WFP EVALUATION

Impact Evaluation for Resilience Learning in the Sahel

Mali Inception Report

WFP World Food Programme

SAVING LIVES CHANGING LIVES

i2i DIME TRANSFORM DEVELOPMENT

OEV/2019/036

October 2022

Contents

Key Personnel for the Evaluation

OFFICE OF EVALUATION (WORLD FOOD PROGRAMME)

Director of Evaluation, Andrea Cook

Evaluation Officer, Jonas Heirman

Evaluation Officer, Hanna Paulose

Evaluation Analyst, Theresa Schneider

Evaluation Analyst, Ola Eltoukhi

DEVELOPMENT IMPACT EVALUATION TEAM (WORLD BANK)

Principal Investigator, Patrick Premand (Lead Researcher, contact: ppremand@worldbank.org)

Co-Principal Investigator, Guigonan Serge Adjognon

Resilience Window Manager, Marcus Holmlund

Research Officer, Chloë Fernandez

Field Coordinator, Dimanche Allo

Research Assistant, Mariana Garcia Martinez

Research Assistant, Kane Borders

WFP MALI COUNTRY OFFICE

Head of Programme, Ibrahima Diallo

RAM Officer, Emmanuel Hakizimfura

1. Introduction

1. In 2020, 155 million people worldwide faced "crisis or worse" levels of food insecurity. Close to 115 million of them lived in countries affected by conflict or weather extremes.[1] Extended food crises are predicted in 2022, which would necessitate continued large-scale humanitarian assistance. Since the beginning of armed rebellion and coup d'état in northern Mali in 2012, which later evolved into an extremist insurgency,[2] 3.6 million people on average (18 percent of the population) experience food insecurity. Conflict, forced population displacements, and climate change exacerbate food insecurity.[3] However, evidence is lacking on how development outcomes are affected by these shocks, and how WFP's programmes support populations to effectively respond to these shocks.

2. The World Food Programme's (WFP) Office of Evaluation, Asset-Creation, Livelihoods and Resilience Unit, and the Climate and Disaster Risk Reduction Unit partnered with the World Bank's Development Impact Evaluation (DIME) department to create the Climate and Resilience Impact Evaluation (IE) Window (see Annex 1 for a summary of the window). WFP's Impact Evaluation Strategy (2019–2026) focuses on delivering impact evaluations that contribute to global evidence and organizational learning. Impact evaluation windows help to achieve this strategy by focusing portfolios of impact evaluations on priority evidence needs identified through literature reviews and extensive consultations.

3. The climate and resilience window aims to understand how the WFP's programmes contribute to the resilience of populations supported. The first round of impact evaluations selected for this window aims to estimate the impacts of integrated packages of resilience activities on households' capacities to absorb shocks (absorptive capacity), adapt to increasing environmental or economic stressors (adaptive capacity), and improve well-being in the long term (transformative capacity).

4. The Mali impact evaluation aims to estimate the impacts of the WFP's Integrated Resilience Programme on absorptive, adaptive, and transformative resilience capacities. The WFP's integrated resilience programme in Mali consists of activities including food assistance for assets (FFA), nutrition support, school feeding, and smallholder agriculture market support (SAMS) activities. The intended direct outcomes of the intervention are increasing both household resilience capacities and food security.

5. This inception report outlines the strategy for assessing the impact of the WFP's resilience programme in Mali on the dynamics of population well-being and resilience. Through this impact evaluation, the WFP and DIME are working together to complement other ongoing efforts and guide future investments and activities related to resilience in the Sahel.

6. This inception report also builds on a pre-analysis plan (PAP) registered with the American Economic Association's registry for randomized controlled trials. The pre-analysis plan includes detailed information on primary outcomes, research design, randomization method, randomization unit, clustering, sample size (total number, number of clusters, and units per treatment arm), and regression specifications. The purpose of the PAP is to outline the set of hypotheses and analyses that will be performed on the data before it is collected.

1 FSIN. 2021. Global Report on Food Crises: Joint Analysis for Better Decisions. [Link]
2 Stockholm International Peace Research Institute. N.d. Mali. [Link]
3 World Food Programme. 2020. Mali Country Strategic Plan (2020–2024). [Link]

2. Evaluation Context

2.1. BACKGROUND AND CONTEXT

7. Mali is a vast, land-locked country in the heart of the Sahel. With social indicators among the lowest in the world, the country ranks 184 out of 189 on the UNDP's 2019 Human Development Index and faces serious challenges in achieving Sustainable Development Goal 2 on zero hunger and improved nutrition. Every year since the 2012 conflict, 3.6 million people (18 percent of the population), on average, experience food insecurity, including 600,000 severely affected people. Food insecurity varies from one region to another, with the north and central regions (Gao, Mopti, and Tombouctou) particularly challenged.[4]

8. Food insecurity in Mali also sees seasonal variations, often peaking before the main cereal harvest for farmers (June–September). The widespread drought occurring in Mali every few years results in the advanced onset of lean seasons.[5] The 2020 lean season pushed 1.3 million people in central and northern Mali into food insecurity, triggering a state of emergency.[6] These variations are further aggravated by conflict and insecurity. The WFP's satellite imagery analysis indicated reduced agricultural activity in 2020, as the conflict expanded to more areas in central Mali.[7] Conflict, and its effect on agricultural activities, already risks year-round food insecurity challenges in the country. Several parts of the country also experience recurrent weather extremes (rains, floods), further adding to the problem. In 2020 alone, heavy rains and floods damaged the livelihoods of more than 11,000 households in the Koulikoro, Menaka, and Segou regions. Figures from 2021 indicate that 907,000 children under five years of age suffer from moderate acute malnutrition, and 309,800 from severe acute malnutrition.[8] Multiple, recurring shocks and stressors both directly threaten food security in the immediate future and affect livelihoods, assets, and basic services, which contribute to sustained food security in the long run.

9. As explained above, the climate and resilience window aims to examine the interaction between shocks and stressors, people's livelihoods, and food security. Given the seasonal nature of Mali's food insecurity, it is particularly important to monitor food insecurity over time to identify vulnerable populations. It is equally critical to understand how and when the WFP's resilience programmes can support people in maintaining food security in the face of shocks and stressors, including unpredictable weather events and conflicts. Past studies demonstrate that multifaceted economic interventions in the region (Niger) can contribute to addressing people's vulnerability to shocks and help them maintain food security.[9] However, past studies do not closely examine the dynamic nature of food security in the region or seasonality and other factors.

10. Given this context and policy environment, DIME and the WFP, with support from Germany's Federal Ministry for Economic Cooperation and Development (BMZ), are collaborating to generate evidence on how multiple interventions can be combined or sequenced to boost the resilience of poor and vulnerable households in Mali. This evidence agenda contributes to the implementation of the WFP's integrated resilience programme in the Sahel and is also of interest for the World Bank programme.

4 World Food Programme Mali. 2021. Country Brief. [Link]
5 World Food Programme. 2020. Mali Country Strategic Plan (2020–2024). [Link]
6 Cadre Harmonisé (CH). March 2020. This figure is higher than the projected figure by the November CH (1.1 million), and represents a 142 percent increase compared with the 2019 lean season, when 553,770 people were estimated to be food-insecure.
7 The imaging analysis was conducted in November 2020. The most affected areas are those where intercommunal tensions led to increased violence in 2020.
8 World Food Programme. 2021. Mali Emergency Dashboard, December 2021. [Link]
9 Bossuroy, T., Goldstein, M., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C., Udry, C., Vaillant, J. & Wright, K. 2021. Pathways Out of Extreme Poverty: Tackling Psychosocial and Capital Constraints with a Multi-Faceted Social Protection Program in Niger. Policy Research Working Paper 9562. World Bank, Washington, DC.

11. This impact evaluation in Mali aims to identify the impact of the WFP's programme on resilience, as measured by people's capacity to maintain food security while experiencing shocks and stressors. The evaluation will also use bimonthly surveys to track food security and coping strategies of households supported by the programme and those who are not part of the programme. This approach will help provide an understanding of when food security peaks in Mali, and which households are vulnerable to becoming food-insecure at different phases. Surveys will be complemented by qualitative data and analysis to understand how the programme is implemented and identify opportunities for future improvements, how the support provided is perceived by beneficiaries, as well as to generate insights about the patterns observed in the quantitative data. The impact evaluation aims to inform planning for the subsequent programme cycle as well as the WFP's next country strategic plan for Mali.

2.2. PROGRAMME DESCRIPTION

12. Given the dynamic food security situation, the WFP's Mali Country Strategic Plan 2020–2024 (CSP) has adopted a two-pronged approach to address both the short-term and long-term needs of beneficiaries in its programming.[10] The CSP stresses the importance of maintaining the WFP's capacity to respond to emergency needs while also increasing its focus on longer-term outcomes (such as diversified livelihoods, agricultural productivity, peace, and social cohesion), to improve households' and communities' capacities to respond to shocks and stressors, and to therefore enhance resilience.

13. Mali's resilience programme is aligned with the CSP objectives and includes interventions that aim to promote the capacities of households and communities to absorb shocks, adapt to risks, transform livelihoods, and, more broadly, in the living environment, to exit poverty. It includes (i) food assistance for assets (FFA), (ii) nutrition/health, (iii) value chain and smallholder agriculture market support (SAMS), and (iv) school feeding. These are implemented in parallel with social safety net interventions aiming to address the immediate needs of the most vulnerable people within the targeted communities (e.g., lean season support and COVID-19 cash transfers). The planning and prioritization of these interventions is supported and guided by the community-based participatory planning (CBPP) process.

14. By introducing a combination of layered and sequenced activities targeting the most vulnerable, the WFP resilience programme in Mali aims to promote the resilience capacities of individuals, households, and communities:

    - At the community level, activities aiming to promote resilience focus on environmental rehabilitation and food systems development. These include the FFA and SAMS components.
    - At the individual and household levels, activities promoting resilience include investments in human capital by providing an integrated package of school feeding and nutrition services, and lean season support for vulnerable households. These activities complement government efforts to strengthen the resilience of vulnerable communities.

15. More specifically, the WFP resilience programme in Mali includes the following:

    - *The food assistance for assets (FFA) component* aims to meet the immediate food needs of households while restoring degraded landscapes, improving water harvesting, reducing the risk of environmental disasters, creating productive assets to secure ecosystem services, increasing productivity and yield, supporting economic development, and strengthening social ties between community members and villages. New assets are selected via a community-wide participatory process and are built through asset-creation activities tied to cash or in-kind support. Beneficiary households for FFA activities are targeted through a community-based participatory planning (CBPP) process. Households are categorized into four socioeconomic groups: very poor, poor, average, and well-off. Households in the very poor and poor categories are eligible to participate in FFA activities. They are paid approximately USD 35 per month (CFA 19,500) for approximately

10 World Food Programme. 2020. Mali Country Strategic Plan (2020–2024). [Link]

three months to work on the assets, typically between March and June before the rainy season, but the period can be extended as needed.

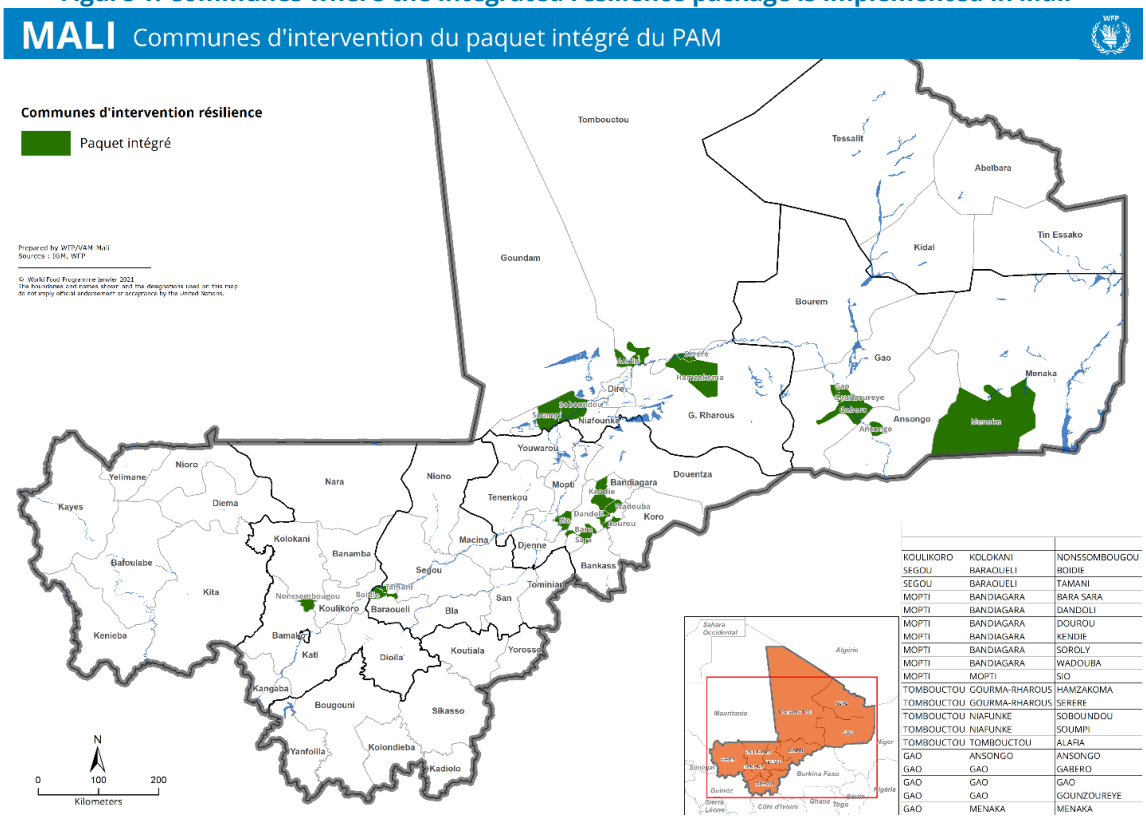
- *The nutrition/health component* aims to prevent malnutrition through a combination of nutrition-specific and sensitive interventions (i.e., a lifecycle approach) that seeks to improve the availability, access, and use of nutrient-dense food and the adoption of key nutrition and healthy family practices. To do so, the WFP provides food supplements to households with children under two years old, and to pregnant or lactating women. This is complemented with intensive capacity-building activities at the community level on how to develop local initiatives that can improve community feeding practices. The WFP also provides incentives to pregnant or lactating women to boost attendance at nutrition sensitization sessions, and to antenatal and postnatal care in health facilities. This component is implemented in synergy with the prevention and treatment of wasting through health facilities and the general food assistance platform. Synergy is also developed with the FFA and SAMS components to nutritionally optimize food production across different assets and support the development of a nutrition-sensitive value chain that increases the availability of locally produced nutritious foods. All of these activities are backed by large sensitization campaigns to promote good practices related to feeding, nutrition, health, and hygiene for infants and young children.
- *The smallholder agriculture market support* (SAMS) *component* aims to complement FFA activities by supporting smallholders in managing assets and increasing their incomes through related activities, such as improving storage (reducing post-harvest losses), agri-food processing, capacity-building, and market access facilitation.
- *The school feeding component* aims to increase access to education and school retention rates by providing nutritious school meals and support to adolescent girls. Complementary activities, such as nutrition education, the creation of school gardens, or trainings on canteen management, leverage schools as a platform to contribute to food diversification and deliver messages on hygiene, family practices, and environmental stewardship. School feeding is a government-led programme, which the WFP currently implements in 650 schools.
- *The lean season support component* is unconditional cash/food assistance provided to extreme poor households to offset the peak hunger and malnutrition period. It amounts to approximately USD 52 (CFA 30,400) for two or three months between June and August. Lean season support helps households to meet their gaps in food needs during the lean season. The support may only be provided in the first year of the resilience programme.

