and within the community/village, public transfers, off-farm wage income. The general formula of a Simpson index is the following:

$$Simpson = 1 - \sum_{i=1}^I sh_i^2 \quad (1)$$

where sh_i is the income share of source i , calculated over total household gross income. The Simpson index ranges between zero and one; a value of zero implies that the household relies only on one income source while a value closer to one reflects a more even distribution of income by source.

The general formula of a Shannon index is instead:

$$Shannon = - \sum_{i=1}^I sh_i \log(sh_i) \quad (2)$$

Where again sh_i is the income share of source i . Values for the Shannon index can range from zero to the value of the log of the highest number of income sources of the household. The Shannon index ranges from 0, which flags households relying on one income source only, to a maximum of $\log I$ (when all shares equal $1/n$).

5.2. Crop diversification index

The crop diversification index is a standardized weighted average of the (positively coded) number of crops planted, a Simpson index of farmland distribution and a Shannon index of farmland diversity by crop. For each of these three indicators, we consider the following crops: maize, pigeon peas, beans, groundnuts, Irish potatoes, cow peas, sweet potatoes, cassava, sorghum, sugar cane, peas, cotton, other cereals, other legumes, and a residual category of other crops. In the Simpson and Shannon index of crop diversification we look at the shares of cultivated land with the above mentioned crops.

5.3. Agricultural practices index

The agricultural practices index is a standardized weighted average of the following indicators: share of farmland under crop rotation, share of farmland with crop residue used to cover land, share of farmland where zero/minimum tillage is practiced, share of farmland where two or more crops have been intercropped in the same plot, dummy variables for whether any water conservation structure has been applied, trees have been planted on farmland, land has been

irrigated, manure has been used, land was left fallow for more than one year in the last five years and finally avoided use of pesticides.

5.4. Food security index

The food security index is a standardized weighted average of the (negatively coded) indicators for whether during the last 12 months there was a time when, because of lack of money or resources, the main respondent: was worried about not having enough food to eat, was unable to eat healthy and nutritious food, ate only a few kinds of foods, had to skip a meal, ate less than they thought they should, thought their household ran out of food, was hungry but did not eat, went without eating for a whole day; and the (positively coded) number of meals, including breakfast, eaten per day by the household members, the number of months the maize from the previous year harvest lasted, the number of months the maize currently in the granary is expected to last.

5.5. Climate information index

The climate information index is a standardized weighted average of the (positively coded) indicators for whether the household received information on sudden catastrophes, slow-onset disasters, pest/disease outbreak, timing of rains, weather forecasts for the next three days and weather forecasts for the next three months.

5.6. Coping strategies index

The coping strategies index is a standardized weighted average of the (negatively coded) indicators for household level ex-post adaptation mechanisms (after the occurrence of climatic shock): children's migration, changes in food consumption habits (relying on less preferred food, reducing the proportion or number of meals per day, etc.), reduction of health and/or education expenditures, and sale of household assets, and the (positively coded) indicators for the following farm-level adaptation strategies after forecast of weather shocks: change in cropping pattern, improved seeds adoption, change in sowing date, increased use of organic compost, increased use of chemical fertilizers, investment in irrigation, greater crops diversification, weather insurance.

6. Estimation strategy

To evaluate the differential impact of the CSA input transfer (T2 vs. T1), the differential impact of the cash transfer (T3 vs. T1) and the relative effectiveness of CSA input package vis-à-vis the cash (T3 vs. T2) we will take advantage of the randomized experimental design and conduct

an intent-to-treat (ITT) analysis, avoiding the bias that may occur due to selection into and out of the project. We estimate the treatment effect using Analysis of Covariance (ANCOVA), which controls for the lagged outcome variable. ANCOVA estimates are preferred to difference-in-difference estimates when the autocorrelation of outcomes is low (McKenzie, 2012). ANCOVA estimates will adjust for baseline imbalances according to the degree of correlation between baseline and follow-up and lead to a more efficient estimation of impact.⁸ Therefore, we will estimate the following ANCOVA model:

$$Y_{ij1} = \alpha_{TA} + \beta_1 CT_j + \beta_2 IT_j + \beta_3 Y_{ij0} + \varepsilon_{ij} \quad (3)$$

