Niger Irrigation and Market Access Project

Baseline Report – Part II: Dosso-Gaya Region

October 19, 2023

Christopher Ksoll, Anthony Louis D'Agostino, Margo Berends, Sam Studnitzer, Poorva Upadhyaya

Submitted to:

Millennium Challenge Corporation
1099 Fourteenth Street, NW, Suite 700
Washington, DC 20005
Project Officer: Hamissou Samari
Contract Number: 95332418F0225

Submitted by:

Mathematica
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Fax: (202) 863-1763
Project Director: Matt Sloan
Reference Number: 50542.01.110.031.000

Contents

Tables

Figures

Acknowledgments

This report was made possible through the contributions of many individuals.

We would like to thank the many individuals from MCC and MCA-N, as well as various members of government agencies in Niger who participated in the workshop to discuss the draft version of this report. Their thoughtful comments and feedback, which helped shape this final version, and our responses are in Appendix G. Special thanks to Hamissou Samari and Sarah Lane (MCC) for their guidance and support throughout the design of this evaluation. We also thank Ahmadou Issaka at MCA-N for coordinating the MCA-N response.

We also greatly appreciate the efforts of Mathematica staff who made helpful contributions: Jill Miller, Walter Bower, Dr. Evan Morier, and Matt Sloan. Finally, we are also grateful to our consultant Amadou Moussa Saidou, for his on-the-ground support in Niger, and to Moussa Diouf, who translated the French version of this report.

List of Acronyms

ASMD	Absolute standardized mean difference
BT	Basse Terrasse
CFI	Chronic food insecurity
EDR	Evaluation design report
ERR	Economic rate of return
FEWS NET	Famine Early Warning Systems Network
GCVI	Green chlorophyll vegetation index
GDP	Gross domestic product
IHS	Inverse hyperbolic sine
IMAP	Irrigation and Market Access Project
IPC	Integrated Food Security Phase Classification
IRB	Institutional review board
IWUA	Irrigation water user association
LER	Land equivalence ratio
MCA-N	Millennium Challenge Account-Niger
MCC	Millennium Challenge Corporation
MCG	Matched comparison group
NDVI	Normalized difference vegetation index
ONAHA	*l'Office Nationale des Aménagements Hydro-agricoles* (National Office for Irrigation Schemes)
PAP	Project affected person
PNFL	*Produits forestier non ligneux* (non-timber forest products)
PPP	Purchasing power parity
RAP	Resettlement Action Plan
RMSE	Root mean square error
SD	Standard deviation(s)
SDI	Société de Développement International
SSI	Small-scale Irrigation
WEAI	Women's Empowerment in Agriculture Index

I. Introduction

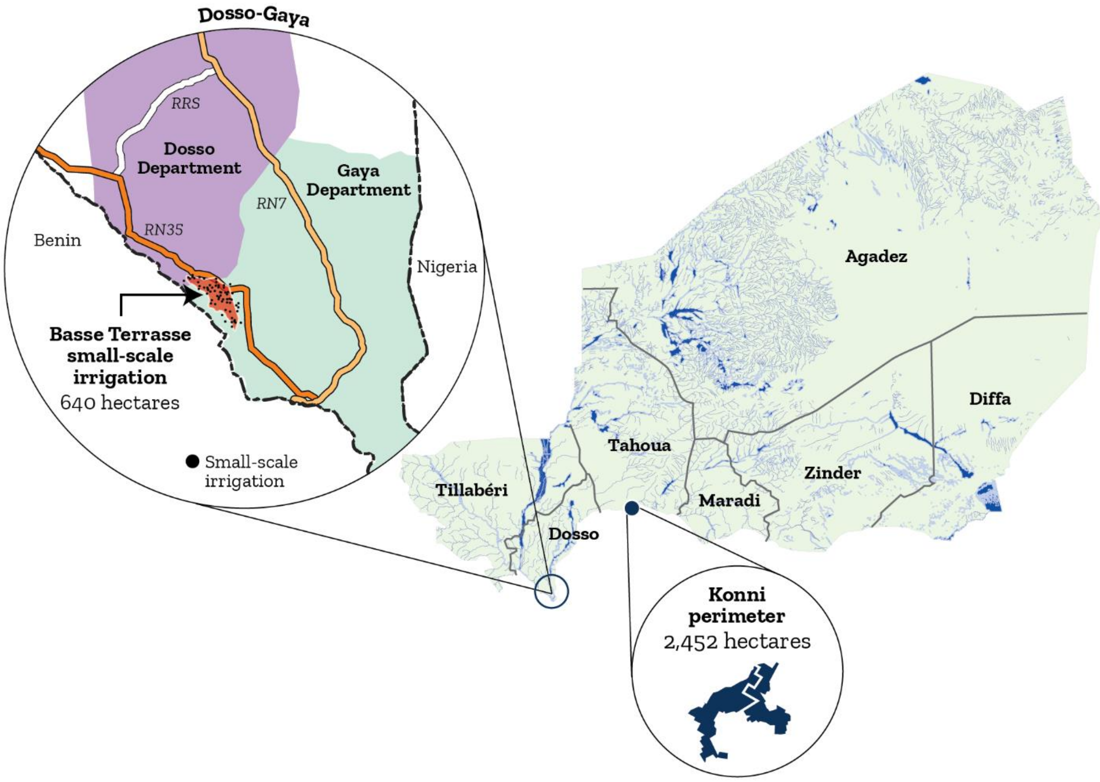
A. Overview of the IMAP

The agricultural sector in Niger is the source of the livelihoods of more than 80 percent of the population and contributes about one-fourth of the country's gross domestic product (CIA 2018). However, agricultural productivity in Niger is among the lowest in West Africa (FAO 2021a). The majority of Niger's agricultural production is rainfed; in 2011, irrigated farmland accounted for less than 1 percent of the total agricultural land in the country (FAO 2016). Without access to irrigation, crop production is vulnerable to droughts, which are frequent in Niger and can cause severe crop losses. Inadequate irrigation infrastructure also constrains production growth in the dry season (World Bank 2013). Productivity gains are further hampered by farmers' lack of market access to improved seeds, low adoption of new technologies, and inadequate extension services (World Bank 2017). There is also a substantial gender gap in yields: agricultural land managed by Nigerien women produces 19 percent less per hectare than land managed by men (Backiny-Yetna and McGee 2015). Low agricultural productivity has broader implications for human development in Niger. More than 1.2 million people in Niger experienced severe food insecurity in 2019, and another 1.5 million of the Nigerien population is chronically food insecure and unable to meet their food needs (WFP 2019). In 2019, Niger was ranked last globally on the United Nations Human Development Index (UN 2019).

To improve Niger's agricultural productivity and increase the incomes of rural farmers, the Millennium Challenge Corporation (MCC) is partnering with the Government of Niger through the $442.6 million Niger Sustainable Water and Agriculture Compact. The compact, which is being implemented between January 2018 through January 2024 , includes two projects. One is the Irrigation and Market Access Project (IMAP), which aims to increase rural incomes through improved agricultural productivity and increased agricultural sales resulting from modernized irrigated agriculture with sufficient trade and market access. IMAP's planned activities include constructing small-scale irrigation (SSI) infrastructure in the part of the Dosso-Gaya area known as the *Basse Terrasse* (BT), rehabilitating irrigation infrastructure in the Konni area, supporting institutions to increase land tenure security, training farmers and facilitating market access, promoting policy reform, implementing sustainable management of irrigation systems, and upgrading rural and national roads to connect the Dosso-Gaya area to the rest of the country and facilitate trade. The other project is Climate-Resilient Communities, which aims to improve agricultural and livestock productivity for livestock-dependent households, preserve natural resources, and increase market sales of targeted commodities. The compact is being implemented by the Millennium Challenge Account-Niger (MCA-N), and the United Nations Office for Project Services (UNOPS) is supporting MCA-N by providing technical services that support the compact's management.

In September 2017, MCC contracted with Mathematica to design and implement an evaluation of IMAP activities. Four overlapping activities make up the $256 million IMAP: (1) the Irrigation Perimeter Development Activity (IPD), (2) the Management Services and Market Facilitation Activity (MSMF), (3) the Roads for Market Access Activity (RMA), and (4) the Policy Reform Activity (PR). The first two activities are being implemented in two areas of Niger: the Konni area and the Dosso-Gaya project area, shown in **Figure I.1**. The Roads for Market Access Activity is taking place only in the Dosso-Gaya project area, and the PR is national.

Figure I.1. Map of project regions

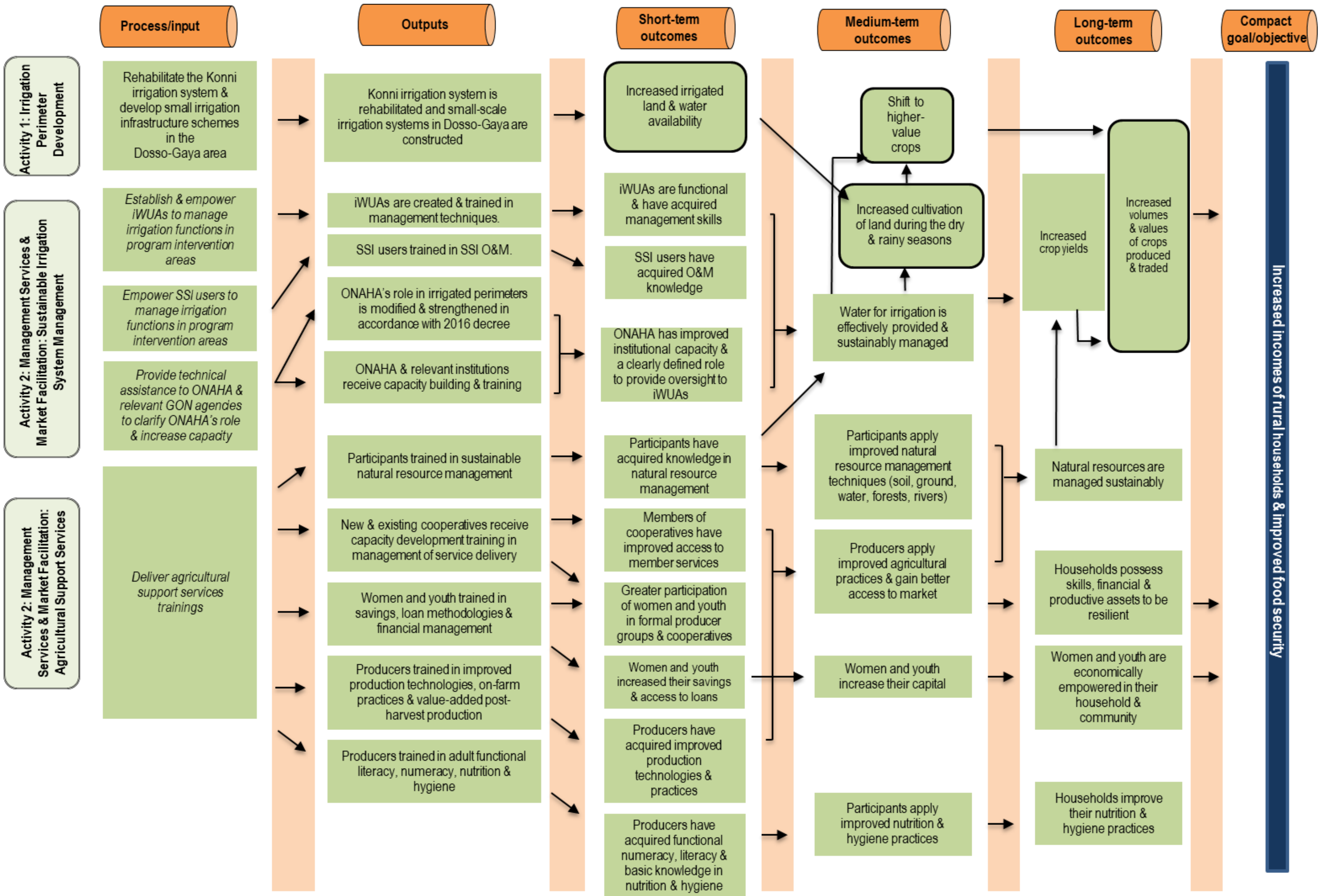


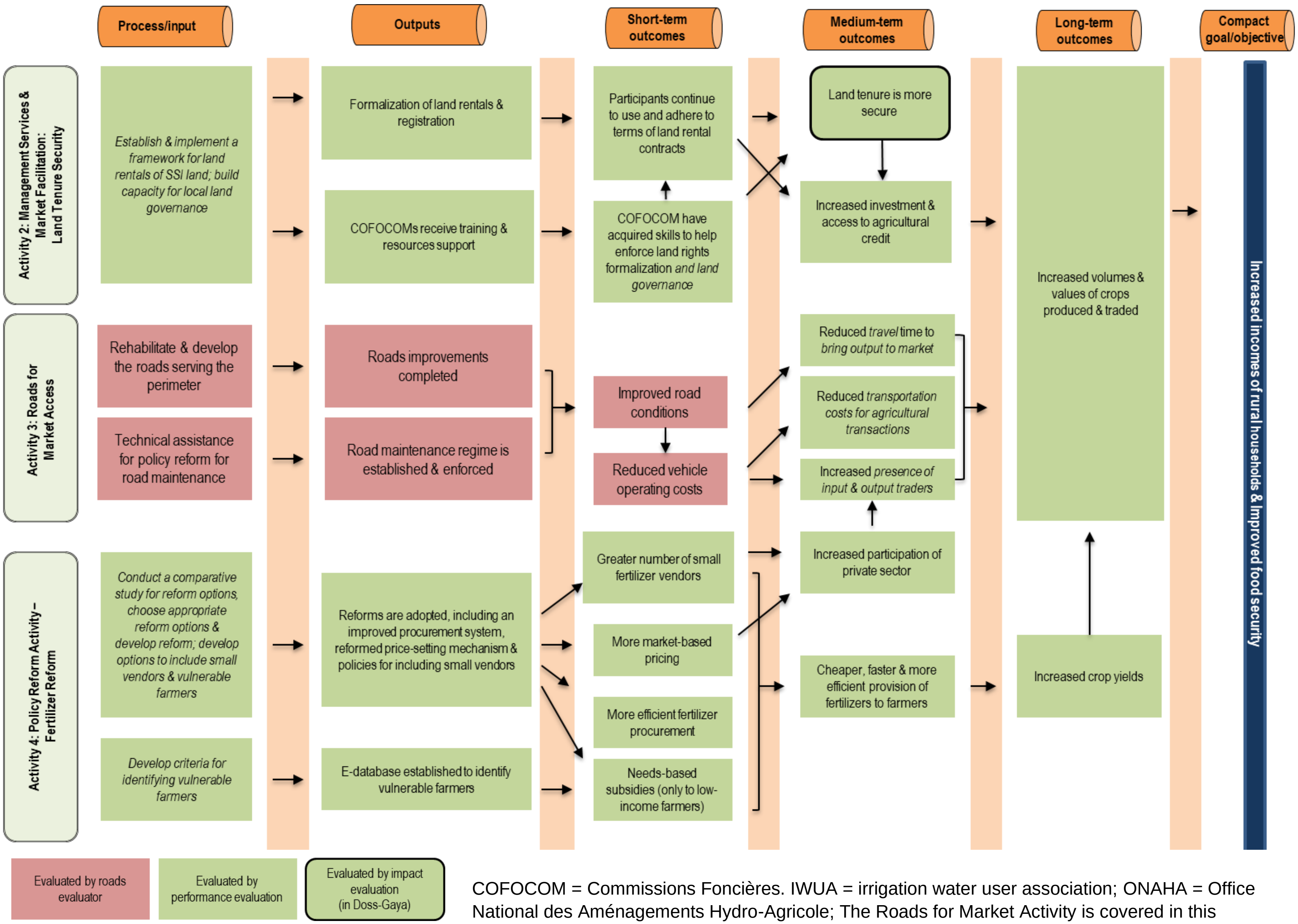
Different implementation timelines for the project activities necessitated separate evaluation design reports and baseline reports for the Konni area and the Dosso-Gaya project area (D'Agostino et al. 2019, 2021). In the first part of the baseline report, Ksoll et al. (2021) provide baseline findings for the Konni perimeter investments and national-level activities.[1] This second part of the baseline report focuses on establishing a baseline and constructing comparison groups for the compact investments in small-scale irrigation in the Dosso-Gaya area which include the construction of SSI infrastructure and complementary capacity-building activities for strengthening land tenure security, and improved agricultural productivity and market access. The report also includes baseline information to support the assessment of the benefits of the road infrastructure investments as they pertain to beneficiaries of the SSI activities.

The project's theory of change stipulates that investing in small-scale irrigation infrastructure will result in increased water availability for project beneficiaries during the rainy and dry seasons (MCA-N 2019). The project will supplement the small-scale irrigation infrastructure investments with technical assistance and training in water management, savings and financial literacy, improved production practices, agricultural marketing, and other complementary skills, such as literacy and numeracy, designed to increase overall production and sales on the perimeters. In addition, the project will provide support to improve land tenure security and test out new rental arrangements. **Figure I.2** shows the pathway from anticipated activities to short-, medium-, and long-term outcomes, which include increased crop yields, increased quantity and value of crops sold, economic empowerment of women and youth, and improved nutrition and hygiene practices of households. The ultimate goal of the project is to increase the incomes and food security of rural households. More information on the project, logic model, and theory of change can be found in the evaluation design report (EDR) focused on the investments in the Dosso-Gaya area (D'Agostino et al. 2021).

[1] MCC's evaluation catalog for the Niger IMAP houses all published materials associated with the evaluation and is available at https://data.mcc.gov/evaluations/index.php/catalog/265/related_materials.

Figure I.2. IMAP Logic Model





COFOCOM = Commissions Foncières. IWUA = irrigation water user association; ONAHA = Office National des Aménagements Hydro-Agricole; The Roads for Market Activity is covered in this evaluation only to the extent that the SSI beneficiaries benefit from access to roads.

B. Overview of evaluation

Mathematica developed an evaluation design to assess MCC's investments in the Dosso-Gaya area (D'Agostino et al. 2021). The EDR includes a mixed-methods approach to evaluate IMAP investments in Dosso-Gaya including: (a) an implementation analysis, (b) qualitative analysis, (c) an engineering assessment of the small-scale irrigation infrastructure, (d) a matched comparison design for the small-scale irrigation investments, and (e) a performance evaluation of the effects of the roads infrastructure on the *BT* beneficiaries. The analysis presented in this report provides a key input for the two quantitative components (d) and (e): it (1) establishes a baseline for the small-scale irrigation assessment and the roads benefits assessment, and (2) develops a matched comparison group for the small-scale irrigation assessments and assesses to what extent the selected comparison group can function as a reliable comparison group to rigorously estimate the impacts of small irrigation and complementary activities. Because no sampling frame was available for households in comparison areas, the proposed matched comparison group design relied on a two-step procedure. First, we selected potential comparison areas that contained possible matches for the likely small-scale irrigation target plots and conducted a listing survey with the households cultivating these plots. Second, we conducted a household survey in the *BT* owning or cultivating the plots likely to be irrigated and a subset of households in comparison areas identified through the listing. After the EDR was approved and finalized in September 2021, Mathematica conducted listing surveys with 3,461 individuals between October 13, 2021, and January 21, 2022, and household surveys with 1,827 households between November 26, 2021, and July 7, 2022.

B.1. Research questions

The EDR defines the comprehensive set of research questions that this evaluation will address. **Table I.1** sets out the research questions that are relevant to the quantitative analysis of outcomes in the *BT* and for which this report will establish baseline values. In addition, **Table I** presents the evaluation method used to address each research question, and the data source and type. The research questions relate to (1) changes in agricultural outcomes and household incomes, (2) outcomes related to the performance of the small-scale irrigation systems, (3) outcomes related to land tenure security, and (4) the cost of fertilizer. All of these research questions investigate specific outcomes of various project activities for a defined group of project participants. In terms of MCC's three overarching evaluation questions (MCC n.d., they all fall under the second evaluation question: "Did the investment produce the intended results? Did it achieve its stated objective in pursuit of MCC's mission to reduce poverty through economic growth?". The primary and secondary outcomes linked to these research questions are discussed in detail in **Chapter II, Section B.**

Table I.1. Evaluation design overview: Summary of research questions, methods, and data sources

Question group		Evaluation method	Data source and type
Overarching questions			
RQ3	What is the impact of SSI investments on beneficiary households' incomes, volumes, and value of agricultural products sold and traded, food and nutritional security, and production of cash crops?	• Difference in Differences with Matching	• Surveys of SSI beneficiary and comparison households • Satellite imagery • Crop cuts

Question group		Evaluation method	Data source and type
Irrigation Perimeter Development Activity			
RQ12	Did irrigated land increase as expected?	• Difference in Differences with Matching	• Satellite imagery • Surveys of SSI beneficiary households
Management Services and Market Facilitation Activity			
RQ22	What is the impact of SSI investments and land formalization on land tenure security and the level and risk of land conflict?	• Difference in Differences with Matching	• Surveys of SSI beneficiary households • COFOCOM administrative data
Roads for Market Access Activity			
RQ40	To what extent did the activity lead to a change in transportation method, travel time, vehicle operating costs, and transportation costs for traders and farmers in the *Basse Terrasse* and surrounding areas?	• Pre-post analysis • Quantitative descriptive analysis	• Surveys of SSI beneficiary and comparison households • Surveys of traders • Surveys with village leaders
RQ41	Are more input and output traders present in the Dosso-Gaya region as a result of the roads improvements?	• Quantitative descriptive analysis	• Market records • Surveys with village leaders
RQ42	To what extent did the activity contribute to increased volumes and values of agricultural products traded from the *Basse Terrasse* area? How has the activity changed the quality of crops, in particular produce, brought to market and the quantity of crops lost in transit post-harvest?	• Pre-post analysis	• Surveys of SSI beneficiary and comparison households • Surveys of traders • Surveys with village leaders •

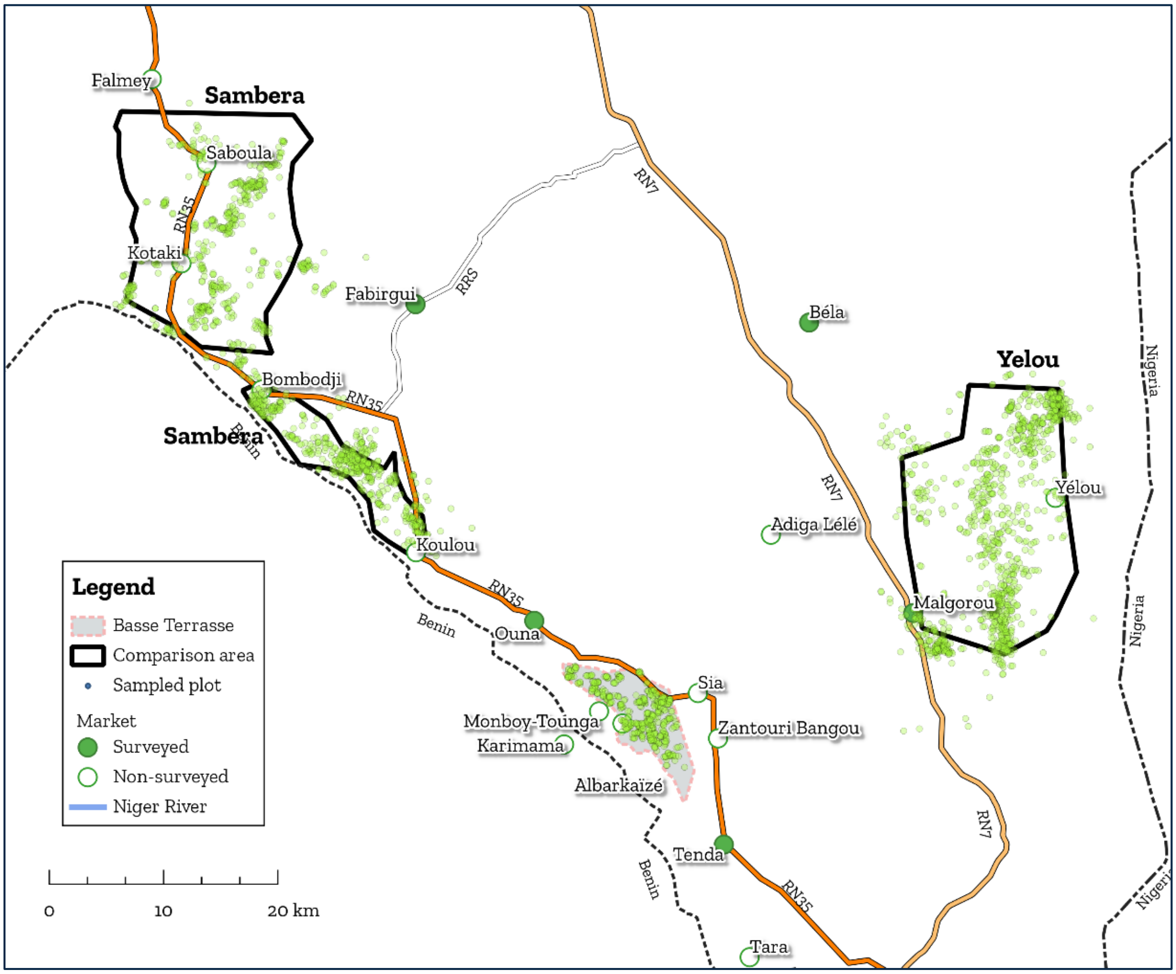
Notes: See **Table A.1.1** for a full presentation of the evaluation design and research questions.

For completeness, **Table A**.1..**1** in **Appendix A** provides a comprehensive list of the research questions related to Dosso-Gaya. Research questions for which the baseline quantitative data are not informative are presented in italics in **Table A.1.1.**

B.2. Quantitative evaluation methodology

The EDR provides information on the different methods we use to address specific research questions and the reasons for choosing a specific evaluation methodology (D'Agostino et al. 2021). We use a pre-post design to assess changes over time for traders and farmers operating in the *BT* and surrounding areas because of the roads improvement, as we have reliable baseline information, but a matched comparison group (MCG) design is not feasible. We use a MCG design to construct a comparison group with similar baseline characteristics as the group of households and plots in the *BT* receiving the SSI package. The MCG design allows us to estimate the impact of compact investments in SSI. This report presents the pre-intervention baseline values for the outcomes of interest, based on household survey data from the Dosso-Gaya area. It also provides information on the extent to which comparison households can be selected to match the treatment households.

Figure I.3. Overview map of project area, including roads targeted for rehabilitation and key agricultural markets



C. Timeline for data collection, evaluation, and project activities

This baseline report is based on two distinct data collection activities: household data collection (which necessitated a prior listing survey and included interviews and plot measurements) and a trader, market, and village leader survey. **Figure I.3** depicts the locations for the household survey (in the *Basse Terasse* and comparison areas) and the trader and market survey in relation to the project area[2] and the roads targeted for rehabilitation (RN7, RN35, and RRS).

Figure I.4 presents the timeline for the surveys in the Dosso-Gaya area and the dry and rainy seasons referenced in the household and plot measurement surveys, as well as the timeline for the implementation activities in the Dosso-Gaya area, beginning in the third quarter of 2020. Because implementation of some project activities began before the end of the 2020 and 2021 rainy dry seasons covered by the baseline survey, the outcomes measured by the baseline survey might reflect some very early effects of the training activities, policy reforms, and land tenure activities. However, since small-scale irrigation

[2] Given several iterations of the project activities in the Basse Terasse, it is important to note that the project area described here corresponds to what is informally known as SK2, which targeted small-scale irrigation for over 600 hectares of land.

construction has not yet begun, the baseline survey should be well placed to serve as a true pre-irrigation construction baseline.

Figure I.4. Timeline for data collection and (planned) implementation activities

Year	2020		2021				2022				2023				2024			
Quarter	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Implementation activities																		
Plot identification for SSI				■														
SSI construction																		
BT training activities				■	■	■	■	■	■	■								
Roads for markets	■	■	■	■	■	■	■	■	■	■								
Policy reforms	■	■	■	■	■	■	■	■	■	■								
Data collection activities																		
Listing survey						■												
Baseline household survey						■	■											
Baseline household plot measurement survey							■	■										
Dry and rainy seasons covered by household survey and plot measurement survey		■	■	■	■													
Trader and market survey							■											
Denotes end of Niger Compact after extension																		
Note: plot identification has begun for the SSI construction activities, but construction has not yet begun.																		

D. Key baseline findings

In this section we preview the key findings from the baseline analysis and the implications we draw for the evaluation. The key findings focus on constraints to higher productivity and incomes at baseline and the potential of the IMAP activities to overcome these constraints.

Small-scale irrigation

- Farmers in the *Basse Terrasse* are primarily engaged in rainfed agriculture and grow traditional crops such as millet, sorghum and cowpeas. Intercropping is common in the rainy season with 40 percent of plots intercropped. Less than 10 percent of plots are cultivated in the dry season; most of these plots grow rice; about 30 percent use surface water as the source of irrigation water, and the remainder draw water from tube wells or other wells. Yields and revenue are low, reflecting the low use of best practices in input use and agricultural practices. Households are poor with an estimated 40 percent of households below the national poverty line. Thirty percent of households state having experienced hunger in the past lean season.
- Access to small-scale irrigation is a significant opportunity to increase farmer's agricultural production and incomes and to improve food security. However, at baseline few to no farmers cultivated cash crops such as tomatoes or onions; and only 10 percent of plots were cultivated in the dry season using irrigation. The low baseline prevalence of higher-value crop cultivation, use of best practices and irrigation access justifies the logic model's inclusion of training activities designed to support farmers to fully benefit from access to irrigation. Farmers will also need (access to) complementary inputs such as seeds and fertilizer highlighting the importance of IMAP activities focused on increasing market presence of input traders.
- Levels of female empowerment are low across many domains of empowerment. The low level of membership in groups and low levels of decision making with respect to assets stand out (negatively)

among the domains of empowerment, and supports the program's logic of raising women's empowerment in households and communities through increases in group membership and accumulation of productive capital.

Implications for the evaluation of SSI investments

- Overall, the process we developed to create a matched comparison group has led to a treatment and matched comparison group that is well balanced in both indicators derived from remote sensing and those derived from household survey data. Most primary and secondary indicators show standardized differences that are smaller (in absolute terms) than the 0.1SD threshold below which the literature considers variables to be balanced. Of the remaining indicators, most are below the 0.25SD threshold that the literature considers to be an acceptable range for removing bias through the inclusion of baseline outcomes.
- The procedure was, however, not effective in finding matches from the Yelou and Sambera comparison areas for the 10 percent of SSI target plots engaged in dry season rice cultivation. The small number of SSI target plots and their matched comparisons remain significantly different according to input use, yields and sales. Given the project's intended focus on non-rice-growing plots, the consequences of these specific imbalances for the evaluation will depend on whether these plots are part of the final group of SSI beneficiary plots. If they are, it will be important to identify additional comparison plots for these target plots.
- While matching was effective in balancing overall levels of empowerment between the *Basse Terrasse* and the matched comparison samples, this was not the case for the different domains of empowerment. To obtain balance for the various empowerment domains the final evaluation will implement a separate matching procedure for women.
- Given delays in project activities, the final scope of the SSI activity and the selection of beneficiary plots are yet undetermined. The possibility of a significant reduction in scope raises questions concerning the power of the evaluation design to identify impacts. Once the scope is finalized, we will update power calculations. The availability of the baseline data collected for this report will ensure these power calculations are reliable.
- While overall response rates are high, we were not able to interview all target SSI beneficiary households nor obtain geo-coded information from all target plots. If project scope is significantly reduced, we will assess whether the smaller number of beneficiary plots, and the area to be irrigated, was covered by the baseline data collection.
- Finally, the baseline report confirms findings in the literature that self-reported plot sizes and the area cultivated are subject to substantial biases. Plots measured using GPS measurements are about 1/3 smaller than the farmer reported size of the same plot. A primary reason is bunching where farmers generously round up to the nearest hectare or a simple fraction of a hectare—such as 1/2 of a hectare. This substantially affects measurements of agricultural productivity that are calculated on a per hectare basis of the area that is cultivated. Given the importance of these biases it is important to continue to collect geo-coded measurements of the area cultivated.

Roads for market access

- The findings from the market, trader, and village leader surveys suggest that rehabilitating the RN7, RN35, and RRS could lead to improvements in the short-, medium-, and long-term outcomes envisioned in the logic model as the current poor road quality does appear to be a binding constraint.

- In particular, the poor quality of roads increases the travel time to markets and transportation costs in the rainy season, a time when more crops are produced, and rain affects the state of many unimproved roads.
- Traders operating in markets near the project area are hesitant to trade perishable crops and experience losses and damages to these crops in transit when they do transport them, resulting in decreased crop sales and trader income. Farmers are similarly hesitant to grow perishable crops, likely because they are harder to sell and transport to market given the challenges faced by traders.
- The roads for market access activity also benefits households benefitting from the SSI investments. Given the SSI investment's focus on shifting to higher value crops, some of which are perishable, the logic model's envisioned positive impact on the presence of traders, lower transportation costs, and decreased travel time to bring crops to markets, would support the anticipated increase in the volumes and values of crops produced and traded to achieve the overall compact goal of increasing farmer incomes.

E. Road map of the report

This baseline report contains three chapters. **Chapter I** provides an overview of the project and evaluation. **Chapter II** contains the primary analysis of this report and (1) presents an overview of the matching process; (2) describes findings for the outcome indicators for the treatment sample and, where relevant, describes differences with the comparison samples; and (3) presents information on traders and markets in the area. **Chapter III** concludes the report with a discussion of the administration of the evaluation, including institutional review board (IRB) approval, data access and privacy, the dissemination plan, and the evaluation team. The appendixes contain supplemental information, including additional research questions and secondary outcome indicators (**Appendix A**); information on the coverage of geo-coded Basse Terasse target plots (**Appendix B**); matching methodology and supplemental matching results(**Appendix C**); yield estimates by method of measurement and by intercropped versus pure-stand cultivation (**Appendix D**); historical rainfall patterns for treatment and control areas (**Appendix E**); and stakeholder comments (**Appendix F**).

II. Dosso-Gaya Baseline Analysis

A. Overview of matched comparison group design, matching procedure, and outcomes

In this section we describe our matching procedure and report on key characteristics of our baseline sample in the treatment and comparison areas. Technical details are in Appendix C.