16. Among these components, FFA is an entry point for targeting and implementing the other activities to strengthen households' and community resilience capacities. The FFA programme guidance manual describes the core functions of FFA,[11] which include, simultaneously, the direct provision of food or cash-based transfers to meet the consumption needs of the most vulnerable (i.e., short-term access to food), as well as the construction/development of household and community assets that reduce the risk of disaster, strengthen livelihoods, and build resilience over time.[12] The strong emphasis on asset-creation, and the impacts of assets on people and communities, distinguishes FFA from other forms of delivering food assistance, such as food for work or cash for work programmes.

17. As of 2021, the integrated resilience programme covered 21 communes across five regions in Mali: Gao, Koulikoro, Menaka, Mopti, and Tombouctou.[13] Figure 1 below highlights the areas where the resilience programme is implemented.

11 See Carucci et al., 2016 in the references for complete publication information on the manual.
12 Carucci, V, Ronchini, S., Gentile, J. Policastro, R., Crahay, P., Primozic, M., Vaquier, D., Gonzalez, R., Gardesten, J. & Mariangeloni, S. 2016. Food Assistance for Assets (FFA) for Zero Hunger and Resilient Livelihoods: A Program Guidance Manual.
13 World Food Programme. 2021. Annual Country report. [Link]

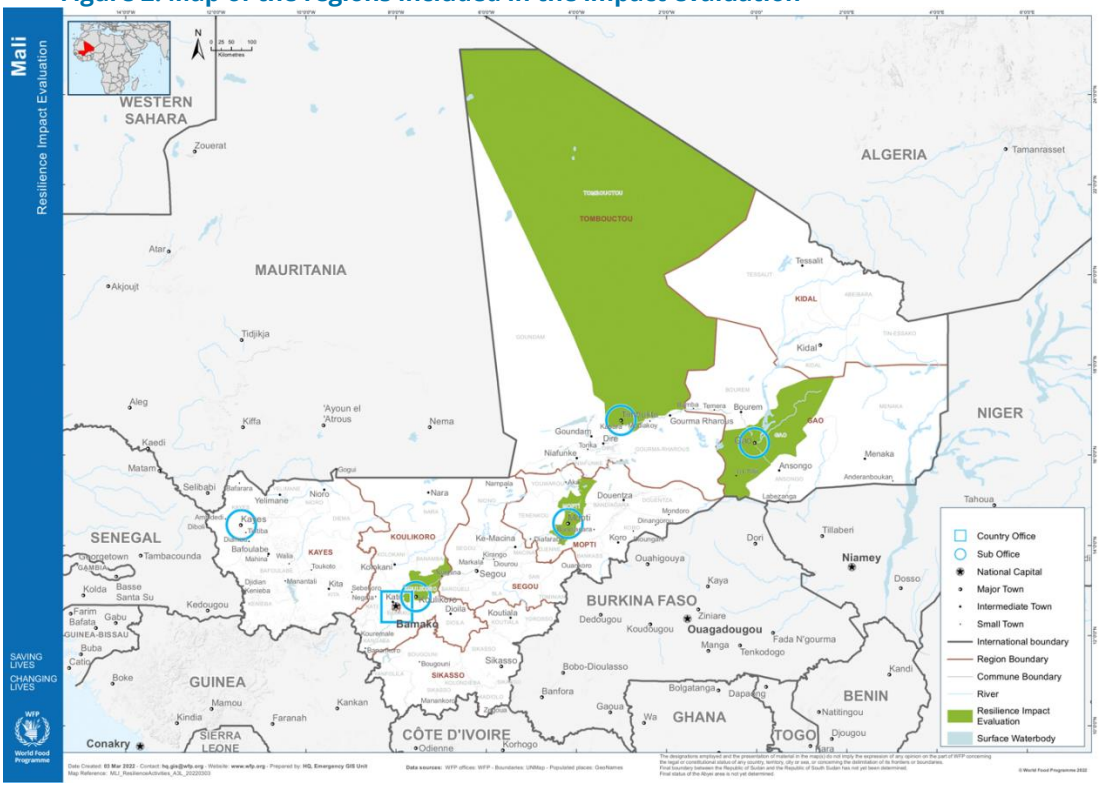
Figure 1: Communes where the integrated resilience package is implemented in Mali



18. The impact evaluation will focus on new sites (Figure 2) added to the resilience programme in 2021 – more precisely, 91 villages across 14 communes in the regions of Gao (communes of Asongo, Gabero, Gao, and Gounzoureye), Koulikoro (commune of Nonssombougou), Mopti (communes of Dandoli, Dourou, Kendie, Soroly, and Wadouba), and Tombouctou (communes of Alafia, Serere, Soboundou, and Soumpi).[14]

14 Menaka is not included in the impact evaluation because no new villages are becoming enrolled in the resilience programme this year.

Figure 2: Map of the regions included in the impact evaluation



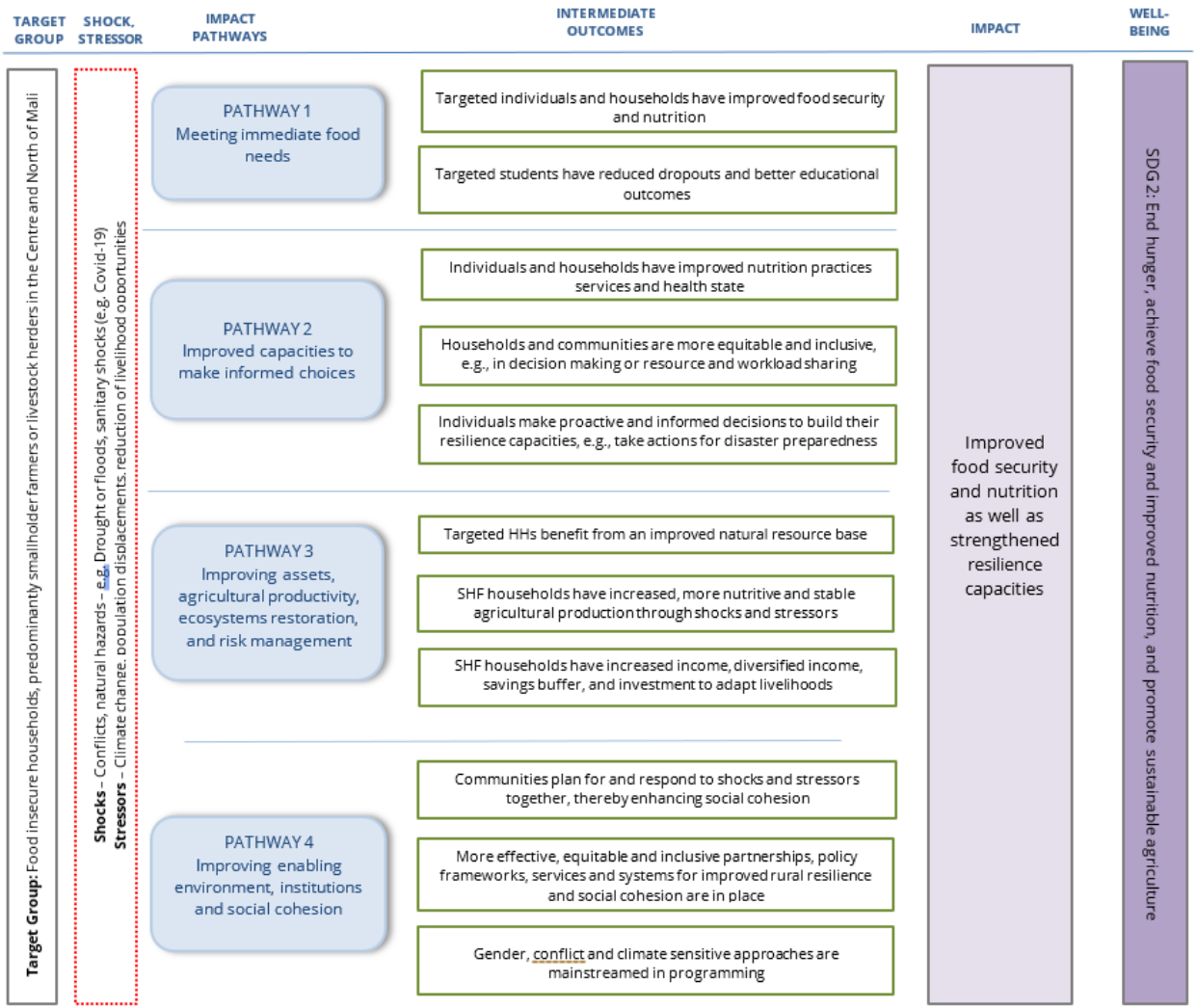
2.3. THEORY OF CHANGE AND HYPOTHESES

19. The theory of change of the resilience programme assumes that supporting communities through multiple activities focusing on various outcomes will (i) support people to ensure their short-term well-being, and (ii) enhance people's capacity to maintain and improve well-being while facing shocks and stressors. Figure 3 below summarizes the key intermediate outcomes and the impact the programme focuses on. This is a simplified version of the full programme theory of change.

20. The resilience impact evaluation in Mali aims to test the following hypotheses.

- Hypothesis 1: In the short term, the WFP's resilience programme will support people in maintaining their food security by meeting a household's immediate food needs that may arise during a shock or stressor. The effect of activities focused on meeting immediate food needs would be reflected mainly in:
    - Household-level food consumption
    - Coping strategies of households.
- Hypothesis 2: In the medium term, the WFP resilience programme will support households by improving capacities associated with maintaining and/or improving food security while experiencing multiple and/or recurring shocks and stressors. These capacities include:
    - Livelihood activities
    - Household assets
    - Financial outcomes (e.g., income, savings, and expenses)
    - Variations in food consumption over time.

Figure 3: Theory of change of the resilience programme in Mali[15]



15 The theory of change of the resilience programme in Mali is developed by the country office.

3. Evaluation Approach and Questions

21. Impact evaluations measure changes in development outcomes of interest for a target population that can be attributed to a specific programme or policy through a credible counterfactual. The WFP's ability to establish a credible counterfactual for programme interventions depends on logistical and financial constraints. Impact evaluations are therefore restricted to focusing on a set of questions that can be answered during a programme cycle using credible counterfactuals.

22. Regional discussions, in-country consultations, and subsequent conversations with the programme and Monitoring and Evaluation (M&E) teams have led to the adoption of a gradual approach to building an impact evaluation learning agenda that is suitable for Mali's context and the WFP's programme implementation plans. A number of primary and secondary questions were developed addressing different aspects of the programme. The questions and the agreed methodological approaches are summarized in Table 1 below.

23. The main focus will be on documenting impacts on food security and related changes in well-being associated with households' resilience capacities. These indicators include:

- Consumption and food security
- Nutritional status
- Financial outcomes and assets
- Assets and livelihoods.

24. The evaluation will also directly assess *how* the resilience programme affects households' ability to mitigate the effects of shocks on their food security and welfare. This is mainly achieved by monitoring the following main outcomes using bimonthly surveys:

- Food consumption
- Coping strategies
- Shocks and stressors experienced.

25. The impact evaluation uses a mixed-methods evaluation design, using quantitative and qualitative data. The design includes three complementary components:

- *Clustered randomized controlled trial (RCT) design:* aims to answer the primary evaluation question outlined in Table 1. The cluster RCT design is complemented by heterogeneity analysis and qualitative analysis. Within this design, 91 villages in four regions of Mali are randomized into treatment and control groups. Participants in the treatment group are expected to receive the integrated package of interventions, including school feeding, food assistance for assets, nutrition/health, and SAMS activities. The RCT constructs credible counterfactuals to identify the impact of the programme on resilience outcomes.
- *Heterogeneity analysis*: Not all the resilience programme activities are suitable for randomised assessment. The data collected within the framework of the RCT will also be used to answer two secondary questions to be examined through this impact evaluation using heterogeneity analysis and qualitative analysis. Additional details of the evaluation methodology can be found in Table 1 below as well as in the following sections.
- *Qualitative analysis*: Qualitative data will be used to understand how the programme is implemented, and how the support provided through the programme is perceived by the

beneficiaries. Additionally, the data will be used to generate insights about the patterns observed in the quantitative data. The qualitative data will be particularly useful for understanding the aspects of the programme that are well implemented, and to identify opportunities for further improvement. Qualitative data will be collected through two main sources: focus groups of willing beneficiaries, and interviews with select village leaders. The topics for the interviews and focus group discussions will be informed by the quantitative survey data and may include: the overall awareness of the programme; level of participation in the programme; perceived changes on key outcomes of the programme; and feedback on programme implementation. The focus groups and interviews will be planned after sufficient time has passed since the start of the implementation (i.e., approximately one year) to collect informed feedback on the programme.

26. All quantitative analysis in the impact evaluation will use data collected through baseline surveys, endline surveys, and high-frequency surveys (conducted every two months) to answer the evaluation questions. The impact evaluation involves baseline and endline data collection, allowing the team to estimate short-term and medium-term impacts (timeline presented in Section 8). Baseline data collection takes place before implementation of the integrated resilience package began. The endline data collection will occur after at least two years of implementation of activities in the treatment groups.

27. The high-frequency surveys will be implemented every two months, starting after the baseline survey, and continuing for a period of at least one year. This high-frequency data collection exercise will focus on collecting data related to food security, shocks experienced, and coping strategies. These surveys, combined with endline data, enable the evaluation team to observe changes in food security over shorter periods of time more frequently, providing a more nuanced picture of fluctuations in food security over various shocks and agricultural seasons. The high-frequency surveys will allow us to examine the characteristics of households whose food security is less stable and understand what periods in the year households will require support the most to maintain or improve their food security.