Where Y_{ij1} and Y_{ij0} are the summary outcome index for farmer/household i from cluster j at the follow-up and at the baseline, respectively. CT_j and IT_j are two binary indicators that equal 1 if cluster j is in the FFS+cash transfer treatment arm and in the FFS+input transfer treatment arm respectively (0 otherwise). Traditional authority-level fixed effects are captured by α_{TA} . β_1 and β_2 represent the ITT estimators, or the effect of being assigned to the specific treatment arm. In all regressions we adjust standard errors for clustering at the cluster which was the level of randomization. To test whether the estimators are statistically different by treatment arm, we conduct tests of equality and report the p-values.⁹ Further, we will conduct robustness checks, by carrying out complementary estimations using a single difference at the follow-up (without including the lagged outcome) or a double difference estimator.

To evaluate the impact of the FFS standard packet (T2 vs. C), we will rely on a double difference approach combined with inverse probability weighting. Inverse-probability-weighted regression adjustment (IPWRA) estimators use weighted regression coefficients to compute averages of treatment-level predicted outcomes, where the weights are the estimated inverse probabilities of treatment. The contrasts of these averages estimate the treatment effects. This method possess the property of being doubly robust (Cattaneo, 2010; Hirano & Imbens, 2001; Wooldridge, 2007). The double robustness for the proposed estimation method implies that if the weights are estimated based on a correct probit/logit specification and/or if the conditional mean of $(Y_{i1} - Y_{i0})$ are correctly specified, the resulting estimator will be consistent. To estimate the treatment model (T2 vs. C groups), we will use LASSO to determine the set of

⁸ For our key variables of interest, we expect moderate autocorrelations between baseline and follow-up, i.e. below 0.5.

⁹ Preliminary baseline data analysis shows the random assignment has been extremely successful in creating equivalent groups in basically all socio-economic characteristics. For this reason, the inclusion of baseline controls is not necessary to obtain unbiased estimates of our parameters of interest.

variables explaining the treatment (Belloni et al., 2014), while for the outcome equations we will re-run the LASSO procedure for each primary outcome to get a separate set of controls. A preliminary list of control variables is given in appendix (Table 4), though we will expand it with other spatial/biophysical factors, such as rainfall precipitations and temperatures, which might explain our treatment and outcome equations. We will conduct robustness checks for our main estimates, by revising the LASSO method used to select the indicators for both the treatment and outcome equations, and by showing results using a simple difference-in-difference approach with control variables but without reweighting.

7. Power and attrition

At the baseline we have interviewed 1,886 households, with 553 comparison households in 25 clusters and 1,380 households in 74 clusters. At the conventional level of significance of 0.05 and a power of 0.8, our sample size would allow for a minimum detectable effect (MDE) between 0.244 and 0.379 standard deviations in the experimental arms for the six main outcomes discussed in section 5. Obviously, at the follow-up it will be unlikely to reach all the respondents initially sampled. However, we expect extremely low levels of attrition in the three experimental arms since households have been followed constantly during the FFS activities. Accounting for either a 5 or 10% attrition, the MDEs will only slightly increase and will be comprised between 0.247 and 0.384 standard deviations (Table 5). We will check whether non-response is correlated with the random assignment. If there is a statistically significant difference in non-response between the FFS only group and the other FFS+ groups, we will follow the procedure proposed by Kling et al. (2007). We will obtain lower bounds of the treatment effect by replacing missing observations in the treatment (control) arms by the corresponding arm's mean value minus (plus) 0.05, 0.10 and 0.20 standard deviations of the control group. Upper bounds of the treatment effects will be constructed in a symmetrical way.