We will use a **matched comparison group (MCG) design** to evaluate the impact of SSI investment on selected outcomes where we include the baseline value of the outcome to conduct an ANCOVA analysis. We will analyze outcomes on two levels: (1) the plot level, using a plot-level MCG, and (2) on the household level, using a household-level MCG. To construct the MCG we used in this baseline analysis and we will use for the evaluation, we performed plot-level matching using a combination of satellite data and plot-level indicators collected from the household survey, and we performed household-level matching using only data collected from the household survey. Our objective was to develop an MCG of plots that resemble treatment plots in key characteristics, and similarly an MCG of households that resemble treatment households. We treat plots as independent of the households that cultivate them, and we construct comparison groups in separate matching processes, such that including a household in the household-level sample does not mean its plot[s] will necessarily be included in the plot-level sample.[3]

With a credible comparison group, this design can support causal claims that the estimated effects are the result of SSI investments, including complementary training activities and land tenure support. Specifically, the comparison group enables us to remove the effects of external shocks and isolate, in beneficiaries' outcomes, changes that are due solely to SSI investments. The ability to account for common shocks affecting both treatment and comparison households is especially valuable in contexts like the *Basse Terrasse,* where livelihoods depend on rainfed agriculture and thus rainfall variations directly affect household welfare. The counterfactual provided by the comparison group will also improve the accuracy of the investment's cost-benefit analysis, which will be based on changes driven by the program and not just changes over time, as would be the case in analyses lacking a comparison group.

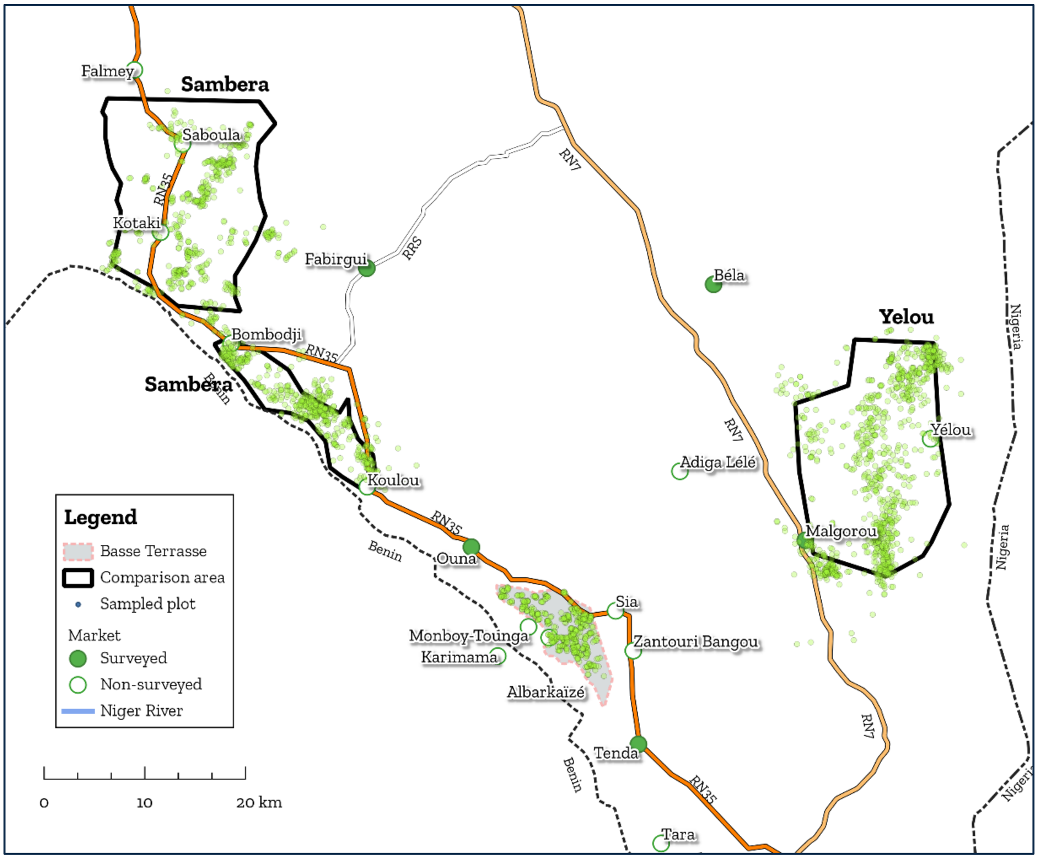
Appendix C provides additional information on our data collection process, our approach for assessing balance between treatment and comparison groups, and balance results for both our plot-level and our household-level samples.

A.1. Definition of sample and coverage

Figure II.1 illustrates the geographic spread of plots that were surveyed as part of the plot outline component of our household survey. *Basse Terrasse* plots are contained within the shell-shaped area in pink that lies adjacent to the Niger River and the border with Benin. The IMAP project selected treatment plots in a multistage process that relied on Light Detection and Ranging (LiDAR) to detect groundwater and on-the-ground activities by a consultant to identify owners and cultivators willing to share access to irrigated land. Due to a reduction in scope, the extent of which is still to be determined, this is not the final group of treatment plots. The comparison plots come from either Sambera (purple) or Yelou (seagrass). Both comparison regions adjoin international borders and national route highways (for example, RN7).

[3] This approach of cross-level independence allows for improved balance performance compared to forcing household-plot pairing across both levels of analysis. Our approach also increases the likelihood of recovering valid impact estimates.

Figure II.1. Map of sampled plots in BT and comparison areas



Note: The comparison area polygons constitute the boundary areas within which all target plots for baseline data collection were located. Plots outside the comparison area polygons are non-target plots also cultivated by survey respondents.

Our baseline survey had a high response rate, as shown in **Table II.1**1. Overall, 83.1 percent of contacted households completed the interview component, and a smaller 79.4 percent completed the plot measurement component. For both the interview and plot measurement components, Sambera households registered the highest response rates and Yelou the lowest.

Table II.1. Household survey response rates by component and zone

Area	Total number of households contacted	Household survey interview component: number completed	Response rate household interviews (%)	Household survey plot measurement component: number completed	Response rate plot measurement (%)
Basse Terrasse	711	596	83.8%	579	81.4%
Sambera	712	622	87.4%	599	84.1%
Yelou	776	609	78.5%	567	73.1%
Total	**2,199**	**1,827**	**83.1%**	**1,745**	**79.4%**

Note: Mathematica calculations.

A.2. Propensity score model and balance

In this section, we summarize the methodology and the criteria for assessing the ability of the matching procedure to create appropriate comparison groups. We then present the analysis, and comment on the ability of this matching procedure to create appropriate comparison groups. In Appendix C, we describe the propensity score model and report in detail on balance between the treatment and comparison groups.

We use three indicators to assess the suitability of the comparison group, separately for plot- and household-level results. We present (1) the distribution of propensity scores, (2) information on the sample size of included observations, and (3) balance test results for a number of variables to demonstrate the suitability of the selected comparison group.

We first show the overlap of propensity scores to show the availability of potential comparison units for different levels of the propensity score. Then we present the number of observations for which suitable matches are available and that are retained for the analysis sample.

Following Imbens and Rubin (2015), we then assess balance for an individual variable by estimating the standardized mean difference between treatment and comparison groups, which is calculated as:

$$d = \frac{\mu_T - \mu_C}{\sqrt{\frac{\sigma_T^2 + \sigma_C^2}{2}}}$$

where the mu terms represent the treatment and comparison means and the denominator is the square root of the average variance.[4] Following work such as Normand et al. (2001) and Stuart et al. (2013), we report the standardized mean difference for variables and use a 0.1 absolute standardized mean difference (ASMD) threshold as constituting acceptable balance.[5] This literature also suggests that variables for which absolute imbalance is between 0.1 SD and 0.25 SD can be accounted for by including the baseline values of these variables in outcome regressions when estimating treatment effects (a procedure known as regression adjustment). Variables with an ASMD of less than 0.1 SD indicated sufficient balance at baseline and therefore do not need to be included as controls in the final analysis.

We extend this approach to primary and secondary outcomes as follows: where standardized differences exceed 0.1SD but are less than 0.25 SD, our regression models for the final analysis will therefore include controls for the baseline value in a regression of these outcomes on a treatment indicator.[6] This is our preferred specification for two primary reasons: (1) including baseline values as controls is efficient and therefore increases power (Lin 2013 List et al 2021), and (2) this approach allows for differences in measurement between the baseline and endline measures.

When differences in primary and secondary outcomes are larger than 0.25 SD, we instead use a difference-in-differences specification for the matched treatment and control groups. This compares the changes over time between the treatment and the control groups.

[4] Imbens and Wooldridge (2009) argue that the standardized difference should be used and not tests of significance because the standardized difference provides a better measure of the difficulty to adjust for the differences through the use of covariates in the outcome regression.

[5] We performed all matching tasks using the "Matching" package (Sekhon 2011), and assessed covariate balance using the "cobalt" package (Greifer 2022), both in R (R Core Team 2022).

[6] More specifically, we include the baseline covariate as well as an interaction with the treatment indicator. In the context of randomized trials, Lin (2013) and List et al. (2021) show this technique reduces bias and has higher power.

We display the results of balance tests in the subsections that follow.[7]

In addition to reporting ASMD, we also graphically examine univariate distributions of the same covariates (See Appendix C). This step offers additional evidence on the similarity (or dissimilarity) of the treatment and comparison groups. Rather than just sharing a similar mean value, the density of values across the covariate's full support should also be comparable.

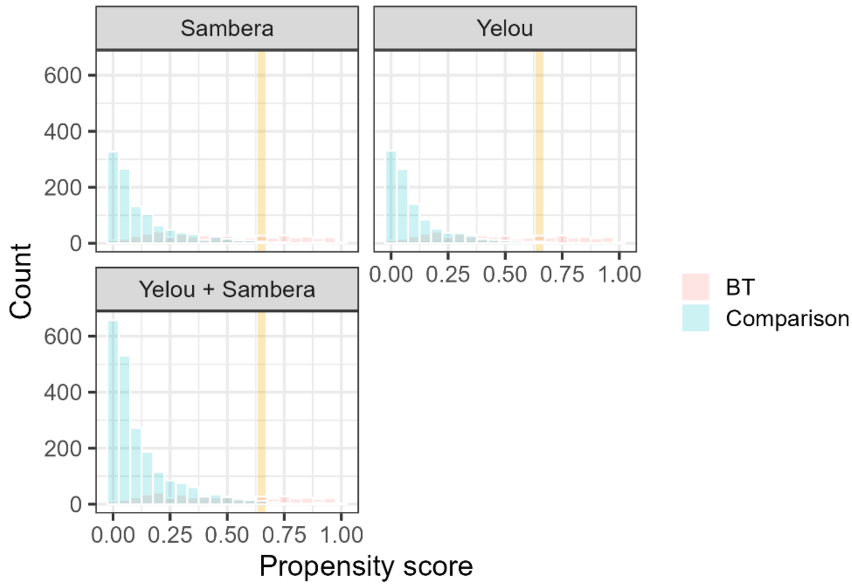
Plot-level balance

Under our preferred specification for estimating propensity scores (in which "treatment" is a function of the linear combination of all plot-level variables under the "PS column in **Appendix Table C.1Error! Reference source not found.**), we find that propensity scores for treatment plots approximate a uniform distribution between 0.1 and 0.7, and propensity scores for comparison plots are right skewed, as seen in **Figure II.**. Both Sambera and Yelou exhibit similar distributions, and because they have comparable sample sizes (**Table II.**), also have similar bin heights across the x-axis range when examined separately. There is significant value in collecting plot-level data from both comparison areas in future data collection as potential matching units for *Basse Terrasse* observations with propensity scores of 0.4 or higher, reducing the importance from any single comparison unit. More concretely, consider the Yelou graph in **Figure II.2**. For the bin centered at a propensity score value of 0.45, treatment plots outnumber comparison plots. However, when Sambera plots also become available, then we have more comparison plots than treatment plots at a given propensity score value.

The common support of propensity scores is roughly 0 to 0.65. As seen in the figure, many treatment plots have propensity score values exceeding 0.65, but few to no comparison units do. Our analysis sample consists of all observations with propensity score values less than or equal to 0.65. Since a larger share of target plots have high propensity score values, the restriction to observations in the area of common support excludes a larger proportion of the available *Basse Terrasse* plots, which shrinks from 485 to 337 plots, a 30.5 percent reduction (**Table II.).** In comparison, because of the common support restriction, fewer than 2 percent of Sambera and Yelou plots are excluded from the MCG design.

[7] Nguyen et al. (2017) find that regression adjustment can be used effectively for variables whose imbalance exceeds 0.1 SD, with diminishing benefits when applied to variables with smaller imbalance. Imbens and Wooldridge (2009) report that when imbalances are larger than 0.25, regression adjustments are more likely to be sensitive to the exact specification used. What Works Clearinghouse (2022) applies a more-exacting 0.05 SD cutoff, with covariates whose ASMD is between 0.05 and 0.25 requiring regression adjustment. The Clearinghouse states that covariates with ASMD values exceeding 0.25 do not satisfy their baseline equivalence standard.

Figure II.2. Plot-level propensity score distributions by comparison region



Note: Mathematica calculations. Vertical gold lines represent the propensity score threshold of 0.65, with observations to the right dropped from the analysis sample. **Appendix Table C.2** Error! Reference source not found.presents the regression specification to estimate propensity scores. N = 2626 plots (485 *BT*, 1103 Sambera, 1038 Yelou).

Table II.2. Sample sizes available for constructing plot-level matched comparison group sample

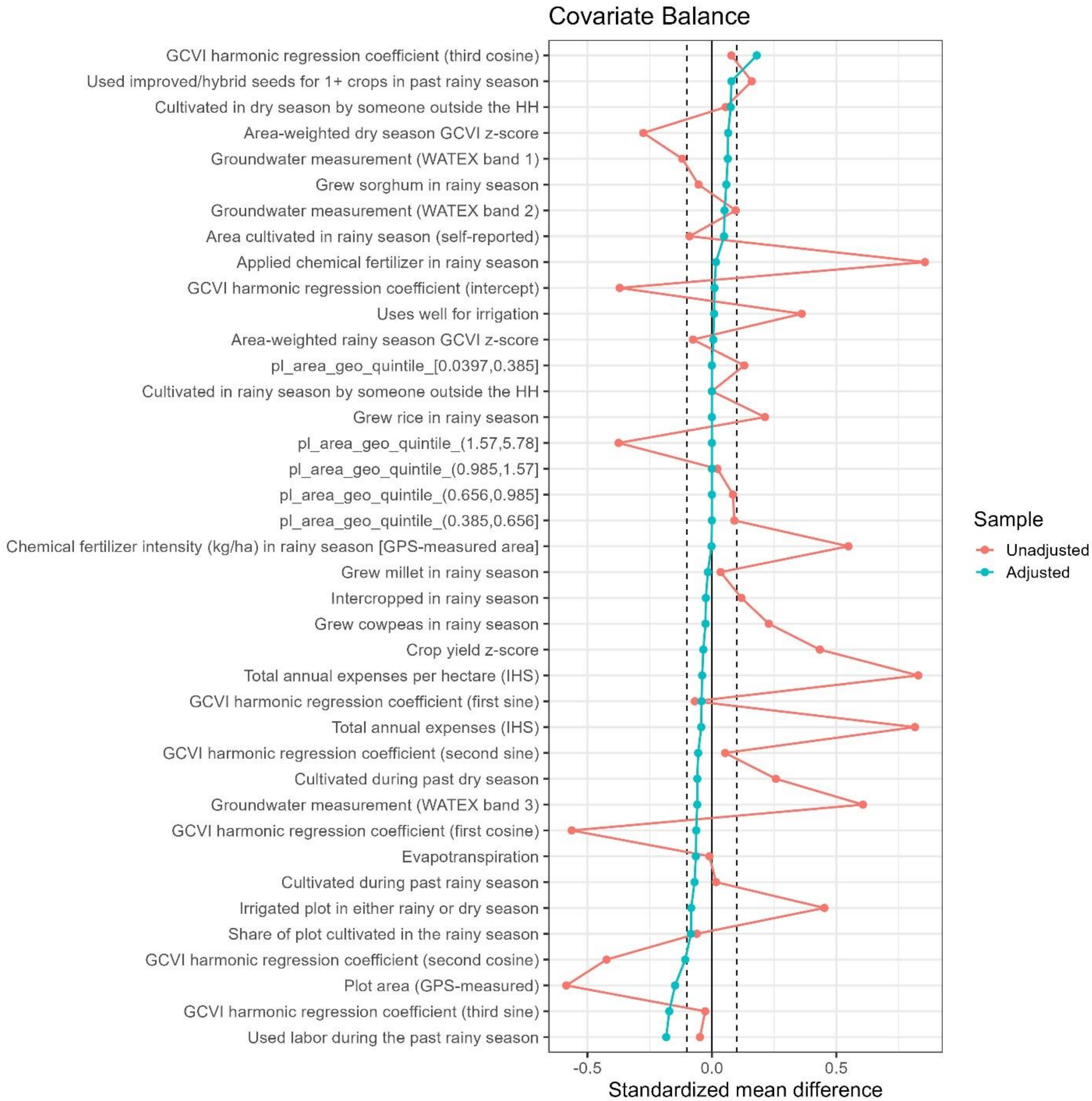
Region	Full sample (N)	Trimmed sample (N)	% of full sample
Basse Terrasse	485	337	69.5
Sambera	1,103	1,088	98.6
Yelou	1,038	1,029	99.1
Total	**2,626**	**2,454**	**93.4**

Note: Mathematica calculations. Analysis sample drops observations with estimated propensity scores > 0.65. Sample sizes do not reflect the actual samples included in the matched comparison group analyses, but instead denote the number of observations available for propensity score matching.

We display covariate balance in **Figure II.** and find a strong degree of balance across numerous plot-level covariates. Peach-colored values denote the unadjusted (that is, not applying weights estimated from the matching process) sample means while the teal values represent the adjusted means. Under our 0.1 AMSD criterion, 34 of the 39 assessed covariates are balanced. Key plot-level characteristics that are *ex ante* likely to correlate with headline project outcomes such as agricultural revenue and food security are balanced, including groundwater accessibility, chemical fertilizer application (binary), dry season cultivation (binary), and crop-specific z-scores of crop yields.

None of the unbalanced variables exceed an adjusted mean difference of 0.2 SD (Error! Reference source not found.). **Figure II.** displays the direction of the imbalance, with teal observations to the right of the solid vertical line, which indicates that the adjusted treatment mean is greater than the adjusted comparison mean. The reverse holds at the bottom of the plot, which is arranged sequentially from largest (most positive) standardized mean difference to smallest. Most notably, one of the GCVI harmonic regression coefficients has a larger adjusted treatment mean than comparison mean (the third cycle, or highest frequency cycle, cosine term), and adjusted means among the comparison group were larger than among the treatment group for rainy season cultivation, the other third-cycle GCVI harmonic regression term, and whether hired labor was employed during the rainy season.

Figure II.3. Treatment-comparison balance of plot-level covariates using 5-nearest-neighbor matching with replacement and common support restriction



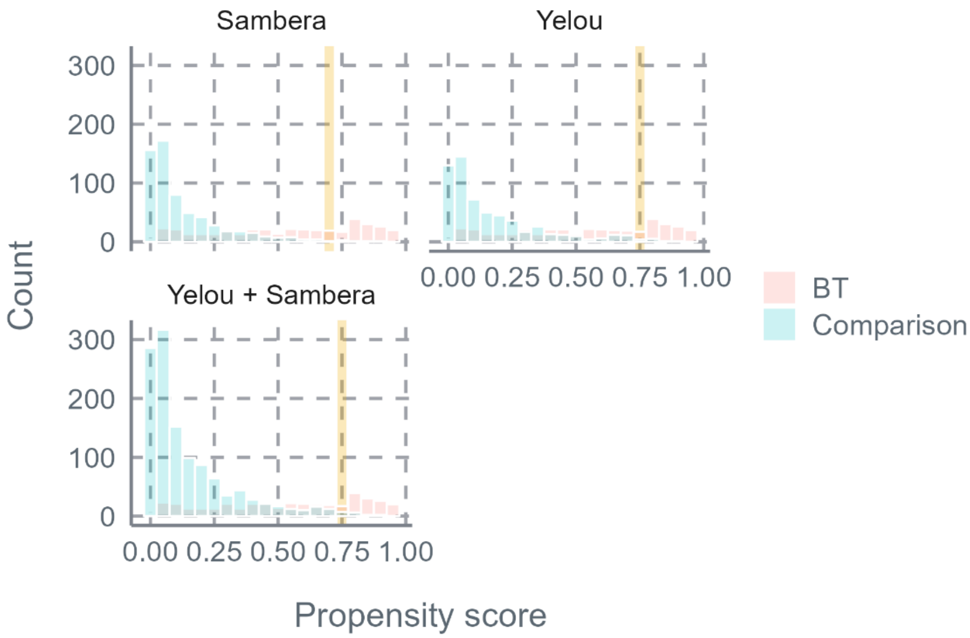
Note: Mathematica calculations. Results are for analysis sample of observations with estimated propensity score values less than or equal to 0.65 calculated using the logistic regression model in Error! Reference source not found.N = 1067 plots (337 treatment, 730 comparison).

In addition to the standardized mean differences displayed above, we also report the adjusted mean and SD for each variable in **Appendix Table C.7**.

Household-level balance

We perform a separate logistic regression (in which "treatment" is a function of the linear combination of all household-level variables under the "PS column in **Appendix Table C.1**Error! Reference source not found.) to estimate propensity scores at the household level, with the propensity scores by region displayed in **Figure II.**. Households in the *Basse Terrasse* have propensity score values across nearly the entire 0 to 1 range. In contrast, the highest propensity scores observed for Sambera and Yelou households are about 0.70 and 0.75, respectively, shown as the vertical gold line. The largest cluster of households in Sambera and Yelou have propensity score values below 0.15. To implement the common support restriction, we retain all observations with a propensity score of 0.75 or less for Yelou and 0.70 or less for Sambera, which provides us with pre-matching, region-specific sample sizes shown in **Table II.**. With this threshold, the impact evaluation will exclude 123 *Basse Terrasse* observations, or about 31.5 percent of the sample, because of the very small number of possible matches in either Sambera or Yelou. (In Yelou or Sambera, around 1 percent of households have a propensity score above the threshold.)

Figure II.4. Household-level propensity score distributions by comparison region



Note: Mathematica calculations. Vertical gold lines represent the propensity score threshold value, with observations to the right of the line dropped from the trimmed sample. The threshold value is 0.70 for Sambera, 0.75 for Yelou, and 0.75 for Yelou + Sambera. See Error! Reference source not found. for the regression specification used to estimate propensity scores. N = 1608 households (391 *BT*, 617 Sambera, 600 Yelou).

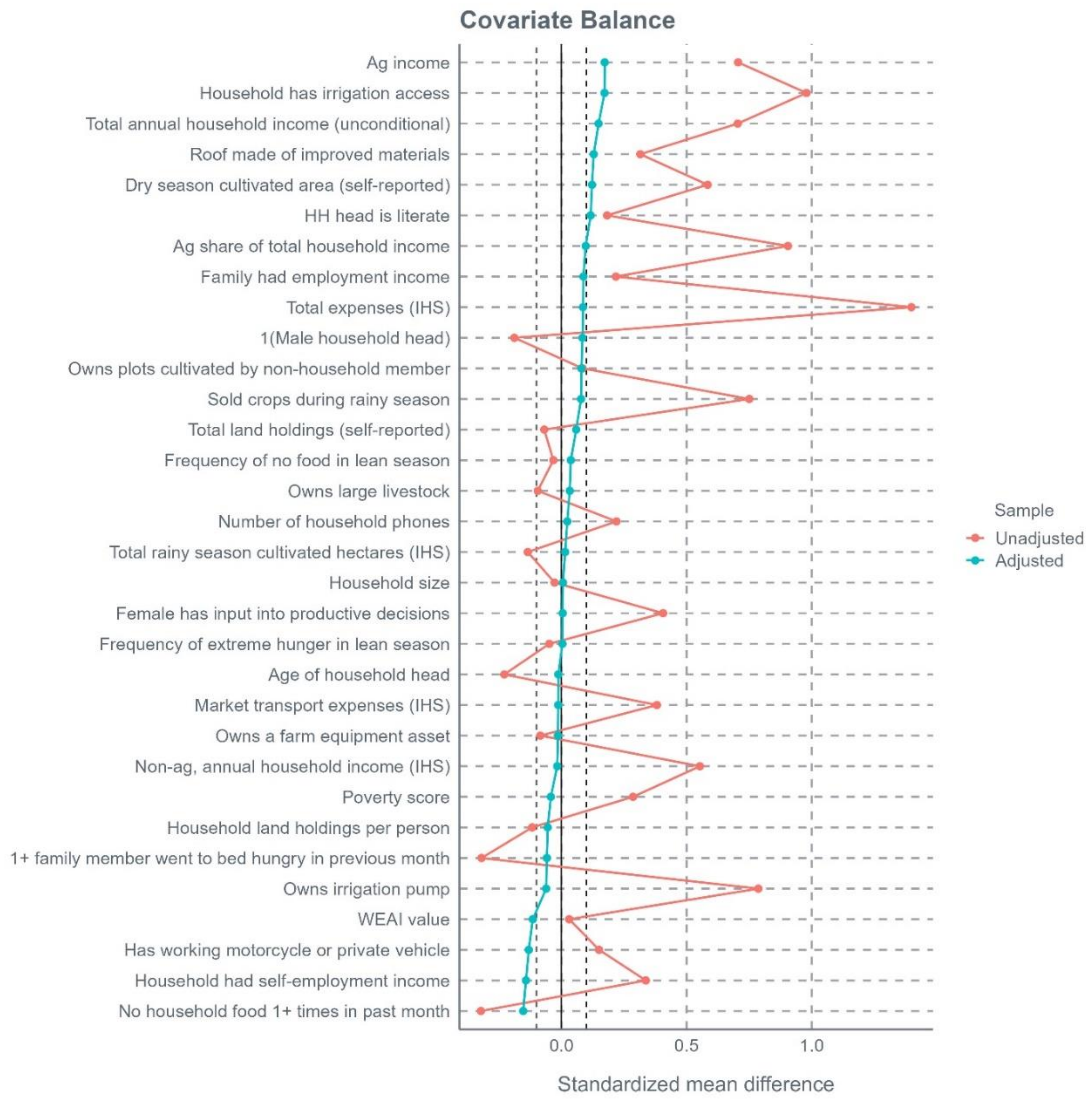
Table II.3. Sample sizes available for constructing household-level matched comparison group sample

Region	Full sample (N)	Trimmed sample (N)	% of full sample
Basse Terrasse	391	268	68.5
Sambera	617	613	99.4
Yelou	600	592	98.7
Total	**1608**	**1473**	**91.6**

Note: Mathematica calculations. Households with propensity scores above 0.70 (for Sambera) and 0.75 (for Yelou) are dropped in the trimmed sample.

Applying the 0.1 SD threshold, we see in **Figure II.** that AMSDs (in teal) are in an acceptable range for important indicators like rainy season cultivated area, poverty likelihood, and several food security metrics. For variables at the top of the figure, adjusted mean values for the treatment group exceed those of the MCG. The figure demonstrates that even after matching, our *Basse Terrasse* households on average experience better outcomes along some dimensions that are likely attributable to improved growing conditions. They enjoy more irrigation access, which leads to higher dry season cultivation and annual agricultural and total household income. The increased agricultural production also incurs more input expenses than incurred by the comparison group. At the bottom of the figure, adjusted mean values for the comparison group exceed those of the treatment group. For all covariates where the comparison group's adjusted mean value exceeds the treatment group's (such as from poverty score to household had self-employment income), differences are always below 0.2 SD.

Figure II.5. Treatment-comparison balance of household-level covariates using 5-nearest-neighbor matching with replacement on a trimmed sample



Note: Mathematica calculations. Results are for trimmed sample of observations with estimated propensity score values less than or equal to 0.70 (for Sambera) or 0.75 (for Yelou). N = 758 households (268 treatment, 490 comparison).

In **Appendix Table C.9**, we report the adjusted means and SD as additional context for interpreting the balance performance displayed above.

B. Data collection

This section describes the two distinct baseline data collection activities—household data collection (which required a prior listing survey and included interviews and plot measurements) and a trader, market, and village leader survey—that form the basis for the analysis presented in this report. It also defines the primary and secondary indicators that we present in the subsequent sections.

Household data collection

Before fielding the household survey, we conducted a listing exercise to collect preliminary information on household's characteristics, landholdings, and agricultural activities to identify households in the comparison areas of Sambera and Yelou that are comparable to households receiving small-scale irrigation in the *Basse Terrasse* to serve as the comparison group for the evaluation. The data from the listing exercise, conducted between October 13, 2021, and January 21, 2022, formed the basis for the final sample for the household and plot measurement survey. A total of 3,461 individuals completed the listing survey – 694 in the *Basse Terrasse,* 1,538 in Yelou, and 1,229 in Sambera.

Because of length, the household survey was split into two components: an interview and a plot measurement. A total of 1,827 households completed the interview component, and 1,745 (96 percent) of those households also completed the plot measurement component. The interview component, which measured household characteristics and agricultural activities, was conducted between November 26, 2021, and March 24, 2022.[8] The plot measurement component, which measured the area of plots and areas cultivated by the household, was conducted between February 16, 2022, and July 7, 2022, including a pause in data collection activities during Ramadan.

Both components of the household survey measured agricultural activities in the dry season from October 2020–May 2021 and the rainy season from June–September 2021. We focused on the 2020 to 2021 dry and rainy seasons in the baseline survey because, at the time of designing and planning for the baseline, the construction of small-scale irrigation infrastructure was scheduled to begin in April–June 2022. We needed to collect information about the dry and rainy seasons prior to the beginning of construction of irrigation for the baseline to serve as a true reference point. Conducting baseline data collection later or collecting information on the following dry season (October 2021–May 2022) risked the baseline reference period overlapping with the beginning of irrigation construction.

Table II. lists and defines the primary outcome indicators for the impact analysis, categorized by land security, irrigation, fertilizer, agricultural production, income, food security, and women's empowerment. We selected the primary indicators based on the IMAP program logic along with the core pre-post questions in the Indicator Tracking Table (MCA-N 2019) and the EDR (D'Agostino et al. 2021). These core levels of data correspond to levels represented in the IMAP program logic (**Figure I.2**). The subsequent analysis presents our results at these different levels and enables us to assess the extent to

[8] Because we were concerned about completing the household survey before the start of irrigation construction in the *Basse Terrasse*, we began the survey in the *Basse Terrasse* before the listing survey was completed in Sambera and Yelou. We then paused household survey data collection until the listing survey was completed in Sambera and Yelou, at which point we completed the survey with the most appropriate households to serve as the comparison group for the *Basse Terrasse* treatment households.

which the baseline data provide support for the constraints to agricultural production that underlie the program logic.

Table II.4. Primary outcome indicators for matched comparison group analysis

Indicator	Definition
Land security and use (target plot)	
Formalized land rights	Indicator of formal land rights documentation and the associated land area measure
Land security	Indicators for experienced land dispute in last year, perception of involuntary loss of land over subsequent five years, and the associated land area measures
Land under cultivation	Total land under cultivation by season and crop type
Irrigation (target plot)	
Use of irrigation	Seasonal indicator of use of irrigation other than rainfall
Irrigation expenditures	Annual household expenditures on irrigation and cost of irrigation per hectare
Fertilizer (target plot and household)	
Fertilizer application	Seasonal indicators of fertilizer use and quantity of fertilizer applied per hectare
Fertilizer expenditures	Annual household expenditures on fertilizer and cost of fertilizer per hectare
Agricultural production outcomes (crop level for target plot)	
Crop yield	Seasonal productivity (t/ha) for focus crops
Crop income per hectare	Seasonal income (crop sales and own consumption net of expenditures) per hectare for focus crops
Income (household level)	
Total income	Annual value of non-agricultural and agricultural income
Agricultural income	Annual income from crop sales, renting out land, and own consumption net of agricultural expenditures
Food insecurity (household level)	
Food inadequacy	Indicator of households that did not have enough food in the previous month
Hunger	Indicator of households where at least one member went to bed hungry in the previous month
Women's Empowerment in Agriculture Index (WEAI) (household level)	
Women's empowerment score	Adaptation of the Women's Empowerment in Agriculture Index (WEAI) that measures women's empowerment based on their role in four different domains: production, resources, income, and leadership

Note: All indicators are constructed from household survey and plot measurement data.

The secondary indicators, which allow us to further explore the research questions and contextualize the data, are presented and defined in **Table A.2**

Trader, market, and village leader surveys

To provide a baseline for the **Roads for Market subactivity,** we present baseline outcomes measured by data collected from trader, market, and village leader surveys. The surveys were conducted from March 11–20, 2022. We interviewed managers from six markets in the Dosso-Gaya region (Béla, Fabirgui, Gaya, Malgorou, Saboula, and Tanda markets), 36 traders (three itinerant and three fixed traders from each market), and 15 village leaders (one each from five villages in the *Basse Terrasse,* five villages in

Sambera, and five villages in Yelou) about perceptions of road quality and the transportation of crops. We purposively sampled markets by selecting large markets and randomly selected villages of differing sizes that depend on the roads being rehabilitated by the IMAP (the RN7, RN35, and Rural Route Sambera). We used the market survey to create our trader sample frame. During each market survey, we asked market managers to provide contact information for five large-volume fixed crop traders and five large-volume itinerant crop traders of both genders and randomly selected three of each type (fixed and itinerant) from that list.

Table II.5 lists and defines the indicators for the pre-post outcomes analysis of the Roads for Market subactivity. Since the Roads for Market activity affects farmers and traders in the *Basse Terrasse* and the two comparison areas, we are unable to conduct an impact analysis for these outcomes but will measure them again at endline and report changes over time. We intended to present traders' vehicle operating and maintenance costs, but only one trader in our sample owned a vehicle. The rest of the traders used shared transportation, mostly commonly a vehicle owned by the market, and thus did not have operating or maintenance costs to report.