28. By virtue of the evaluation design, the data collected will be disaggregated by the gender of the respondent. Importantly, the evaluation does not consider a 'household' as one unit, but rather considers individuals within the households separately. As such, some components of the questionnaire are directed at female respondents of reproductive age (such as minimum dietary diversity) or at children aged 6 to 23 months (such as vaccination information, and minimum acceptable diet), among other age categories. Additionally, the evaluation will also disaggregate households based on the gender of the households, to understand how gender influences households' access to services, income-generating opportunities, and their well-being.

Table 1: Overview of evaluation questions and methods

Questions answered through a randomized controlled trial (RCT)	**Details of evaluation methodology**
What is the impact of the integrated WFP resilience package (FFA, SAMS, nutrition/health, and school feeding) on the resilience of recipient communities and households?	This question is answered by comparing households in villages that participate in the integrated resilience programme with households in villages that do not participate in the programme.
Questions answered through heterogeneity analysis	**Details of evaluation methodology**
How does the effectiveness of the WFP's resilience package vary depending on households' initial poverty and food security levels?	This question is answered by disaggregating the data collected based on socioeconomic and demographic characteristics, such as gender. This will allow us to assess social determinants of resilience outcomes, in particular when looking at cross-country patterns (e.g., decision-making of women, shock exposure of different social groups, etc).
Does the resilience package have any observable environmental impacts on site-level outcomes, such as vegetation indices, around the sites where FFA activities recuperate land? For example, do impacts differ for sites involving a water project or pond compared with those that involve land recuperation only?	This question is answered by combining household-level survey data collected through baseline, endline and high-frequency surveys, with environmental indicators, such as the normalized difference vegetation index (NDVI), that are publicly available. Adding environmental data into the analysis provides a more nuanced picture about the shocks affecting the impact evaluation areas during the time of the evaluation. This may also inform the food security patterns being observed in the survey data.

4. Evaluation Methodology

29. The evaluation is taking place between 2021 and 2023. In 2021, the impact evaluation design was set up and started collecting data in 91 villages in the four regions of Mali.

30. As outlined in Section 3, the impact evaluation utilizes a cluster RCT design that is complemented by heterogeneity analysis and qualitative analysis.

4.1. CLUSTER RANDOMISED CONTROLLED TRIAL (RCT) DESIGN

31. In a cluster RCT design, communities are randomly assigned to one of the comparison groups (i.e., treatment and control groups). The approach is depicted in Figure 4 below. The RCT analysis will compare treatment and control groups to estimate the credible and unbiased treatment effects of the resilience package.

32. In Mali, villages are important entry points for programme targeting and implementation. Many activities critical to the programme are implemented at the village level, as opposed to household- or individual-level interventions. Therefore, to identify the causal impact of the resilience programme on different comparison groups, the impact evaluation utilizes village-level randomization. This approach is depicted in Figure 4 below.

33. The cluster RCT design is meant to study the impact of the integrated resilience package on intended beneficiary communities and households. The clustered RCT will make use of a pool of villages that meet the criteria for programme participation and are therefore equally eligible to receive support through the programme. This ensures that villages in different comparison groups have similar core characteristics. After the Mali country office identifies eligible villages for the resilience programme using the WFP's targeting criteria, eligible villages are randomly assigned to the comparison groups. Since FFA activities are entry points for implementing the resilience programme in a village, these activities are used as the reference for randomization in this RCT.

Treatment and control groups

34. The resilience programme in Mali covered 59 villages where FFA activities had been active since 2018 – with a break in 2019 – before the impact evaluation was designed. In 2021, the programme planned an expansion to 45 additional villages in the vicinity of the 59 existing ones. We leverage this programme expansion for the impact evaluation strategy. A set of 174 villages (clusters) eligible for the programme expansion across four regions are randomly assigned to the following groups:

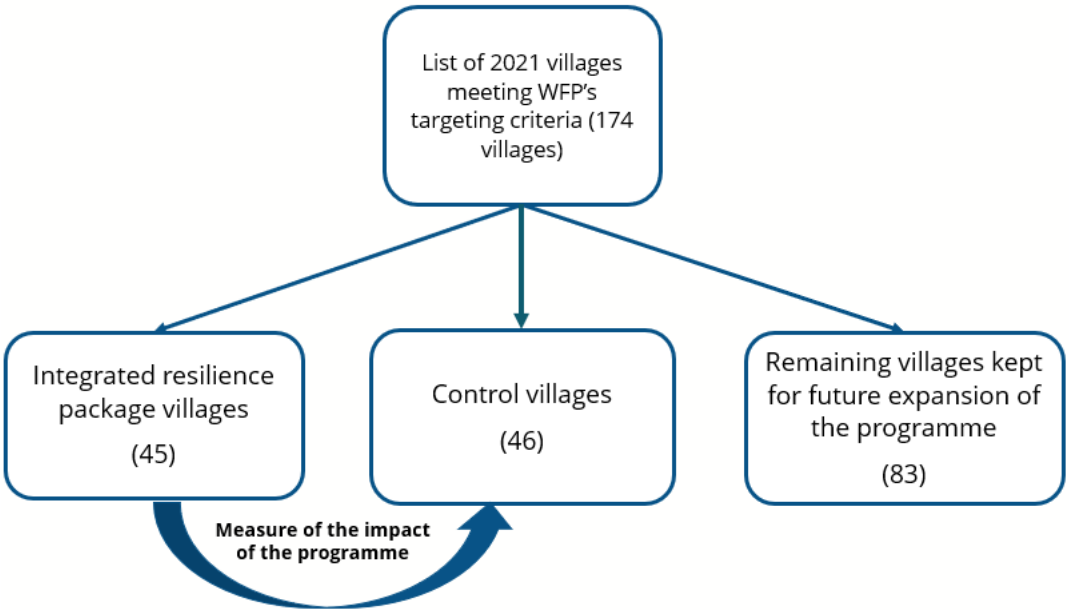
- Group A: treatment group – villages that will receive the integrated resilience package [45 villages]
- Group B: control group – villages that will not receive the integrated resilience package during the impact evaluation period [46 villages]
- Group C: waitlist group – villages that are outside of the impact evaluation sample and will not be surveyed for the impact evaluation [83 villages].

35. The randomized assignment of the treatment (i.e., the integrated resilience package) to the 45 villages leverages the resource constraints of the programme for learning. A lack of resources in 2021 prevented the programme from enrolling all 174 eligible villages in the resilience programme. After randomly assigning the 45 villages, the remaining villages are split into the control group and the waitlist group, anticipating that future financial availability may allow more villages to receive the programme. The waitlist group of 83 villages will be prioritized for programme participation if new funding becomes available. The control group of 46 villages is the main control group for the impact evaluation and will be considered for programme participation only after the waitlist group has been incorporated. This approach is depicted in Figure 4 below. The evaluation will not impose any artificial constraints on potential beneficiaries receiving programme benefits. Instead, the randomized assignment is an

objective and unbiased mechanism to decide which of the eligible villages, all meeting the same eligibility criteria for support, should receive the programme first. Respecting a sufficient sample size, the randomization eliminates any systematic differences between the treatment and control group and thus creates a valid counterfactual.

36. The data collected by the impact evaluation will also allow us to explore the heterogeneous effects of the programme on different groups within the population, along various dimensions (e.g., male- versus female-headed households, and food-insecure versus food-secure). We will be able to address these parts of the evaluation question by performing heterogeneity analysis that identifies specific categories within the data.

Figure 4: Resilience package experimental design



Introduction of safety nets programme in the impact evaluation area

37. It is important to note that outcomes measured in villages supported by the WFP resilience programme will also be influenced by any other interventions in the area. WFP Mali is planning a safety net programme to support vulnerable communities impacted by the COVID-19 shock. These safety net interventions will be implemented in the same geographic areas as the integrated resilience programme.

38. Since the safety net interventions are implemented in both treatment and control villages, they present no risk of bias for impact estimates. However, this means that the impact of the resilience programme captured through the evaluation will be above and beyond the effects of these safety net interventions. In general, we would expect other interventions, even if from other organizations, to be happening in the background of an impact evaluation in such a fragile context. Therefore, we will capture these details during data collection and interpret our findings accordingly.

39. The key activities of the COVID-19 safety net programme include:

- Unconditional cash transfers (UCT)

- Eligible households within targeted villages will receive a direct cash transfer of CFA 15,000 (USD 30) monthly for at least six months.[16]
- All villages (in treatment and control groups) in the resilience impact evaluation sample are included.
- Eligibility within targeted households is based on the food consumption score (FCS) as measured by the *Registre Social Unifié* (RSU)[17].

- Nutrition support
  - This includes one-time top-up payments that will complement the UCT and provide assistance to households with children less than 2 years old (USD 70 top-up), or with pregnant or lactating women (USD 90 top-up).
  - Eligibility is limited to households eligible for the safety nets and with children less than 2 years old, or with pregnant or lactating women.

4.2. HETEROGENEITY ANALYSIS

40. Not all resilience outcomes can be captured by only examining differences between treatment and comparison groups. The impact evaluation uses heterogeneity analysis to understand the secondary evaluation questions identified during the design phase. Heterogeneity analysis harnesses the cluster RCT design as a basis for identifying treatment and comparison communities. The approach allows us to compare well-being outcomes (e.g., food security) of household groups with different characteristics or varying levels of vulnerability. The impact evaluation employs heterogeneity analysis to examine the following interactions and outcome areas.

Resilience across varying vulnerability levels

41. Within the impact evaluation sample, the heterogeneity analysis will be used to understand how social or demographic characteristics such as livelihood types, degree of exposure to shocks, gender of the households, etc, will affect resilience outcomes. Trends in food security can be observed within different sub-categories of the sample.

42. This type of analysis is made feasible by the extensive baseline survey, which collects a list of outcomes ranging from households' livelihood sources and their frequency of cultivation in a year, to their access to social safety nets and community support mechanisms. The baseline characteristics, combined with bimonthly high-frequency surveys, allow us to observe food security pathways of different sub-categories within the sample.

Impact on village-level environmental outcomes

43. It is possible that larger infrastructure construction projects (e.g., dam construction, large-scale land restoration, etc) may influence environmental outcomes such as rainfall density and water availability. These environmental outcomes could in turn influence village-level exposure to different shocks or stressors. The impact evaluation team will examine the Water Requirement Satisfaction Index (WRSI) data and Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) data. WRSI provides information about crop performance based on the availability of water at different stages of the crop

16 Households received on average approximately CFA 54,610 (approximately USD 94) of FFA transfers in 2021. Their average yearly consumption at baseline was CFA 508,385 (USD 877). Therefore, these FFA cash transfers represent approximately 10.7 percent of average yearly consumption (CFA 54,610/508,385).

17 The RSU is the government's social protection tool that will be used by the WFP and other partners to select beneficiaries for the social protection programme. The RSU provides a complete listing of households in all participatory and control villages, as well as a classification of households, with the objective of identifying the different socioeconomic strata and wealth distributions within villages.

cycle. CHIRPS is a dataset that includes more than 35 years of rainfall data from select regions across the globe. These datasets will inform the impact evaluation by providing valuable information about crop growth and climatic shocks in the impact evaluation areas. As most geographic areas under the impact evaluation are predominantly agriculture-focused, these datasets will be important to better understand the shocks experienced by the populations at different periods during the impact evaluation.

4.3. QUALITATIVE ANALYSIS

44. In addition to the quantitative analysis, the impact evaluation will examine important process-related questions, such as:

- How did the process of programme implementation contribute to, or hinder, the achievement of measured outcomes? To what extent were programme interventions implemented as planned?
- How did intended beneficiaries supported by the programme experience participation in selected interventions?

45. In 2020 and 2021, a barrier to using additional qualitative data collection methods, such as focus group discussions, was the institutional review board (by Solutions IRB, used for this impact evaluation) requirement to limit "research activities" that increase the chance of group-based spread of COVID-19.

46. If conditions allow, there are two planned qualitative data collection activities under the impact evaluation, prior to the endline survey: interviews and focus group discussions. The impact evaluation uses semi-structured interviews with implementing partners to capture information about the process of programme implementation and the experience of programme participants. The structure of focus group discussions will be informed by the quantitative survey data and may include: the overall awareness of the programme; level of participation in the programme; perceived changes on key outcomes of the programme; and feedback on programme implementation. The focus groups and interviews will be planned after sufficient time has passed since the start of the implementation (i.e., approximately one year) to collect informed feedback on the programme.

5. Data Collection and Measurement

47. This section provides an overview of data collected by the impact evaluation, including the sample sizes and outcomes measured by household surveys.

5.1. DATA COLLECTION FOR THE RCTS

48. This impact evaluation relies on detailed baseline and endline surveys, as well as shorter high-frequency surveys every two months. The baseline and endline surveys will enable us to measure outcomes before and after the intervention, and to examine whether the well-being of beneficiaries improved during the programme period. The high-frequency surveys will enable us to capture variations in outcomes, such as food security across seasons, and as households encounter shocks or stressors. The high-frequency surveys will generate additional information about which types of households need assistance and when. We present our sampling strategy for both exercises below.

49. A sufficient sample size in an impact evaluation has two important roles: (i) ensuring that households in both the treatment and the control groups are, on average, similar across the main characteristics that would influence outcomes; and (ii) ensuring that the sample households in the evaluation are, on average, representative of the population they were drawn from.

50. Insufficient sample sizes pose the risk of falsely detecting, or not detecting, programme impacts. As it is logistically and financially impossible to survey the entire population, the power calculations applied for this evaluation follow standard research norms to estimate the minimum sample size needed to minimize the risk of biased estimates.

51. We conduct the power calculations using food security outcomes, as we believe that these are the outcomes most likely to change because of the programmes, and for which we have comparable data that can be used for more precise calculations. This applies to both livelihoods and education designs.

