WFP
World Food Programme

SAVING LIVES
CHANGING LIVES

Impact Evaluation for Resilience Learning in South Sudan

Inception report

unicef

i2i DIME
TRANSFORM DEVELOPMENT

August 2022

Contents

Key Personnel for the Evaluation

WFP OFFICE OF EVALUATION

Andrea Cook, Director of Evaluation

Jonas Heirman, Evaluation Officer

Hanna Paulose, Evaluation Officer

WORLD BANK DEVELOPMENT IMPACT EVALUATION (DIME) DEPARTMENT

Paul Christian, Principal Investigator

Erin Kelley, Principal Investigator

Greg Lane, Principal Investigator

Marcus Holmlund, Research Manager

Eric Jospe, Research Analyst

Julia Ashikbayeva, Research Analyst

Roxana Elena Manea, Field Coordinator

UNICEF SOUTH SUDAN COUNTRY OFFICE

Alwin Nijholt, Social Policy Officer

WFP SOUTH SUDAN COUNTRY OFFICE

Wilson Kaikai, Monitoring & Evaluation Officer

Anna Soper Programme Policy Officer

Photocredits:

© WFP/Gabriela Vivacqua

1. Introduction

1. In 2020, 155 million people worldwide faced "crisis or worse" levels of food insecurity. Almost 115 million of them lived in countries affected by conflict or weather extremes.[1] Extended food crises were predicted in 2021, which would necessitate continued large-scale humanitarian assistance. In South Sudan, 7.4 million people faced acute food insecurity in April-July 2021.[2] Climate shocks such as floods,[3] as well as continuing conflicts, are identified to have significantly affected food security, and the COVID-19 pandemic has further aggravated the crisis.

2. Close to 70 percent of the population of South Sudan (8.3 million people) need some form of humanitarian assistance. Of these, close to 108,000 people face "catastrophic" levels of food insecurity – the highest level in the Integrated Food Security Phase Classification System (IPC).[4] Approximately 1.4 million children were expected to experience acute malnutrition in 2021,[5] and more than 2 million children were out of school. To respond to these development challenges, the United Nations Children's Fund (UNICEF), the World Food Programme (WFP), and partners with support from the German Federal Ministry for Economic Cooperation and Development (BMZ) have created a multi-year Joint Programme to strengthen resilience in urban and peri-urban communities in South Sudan. Through a four-year commitment, the programme is intended to enhance resilience to shocks by meeting immediate food and nutrition needs, strengthening livelihoods, and improving access to basic services such as education and health care. However, there is a lack of evidence on how development outcomes are affected by these shocks, and how UNICEF's and WFP's programmes support populations to effectively respond to these shocks.

3. The World Food Programme's (WFP) Office of Evaluation (OEV), Asset Creation and Livelihood Unit, and 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. WFP's Impact Evaluation Strategy (2019–2026) focuses on delivering impact evaluations that contribute to global evidence as well as 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.

4. The Climate and Resilience Window was designed to help with understanding of how WFP's programmes contribute to the resilience of the populations supported. The first round of impact evaluations selected for this window were designed to estimate the effects of integrated packages of resilience activities on households' capacity to absorb shocks (absorptive capacity), adapt to increasing environmental or economic stressors (adaptive capacity), and improve well-being in the long term (transformative capacity).

5. The impact evaluation for South Sudan is intended to estimate the impacts of the UNICEF-WFP joint resilience programme on absorptive, adaptive, and transformative resilience capacities. The joint resilience programme in South Sudan comprises a range of activities, including health and nutrition interventions, education programming, and Food Assistance for Assets (FFA) activities. The intended direct outcomes of the intervention are increased household resilience capacities, food security, nutrition, health, and education.

6. This Inception Report outlines the planned strategy for assessing the impact of the UNICEF-WFP resilience programme in South Sudan on the dynamics of population well-being and resilience. Through this impact evaluation, UNICEF, WFP, and DIME are working together to complement other ongoing efforts and to guide future investments and activities related to resilience in South Sudan.

---

[1] WFP. 2021. Global Report on Food Crises: Joint analysis for better decisions

[2] WFP. 2021, WFP South Sudan Situation Report #294

[3] WFP. 2021 WFP South Sudan Country Brief, July 2021, In 2021 alone, more than 115,000 people have been affected by flooding in South Sudan.

[4] WFP. 2021 WFP South Sudan Country Brief, July 2021

[5] UNICEF South Sudan. 2021. All together to prevent child malnutrition

7. This Inception Report also builds on a pre-analysis plan (PAP) that was registered with the American Economic Association's registry for randomized controlled trials. The pre-analysis plan includes detailed information on primary outcomes, research design - which includes details of randomized controlled trials and heterogeneity analysis, the randomization method, the randomization unit, clustering, sample size (total number, number of clusters, and units per treatment arm), and the 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.

2. Evaluation context

BACKGROUND AND CONTEXT

8. South Sudan became an independent country in 2011 following decades of war and conflicts. Independence was followed by a return to widespread armed conflict and insecurity, which has resulted in poor economic growth, displacement, and worrying development trends.