Table II.5. Primary outcome indicators for the Roads for Market pre-post analysis

Indicator	Data source	Definition	Estimated exposure period[a]
Presence of traders			
Traders present at market	• Market survey	• Median number of fixed male, fixed female, itinerant male, itinerant female, input, livestock, crop, perishable crop, and non-timber forest product traders present at market	1–2 years
Traders present at village	• Village leader survey	• Median number of input, livestock, crop, and total traders coming to the village or farmgate in each season	1–2 years
Transportation			
Transportation method for crops to market	• Village leader survey	• Indicator for transportation method (motorized vehicle or animal-pulled cart) most frequently used by farmers to transport crops to market for sale	6 months to 1 year
Trader transportation method	• Trader survey	• Indicator for type of vehicle (truck known as *camion* or mini-truck known as *dogonbaro*) used by trader to transport crops	6 months to 1 year
Travel time	• Village leader survey • Trader survey	• Average travel time (minutes) from village to two nearest markets in each season • Trader's travel time (minutes) from point of purchase to point of sale during last trip (all crops and perishable crops)	Immediate (and additional changes over 6 months to 1 year as transportation methods change)
Transportation cost	• Village leader survey • Trader survey	• Average transportation cost (CFA) for 100kg bag of sorghum from village to two nearest markets in each season • Trader's transportation cost (CFA) for most recent trip (all crops and perishable crops)	Immediate for traders and 1–2 years for farmers
Transportation cost per kilometer to market	• Village leader survey	• Average transportation cost (CFA/km) to transport 100kg sac of sorghum to two nearest markets in each season	1–2 years for farmers

Indicator	Data source	Definition	Estimated exposure period[a]
Value of crops transported by traders	• Trader survey	• Value (CFA) of crops transported by trader on most recent trip (all crops and perishable crops)	n.a.
Road quality			
Perceptions of road quality	• Village leader survey • Trader survey • Trader survey	• Indicator for roads from the village to the nearest two markets are poor or extremely poor • Indicator for quality of RN7, RN35, or Rural Route Sambera is poor or extremely poor • Indicator for quality of roads used on most recent trip is poor or extremely poor	Immediate
Road quality affects which crops are produced by farmers	• Village leader survey	• Indicator for the quality of roads to markets affects which crops are produced by village farmers	n.a.
Road quality affects which crops are bought by traders	• Trader survey	• Indicator for the quality of the RN7, RN35, or Rural Route Sambera affects which crops the trader buys	n.a.
Road quality affects quality of crops sold by farmers	• Village leaders survey	• Indicator for the quality of roads from village to nearest two markets affects the quality of crops sold	Immediate to 6 months
Crops sold			
Trader transports perishable crops	• Trader survey	• Indicator for whether a trader transports perishable crops (fruits and vegetables including potatoes, onions, tomatoes, peppers, lettuce, and cabbage)	n.a.
Locations where trader sells crops	• Trader survey	• Indicator for whether a trader sells crops within the Gaya department, outside the Gaya department but within the Dosso region, and/or outside of Niger	n.a.
Crops damaged in transit			
Loss of or damage to perishable crops in transit	• Trader survey	• Indicator for trader experienced loss or damage to perishable crops in transit during most recent trip	Immediate
Percentage of perishable crops lost or damaged in transit	• Trader survey	• Percentage of trader's perishable crops lost or damaged in transit during most recent trip	Immediate
Decreased crop sales price	• Trader survey	• Indicator for decreased sales price due to damage to crops in transit during most recent trip (among those who experienced damages or losses)	Immediate to 6 months
Percentage decrease in crop sales price	• Trader survey	• Percentage decrease in sales price due to damage to crops on RN7, RN35, or Rural Route Sambera (all crops and perishable crops)	Immediate to 6 months

[a] Exposure periods were estimated based on a review of several pieces of literature on the impacts of road improvements on transportation and economic activity (Harris et al. 2020a; Harris et al. 2020b; Fortson et al. 2015).

n.a. = not available.

At the time of these surveys, work on the three roads being rehabilitated under the Roads for Market Access activity (the RN7, RN35, and RRS) was already in progress.[9] Therefore, the outcomes related to the Roads for Market Access activity measured by these surveys might already reflect early effects of the project activities.

C. Household-, plot- and crop-level outcomes at baseline

In this section, we report descriptive statistics from the household survey for primary and secondary outcomes prior to the intervention for the *Basse Terrasse* households in the MCG sample. These results enable us to lay the groundwork for addressing the core impact questions at the heart of the quantitative MCG analysis (presented in **Table I.1**) in the future.

In subsection C.1 we introduce the households in the *Basse Terrasse*, describe their poverty status relative to the country as a whole, and their levels of female empowerment. The following subsections investigate constraints faced by these households from various agricultural inputs—land, irrigation, fertilizer, seeds, credit—to agricultural results, overall household income and women's empowerment. In each subsection, we first provide a link to the logic model, then discuss the outcomes of *Basse Terrasse* households and the constraints to higher incomes these households might face. Finally, we discuss whether there are large imbalances between *Basse Terrasse* households and their matched counterparts, and if so whether these can be accounted for in the final impact evaluation through inclusion of appropriate covariates or whether these imbalances are too large to reliably remove potential biases.

C.1. Household demographics

Households in the *Basse Terrasse* are predominantly male-headed and larger than the average Nigerien household (**Table II.**). The average household has between 9 and 10 household members, 5 of whom are children under 16 years of age; this is larger than the average household size in Niger (5.5 members) or the Dosso region (5.7 members) (LSMS 2014).

Basse Terrasse households' poverty status appears to be about average compared to the country. The average poverty score[10] among *Basse Terrasse* households is 28.85. This score translates to, on average, a 39-40. percent likelihood of households falling below the 2011 national poverty line and means that households have a 14 percent likelihood of falling in the poorest half of Nigeriens (Schreiner 2018. See also the detailed discussion of income and poverty in Sections C.7 and C.8).

Most women remain unempowered, defined as lacking agency and autonomy over critical parts of life, including production, resources, income, and leadership. To measure empowerment, we adapted questions from the Women's Empowerment in Agriculture Index (WEAI) and administered them to the woman who makes the most important decisions in the household. The average adequacy of *Basse Terrasse* women is 52 percent (of 100 percent) with only 14 percent of *Basse Terrasse* women achieving

[9] Construction on the RN35 and RRS began in the third quarter of 2020, and work on the RN7 began in January 2021. As of the writing of this report in October 2022, work on all three roads was ongoing.
[10] The poverty score was calculated using the Scorocs Simple Poverty Scorecard for Niger (Schreiner 2018), which aggregates nine poverty indicators to estimate consumption-based poverty rates. These nine components are (1) region, (2) number of household members, (3) number of rooms in house, (4) roof construction material, (5) type of toilet, (6) main source of lighting, (7) ownership of a lounge chair, (8) ownership of a cell phone, and (9) ownership of a bicycle, motorcycle, or private vehicle.

empowerment. Women have a multidimensional empowerment score of 63 percent (of 100 percent). We present information on indicators of women's empowerment in more detail in **Table II.8.**

In terms of balance between *Basse Terrasse* and comparison households, the matching procedure we used was successful in selecting households that are similar along demographic characteristics. This is reflected in the small standardized differences as well as the high *p*-values, which indicate that there are no precisely estimated differences between the two groups.

Table II.6. Household demographics

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Age of HH Head	44	44	-0.01	0.88
Head of HH reads or writes (0/1)	0.51	0.45	0.12	0.19
Female head of HH (0/1)	0.06	0.09	-0.08	0.41
Number of HH members	9.37	9.34	0.01	0.94
Number of adults in HH (age 16+)	4.17	4.34	-0.07	0.44
Number of children in HH	5.46	5.23	0.05	0.52
Poverty score	28.85	29.31	-0.04	0.63
Women's empowerment score+ (%)	0.52	0.57	n.a.	n.a.
Sample size	**268**	**490**		

Notes: + = Score is calculated for either treatment or comparison group as 1 – (weighted percentage of households in group that are unempowered * weighted mean adequacy score among unempowered in group). Sample sizes shown are for the largest sample. The following indicators are top-coded at the 99th percentile: Number of HH members, Number of children in HH.

HH = household; n.a = not applicable.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

C.2. Land holding, land use, and land tenure security

A key component of the IMAP's program logic is the increase in (secure) access to irrigated land that allows for cultivation in the rainy and dry seasons and the switch to higher-value crops. **Table II.2** presents information on household total and irrigated land holdings, the size of the matched plot, and tenure security.

BT households, on average, own or cultivate two and one-quarter plots, with a total area of 1.61 hectares.[11] These plots are almost always owned and cultivated. About one-third of a hectare of this land is irrigated, the majority of plots (67 percent) are obtaining water for irrigation from wells. Other sources of irrigation include canals (21 percent), river (10 percent), and flooding (3 percent), which would include surface water or swamp areas. The average self-reported SSI target plot size is about 1.6 hectares of land, while GPS measurements of plot outlines indicate a more modest size of one hectare. (Appendix D further compares the self-reported area and the GPS measurements). On average, SSI target plots thus make up about three-fifths of a household's total land holdings.

Fewer than 5 percent of *BT* plots have formalized land rights. Despite this low level of formalized rights, self-reported land disputes are rare. Only about 1 percent of *BT* plots experienced a land dispute in the past year. Despite the lack of formal documentation for most plots, land disputes and concerns over the

[11] FEWS NET reports that households in Southern Dosso region have access to approximately 1.5 hectares of land (FEWS NET 2019).

involuntary loss of land are rare (1–2 percent). Although recent disputes are few, owners of 10 percent of plots in our sample, in both *BT* and the comparison area, fear they are at risk of involuntary loss within the next five years.

Plot rentals are rare in the Basse Terasse. Only 1 percent of plots are cultivated or owned by someone outside the household. There are only slight differences between the dry and rainy seasons in the share of plots either owned or cultivated by someone outside the household.

Overall, the SSI target households and their comparison counterparts appear balanced with respect to landholdings, the target plot sizes, and tenure security, with one exception. The imbalance in the total landholdings is 0.32 SD. Although not negligible, the imbalance in irrigated landholdings, which is the variable that would be of primary concern, is less than the 0.25 SD cutoff that the literature suggests can be addressed by including control covariates in the impact regression.

Table II.2. Landholdings, matched plots and land tenure security

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Household landholdings				
Number of plots	2.28	2.26	0.02	0.84
Number of plots cultivated	2.21	2.22	-0.01	0.92
Number of plots owned	2.27	2.24	0.03	0.77
Total HH landholdings (ha)	1.61	2.09	-0.32	0.00***
Total irrigated landholdings in dry season (ha)	0.28	0.38	-0.12	0.34
Household sample size	268	490		
Plot				
Size of matched plot – self-report (ha)	1.59	1.42	0.26	0.00***
Size of matched plot – plot measurement (ha)	1.54	1.69	-0.13	0.19
Plot tenure				
Formalized land rights for plot (0/1)	0.04	0.07	-0.13	0.25
Experienced land dispute in last year over plot (0/1)	0.01	0.00	0.05	0.56
Perceived risk of involuntary loss of plot in next 5 years (0/1)	0.10	0.15	-0.18	0.05**
Plot cultivated by someone outside the household in last year (0/1)	0.01	0.00	0.11	0.16
Plot owned by someone outside the household (0/1)	0.01	0.02	-0.06	0.42
Plot sample size	337	730		

Notes: Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values. All continuous indicators are top-coded at the 99th percentile.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

Table II.8 presents information on cultivation and cropping patterns in each season separately for the *Basse Terrasse* and comparison groups. Dry season cultivation is rare (9 percent of *BT* plots, and 13 percent of comparison plots) relative to rainy season cultivation, confirming the need for access to irrigation in the dry season through the IMAP. Almost all plots (96 percent of *BT* plots, 96 percent of comparison plots) were cultivated in the rainy season.

Table II.8. Land use and crop choice

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Dry season (Oct 2020–May 2021)				
Plot was cultivated (0/1)	0.09	0.13	-0.13	0.21
Plot area cultivated (ha)	0.07	0.08	-0.02	0.86
Rice grown on plot (0/1)	0.06	0.07	-0.02	0.79
Cassava grown on plot (0/1)	0.00	0.03	-0.26	0.11
Plot sample size	337	730		
Rainy season (June–Sept 2021)				
Plot was cultivated (0/1)	0.96	0.96	-0.01	0.90
Plot area cultivated (ha)	1.39	1.57	-0.17	0.12
Plot was inter-cropped (0/1)	0.48	0.47	0.01	0.91
Millet grown on plot (0/1)	0.77	0.80	-0.08	0.33
Cowpea grown on plot (0/1)	0.38	0.41	-0.07	0.44
Sorghum grown on plot (0/1)	0.32	0.25	0.17	0.10*
Rice grown on plot (0/1)	0.15	0.15	0.01	0.87
Plot sample size	336	729		

Notes: Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values. All continuous indicators are top-coded at the 99th percentile. Crops that were grown by less than 2 percent of treatment or comparison plots are not reported.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

Crops grown in the dry season are a subset of the crops grown in the rainy season. Dry season cultivation on the small number of plots cultivated consists primarily of rice, whereas households use the rainy season to grow traditional crops (sorghum, millet, and cowpea) and rice. The dry season thus presents the greatest opportunity for income-generating cultivation since 90 percent of plots are not cultivated. Close to half of rainy season plots (48 percent in *BT*, 47 percent in the comparison areas) were intercropped, which is why the summed share of plots growing millet, cowpea, rice, or sorghum exceeds 100 percent in both regions. Crops not displayed in **Table II.8** each account for no more than 1 percent of cultivated plots, such as cowpea (1 percent) and sorghum (1 percent) in the dry season, and corn, cassava, and sesame (each 1 percent) in the rainy season.

Most farmers in the *Basse Terrasse* thus grow traditional food crops, and about 15 percent grow some rice. Few, however, grow the higher-value cash crops that would justify SSI investments. This suggests the relevance in the logic model of the planned training activities to support the shift to higher-value crops.

When we investigate differences between the *Basse Terrasse* households and the comparison group, only one indicator in this subsection appears somewhat imbalanced: the proportion of plots on which sorghum was grown in the dry season. This imbalance is less than the 0.25 SD cutoff that the literature suggests can be addressed by including control covariates in the impact regression.

C.3. Irrigation use and availability for cultivated plots

Although most plots were not cultivated during the dry season, most of those that were cultivated used irrigation (69 percent). During the rainy season, about 12 percent of plots used irrigation. **Table II.9** provides summary statistics on irrigation use and availability in each season.

The fact that few plots were cultivated in the dry season (9 percent, see **Table II.8**) suggests that there might be considerable room for improved access to irrigation in the dry season.

There is a significant imbalance between treatment and comparison plots that are cultivated in the dry season regarding whether irrigation was always available when needed; this imbalance is over four times 0.25 SD recommended cutoff. As a result, it is unclear to what extent regression adjustment will successfully address potential biases in the analysis of dry season practices *when conditioning on baseline dry season cultivation*. However, because the primary effect of the investments in small scale irrigation will be the increase in the number of plots that are irrigated—what is known as the extensive margin—the pre-existing differences in the small number of plots that were irrigated at baseline will have very little effect on differences in the final evaluation.

Table II.9. Use and availability of irrigation

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Dry season (Oct 2020–May 2021)				
Used irrigation, conditional on cultivation (0/1)	0.69	0.76	-0.15	0.66
Irrigation always available when needed, conditional on using irrigation (0/1)	0.99	0.60	1.10	0.01***
Plot sample size	36	84		
Rainy season (June–Sept 2021)				
Used irrigation, conditional on cultivation (0/1)	0.12	0.19	-0.18	0.15
Irrigation always available when needed, conditional on using irrigation (0/1)	0.81	0.71	0.23	0.46
Plot sample size	314	699		

Note: Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values or conditionality.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

C.4. Fertilizer

In addition to enhanced access to irrigation, the program logic anticipates farmers in the *Basse Terrasse* using improved technologies and inputs, including increased use of fertilizers.

Although use of inorganic and organic fertilizer is widespread, the amount of inorganic fertilizer applied is low. **Table II.30** shows rates of inorganic and organic fertilizer application in each season. A large majority of cultivated *Basse Terrasse* plots used inorganic fertilizer: 84 percent in the dry season and 68 percent in the rainy season. Most farmers also used organic fertilizer in the rainy season, when 69 percent of cultivated plots used organic fertilizer. Higher quantities of inorganic fertilizer per hectare were applied in the dry season: 0.41 t/ha in the dry season and 0.12 t/ha in the rainy season. In contrast, for organic fertilizer, vastly larger quantities per hectare were applied on the perimeter in the rainy season (3.75 t/ha) compared to the dry season (0.39 t/ha).

The current low use of inorganic fertilizer application in the *Basse Terrasse* suggests that project activities aimed at increasing fertilizer use would improve yields.

Although plots in both treatment and comparison households are managed with somewhat similar practices, comparison households used more organic fertilizer per hectare relative to treatment households: 4.54 t/ha for comparison households compared to 3.75 t/ha for treatment households.

Table II.3. Fertilizer

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Dry season (Oct 2020–May 2021)				
Organic and/or inorganic fertilizer applied, conditional on cultivation (0/1)	0.87	0.78	0.24	0.49
Inorganic fertilizer applied (0/1)	0.84	0.72	0.30	0.37
Quantity of inorganic fertilizer applied (t/ha)	0.41	0.25	0.37	0.06*
Organic fertilizer applied (0/1)	0.08	0.34	-0.66	0.06*
Quantity of organic fertilizer applied (t/ha)	0.39	0.90	-0.18	0.37
Plot sample size	38	85		
Rainy season (June–Sept 2021)				
Organic and/or inorganic fertilizer applied, conditional on cultivation (0/1)	0.91	0.92	-0.06	0.49
Inorganic fertilizer applied (0/1)	0.68	0.69	-0.02	0.84
Quantity of inorganic fertilizer applied (t/ha)	0.12	0.09	0.16	0.06*
Organic fertilizer applied (0/1)	0.69	0.75	-0.14	0.12
Quantity of organic fertilizer applied (t/ha)	3.75	4.54	-0.14	0.10*
Plot sample size	317	699		

Note: Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values. Indicators for the quantity of organic fertilizer applied per hectare in the rainy and dry season are top-coded at 20 t/ha, based on input from a Niger-based consultant. Indicators for the quantity of inorganic fertilizer applied per hectare in the rainy and dry season are top-coded at the 95th percentile.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

C.5. Seeds and improved agricultural practices

The IMAPs program logic anticipates that farmers will use improved farming practices and plant improved seeds to complement the increased access to irrigated land and use of fertilizer.

In **Table II.4** we present information on the types of seeds sown and the prevalence of improved inputs or agricultural practices, including improved water and soil management techniques. Farmers sowed few cultivated plots with purchased seeds, with rates very similar in the small number of plots cultivated in the dry season (26 percent) and those cultivated in the rainy season (23 percent). Farmers sowed an even smaller proportion of plots (21 percent dry season, 20 percent rainy season) with improved open-pollinated or hybrid seeds.

Few cultivated plots (19 percent) are cultivated using an improved water and soil management technique. The different types of improved water and soil management techniques are zaï, tassa, agricultural half-moon, fences, stone walls, silviculture benches, and adding lime to soil.[12] Almost all plots (97 percent) were cultivated with at least one improved input or practice (the most common practice, used by over 91 percent, is row planting of seeds, though the average number of applied inputs or practices of 2.65 out of 9 was low). The nine categories of improved inputs or practices are zero tillage land preparation, planting seeds in rows, improved open pollinated or hybrid seeds, improved water and soil management techniques (detailed above), mechanized equipment, inorganic fertilizer, pesticides or herbicides, processing crops after harvest, and storing crops in hermetic bags.

The very low number of improved inputs and practices by *Basse Terrasse* farmers suggest significant scope for the IMAP's farmer training activities to increase agricultural productivity.

In assessing *Basse Terrasse* comparison plot differences, we find several that are large relative to baseline means and SD. In particular, *Basse Terrasse* households use a slightly larger number of improved inputs and practices. Because the baseline number of improved practices—2.65 of 9—is very low, we anticipate that the increase in the usage of best practices among *Basse Terrasse* households benefiting from training will dwarf any possible biases that arise from the baseline imbalances.

Table II.4. Seeds and improved agricultural practices

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Dry season (Oct 2020–May 2021)				
Sowed purchased seeds, conditional on cultivation (0/1)	0.26	0.41	-0.32	0.36
Share of plot area with purchased seeds (%)	0.21	0.17	0.12	0.74
Sowed improved seeds (0/1)	0.35	0.17	0.41	0.22
Plot sample size	38	85		
Rainy season (June–Sept 2021)				
Sowed purchased seeds (0/1)	0.23	0.23	0.01	0.93
Share of plot area with purchased seeds (%)	0.20	0.21	-0.02	0.84
Sowed improved seeds (0/1)	0.15	0.12	0.07	0.56
Plot sample size	317	698		
Growing year				
Applied improved inputs or practices (0/1)	0.97	0.99	-0.16	0.08*
Applied improved water and soil management techniques (0/1)	0.19	0.10	0.25	0.01***
Number of improved inputs or practices (out of 9)	2.65	2.32	0.29	0.00***
Plot sample size	329	716		

Notes: Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values. The different types of improved water and soil management techniques are zaï, tassa, agricultural half-moon, fences, stone walls, silviculture benches, and adding lime to soil. The nine categories of improved inputs or practices are zero tillage land preparation, planting seeds in rows, improved seeds, improved water and soil management techniques, using mechanized equipment, applying

[12] Zaï, tassa, and agricultural half-moon are agricultural techniques that involve digging pits in the soil prior to planting to accumulate water. Fences, stone walls, and silviculture benches reduce soil erosion by managing water flow. Adding lime to soil makes the soil less acidic, which helps improve the availability of nutrients for crops.

inorganic fertilizer, applying pesticides or herbicides, processing crops after harvest, and storing crops in hermetic bags.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

C.6. Credit and expenditures

The inability to invest in agricultural inputs can be a key constraint to improving agricultural outcomes and might depend on farmers' lack of access to credit.

Although many parts of Niger lack adequate access to credit, accessing credit in the *Basse Terrasse* is common. In **Table II.5** we provide descriptive statistics on household access to credit, the amounts and number of loans taken out, and total household agricultural expenditures in the past year. Most households could access credit (61 percent), and half took out a loan (46 percent) in the past year. Plots were rarely used as collateral for loans.

Expenditures households incur for agricultural production include expenses for irrigation, fertilizer, seeds, labor, animals, equipment, pesticides/herbicides, canal cleaning, preparation of crops for sale, and transport for sale. Total average annual agricultural expenditures were 75,500 FCFA (about $136 USD[13]) across all households. This represents roughly 25 percent of the value of crop sales (see **Table II.7**). Expenditures are dominated by labor, fertilizer, and seeds; expenses for irrigation are low in comparison. Per hectare costs follow a similar pattern.

In reference to the logic model, it does not seem that lack of credit options are primary constraints for farmers in the *Basse Terrasse* to increase agricultural investments.

In terms of balance, there are imbalances in total annual expenditures for target plots—primarily coming from higher fertilizer expenditures in the *Basse Terrasse*. The per-hectare costs (the more relevant measure for the evaluation as planned irrigation area from the SSI infrastructure is 1 hectare) show smaller imbalances.

Table II.5. Credit and expenditures

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Credit (household)				
Household can access credit (0/1)	0.61	0.51	0.20	0.03**
Loan taken out in the past year (0/1)	0.46	0.35	0.22	0.01***
Total value of loan(s) taken out in the past year, not conditional on borrowing (FCFA)	36,423	32,815	0.04	0.70
Household sample size (credit)	265	486		
Collateral (target plots)				
Used plot as collateral for credit in past dry or rainy season (0/1)	0.04	0.03	0.08	0.40
Would consider using plot as collateral (0/1)	0.02	0.01	0.08	0.34
Plot sample size	334	719		

[13] Conversions between FCFA and USD are based on a historical average exchange rate of 0.0018 FCFA per USD for the period October 2020 to September 2021, which covers the past dry and rainy season we asked about in our household survey (Exchange Rates UK 2022a, 2022b).

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Annual expenditures (target plots)				
Total agricultural expenditures (FCFA)	75,589	49,580	0.31	0.00***
Irrigation expenditures (FCFA)	1,454	2,593	-0.08	0.41
Fertilizer expenditures (FCFA)	45,774	28,364	0.31	0.01***
Seed expenditures (FCFA)	3,845	6,902	-0.17	0.06*
Labor expenditures (FCFA)	10,989	5,742	0.22	0.02**
Preparation and processing expenditures (FCFA)	1,468	1,209	0.04	0.64
Transportation expenditures (FCFA)	554	272	0.10	0.35
Animals or mechanized equipment expenditures (FCFA)	7,287	3,192	0.26	0.01***
Pesticides or herbicide expenditures (FCFA)	3,468	2,761	0.09	0.55
Cost estimate for unpaid household labor[a]				
Plot sample size	336	711		
Annual per-hectare costs (target plots)				
Annual irrigation cost per hectare, dry and rainy seasons (FCFA/ha)	692	1,616	-0.11	0.15
Annual fertilizer cost per hectare, dry and rainy seasons (FCFA/ha)	31,388	21,980	0.29	0.00***
Annual seed cost per hectare, dry and rainy seasons (FCFA/ha)	3,116	3,244	-0.01	0.88
Annual labor cost per hectare, dry and rainy seasons (FCFA/ha)	9,262	5,548	0.16	0.03**
Annual preparation and processing cost per hectare, dry and rainy seasons (FCFA/ha)	1,019	1,090	-0.02	0.85
Annual transportation cost per hectare, dry and rainy seasons (FCFA/ha)	307	196	0.07	0.37
Annual animals or mechanized equipment cost per hectare, dry and rainy seasons (FCFA/ha)	5,828	2,790	0.25	0.00***
Pesticide or herbicide cost per hectare, dry and rainy seasons (FCFA/ha)	2,988	1,403	0.32	0.00***
Plot sample size	322	683		

Notes: Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values. Expenditure sources are top-coded on a per-hectare basis based on input from a Niger-based consultant. The indicator for the total value of loan(s) taken out in the past year is top-coded at the 95th percentile. Expenditures and costs per hectare for irrigation, fertilizer, seeds, labor, preparation and processing, transportation, animals or mechanized equipment, and pesticides or herbicides are not conditional on using the specified input.

[a] The cost estimate for unpaid household labor is a self-reported estimate from the plot decision maker based on their estimate of how much it would have cost to hire non-household members to perform the unpaid labor household members performed.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

C.7. Agricultural productivity and profitability

To justify the IMAP investment in small-scale irrigation, the value of production on fields benefiting from SSI must increase. This can be achieved by cultivating a larger area of land, obtaining a second harvest during the dry season, shifting to more profitable crops, increasing agricultural productivity, or achieving higher prices and thus incomes from crop sales. Previous sections have presented information on area cultivated in both seasons and crop choice which are the primary channels through which *Basse Terrasse* farmers' incomes are expected to increase; in this section, we describe baseline levels for yields and crop sales. We note that because crop choice is expected to change significantly with SSI, these baseline levels are more useful in understanding how farmers' outcomes are influenced by the current context rather than providing a useful reference point to study changes over time.

Table II.6 presents yields (t/ha) and per-hectare income (FCFA/ha) for the crops that are most commonly grown in each season. In the dry season, rice is the most frequent crop grown on the small number of target plots cultivated in the dry season in the *Basse Terrasse* (31 of 36 plots) and comparison areas (50 of 64 plots). Millet, cowpeas, rice, and sorghum are common in the rainy season.

In

Table II.6 we present yields from pure-stand and intercropped plots. Income was calculated as revenue from crop sales plus the estimated value of own consumption net of agricultural expenditures. Income per hectare and yields are calculated by dividing by the plot area cultivated with a crop to obtain per (cultivated) hectare values. Estimates for rice plot yields in the dry season are imprecisely estimated, with a sample of 31 plots. Income was calculated as revenue from crop sales plus the estimated value of own consumption net of agricultural expenditures. Income was then divided by total area cultivated to calculate income per hectare.

Rice in both seasons earned the highest income per hectare, followed by sorghum. Millet seems to have low income per hectare relative to rice, sorghum, and cowpeas. To provide some, albeit imperfect, context for how high or low per-hectare yields and incomes in the *Basse Terrasse* are, we can compare these outcomes to results obtained in Konni that we measured in for the 2018 rainy and 2018/2019 dry seasons. Although the rainy season yield and income per hectare for the small number of sorghum pure-stand plots is about as high as in Konni, for millet and cowpeas, yields are about half those in Konni, and income per hectare is 27 and 51 percent the income obtained by farmers on the Konni perimeter (there is no rice planting in Konni). Dry season incomes are dwarfed by results obtained in Konni on fields growing high-value crops (primarily tomatoes, anise, cabbage, onion, and wheat) where incomes range between 1 and 4 million CFA per hectare.

Comparison yields and incomes are lower than *Basse Terrasse* yields and incomes. For the small number of fields with rice cultivation, the imbalances are too large to be addressed fully by including covariates in the impact estimations. For cowpeas grown in the rainy season, the imbalances for yields and on incomes also appear very large. **Table D.2** presents results that exclude inter-cropped plots..

Table II.6. Crop yield, income, and income per hectare

Indicator	Treatment mean	Comparison Mean	Standardized difference	*p*-value
Yield in dry season (Oct 2020–May 2021) (t/ha)				
Rice	2.77	1.63	0.78	0.01***
Income per hectare in dry season (Oct 2020–May 2021) (FCFA/ha)				
Rice	487,641	276,661	0.39	0.06*
Plot sample size	31	50		
Yield in rainy season (t/ha)				
Millet	1.54	1.58	-0.03	0.76
Rice	2.14	1.52	0.41	0.04**
Cowpeas	0.75	0.39	0.34	0.01***
Sorghum	1.39	1.40	-0.01	0.95
Income per hectare in rainy season (FCFA/ha)				
Millet	143,155	160,021	-0.11	0.25
Rice	252,389	229,733	0.08	0.70
Cowpeas	102,490	27,317	0.38	0.00***
Sorghum	135,228	162,574	-0.13	0.48
Plot sample size	313	674		

Notes: Sample sizes shown are at the plot level, but individual crops have smaller sample sizes. Because sample sizes are small, some crop yields and incomes per hectare are not reported. Yields and income per hectare for cowpea, sorghum, and millet include both pure-stand and intercropped cropping patterns. For intercropped cropping systems, subplot areas are allocated based on the ratio of the average partial land equivalence ratio (LER) for a crop intercropping system relative to the average LER for the whole cropping system from Namatsheve et al. (2020). Pure-stand crop yields are top-coded based on INRAN yield potentials. To top-code intercropped crop combinations, INRAN yield potentials are adjusted by the average partial LERs from Namatsheve et al. (2020). Indicators for crop income per hectare are top-coded at the 95th percentile.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

C.8. Household income, sales, and profits

The ultimate objective of MCC's investments is to increase household incomes and food security through increases in agricultural productivity and sales. In this section, we describe baseline levels of agricultural sales, agricultural profits, and household income. We put baseline household income into perspective by contrasting it with national and international poverty lines.

Table II.7 shows the breakdown of income and profit across sources, including different types of crops and non-agricultural sources, as well as how crop sales vary by season.

The average household income per person of *Basse Terasse* households remains significantly below the global poverty line of $3.10 USD per person per day. Total average annual household income from agricultural and non-agricultural sources (including the estimated value of own consumption) was 547,608 FCFA (about $1,000 USD), which translates to 157 FCFA ($0.29 USD) per person per day for the average household size of 9.5 members in our household data. The 2014/2015 consumption-based poverty line for rural Dosso Niger (the region in which the *Basse Terasse* is located) was 431 FCFA ($0.79 USD) per person per day (Schneider 2019). This translates to a poverty line of about 1,500,200

FCFA ($2761 USD) per household per year for the average household on the Konni perimeter. Therefore, average household income is about 37 percent of the rural poverty line in Dosso.

Basse Terrasse households consume about half their agricultural production while selling the other half. (To estimate the value of crops for own consumption, we impute sales prices for the share of the harvest that is consumed by the household itself). About 40 percent of their total agricultural sales come from target plots.

Non-agricultural income through employment and self-employment activities is also an important source of household income. Averaging close to 150,000 FCFA ($270USD) annually, it accounts for roughly 28 percent of total household income. Thirteen percent of households engaged in non-agricultural employment, and 52 percent engaged in self-employment during the past year.

While crop sales from the rainy season generate incomes about as high for *Basse Terrasse* households as for comparison households, *Basse Terrasse* households derive twice as much revenue from dry season sales than comparison households. This imbalance, and the resulting imbalance in total sales, stands out in assessing balance along overall economic dimensions. The imbalance in dry season sales is driven by the irrigated rice plots that some *Basse Terrasse* households cultivate in the dry (and also rainy) seasons. These can include plots on the lists we received from SONED as target plots for small-scale irrigation and plots that are not on the lists. We are unlikely to fully address the issue of pre-existing dry season revenue through the inclusion of baseline covariates alone. To address these imbalances, we propose to implement differences-in-differences estimation combined with matching. This methodology subtracts the baseline value from future outcomes, such as those estimated based on endline data, and compares the changes over time between treatment and comparison households. This would account for higher baseline dry season revenue of treatment households.

Table II.7. Household income, revenue, and sales

Indicator	Treatment Mean	Comparison mean	Standardized difference	*p*-value
Annual				
Total agricultural and non-agricultural income, including own consumption (FCFA)	547,608	521,488	0.06	0.56
Agricultural income, including own consumption (FCFA)	416,615	379,013	0.09	0.33
Agricultural expenditures (FCFA)	122,254	71,636	0.40	0.00***
Value of crops for own consumption (FCFA)	310,212	297,444	0.04	0.63
Revenue from agricultural sales (FCFA)	247,052	152,623	0.28	0.00***
Non-agricultural income (FCFA)	150,829	149,964	0.00	0.97
Employment income (FCFA)	16,331	11,801	0.06	0.54
Self-employment income (FCFA)	136,625	134,290	0.01	0.92
Household sample size	263	483		
Dry season (Oct 2020–May 2021) crop sales				
Sold a crop (0/1)	0.29	0.25	0.08	0.41
Revenue from crop sales (FCFA)	124,716	51,918	0.34	0.00***
Revenue from crop sales of target plots (FCFA)	39,788	n.a.	n.a.	n.a.
Household sample size	267	483		

Indicator	Treatment Mean	Comparison mean	Standardized difference	*p*-value
Rainy season (June–Sept 2021) crop sales				
Sold a crop (0/1)	0.46	0.43	0.06	0.47
Revenue from crop sales (FCFA)	127,371	101,487	0.12	0.18
Revenue from crop sales of target plots (FCFA)	68,078	n.a.	n.a.	n.a.
Household sample size	267	482		

Notes: Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values. Indicators for agricultural income and revenue are top-coded at the 95th percentile. Expenditure sources are top-coded on a per-hectare basis at the plot level, based on input from a Niger-based consultant, before aggregating to the household level. Before calculating the value of crops for own consumption at the household level, the total quantity of crops set aside for household consumption is capped based on the maximum possible harvest for a crop based on INRAN yield potentials.

n.a. = not applicable.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

C.9. Household food security, poverty, and women's empowerment

IMAP's program logic also anticipates improved outcomes for several other socioeconomic indicators, including increased women's empowerment and assets, as well as improvements in terms of food security—the second goal of the program logic.

Table II.8 indicates the frequency with which households experience food insecurity[14] and household poverty. We asked about two periods during the past 12 months, the last month[15] and the last lean season, as well as households' expectations of food security in the upcoming 2022 lean season. The latter two food insecurity measures reflect the severity of food insecurity that may be experienced in the lean season before harvests in the rainy season.