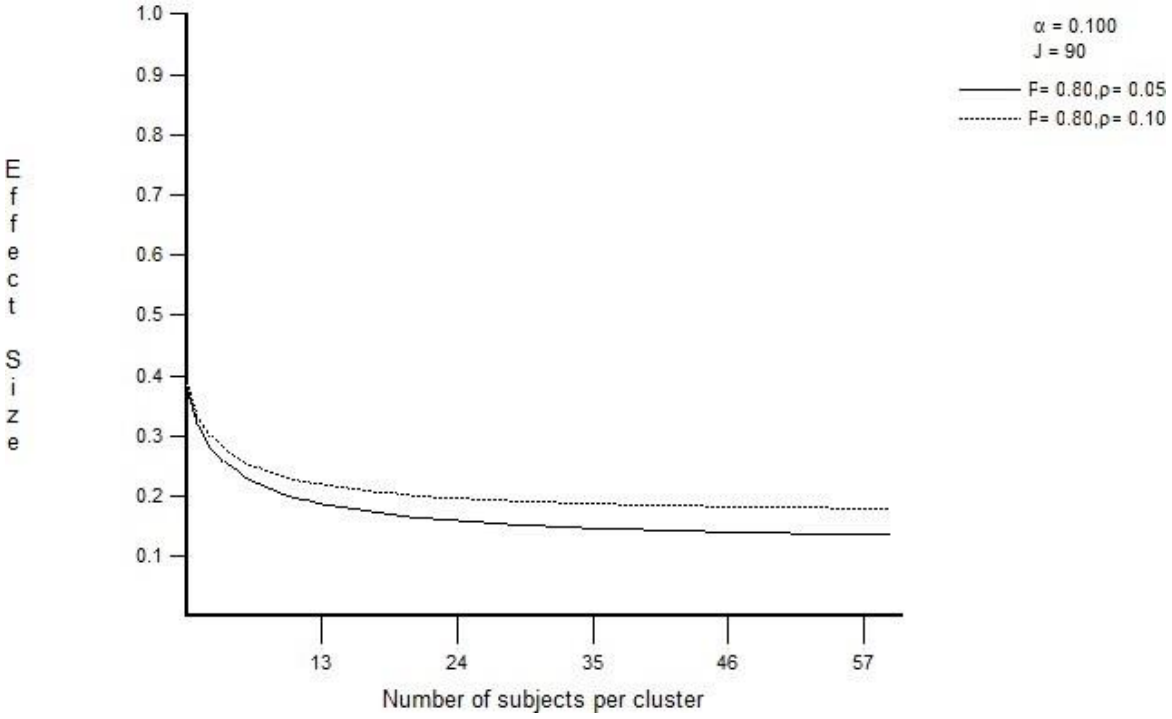
52. The RCT will rely on baseline and endline surveys, as well as repeated high-frequency measures (every two months). Power calculations were used to determine the ideal number of clusters, sample size, and frequency of data collection (see Annex 4). We present our sampling strategy for both exercises below.

Baseline and endline surveys

53. The food consumption score (FCS) was used as the main outcome for power calculations as it is a primary outcome for the impact evaluation and will be measured in all surveys. The team used FCS data collected from the Sahel for another study for the power calculations. Specifically, we calculated the number of households we need to survey in each village to detect effect sizes of 0.2 standard deviations in FCSs (which is a reasonable minimum detectable effect for Mali as determined in the literature), with a power of 0.8. Additional factors considered in the power calculations include intra-cluster correlations (ICC) of 0.05 and 0.1, and a significance level of 0.1.

54. As Figure 5 shows, below, with 91 clusters (villages) evenly split between the treatment (45 villages) and control (46 villages) groups, we would need 22 subjects per cluster to capture an effect size of 0.2 with an intra-cluster correlation of 0.1. Eleven subjects are sufficient with an ICC of 0.05. Therefore, the evaluation requires a least 22 subjects per cluster for each subgroup (e.g., men and women) of interest.

Figure 5: Power calculations using optimal design, power by number of observations per cluster



55. Based on the calculations, we surveyed approximately 60 households per village, across 91 villages. This gives us enough power to explore heterogeneous effects along several dimensions, such as socioeconomic status. More importantly, we expect enough people between households eligible and non-eligible for COVID-19 safety nets to be able to perform our analyses within each stratum with sufficient statistical power.

56. The data collection efforts in Mali started with the RSU, a full census of households in the 91 study villages. Data collection for the RSU in the 91 study villages was conducted in December 2020 by DIME. Overall, 22,445 unique households were identified as part of the RSU data collection. For the baseline survey, the evaluation team randomly sampled 60 (plus five replacement) households per village using data from the RSU, resulting in 5,093 households. It is important to note that some villages had fewer than 60 households; in those cases, the research team sampled all households. Of the 5,093 households sampled, 4,841 households were found and consented to be interviewed at baseline (a 95 percent response rate).

High-frequency surveys

57. High-frequency data collection is a relatively new approach to measuring resilience in Mali, and previous datasets that would inform power calculations are not available in the Mali context. Therefore, we use a Madagascar dataset to assess the size of the sample required and assess how frequently data should be collected to detect reasonably small changes in outcomes, such as the food consumption score (FCS).[18]

18 Depending on the size of change expected in an outcome during each period, in this case high-frequency rounds, an evaluation may need to survey more or fewer households to detect impact on a specific outcome, such as food security. To estimate expected changes, evaluations try to use previously collected data from the same context when available (e.g., national surveys, etc). In the absence of available high-frequency data on household food security from Mali, this impact evaluation uses data from Madagascar to support power calculations and estimate the sample sizes needed for high-frequency survey rounds. At the time of power calculations, the high-frequency panel from Madagascar was a rare example of the food security data required.

The Madagascar data is unique because it collects three common food security indicators: household hunger scale (HHS), food consumption score (FCS), and household dietary diversity score (HDDS). We take this data as our starting point and assess the role of survey frequency on power to compute changes in these measures over time.

58. To make recommendations for sample size and power for the high-frequency surveys, we used data collected in Madagascar on 601 households (HHs) from 32 communities surveyed every month for 18 months. We use the first 12 of these 18 months so that we are consistent in using one full year as the relevant period.[19] Then, using a data expansion approach, we increased the number of clusters to 90 communities and 1,600 households.

59. We model, through simulations, a hypothetical experiment that assigns half of the 90 communities to treatment. All households in treated communities experience one of three treatment effects:

- Increases in the mean of high-frequency measures by X percent of the control mean, keeping other parameters constant.
- Decreases in the standard deviation (SD) of food security measures for a household over time by X percent of baseline control SD, keeping other parameters constant.
- Decreases in the share of the year spent in poverty by X percent of the control proportion in poverty (as defined by standard thresholds for each indicator).

60. These simulations allow us to estimate the power needed for detecting the effects of programmes that may make households less food-insecure on average but not change variability around that mean (variation in food security) or vice versa. For each of these effects, we replicate the hypothetical experiment with the assigned effect size for a given parameter 1,000 times, regress the measure on treatment, and calculate the proportion of the 1,000 hypothetical experiments in which we can reject the null hypothesis of no impact of treatment at the 10 percent level. This proportion is our estimate of the statistical power of an experiment with this sample size to estimate the effect. The goal of these simulations is to give guidance on how frequently countries need to collect food security data in order to identify the impact on food consumption scores and other measures.

61. Power calculation tables for 91 clusters and 800–1,600 households are provided in Annex 4.

62. Table 7 in Annex 4 presents the results of power calculations needed to detect a 15 percent effect size for each of the three outcome measures with varying frequencies of data. A 15 percent effect on power gains in increasing frequency from bimonthly to monthly frequency is relatively small, but the power losses in going from a quarterly to semi-annual schedule are large. We therefore focus on comparisons with the bimonthly and quarterly schedules and compare effect sizes needed to obtain 80 percent power to guide the decisions on whether to plan for quarterly or bimonthly data collection.

63. Table 8 in Annex 4 repeats the power exercise for different effect sizes for bimonthly and quarterly schedules. We aim for bimonthly data collection for 1,600 households in 90 communities, which is sufficient to detect a 20 percent change in either the mean or standard deviation of food security at 80 percent power for all three measures. This frequency is sufficient to detect impacts on both means and standard deviations of 20 percent of control averages with 80 percent power for two of these three food security measures. In such cases, additional calculations will be performed to determine whether adding additional households or villages is necessary to increase power for expected effects.

64. Based on these power calculations, we establish a sample size of 18 households per cluster to detect effect sizes of 0.2 standard deviations with a power of 0.8 surveyed bimonthly. This amounts to 1,638 households per round distributed over the 91 clusters. We aim to have two cohorts (A and B) of 819 households (9 in each of the 91 clusters) that would be surveyed every two months using the schedule

19 Outcomes such as food consumption scores are expected to vary based on seasonal changes and agricultural cycles. Therefore using 12 months of data will allow us to account for the seasonal variations in the outcome.

in Table 2, below. Within each cohort, half of the sample is expected to include households eligible for the safety nets, and half would be non-eligible for the safety nets.

Table 2: High-frequency data collection schedule

Cohort	Apr '21	May '21	June '21	July '21	Aug '21	Sept '21	Oct '21	Nov '21	Dec '21	Jan '22
A	819 HHs		819 HHs		819 HHs		819 HHs		819 HHs	
B		819 HHs		819 HHs		819 HHs		819 HHs		819 HHs

5.2. DATA USED FOR HETEROGENEITY ANALYSIS

65. The heterogeneity analysis will utilize the data collected through baseline, endline, and high-frequency surveys. Outcomes such as food consumption score, nutrition, health, coping strategies, etc, will be used to understand the changes in well-being among households with varying characteristics, and receiving varying packages of types of programme support. Additionally, data needed to understand the effect of physical distance from basic service points, behavioural outcomes related to vaccine take-ups, school attendance, frequency of seeking medical assistance, etc, will be collected at baseline and endline. Heterogeneity analysis based on access to programme activities will also utilize the data collected from implementing teams and cooperating partners on location of facilities such as schools, health centres, distribution centres, etc.

66. Additional information required for creating sub-groups of households (e.g., children of school-going age, etc) will be collected at the baseline and follow-up surveys as relevant. It is important to note that the criteria for creating the sub-groups cannot be predetermined. Therefore, until we collect this information, we will not know how many households fall into each category, or whether we would have enough households in each group to detect any differences.

5.3. OUTCOMES MEASURED

67. The resilience programme will support households' ability to cope when hit by shocks and stressors. Typically, a programme's ability to buffer against shocks is assessed by examining the interaction between the changes in outcome and exposure to shock.[20] A growing body of resilience literature has relied on measuring the impacts of resilience programmes at single points in time, and documents positive gains in well-being.[21] However, households are systematically exposed to seasonal fluctuations and shocks, such as changes in precipitation or agricultural productivity, that affect well-being over time.

68. The impact evaluation considers the fact that people who are poor today may not be the poorest tomorrow. The capacities needed to improve and sustain well-being are also likely to evolve over time depending on the type and severity of shocks encountered. Evaluating the effect of programmes on resilience requires measuring well-being and absorptive, adaptive, and transformative capacities across seasons, and before and after shocks.

69. Building on proposals from Barrett and Constas (2014)[22] and Cissé and Barrett (2018)[23] to conceptualize resilience as avoidance of poverty in the face of shocks and stressors, each evaluation in the climate and resilience window directly measures welfare dynamics to understand resilience outcomes. These

20 Gunnsteinsson et al., 2019. *Protecting Infants from Natural Disasters*. NBER Working Papers. 35; Macours, Premand, and Vakis. 2020. Transfers, Diversification and Household Risk Strategies. Working Paper; Premand and Stoeffler. 2020. *Do Cash Transfers Foster Resilience?* Policy Research Working Paper No. 9473. World Bank, Washington, DC.
21 Macours, K., Premand, P., & Vakis, R. 2020.
22 Barrett, C., & Constas, M. 2014. Toward a Theory of Resilience for International Development Applications. Proceedings of the National Academy of Sciences of the United States of America 111 (40): 14625–14630.
23 Cissé, J., & Barrett, C. 2018. Estimating Development Resilience: A Conditional Moments-Based Approach. Journal of Development Economics 135: 272–284.

measures are calculated from a minimum set of indicators collected at higher frequencies in each country supported. Annex 2 provides additional details on the resilience measurement approach.

70. A wider range of likely outcomes is considered when answering the main evaluation questions. Annex 3 summarizes and briefly defines the key outcomes of interest for the impact evaluation in Mali.

71. The indicators were selected in collaboration with the WFP country office and the following three issues were considered: (i) operational relevance and importance to the programme components; (ii) a review of relevant literature; and (iii) evidence-generation across the portfolio of climate and resilience window evaluations.

72. The primary set of outcomes are food security indicators. For example, food consumption score (FCS), Food Insecurity Experience Scale (FIES), and household food consumption expenditures (measured at household and individual levels). A set of secondary outcomes will also be captured to understand the mechanisms of impacts, and other benefits beyond the immediate food security effects of the package of interventions.

73. The outcomes are measured during the baseline data collection, high-frequency surveys (bimonthly surveys following the baseline), and at the endline (after at least 24 months from the beginning of project implementation). A key feature of the resilience measurement approach adopted for this evaluation is reliance on high-frequency data to explore the dynamics of well-being throughout the evaluation period.

5.4. SURVEY IMPLEMENTATION

74. Data will be collected using baseline, high-frequency (HF), and endline surveys. The surveys will be identical in structure for all countries in the climate and resilience window and will only be adapted to reflect the different contexts.

- Baseline and endline surveys: Estimated to last two hours to administer, on average, for the median household.
- HF surveys: Estimated to last 30 minutes, on average, for the median household. The HF survey will be implemented every month for a year. The sample will be divided into two cohorts. Each month, one cohort will be surveyed.

75. The baseline survey is administered in the first year of the programme in 2021, before beneficiaries receive any of the treatments listed above. Baseline data collection activities include creating an exhaustive household listing in all participatory villages and a household classification to identify village distributions of wealth and socioeconomic layers. Typically, baseline surveys will occur at the same time beneficiaries are registered with the WFP a few months or a few weeks before the first cash transfer. Annex 5 lists the modules covered in the baseline and provides a link to the baseline survey.

76. The data collected at baseline is important to inform about the pre-programme situation and therefore serves as a point of reference for the impact evaluation. It is also used to verify that indicators that potentially affect the main outcomes of impact (i.e., food consumption, and food and nutrition security) are balanced and thus assure that the randomization process was successful. Furthermore, baseline data provides a last-resort opportunity to assess programme impacts when there is imperfect randomization, in which endline data alone will not be sufficient to assess the programme's impact. Then, baseline data can be used to account for observable differences between treatment and control groups to assess the programme's impact.

77. We will implement the endline survey after at least two years of programme implementation to measure changes in the outcomes of interest. We complement these yearly rounds of data collection with high-frequency surveys that ask a smaller set of questions at more regular intervals (see Table 2).

78. A key feature of high-frequency measurements is capturing intra-annual dynamics of well-being through high-frequency food security data. This strategy allows us to better understand the resilience impacts of these interventions by exploring not only the differences between beneficiaries and non-beneficiaries

before and after the programme, but also by capturing the dynamics of food security throughout the evaluation period. This will help us to understand how individuals absorb shocks, adapt to changing situations, and improve well-being over time. It is thus an important measurement strategy for understanding how the WFP's resilience programme contributes to resilience capacities in Mali.

5.5. MANAGEMENT OF DATA QUALITY

79. Multiple steps are undertaken to ensure the high quality of data collected through the impact evaluation in Mali. Factors to consider and measures taken at each stage of the data collection are summarized below.

Questionnaire development

80. Data will be collected using multi-module household surveys covering a range of outcomes. The planned baseline and endline will be identical in structure and format. The high-frequency data collection will also follow the same format for multiple rounds of data collection throughout the study period. Questions will be repeated across surveys to be able to create panel data on outcomes. The data collection instruments will be piloted extensively in each country to ensure context-specific details and option sets for each question are appropriately identified.