While it is relatively straightforward carrying-out power calculations for randomized control trials, doing power calculations for quasi- and non-experimental impact evaluations requires more judgement. While for IPWRA estimators we do not have specific guidance to follow, recently Hu & Hoover (2018) studied power / sample size estimation methods for non-randomized DiD designs. Further, Schochet (2022) developed new closed-form variance expressions for power analyses for commonly used DID panel data estimators, with formulas also accounting for other key design features that arise in practice, such as autocorrelated errors,

unequal measurement intervals, and clustering due to the unit of treatment assignment.¹⁰ Besides the cluster size, the number of clusters and the ICC coefficient, other parameters must be considered in a DiD design. For the FFS evaluation, we rule out staggered timing (so we have one treatment group occurring after the baseline), while we assume autocorrelation follows an AR(1) process with autoregressive parameter equal to 0.4. While the power literature in short panels assumes constant autocorrelations for pooled estimators (Frison & Pocock, 1992; McKenzie, 2012), here correlations are larger for cluster observations closer in time than further apart. We do not have a precise idea concerning this parameter for the main outcome indicators considered for this study. McKenzie (2012) provides a useful indication for a number of similar economic outcomes, though in different countries and contexts. Education outcomes such as math and language test scores tend to have autocorrelation coefficients above 0.5 and 0.6 even when the time interval of the measurement is 1 or 2 years. Instead, income and expenditure measured at 6-months intervals have lower coefficients, and vary between 0.1 and 0.4. In the absence of a precise reference, we carried out the power calculations assuming an autocorrelation parameter equal to 0.2, 0.4 and 0.6 (Table 6). Clearly, the MDE for the six outcome indicators increases considerably when the autocorrelation parameter is low and it's the highest for the food security summary index, which is the indicator with the largest intra-cluster coefficient.

Finally, in this study we will also adjust the p-values for the fact that we are testing the impact on six outcomes. We will calculate q-values using the Benjamini-Hochberg step-up method, which minimizes the false discovery rate (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001). The false discovery rate method entails that the M p-values of the i hypotheses are ordered from low to high and that the critical value of the p-value is then $p(i) = \alpha * i / M$. Therefore, with 6 outcomes and hypotheses and a significance level (α) of 0.05, the critical p-value would be 0.0083 for the one with the lowest p-value ($0.05 * 1/6$), which coincides with the most restrictive Bonferroni correction. For the second hypothesis, the critical p-value is 0.01666667 ($0.05 * 2/6$) and for the seventh it is 0.05 ($0.05 * 6/6$). Thus, correcting for the false discovery rate increases the MDEs, which will be comprised between 0.304 and 0.472 standard deviations in the experimental evaluation.

Obviously, we will also look at each variable composing the outcome measures, to open the “black box” represented by the summary indices. The main advantage for using the Benjamini-Hochberg procedure is that it has more power to detect real differences with the same

¹⁰ A Shiny R dashboard performs the sample size calculations for the estimators considered by Schochet (2022).

uncorrected p-value, especially if the number of measured parameters is large. Further, it is less conservative as it allows for correlation across test statistics, while other methods such as Bonferroni are based on the assumption of independence. This is unlikely to be the case, especially within the summary indices “family”.

References

- Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481–1495.
<https://doi.org/10.1198/016214508000000841>
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2), 608–650. <https://doi.org/10.1093/restud/rdt044>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4).
<https://doi.org/10.1214/aos/1013699998>
- Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, 155(2), 138–154.
<https://doi.org/10.1016/j.jeconom.2009.09.023>
- Frison, L., & Pocock, S. J. (1992). Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in Medicine*, 11(13), 1685–1704. <https://doi.org/10.1002/sim.4780111304>
- Hirano, K., & Imbens, G. W. (2001). Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization. *Health Services and Outcomes Research Methodology*, 2(3/4), 259–278.
<https://doi.org/10.1023/A:1020371312283>

- Hu, Y., & Hoover, D. (2018). Simple Power and Sample Size Estimation for Non-Randomized Longitudinal Difference in Differences Studies. *Journal of Biometrics and Biostatistics*, 9(5), 415.
- Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1), 83–119. <https://doi.org/10.1111/j.1468-0262.2007.00733.x>
- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2), 210–221. <https://doi.org/10.1016/j.jdeveco.2012.01.002>
- Schochet, P. Z. (2022). Statistical Power for Estimating Treatment Effects Using Difference-in-Differences and Comparative Interrupted Time Series Estimators With Variation in Treatment Timing. *Journal of Educational and Behavioral Statistics*, 47(4), 367–405. <https://doi.org/10.3102/10769986211070625>
- Schwab, B., Janzen, S., Magnan, N. P., & Thompson, W. M. (2020). Constructing a summary index using the standardized inverse-covariance weighted average of indicators. *The Stata Journal: Promoting Communications on Statistics and Stata*, 20(4), 952–964. <https://doi.org/10.1177/1536867X20976325>
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Simpson, E. H. (1949). Measurement of Diversity. *Nature*, 163(4148), 688–688. <https://doi.org/10.1038/163688a0>
- Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141(2), 1281–1301. <https://doi.org/10.1016/j.jeconom.2007.02.002>