About a quarter of households in the *Basse Terrasse* experience food insecurity. Among those experiencing some degree of food inadequacy, hunger, or extreme hunger, the majority rarely experienced food insecurity (defined as once or twice in the past month).[16] Only 1 percent of households experienced food inadequacy, hunger, or extreme hunger more than 10 times in the past month. We note that food insecurity in the previous lean season was substantially higher, particularly for the comparison group. Whereas 24 percent of comparison households experienced hunger in the past month, 44 percent did in the past lean season.

The results from the analysis of asset-based poverty scores show that *Basse Terrasse* households selected for the irrigation investments appear to be about average among households within Niger at large. As we show in **Table II.8**, the average poverty score among *Basse Terrasse* households is 28.85, which

[14] Food insecurity is defined as not having any food to eat in the household at some point in the past 30 days or four weeks, due to a lack of resources to get food.
[15] Because of the timing of the baseline survey between November, 2021 and March, 2022, the food insecurity questions correspond approximately to the period October, 2021 to February 2022.
[16] In comparison, Famine Early Warning Systems Network (FEWS NET) finds that 29 percent of households in Southern Dosso experience moderate-severe food security deficits based on the Integrated Food Security Phase Classification (IPC)—Chronic Classification (FEWS NET 2019). The IPC contains four severity levels from minimal/no chronic food insecurity (CFI) to severe CFI and spans a "common year" (IPC 2022). Using data from the most recent World Bank Living Standard Measurement Survey, Kafle and Balasubramanya (2022) estimate that 55 percent of Nigerien households experienced food insecurity in 2014 based on 7-day and 12-month recall periods.

represents a 39 to 40 percent likelihood of households falling below the 2011 national poverty line and a 14 percent likelihood of falling in the poorest half of Nigeriens (Schreiner 2018). (A higher poverty score indicates a lower likelihood of being poor).

In **Table II.8,** we also highlight some components of the poverty score. Six percent of households have an improved latrine or a flush toilet. Few to no *Basse Terrasse* households have electricity, and 26 percent have a motorcycle or private vehicle.

Overall, the comparison group seems to experience somewhat more frequent food insecurity in the lean season, with 44 and 31 percent of comparison households having experienced food inadequacy or extreme hunger compared to 30 and 19 percent among *Basse Terrasse* households. Five percent of comparison households have electricity, relative to 0 percent of treatment households. With the exception of food insecurity in the past lean season, these differences are below the 0.25 SD threshold and can be addressed adequately by regression adjustment.

Table II.8. Household food security and poverty

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Food insecurity in past month				
Food inadequacy (0/1)	0.24	0.32	-0.18	0.05**
Rare food inadequacy (1 or 2 times)	0.15	0.22	-0.18	0.04**
Sometimes food inadequacy (3–10 times)	0.08	0.10	-0.06	0.49
Often food inadequacy (more than 10 times)	0.01	0.00	0.10	0.24
Hunger (0/1)	0.21	0.24	-0.08	0.37
Rare hunger (1 or 2 times)	0.14	0.16	-0.04	0.62
Sometimes hunger (3–10 times)	0.06	0.08	-0.08	0.39
Often hunger (more than 10 times)	0.01	0.00	0.04	0.64
Extreme hunger (0/1)	0.15	0.22	-0.17	0.06*
Rare extreme hunger (1 or 2 times)	0.11	0.15	-0.11	0.22
Sometimes extreme hunger (3–10 times)	0.04	0.06	-0.09	0.30
Often extreme hunger (more than 10 times)	0.00	0.01	-0.13	0.13
Food insecurity in past lean season				
Hunger (0/1)	0.30	0.44	-0.30	0.00***
Rare hunger (1–2 times)	0.23	0.35	-0.29	0.00***
Sometimes hunger (3–10 times)	0.06	0.08	-0.06	0.54
Often hunger (more than 10 times)	0.01	0.01	0.01	0.90
Extreme hunger (0/1)	0.19	0.31	-0.28	0.00***
Rare extreme hunger (1–2 times)	0.19	0.31	-0.28	0.00***
Sometimes extreme hunger (3–10 times)	0.05	0.08	-0.11	0.22
Often extreme hunger (more than 10 times)	0.00	0.01	-0.05	0.53

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Food insecurity in upcoming 2022 lean season				
Expects hunger (0/1)	0.26	0.34	-0.18	0.07*
Expects rare hunger (1–2 times)	0.19	0.27	-0.20	0.05**
Expects sometimes hunger (3–10 times)	0.06	0.05	0.01	0.91
Expects often hunger (more than 10 times)	0.01	0.01	0.01	0.94
Expects extreme hunger (0/1)	0.23	0.30	-0.15	0.11
Expects rare extreme hunger (1–2 times)	0.17	0.24	-0.17	0.08*
Expects sometimes extreme hunger (3–10 times)	0.05	0.05	0.01	0.96
Expects often extreme hunger (more than 10 times)	0.01	0.01	-0.02	0.84
Poverty				
Poverty score (0–100)	28.85	29.31	-0.04	0.63
Improved roof materials (0/1)	0.34	0.28	0.13	0.14
Number of rooms	3.52	3.81	-0.14	0.10*
Improved toilet (0/1)	0.06	0.07	-0.03	0.75
Electricity (0/1)	0.00	0.05	-0.29	0.00***
Number of cell phones	1.79	1.76	0.02	0.78
Owns motorized transportation (0/1)	0.26	0.32	-0.13	0.15
Household sample size	282	285		

Note: Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

Increasing women's economic empowerment is an anticipated long-term outcome of IMAP's logic model. During the design and questionnaire development phases of the evaluation, MCC also highlighted an interest in understanding changes in women's empowerment more broadly beyond economic empowerment. To explore this, we adapted a commonly used tool to measure empowerment, the Women's Empowerment in Agriculture Index (WEAI). The WEAI methodology conceptualizes empowerment as agency and autonomy over critical parts of life, including production, resources, income, and leadership, and develops a set of questions that allow for the construction of a quantitative empowerment score. Because the cumulative respondent burden of administering the WEAI in conjunction with the modules designed to collect information on agricultural outcomes would have been too onerous, we shortened and adapted the WEIA modules.

Relative to a standard WEAI index, our empowerment score includes four out of five WEAI domains (production, resources, income, and leadership) but excludes time use. Because we ask only a female member of the household about empowerment, we cannot construct a gender parity measure. In addition, we have selected questions in the productive income, control over household income, resources, and leadership domains that are most relevant to the IMAP context and have adapted them to the local context as well.

Following the WEAI methodology (IFPRI 2012; Alkire et al. 2013), empowerment is defined as a weighted adequacy score of 80 percent or higher, indicating that women have achieved a weighted adequacy in at least 80 percent of a possible 100 percent of input indicators across the production, resources, income, and leadership domains. Domains are weighted so that they contribute equally (in our adaptation with four domains at 25 percent each) to the total adequacy score. Similarly, within each domain, input indicators are weighted so they contribute equally to the domain. These weights vary within each domain depending on the number of input indicators.

The vast majority of women in *Basse Terrasse* households remain unempowered. We characterize women's empowerment in **Table II.**, disaggregating important components of the empowerment score. Women in *Basse Terrasse* households have an empowerment score of 53 percent (out of a potential 100 percent). This score reflects that 13 percent of women are empowered across the four domains in the modified index (input into productive income, control over household income, resources, and leadership), 87 percent are unempowered, and unempowered women have an average adequacy score of 45 percent.

Fifty-two percent of women report having input into productive decisions, 63 percent report control over household income, 45 percent report adequacy with respect to resources (with only 24 percent making decisions about purchases, sales or transfers of assets), and 44 percent report leadership adequacy (with 28 percent reporting community group membership). These values suggest a strong opportunity for the IMAP to increase women's empowerment as a result of planned project activities.

Compared to their *Basse Terrasse* counterparts, women in comparison households have higher resource adequacy. This is due to higher ownership of assets, primarily higher female ownership of small livestock and poultry, and higher levels of financial decision making. While women's overall empowerment levels appear relatively similar, the large standardized differences in resource domain adequacy seem to require creating an MCG of women from the comparison areas that is better matched to treatment women along the WEIA subindices.

Table II.16. Women's empowerment in agriculture

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Women's Empowerment in Agriculture Index				
Women's empowerment score+ (%)	0.52	0.57	n.a.	n.a.
Empowered (%)	0.13	0.13	0.00	0.98
Unempowered (%)	0.87	0.87	-0.00	0.98
Adequacy score (0/1)	0.51	0.56	-0.18	0.05**
Adequacy score among unempowered (0/1)	0.45	0.51	-0.22	0.02**
Input into productive decisions (0/1)	0.52	0.56	-0.09	0.30
Control over household income (0/1)	0.63	0.71	-0.17	0.05**
Resource domain adequacy (0/1)	0.44	0.55	-0.32	0.00***
Ownership of assets (0/1)	0.57	0.75	-0.39	0.00***
Makes decisions about purchase, sale, or transfer of assets (0/1)	0.24	0.21	0.08	0.38
Makes decisions about borrowing or using credit (0/1)	0.52	0.70	-0.38	0.00***
Leadership domain adequacy (0/1)	0.44	0.42	0.04	0.65
Community group membership (0/1)	0.28	0.22	0.13	0.14
Comfortable speaking in groups or in public (0/1)	0.59	0.63	-0.08	0.40

Indicator	Treatment mean	Comparison mean	Standardized difference	*p*-value
Women's participation in agriculture				
Household has at least one female plot decision maker (0/1)	0.06	0.09	-0.10	0.34
Household has at least one female plot owner (0/1)	0.08	0.08	0.02	0.84
Participates in food crop farming (0/1)	0.34	0.15	0.45	0.00***
Participates in cash crop farming (0/1)	0.22	0.17	0.13	0.18
Participates in livestock raising (0/1)	0.29	0.36	-0.15	0.09*
Member of agriculture, livestock, or fisheries producer's group (0/1)	0.11	0.06	0.20	0.03**
Board member of agriculture, livestock, or fisheries producer's group (0/1)	0.06	0.02	0.19	0.05**
Household sample size	262	455		

Note: $^{+}$ = Score is calculated for either treatment or comparison group as 1 – (weighted percentage of households in group that are unempowered * weighted mean adequacy score among unempowered in group). Sample sizes shown are for the largest sample, but some variables might have smaller sample sizes as a result of missing values.

n.a. = not applicable.

Significantly different from zero at the *.10 level, **.05 level, ***.01 level, two-tailed test.

D. Road quality and market access

The Roads for Market Access Activity is expected to improve road conditions, leading to an increase in the presence of input and output traders, reduced transportation and vehicle operating costs, and reduced travel time to bring crops to markets, culminating in an increase in the volumes and values of crops produced and sold by farmers in the Dosso-Gaya region. To evaluate these outcomes at baseline, we surveyed market managers at six different markets in the Dosso-Gaya region accessed by the roads targeted for rehabilitation under the RMA, 36 traders (6 from each of the markets), and village leaders from 15 different villages (5 in the *Basse Terrasse*, 5 in Sambera, and 5 in Yelou) in March 2022.[17] Because the sample sizes for all three surveys are small, the results should be interpretated with caution. They provide descriptive evidence of the experiences and opinions of the individuals surveyed rather than conclusive evidence that is representative of traders and farmers in the Dosso-Gaya region.

We first describe the number and type of traders present at these markets and villages before describing perceptions of road quality, transportation times and costs for both farmers and traders, and damage to crops in transit, and the resulting effect on sales price among traders.

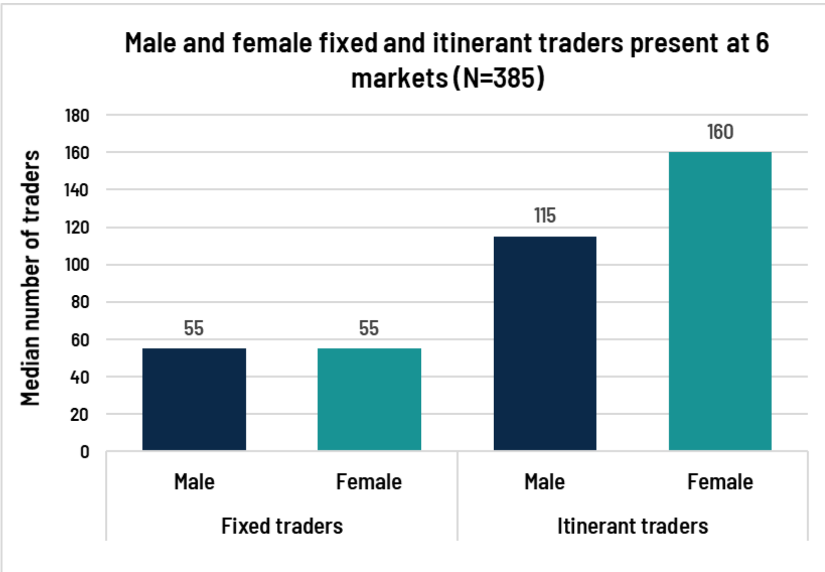
D.1. Trader presence in the Dosso-Gaya area

Figure II.6 presents the median number of male and female fixed and itinerant traders at six different markets in the Dosso-Gaya region: Béla, Fabirgui, Gaya, Malgorou, Saboula, and Tanda markets. The median number of total traders present at each market is 385. There are an equal median number of male and female fixed traders, and more female itinerant traders than male itinerant traders. **Figure II.7** presents the median number of traders selling different types of goods at each of these six markets.

[17] **Figure I.3** shows the location of the markets. To maintain respondent confidentiality, we do not indicate selected villages.

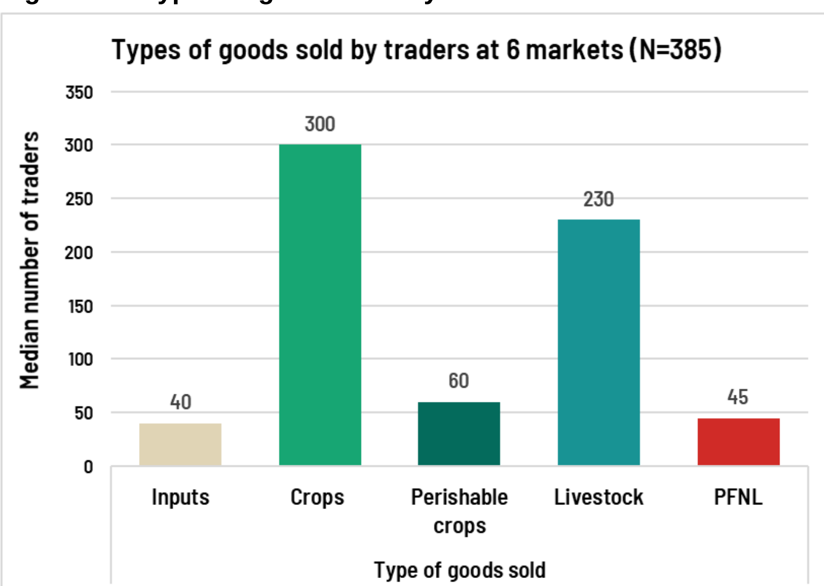
Traders deal in farming inputs, crops (including perishable crops[18]), livestock, and/or non-timber forest products (PNFL). Each trader can buy and sell more than one type of goods, and most traders buy and sell at least two types. Almost all traders buy and sell crops, and the majority also buy and sell livestock. A small minority buy and sell inputs or perishable crops.

Figure II.6. Number of male and female fixed and itinerant traders



Source: Surveys with six market managers.

Figure II.7. Types of goods sold by traders at markets



Source: Surveys with six market managers.

Note: One trader can deal in multiple types of goods. Perishable crops are defined as fruits and vegetables and do not include grains.

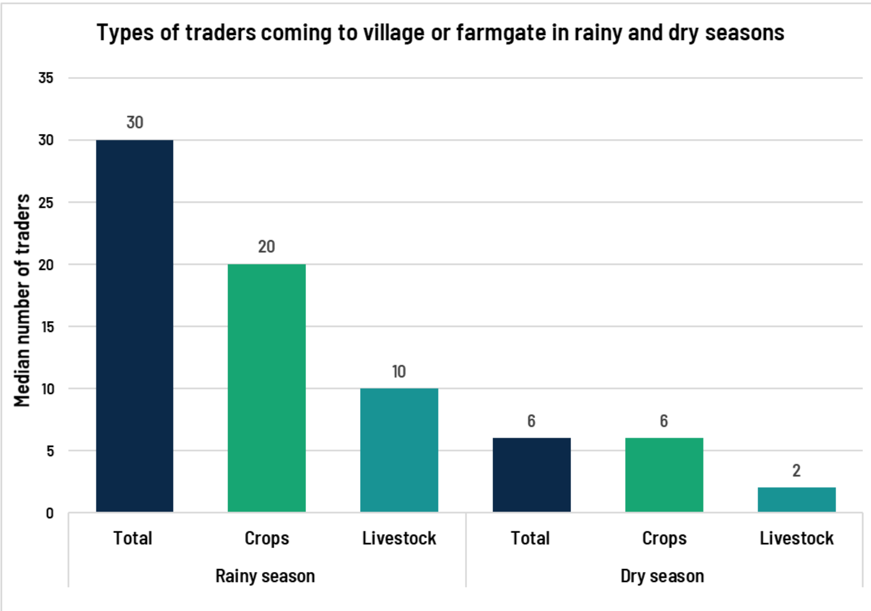
PFNL = non-timber forest products (*produits forestiers non ligneux*)

[18] Perishable crops are defined as fruits and vegetables and do not include grains. Surveyed traders of perishable crops reported transporting sweet potatoes, onions, peppers, tomatoes, lettuce, and cabbage.

D.2. Trader farmgate purchases and transportation of crops – trader survey

Traders that come directly to villages or farmgates to buy and sell goods deal exclusively in crops and livestock. As illustrated in **Figure II.8,** more traders come to villages during the rainy season than during the dry season. In both seasons, most traders come to buy crops, and fewer than half buy livestock. There were no traders selling farming inputs, such as fertilizer or seeds, directly at villages in either season (not shown).

Figure II.8. Types of goods bought by traders at villages in the rainy and dry seasons



Source: Surveys with 15 village leaders.

Note: One trader can buy both livestock and crops.

In **Table II.17** we present outcomes related to traders' transportation of crops in the Dosso-Gaya region at baseline in March 2022, after some rehabilitation activities to the RN7, RN35, and RRS were already under way. To transport crops, most traders we interviewed use a *dogonbaro* (a three-wheeled motorcycle with a bed similar to that of pickup truck) and a minority use a *camion* (pickup truck). None of the traders we surveyed owned their own vehicle (they all used vehicles owned by the markets or shared transportation), so we are unable to report on vehicle operating or maintenance costs. Most traders (77 percent) sell crops within the Gaya department, half sell crops outside the Gaya department but within the Dosso region, and only 14 percent sell crops outside Niger.[19]

Traders report that the quality of the RN7, RN35, and RRS is poor and results in damage to or loss of crops in transit. More than 80 percent of traders taking the RN7 and all traders taking the RRS rated the road quality as poor or extremely poor. The RN35 is slightly better: about half the traders using the RN35 to transport crops rated the road quality as poor or extremely poor.[20] It is important to note that of these three roads, construction on the RN35 started the earliest, in 2020, so the better quality of the RN35

[19] Traders sell crops in multiple locations, so these categorizations are not mutually exclusive.

[20] We did not specifically ask traders about road quality in the dry season versus the rainy season. Since we surveyed them in March 2020, the middle of the dry season, they might have been thinking primarily about road conditions in the dry season as they answered this question. Road conditions are generally worse in the rainy season than the dry season.

relative to the RN7 and RRS might reflect, at least partially, some of the improvements made to the road under the RMA. As a result of damage to or loss of crops in transit, most traders using the RN7, RRS, or RN35 reported obtaining sales prices for crops that were 14 to 19 percent lower than what they could otherwise have obtained absent damages and losses. Moreover, they reported that the quality of the road affects which crops they buy; in particular, the majority of traders reported not buying certain perishable crops because of concerns over damages and losses in transit.

On their most recent trip transporting crops, the average travel time for traders from point of purchase to point of sale was almost 3 hours, and the average transportation cost was 8,628 FCFA. The average value of crops transported was 60,625 FCFA.

About one-third of surveyed traders transport perishable crops. Among these traders, average travel time on their most recent trip was closer to 2 hours, transportation costs were lower, and the average value of crops transported was almost half as much: 37,591 FCFA. Eighty percent of these traders experienced crop loss or damage to perishable crops during their most recent trip, resulting in an average loss of 13.6 percent of the perishable crops and an average decrease in sales price of 18.9 percent among those experiencing losses. This suggests that the poor road quality might affect how far traders are willing to transport perishable crops and the volume of perishable crops traded. Taken together with reports of crop damage in transit and decreased sales price, these results suggest that perishable crops are a less-profitable commodity relative to non-perishable crops owing to road quality issues in the Dosso-Gaya region. This also helps explain why only about one-third of traders transport perishable crops.

Table II.17. Trader transportation and crops lost in transit

Indicator	Mean
Uses *camion* (pickup truck) to transport crops (%)	0.11
Uses *dogonbaro* (moto-tricycle with a bed like a pickup truck) to transport crops (%)	0.86
Transports perishable crops (0/1)	0.31
Sells crops within Gaya department (0/1)	0.77
Sells crops outside Gaya department but within Dosso region (0/1)	0.51
Sells crops outside of Niger (0/1)	0.14
Sample size – all traders	36
Road quality, RN7 between Dosso and Gaya	
Believes quality of RN7 is poor or extremely poor (%)	0.81
Quality of RN7 affects which crops trader buys (0/1)	0.69
Percentage decrease in sale price due to damage to crops on RN7 (%)	15.0%
Sample size – traders using the RN7	27
Road quality, RN35 between Margou and Gaya	
Believes quality of RN35 is poor or extremely poor (%)	0.47
Quality of RN35 affects which crops trader buys (0/1)	0.76
Percentage decrease in sale price due to damage to crops on RN35 (%)	18.9%
Sample size – traders using the RN35	19
Road quality, Rural Route Sambera (RRS)	
Believes quality of RRS is poor or extremely poor (%)	1.00
Quality of RRS affects which crops trader buys (0/1)	1.00
Percentage decrease in sale price due to damage to crops on RRS (%)	14.0%
Sample size – traders using the RRS	10

Indicator	Mean
Most recent trip, all crops	
Travel time from point of purchase to point of sale (minutes)	174
Transportation cost (cfa)	8,628
Total value of crops transported (cfa)	60,625
Believes quality of roads used on most recent trip is poor or extremely poor (%)	0.72
Sample size – all traders	36
Most recent trip, perishable crops	
Travel time from point of purchase to point of sale (minutes)	117
Transportation cost (cfa)	6,800
Total value of crops transported (cfa)	37,591
Experienced crop loss or damage in transit during most recent trip (%)	0.80
Percentage of crops lost or damaged in transit (%)	13.6%
Decreased price due to damage to crops (among those who experienced damage) (0/1)	1.00
Percentage decrease in sale price due to damage to crops (%)	18.9%
Sample size – traders buying and selling perishable crops	11

Source: Surveys with 36 traders. Sample sizes for some indicators are smaller owing to missing data.

Note: Perishable crops are defined as fruits and vegetables and exclude grains. Surveyed traders who trade perishable crops reported transporting sweet potatoes, onions, peppers, tomatoes, lettuce, and cabbage.

D.3. Farmer transportation of crops to market – village leader survey

Poor road quality also affects market access for farmers in the *Basse Terrasse*, Sambera, and Yelou (**Table II.18**). Almost all village leaders report that the quality of roads affects the quality of crops produced and sold and which crops farmers produce. According to village leaders, most farmers (70 percent) use animal-pulled carts to transport crops to market.

Although road quality is poor in both the rainy and dry seasons, the quality of roads from villages to the nearest markets is worse in the rainy season, resulting in longer travel times and higher transportation costs. In the rainy season, it takes longer and costs more to transport crops to nearby markets. Because of the limited number of traders coming directly to villages relative to the number of traders present at markets (see **Figure II.6** and **Figure II.7**), farmers' access to nearby markets greatly influences their ability to access inputs and get competitive sales prices for their crops.

Table II.18. Transportation and road quality from villages to markets

Outcome	Mean
Transportation to market (both seasons)	
Most common transport method used by farmers to transport crops to market is motorized vehicle (%)	0.23
Most common transport method used by farmers to transport crops to market is animal pulled cart (%)	0.70
Believes quality of roads affects the quality of crops produced and sold (0/1)	0.97
Believes quality of roads affects which crops farmers produce (0/1)	0.93
Rainy season	
Believes quality of roads to nearest market is poor or extremely poor	0.90
Travel time from village to nearest market (minutes)	118
Farmer transportation cost to get 100kg bag of sorghum to market (cfa)	690
Farmer transportation cost per km to get 100kg bag of sorghum to market (cfa/km)	60

Outcome	Mean
Dry season	
Believes quality of roads to nearest market is poor or extremely poor	0.17
Travel time from village to nearest market (minutes)	94
Farmer transportation cost to get 100kg bag of sorghum to market (cfa)	612
Farmer transportation cost per km to get 100kg bag of sorghum to market (cfa/km)	58
Sample size (village leaders)	*15*

Source: Surveys with 15 village leaders.

E. Discussion of findings

In this section we discuss the key findings from the baseline analysis and draw implications for the evaluation. The key findings focus on constraints to higher productivity and incomes at baseline and the potential of the IMAP activities to overcome these constraints.

Small-scale irrigation

- Farmers in the *Basse Terrasse* are primarily engaged in rainfed agriculture and grow traditional crops such as millet, sorghum and cowpeas. Intercropping is common in the rainy season with 40 percent of plots intercropped. Less than 10 percent of plots are cultivated in the dry season; most of these plots grow rice. Yields and revenue are low, reflecting the low use of best practices in input use and agricultural practices. Households are poor with an estimated 40 percent of households below the national poverty line. Thirty percent of households state having experienced hunger in the past lean season.
- Access to small-scale irrigation is a significant opportunity to increase farmer's agricultural production and incomes and to improve food security. However, at baseline few to no farmers cultivated cash crops such as tomatoes or onions; and only 10 percent of plots were cultivated in the dry season using irrigation. The low baseline prevalence of higher-value crop cultivation, use of best practices and irrigation access justifies the logic model's inclusion of training activities designed to support farmers to fully benefit from access to irrigation. Farmers will also need (access to) complementary inputs such as seeds and fertilizer highlighting the importance of IMAP activities focused on increasing market presence of input traders.
- Levels of female empowerment are low across many domains of empowerment. The low level of membership in groups and low levels of decision making with respect to assets stand out (negatively) among the domains of empowerment, and supports the program's logic of raising women's empowerment in households and communities through increases in group membership and accumulation of productive capital.

Implications for the evaluation of SSI investments

- Overall, the process we developed to create an MCG has led to a treatment group and a matched comparison group that are well balanced both on indicators derived from remote sensing and those derived from household survey data. Most primary and secondary indicators show standardized differences that are smaller (in absolute terms) than the 0.1 SD threshold below which the literature considers variables to be balanced. Of the remaining indicators, most are below the 0.25 SD threshold that the literature considers to be an acceptable range for removing bias through the inclusion of baseline outcomes.

- The procedure was, however, not effective in finding matches from the Yelou and Sambera comparison areas for the 10 percent of SSI target plots engaged in dry season rice cultivation. The small number of SSI target plots and their matched comparisons remain significantly different according to input use, yields and sales. Given the project's intended focus on non-rice-growing plots, the consequences of these specific imbalances for the evaluation will depend on whether these plots are part of the final group of SSI beneficiary plots. If they are, it will be important to identify additional comparison plots for these target plots.
- While matching was effective in balancing overall levels of empowerment between the *Basse Terrasse* and the matched comparison samples, this was not the case for the different domains of empowerment. To obtain balance for the various empowerment domains the final evaluation will implement a separate matching procedure for women.
- Given delays in project activities, the final scope of the SSI activity and the selection of beneficiary plots yet undetermined. The possibility of a significant reduction in scope raises questions concerning the power of the evaluation design to identify impacts. Once the scope is finalized, we will update power calculations. The availability of the baseline data collected for this report will ensure these power calculations are reliable.
- While overall response rates are high, we were not able to interview all target SSI beneficiary households nor obtain geo-coded information from all target plots. If project scope is significantly reduced, we will assess whether the smaller number of beneficiary plots, and the area to be irrigated, was covered by the baseline data collection.
- Finally, the baseline report confirms findings in the literature that self-reported plot sizes and the area cultivated are subject to substantial biases. Plots measured using GPS measurements are about 1/3 smaller than the farmer reported size of the same plot. A primary reason is bunching where farmers generously round up to the nearest hectare or a simple fraction of a hectare—such as 1/2 of a hectare. This substantially affects measurements of agricultural productivity that are calculated on a per hectare basis of the area that is cultivated. Given the importance of these biases it is important to continue to collect geo-coded measurements of the area cultivated.

Roads for market access

- The findings from the market, trader, and village leader surveys suggest that rehabilitating the RN7, RN35, and RRS could lead to improvements in the short-, medium-, and long-term outcomes envisioned in the logic model as the current poor road quality does appear to be a binding constraint.
- In particular, the poor quality of roads increases the travel time to markets and transportation costs in the rainy season, a time when more crops are produced, and rain affects the state of many unimproved roads.
- Traders operating in markets near the project area are hesitant to trade perishable crops and experience losses and damages to these crops in transit when they do transport them, resulting in decreased crop sales and trader income. Farmers are similarly hesitant to grow perishable crops, likely because they are harder to sell and transport to market given the challenges faced by traders.
- The roads for market access activity also benefits households benefitting from the SSI investments. Given the SSI investment's focus on shifting to higher value crops, some of which are perishable, the logic model's envisioned positive impact on the presence of traders, lower transportation costs, and decreased travel time to bring crops to markets, would support the anticipated increase in the volumes and values of crops produced and traded to achieve the overall compact goal of increasing farmer incomes.

III. Evaluation Administration

A. Summary of IRB requirements and clearances

Mathematica is committed to protecting the rights and welfare of human subjects. We have ensured that the study meets all U.S. and Niger research standards for ethical clearance. Mathematica used Health Media Lab as our IRB because of our positive experience with it on other MCC projects. IRB approval required three sets of documents: (1) a research protocol that described the purpose and design of the research and provided information about our plans for protecting study participants and their confidentiality and human rights, including how we acquired consent for their participation; (2) copies of all data collection instruments and consent forms used for the evaluation; and (3) a completed IRB questionnaire that provided information about the research protocol, how we will securely collect and store our data, our plans for protecting participants' rights, and any possible drawbacks for participants that might result from any breach of data confidentiality. We also collaborated with our data collection firm, Société de Développement International, POKET, and ECDT DSEA, to obtain approval for conducting fieldwork from the National Statistics Institute in Niger.

B. Preparing data files for access, privacy, and documentation

All data collected for this evaluation are securely transferred from the collection firm to Mathematica, are stored on Mathematica's secure server, and are accessible only to project team members who use them. After producing and finalizing each of the evaluation reports, including this baseline report, we will prepare corresponding de-identified data files, user manuals, and codebooks based on the quantitative survey data. We understand that these files could be made available to the public, so we will de-identify the data files, user manuals, and codebooks according to MCC's most recent guidelines. Public-use data files will be free of personal or geographic identifiers that would permit unassisted identification of individual respondents or their households. In addition, we will remove or adjust variables that introduce a reasonable risk of deductive disclosure of the identities of individual participants.

For internal control and audit purposes, the local data collection firm will retain the data files, in both paper and electronic form, for the entire duration of the project, including the base contract and the subsequent option contracts. All the collected data and databases are the property of Mathematica and will be delivered to us at the end of the contract. We will also recode unique and rare data by using top and bottom coding or replacing affected observations with missing values. If necessary, we will also collapse any variables that make an individual highly visible because of geographic or other factors into less easily identifiable categories.

C. Dissemination plan

The Mathematica team will present evaluation findings remotely at both MCC and MCA-N headquarters. We will also participate in any other MCC-financed dissemination and training events related to the findings from the baseline and subsequent evaluation reports. To ensure that the results and lessons from the evaluation reach a wide audience, we will work with MCC to increase the visibility of the evaluation and findings within the agriculture sector, especially for policymakers and practitioners. After acceptance of the final evaluation report, the team will develop a policy brief with findings and analysis relevant to decision makers of MCC and the Government of Niger. We expect the broader research community to have a strong interest in the evaluation findings. To help disseminate results and lessons, we will collaborate with MCC and other stakeholders to identify additional forums (conferences, workshops, and publications) for disseminating the results.

D. Evaluation team: Roles and responsibilities

Evaluation team members	Role	Responsibility
Mr. Matt Sloan	Program manager	Overseeing the project team, providing quality assurance
Dr. Christopher Ksoll	Project director/primary point of contact for MCC	Leading the evaluation design and data analyses, overseeing the execution of the quantitative components of the design and data collection, managing quantitative data analysis. Communicating with client, coordinating with key stakeholders in the Niger agriculture sector, overseeing evaluation budget, overseeing data collection, managing evaluation team staffing and priorities, and being primarily responsible for delivering high quality products that meet MCC's and other stakeholders' needs
Dr. Anthony Louis D'Agostino	Senior analyst	Working on the design of the performance analysis and the analysis. Conducting data quality checks and overseeing the programming
Ms. Margo Berends	Analyst	Supporting the analysis and data collection and drafting of the baseline report, coordinating data collection and subcontractors
Mr. Samuel Studnitzer	Research assistant	Supporting data analysis
Dr. Evan Morier	Analyst	Supporting GIS data analysis
Mr. Saidou Amadou	In-country coordinator	Overseeing data collection fieldwork, monitor data quality, coordinate site visits, assist in communications with MCA-N
Ms. Poorva Upadhyaya	Project manager	Managing the project internally, tracking implementation documentation, coordinating with subcontractors, invoicing, communication with MCC

References

Alkire, S., R. Meinzen-Dick, A. Peterman, A. Quisumbing, G. Seymor, and A. Vaz. "The Women's Empowerment in Agriculture Index." *World Development*, vol. 52, 2013, pp. 71–91.

Backiny-Yetna, Prospere, and Kevin McGee. "Gender Differentials and Agricultural Productivity in Niger." Policy Research Working Paper 7199. Washington, DC: World Bank, 2015.

Carletto, Calogero, Dean Jolliffe, and Raka Banerjee. "From Tragedy to Renaissance: Improving Agricultural Data for Better Policies." *Journal of Development Studies*, vol. 51, no. 2, 2015, pp. 133–148.

Central Intelligence Agency (CIA). "The World Factbook: Niger." December 10, 2018. Available at https://www.cia.gov/library/publications/the-world-factbook/geos/ng.html. Accessed January 4, 2019.

D'Agostino, Anthony Louis, Christopher Ksoll, and Margo Berends. "Niger Irrigation and Market Access Project: Evaluation Design Report for Small-Scale Irrigation Activities and Complementary Investments in the Dosso-Gaya Area." Washington, DC: Mathematica, September 2021.