Tracking participants over multiple survey rounds

81. The high-frequency and endline surveys will involve revisiting as many baseline households as possible to create a panel, allowing us to control for differences in initial levels of key outcome indicators. Information required to track households over time will be collected to allow for the possibility of revisiting some or all of the households following the first round of the survey. Collecting identifiable data is necessary to verify the identity of respondents and merge data across survey rounds. Furthermore, it is needed to locate respondents for subsequent survey rounds. Participants may skip these questions if they're not comfortable answering them. Relevant data protection guidelines will be followed for all data collection exercises (See "Enumerator management and training" below for additional details). To track respondents over time and construct social networks, the following direct identifiers will be recorded.

High-frequency surveys:

- Names, addresses, and phone numbers of study participants.
- Names and phone numbers of alternative contacts to assist in the location of study participants.
- GPS coordinates of respondent's household.

Baseline and detailed follow-up surveys:

- Names, addresses, and phone numbers of study participants.
- Names, approximate addresses, and phone numbers for members of study participants' social networks (who will then be enrolled in the study as part of the first face-to-face follow-up survey).
- Names and phone numbers of alternative contacts to assist in the location of study participants for the next survey round. This will be collected both for the original study participants (treatment and control group) and members of their social network enrolled as part of the first face-to-face follow-up survey.

Enumerator management and training

82. In parallel to the development of the questionnaire, a suitable third-party monitoring agency is identified for carrying out data collection on the ground. The criteria for selecting these agencies in Mali include (i) prior experience in collecting high-quality survey data, (ii) experience in Mali, and (iii) capacity to deliver multiple household surveys in the country over the study period. While the third party will be responsible for hiring the enumerators and managing them in Mali, all data collection activities will be supported by the DIME and Office of Evaluation impact evaluation team. Detailed protocols are developed to guide the

data collection. These are developed by the impact evaluation team, who will also lead the enumerator training.

83. Enumerator training will include classroom and field training. Enumerators will be selected based on their performance during the training. The training is divided into three stages and will take approximately one week to complete:

- Review the survey's content: the team will guide enumerators through each section of the survey, eliciting their feedback about the content and answering any questions they may have about how to administer the questions to respondents. This process ensures that any ambiguities about the questionnaire are resolved ahead of time.
- Mock surveys: once the survey has been reviewed, the team will ask the enumerators to pair up and conduct "mock surveys" where they administer questions to each other. This session is followed by a question-and-answer period to review any additional concerns or questions, and to provide feedback on individual enumerators' performance.
- Review best practices: once the mock surveys are complete, the team discusses best practices for engaging with respondents and recording their answers with the software. This includes a review of:
    - How to record survey responses.
    - How to provide alternative phrasing so respondents understand the question.
    - How to ensure a smooth transition in telephone surveys, especially when the survey will be broken up into several telephone calls.

Confidentiality of data

84. Since the survey collects information about sensitive topics, strict data confidentiality protocols will be maintained throughout the evaluation. Data will be synced from the field to servers protected by passwords so that individual enumerators do not have access to the data. The data will be de-identified for analysis. Only the principal investigators (PIs) will have the key to link anonymized data to individually identifiable information; the PIs will consequently be responsible for ensuring the security of this key. No individual-level results will be reported, and all results will be aggregated to protect the identities of individuals.

Data quality protocols

85. Data will be collected electronically using the computer-assisted personal interviewing (CAPI) platform; CAPI surveys will reduce logical inconsistencies in the questionnaire. It also allows us to programme consistency checks into the survey and perform quality checks daily.

86. High-frequency checks look out for the following instances:

- Too many missing observations
- Duplicate observations
- Unusual survey duration (too short or too long)
- Too many respondents stating "no consent"
- Inconsistent patterns in the data.

87. Any anomalies that we detect through this process will be flagged to the data collection team immediately. In addition, the team will perform a set of back-checks (drawing a random 10–20 percent sample of households and calling them back to validate some of their answers). Cross-checking the data will allow us to provide immediate feedback to the field teams in case of divergences or other problems.

Internal team coordination

88. The evaluation process will be overseen by the project PIs, including coordination with programme counterparts, data collection, and analysis. The analysis will be performed by the project research assistant with close supervision from the PIs, and it will be completed in a replicable and reproducible manner.

5.6. IMPLEMENTATION MONITORING SYSTEM

89. The WFP Mali country office, the Office of Evaluation, and DIME are working together to ensure that beneficiaries receive the scheduled WFP programming on time. The country office regularly tracks when transfers are made to programme recipients, as well as whether work requirements are met. The Office of Evaluation and DIME are complementing these efforts by ensuring that the programme variations we introduce are properly followed. More specifically, the Office of Evaluation and DIME are monitoring treatment compliance in the following ways:

    - The evaluation team ensures that the unique identifier used in the survey is aligned with the beneficiary ID used in the programme.
    - The team cross-checks periodically with field teams to ensure that initial randomization plans are adhered to. Any deviation is recorded and documented systematically to be considered during the analysis stage. The high-frequency surveys are used to track self-reported participation in programme activities. This is then cross-checked with administrative data to understand adherence to the randomization and initial programme plan. The team will support country offices in establishing protocols to obtain relevant administrative data on programme participation.

90. The evaluation team also monitors any new activities that may be introduced into the treatment or control communities and, where possible, captures the impact of these activities through the measurement framework.

6. Data Processing and Analysis

6.1. DATA CODING, ENTRY, AND EDITING

91. Data will be collected by enumerators recruited by a survey firm contracted by the World Bank. The criteria used to select the survey firm include experience collecting surveys in the specific impact evaluation areas, ability to hire enumerators who know the local contexts and languages, and capacity to securely collect and manage high-quality data. All data will be collected using tablets and will be stored on SurveyCTO servers. As soon as an enumerator marks a filled-out form as "finalized", the form's content is encrypted. Whenever form data is transmitted via a 3G or other Internet network, it is encrypted in transit using SSL as well. Finally, any data downloaded from the server will be encrypted or purged of any personal identifiers before analysis.

92. Daily high-frequency checks on data quality will be implemented by the impact evaluation field coordinator and DIME research assistant with regular reports to the IE teams and field teams. The high-frequency surveys will pilot methods for reaching respondents by phone or in-person visits to determine which method is more cost-effective for minimizing non-response.

6.2. PROGRAMME-SPECIFIC QUANTITATIVE DATA ANALYSIS

Resilience package

93. To measure the impact of the FFA and integrated resilience package against the control group, our primary means of analysis is a simple regression of resilience outcomes on treatment status. A dummy variable (1/0) will be used for randomized treatment at the community level (community receives FFA/resilience or is assigned to the control group).

Descriptive targeting analysis

94. The study will document the profile of select beneficiaries along with a wide range of indicators collected at baseline. This will provide descriptive information to the WFP about the efficiency of its targeting protocols and its ability to identify the poorest households. The first rounds of high-frequency data collection may include questions about households' satisfaction with targeting to shed further light on the legitimacy of targeting among local populations.

Heterogeneity to shocks

95. A feature underlying household resilience is the ability to anticipate, absorb, adapt (e.g., avoid), cope with, and recover from shocks, while improving the trajectory for well-being (e.g., transform). Many programmes are designed to help households mitigate the impacts of shocks but evaluating this ability can be difficult. Typically, assessing the ability of a programme to buffer against shocks is done by interacting a treatment effect with a variable measuring exposure to a shock (Gunnsteinsson et al., 2019; Macours et al., forthcoming; Premand and Stoeffler, 2020). However, evaluations that measure impacts by using only baseline and endline surveys capture only a single period of the recovery trajectory, meaning that most evaluations fail to measure the full depth of welfare costs associated with shocks, full recovery, or both. Moreover, shocks are rarely pre-specified in experiments, meaning that the literature on shock mitigation may be vulnerable to publication bias.

96. To determine the differential impact of the programme based on whether a household was exposed to a shock, we will run a regression interacting programme participation with a list of pre-specified context-specific shocks that will include natural events (e.g., droughts as defined by rainfall during main cultivation months falling below a defined threshold); conflict (e.g., as defined by a recorded conflict in standardized data, such as ACLED); and economic shocks. In Mali, based on previous work in the region, information is collected on shocks related to natural disasters, increase in prices, family, conflict, and other miscellaneous events. The high-frequency data will allow us to run this regression for multiple points in the year while accounting for different types and severities of shock.

Sampling and specification

97. The sampling frame will be the lists of project sites and households as provided by the Mali WFP team. The sample will be households identified to receive benefits. Identification of recipients before implementation in all treatment arms will ensure that we can estimate intent-to-treat effects on recipient households, or likely recipient households, in pure control groups even in the event of endogenous take-up. Across all specifications, we use double-selection LASSO to select controls for precision, and we control for baseline measures of our outcomes when they are available through an ANCOVA specification. We cluster standard errors at the community level whenever the treatment of interest is assigned at the community level. In the event of non-random attrition, we will report Lee bounds on primary impacts.

6.3. PROGRAMME-SPECIFIC QUALITATIVE DATA ANALYSIS

98. The impact evaluation will use semi-structured interviews to capture information about the programme implementation process and the experience of programme participants. The questions will be developed based on the information gaps identified through the high-frequency surveys and discussions with country offices. We will be collecting qualitative information relating to the implementation process as described in Section 5. The qualitative data collected, specifically on programme implementation, will be used to support and strengthen the analysis of high-frequency survey data and to better understand the changes in outcomes observed in the impact evaluation.

99. Additionally, through interactions with the programme teams and cooperating partners, information will be collected on implementation progress, barriers to effective programme implementation, and participation levels in different activities. This will be used to gain better understanding of the context in which the programme is being implemented.

7. Ethical Considerations

100. A key goal of the WFP's impact evaluation strategy is to increase the use of rigorous evidence in delivering interventions, both in the countries directly involved in this evaluation and in other parts of the world where humanitarian and development programmes are delivered. Guided by this overarching principle, the evaluation will take into account several ethical considerations and put in place a number of relevant practices.

IRB APPROVAL

101. The evaluation team obtained international approval (on 12 November 2020) from an institutional review board for the Climate and Resilience IE window design as well as for the specific design and measurement elements in Mali. In addition, the evaluation team will obtain approvals from local institutions in Mali to ensure that the evaluation complies with local regulations and does not violate any local laws.

COMMUNICATION WITH PARTICIPANTS

102. Given that the evaluation is taking place in a context of heightened inter-communal tensions and extreme vulnerability, an evaluation risk is the perception that some groups receive benefits at the expense of others solely for research. To mitigate this risk, DIME and the WFP are working together to ensure transparent and clear communication to communities.

INFORMED CONSENT

103. The evaluation and survey teams will ensure that enumerators are fully trained to obtain informed oral consent from all evaluation participants. Every participant must consent to take part in our surveys. We are very explicit that refusal to respond to our survey does not come with any consequences for their participation in the WFP's resilience programming. The head of the household is the primary respondent for the survey. While most survey questions are addressed to the head of the household, a few questions may be directed to other household members, including women (e.g., questions on women's empowerment and food consumption for children aged 6–23 months). To avoid respondent discomfort during surveys, we will take precautions to ensure that questions are being asked keeping in mind the privacy and comfort of respondents:

- Participants may skip any questions they do not wish to answer or withdraw from the survey at any time.
- Interviews will be done at participants' homes to help them feel comfortable answering questions.
- Finally, all enumerators will complete training lasting one to two weeks. Following the training, the surveys will be piloted in the impact evaluation areas. The goals of the training are to ensure that enumerators follow survey best practices in terms of protocols and ethics, and that questions are asked in a uniform and contextually appropriate manner.

CONFIDENTIALITY

104. The evaluation team will ensure complete anonymity and confidentiality of all data collected from study participants. This means that the identity of study participants will remain hidden in all forms of data construction and analysis, and sensitive information will not be shared with anyone outside the evaluation team.

TRANSPARENCY IN EVALUATION DESIGNS

105. To increase the transparency of the work, the evaluation is registered through the American Economics Association's (AEA's) trial registry.

CONSIDERATIONS FOR REWARDING PARTICIPATION

106. The evaluation team considered providing small cash transfers to participants in the high-frequency survey. However, following discussion with the country office, it was decided it would be preferable to provide small in-kind payments (e.g., soap), which will be provided for each round of high-frequency data collection.

8. Risks and Limitations

RISKS TO INTERNAL VALIDITY

107. The primary risks to an internally valid, causal estimate of programme impact, in combination and in isolation, are statistical power for estimating multiple treatment arms with limited scope for *ex ante* identification of programme sites/participants combined with rigorous methods to create a counterfactual, such as phase-in randomization. When sites and participants have already been selected, options to create counterfactuals are limited, and with a small number of planned expansion sites, not all possible treatment arms can be implemented simultaneously. To address this problem, we use the most rigorous impact evaluation method available – a randomized controlled trial. In addition, as with any in-field RCT, spillover across communities and differential attrition are potential risks for the evaluation. However, the team will work closely with the implementing partners on the ground to monitor potential spillover risks and design clear and direct implementation protocols. We expect differential attrition to be less common than in other contexts since the control group has been made aware that they will receive the FFA intervention in the programme's second year.

108. Another related risk is low response rates due to respondents' fatigue. The provision of small in-kind payments (e.g., soap) will mitigate this risk, along with close monitoring of response rates and data quality. Previous work in the study area has shown that differential attrition is not a major concern.

RISKS TO EXTERNAL VALIDITY

109. One of the evaluation's limitations may be that the results of a single study might not be externally valid. However, we can indicate the robustness of our findings in external contexts through the formal synthesis of findings from all countries that participate in the climate and resilience window. Furthermore, comparing the findings with a similar ongoing study in Niger will be particularly interesting as there may be more similarities in programme mechanisms and impacts within the same region.

110. The use of coordinated survey instruments and data collection protocols will help ensure that data collected from Mali will be comparable with other countries in the window and in other WFP-supported evaluation windows to maximize the potential for externally valid inferences.

RISKS DUE TO INSTABILITY

111. A further risk is that a crisis (e.g., conflict, political instability, or natural disaster) impedes programme progress or the ability of implementing teams to follow the planned evaluation design. To mitigate the consequences of unforeseen issues, the evaluation team will work with the implementing partners to proactively resolve potential delays *ex ante*, including by supporting the planning and implementation of operational activities and the timely launch of procurement processes. Furthermore, field coordinators will work closely with DIME, the WFP, and the implementing partners to ensure that programme activities are being carried out according to the planned standards and protocols, and to alert the evaluation team in a timely fashion about deviations and other implementation challenges.