Appendix

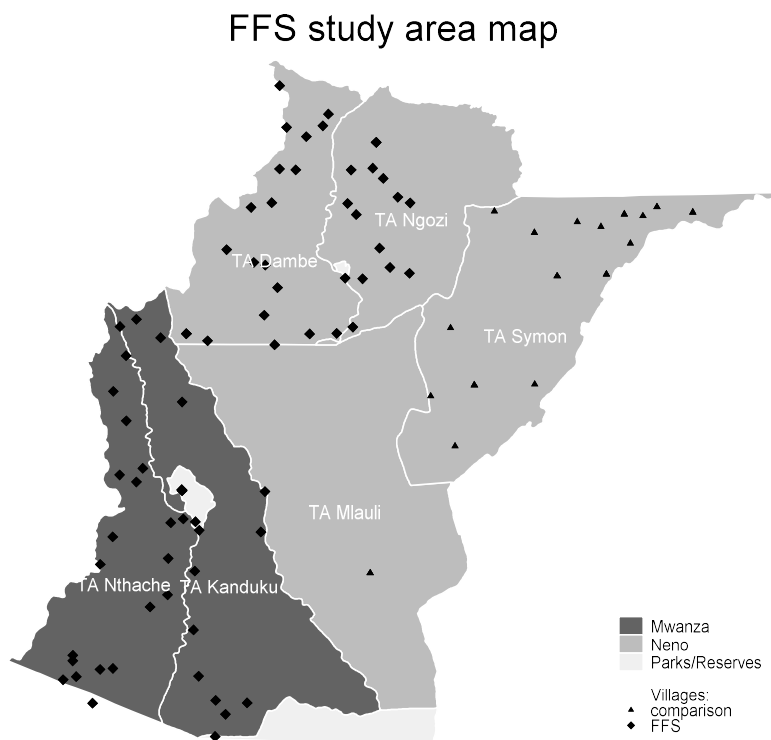
Figure 1: Project districts map



Note: Authors' elaboration from Malawi subnational administrative boundaries data.

Source: National Statistics Office of Malawi. 2020. Malawi - Subnational Administrative Boundaries. <https://data.humdata.org/dataset/cod-ab-mwi>

Figure 2: Location of FFS and comparison villages



Note: Authors' elaboration from FAO monitoring and spatial data. FFS = Farmer Field Schools

Source: World Agroforestry Centre. 2015. Malawi villages.

<http://landscapesportal.org/layers/geonode:villagesgeo>

Table 1: Sample size of households, individuals, and communities, by treatment group and Traditional Authority

	Chekucheku	Dambe	Govati	Kanduku	Nthache	Mlauli	Symon	total
comparison						22	531	553
						<i>103</i>	<i>2,474</i>	<i>2,577</i>
						(1)	(24)	(25)
FFS only	133	87	36	108	72			436
	<i>610</i>	<i>404</i>	<i>183</i>	<i>529</i>	<i>324</i>			<i>2,050</i>
	(7)	(5)	(2)	(6)	(4)			(24)
FFS+inputs	35	182	36	126	78			457
	<i>144</i>	<i>899</i>	<i>176</i>	<i>577</i>	<i>400</i>			<i>2,196</i>
	(2)	(10)	(2)	(7)	(4)			(25)
FFS+cash	68	125	54	90	103			440
	<i>335</i>	<i>608</i>	<i>257</i>	<i>438</i>	<i>500</i>			<i>2,138</i>
	(4)	(7)	(3)	(5)	(6)			(25)
total	236	394	126	324	253	22	531	1,886
	<i>1,089</i>	<i>1,911</i>	<i>616</i>	<i>1,544</i>	<i>1,224</i>	<i>103</i>	<i>2,474</i>	<i>8,961</i>
	(13)	(22)	(7)	(18)	(14)	(1)	(24)	(99)

Note: Authors' elaboration from survey data. Figures in italic (grey shaded rows) represent individuals. Number of clusters and community survey interviews in parentheses.