D'Agostino, Anthony Louis, Christopher Ksoll, Patricia Costa, Galina Lapadatova, and Matt Sloan. "Niger Irrigation and Market Access: Evaluation Design Report." Washington, DC: Mathematica, December 2019.

Desiere, Sam, and Dean Jolliffe. "Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?" *Journal of Development Economics*, vol. 130, 2018, pp. 84–98.

Exchange Rates UK. "West African CFA franc to US Dollar Spot Exchange Rates for 2020." November 2022a. Available at https://www.exchangerates.org.uk/USD-XOF-spot-exchange-rates-history-2020.html. Accessed November 6, 2022.

Exchange Rates UK. "West African CFA franc to US Dollar Spot Exchange Rates for 2021." November 2022b. Available at https://www.exchangerates.org.uk/USD-XOF-spot-exchange-rates-history-2021.html. Accessed November 6, 2022.

Famine Early Warning Systems Network (FEWS NET). "Assessment of Chronic Food Insecurity in Niger." March 2019. Available at https://fews.net/sites/default/files/documents/reports/Chronic%20FI%20Niger%20report%20Final%20English_0.pdf. Accessed: December 29, 2022.

Food and Agriculture Organization of the United Nations (FAO). "Countries by Commodity." 2020a. Available at http://www.fao.org/faostat/en/#rankings/countries_by_commodity. Accessed October 21, 2020.

Food and Agriculture Organization of the United Nations (FAO). "Land Use Indicators." 2020b. Available at http://www.fao.org/faostat/en/#data/EL._A Accessed October 21, 2020.

Fortson, Kenneth, Randall Blair, and Kathryn Gonzalez. "Evaluation of a Rural Road Rehabilitation Project in Armenia." Washington, DC: Mathematica, March 2015.

Funk, C., Peterson, P., Landsfeld, M., et al. "The Climate Hazards Infrared Precipitation with Stations—A New Environmental Record for Monitoring Extremes." *Scientific Data*, vol. 2, no. 150066, 2015). https://doi.org/10.1038/sdata.2015.66

Gourlay, Sydney, Talip Kilic, and David B. Lobell. "A New Spain on an Old Debate: Errors in Farmer-Reported Production and Their Implications for Inverse – Productivity Relationship in Uganda." *Journal of Development Economics*, vol. 141, 2019, 102376.

Greifer, N. Cobalt: Covariate Balance Tables and Plots. R package version 4.4.0. 2022.

Harris, Anthony, Laura Meyer, Delia Welsh, Anu Rangarajan, Alejandra Aponte, Joshua Claxton, Naomi Dorsey, et al. "Final Report: Evaluation of the Ghana Roads Activities." Washington, DC: Mathematica, September 2020a.

Harris, Anthony, Delia Welsh, Alejandra Aponte, Laura Meyer, Joshua Claxton, Naomi Dorsey, Roberto Armijo Zarricueta, et al. "Final Report: Evaluation of the Tanzania Transport Sector Project." Washington, DC: Mathematica, September 2020b.

Imbens, G.W., and Rubin, D.B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, 2015. https://doi.org/10.1017/CBO9781139025751

Imbens, Guido W., and Jeffrey M. Wooldridge. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature,* vol. 47, no. 1, 2009, pp. 5–86.

International Food Policy Research Institute (IFPRI). "Women's Empowerment in Agriculture Index Brief." Washington, DC: International Food Policy Research Institute, 2012. Available at https://www.ifpri.org/publication/womens-empowerment-agriculture-index.

Integrated Food Security Phase Classification (IPC). "Understanding the IPC Scales." 2022. Available at https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/docs/communication_tools/brochures/IPC_Brochure_Understanding_the_IPC_Scales.pdf. Accessed December 29, 2022.

Kafle, Kashi, and Soumya Balasubramanya. "Reducing Food Insecurity Through Equitable Investments in Irrigation: The Case of Niger." *Journal of the Agricultural and Applied Economics Association,* vol. 1, 2022, pp. 494–515. https://doi.org/10.1002/jaa2.40.

Lin, W. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics,* vol. 7, no. 1, 2013, pp. 295–318.

List, John A., Azeem M. Shaikh, and Atom Vayalinkal. "Multiple Testing with Covariate Adjustment in Experimental Economics." No. 00732. The Field Experiments website, 2021.

Lobell, D.B., G. Azzari, M. Burke, S. Gourlay, Z. Jin, T. Kilic, and S. Murray. "Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis." *American Journal of Agricultural Economics,* vol. 102, 2020, pp. 202–219.

Maidment, R.I., D. Grimes, E. Black, E. Tarnavsky, M. Young, H. Greatrex, R.P. Allan, et al. "A New, Long-Term Daily Satellite-Based Rainfall Dataset for Operational Monitoring in Africa." Scientific Data 4: 170063 DOI:10.1038/sdata.2017.63.

Maidment, R., D. Grimes, R. Allan, E. Tarnavsky, M. Stringer, T. Hewison, R. Roebeling, and E. Black. The 30-year TAMSAT African Rainfall Climatology and Time-series (TARCAT) Data Set. *Journal of Geophysical Research: Atmospheres,* vol. 119, 2014, pp. 10,619–10,644. DOI: 10.1002/2014JD021927.

Millennium Challenge Account Niger (MCA-N). "Monitoring and Evaluation Plan of the Niger Compact Between the United States of America, Acting Through the Millennium Challenge Corporation and the Republic of Niger." Niamey, Niger: Millennium Challenge Account Niger, July 2019.

Millennium Challenge Corporation (MCC). "Evaluation." n.d. https://www.mcc.gov/our-impact/independent-evaluations.

Millennium Challenge Corporation (MCC). "Millennium Challenge Compact Between the United States of America Acting Through the Millennium Challenge Corporation and the Republic of Niger Acting Through the Ministry in Charge of Foreign Affairs and Cooperation." Washington, DC: Millennium Challenge Corporation, July 29, 2016.

Millennium Challenge Corporation (MCC). "Request for Quotation: Irrigation and Market Access Evaluation in Niger." Solicitation Number: MCC-17-RFQ-0123. Issued July 26, 2017.

Namatsheve, T., R. Cardinael, M. Corbeels, and R. Chikowo. "Productivity and Biological N2-Fixation in Cereal-Cowpea Intercropping Systems in Sub-Saharan Africa. A review." *Agronomy for Sustainable Development*, vol 40, no. 30, 2020. https://doi.org/10.1007/s13593-020-00629-0

Nguyen, T.L., G.S Collins, J. Spence, et al. "Double-Adjustment in Propensity Score Matching Analysis: Choosing a Threshold for Considering Residual Imbalance." *BMC Medical Research Methodology*, vol. 17, no. 78, 2017. https://doi.org/10.1186/s12874-017-0338-0

Normand, S.L.T., M.B. Landrum, E. Guadagnoli, J.Z. Ayanian, T.J. Ryan, P.D. Cleary, and B.J. McNeil. "Validating Recommendations for Coronary Angiography Following an Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores. *Journal of Clinical Epidemiology*, vol. 54, 2001, pp. 387–398.

Paliwal, Ambica, and Meha Jain. "The Accuracy of Self-Reported Crop Yield Estimates and Their Ability to Train Remote Sensing Algorithms." *Frontiers in Sustainable Food Systems*, vol. 4, 2020.

R Core Team. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2002. https://www.R-project.org/.

Rosenbaum, Paul R., and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, vol. 70, no. 1, 1983, pp. 41–55.

Rubin, D.B. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology*, vol. 2, 2001.

Schreiner, Mark. "Simple Poverty Scorecard® Tool: Niger." Scorocs, April 2018.

Sekhon, Jasjeet S. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." *Journal of Statistical Software*, vol. 42, no. 7, 2011, pp. 1–52. https://doi.org/10.18637/jss.v042.i07

Tarnavsky, E., D. Grimes, R. Maidment, E. Black, R. Allan, M. Stringer, R. Chadwick, and F. Kayitakire. "Extension of the TAMSAT Satellite-Based Rainfall Monitoring over Africa and from 1983 to Present." *Journal of Applied Meteorology and Climatology*, vol. 53, no. 12, 2014, pp. 2805–2822. doi: 10.1175/JAMC-D-14-0016.1.

Turiansky, Abbie, Christopher Ksoll, Caroline Lauver, and Matt Sloan. "Niger Irrigation and Market Access: Evaluability Assessment." Washington, DC: Mathematica, November 2018.

Wahab, Ibrahim. "In-Season Plot Area Loss and Implications for Yield Estimation in Smallholder Rainfed Farming Systems at the Village Level in Sub-Saharan Africa." *GeoJournal*, vol. 85, 2020, pp. 1553–1572.

Wang, Sherrie, Stefania DiTommaso, Jillian M. Denies, and David B. Lobell. "Mapping Twenty Years of Corn and Soybean Across the US Midwest Using the Landsat Archive." *Scientific Data*, vol. 7, 2022, no. 307.

What Works Clearinghouse. *What Works Clearinghouse Procedures and Standards Handbook,* Version 5.0. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2022.

World Bank. "Agricultural Sector Risk Assessment in Niger: Moving from Crisis Response to Long-Term Risk Management." Washington, DC: World Bank, 2013.

World Bank. "Monitoring Global Poverty: Report of the Commission on Global Poverty." Washington, DC: World Bank, 2017a. Available at https://openknowledge.worldbank.org/bitstream/handle/10986/25141/9781464809613.pdf. Accessed November 9, 2020.

World Bank. "PPP Conversion Factor, GDP (LCU per international $) – Niger." Available at https://data.worldbank.org/indicator/PA.NUS.PPP?locations=NE. Accessed October 29, 2020.

World Bank. "Republic of Niger: Priorities for Ending Poverty and Boosting Shared Prosperity—Systematic Country Diagnostic." Report no. 115661-NE, 2017b.

World Food Programme. "Niger Factsheet." Available at http://www1.wfp.org/countries/niger. Accessed August 10, 2018.

Wouterse, Fleur. "The Role of Empowerment in Agricultural Production: Evidence from Rural Households in Niger." *Journal of Development Studies*, vol. 55, no. 4, 2019, pp. 565–580. https://doi.org/10.1080/00220388.2017.1408797

Appendix A:

Evaluation Design Overview and Secondary Outcome Indicators

Table A.1. Evaluation design overview

#	Research question	Evaluation method	Data source and type
A. Overarching impact evaluation questions			
RQ1	Did the project components interact as envisioned during project design to reach a common objective? If yes, what facilitated the interaction, and if no, why not? Was there close coordination and planning among the different contractors designing and implementing the activity (land governance, infrastructure, training in infrastructure management, and agricultural services)? Did UNOPS in the role of project management consultant facilitate the rollout and coordination of activities?	• Implementation analysis	• Project documentation • KIIs with MCA-N, UNOPS, and program implementers • FGDs with beneficiaries
RQ2	*To what extent did the project interact with the grant facility of the Climate-Resilient Communities Project? What facilitated the interaction, and what didn't?*	• *Implementation analysis*	• *KIIs with program implementers and key stakeholders* • *FGDs with beneficiaries* • *Project documentation*
RQ3	What is the impact of SSI investments on beneficiary households' incomes, volumes, and value of agricultural products sold and traded, food and nutritional security, and production of cash crops?	• Impact analysis	• Surveys of SSI beneficiary and comparison households • Satellite imagery • Crop cuts
RQ4	Do stakeholders believe the project was well designed to achieve the project objective? What changes to implementation occurred, and why?	• Implementation analysis	• Project documentation • KIIs with MCA-N, program implementers, GoN stakeholders
RQ5	If the project produced results, are they expected to be sustained?	• Sustainability analysis • Infrastructure assessment	• KIIs with MCA-N, program implementers, GoN stakeholders • Site visits
RQ6	What lessons can be drawn to inform future projects?	• Synthesis of evaluation analyses	• Mathematica evaluation analyses • Compact closeout documents • KIIs with MCA-N, program implementers, GoN stakeholders
RQ7	What is the post-compact ERR of the project (except for the Roads for Market Access Activity)?	• Cost-benefit analyses	• Surveys of SSI beneficiary and comparison households • Project financial data • Satellite imagery • KIIs with market actors • Crop cuts

#	Research question	Evaluation method	Data source and type
RQ8	Were IPD project activities implemented as planned? If not, what changes occurred?	• Implementation analysis	• Project documents • KIIs with MCA-N, program implementers, Ministry of Water and Sanitation • FGDs with beneficiaries
RQ9	Were the expected outputs produced by the IPD activity?	• Infrastructure assessment • Qualitative analysis	• Program monitoring data • KIIs with MCA-N, program implementers, GoN stakeholders • FGDs with beneficiaries • Project area visits
RQ12a	Did irrigated land increase as expected?	• Impact analysis	• Satellite imagery • Surveys of SSI beneficiary households
RQ14	Were MSMF project activities implemented as planned? If not, what changes occurred?	• Implementation analysis	• Project documentation • KIIs with MCA-N, program implementers, Ministry of Agriculture • FGDs with beneficiaries
RQ15	Were the expected outputs produced by the MSMF activity?	• Implementation analysis	• Program monitoring data • KIIs with MCA-N, program implementers, GoN stakeholders • FGDs with beneficiaries
RQ22a	What is the impact of SSI investments and land formalization on land tenure security and the level and risk of land conflict?	• Impact analysis	• Surveys of SSI beneficiary households • COFOCOM administrative data
B. Irrigation Perimeter Development Activity evaluation questions			
RQ10	Is the new infrastructure operating and functioning properly?	• Infrastructure assessment • Qualitative analysis	• Site visits and irrigation assessment • KIIs with MCA-N, program implementers, GoN stakeholders • FGDs with beneficiaries
RQ11	Is water for irrigation in farmers' plots available as expected from the small-scale irrigation systems, including frequency, timing, and amount as planned? If not, why not?	• Quantitative descriptive analysis • Qualitative analysis	• Surveys of SSI beneficiary households • FGDs with beneficiaries • Satellite imagery (such as Soil Moisture Active Passive) • KIIs with GoN stakeholders • FGDs with beneficiaries
RQ12b	If irrigated land did not increase as expected, then why?	• Qualitative analysis	• Surveys of SSI beneficiary households • KIIs with GoN stakeholders

#	Research question	Evaluation method	Data source and type
RQ13	What is the cost of irrigation, including any fuel costs for pumping water? If water was available before the SSI system was built, how did the cost of irrigation water change?	• Quantitative descriptive analysis	• Surveys of SSI beneficiary households
RQ14	*Were project activities implemented as planned? If not, what changes occurred?*	• *Implementation analysis*	• *KIIs and FGDs* • *Project documentation*
RQ15	*Were the expected outputs produced by the activity?*	• *Qualitative outcomes analysis*	• *KIIs and FGDs* • *Monitoring data* • *Project documentation*
C.	**Management Services and Market Facilitation Activity: Sustainable Irrigation Systems Management Sub-Activity evaluation questions**		
RQ16	Did the project support the institutions or market actors responsible for O&M as planned? What is the capacity of these institutions or market actors, and the government oversight institutions?	• Qualitative analysis	• KIIs
RQ17	*What was the profile of the participants (total number of participants disaggregated by sex and age)?*	• *Quantitative descriptive analysis*	• *Monitoring data*
RQ18	*What percentage of IWUA leadership committee members at the end of the Compact were women?*	• *Quantitative descriptive analysis*	• *Administrative data*
RQ19	Is the small-scale irrigation infrastructure being maintained properly?	• Infrastructure assessment • Qualitative analysis	• KIIs with MCA-N, program implementers, GoN stakeholders • FGDs with beneficiaries • Site visits
D.	**Management Services and Market Facilitation Activity: Land Tenure Security Sub-Activity evaluation questions**		
RQ3a	Do agricultural input use, crop choice, agricultural techniques, and agricultural income, volumes and value of agricultural products differ between landowners and renters or renter groups? If so, why?	• Quantitative descriptive analysis • Qualitative analysis	• Surveys of SSI beneficiary and comparison households • Satellite imagery • KIIs with GoN stakeholders • FGDs with beneficiaries
RQ20	Is the land registry used as a tool by local authorities to record continual changes in landholdings? Do landholders have access to the correct documentation according to the project plan?	• Qualitative analysis • Quantitative descriptive analysis	• KIIs with MCA-N, program implementers, GoN stakeholders • COFOCOM administrative data • Program monitoring data • FGDs with beneficiaries • Surveys of SSI beneficiary households
RQ21	Are the local land commissions in the project zone better equipped to ensure sustainable management of land rights in/around the *Basse Terrasse* project area?	• Sustainability analysis	• Project documentation • Budget outlays • KIIs with MCA-N, program implementers, GoN stakeholders

#	Research question	Evaluation method	Data source and type
RQ21a	Is the formal land rental process used by landholders? How is it functioning?	• Qualitative analysis • Quantitative descriptive analysis	• KIIs with MCA-N, program implementers, GoN stakeholders • COFOCOM administrative data • Program monitoring data • FGDs with beneficiaries • Surveys of SSI beneficiary households
RQ21b	What are the contractual terms between landowners and land renters or renter groups? How are input costs—including costs for the irrigation operation and maintenance—and agricultural outputs shared between owners and renters or renter groups? Have there been disagreements over land usage and contract terms?	• Quantitative descriptive analysis	• Surveys of SSI beneficiary households
RQ22a	What is the effect of SSI investments and land formalization on land tenure security, and what is the level and risk of land conflict?	• Qualitative analysis	• FGDs with beneficiaries
RQ22b	How do perceptions of land tenure security, risk of land conflict, access to credit, and agricultural inputs, investments, and outputs compare among landholders and tenants, and vary among tenants with different contractual terms?	• Quantitative descriptive analysis	• Surveys of SSI beneficiary households
E.	**Management Services and Market Facilitation Activity: Agricultural Support Services Sub-Activity evaluation questions**		
RQ24	*What percentage of participants of adult functional literacy and numeracy trainings report improvement in their skills (basic reading and writing) after the training? What percentage indicate improved knowledge of nutrition and hygiene, and budgeting and record keeping (inasmuch as these concepts were introduced as part of the literacy and numeracy training)?*	• *Quantitative descriptive analysis* • *Qualitative outcomes analysis*	• *Surveys of households* • *Monitoring data* • *FGDs*
RQ25	*What percentage of participants' self-report increased knowledge of sustainable land and water resources management?*	• *Quantitative descriptive analysis*	• *Surveys of households*
RQ26	Did participants perceive that they learned new skills/knowledge? Did this vary by subgroup? If they didn't perceive learning/acquiring new knowledge, why didn't they?	• Quantitative descriptive analysis • Qualitative analysis	• Surveys of SSI beneficiary households • Program monitoring data • KIIs with MCA-N, program implementers, GoN stakeholders • FGDs with beneficiaries • Program implementer reports

#	Research question	Evaluation method	Data source and type
RQ27	What percentage of participants of adult functional literacy and numeracy report improvement in their skills (basic reading and writing) after the training? What percentage of them indicate improved knowledge of nutrition and hygiene, and of budgeting and record keeping?	• Quantitative descriptive analysis	• Surveys of SSI beneficiary households • Program monitoring data
RQ28	What percentage of participants' self-report increased knowledge of sustainable land and water resources management?	• Quantitative descriptive analysis	• Surveys of SSI beneficiary households • Program monitoring data
RQ29	What percentage of participants show an active knowledge of improved agricultural practices that they did not know before the training?	• Quantitative descriptive analysis	• Surveys of SSI beneficiary households • Program monitoring data
RQ30	What percentage of members of *comites de gestion* within the producer groups indicate improved knowledge of producer group management?	• Quantitative descriptive analysis	• Surveys of SSI beneficiary households • Program monitoring data
RQ31	Have participants applied new practices and technologies? Was this different for women/men or youth/non-youth participants? If knowledge was not applied, why wasn't it?	• Quantitative descriptive analysis • Qualitative analysis	• Surveys of SSI beneficiary households • Program monitoring data • KIIs with MCA-N, program implementers, GoN stakeholders • FGDs with beneficiaries • Program implementer reports
RQ32	Were savings and loans groups created and fostered by the project? Based on their participation, have group participants indicated they have improved access to credit?	• Quantitative descriptive analysis • Qualitative analysis	• Program implementer reports • Program monitoring data • KIIs with MCA-N, program implementers, GoN stakeholders • FGDs with beneficiaries • Program implementer reports
RQ33	How are producer groups applying knowledge?	• Qualitative analysis	• KIIs with MCA-N, program implementers, GoN stakeholders • FGDs with beneficiaries and producer groups • Program implementer reports

#	Research question	Evaluation method	Data source and type
F. Roads for Market Access Activity evaluation questions			
RQ40	To what extent did the activity lead to a change in transportation method, travel time, vehicle operating costs, and transportation costs for traders and farmers in the *Basse Terrasse* and surrounding areas?	• Pre-post analysis • Quantitative descriptive analysis	• Surveys of SSI beneficiary and comparison households • Surveys of traders • Surveys with village leaders
RQ41	Are more input and output traders present in the Dosso-Gaya region as a result of the road improvements?	• Quantitative descriptive analysis	• Market records • Surveys with village leaders
RQ42	To what extent did the activity contribute to increased volumes and values of agricultural products traded from the *Basse Terrasse* area? How has the activity changed the quality of crops, in particular produce, brought to market and the quantity of crops lost in transportation post-harvest?	• Pre-post analysis • Quantitative descriptive analysis • Qualitative analysis	• Surveys of SSI beneficiary and comparison households • Surveys of traders • Surveys with village leaders • KIIs with traders

Notes: CAIMA = *Centrale d'Approvisionnement en Intrants et Matériels Agricoles*; FGD = focus group discussion; GoN = Government of Niger; IPD = Irrigation Perimeter Development Activity; KII = key informant interview; LTS = Land Tenure Security; MCA-N = Millennium Challenge Account-Niger; MSMF = Management Services and Market Facilitation Activity; ONAHA = *l'Office National des Aménagements Hydroagricoles*; PAP = project affected person; PR = Policy Reform Activity; SAA = Agricultural Support Services Sub-Activity; SISM = Sustainable irrigation System Management; UNOPS = United Nations Office for Project Services.

Table A.2. Secondary outcome indicators for impact analysis

Key indicator	Definition
Land security and use (plot level)	
Cropping patterns	Seasonal indicators of plots with mono- or inter-crop cultivation
Irrigation (plot level)	
Irrigation available when needed	Seasonal indicators of irrigation availability when needed most
Fertilizer (plot level)	
Fertilizer application area	Seasonal share of plot area where fertilizer was applied
Type of fertilizer used	Seasonal indicators of fertilizer use and quantities for chemical and organic fertilizer
Agricultural inputs (plot and household level)	
Type of seeds	Seasonal indicators of purchased and/or improved seeds and the associated share of plot area sown
Seed expenditures	Annual household expenditures on seeds and cost of seeds per hectare
Improved inputs or practices	Indicators for use of improved inputs or practices including improved water and soil management techniques
Credit and expenditures (household and plot level)	
Credit access	Indicators of access to credit or taking out a loan in the past year and the associated value of loans
Plot collateral	Indicators of using or considering using plot as collateral for credit
Agricultural expenditures	Annual household agricultural expenditures
Labor expenditures	Annual household expenditures on labor and cost of labor per hectare
Preparation and processing expenditures	Annual household expenditures on preparing and processing crops for sale and cost per tonne
Income (household level)	
Agricultural revenue	Annual value of revenue from crop sales and rent
Crop sales	Seasonal and annual values of revenue from crop sales including cash crops, traditional crops, and dual purpose crops
Non-agricultural income	Annual value of employment and self-employment income
Food insecurity (household level)	
Extreme hunger	Indicator of households where at least one member did not eat for an entire day in the previous month
Food insecurity frequency	Number of times in the past month that household experienced food inadequacy, hunger, or extreme hunger
Food inadequacy in the past lean season	Indicator of households that did not have enough food in the past lean season (defined as the 30 days prior to the first rain of the past rainy season)
Hunger in the past lean season	Indicator of households where at least one member went to bed hungry during the previous lean season
Anticipated food inadequacy in the upcoming lean season	Indicator of households that expect they will not have enough food in the upcoming lean season (defined as the 30 days prior to the first rain of the next rainy season)
Anticipated hunger in the upcoming lean season	Indicator of households that expect that at least one member will go to bed hungry during the upcoming lean season

Key indicator	Definition
Poverty (household level)	
Poverty score	Scorocs Simple Poverty Scorecard for Niger based on nine poverty indicators to estimate consumption-based poverty rates
Poverty indicators	Selected indicators related to the poverty score: improved roof materials, number of rooms, improved toilet, electricity, number of cell phones, and motorized transport
Women's Empowerment in Agriculture Index (WEAI) (household level)	
Percentage empowered	Share of women who are empowered (at or above 80 percent adequacy) based on production, resources, income, and leadership domains
Percentage unempowered	Share of women who are unempowered (below 80 percent adequacy) based on production, resources, income, and leadership domains
Adequacy score	Weighted measure of adequacy (the extent of empowerment)
Adequacy score among the unempowered	Weighted measure of adequacy (the extent of empowerment) among unempowered women

Note: Data for all secondary indicators are collected through the household surveys.

Appendix B:

Coverage of Geo-coded *Basse Terrasse* Target Plots

Table B.1. Coverage of geo-coded *Basse Terrasse* target plots

Type	SONED	Baseline information coverage
Area		
Area of *BT* target plots covered (ha)	634.6	441
Area of *BT* target plots covered (%)	100	69.5

Note: Calculated based on SONED data delivered in June 2022. We exclude from the baseline sample all *BT* households that did not have at least 30 percent overlap with any treatment parcel included in the SONED data, which account for GPS and other measurement errors.

Appendix C:

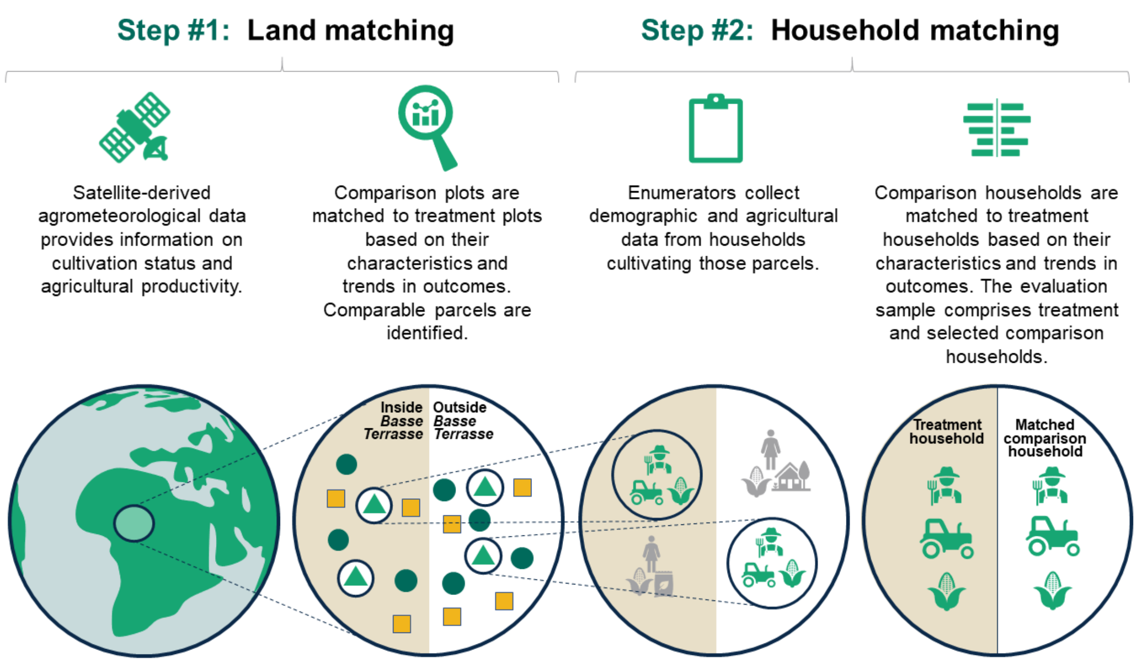
Matching Methodology and Supplementary Results

In this appendix, we provide additional methodological details and results for our matching procedure. In the following subsections, we (1) summarize our process for identifying plausible comparison group units; (2) present the propensity score models for the plot- and household-level matching models; (3) describe the matching procedure used to create the comparison group using the propensity scores; (4) describe our approach for assessing balance between the treatment and comparison groups, which is essential for judging the validity of our research design; and (5) report the balance performance for our plot-level sample and then for our household-level sample.

A. Process for identifying plausible comparison group units

Appendix Figure C.1 Error! Reference source not found. summarizes our process for identifying and selecting plots and households that would support a matched comparison group (MCG) design. Because project implementers had collected data only from potential beneficiaries, our process involved determining which non-beneficiaries were most likely to be comparable to beneficiaries across characteristics important to the analysis, such as agricultural productivity, household income, crop-marketing activities, and food security status. In the first step of our process—land matching—we used remotely sensed land cover and groundwater data to shortlist 1-hectare land parcels with hydrogeological conditions and vegetation index properties that approximated those of *BT* target parcels. This process combined both longitudinal (for example, time-series normalized difference vegetation index [NDVI] values) and cross-sectional (for example, a radar-derived measure of access to groundwater) data to ensure that prospective matches were as similar as possible to treatment parcels in recent years and over a longer period. D'Agostino et al. (2021) describes this process and specifies the variables used to identify potential comparison parcels.

Figure C.1. Overview of matching process to select comparison plots and households



The first step resulted in a list of locations which, from the vantage point of satellite observation, appeared comparable to *BT* target parcels. We provided enumerators with the shortlisted locations' GPS coordinates and for each location the enumerators were tasked with identifying and locating the relevant cultivator, since cadastral data were unavailable for these areas. We administered a brief, in-person survey through the POKET app on the Android OS platform to cultivators who consented.[21] Using POKET, enumerators collected information on crop choices, land holdings, and input costs, among other data fields. Enumerators also collected GPS outlines for cultivators' plots by walking plot perimeters and recording location data in the app. This data collection yielded a sampling frame of plot- and household-level information.

In the second step, we selected a subset of comparison group households from our POKET-collected sampling frame that were most comparable to *BT* households. These households, along with those of *BT*, were revisited and asked to complete a multi-topic household survey. Using the data collected, including GPS outlines of plots and subplots, along with satellite data corresponding to the GPS outlines, we performed a final iteration of the two steps displayed in **Figure C.1**Error! Reference source not found..

B. Propensity score model and approach for testing balance

In this sub-section, we describe the propensity score model and report on balance between the treatment and comparison groups. We focus on sets of variables that are plausibly related to treatment assignment or could influence the key outcomes of interest such as farm productivity, household income, and food security status. We included these variables in the estimation of a propensity score (Rosenbaum and Rubin 1983), estimated separately at the plot and household levels. Such propensity scores represent the probability of assignment to the treatment group that is generated by regressing a logit model of actual treatment assignment on a set of covariates that are subjectively believed to influence treatment compliance and/or our other outcomes of interest like crop productivity, and iteratively assessing balance before modifying the propensity score model. After examining the degree of common support—when both treatment and comparison group units have common propensity score values—we manually selected an upper bound beyond which observations were excluded from the analysis sample. Without restricting observations to those with common support, treatment observations with high propensity scores are either matched with untreated comparison units with much lower propensity scores or matched many times to the same few comparison observations, which would increase the variance of estimates (Rubin 2001). This would introduce two types of risks: (1) an evaluation risk, because future data collection rounds might be unable to collect data for that specific observation; and (2) a variability risk, because of the outsized weight attached to the few comparison observations at high levels of propensity scores. We matched treatment observations to the five nearest comparison neighbors, with replacement.

For both plot- and household-level results, we share (1) the distribution of propensity scores, (2) information on the sample size of included observations, and (3) balance test results for a number of variables to demonstrate the suitability of the selected comparison group.

Appendix Table C.1 lists all variables used our analysis, with separate columns to denote whether (1) the variable was an input into estimating the propensity score (categorized separately by plot or household level), (2) the variable was an input into the matching algorithm, and/or (3) balance tests were performed on the variable. Most indicators were collected from the household survey, but a subset of variables in our plot-level analysis were derived from satellite or radar data, as stated in the "source" column. For each indicator, we note whether it was an input into the propensity score model ("PS"), was used in matching ("M"), or was analyzed in a balance test ("B").

[21] The POKET app website is at https://www.poketapp.com/.

Table C.1. Plot- and household-level variables used in analysis

Variable	Source	Application		
		PS	M	B
Plot-level				
Applied chemical fertilizer in rainy season (0/1)	Household survey	✓	✓	✓
Area cultivated in rainy season (self-reported)	Household survey	✓	✓	✓
Area-weighted dry season (GCVI) z-score	Sentinel-2			✓
Area-weighted rainy season green chlorophyll vegetation index (GCVI) z-score	Sentinel-2	✓	✓	✓
Chemical fertilizer intensity (kg/ha) in rainy season [GPS-measured area]	Household survey; plot measurement survey	✓	✓	✓
Crop yield z-score	Household survey	✓		✓
Cultivated during past dry season (0/1)	Household survey			✓
Cultivated during past rainy season (0/1)	Household survey			✓
Cultivated in dry season by someone outside the HH (0/1)	Household survey			✓
Cultivated in rainy season by someone outside the HH (0/1)	Household survey			✓
Evapotranspiration	FAO WaPOR	✓	✓	✓
Fallow in past rainy season (0/1)	Household survey			✓
GCVI harmonic regression coefficients (three cycles)[a]	Sentinel-2	✓	±	✓
Grew cowpeas in rainy season (0/1)	Household survey	✓	✓	✓
Grew millet in rainy season (0/1)	Household survey	✓		✓
Grew rice in rainy season (0/1)	Household survey	✓		✓
Grew sorghum in rainy season (0/1)	Household survey			✓
Groundwater measurement (3 bands)	WATEX	✓	±	✓
Intercropped in rainy season (0/1)	Household survey			✓
Irrigated plot in either rainy or dry season (0/1)	Household survey	✓	✓	✓
Plot area (GPS-measured)	Plot measurement survey	✓	✓	✓
Share of plot cultivated in the rainy season	Household survey			✓
Total annual expenses (IHS)	Household survey			✓
Total annual expenses per hectare (IHS)	Household survey	✓	✓	✓
Used improved/hybrid seeds for 1+ crops in past rainy season (0/1)	Household survey			✓
Used labor during the past rainy season (0/1)	Household survey			✓
Household-level				
1+ family member went to bed hungry in previous month (0/1)	Household survey	✓	✓	✓
Age of household head	Household survey			✓
Agricultural income	Household survey			✓
Agricultural income share of total household income	Household survey	✓	✓	✓
Dry season cultivated area (self-reported)	Household survey			✓
Experiences extreme hunger in lean season sometimes/often (0/1)	Household survey			✓
Family had employment income (0/1)	Household survey			✓
Female has input into productive decisions (0/1)	Household survey	✓	✓	✓
Food in lean season sometimes/often unavailable (0/1)	Household survey			✓

Variable	Source	Application		
		PS	M	B
Has working motorcycle or private vehicle (0/1)	Household survey			✓
Household had self-employment income (0/1)	Household survey			✓
Household has irrigation access (0/1)	Household survey			✓
Household head is literate (0/1)	Household survey			✓
Household land holdings per person	Household survey			✓
Household size	Household survey			✓
Male household head (0/1)	Household survey	✓	✓	✓
Market transport expenses (IHS)	Household survey			✓
No household food 1+ times in past month (0/1)	Household survey			✓
Non-agricultural, annual household income (IHS)	Household survey	✓	✓	✓
Number of household phones	Household survey			✓
Owns a farm equipment asset (0/1)	Household survey	✓	✓	✓
Owns irrigation pump (0/1)	Household survey	✓	✓	✓
Owns large livestock (0/1)	Household survey	✓	✓	✓
Owns plots cultivated by non-household member (0/1)	Household survey			✓
Poverty score	Household survey	✓	✓	✓
Roof made of improved materials (0/1)	Household survey			✓
Sold crops during rainy season (0/1)	Household survey	✓	✓	✓
Total annual household income (unconditional)	Household survey			✓
Total expenses (IHS)	Household survey	✓	✓	✓
Total land holdings (self-reported)	Household survey	✓	✓	✓
Total rainy season cultivated hectares (IHS)	Household survey			✓
WEAI value	Household survey			✓

B = variable tested for balance; GCVI = Green Chlorophyll Vegetation Index; HH = Household; IHS = inverse hyperbolic sine transformation applied to variable; M = variable was input into matching algorithm; PS = variable used in estimating the propensity score.