RISKS DUE TO COVID-19

112. As a result of COVID-19, in the early phase of the impact evaluation, the country office implemented all of its programmes with third-party NGOs that are now responsible for all field-related activities. This creates additional monitoring challenges as the evaluation team must ensure the NGOs comply with the original design (registering dual-headed households, respecting the randomization of communities to treatment arms, and delivering cash and assets on time). The evaluation team has developed a strong working relationship with the country office and frequently communicates with the country office and the NGOs to monitor these dynamics.

113. In addition, traditional in-person surveys may become difficult to implement if national authorities require social distancing. Survey activities will comply with national policies, and, if social distancing is required, alternative means of data collection, such as remote surveys by telephone or similar, will be used. These forms of data collection are currently being explored by the WFP in multiple countries.

RISKS DUE TO LOW PROGRAMME PARTICIPATION

114. In contexts of insecurity, participation in the programme activities may be reduced due to difficulties accessing activity areas (e.g., asset-creation sites and nutrition centres). It is also possible that beneficiaries find alternative income sources that are more suitable for their needs and choose not to participate in the resilience programme. If programme participation is low, it is difficult to detect statistically significant effects of the programme based on the original survey sample. It is also crucial to detect any reduction in participation early on in the evaluation and try to identify possible reasons.

115. The evaluation team is working closely with the country office to establish programme monitoring systems to closely track participation in various programme activities. Data obtained through administrative sources will be cross-checked with self-reported programme participation through high-frequency surveys to better understand programme implementation progress.

9. Quality Assurance and Peer Review

116. The WFP's Impact Evaluation Quality Assurance System (IEQAS) provides guidance on definitions, methods, processes, and procedures for ensuring that impact evaluation outputs provide robust and credible evidence about impact. The IEQAS consists of process guidance, quality checklists, templates, technical notes, and other reference materials to guide evaluation teams and partners throughout the evaluation process. Quality assurance measures will be systematically applied throughout the evaluation phases. These include preparation and selection, design, data collection,[24] consistency of programme implementation with the evaluation design, analysis, and reporting.

117. Climate and resilience window pre-analysis plans, which include each country using a similar impact evaluation design, are reviewed by the Steering Committee and Technical Advisory Group, and by external quality support peer reviewers before registration. Following registration, country-specific evaluation reports published by the WFP – including inception, baseline, and final reports – are prepared by the evaluation team. All country-specific evaluation reports are reviewed by the Evaluation Committee (see Table 5) and shared with the window's steering committee for comments. Final evaluation reports are also reviewed by external peer reviewers. In addition to WFP-published reports, the impact evaluation team produces a window-level meta-analysis and peer-reviewed journal articles. All reports and articles are reviewed by the WFP head of impact evaluation.

118. The WFP Director of Evaluation approves all of the reports before they are submitted for publication. In addition, all final evaluation reports will be subjected to a post hoc quality assessment by an independent entity through a process managed by the Office of Evaluation. The overall rating category of the reports will be made public alongside the evaluation reports.

24 This includes using high-frequency data quality checks routinely throughout the data collection phases, and ensuring the baseline and endline reports adhere to predesignated standards set by the Office of Evaluation.

10. Communication Plan

119. In Mali, the WFP's integrated resilience programme aims to address chronic food insecurity and to support communities in responding and adapting to climatic shocks and conflicts. The evidence generated from the impact evaluation will inform the programme's future scale-up or expansion. The impact evaluation will also provide insights on the most effective approaches to targeting the most vulnerable and providing support at the most effective times.

120. More broadly, the impact evaluation evidence will also contribute to the planning of the next country strategic plan for the WFP in Mali by supporting the country office in identifying resilience priorities and focus areas. Finally, the WFP's resilience programme in Mali is part of the regional WFP resilience initiative, and the impact evaluation will contribute to resilience learning in the region, as well as the development of a regional data ecosystem.

121. Considering these objectives, the impact evaluation team developed a communication plan to ensure timely dissemination of the evidence, and to allow for its use in programme design and delivery. DIME and the WFP will ensure that the WFP regional bureau and country offices are full partners in discussing and using the evidence created by the impact evaluation. More specifically, there are three complementary avenues envisioned for dissemination:

- **Active engagement with programme teams:** This includes evaluation workshops and country-level engagements. Upon completion of the evaluation phases, we will work closely with all stakeholders to elaborate relevant and visually appealing policy briefs, social media communications, and dissemination events. The evaluation team will produce a report to be shared with operational teams and policymakers in each country to summarize learning, solicit suggestions and improvements, and generate new uses for the resulting data.
- **Harnessing the WFP's and DIME's global networks:** The impact evaluation team is working closely with many different stakeholders in the development arena. The network brings together governments, donors, and academics. Evaluation results will be disseminated widely across the community of practice through workshops. In addition, we plan to make our findings broadly available to other World Bank teams and cash transfer/social protection-related projects to emphasize the role of community targeting of public goods. DIME hosts or participates in multiple workshops each year on using impact evaluations to improve project learning. Each of these workshops will be an opportunity to share evaluation findings and lessons with other agriculture and rural development projects from both within and outside of the World Bank and the WFP.
- **Academic publications:** Papers are planned based on the RCT in Mali. The impact evaluation team will engage broader academic communities to both contribute to and shape the knowledge generated by the impact evaluation. All data collected as part of the set of evaluations will be made available online through the impact evaluation database, following the World Bank's open data policy.

11. Organization of the Evaluation

122. The impact evaluation will be delivered through a partnership between the WFP and the World Bank's Development Impact Evaluation Department (DIME). DIME and the WFP will deliver the impact evaluation through the existing memorandum of understanding between the Office of Evaluation of the WFP and the World Bank. Key governing and management structures within the partnership are outlined below.

EVALUATION TEAM

123. The evaluation team will consist of principal investigators and focal points from DIME and the WFP. The composition of the team is summarized in Table 3 below.

124. The responsibilities of the evaluation team include:

- Preparation of the impact evaluation concept note and workplan
- Delivery of all activities set out in the impact evaluation workplan
- Monitor and report the progress made in delivering the workplan to the evaluation steering committee
- Prepare annual progress reports.

Table 3: Evaluation team and main counterparts

NAME	ROLE	ORGANIZATION/UNIT
Patrick Premand	Principal investigator (lead researcher)	World Bank/DIME
Guigonan Serge Adjognon	Principal investigator	World Bank/DIME
Marcus Holmlund	Resilience window manager	World Bank/DIME
Chloë Fernandez	Research officer	World Bank/DIME
Jonas Heirman	Evaluation officer	WFP/OEV
Hanna Paulose	Evaluation officer	WFP/OEV
Ola Eltoukhi	Evaluation analyst	WFP/OEV
Mariana Garcia Martinez	Research assistant	World Bank/DIME
Kane Borders	Research assistant	World Bank/DIME
Dimanche Allo	Field coordinator	World Bank/DIME

EVALUATION COMMITTEE

125. The evaluation committee will include the evaluation field coordinator, and representatives of the WFP Mali country office and the Regional Bureau in Dakar. The committee will be responsible for monitoring the progress and advising on broad strategic issues at each stage of the impact evaluation (e.g., concept notes, data collection, reviewing reports, etc). The committee will meet annually or on the sidelines of the evaluation learning workshops. Key members of the evaluation committee are listed in Table 4 below.

Table 4: Evaluation committee

NAME	ROLE	ORGANIZATION/UNIT
Ibrahima Diallo	Head of programme	WFP Mali
Emmanuel Hakizimfura	RAM officer	WFP Mali
Claudia Schwarze	Regional evaluation officer	WFP Regional Office in Dakar
Edoxi Kindane	Evaluation consultant	WFP Regional Bureau in Dakar
Allo Dimanche	Field coordinator	DIME

WORK PLAN AND DELIVERABLES

Table 5: Milestones, deliverables, and estimated timeline

MILESTONES	DELIVERABLES	TIMELINE
Agreement on the impact evaluation design	Concept note	August 2020
Data collection plan and pilot	Terms of references (TORs) Questionnaires	November 2020–February 2021
Data collection (baseline)	Cleaned data Dictionaries	February–September 2021
First data analysis	Presentation Data file Do files Baseline report	November 2021–January 2022
Implementation of intervention aligned to evaluation	Rollout plan Monitoring reports verifying treatment and control status	Starting February 2021
High-frequency survey data collection plan	TORs Questionnaire Sampling plan	March 2021
Data collection (high-frequency surveys)	Cleaned data dictionaries Datasets	Bimonthly April 2021–August 2022
Draft inception report	Peer-reviewed methodology note	October 2020
Final inception report	Report published on WFP.org	September 2022
Follow-up data collection plan	TORs Questionnaire	September 2022
Baseline report	Report published on WFP.org	October 2022
Data collection (follow-up)	Cleaned data Dictionaries	July 2023/aligned with the programme timeline
Final report and policy notes	Technical note Policy note Data file Do files	August 2023/aligned with the programme timeline
Dissemination of findings	Presentations	December 2023/aligned with the programme timeline

Annexes

Annex 1: Window Summary

126. The climate and resilience Window was developed by the WFP's Office of Evaluation (OEV) in partnership with the WFP's Asset Creation and Livelihoods Unit, and Climate and Disaster Risk Reduction Unit, as well as the World Bank's Development Impact Evaluation (DIME) department. The window will coordinate a portfolio of impact evaluations to measure the impact of the WFP's resilience programmes on household resilience across a set of countries.

127. Windows are coordinated portfolios of impact evaluations on a specific evidence area – in this case, climate and resilience. Windows are part of the WFP's impact evaluation strategy and are coordinated by the WFP's Office of Evaluation and DIME department of the World Bank. They inform country offices about the effectiveness of their programming and contribute to a global evidence base by examining similar questions about the same concept in multiple programming contexts. Through a coordinated portfolio of impact evaluations, windows aim to increase the power of evidence generated and expand its ability to be generalized across contexts, thereby improving external validity.

128. The concept of resilience has gained attention in the development and humanitarian sectors because it recognizes that a household's well-being depends on social, economic, human, and environmental capital, as well as exposure to – and ability to cope with – shocks and stressors. Therefore, it is centred around addressing shorter-term humanitarian needs while simultaneously supporting communities to face future crises induced by climate change, conflict, and other factors. Many institutions, including the United Nations Children's Fund (UNICEF), the World Bank, and the World Food Programme (WFP), have increasingly used the concept as a basis for their programming.

129. The WFP's resilience policy uses the definition of the Technical Working Group of the Food Security Information Network (FSIN) for its resilience programming, which defines resilience as "the capacity to ensure that shocks and stressors do not have long-lasting adverse development consequences".[25] The set of capacities, required before, during, and after the onset of shocks and stressors, supports the ability to (i) absorb shocks and stressors; (ii) adapt to change through making proactive choices; and (iii) transform, thus changing the available choices. The capacities contribute to maintaining development gains during shocks and stressors.

130. To strengthen resilience, the WFP employs an integrated approach to programming, where multiple forms of support are provided to the same community. These integrated packages of interventions aim to improve food security and nutrition by smoothing and improving food consumption in the short term, while supporting livelihoods and addressing barriers to development (e.g., better climate information, access to markets, education, Water, Sanitation and Hygiene (WASH), and community ownership and leadership) in the long term. Rigorous evidence on how these interventions contribute to resilience is needed to design programmes that simultaneously address the root causes of food insecurity and malnutrition while meeting immediate food needs.

131. The climate and resilience impact evaluation window will support resilience programme teams in designing impact evaluations to understand how the integrated packages of interventions, and activities within the package, contribute to resilience. Currently, resilience impact evaluations in four countries (Mali, Niger, Rwanda, and South Sudan) are part of the window, with the expectation that at least one more impact evaluation will be added.

25 FSIN Resilience Measurement Technical Working Group. 2014. Resilience Measurement Principles: Toward an Agenda for Measurement Design. Rome, FAO & WFP.

132. Each window is guided by one or more pre-analysis plan(s) (PAP). The first climate and resilience window PAP describes how evaluations will estimate the impacts of experimentally varying livelihoods, education, health, and complementary activities on resilience. Resilience is measured in the window through baseline, endline, and high-frequency surveys that capture changes in household well-being, defined in terms of food consumption and food and nutrition security. Climate and resilience window impact evaluations also examine the timing and sequencing of activities, as well as their targeting modalities, to understand if, and how, programme designs can be most effective.

Annex 2: Defining Resilience

133. This annex describes how we plan to conceptualize resilience by measuring dynamic outcomes, such as food security, school attendance, and labour outcomes. The material in this annex is closely adapted from the WFP climate and resilience pre-analysis plan.

DEFINING RESILIENCE THROUGH HIGH-FREQUENCY MEASUREMENT

134. Measurement of resilience has mostly taken one of three approaches in the literature. The first is to define *ex ante* characteristics of households that are expected to be associated with lower resilience, and construct a "resilience index". This is the approach of the Food and Agriculture Organization's (FAO) RIMA index or the TANGO resilience index, as well as examples of resilience evaluations that use characteristics like diversification of livelihood strategies as a proxy for resilience. The second is to regress outcomes on measures of shocks to isolate the contribution of shocks to food security. The third is to use the measurement of different households' food security at different times to impute a given household's food security path and then measure parameters of the imputed distribution.

135. Our measurement framework extends these existing imputation-based measures of food security dynamics by allowing idiosyncratic shocks that are not shared across households. The measures of interest are closely related to proposed measures of vulnerability, but we aim to measure underlying consumption smoothing behaviour rather than the welfare consequences of such behaviour. Resilience is best described not by a single index, but by the following simple structural equation for household welfare:

$$y_{it}=\alpha_i+f_i(d)+\delta_i t+\epsilon_{it} \quad \text{yit=αi+fid+δit+ϵit}$$

where yit is a measure of well-being, such as aggregate consumer expenditure, food security, or poverty status, for an observation unit i at time t. Since the programmes included in the study primarily focus on improving food security and nutrition outcomes, selected food security indicators will be used as measures of well-being. The four components of this equation determine a household's ability to avoid food insecurity over time and can be estimated as a regression of household food security on time and survey dates. To understand this equation, imagine using this framework to estimate a household's level of resilience. Specifically, αi, the household-specific fixed effect, measures a household's reference level of food security. The second term is a function of the calendar date on which food security is measured, and measures seasonality. The third term is a trend measuring how quickly a household is improving food security over time t. Finally, ϵit measures exposure to shocks not systematically correlated with survey dates. Figure 8 shows how this looks in a plot, where we measure a household's consumption or food security status in every period from t = 0 to some period t = T.

136. Impact evaluations typically focus on measuring a household's consumption at one point in time, with the view that a single observation is a sufficient statistic for that household's reference level of well-being for a given year. In panel A, the red and blue households differ only in their value of α. The household whose consumption is depicted by the red line is always "more food-insecure" than the household whose consumption trajectory is shown by the blue line, meaning that for any given food security threshold, the blue household will be food-insecure if, and only if, the red household is also food-insecure.