Table 2: Topics in the survey questionnaires

<u>Household survey</u>	<u>Community survey</u>
Roster, time use and wage labor	Community roster
Land	Agricultural inputs price
Crop production	Wages
Fruits and vegetables	Distances
Livestock holding	
Livestock by-products	
Agricultural assets	
Agricultural inputs expenses	
Non-farm enterprises	
Hired labor	
Non-timber forest products	
Transfers	
Decision-making	
Access to information	
Climate	
COVID-19	
Food insecurity	
Housing	
Credit	

Table 3: Stata code for randomization of the three experimental groups

```
/* [SETS STATA VERSION] */
version 12
/* [SETS THE RANDOM SEED FOR REPLICATION, 265 is Malawi phone code] */
set seed 265
/* we randomly allocate the treatment only to FFS villages */
keep if ffs==1
duplicates drop cluster_id, force
isid cluster_id, sort
* Assign random numbers to the observations and rank them from the smallest to the
largest
/* [GENERATES A RANDOM NUMBER BETWEEN 0 AND 1] */
gen random_number = uniform()

egen ordering = rank(random_number)
count
local first = r(N)/3
local second = 2*(r(N)/3)
* Assign observations to the treatment groups based on their ranks
gen treatarm = .
/* [ASSIGNS TREATMENT STATUS TO FIRST THIRD OF SAMPLE] */
replace treatarm = 1 if ordering <= `first'
/* [ASSIGNS CONTROL STATUS TO SECOND THIRD OF SAMPLE] */
replace treatarm = 2 if ordering > `first' & ordering <= `second'
/* [ASSIGNS CONTROL STATUS TO LAST THIRD OF SAMPLE] */
replace treatarm = 3 if ordering > `second'
lab def arm 0 "comparison" 1 "FFS only" 2 "FFS+CSA packet voucher" 3 "FFS+cash",
modify
lab val treatarm arm
keep cluster_id treatarm
```

Table 4: List of control variables screened with LASSO

variable name	variable label
hh0_5	# hh members 0-5 yrs
hh6_12	# hh members 6-12 yrs
hh13_17	# hh members 13-17 yrs
hbm18_59	# male hh members 18-59 yrs
hhf18_59	# female hh members 18-59 yrs
hh60	# hh members 60+ yrs
femhd	female headed hh
agehd	head of hh age
mrrhd	head of hh married
wdwhd	head of hh widow
disable	# disabled hh members
edhd	head of hh yrs of education
edhigh	highest yrs of education in hh
prmedhd	head of hh completed primary school
dtotvhrv	hh harvesting crop
totvhrv	total value of harvest
dsalesveg	hh selling vegetables
salesveg	value of vegetable sales
dsalesfrt	hh selling fruits
salesfrt	value of fruits sales
dlvstcksls	hh selling livestock
lvstcksls	value of livestock sales, MKW
dsls_lvstckbyprod	hh selling livestock by-products
sls_lvstckbyprod	value of livestock by-products sales, MKW
dnfbus_rev	hh engaged in a non-farm business
nfbus_rev	non-farm business revenues last 12 months, MKW
dsls_forest	hh selling non-timber forest products
sls_forest	value of sales of non-timber forest products, MKW
dhhtotwage	hh with at least one member in wage employment
hhtotwage	hh total annual wage
dprivtransf	hh received private transfers
privtransf	value of private transfers received, MKW
dpubtransf	hh received public transfers
pubtransf	value of public transfers received, MKW
ppea	hh cultivates pigeon pea
bns	hh cultivates beans
grnd	hh cultivates groundnut
cowp	hh cultivates cow pea
srgh	hh cultivates sorghum
dlandfllw	hh left land fallow
dlandirr	hh irrigated land
residue_cover	hh uses crop residue to cover land
lprep_zerotill	hh prepares land with zero/minimum tillage