Notes: ± A subset of bands were used in the matching algorithm (WATEX bands 1 and 3, GCVI intercept, sine *t* and cosine *t*, and sine 2*t*, and cosine 2*t* terms.) [a] Our three-cycle harmonic regression takes the form of $y_{it} = \alpha_i + \sum_{k=1}^{3} \beta_{i,k} \cos(2\pi\omega kt) + \gamma_{i,k} \sin(2\pi\omega t)$ where y_{it} is the average GCVI of plot *i* at time *t*. Wang et al. (2020) recommend GCVI to normalized difference vegetation index (NDVI) because it does not saturate at high leaf area values and is also derived from Sentinel-2. We selected k = 3 cycles and w = 1 as generating better observed model fit than alternative parameter values.

In addition to reporting the absolute standardized mean difference (ASMD), we also graphically examine univariate distributions of the same covariates (**Appendix Table C.11**). This step offers additional evidence on the similarity (or dissimilarity) of the treatment and comparison groups; rather than just sharing a similar mean value, the density of values across the covariate's full support should be comparable.

Appendix Table C.2 and **Appendix Table C.3** present the propensity score model results for plots and households, respectively.

Table C.2. Logistic regression model results for estimating plot-level propensity score

	β (SE)
Plot area quintile 2 (GPS-measured)	0.30 (0.21)

	β (SE)
Plot area quintile 3 (GPS-measured)	0.21 (0.21)
Plot area quintile 4 (GPS-measured)	0.20 (0.22)
Plot area quintile 5 (GPS-measured)	-0.28 (0.23)
Crop yield z-score	0.06 (0.09)
Area cultivated in rainy season (self-reported)	-0.35 *** (0.10)
Total annual expenses per hectare (IHS)	0.08 *** (0.02)
Applied chemical fertilizer in rainy season (0/1)	0.72 *** (0.18)
Chemical fertilizer intensity (kg/ha) in rainy season [GPS-measured area]	1.71 *** (0.33)
Irrigated plot in rainy or dry season (0/1)	1.42 *** (0.18)
Grew rice in rainy season (0/1)	0.51 ** (0.20)
Groundwater measurement (WATEX band 1)	0.02 *** (0.00)
Groundwater measurement (WATEX band 3)	0.02 *** (0.00)
Area-weighted rainy season GCVI z-score	-0.11 (0.09)
Grew millet in rainy season (0/1)	0.69 *** (0.19)
Evapotranspiration	0.00 (0.00)
Grew cowpeas in rainy season (0/1)	0.86 *** (0.14)
GCVI harmonic regression coefficient (intercept)	-5.79 *** (0.81)
GCVI harmonic regression coefficient (first sine)	-2.02 * (1.08)
GCVI harmonic regression coefficient (first cosine)	-8.79 *** (1.10)
GCVI harmonic regression coefficient (second sine)	-4.98 *** (1.08)
GCVI harmonic regression coefficient (second cosine)	-2.79 *** (1.06)
Constant	-2.51 ** (0.98)
N	2625

	β (SE)
Log likelihood	-838.41
AIC	1722.83

Note: Mathematica calculations. Standard errors are reported in parentheses.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table C.3. Logistic regression model results for estimating household-level propensity score

	β (SE)
Male household head	-1.30 *** (0.34)
Total land holdings (self-reported)	-0.20 *** (0.05)
Female has input into productive decisions	0.60 *** (0.16)
Poverty score	0.01 * (0.01)
Non-ag, annual household income (IHS)	0.09 *** (0.01)
Total expenses (IHS)	0.17 *** (0.03)
1+ family member went to bed hungry in previous month	-0.65 *** (0.18)
Owns a farm equipment asset	-0.49 *** (0.18)
Owns large livestock	-0.99 *** (0.19)
Sold crops during rainy season	-0.26 (0.22)
Owns irrigation pump	1.50 *** (0.19)
Ag share of total household income	1.97 *** (0.28)
(Intercept)	-2.19 *** (0.51)
N	1,608
logLik	-583.44
AIC	1193.87

Note: Mathematica calculations. Standard errors are reported in parentheses.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

C. Matching procedure

Stuart (2010) describes the tradeoffs to consider in deciding between various methods to create matched comparison groups. We chose the nearest neighbor method as a simple, yet effective method for identifying control observations, in particular where there are decisions to be made about future follow-up of respondents.

Having implemented a simple nearest neighbor matching without replacement (see **Appendix Table C.4** and **Appendix Table C.5**), we considered the samples not balanced enough. Stuart (2010) also discusses k:1 matching, where multiple control observations are matched to a treatment observation, matching with and without replacement, as well as limiting the matches to the region of common support. In our second, and preferred approach, we considered the tradeoffs discussed. Matching with replacement is "particularly helpful in settings where there are few control individuals comparable to the treated individuals" (Stuart 2010), as is the case for the observations with higher propensity scores. We limited the matches to the area of common support. We also considered matching with 5 and 10 matches and concluded that matching with 5 matches performed better. Our primary matching model used 5-nearest neighbors with replacement, restricting the sample to the area of common support.

Table C.4. Balance table of key plot-level covariates for untrimmed, 1nn without replacement sample

	Comparison		Treatment			
Covariate	Mean	SD	Mean	SD	Standardized diff	Variance Ratio
Applied chemical fertilizer in rainy season	0.67	0.47	0.75	0.43	0.19	
Area cultivated in rainy season (self-reported)	1.28	0.64	1.27	0.71	-0.01	1.21
Area-weighted dry season GCVI z-score	-0.05	0.34	-0.10	0.33	-0.14	0.97
Area-weighted rainy season GCVI z-score	-0.07	0.99	-0.03	0.76	0.06	0.59
Chemical fertilizer intensity (kg/ha) in rainy season [GPS-measured area]	0.16	0.27	0.24	0.32	0.26	1.35
Crop yield z-score	0.08	0.77	0.20	0.80	0.15	1.09
Cultivated during past dry season	0.13	0.34	0.21	0.40	0.18	
Cultivated during past rainy season	0.97	0.18	0.95	0.22	-0.08	
Cultivated in dry season by someone outside the HH	0.00	0.06	0.01	0.08	0.03	
Evapotranspiration (000s)	7.26	1.59	7.28	1.12	0.01	0.49
GCVI harmonic regression coefficient (first cosine)	0.04	0.12	0.04	0.09	-0.02	0.56
GCVI harmonic regression coefficient (first sine)	-0.47	0.32	-0.48	0.14	-0.03	0.20
GCVI harmonic regression coefficient (intercept)	1.25	0.24	1.25	0.13	-0.04	0.30
GCVI harmonic regression coefficient (second cosine)	-0.12	0.23	-0.14	0.09	-0.20	0.16
GCVI harmonic regression coefficient (second sine)	0.04	0.15	0.03	0.11	-0.12	0.55
GCVI harmonic regression coefficient (third cosine)	0.05	0.07	0.06	0.08	0.16	1.27
GCVI harmonic regression coefficient (third sine)	0.05	0.21	0.04	0.07	-0.13	0.11
Grew cowpeas in rainy season	0.36	0.48	0.36	0.48	0.00	
Grew millet in rainy season	0.74	0.44	0.67	0.47	-0.16	
Grew rice in rainy season	0.19	0.40	0.29	0.45	0.21	

	Comparison		Treatment			
Covariate	Mean	SD	Mean	SD	Standardized diff	Variance Ratio
Grew sorghum in rainy season	0.22	0.41	0.23	0.42	0.02	
Groundwater measurement (WATEX band 1)	104.78	66.67	98.61	63.38	-0.10	0.90
Groundwater measurement (WATEX band 2)	149.15	66.24	145.95	65.12	-0.05	0.97
Groundwater measurement (WATEX band 3)	150.32	66.35	157.74	60.71	0.12	0.84
Intercropped in rainy season	0.42	0.49	0.40	0.49	-0.05	
Irrigated plot in rainy or dry season	0.18	0.39	0.31	0.46	0.28	
Plot area (GPS-measured)	1.01	0.97	0.91	0.70	-0.14	0.52
Propensity score	0.33	0.18	0.47	0.28	0.48	2.27
Share of plot cultivated in the rainy season	0.95	0.19	0.92	0.24	-0.12	1.59
Total annual expenses (IHS)	9.68	3.46	10.20	3.77	0.14	1.19
Total annual expenses per hectare (IHS)	9.50	3.42	10.06	3.74	0.15	1.20
Used improved/hybrid seeds for 1+ crops in past rainy season	0.09	0.28	0.13	0.34	0.13	
Used labor during the past rainy season	0.81	0.40	0.70	0.46	-0.23	
Uses well for irrigation	0.11	0.32	0.21	0.41	0.23	
Plot area quintile 1 (GPS-measured)	0.23	0.42	0.23	0.42	0.00	
Plot area quintile 2 (GPS-measured)	0.22	0.41	0.22	0.41	0.00	
Plot area quintile 3 (GPS-measured)	0.20	0.40	0.20	0.40	0.00	
Plot area quintile 4 (GPS-measured)	0.19	0.40	0.19	0.40	0.00	
Plot area quintile 5 (GPS-measured)	0.15	0.36	0.15	0.36	0.00	

Note: Mathematica calculations. Variance ratios are not available for binary variables. N = 970 plots (485 treatment, 485 comparison).

Table C.5. Balance table of key household-level covariates for untrimmed, 1nn without replacement sample

Covariate	Comparison		Treatment		Standardized diff	Variance Ratio
	Mean	SD	Mean	SD		
Male household head	0.95	0.22	0.93	0.26	-0.09	
1+ family member went to bed hungry in previous month	0.21	0.41	0.18	0.38	-0.09	
Ag income (000,000s)	1.36	2.55	4.18	5.19	0.54	4.12
Ag share of total household income	0.38	0.42	0.53	0.40	0.38	0.90
Age of household head	45.16	12.69	43.20	13.84	-0.14	1.19
Dry season cultivated area (self-reported)	0.32	0.69	0.70	0.91	0.41	1.76
Family had employment income	0.11	0.32	0.15	0.36	0.11	
Female has input into productive decisions	0.55	0.50	0.59	0.49	0.07	
Frequency of extreme hunger in lean season	0.04	0.20	0.04	0.20	0.00	
Frequency of no food in lean season	0.05	0.22	0.05	0.22	0.01	
HH head is literate	0.46	0.50	0.47	0.50	0.02	
Has working motorcycle or private vehicle	0.29	0.45	0.28	0.45	-0.01	
Household had self-employment income	0.60	0.49	0.60	0.49	0.01	
Household has irrigation access	0.34	0.47	0.63	0.48	0.62	
Household land holdings per person	0.42	0.32	0.40	0.30	-0.07	0.87
Household size	9.45	5.33	9.42	4.92	-0.01	0.85
Market transport expenses (IHS)	1.30	2.98	1.99	3.65	0.19	1.50
No household food 1+ times in past month	0.29	0.46	0.22	0.41	-0.18	
Non-ag, annual household income (IHS)	7.05	6.42	7.94	6.23	0.14	0.94
Number of household phones	1.81	1.67	1.92	1.68	0.06	1.02
Owns a farm equipment asset	0.24	0.42	0.22	0.41	-0.04	
Owns irrigation pump	0.19	0.39	0.46	0.50	0.55	
Owns large livestock	0.79	0.41	0.75	0.43	-0.08	
Owns plots cultivated by non-household member	0.00	0.00	0.01	0.11	0.11	
Poverty score	29.43	10.12	29.61	10.70	0.02	1.12
Propensity score	0.34	0.19	0.54	0.28	0.70	2.21
Roof made of improved materials	0.26	0.44	0.35	0.48	0.19	
Sold crops during rainy season	0.42	0.49	0.57	0.50	0.28	
Total annual household income (000,000s)	3.18	4.44	6.26	6.69	0.46	2.27
Total expenses (IHS)	11.17	2.27	11.85	2.73	0.25	1.44
Total land holdings (self-reported)	2.99	1.52	3.05	1.80	0.04	1.39
Total rainy season cultivated hectares (IHS)	1.52	0.47	1.49	0.56	-0.05	1.43
WEAI value	0.66	0.48	0.69	0.65	0.04	1.84

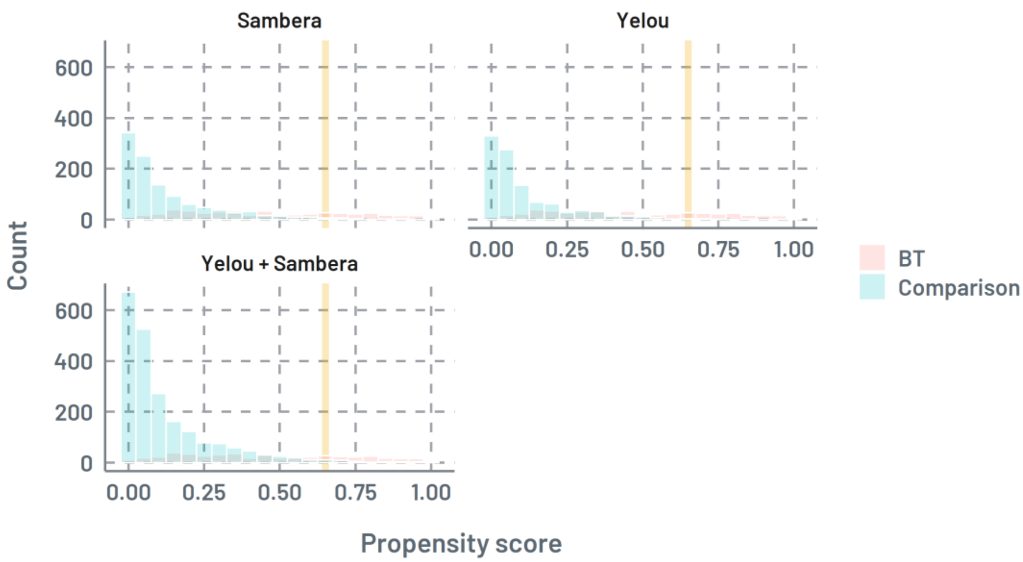
Note: Mathematica calculations. Variance ratios are not available for binary variables. N = 782 households (391 treatment, 391 comparison).

D. Balance results

Plot-level balance

The common support of propensity scores is roughly 0 to 0.65. As seen in **Appendix Figure C.2**, many treatment plots have propensity score values exceeding 0.65, but few to no comparison units do. Our analysis sample consists of all observations with propensity score values less than or equal to 0.65. Since a larger share of target plots have high propensity score values, the restriction to observations in the area of common support excludes a larger proportion of the available *Basse Terrasse* plots, which shrinks from 485 to 337 plots, a 30.5 percent reduction (**Table II.).** In comparison, because of the common support restriction, fewer than 2 percent of Sambera and Yelou plots are excluded from the MCG design.

Figure C.2. Plot-level propensity score distributions by comparison region



Note: Mathematica calculations. Vertical gold lines represent the propensity score threshold of 0.65, with observations to the right dropped from the analysis sample. Error! Reference source not found.**Appendix Table C.2** presents the regression specification to estimate propensity scores. N = 2,626 plots (485 *BT*, 1,103 Sambera, 1,038 Yelou).

Table C.6. Sample sizes available for constructing plot-level matched comparison group sample

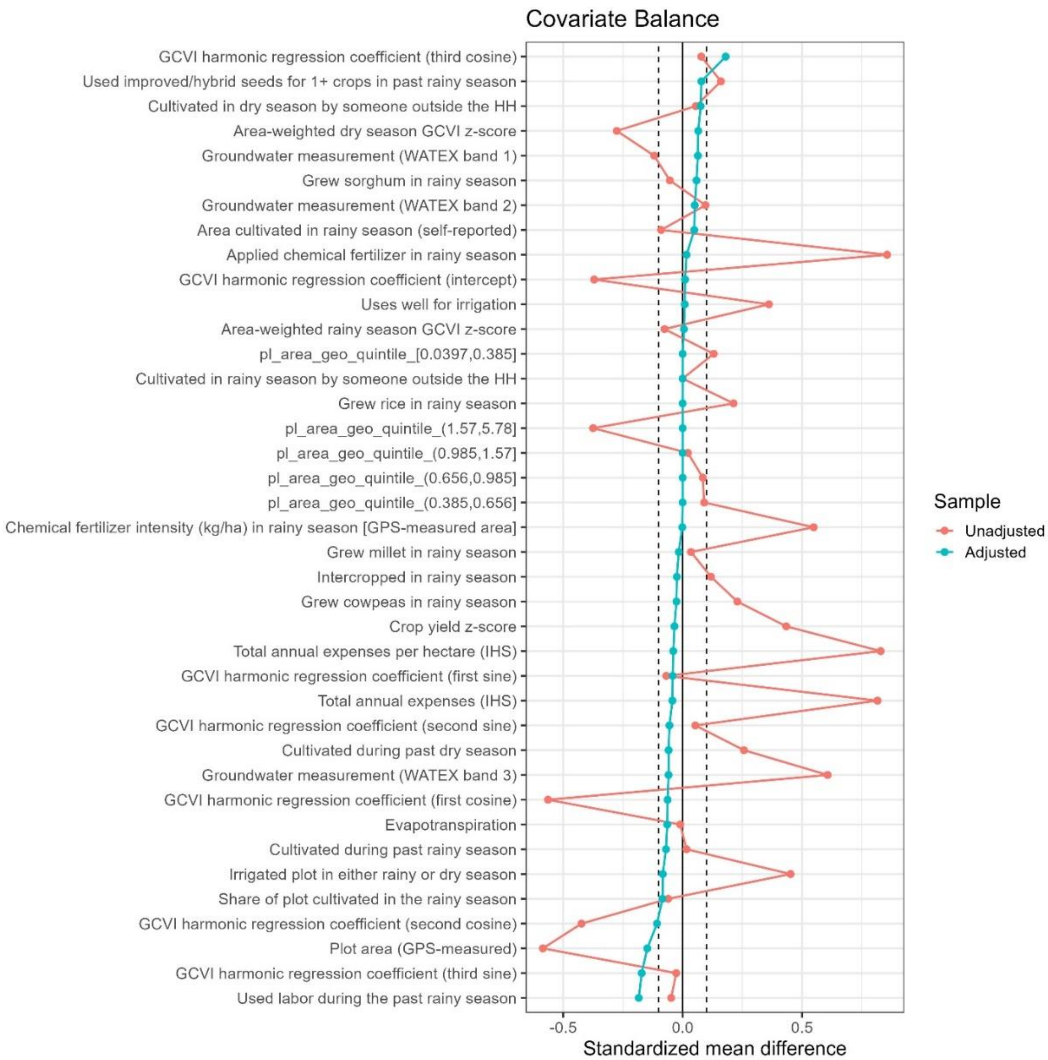
Region	Full sample (N)	Trimmed sample (N)	% of full sample
Basse Terrasse	485	337	69.5
Sambera	1,103	1,088	98.6
Yelou	1,038	1,029	99.1
Total	**2,626**	**2,454**	**93.4**

Note: Mathematica calculations. Analysis sample drops observations with estimated propensity scores > 0.65. Sample sizes do not reflect the actual samples included in the matched comparison group analyses, but instead denote the number of observations available for propensity score matching.

We display covariate balance in **Appendix Figure C.3** and find a strong degree of balance across numerous plot-level covariates. Peach-colored values denote the unadjusted sample means (that is, not applying weights estimated from the matching process) and the teal values represent the adjusted means. Under our 0.1 AMSD criterion, 34 of the 39 assessed covariates were balanced. Key plot-level characteristics that were *ex ante* likely to correlate with headline project outcomes such as agricultural revenue and food security were balanced, including groundwater accessibility, chemical fertilizer application (binary), dry season cultivation (binary), and crop-specific z-scores of crop yields.

None of the unbalanced variables exceeded an adjusted mean difference of 0.2 SD (**Appendix Table C.7**). **Appendix Figure C.3** displays the direction of the imbalance, with teal observations to the right of the solid vertical line, which indicates that the adjusted treatment mean was greater than the adjusted comparison mean. The reverse holds at the bottom of the plot, which is arranged sequentially from largest (most positive) to smallest standardized mean difference. Most notably, one of the GCVI harmonic regression coefficients had a larger adjusted treatment mean than comparison mean (the third cycle, or highest frequency cycle, cosine term), and adjusted means among the comparison group were larger than among the treatment group for rainy season cultivation, the other third-cycle GCVI harmonic regression term, and whether hired labor was employed during the rainy season.

Figure C.3. Treatment-comparison balance of plot-level covariates using 5-nearest-neighbor matching with replacement and common support restriction



Note: Mathematica calculations. Results are for analysis sample of observations with estimated propensity score values less than or equal to 0.65 calculated using the logistic regression model. Error! Reference source not found.N = 1,067 plots (337 treatment, 730 comparison).

In addition to the standardized mean differences displayed above, we also report the adjusted mean and SD for each variable in **Appendix Table C.7**. Values here are helpful in interpreting the materiality of differences displayed in the covariate balance plot. For example, "cultivated during past rainy season" appears as an unbalanced variable, but it is important to recognize that the adjusted means for the comparison and treatment groups are respectively, 96 and 93 percent. For this variable, the small-group-wise SDs are due to a high level of clustering at 100 percent for both groups, which also marks the ceiling value for this variable, and therefore there is little room for post-baseline growth.

Table C.7. Balance table of key plot-level variables, adjusted results only

	Comparison		Treatment			
Covariate	Mean	SD	Mean	SD	Mean diff	Variance ratio
GCVI harmonic regression coefficient (third cosine)	0.05	0.07	0.07	0.07	0.18+	1.13
Used improved/hybrid seeds for 1+ crops in past rainy season	0.09	0.29	0.12	0.32	0.08	
Cultivated in dry season by someone outside the HH	0.00	0.05	0.01	0.09	0.08	
Area-weighted dry season GCVI z-score	-0.06	0.34	-0.04	0.27	0.07	0.62
Groundwater measurement (WATEX band 1)	102.91	64.97	106.96	66.87	0.06	1.06
Grew sorghum in rainy season	0.22	0.41	0.24	0.43	0.06	
Groundwater measurement (WATEX band 2)	149.13	66.79	152.43	66.25	0.05	0.98
Area cultivated in rainy season (self-reported)	1.28	0.64	1.32	0.73	0.05	1.32
Applied chemical fertilizer in rainy season	0.65	0.48	0.65	0.48	0.02	
Propensity score	0.31	0.17	0.31	0.18	0.01	1.05
GCVI harmonic regression coefficient (intercept)	1.25	0.23	1.26	0.14	0.01	0.36
Uses well for irrigation	0.12	0.32	0.12	0.32	0.01	
Area-weighted rainy season GCVI z-score	-0.08	1.00	-0.07	0.76	0.01	0.58
Plot area quintile 1 (GPS-measured)	0.20	0.40	0.20	0.40	0.00	
Grew rice in rainy season	0.19	0.39	0.19	0.39	0.00	
Plot area quintile 2 (GPS-measured)	0.19	0.39	0.19	0.39	0.00	
Plot area quintile 3 (GPS-measured)	0.21	0.41	0.21	0.41	0.00	
Plot area quintile 4 (GPS-measured)	0.21	0.41	0.21	0.41	0.00	
Plot area quintile 5 (GPS-measured)	0.19	0.39	0.19	0.39	0.00	
Chemical fertilizer intensity (kg/ha) in rainy season [GPS-measured area]	0.15	0.27	0.14	0.20	0.00	0.55
Grew millet in rainy season	0.72	0.45	0.72	0.45	-0.02	
Intercropped in rainy season	0.41	0.49	0.39	0.49	-0.02	
Grew cowpeas in rainy season	0.36	0.48	0.34	0.48	-0.03	
Crop yield z-score	0.04	0.73	0.01	0.76	-0.03	1.10
Total annual expenses per hectare (IHS)	9.43	3.49	9.28	4.20	-0.04	1.45
GCVI harmonic regression coefficient (first sine)	-0.47	0.28	-0.47	0.14	-0.04	0.25
Total annual expenses (IHS)	9.63	3.54	9.46	4.27	-0.04	1.46

Covariate	Comparison Mean	Comparison SD	Treatment Mean	Treatment SD	Mean diff	Variance ratio
GCVI harmonic regression coefficient (second sine)	0.04	0.14	0.03	0.10	-0.06	0.56
Cultivated during past dry season	0.14	0.34	0.11	0.32	-0.06	
Groundwater measurement (WATEX band 3)	150.00	65.28	146.42	61.98	-0.06	0.90
GCVI harmonic regression coefficient (first cosine)	0.05	0.11	0.05	0.08	-0.06	0.55
Evapotranspiration (000s)	7.27	1.68	7.20	1.13	-0.06	0.45
Cultivated during past rainy season	0.96	0.20	0.94	0.23	-0.07	
Irrigated plot in rainy or dry season	0.20	0.40	0.16	0.37	-0.08	
Share of plot cultivated in the rainy season	0.94	0.21	0.92	0.25	-0.08	1.41
GCVI harmonic regression coefficient (second cosine)	-0.12	0.19	-0.13	0.08	-0.11+	0.17
Plot area (GPS-measured)	1.10	1.03	1.00	0.74	-0.15+	0.51
GCVI harmonic regression coefficient (third sine)	0.05	0.17	0.03	0.06	-0.17+	0.13
Used labor during the past rainy season	0.76	0.43	0.68	0.47	-0.18+	

Note: Mathematica calculations. Results from the 5nn with replacement matching model on the trimmed sample. Variance ratios are not available for binary variables. N = 1,067 plots (337 treatment, 730 comparison).

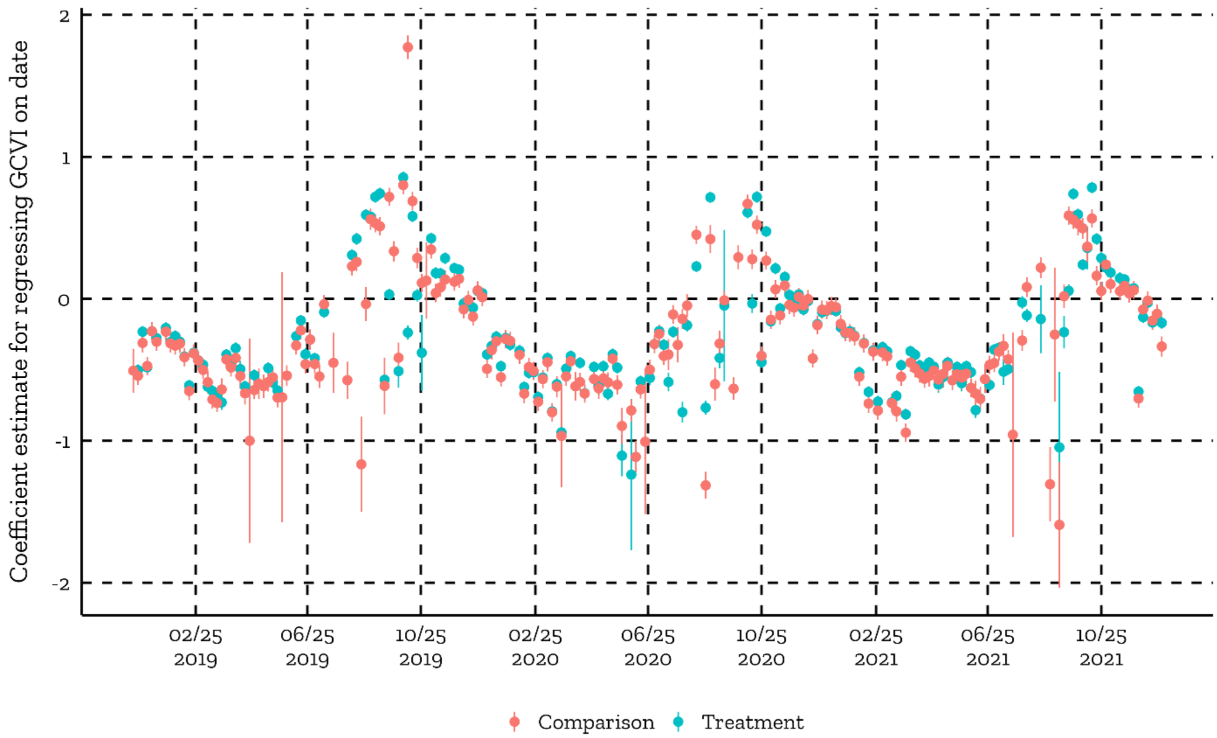
+ Absolute standardized difference exceeds 0.1 SD.

Testing for pre-trends

Although nearly all constructs on which we can perform balance tests either do not change over time (for example, plot size or highest level of education of household head) or are limited to the time period covered in the baseline survey (for example, whether a plot was intercropped during the preceding rainy season), both GCVI and rainfall are continuously available for several years preceding treatment and offer additional opportunities to test for pre-trends. An additional test is to assess whether there is evidence for "parallel trends," whereby time-series data for the treatment group follow the same temporal evolution as comparison group data over the course of the pre-intervention period. This assumption is violated when the two pre-trends are not parallel. One example of such a violation would be if the treatment group's pre-trend slopes upwards while the comparison group's trend slopes downward. This lack of parallel trends suggests that non-project factors were already contributing to differences in outcomes between the treatment and comparison groups. In this context, if a positive treatment effect were estimated it could not be reliably or causally linked to the treatment group's participation in the intervention.

We did not observe any pre-trends in either GCVI or monthly rainfall that would indicate the two areas are on different trends. Appendix **Figure C.4** plots the results from regressing GCVI on the series of dates for which GCVI values could be constructed—that is, days when Sentinel-2 images were sufficiently cloud-free. The plotted coefficients are from separate estimations of treatment and comparison plots that were selected through the 5-nearest neighbor with replacement matching procedure. Each estimate is the daily mean GCVI value, among either treatment or comparison plots, after subtracting the sample mean GCVI value. The results depict a typical signature for cropland with a single agricultural cycle with peak pre-harvest GCVI values occurring in the early fall each year. In general, the patterns for treatment and comparison plots are extremely comparable and do not provide any indication that the two groups were on different GCVI trajectories leading into the start of the SSI intervention.

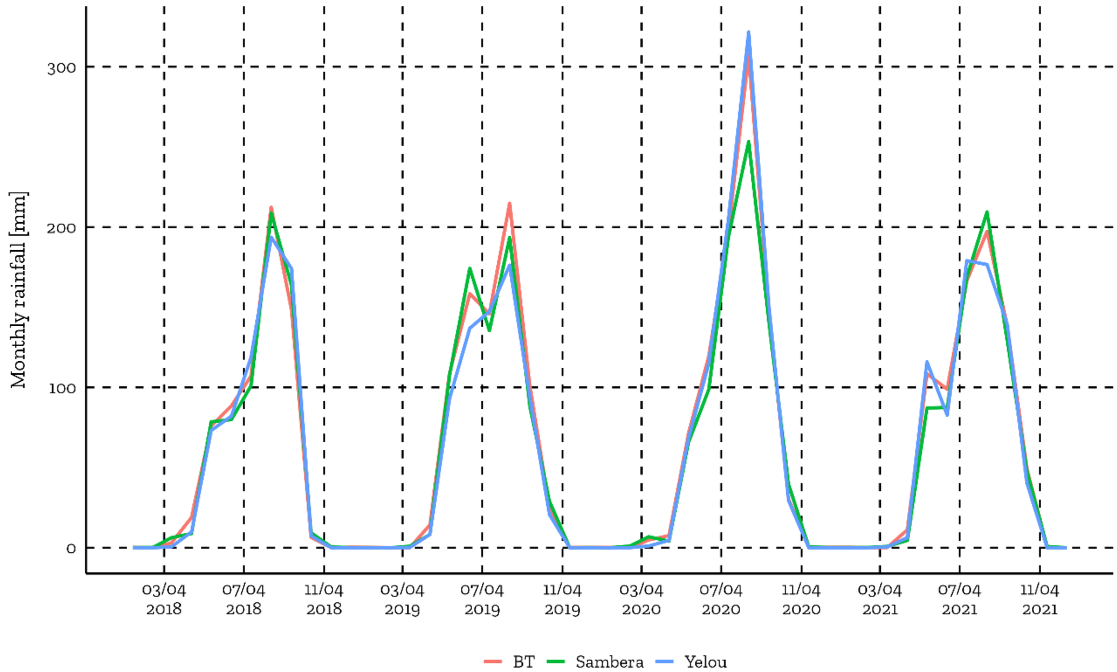
Figure C.4. Visual test for pre-trends in GCVI values



Source: Mathematica calculations using Sentinel-2 data.

We also observed that pre-treatment rainfall between BT and the comparison regions followed comparable patterns; weather conditions across plots are unlikely to be a relevant factor in explaining outcomes.[22] **Appendix Figure C.5** displays monthly average rainfall for the three regions and reveals strong concordance. For example, the 2020 rainy season had the highest monthly rainfall for the period across BT, Sambera, and Yelou, and monthly values were nearly equivalent in all months.