137. However, the average food security of the household over the period (αi) only captures one feature of the consumption function that is important for welfare analysis. The blue household in panel B has a steeper δ, indicating a steeper trend in food security, meaning that this household will move above the poverty line and/or farther away from it. The blue household in panel C has a seasonal pattern with greater variability than the household with a red line. Seasonality could lead to households falling below

a food security threshold in the lean season. In panel D, both the red and blue households experience a shock at the same point.

138. Given the structure of the equation of motion for consumption above, each component could be estimated if data were collected every day from t = 0 to T. However, such data is virtually impossible to collect and may not be necessary to distinguish impacts arising from influencing different components of the well-being equation. We propose to operationalize resilience measurement by repeated sampling of the same household on different dates within a predefined period and estimating key household-specific parameters of the structural consumption equation from this sample of consumption at different dates.

OPERATIONALIZING FEASIBLE MEASURES OF RESILIENCE

139. These impact evaluations will estimate welfare trajectories within a one-year period following the start of a programme. Figure 6 shows a hypothetical consumption path for a household in a period t = 0…T. The dynamics shown could represent either a seasonal consumption path with one lean season and one peak season, or a household that experiences one positive and one negative shock.

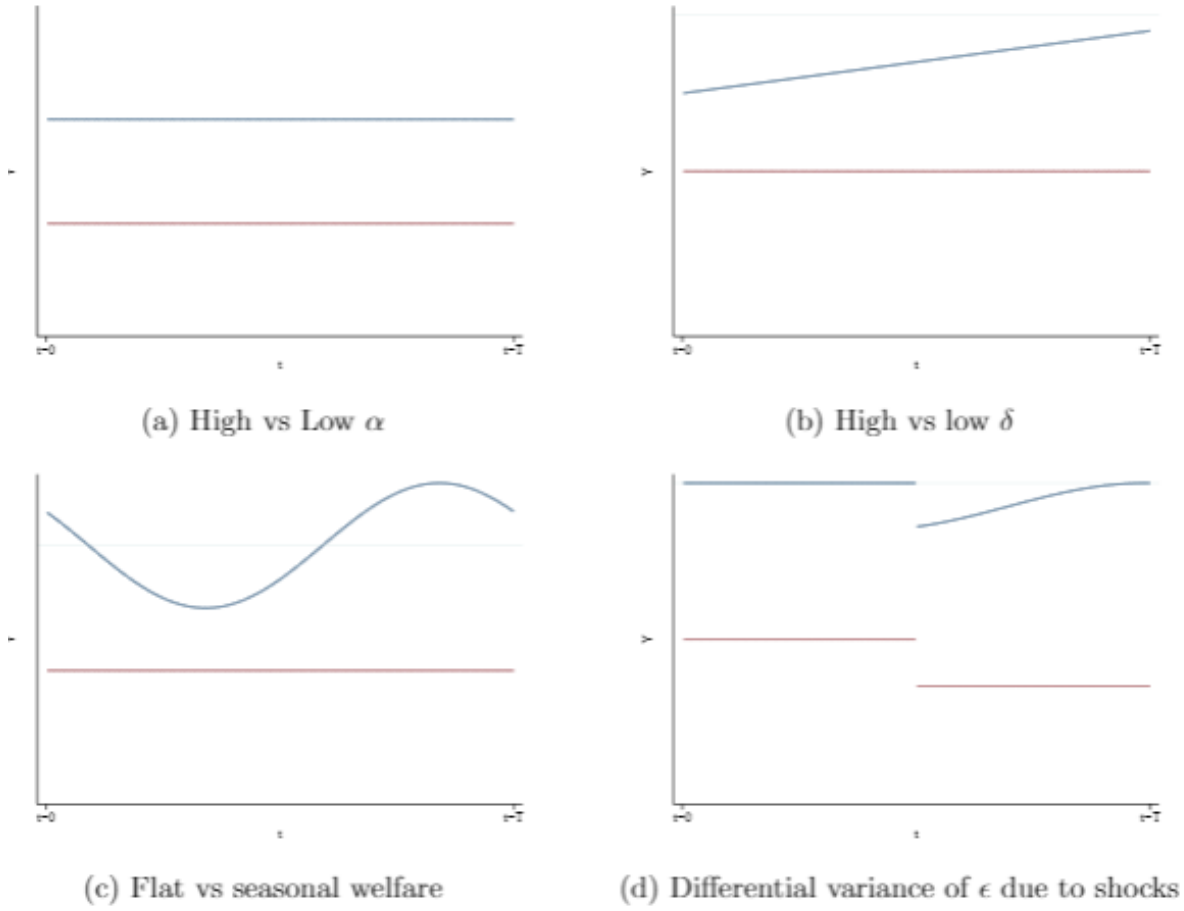
140. The first measure of the consumption equation we are concerned with is the household's intra-annual reference level of consumption – this is αi in the structural equation. If we observed a household's consumption value every day, this would be measured as a household's average food security status over the period – as shown by *m* in Figure 7 (panel A). Next, we consider the household's intra-annual standard deviation, the average of the household's deviations from the reference mean (Figure 7 – panel B). The standard deviation captures the combined influence of both $f(d)$ and (ϵ) on household welfare trajectories. This single indicator summarizes the variability associated with both seasonality and shocks within the period. The third measure is the time trend. However, by limiting the comparison within a year, we do not consider a year-on-year trend in welfare. The final measure we consider is the share of the period the household spends below a poverty line or food security range. This is the number of days covered below the poverty line divided by the total number of days in the period of interest (Figure 7 – panel C). Resilience is then defined as the ability of a household to avoid poverty over time, which we operationalize in the following way:

- A household with a higher *m* is on average higher above or less below the food security threshold. So, households with higher *m* are more resilient than households with lower *m*. The intra-annual reference mean of food security is measured by $mi\hat{}=1ni\sum Tt=0yit$mi^=1ni∑t=0Tyit
- Conditional on *m*, having a higher standard deviation will increase the likelihood of falling below a food security threshold, the share of time spent below the poverty threshold, and/or the number of days relatively far below the food security threshold. Conditional on *m*, households with a higher standard deviation are less resilient. The intra-annual reference standard deviation of food security is measured by: $s\hat{}=1ni\sqrt{\sum Tt=0(yit-mi)2}$s^=1ni∑t=0Tyit−mi2
- Households that spend more time below the threshold are less resilient than households that spend less time below the line. The share of observations below a poverty line is measured by $sharei\hat{}=1ni\sum Tt=1\mathbb{1}(yit<y-)$
- where ni is the number of times community, household, or individual *i* is surveyed; *T* is the length of the period over which resilience is measured, yit is a measure of household food security status, and $y-$ is a threshold below which a unit is considered poor or food-insecure. These three measures, defined for a selected set of food security indicators, will be our main welfare outcomes. Below we consider power and describe how frequently we need to measure outcomes to detect changes in these outcomes associated with interventions.

141. Figure 8 shows what the measures look like for the household with the hypothetical sinusoid function shown so far, assuming a quarterly data collection schedule in which food security status is observed at

three-month intervals. For this household, the reference level of consumption *m* (shown by the red dashed line) is simply the average of the four points. The intra-annual standard deviation is estimated by calculating the standard deviation of the four points, the average of the solid red lines. The range is the difference between the highest of the four values and the lowest, the difference between the dashed black lines. And the share of the period spent below the poverty line is the number of observations that fall below the poverty line (the grey dashed line) divided by the total number of observations (i.e., the number of grey dots divided by the number of blue dots).

Figure 6: Examples of capacities over time



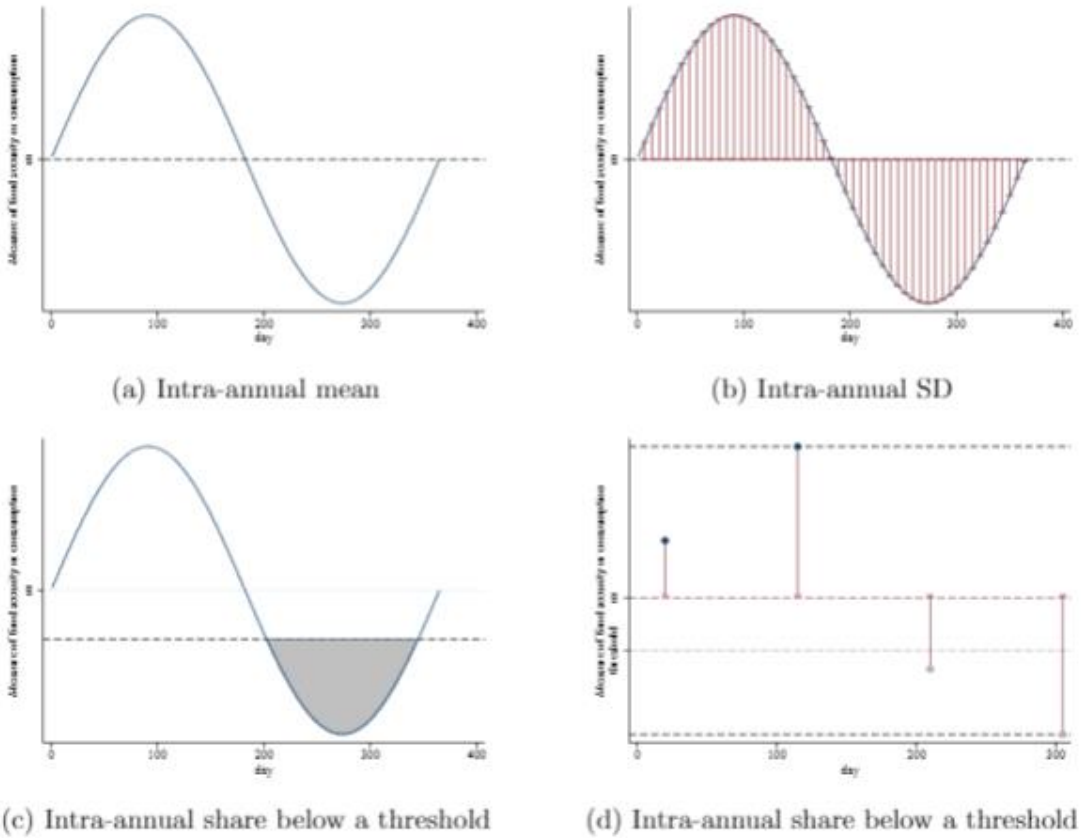
(a) High vs Low α

(b) High vs low δ

(c) Flat vs seasonal welfare

(d) Differential variance of ϵ due to shocks

Figure 7: Measures of capacities



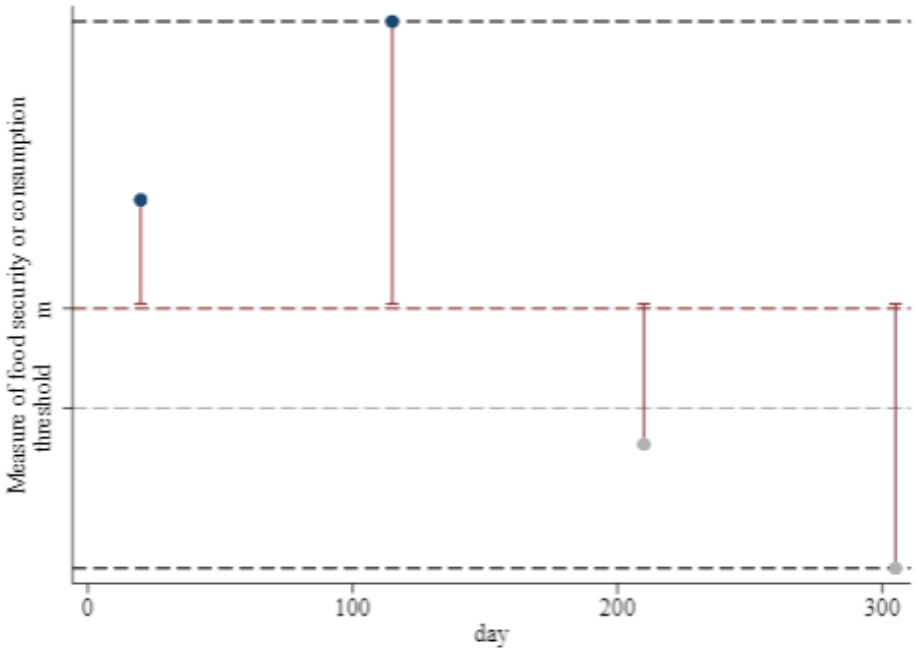
(a) Intra-annual mean

(b) Intra-annual SD

(c) Intra-annual share below a threshold

(d) Intra-annual share below a threshold

Figure 8: Feasible measurement of capacities



Annex 3: Main Outcomes of Interest

Table 6: Main outcomes of interest

Outcome type	Outcome name	Definition	Measurement level	Source
Primary	Consumption and food security	FCS/FIES/expenditure.	Household/ individual	Baseline, endline, and high-frequency surveys
Secondary	Time use	Activities and time spent at points of the day for selected household members.	Household/ individual	Baseline, endline, and high-frequency surveys
Secondary	Assets	Number and value of assets owned by the household from a contextually predefined list.	Household	Baseline and endline surveys
Secondary	Income-generating activities	Participation in non-farm business, agriculture and livestock, or wage employment and revenue from these activities.	Household/ individual	Baseline, endline, and high-frequency surveys
Secondary	Shocks and coping mechanisms	Shocks encountered by the household, including the severity of shocks, and coping strategies used. Selection of shocks from a predefined list.	Household/ individual	Baseline, endline, and high-frequency surveys
Secondary	Financial outcomes	Current savings levels, the number of loans taken and current outstanding debt, insurance products currently owned, and cash transfers undertaken, including remittances.	Household/ individual	Baseline and endline surveys
Secondary	Migration	Migration of household members.	Household/ individual	Baseline and endline surveys
Secondary	Psychosocial well-being	Stress, life satisfaction, self-efficacy, aspirations, Center for Epidemiological Studies Depression scale.	Household/ individual	Baseline and endline surveys
Secondary	Women's empowerment	As defined by CBT/gender window (normative and positive time use and agency).	Household/ individual	Baseline and endline surveys
Secondary	Social capital	Social cohesion, closeness of community index, financial support index, collective action index.	Household/ individual	Baseline and endline surveys
Secondary	Safety nets	Amount and source of transfers from other NGO and government sources.	Household/ individual	Baseline and endline surveys

Secondary	Reservation wages	The minimum hourly wage that selected household members would accept to engage in short-term labour and duration they would be willing to work.	Household/ individual	Baseline, endline, and high-frequency surveys

Annex 4: Power Calculations

Table 7: Power calculations to detect a 15 percent effect size

FCS

Frequency	Mean	SD	Share of obs < threshold
Monthly	0.997	0.932	0.526
Bi-monthly	0.996	0.884	0.420
Quarterly	0.958	0.712	0.365
Semi-annually	0.962	0.387	0.302

HDDS

Frequency	Mean	SD	Share of obs < threshold
Monthly	0.991	0.925	0.861
Bi-monthly	0.973	0.847	0.713
Quarterly	0.972	0.700	0.802
Semi-annually	0.948	0.455	0.531

HHS

Frequency	Mean	SD	Share of obs < threshold
Monthly	0.421	0.832	0.330
Bi-monthly	0.394	0.697	0.307
Quarterly	0.372	0.537	0.285
Semi-annually	0.262	0.268	0.245

Table 8: Power calculations to detect different effect sizes for bimonthly and quarterly schedules

FCS

Bimonthly				Quarterly			
Effect size	Mean	SD	Share of obs < threshold	Effect size	Mean	SD	Share of obs < threshold
15%	0.996	0.884	0.420	15%	0.958	0.712	0.365
20%	1.000	0.990	0.600	20%	1.000	0.922	0.546
25%	1.000	1.000	0.777	25%	1.000	0.984	0.683
30%	1.000	1.000	0.881	30%	1.000	1.000	0.851

HDDS

Bimonthly				Quarterly			
Effect size	Mean	SD	Share of obs < threshold	Effect size	Mean	SD	Share of obs < threshold
15%	0.973	0.847	0.713	15%	0.972	0.700	0.802
20%	0.999	0.973	0.915	20%	1.000	0.892	0.968
25%	1.000	0.998	0.992	25%	1.000	0.971	0.997
30%	1.000	1.000	1.000	30%	1.000	0.996	1.000

HHS

Bimonthly				Quarterly			
Effect size	Mean	SD	Share of obs < threshold	Effect size	Mean	SD	Share of obs < threshold
15%	0.394	0.697	0.307	15%	0.372	0.537	0.285
20%	0.523	0.892	0.423	20%	0.480	0.752	0.362
25%	0.603	0.982	0.527	25%	0.548	0.859	0.436
30%	0.725	0.992	0.654	30%	0.653	0.940	0.563

Annex 5: Questionnaires

A link to the baseline questionnaire is available here. The modules included in the survey are summarized in Table 9.