dfert_manure	hh used manure fertilizer
dnonpest	hh not using pesticides
rotation	hh adopts crop rotation
intercrop	hh practices intercropping
dtrees	hh planted trees on operated land
dtrees	hh planted trees on operated land
dwatercons	hh investing in water conservation structures
fies_raw	raw fies score
nmeals	number of meals take per day in the HH
nmonths_lasted	months maize harvest lasted in the last year (2019-2020)
nmonths_grainery	months will last maize currently in the grainery
info_suddencat	hh received info about sudden catastrophes
info_slowdis	hh received info about slow-onset disasters
info_pestout	hh received info about pest outbreak
info_rains	hh received info about rains
info_dweathfore	hh received info about weather forecasts in 2/3 days
info_mweathfore	hh received info about weather forecasts in 2/3 months
wshock_childmigr	because of weather shock hh sent children living elsewhere
wshock_foodhabit	because of weather shock hh changed food habits
wshock_healthedu	because of weather shock hh reduced health and education expenses
wshock_soldassets	because of weather shock hh sold assets
wfrcst_cropptrn	hh strategies due to weather events: change in cropping pattern
wfrcst_imprvseed	hh strategies due to weather events: improved seeds adoption
wfrcst_sowdate	hh strategies due to weather events: change in sowing date
wfrcst_orgcomp	hh strategies due to weather events: increased use of org compost
wfrcst_chfert	hh strategies due to weather events: increased use of chemical fertilizer
wfrcst_invirrig	hh strategies due to weather events: investment in irrigation
wfrcst_cropdiv	hh strategies due to weather events: greater diversification of crops
wfrcst_insure	hh strategies due to weather events: crop insurance
treesamp	hh acquired tree samplings
treefer	hh acquired fertilizer trees
treefod	hh acquired fodder trees
treefrt	hh acquired fruit trees
treefuel	hh acquired fuel wood trees
sp_treesamp	total expenditure on tree samplings
sp_treefer	total expenditure on fertilizer trees
sp_treefrt	total expenditure on fruit trees
sp_treefuel	total expenditure on fuel wood trees
ntrees	# trees planted on operated land
info_agro	hh received info on agro-forestry
assets_own	assets owned by the hh
handhoe	hh owns hand hoe
slasher	hh owns slasher
axe	hh owns axe
oxcart	hh owns ox cart

okplough	hh owns ox plough
mot_pump	hh owns generator or motorised pump
scotchcart	hh owns scotchcart
tractor	hh owns tractor
sprayer	hh owns sprayer
pangaknif	hh owns panga knife
sol_pump	hh owns micro-solar water pumps
sickle	hh owns sickle
tr_pump	hh owns treadle pump
watercan	hh owns watering can

Table 5: Minimum detectable effects for main outcome variables in the experimental evaluation

Summary index	Attrition rate (%)		
	0	5	10
income diversification	0.285	0.288	0.292
crop diversification	0.244	0.246	0.248
agricultural practices	0.250	0.252	0.254
food security	0.379	0.381	0.384
climate information	0.306	0.309	0.312
coping strategies	0.246	0.250	0.254

Notes: Calculations done using the baseline data collected for the project in June/July 2021. Significance level α and power of the test β equal 0.05 and 0.8. We consider 24 and 25 clusters for the treatment (FFS+) and control (FFS only) arms, with average cluster size of 18 households per cluster. Intra-cluster correlations calculated at the outcome level with the Stata command “loneway”. Power calculations carried out with Stata command “power twomeans”.

Table 6: Minimum detectable effects for main outcome variables in the non-experimental evaluation

Summary index	AR parameter		
	0.2	0.4	0.6
income diversification	0.367	0.333	0.290
crop diversification	0.429	0.383	0.333
agricultural practices	0.398	0.358	0.313
food security	0.452	0.403	0.347
climate information	0.372	0.339	0.300
coping strategies	0.318	0.285	0.267

Notes: Calculations done using the baseline data collected for the project in June/July 2021. Significance level α and power of the test β equal 0.05 and 0.8. We use 18 observations for each cluster, 25 clusters per treatment arm, no staggered treatment timing, 2 time periods, a longitudinal design with 1 pre- and 1 post-treatment periods, and intra-cluster correlation coefficient calculated with Stata loneway command. Power calculations carried out with R-shiny dashboard (Schochet, 2022)