Figure C.5. Visual test for pre-trends in monthly rainfall



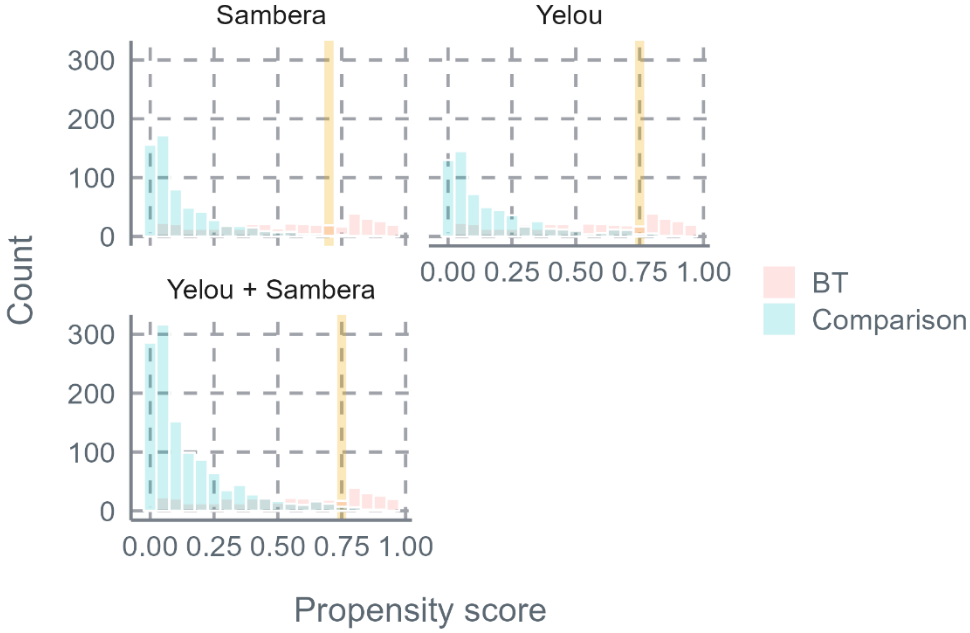
Source: Mathematica calculations using CHIRPS rainfall data (Funk et al. 2015).

Household-level balance

We performed a separate logistic regression (**Appendix Table C.3**) to estimate propensity scores at the household level, with the propensity scores by region displayed in **Appendix Figure C.6**. Households in the *Basse Terrasse* had propensity score values across nearly the entire 0 to 1 range. In contrast, the highest propensity scores observed for Sambera and Yelou households were about 0.7 and 0.75 respectively, shown as the vertical gold line. The largest cluster of households in Sambera and Yelou had propensity score values below 0.15. To implement the common support restriction, we retained all observations with a propensity score of 0.7 or less for Sambera and 0.75 or less for Yelou, which provided us with pre-matching, region-specific sample sizes shown in **Appendix Table C.8**. With this threshold, the impact evaluation will exclude 123 *Basse Terrasse* observations, or about 31.5 percent of the sample, because of the very small number of possible matches in Sambera or Yelou. (In Yelou or Sambera, around one percent of households had a propensity score above the threshold.)

[22] Precipitation has high spatial covariance and many plots within a single region are assigned the same values given the spatial resolution (roughly 6-kilometers) of the CHIRPS data. We therefore do not present plot-level data as with **Appendix Figure C.4**.

Figure C.6. Household-level propensity score distributions by comparison region



Note: Mathematica calculations. Vertical gold lines represent the propensity score threshold value of 0.70 (for Sambera) or 0.75 (for Yelou, Yelou + Sambera), with observations to the right of the line dropped from the trimmed sample. See Error! Reference source not found. Appendix Table C.5 for the regression specification used to estimate propensity scores. N = 1,608 households (391 *BT*, 617 Sambera, 600 Yelou).

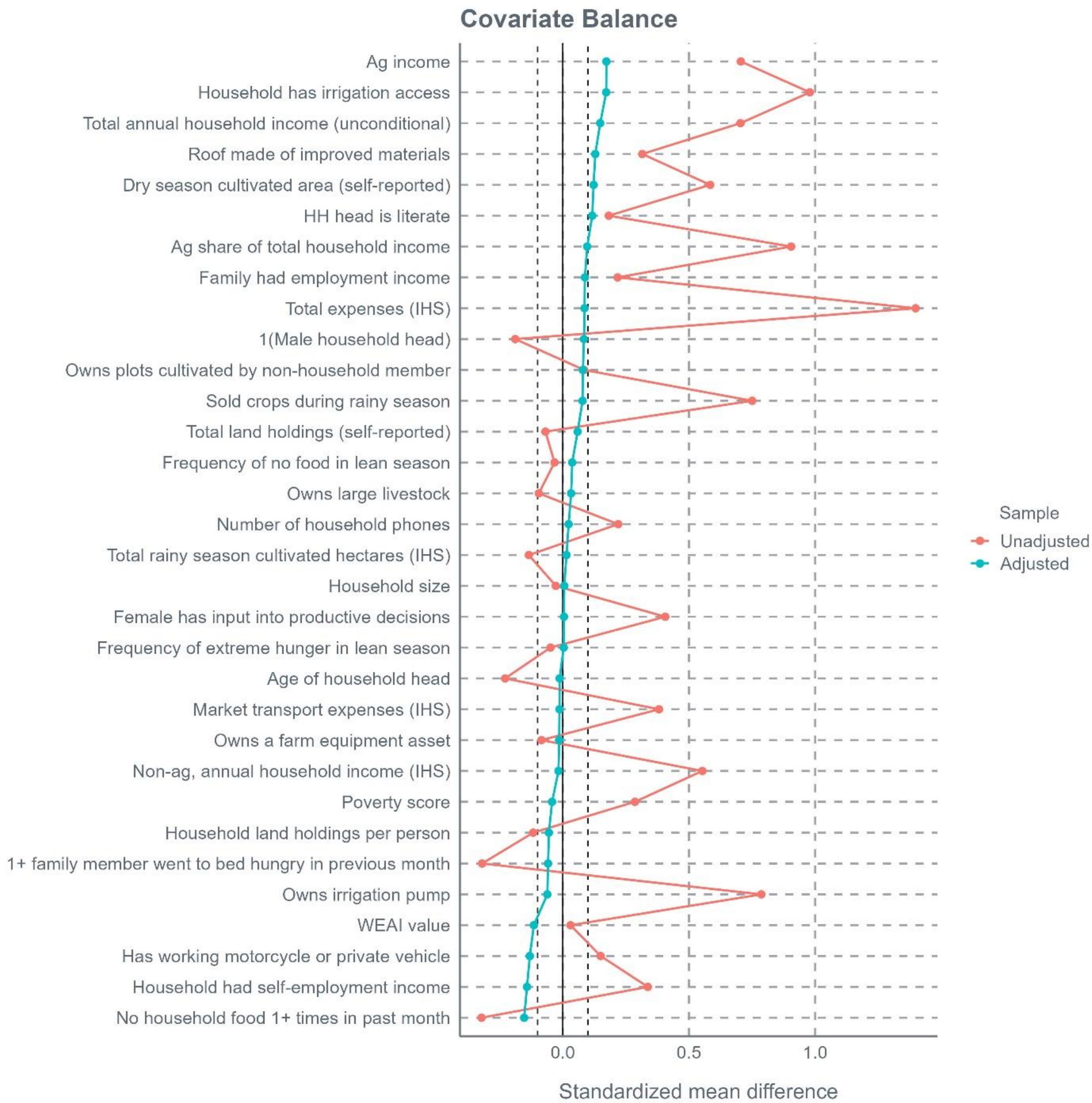
Table C.8. Sample sizes available for constructing household-level matched comparison group sample

Region	Full sample (N)	Trimmed sample (N)	% of full sample
Basse Terrasse	391	268	68.5
Sambera	617	613	99.4
Yelou	600	592	98.7
Total	**1,608**	**1,473**	**91.6**

Note: Mathematica calculations. Households with propensity scores above 0.70 or 0.75 thresholds are dropped in the trimmed sample.

Applying the 0.1 SD threshold, we see in **Appendix Figure C.7** that AMSDs (in teal) are in an acceptable range for important indicators like rainy season cultivated area, poverty likelihood, and several food security metrics. For variables at the top of the figure, adjusted mean values for the treatment group exceed those of the matched comparison group. The figure demonstrates that, even after matching, our *Basse Terrasse* households on average experience better outcomes along some dimensions that are likely attributable to improved growing conditions. They enjoy more irrigation access, which leads to higher dry season cultivation and annual agricultural and total household income. The increased agricultural production also incurs more input expenses than incurred by the comparison group. At the bottom of the figure, adjusted mean values for the comparison group exceed those of the treatment group. For all covariates where the comparison group's adjusted mean value exceeds the treatment group's (such as from poverty score to household had self-employment income), differences are always below 0.1 SD.

Figure C.7. Treatment-comparison balance of household-level covariates using 5-nearest-neighbor matching with replacement on a trimmed sample



Note: Mathematica calculations. Results are for trimmed sample of observations with estimated propensity score values less than or equal to the 0.70 or 0.75 thresholds. N = 758 households (268 treatment, 490 comparison).

In **Appendix Table C.9,** we report the adjusted means and SD as additional context for interpreting the balance performance displayed above. Rows in which the mean difference is close to 0, such as the agricultural share of total household income and owning an irrigation pump, indicate near equality in means between the two groups. The largest standardized differences are less than 0.18 SD. Variance ratios (last column) are highest for financial variables, such as agricultural income and total annual household income. Both examples exhibit substantial dispersion in both treatment and comparison groups, evidenced by SDs that are much larger than mean values. Covariates for which absolute standardized differences exceed 0.1 SD are noted accordingly.

Table C.9. Balance table of key household-level covariates, adjusted results only

Covariate	Comparison		Treatment		Mean difference	Variance Ratio
	Mean	SD	Mean	SD		
Ag income ('000,000)	1.56	2.53	2.46	4.02	0.17+	2.53
Household has irrigation access (0/1)	0.40	0.49	0.49	0.50	0.17+	
Total annual household income ('000,000)	3.22	4.03	4.22	5.39	0.15+	1.79
Roof made of improved materials (0/1)	0.28	0.45	0.34	0.47	0.13+	
Dry season cultivated area (self-reported)	0.39	0.73	0.50	0.84	0.12+	1.32
HH head is literate (0/1)	0.45	0.50	0.51	0.50	0.12+	
Ag share of total household income (%)	0.38	0.43	0.42	0.42	0.10	0.99
Family had employment income (0/1)	0.10	0.30	0.13	0.34	0.09	
Total expenses (IHS)	11.10	2.59	11.33	3.09	0.09	1.42
Male household head (0/1)	0.91	0.28	0.94	0.24	0.08	
Owns plots cultivated by non-household member (0/1)	0.00	0.05	0.01	0.11	0.08	
Sold crops during rainy season (0/1)	0.43	0.49	0.47	0.50	0.08	
Total land holdings (self-reported)	3.00	1.67	3.10	1.89	0.06	1.28
Frequency of no food in lean season (0/1)	0.05	0.22	0.06	0.24	0.04	
Owns large livestock (0/1)	0.78	0.42	0.79	0.41	0.03	
Number of household phones	1.76	1.60	1.79	1.64	0.02	1.05
Total rainy season cultivated hectares (IHS)	1.49	0.51	1.50	0.56	0.01	1.22
Household size	9.34	5.13	9.37	4.92	0.01	0.92
Female has input into productive decisions (0/1)	0.52	0.50	0.53	0.50	0.00	
Experiences extreme hunger in lean season sometime/often (0/1)	0.05	0.21	0.05	0.21	0.00	
Propensity score	0.39	0.22	0.39	0.22	0.00	1.00
Age of household head	43.91	13.17	43.73	14.08	-0.01	1.14
Market transport expenses (IHS)	1.65	3.28	1.60	3.35	-0.01	1.04
Owns a farm equipment asset (0/1)	0.24	0.43	0.23	0.42	-0.01	
Non-ag, annual household income (IHS)	7.03	6.39	6.92	6.29	-0.02	0.97
Poverty score	29.30	10.23	28.85	10.88	-0.04	1.13
Household land holdings per person	0.42	0.33	0.40	0.30	-0.06	0.81
1+ family member went to bed hungry in previous month (0/1)	0.23	0.42	0.21	0.40	-0.06	
Owns irrigation pump (0/1)	0.28	0.45	0.25	0.44	-0.06	
WEAI value	0.68	0.50	0.60	0.57	-0.11+	1.30
Has working motorcycle or private vehicle (0/1)	0.32	0.47	0.26	0.44	-0.13+	
Household had self-employment income (0/1)	0.59	0.49	0.52	0.50	-0.14+	
No household food 1+ times in past month (0/1)	0.30	0.46	0.24	0.43	-0.15+	

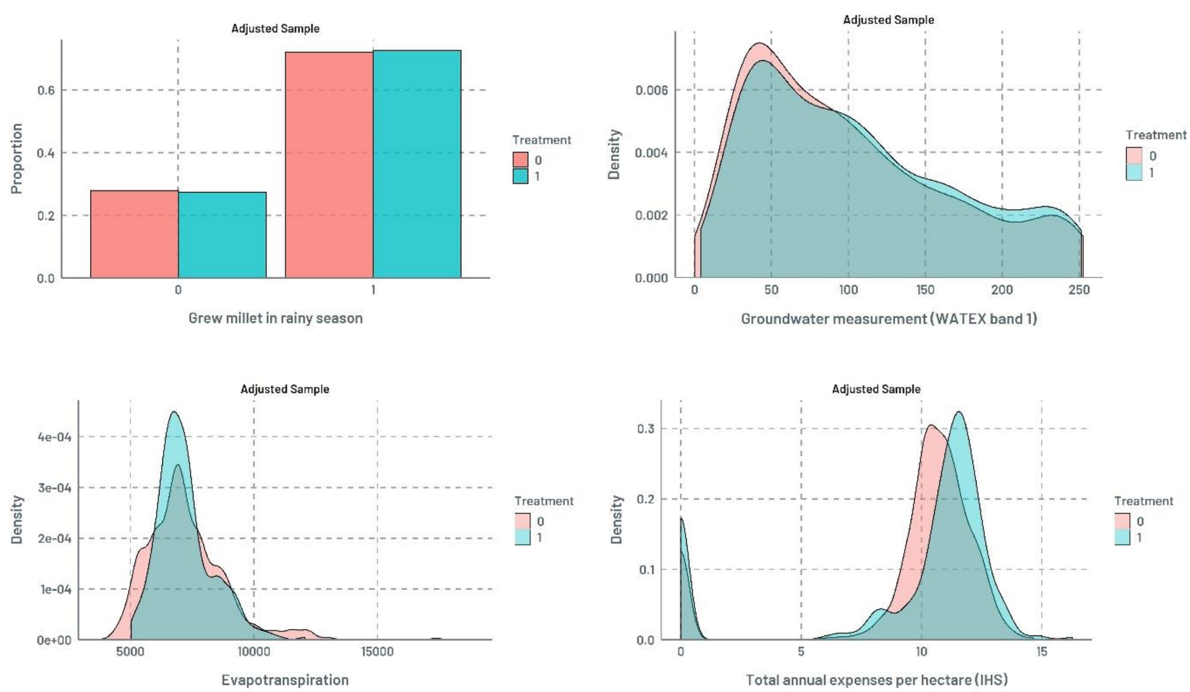
Note: Mathematica calculations. Results from a 5nn with replacement match model on the common support sample. Variance ratios are not presented for binary variables as they do not provide additional information beyond the means. Values are imputed for missing variables, which can lead to differences with the results presented in Section C. N = 758 households (268 treatment, 490 comparison).

+ Absolute standardized difference exceeds 0.1 SD.

Plot-level variable distributions

Aside from assessing balance in the group-wise adjusted means across covariates, we also examined comparability in the *distribution* of adjusted values for covariates on an individual basis. For each variable listed in **Appendix Table C.1**, we produced either a bar plot or kernel density estimate curves to compare the distribution of values, as shown in **Appendix Figure C.8**. For most variables, we observed plots that were like the top two graphs. The left graph, a binary indicator for millet being grown on the plot in the rainy season, shows strong balance in adjusted values. We note that this univariate assessment does not add information beyond what was shown in the figure and tables above, since the mean of a binary provides complete information about the variable's distribution. In contrast, the top-right graph provides an example of the distributions for treatment and comparison groups having a high degree of similarity in the case of a continuous variable. The data are displayed for the first of the three bands of WATEX data, which were generated by RTI Exploration to map groundwater resources through a combination of optical and radar data.[23] While there is some divergence in clustering, with a larger share of comparison units at lower values and more treatment units at higher values, the overall patterns between treatment and comparison groups across the 0–255 range are extremely similar. We note that most covariates exhibited a graphical relationship that most approximates that of WATEX band 1. In a number of cases, we observed less comparability, as seen in the lower two graphs. Evapotranspiration values were more dispersed among comparison plots, even while a common modal value was shared with treatment plots. For annual agricultural expenses per hectare, treatment plots were more dispersed, with a larger chunk indicating no expenses and a more prominent mass of observations reporting higher expenses than the modal respondent in the comparison group.

Figure C.8. Example univariate balance plots using plot-level adjusted sample



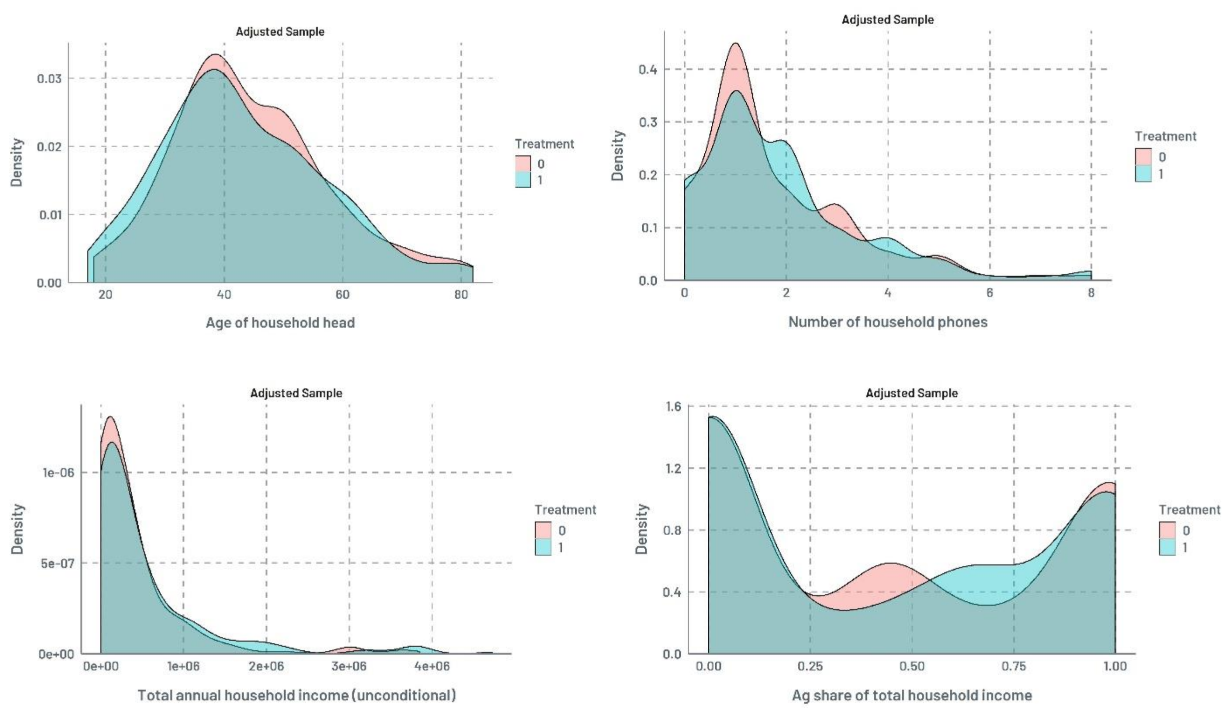
[23] More information about the WATEX System is available at https://www.rtiexploration.com/watex-tm-system.html.

Note: Mathematica calculations. Only adjusted results are shown, based on a 5nn with replacement trimmed sample. N = 1,096 plots (346 treatment, 750 comparison) for all graphs.

Household-level covariate distributions

In addition to assessing balance through means and SD, we also graphically compare the distributions of adjusted values for each individual covariate, with examples shown in **Appendix Figure C.9**. The majority of continuously valued (that is, not binary) covariates exhibited patterns that are like the examples shown. These examples also illustrate the relationship between distributional differences and differences in means. The top-left graph is based on "age of household head," which from **Appendix Table C.9** has an adjusted mean difference of -0.01 (acceptable under our chosen threshold). While both treatment and comparison groups have a late-30s modal value, there is more mass of comparison units in the 40–50 and the 70–80 ranges. Furthermore, the distribution of comparison households has shifted to the right on the left edge of the support; from the youngest reported household heads through to the mid-30s, young household heads account for a larger share of treatment households, which leads to the treatment group having a lower mean. The bottom-right graph of "total annual household income" demonstrates that the mean difference in that variable (0.15 SD) is driven by a longer right-tail and clusters at around 2 million and 3.7 million CFA, which are more pronounced than among the comparison households.

Figure C.9. Example univariate balance plots using household-level adjusted sample



Note: Mathematica calculations. Only adjusted results are shown, based on a 5nn with replacement trimmed sample. N = 811 households (289 treatment, 522 comparison) for all graphs.

Appendix D:

Comparative Analysis of Area Cultivated and Yields Between Survey Data and Area Measured Using Handheld-Devices Analysis

This appendix presents information on several comparisons of plot size and yields, (1) a comparison of self-reported plot sizes and plot sizes measured using GPS measurements, (2) a comparison of crop yields based on self-reported and GPS-measured area cultivated for the main crops grown, and (3) a comparison of crop yields for plots in which crops are grown on their own (pure-stand) and those in which several crops—typically cowpeas with millet— are intercropped.

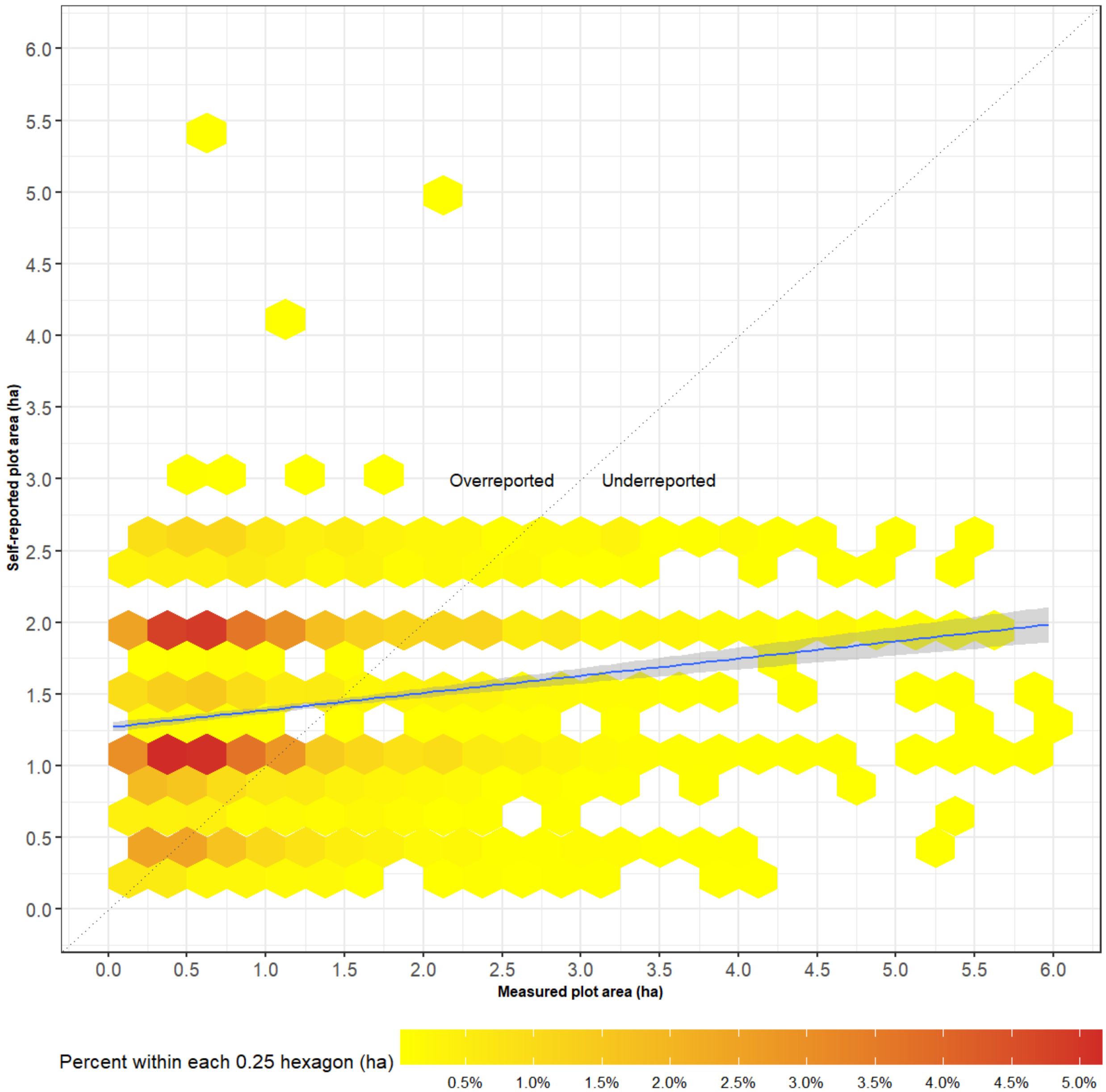
Figure D.1 shows that self-reported plot areas tend be weakly correlated with measured plot areas. As found in the literature, farmers tend to overreport plot areas on average, but the accuracy and directionality of reporting varies by plot size. For small-to-average sized measured plot areas (below 1.5 ha), farmers tend to overstate plot areas, whereas for larger measured plot areas, farmers tend to underreport self-reported areas. Some variation could possibly be explained by how farmers estimate plot size when asked to self-report report area. Farmers typically report plot sizes to the nearest half- hectare (89 percent of self-reported plot areas are recalled by farmers at a hectare or half of a hectare reporting level). However, for around 64 percent of plots, the difference between measured and self-reported plot size falls outside of half of a hectare range, suggesting that internal rounding by farmers may only partially explain discrepancies in reported plot area.

Table D presents the yields estimated based on self-reported and measured data side-by-side for the main crops grown in the dry and rainy season. The first three panels split the evidence by whether the plot was inter-cropped or a pure-stand plot.

Table D complements

Table II.63 and shows that the imbalances for standardized differences in yields are much smaller for cowpeas when intercropped plots are included.

Figure D.1. Comparison of self-reported and measured plot areas



Notes: The figure compares self-reported plot areas against measured plot areas for 3,103 collected plots. The shading gradient captures the percent of plots where the self-reported and measured plot intersect at each possible 0.25 ha bandwidth. Plot areas above 6.0 ha are not included in the figure. The linear prediction is represented by the sold blue line.

Table D.1. Comparison of yields based on self-reported and GPS-measured cultivation areas for pure-stand, intercropped, and all cropping methods

Indicator	Mean, measured	Mean, self-reported
Yields for pure-stand crops in rainy season (June–Sept 2021) (t/ha)		
Millet	1.25	1.06
Cowpeas	0.49	0.28
Sorghum	1.21	0.95
Rice	2.22	1.45
Plot sample size	2,125	2,067
Yields for intercropped crops in rainy season (June–Sept 2021) (t/ha)		
Millet	2.08	1.58
Cowpeas	0.72	0.49
Sorghum	1.72	1.18
Plot sample size	1,096	1,073
Yields for all cropping methods in rainy season (June–Sept 2021) (t/ha)		
Millet	1.67	1.33
Cowpeas	0.68	0.46
Sorghum	1.49	1.08
Rice	2.22	1.45
Plot sample size	3,106	3,025
Yields for all cropping methods in dry season (Oct–May 2021) (t/ha)		
Rice	3.36	2.28
Plot sample size	533	515

Notes: Plot sample sizes represent the number of plots with sufficient information to calculate yields from subplot areas. All collected plots are included. For intercropped cropping systems, subplot areas are allocated based on the ratio of the average partial LER for a crop intercropping system relative to the average LER for the whole cropping system from Namatsheve et al. (2020). Pure-stand crop yields are top-coded based on INRAN yield potentials. To top-code intercropped crop combinations, INRAN yield potentials are adjusted by the average partial LERs from Namatsheve et al. (2020). Rice is not intercropped. There is no intercropping in the dry season.

Table D.2. Crop yields based on GPS-measured subplot areas for pure-stand, intercropped, and all cropping methods

Indicator	Treatment mean	Comparison mean	Standardized difference	P-value
Yields for pure-stand crops in rainy season (June–Sept 2021) (t/ha)				
Millet	1.17	1.17	0.00	0.99
Cowpeas	0.71	0.35	0.50	0.19
Sorghum	1.06	1.30	-0.27	0.33
Plot sample size	239	419		
Yields for intercropped crops in rainy season (June–Sept 2021) (t/ha)				
Millet	1.82	1.84	-0.02	0.90
Cowpeas	0.76	0.40	0.31	0.03**
Sorghum	1.59	1.45	0.09	0.71
Plot sample size	133	287		
Yields for all cropping methods in rainy season (June–Sept 2021) (t/ha)				
Millet	1.54	1.58	-0.03	0.76
Cowpeas	0.75	0.39	0.34	0.01***
Sorghum	1.39	1.40	-0.01	0.95
Plot sample size	313	674		

Notes: Plot sample sizes represent the number of plots that grew millet, sorghum, or cowpeas in the past rainy season for a specified cropping method and provided enough information to calculate yields from measured subplot areas. For intercropped cropping systems, subplot areas are allocated based on the ratio of the average partial LER for a crop intercropping system relative to the average LER for the whole cropping system from Namatsheve et al. (2020). Pure-stand crop yields are top-coded based on INRAN yield potentials. To top-code intercropped crop combinations, INRAN yield potentials are adjusted by the average partial LERs from Namatsheve et al. (2020).

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test

Appendix E:

Historical Precipitation Totals for *Basse Terrasse* and Comparison Areas

In this appendix, we examine historical rainfall to determine whether crop yields collected for the dry and rainy seasons of the 2020–2021 agricultural campaign might be attributable to anomalous rainfall totals. We computed monthly rainfall for each of the three regions from 1983 through 2021 as the spatial average over a region's bounding box using data from Tropical Applications of Meteorology using SATellite data and ground-based observations (TAMSAT) (Maidment et al. 2014; Tarnavsky et al. 2014; Maidment et al. 2017).[24] We then estimated differences from the region-specific climatology or long-term mean – using the 1983—2019 period as the long-term mean. These differences are called *anomalies* in the literature. Positive anomalies are when annual rainfall for a region-year exceeds that region's long-term mean annual rainfall, and negative anomalies are the opposite.

For all three regions, annual rainfall totals for 2020 and 2021 exceeded their long-term average totals, as shown in **Table E.1**. For all three regions, rainfall in 2020 and 2021 was higher than their long-term means for 1983–2019. In 2020, that difference ranged from being 72 mm higher (Sambera) to 81 mm higher (*Basse Terrasse*), or about 12 percent of mean annual rainfall. The 2021 anomalies compared to the long-term mean were smaller, ranging from 45 mm (Yelou) to 63 mm (*Basse Terrasse)*, or 7–9 percent of mean annual rainfall.

Table E.1. Comparison of 2020/2021 precipitation against long-term means

Region	Annual precipitation (mm)				
	Long-term mean	2020	2020 anomaly	2021	2021 anomaly
Basse Terrasse	687	768	+81	750	+63
Sambera	655	727	+72	716	+62
Yelou	653	729	+76	698	+45

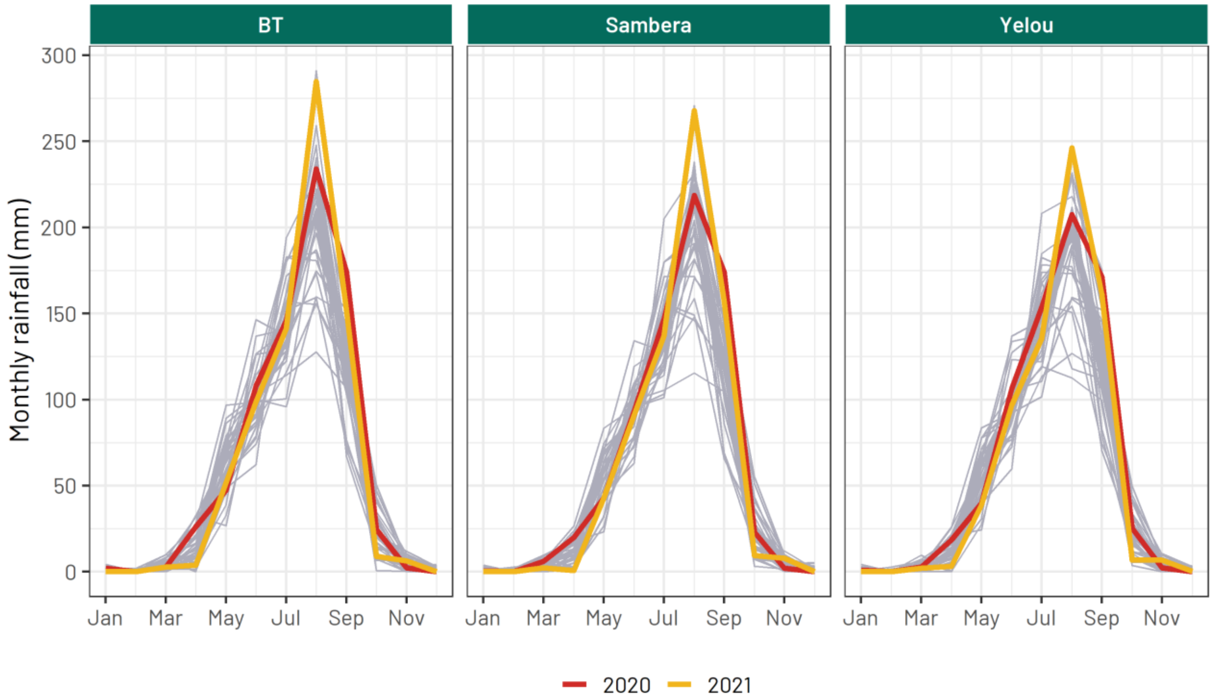
Source: Mathematica calculations using TAMSAT rainfall data (Maidment et al. 2014; Tarnavsky et al. 2014; Maidment et al. 2017).

Note: Long-term mean is estimated from the 1983–2019 period and based on zonal averaging over each region's bounding box constructed from plot outlines included in the full sample. All values represent January 1 to December 31 periods. N = 39 annual observations.