Table 9: List of modules included in the baseline questionnaire

Module	Description
A	Introduction
B	Consent
C	HH roster
D	Education & employment
E	Income-generating activities – non-agricultural business
F	Income-generating activities – agriculture & livestock
G	Food consumption score (FCS)
H	Food Insecurity Experience Scale (FIES)
I	Consumption expenditure (food & non-food)
J	Asset index plus access to basic services
K	Psychosocial & mental health
L	Shocks
M	Coping strategies
N	Migration
O	Financial outcomes (savings, loans, insurance, cash transfers)
P	Time-use
Q	Safety nets
R	Social capital
S	Women's empowerment
T	Women's dietary diversity
U	Child health
	End of survey

Annex 6: Detailed Stakeholder Analysis

142. Stakeholders and users of this evaluation are defined as those actors who may influence the evaluation, and those who may be influenced by it. This includes internal, external, and national actors and programme beneficiaries. The WFP country office in Mali is intended to be the primary user of this evaluation. In addition, the WFP Mali country office provides technical guidance at the national level to inform national policy and dialogue on social protection, and the country office has expressed interest in using the results of this evaluation to support this technical advisory capacity.

143. The various categories of stakeholders include:

- Internal Mali-based stakeholders: the Country Director and deputy director, the head of programme, and all technical and management personnel.
- Internal stakeholders outside of Mali: the Office of Evaluation, the Regional Bureau for Central and Western Africa in Dakar, and the resilience divisions and the WFP.
- Population groups in need (affected populations): resident communities and migrants of different sexes and age groups.
- External stakeholders, including international non-governmental organizations, donors, United Nations' agencies, and forums in Mali.
- National stakeholders, including national and subnational government actors, and non-governmental organizations.

144. The main users of the evaluation (country office management and staff in-country) may be much affected by the evaluation and are actively engaged in its development. Populations in need of WFP assistance will also have a high stake in the results and will be the primary providers of data for the evaluation.

145. Stakeholder engagement will vary depending on the category, but may include:

- Reviewing and commenting on the draft inception report.
- Active monitoring of the evaluation design during programme implementation.
- Participation in the final learning workshop.
- Reviewing and commenting on the draft evaluation report.
- Reading the final evaluation report and other evaluation communication products.

146. More detailed information about evaluation users is provided in Table 10 below. This table introduces all categories of stakeholders, the degree to which they have expressed an interest to be included in the evaluation, how they might be engaged, and how they are expected to use the evaluation results.

Table 10: Stakeholder analysis

Who are the stakeholders?	What is their role in the intervention?	What is their interest in the evaluation?	How should they be involved in the evaluation (be informed, act as key informant, be part of a focus group interview, be part of a reference group, etc)?	At which stage should they be involved?	How important is it to involve them in the evaluation (high, medium, low)?
WFP internal stakeholders					
WFP country office	The main implementer of the programme under evaluation	To inform upcoming country strategic plan and relevant programming	The country office is responsible for implementing the programme according to the evaluation design. It actively provides feedback on the tools and outputs of the evaluation.	From the scoping stage	High
WFP regional bureau	Governance and technical advisory role	To inform regional programme strategies, to support other country offices in evidence generation	As members of the evaluation committee, technical advisors on relevant portions of the questionnaire, data collection activities and implementation.	From the scoping stage, with regular meetings to provide feedback on tools and outputs	High
Office of Evaluation	Coordination of impact evaluation and liaisons with the country office	As coordinators of the impact evaluation and for the WFP, in alignment with impact evaluation strategy (2019–2026)	The Impact evaluation team will be involved in the field coordination meetings and evaluation committee meetings as support to the country office and impact evaluation team.	From the scoping stage	High
External stakeholders					
Affected communities	Affected communities, including men, women, boys, and girls, will be the primary	Beneficiaries will likely have a strong interest in any changes in targeting, reach, or effectiveness of future	Beneficiaries and non-beneficiaries will provide the primary source of data on effectiveness.	From the targeting and selection stage	High

	participants of the intervention	programming as a result of the evaluation and recommendations. Women and girls have a particular stake in the results meant to shed light on recommendations for improving gender equality			
Government at the local and central levels in Bamako, Gao, Koulikoro, Mopti, and Tombouctou	National and local government structures provide ethical and administrative clearance for programming and evaluation efforts and oversee local development initiatives and national social protection programmes	The WFP has an established relationship with the national government to provide technical support on food security and nutrition; evaluation results will support these efforts	The evaluation receives national-level clearance before inception.	At the initial scoping for the intervention & dissemination of findings	Medium
Local non-governmental organizations: Jeunes Sans Frontières (JSF), G Force, Yam-Giribolo-Tumo (YAGTU), Association Malienne pour la Securité et la Souveraineté Alimentaire (AMASSA), GAAS Mali G Gorfe Boidiè, ADAZ Tbt, Association pour la Promotion du Monde Rural au Sahel (APROMORS), Oeuvre	As implementing partners for the programme under evaluation	Evaluation results can inform their livelihood and gender transformation programming	As cooperating partners, JSF, G Force, YAGTU Dandoli, AMASSA-Dourou, GAAS-Soroly, G Gorfe Boidiè, GAAS-Tamani, ADAZ Tbt, APROMORS, and GRADP are responsible for ensuring the programme is implemented in line with the evaluation design.	At the initial scoping for the intervention & dissemination of findings	High

Malienne pour le Développement des Zones Arides (OMADEZA), GRADP					
International non-governmental organizations: The Children and Community Initiative for Development (CAID), SOS Sahel, Afrique Verte International, SOCODEVI, Association pour la Promotion du Monde Rural au Sahel (APROMORS)	Good Neighbors is an implementing partner for the programme under evaluation	Evaluation results can inform their livelihood and gender transformation programming	As a cooperating partners, Amassa Afrique Verte/OMADEZA, SOCODEV, CAID, SOS Sahel, Are responsible for ensuring the programme is implemented in line with the evaluation design.	At the initial scoping for the intervention & dissemination of findings	High
World Bank	Development Impact Evaluation (DIME) department	In line with the Office of Evaluation–DIME partnership, DIME is interested in producing and disseminating the evaluation results as part of a broader research portfolio	As the primary investigators and research analysts.	At the initial conceptualization of the window	High
Federal Ministry for Economic Cooperation and Development (BMZ)	Primary funder of the intervention	As a user of the evaluation	BMZ and Kfw are informed at key milestones in the evaluation. They have an interest in using the results as evidence for other funded projects.	At the proposal stage of the intervention	Medium

Annex 7: Detailed Evaluation Process

Table 11: Detailed evaluation timeline

Phase 1 – Preparation	Involved	Estimated date
Regional discussion on impact evaluation	CO/OEV	March 2019
Discussions with BMZ	BMZ/CO/OEV/DIME	May 2019
Impact evaluation approved by BMZ	BMZ	August 2019
Impact evaluation workshop	CO/DIME/OEV	September 2019
Impact evaluation inception mission to Mali	CO/OEV/DIME	February 2020
Phase 2 – Inception report		
Regional discussion on impact evaluation	DIME/OEV	March 2020
Partnership Steering Committee	OEV	May 2020
Impact evaluation design discussions & adapting to COVID-19	CO/OEV/DIME	April–October 2020
Inception report drafting, quality assurance & peer review	OEV/DIME	January–December 2021
Approval and dissemination of the inception report with country office, regional bureau, evaluation committee, window's reference group, steering committee, online/social media as appropriate	DIME/OEV	March 2022
Phase 3 – Baseline and high-frequency data collection		
Pre-baseline wealth ranking and baseline data collection preparations	CO/OEV/DIME	October–December 2020
Baseline data collection	DIME	December–January 2021
Data collection through high-frequency surveys	DIME/OEV/CO	April 2021 – August 2022
Phase 4 – Baseline report		
Data analysis and baseline report drafted by the impact evaluation team, and submitted for quality assurance and revisions	DIME/OEV	May 2021–March 2022
Publication of the baseline report	OEV	March 2022
Dissemination of the baseline report with survey respondents, country office, regional bureau, evaluation committee (and other evaluation stakeholders), window's reference group, steering committee, online/social media as appropriate	DIME/OEV	March 2022
Phase 5 – Programme implementation		
Randomization	DIME	March 2021
Assignment intervention and comparison sites	DIME/CO	March 2021
Rollout programme activities as per randomization	CO	April 2021

Monitoring programme activities verifying treatment and control status	CO/DIME	April 2021–endline
Phase 6 – Endline data collection		
Preparation of data collection tools, including survey questionnaire, digital devices, sampling strategy, and training material	DIME/OEV/CO	December 2022
Pilot and finalization of data collection tools	DIME	March 2023
Recruitment of enumerators/data collection firm	CO	April 2023
Enumerator training	CO	April 2023
Data collection process and live monitoring data quality checks	DIME	May 2023 (aligned with the programme timeline)
Feedback/data-sharing mechanisms, as appropriate/possible		July 2023
Phase 7 – Final evaluation reports		
Data analysis and final evaluation report drafted by the impact evaluation team, and submitted for quality assurance and revisions	DIME/OEV	August 2023
Publication of the final evaluation report	OEV	December 2023
Dissemination of the final evaluation report with survey respondents, country office, regional bureau, evaluation committee (and other evaluation stakeholders), window's reference group, steering committee, online/social media as appropriate	OEV/DIME/CO	September–December 2023
Final evaluation report reviewed by post hoc quality assessment	OEV	December 2023
Phase 8 – Management response		
Based on the findings, country office to develop a management response	CO	August 2023
The Office of Evaluation to review and, if needed, respond to the management response	OEV	Sept 2023
Publication of the management response	OEV	October 2023
Phase 9 – Dissemination and learning		
Webinar presenting the findings	OEV/DIME	August 2023
Blogs, summary briefs, other relevant communication products	OEV/DIME	September 2023
Considerations for academic publications	DIME/OEV	November 2023

Annex 8: References

Barrett, C. & Constas, M. 2014. Toward a Theory of Resilience for International Development Applications. *Proceedings of the National Academy of Sciences of the United States of America*, 111(40): 14625–14630.

Bossuroy, T., Goldstein, M., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C., Udry, C., Vaillant, J. & Wright, K. 2021. *Pathways Out of Extreme Poverty: Tackling Psychosocial and Capital Constraints with a Multi-Faceted Social Protection Program in Niger*. Policy Research Working Paper 9562. World Bank, Washington, DC.

Carucci, V., Ronchini, S., Gentile, J., Policastro, R., Crahay, P., Primozic, M., Vaquier, D., Gonzalez, R., Gardesten, J. & Mariangeloni, S. 2016. Food Assistance for Assets (FFA) for Zero Hunger and Resilient Livelihoods: A Program Guidance Manual.

Cissé, J. & Barrett, C. 2018. Estimating Development Resilience: A Conditional Moments-Based Approach. *Journal of Development Economics,* 135: 272–284.

FSIN. 2021. Global Report on Food Crises: Joint Analysis for Better Decisions. [Link]

FSIN Resilience Measurement Technical Working Group. 2014. *Resilience Measurement Principles: Toward an Agenda for Measurement Design*. FAO & WFP, Rome.

Gunnsteinsson, S., Adhvaryu, A., Christian, P., Labrique, A., Sugimoto, J., Shamim, A. & West, K. 2019. *Protecting Infants from Natural Disasters: The Case of Vitamin A Supplementation and a Tornado in Bangladesh*. NBER Working Paper 25969.

Macours, K., Premand, P. & Vakis, R. Forthcoming. Transfers, Diversification and Household Risk Strategies: Experimental Evidence with Lessons for Climate Change Adaptation. *The Economic Journal.*

Premand, P. & Stoeffler, Q. 2020. *Do Cash Transfers Foster Resilience? Evidence from Rural Niger*. Policy Research Working Paper 9473. World Bank, Washington, DC.

Stockholm International Peace Research Institute. N.d. Mali. [Link]

World Food Programme. 2020. Mali Country Strategic Plan (2020–2024). [Link]

World Food Programme. 2021. Mali Emergency Dashboard, December 2021. [Link]

World Food Programme Mali. 2021. Country Brief. [Link]

Acronyms

CAPI	Computer-assisted personal interviewing
CBT	Cash-based transfer
CO	Country office
CSP	Country strategic plan
DIME	Development Impact Evaluation (DIME) department (World Bank)
FFA	Food assistance for assets
HFC	High-frequency checks
IE	Impact evaluation
IRB	Institutional review board
OEV	Office of Evaluation (World Food Programme)
PAP	Pre-analysis plan
RCT	Randomized controlled trial
WFP	World Food Programme

Office of Evaluation

World Food Programme

Via Cesare Giulio Viola 68/70
00148 Rome, Italy
T +39 06 65131 wfp.org