Though the monthly distribution of precipitation values for 2020 were comparable to the long-term climatology (**Figure E.1.**), rainfall in 2021 displays two key differences against historical patterns. The rainy season onset occurred one month later than usual, with April experiencing no precipitation compared to a typical year in which up to 30 mm would fall. Second, August 2021 was among the rainiest months experienced over the 1983–2021 period for these locations, shown in gold. While this was followed by a particularly rainy September, the end of the rainy season was more abrupt in 2021 than in other years. October saw very little rain, whereas on average 20–40 mm of rain would fall in that month.

[24] TAMSAT provides a time and area subsetting tool at http://www.tamsat.org.uk/data-subset/index.html to minimize data-processing requirements. The bounding boxes for *Basse Terrasse*, Sambera, and Yelou are constructed using the following [(xmin, ymin, xmax, ymax)] values respectively: [(3.181651, 12.043130, 3.285926, 12.148677)], [(2.823148, 12.216779, 3.111924, 12.576128)], and [(3.438265, 12.078008, 3.611648, 12.361487)].

Figure E.1. Monthly precipitation for the BT and comparison areas by year, 1983–2021



Source: Mathematica calculations using TAMSAT rainfall data (Maidment et al. 2014; Tarnavsky et al. 2014; Maidment et al. 2017). Each gray line denotes estimated average monthly rainfall for the given region for a year in the 1983–2021 period. N = 468 monthly observations.

It is possible that yields measured for the 2020-2021 period would be above average because of these positive rainfall anomalies. Since above-average rainfall also fell on the comparison areas, our matched comparison group design will still enable us to estimate an unbiased treatment effect regardless of which weather conditions are experienced during the interim or endline data collection periods.

Appendix F:

Stakeholder Comments

Table F.1. MCC Stakeholder comments

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
1	1	*Overview of IMAP*	Isabel Dillener, DCO/SEC-HCD	I believe the Compact value is now $442M	We have updated the compact value.
2	1	*Overview of IMAP*	Isabel Dillener, DCO/SEC-HCD	Do we want to mention the literacy training?	We have included financial literacy in the paragraph describing the theory of change.
3	1	*Overview of IMAP*	Isabel Dillener, DCO/SEC-HCD	Please review IMAP current value, I'm not sure what it is, but I don't think it is $250M	We have updated the current value using MCC input.
4	1	*Overview of IMAP*	Kaj Gass, DCO/SEC-AL	It is rather late to be introducing the following aspects as key to SK2 but it would be good for the records to have some mention of the following: - An innovative approach for determining where to find GW using Lidar was used - This area is a sensitive ecosystem (RAMSAR) and MCC went to particular efforts to make sure no detrimental impacts were caused.	We plan to conduct qualitative interviews with project implementers in the fall of 2023 to obtain information on project design and implementation. These two aspects fall under the quality of the design process and implementation and will be covered by the protocols.
5	1	*Overview of IMAP*	Kaj Gass, DCO/SEC-AL	Here is where literacy should certainly be mentioned. Strengthening of land tenure seems important to add to context as well, as land sharing is an element that is being tested.	We have included the reference to literacy and financial literacy, as well as land tenure. Given the updates we received on sharing of irrigated land, we also refer to land sharing arrangements.
6	4	*Research Questions, Table I.1*	Hamissou Samari, DPE/EE-ME	I know this might be hard to do, but any chance these questions could be grouped into the 2 or 3 EQ categories as per the new guidelines and template? Also, any specific EQs that do not apply to both Konni and SK2? In other words are there EQs that only apply to any of the 2 sites? Or are they identical?	We note in the text that all of these research questions fall into the rubric of the second of MCC's evaluation questions. Did the investment produce the intended results? Did it achieve its stated objective in pursuit of MCC's mission to reduce poverty through economic growth? We compare the research questions specific to SK2 and Konni with in the EDR.
7	4	*Research Questions, Table I.1*	Aaron Szott, DPE/EE-EA/PSC	These are all interesting, but what about the value of production as well? To the extent that output is consumed at home rather than traded, focusing only on e.g. quantities sold will understate beneficiary HH living standards.	While the value of home consumption is not specifically mentioned in the indicators, we do include the estimated value of home consumption in the relevant indicators. We update the relevant variable label to clarify this: "Total agricultural and non-agricultural income, including own consumption (FCFA)", "Agricultural income, including own consumption (FCFA)". The total value of production less expenditures constitute the overwhelming part of agricultural income.

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
8	6	*Fig I.3, Overview map of proejct area*	Kaj Gass, DCO/SEC-AL	How far back has mathematica gone in ensuring that similar programming was not given to these areas? Another question that pops into my mind when looking at this map is around where groundwater is present. One of the big innovative aspects of this project was to use Lidar to better target investments. Even within zones we are seeing quite different access to water. All of this is to say that many donors have opted for irrigation investments that are surface water. It would be good to see the comparison of outcomes between these producers and SK2 producers who should be more likely to produce HVA over rice	On the question of similar programming being given to the comparison areas, our local consultant spoke with local government officials about other projects in the areas. The local government officials confirmed that there were no ongoing or previous programs that they recalled. However, we did not ask about a specific time period during these conversations. On the second question, we agree that in the endline evaluation it would be interesting to compare the outcomes from plots that use flooding as a means of irrigation and those that use drip irrigation. We can include this analysis at endline.
9	6	*Timeline for data collection, evaluation and project activities*	Hamissou Samari, DPE/EE-ME	Evaluation is based on the initial geographical boundaries of SK2. Prior to SK2.1	We address your request for clarifying the project area by writing: Given several iterations of the project activities in the Basse Terasse, it is important to note that the project area described here corresponds to what is informally knows as SK2 which targeted small-scale irrigation for over 600 hectares of land.
10	7	*Key baseline findings, SSI*	Kaj Gass, DCO/SEC-AL	Do plots that cultivated rice in the dry season use surface water?	Among treatment plots that cultivated rice in the dry season (and not restricted to only those plots selected in the MCG analysis sample), 30% (of 100 plots) reported irrigating using surface water sources -- the remainder irrigated using tube wells or other wells. We have added this information in the relevant paragraph.
11	8	*Key baseline findings, Implications for evaluation of SSI investments*	Kaj Gass, DCO/SEC-AL	MCA's consultant SONED has very exact land tenure arrangements mapped out. Although, a unique aspect of this work will again be land sharing where "landless" are brought on to irrigated sites but without changing titles.	The updates to the project scope and land rental and sharing arrangements since this comment suggest that having precise geo-coded information on land cultivated by respondents remains important.
12	10	*OVerview of matched comparison group design*	Sarah Lane, DPE/EE-ME	I wonder if this level of detail is needed on the matched comparison design. Is there a way to explain the design and put some of the more technical pieces in an annex?	We have restructured the text and shifted much of the technical material to a newly revised Appendix C.
13	10	*OVerview of matched comparison group design*	Aaron Szott, DPE/EE-EA/PSC	Agrees that this level of detail on matched comparison design should go in annex. It would be interesting to have a sense of how well-matched the comparison groups are to the treatment group.	We have shifted much of the technical content explaining the MCG and the results to an expanded version of Appendix C. The appendix features tables that report out the degree of balance between treatment and comparison groups, as well as balance plots that visualize the same information, and allow for comparisons between the adjusted and unadjusted samples.

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
14	10	*OVerview of matched comparison group design*	Hamissou Samari, DPE/EE-ME	In the revised version, would it be possible to adjust the methodology to the new version of SK2? Will the rescoping affect the applicability of the current methodologies? We expect a smaller geographical area and a smaller target population as a result of the recent changes With a smaller geographical area comes a smaller target population, which might affect the sampling power. Any chance the proposed methodology might not apply to SK2.1 due to the size?	After discussion with the MCC project monitor, we agreed to finalize this baseline report with the larger set of potential treatment plots, as the smaller scope will necessitate an evaluation update. The updated matched groups will be part of the evaluation update.
15	10	*Overview of matched comparison group design*	Bob Fishbein, DCO/IEPS-WSI	I agree with Hamissou. It would be useful to eventually incorporate a short summary of the final scope. For the moment, we have the initial 100 sites, with a final version likely available by end of January, as a result of the groundwater assessment.	
16	11	*Overview of matched comparison group design*	Aaron Szott, DPE/EE-EA/PSC	Assuming this is correct, it might be helpful to simply state that the matched comparison group is meant to resemble the treatment group with respect to both plot-level and HH factors.	Thank you - we have added this as a clarifying statement in the text.
17	13	*A.2 Propensity score model and balance*	Aaron Szott, DPE/EE-EA/PSC	Why is treating plots as independent of the households preferable to simply matching on both plot and HH characteristics	We explain in the accompanying footnote that separating plots from households allows for improved balance and a higher chance of calculating a valid estimate for the project impacts.
18	25	*Table II.9 Primary outcome indicators for matched comparison group analysis*	Andrew Tarter, DCO/SEC-GSI	Earlier in this report you noted "To obtain balance for the various empowerment domains the final evaluation will implement a separate matching procedure for women." (making this note for myself)	No action taken.

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
19	27	*Table II.10. Primary outcome indicators for the Roads for Market pre-post analysis*	Sarah Lane, DPE/EE-ME	Is it possible to flag the likely exposure period for some of these indicators? I wondered if many of them will require a longer exposure to see change.	We are providing our best estimate of the amount of time it would take to see effects on these indicators based on the literature: - Traders present at markets and villages: 1-2 years as traders take advantage of lower vehicle operating costs and opportunities to profit from areas that are newly better connected to markets. - Transportation method : 6 months to 1 year to adopt new methods of transportation - Travel time: immediate change but could also decrease further over 6 months to 1 year as traders and farmers adopt new vehicle types. - Transportation cost and cost per kilometer: costs for traders should go down almost immediately. Price charged to farmer for transporting goods should take about 1-2 years to decrease. This depends on the adopting of new types of vehicles and an increase in the number of people providing transportation serves so it takes a bit longer to observe. - Value of crops transported by traders: difficult to say because it depends on so many different factors. - Perceptions of road quality: immediate - Crop types produced by farmers: difficult to say but will likely take several years. Requires time for a price signal and investments by farmers to change the types of crops they are growing. Likely also requires farmers to observe higher prices for certain crop types for several years to determine that the pattern is sustainable and the investment/transition to different crops is worthwhile. - Crop types bought by traders and sold by farmers depends on the change in crop types produced by farmers. Likely requires several years. - Damage to perishable crops in transit: immediate - Decreased crop sales price due to crop damage in transit: should be immediate. Could take a bit longer (6 months to 1 year) if buyers don't recognize the decrease in damage to crops in transit.
20	28	*Household demographics*	Hamissou Samari, DPE/EE-ME	How do you define "household" in the context of this survey? The difference might stem from the way each survey defines the term	We define a household as "the group of people who live together, eat food prepared in the same pot, and recognize one person as the head of household. Every person who has lived at least 6 months in the household or intends to stay for at least 6 months in the household is considered a member of the household." This is the same definition used by the Living Standards Measurement Survey (LSMS). The Population Reference Bureau appears to have removed their data on household size from their website. We have therefore updated this paragraph to use the Niger LSMS data instead since the LSMS uses the same definition of a household as our survey and 2014 is the most recent reliable data we could find on household size.
21	28	*Household demographics*	Andrew Tarter, DCO/SEC-GSI	(Correction of table no. in text) It's Table II.1 in text but Table II.11 in title of the table.	We corrected the table number.

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
22	28	*Table II.11 Household demographcs*	Hamissou Samari, DPE/EE-ME	P-values seem quite high on most indicators. Concerns about statistical significance? Maybe sample sizes too small?	Large p-values here are preferred, since they would indicate that there are not statistically meaningful differences between the two groups and therefore, they are comparable. Baseline comparability improves our confidence that any endline differences are due to the project. We have included an explanation in the text.
23	28	*Table II.11 Household demographcs*	Isabel Dillener, DCO/SEC-HCD	Just wondering why 16 was chosen here; is that the age of adulthood in Niger?	The age of majority in Niger is 21 but the age to consent to marry is 18 for adolescent males and 15 for adolescent girls. We did not want to treat males and female differently, so we settled on 16. Males as young as 16 are sometimes plot decision makers or the head of the household (when older male household members are deceased) so it is reasonable to count them as adults. We used this same cutoff of age 16 in Konni as well.
24	28	*Table II.11 Household demographcs*	Andrew Tarter, DCO/SEC-GSI	What does this asterix next to women's empowerment score refer the reader to?	We have added supplemental text in the footnote to explain that reported values are not simple averages, but rather-following how this is defined in the WEIA-constructed at the level of the subgroup (either treatment or comparison). We use the following equation: Subgroup Score = 1 – (weighted percent of households in subgroup that are unempowered * weighted mean adequacy score among unempowered in subgroup).
25	29	*C.2. Land holding, land use, and land tenure security*	Kaj Gass, DCO/SEC-AL	This must be the surface water/swamp areas. The term irrigated might be a bit misleading in that sense.	The survey data indicate that of the BT plots scheduled to receive SSI that reported using irrigation in the past dry season, the majority (67%) are obtaining water for irrigation from wells. Other sources of irrigation include canals (21%), river (10%), and flooding (3%) which would include surface water or swamp areas. Keep in mind that this figure of 0.31 ha of irrigated land in the dry season is at the household level. Given average household landholdings of 1.63ha, this indicates than less than 20% of BT household land is irrigated in the dry season.
26	29	*C.2. Land holding, land use, and land tenure security*	Hamissou Samari, DPE/EE-ME	IS this finding of fomalized plot rights coming from producers or landowners? Has this been confirmed by COFOCOM? The definition of land dispute might vary depending on the respondent. Another potential outcome of the intervention would be a spike on disputes as more and more respondents become aware of their rights and what they stand to gain/lose by defending and protecting those rights (@Elbow, Kent M (DCO/SEC-AL/Contractor) would correct me if I'm wrong here)	This information is self-reported by producers (who are also landowners in most cases) and not confirmed by COFOCOM. We asked all the plot decision makers what kind of documentation they had for their plot. But almost all the plot decision makers (or another household member) are landowners because there are almost no renters at baseline. We categorize formal land rights as having a rental document, contract to cultivate, contract to occupy, or sale document. Regarding land disputes, we agree the definition of disputes may vary depending on the respondent. However, if someone reported a dispute, we ask follow up questions about who the dispute was with, what it was over, and what the dispute involved (verbal disagreement, threats of violence etc.) which helps us provide additional context. At endline we can see if there is a change in these characteristics of disputes in addition to measuring changes in the prevalence of disputes. If disputes do spike due to increased awareness of land rights, we should be able to provide some insight.
27	29	*C.2. Land holding, land use, and land tenure security*	Kent Elbow	In response to Hamissou: Agree that a spike in reported conflicts is a potential outcome, but continued low level of conflict is also a potential outcome.	No action taken.
28	29	*C.2. Land holding, land use, and land tenure security*	Kaj Gass, DCO/SEC-AL	Surprised fear of risk of involuntary loss over next five years isnt more given the SK1 intervention	No action taken.

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
29	29	*C.2. Land holding, land use, and land tenure security*	Sarah Lane, DPE/EE-ME	Is the imbalance in total land holdings less than 0.25 SD mean something needs to happen in the analysis due to this?	In the methodological section 4.2, we describe how we approach unbalanced variables in the final evaluation: For outcomes whose standardized difference is less than 0.25 SD, our regression models for the final analysis include control for the baseline value in a regression of these outcomes on a treatment indicator. This regression model is the preferred means to adjust for imperfect balance at baseline. When differences are larger than 0.25SD, we instead use a difference-in-differences specification for the matched treatment and control groups. This compares the changes over time between the treated and the control group. The first specification is our preferred specification for two primary reasons: 1) including baseline values as controls is efficient and therefore increases power (McKenzie, 2012), and 2) this approach allows for differences in measurement between the baseline and endline measures.
30	30	*Table II.13. Land use and crop choice*	Hamissou Samari, DPE/EE-ME	Interesting that both treatment and comparison are quite similar.	Our plot-level matching approach was motivated by the goal of the treatment and comparison plots being as similar as possible for the outcomes of interest.
31	30	*Table II.13. Land use and crop choice*	Hamissou Samari, DPE/EE-ME	Treatment mean of 0.4 for intercropped plots is surprisingly low	We will be able to validate the share of intercropped plots when we conduct crop cuts.
32	31	*C.3. Irrigation use and availability for cultivated plots*	Sarah Lane, DPE/EE-ME	Can we discuss more about the regression adjustment to address potential biases in the analysis of dry season practices when conditioning on baseline dry season cultivation?	In the methodological section 4.2, we describe how we approach unbalanced variables in the final evaluation: We propose to include the unbalanced variables in the outcome regressions following the approach suggested by Lin (2012) and List et al. (2021) who show bias reduction and higher power of this technique. Lin (2012) "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." The Annals of Applied Statistics 7, no. 1 (2013): 295–318. List, John A., Azeem M. Shaikh, and Atom Vayalinkal. (2021) Multiple testing with covariate adjustment in experimental economics. No. 00732. The Field Experiments Website, 2021.
33	32	*Table II.15. Fertilizer*	Aaron Szott, DPE/EE-EA/PSC	Is the unit for Quantity of inorganic fertilizer applied t/ha?	We corrected the indicator name. It is indeed t/ha.
34	33	*C.5. Seeds and improved agricultural practices*	Kaj Gass, DCO/SEC-AL	How were the nine categories of improved inputs/practices determined? Do they link up with what SAA/Cowater tried to impart on beneficiaries?	The nine categories of improved inputs or practices are: zero tillage land preparation, planting seeds in rows, improved seeds (i.e. not re-using seeds from previous harvest), improved water and soil management techniques (this includes are zaï, tassa, agricultural half-moon, fences, stone walls, silviculture benches, and adding lime to soil), using mechanized equipment, applying inorganic fertilizer, applying pesticides or herbicides, processing crops after harvest, and storing crops in hermetic bags.
35	34	*C.6. Credit and expenditures*	Kaj Gass, DCO/SEC-AL	Maybe this is a question to be posed within MCC but how does the avg annual agricultural expenditure compare to ERR baseline costs?	No change made.

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
36	34	*C.6. Credit and expenditures*	Aaron Szott, DPE/EE-EA/PSC	(In response to Kaj) First, I wonder if 4*$142=$568 gives the value of crop sales or the value of crop production. Anyway, these revenues and costs give an ag profit/ha around $426, whereas I have them as equal to $265 in the CBA. (Ag profits/ha are closer to $322 in the control sample though.) So the difference is non-trivial, but then again for CBA purposes a couple hundred dollars does not make a big difference (even if the ERR should perhaps be a bit lower given this information).	In the revised version of the baseline report, the estimated profit/hectare total crop production in the rainy season is 135,000 CFA or USD240 during the rainy season. The approximately 10 percent of plots cultivated in the dry season produce and additional 500,000 CFA/ha or USD816 (at a current exchange rate of about 600CFA/USD). We note that both 2020 and 2021 had above-average total rainfall (Appendix E) which may be potential contributors to the higher-than-expected crop revenues.
37	35	*Table II.17. Credit and expenditures*	Kaj Gass, DCO/SEC-AL	I suppose in order to make sense of the Annual fertilizer cost/ha number, the calculation for getting it needs to be made more clear. Only because we are so absorbed in the fertilizer sector can I say that 60,000CFA should easily buy 100kg of fertilizer (that is with recent price spikes considered). 100kg wuold actually be between 50-100% of recommended dosage. Long winded way of saying that the value is either implausibly high (I'm sure it is extrapolated from smaller amounts that are applied to a very limited production of HVA) OR fertilizer is extremely scarce and arbitrage for small amounts makes it this expensive. The latter would be of huge interest to us to know that smallholder farmers pay this high of a price! I should also mention that this type of signal would also be a reason why farmers would be incredibly hesitant to get into the business of HVA. Seems worth exploring to see if we are confident on the data.	In the revised version of the baseline report, we weight values by the size of the plot. As a result, the estimated fertilizer expenditure for the rainy and dry season combined is 45,000, with an average (reported) application per hectare of 120kg in the rainy season (see Table II.15). The small number of farmers who report dry season cultivation report significantly higher fertilizer application. This suggests that farmers cultivating higher value crops (in this case rice) indeed use more fertilizer. With respect to your hypothesis that fertilizer might be a constraint to adoption of higher value agriculture, we think that that is indeed plausible, but cannot confirm this in our data.
38	35	*Table II.17. Credit and expenditures*	Hamissou Samari, DPE/EE-ME	Farmers in the Comparison areas spending less on fertilizers due to higher cost of seeds and irrigation?	That is correct, farmers in the comparison areas are spending less on fertilizer and slightly more on seeds and irrigation relative to BT farmers. However, seeds and irrigation expenses account for only around 10% of total expenses. We can't draw the conclusion that the lower fertilizer expenditures are due to higher seed and irrigation expenses. There could be other factors at play such as fertilizer access, fertilizer prices, training/knowledge on fertilizer usage, etc.
39	36	*Table II.17. Credit and expenditures*	Kaj Gass, DCO/SEC-AL	The pesticide cost/ha does not quite follow fertilizer in that this cost would not cover a hectare and seems more reasonable for what farmers might spend--probably on raticide for stored production.	This cost is not conditional on using herbicides, pesticides, or insecticides, therefore the low cost per hectare reflects the low usage rates. Only about 20% of plots used pesticides, herbicides or insecticides. We have added a table note to clarify this. Given the wording of the question from which this information is derived, which asks specifically about the application of herbicides, pesticides, or insecticides on the plot, we do not think this could have been misinterpreted to be asking about raticide for stored production.

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
40	36	*C.7. Agricultural productivity and profitability*	Aaron Szott, DPE/EE-EA/PSC	Would it be useful to consider plots other than the pure-stand ones? I see from the table that economically substantial differences between treatment and comparison plots are not statistically significant, and I wonder how useful any of these data will be when we consider the final (considerably smaller) footprint.	In the revised report we present yields from both intercropped and non-intercropped plots as our main results. Pure-stand and intercropped yields are presented side-by-side in Appendix Table D.1. and D.2. We will conduct power calculations to document necessary beneficiary sample sizes to detect reasonable yield improvements.
41	37	*C.7. Agricultural productivity and profitability*	Kaj Gass, DCO/SEC-AL	On rice income: A bit of a shock to me because rice is typically low yielding and inensive on inputs. Understood that production is low input but it needs to be considered that the CBA also calculates labor as a cost. Between rice, sorghum and millet, it seems odd that rice comes on top--can someone explain this in local terms how we get to this outcome?	Our local consultant confirmed that rice cultivation is more profitable on a per hectare basis in the Basse Terasse (where it is possible to grow it). However, it is challenging to account for household labor. Our analysis only accounts for paid labor, not household labor for which the data is incomplete and unreliable.
42	37	*C.7. Agricultural productivity and profitability*	Sarah Lane, DPE/EE-ME	Clarifying question if the word should be "planting" instead of "planning"	We have corrected the text.
43	37	*Table II.18. Crop yield, income, and income per hectare*	Kaj Gass, DCO/SEC-AL	Back to rice and fertilizer. Are these farmers actually paying for fertilizer and seed? I have a hard time believing these yields either. Sorghum also seems far too high. The others are high but within reason. We will send you Cowater's diagnostic of the area.	Your comment raises several issues: 1) the issue of payment for fertilizer and seeds. Almost no one is paying for organic fertilizer (less than 1% of those using organic fertilizer paid for it). Among those using chemical fertilizer, almost all of them reported paying for the fertilizer. As we report in Table II.16, the majority of farmers do not purchase seeds. In both seasons only around 20% of plots used purchased seeds. Regarding the yields, in the revised data analysis we implement additional top coding for cases where yields are outside potential yields as defined by INRAN for the project area.
44	38	*C.8. Household income, sales, and profits*	Kaj Gass, DCO/SEC-AL	More a curiosity than a real concern on imputing sales price since it should all average out but are prices weighted depending on the season? Rice in rainy season estimated to be 2.2MT/ha. The INRAN max for rice is 4.5 for reference.	We apply season-specific imputed prices derived from using data on total sales and quantity sold.
45	38	*C.8. Household income, sales, and profits*	Kaj Gass, DCO/SEC-AL	On non ag income: Perhaps something that should be included in future surveys but do we ever ask about time spent between the two types of activities? Again, labor plays a relatively important role in the ERR calculation.	Unfortunately, we do not ask for time spent on non-agricultural activities. That is a possible addition to the endline survey but is relatively time-intensive to collect.

	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
46	38	*C.8. Household income, sales, and profits*	Aaron Szott, DPE/EE-EA/PSC	On implementing diff-in-diff: Under what assumptions would control HHs then be a good comparison for treatment HHs, and can you evaluate whether those assumptions hold or not?	The key identifying assumption for the validity of our diff-in-diff strategy is that comparison and treatment groups exhibit parallel trends prior to the start of the intervention. This assumption is testable for only a couple of variables because our outcomes of interest that were collected in the baseline survey were for either only the reference period or the preceding period. Longer series of pre-treatment data for outcomes of interest are therefore not possible for the majority of constructs, and so we rely on comparability in baseline values to validate our design. We show in Appendix C.D our balance statistics which demonstrate a high degree of comparability across numerous variables that ex ante we believe are influential factors through which agricultural outcomes are determined. For two variables where a longer time-series is available -- green chlorophyll vegetation index (GCVI) which can serve as a proxy for yields (Burke and Lobell 2017) and monthly rainfall -- we conducted a visual test for parallel trends which appears under a new subsection in Appendix C titled "Testing for pre-trends." For both GCVI and monthly rainfall, we observe similar pre-treatment performance among both treatment and comparison groups which bolsters the claim that a difference-in-difference design is an appropriate research method to apply in this context.
47	40	*C.9. Household food security, poverty, and women's empowerment*	Kaj Gass, DCO/SEC-AL	On food insecurity: Can this be referenced with another study? Is this in the ballpark of what other studies have found?	We have added comparison points of food security levels for both Southern Dosso and Niger-wide.
48	40	*C.9. Household food security, poverty, and women's empowerment*	Aaron Szott, DPE/EE-EA/PSC	Is food insecurity driven to a large extent by the timing of the survey? It seems like hunger in this context would be seasonal.	We agree - we added text to describe how food insecurity is higher in the lean season.
49	42	*C.9. Household food security, poverty, and women's empowerment*	Kaj Gass, DCO/SEC-AL	What about actual opportunities for women to farm? One goal of SK2 is to encourage more womens' groups to gain access to land.	Please see the additional indicators we have added to Table II.20. We will include these in the endline report as well.
50	46	*D.2. Trader farmgate purchases and transportation of crops – trader survey*	Kossi Adjaka, DCO/IEPS-TVS	Construction of RN7 and RN35 started about the same time: RN7 in November 2019 and RN35 in January 2020. RN7 was in extremely poor conditions (large potholes) forcing road users to go on deviations around the main travel ways.	Thank you for the clarification regarding the timeline of the road rehabilitation. We have revised the report to clarify how implementation to-date may have affected traders' responses about road conditions.

Table F.2. MCA Stakeholder comments

#	Page	*Section*	Reviewer Sector	Comments & Questions	Mathematica Response
1	viii	*Acknowledgements*	Harouna Hamidou	Il faut actualiser le nouveau DSE DR. Amadou Issaka	We have updated the report.
2	6	*Research Questions*	Issaka Sani Dodo	What type of data is collected for Reduced harvests	Farmers self-report the quantity of their crop that is lost pre- and post-harvest.
3	9	*Key Results*	Issaka Sani Dodo	Yields and incomes are low compared to which reference? The general level? Medium? Regional? National?	Yields are low compared to potential theoretical yields and both yields and incomes are low compared to the regional averages.
4	12	*Baseline analysis*	Issaka Sani Dodo	How will the effects of other projects in the area be treated? Will the counterfactual and the treatment include effects of other projects.	Our local consultant has been in contact with the agricultural directorate. No irrigation projects have been implemented in the comparison areas or are planned for the comparison areas. At endline, we will again assess whether any irrigation or other agriculture-related projects have taken place in the comparison areas and contextualize our findings accordingly.
5	13	*Sample selection and matching*	Issaka Sani Dodo	Can you list he main variables used in this?	We have updated the tables and made explicit which variables were used in the matching algorithm. In the updated Appendix Table C.1, the columns denote whether a variable was used for estimating the propensity score, an input into the matching algorithm, and/or a variable for which balance tests were conducted.
6	15	*Household survey response rates*	Harouna Hamidou	For the calculation of the sample, the base population is not shown. The sample calculation formulais not entirely clear. The methodology does not describe how respondents are selected	The targeted respondents in the treatment area come from SONED who constituted a list of potential participants in the small-scale irrigation project that meet certain criteria. We selected all of these potential participants for inclusion in the survey. Once the actual beneficiaries are known, we will only consider them as treatment respondents in the final evaluation.
7	16	*Variables utilisées dans l'estimation du score de propension, par niveau d'analyse*	Issaka Sani Dodo	For the qualitative variables selected, can a column be inserted that gives the modalities of the variables by specifying the reference modality for log regression)	We have updated the tables to indicate which variables are binary by specify "(0/1)" next to the variable description. No non-binary categorical variables are included in the analysis.
8	16	*Variables utilisées dans l'estimation du score de propension, par niveau d'analyse*	Issaka Sani Dodo	How is the poverty line set and the poverty score calculated?	Please see footnote #10 and the Simple Poverty Scorecard Tool for Niger (available here: https://www.simplepovertyscorecard.com/NER_2014_ENG.pdf) for information on how the poverty score is calculated. The poverty line is set by the government of Niger. The conclusions in the report are based on the poverty line from 2011.

#	Page	*Section*	Reviewer Sector	Comments & Questions	Mathematica Response
9	27	*Data collection*	Issaka Sani Dodo	How was the sample selected for the household data coll? Was there a sampling frame?	For the treatment households in the Basse Terrasse, our sample included all households with plots scheduled to receive SSI under the project based on SONED data. For the comparison household, please see section II.A of the report where we explain the matching procedures used to determine the sample of comparison households for the household survey.
10	29	*Trader market survey*	Issaka Sani Dodo	Why was the no. 36 chosen for traders? Will this be representative of all traders from the point of view of variables of interest for market research?	We selected 36 traders based on interviewing 6 from each of the 6 markets we visited. We specifically surveyed large-volume traders who may be the most effected by the roads and account for the majority of trade in the region. They are not representative of all traders in the region but should be relatively representative of large-volume traders in the region. Given that the evaluation of the Roads for Market Access activity is a pre-post evaluation rather than an impact evaluation (since the comparison group households are also affected by the activity) we did not need to ensure a sufficiently large sample to detect statistically significant differences but rather just needed to speak with enough traders to capture variety in experiences and practices. 36 traders is a large enough sample to accomplish this.
11	39	*Credit & Expenditures*	Issaka Sani Dodo	Please specify the main sources of funds obtained for the households.	The main source of credit is family and friends. 82% of treatment households and 76% of comparison households who took out a loan in the past year borrow from family or friends. The next most common sources are informal lenders (17% of treatment households and 9% of comparison households) and NGOs (7% of treatment households and 17% of comparison households).
12	39	*Credit & Expenditures*	Issaka Sani Dodo	Disaggregate the 78000 CFA value by gender of the head of household or respondent	We are unable to disaggregate by gender due to the small number of female-headed households and female plot decision makers (respondents). In the BT, only 6% of households have a female head and only 6% of households have at least one plot with a female decision maker.
13	41	*Agriculture productivity*	Issaka Sani Dodo	Redundant sentence	We have corrected this.
14	43	*Household income, sales and profits*	Issaka Sani Dodo	I find the reference of the data quite old compared to the period of the study (more than 5 years). A lot of living condition data may have varied. How will you ensure the stability of the parameters? Can we look at a poverty analysis based on the survey data (with the economic variables collected)?	We acknowledge that this poverty line feels outdated, but the international best practice is to convert this poverty line into today's values (or on the other hand convert today's values back to that poverty line). The WFP also displays figures using this 2011 National Poverty Line (as of February 2021): https://www.wfp.org/countries/niger . The alternative would be to establish a new poverty line, but that task is outside the scope of this data collection and evaluation.
15	45	*Household food security, poverty and women's empowerment*	Issaka Sani Dodo	What definition of food insecurity is applied for the calculation? Add this to the paragraph	We have added this definition to the report as a footnote.

#	Page	Section	Reviewer Sector	Comments & Questions	Mathematica Response
16	47	*Household food security, poverty, and women's empowerment*	Issaka Sani Dodo	What weight is assigned to each of the 4 IAFA dimensions in order to obtain the overall index?	They are weighted equally. Each receives a 1/4 weight.
17	51	*Trader purchases and transportaion*	Issaka Sani Dodo	The point that traders don't sell agricultural inputs directly into villages should be included as a recommendation esp for the fertilizer reform	It is beyond the scope of the evaluation to make recommendations, but we agree that it is an important element of the program logic and causal chain for farmers to have access to high quality, consistently available fertilizer, and the lack of direct sales of fertilizer in villages could hinder this.
18	56	*Implications for the evaluation of the SSI*	Issaka Sani Dodo	How will the change in the forecast area to be developed with SK2.1 be taken into account. Will the parameters of this study be updated to have a revised Baseline?	There are several options that can be chosen once the scope of the small-scale irrigation project are known, including revising the baseline or presenting the baseline values together with the endline values during the final evaluation report. At this point in time, no decision has been taken.
19	56	*Implications for the evaluation of the SSI*	Harouna Hamidou	In view of the delays in starting work, what recommendations do you make and what action should be taken in order to implement this action within the optimal time frame?	Unfortunately, this is beyond the scope of the evaluation. As the independent evaluator, our role is to assess how the project was implemented and the outcomes and impact of implementation. We are unable to make recommendations about how implementation should proceed.
20	59	*Dissemination plan*	Issaka Sani Dodo	Given the volume of the report, it will be useful to produce a summary version of this report with about ten pages focusing on the main results. This will make it easier for a wider audience to understand the study.	In coordination with MCC, we have decided to limit the changes to the baseline report for efficiency purposes.
21	59	*Dissemination plan*	Harouna Hamidou	For dissemination, please add beyond MCC, MCA implementing entity through the management service, the DSE, and stakeholders such as the Ministry of Agriculture, the local depts of Agriculture of Dosso and Gaya and Tanda	We will include these stakeholders for dissemination of future evaluation materials. Due to the change in scope of the project, the current baseline report might not be very useful.
22	59	*Dissemination plan*	Harouna Hamidou	It is also important to have a summary table of all project indicators and their values to date	The focus of this baseline report is to present information on beneficiaries of the project as opposed to providing an update on the project progression.
23	E.2	*Table E.1. Comparison of 2020/2021 precipitation with long-term averages*	Harouna Hamidou	Anomaly or different rainfall patterns. If not, how does this constitute an anomaly?	We have added text to Appendix E to clarify this. For a given region, the technical term anomaly is defined as the difference between annual realized rainfall and the long-term mean annual rainfall.