9. In 2021 alone, nearly 70 percent of the population required some form of humanitarian assistance with more than 100,000 people facing "catastrophic" (IPC Phase 5) levels of food insecurity.[6] In the same year, UNICEF estimated that 313,000 South Sudanese children under five years of age would be affected by severe acute malnutrition.[7] Nearly one in ten children die before the age of five due to health-related complications, and only 44 percent have received the necessary immunization.[8]

10. In 2017, famine was declared in two counties of the country's Unity State. South Sudan's population is highly exposed to climatic shocks, such as floods and drought. Food production has declined since the start of the conflict in 2014, hitting its lowest level in 2017: it has since increased slightly (but adjustment to population growth reveals a consistent reduction in production per capita). In most parts of South Sudan, households have the potential to produce surplus agricultural commodities but have challenges due to weak physical access to inputs and markets, high prices for agriculture inputs, inadequate structures to mitigate climatic shocks, and poor payment terms. In many cases insecurity has prevented farmers from accessing lands during planting and harvesting. The conflict has further constrained the private sector market and, with fewer traders in the market, farmers' terms of trade have further eroded. Floods also washed away much of the crop in 2019 and 2020.

11. The COVID-19 pandemic has also led to unique challenges to nutrition service delivery, as access to sites for treatment was limited because of the risk of COVID-19 infections, and the anthropometric measurement (weight and height) used to diagnose child malnutrition was suspended. The pandemic required the suspension of preventive activities such as the Vitamin A supplementation campaign and mass mid-upper arm circumference (MUAC) screening, which have since been initiated again.

12. South Sudan has received substantial humanitarian assistance over the years. In 2020 alone, the South Sudan humanitarian response totalled US$1.2 billion.[9] A range of interventions has been implemented in response to shocks and seasonal food insecurity, including cash or food transfers during the lean agricultural season, and other forms of health and nutrition support. The collaboration between UNICEF and WFP in South Sudan brings together UNICEF's expertise in the education; child protection; water, sanitation and hygiene (WASH); health; and nutrition sectors; along with WFP's expertise in addressing acute and persevering food security and nutrition needs, as well as its logistical reach and expertise in building community assets and livelihoods.

[6] WFP. 2021. WFP South Sudan Country Brief, July 2021

[7] UNICEF. 2021. Combating malnutrition in South Sudan, one child at a time

[8] UNICEF. 2021. Health in South Sudan: Briefing note

[9] OCHA. 2021. 2020 South Sudan Humanitarian Response in Review

13. Consistent with these broader efforts at national level, a key focus of WFP's Interim Country Strategic Plan in South Sudan (2018-2020 and extended to 2022)[10] is to implement integrated resilience activities to protect livelihoods and foster sustainable development in the long term. WFP has established a resilience programme that layers a set of integrated interventions on Food Assistance for Assets (FFA) interventions. The main objective of WFP's resilience programme is to strengthen the socio-economic resilience of smallholder farmers and vulnerable populations. The programme is intended to build the resilience of food systems and livelihood of targeted communities, while also strengthening community structures to support social cohesion and thus contribute to conflict prevention and prospects for peace. WFP has been working in Sudan since 1963 and upgraded its regional office in Juba to a Country Office after South Sudan's independence in 2011. WFP is actively involved in the food security sector and has presence throughout the country with 15 field offices located in all ten states and with hard-to-reach areas covered by the Rapid Response Missions – the widest footprint of any humanitarian agency in South Sudan. With the goals of saving lives, reducing food insecurity, stabilizing malnutrition rates, and helping to restore and enhance the livelihoods of vulnerable and shock-affected populations, WFP has projects throughout the country, with an expert logistics team as well as an early warning and food security monitoring network. WFP's food assistance activities support the objectives and expected outputs of the Food Security and Livelihood Cluster (FSLC), Nutrition Cluster, Education Cluster, and multi-sector refugee response. WFP co-leads the FSLC with the United Nations Food and Agriculture Organization (FAO) and actively supports cluster leads in the nutrition and education clusters.

14. UNICEF's Country Programme document (2019 – 2021 and extended to 2022) sets out its vision for "enhanced and more equitable outcomes achieved for the children of South Sudan."[11] To this end, UNICEF has been at the forefront since South Sudan's independence in 2011 (and as part of Sudan since 1989) in providing a multisector response – including the education, health, nutrition, child protection, and WASH sectors – to the multiple humanitarian crises in South Sudan. UNICEF's 13 field offices located in the ten states enable wide coverage and quality programming. UNICEF is working towards building a protective environment in which children's rights are respected and, to the extent possible, opportunities are created for children to develop their potential. In stable areas, UNICEF works with partners to implement programmes for longer-term recovery and resilience among affected communities. UNICEF works to strengthen cross-sectoral, integrated responses at the national and subnational levels, while seeking local solutions and community engagement using innovative approaches to access communities requiring humanitarian responses. Efforts will be undertaken to ensure that the sectors mainstream protection and that the "do no harm" principle is fully respected. UNICEF's activities support the objectives and expected outputs of the Nutrition Cluster, Education Cluster, WASH Cluster, Health Cluster, and Child Protection Sub-Cluster. UNICEF co-leads the nutrition, education, and WASH clusters and the Child Protection Area of Responsibility.

15. Given this context and policy environment, UNICEF, WFP, and DIME, with support from BMZ, are collaborating to build evidence on how multiple interventions can be combined or sequenced to boost the resilience of poor and vulnerable households in South Sudan. This evidence agenda is being implemented as part of the UNICEF-WFP joint resilience programme in South Sudan.

PROGRAMME DESCRIPTION

16. The UNICEF's current Country Programme Document (2019-2021)[12] and WFP's South Sudan Interim Country Strategic Plan 2018–2022, intend to directly contribute to the goals outlined in South Sudan's National Development Strategy and the United Nations Cooperation Framework (2019-2021) in South Sudan. UNICEF and WFP's Programme of Cooperation in South Sudan contributes to the four agreed outcome areas of the Cooperation Framework: building peace and good governance; strengthening food security and recovering livelihoods; strengthening social services; and empowering women and young people.

[10] WFP. 2017. South Sudan Interim Country Strategic Plan (2018–2022)

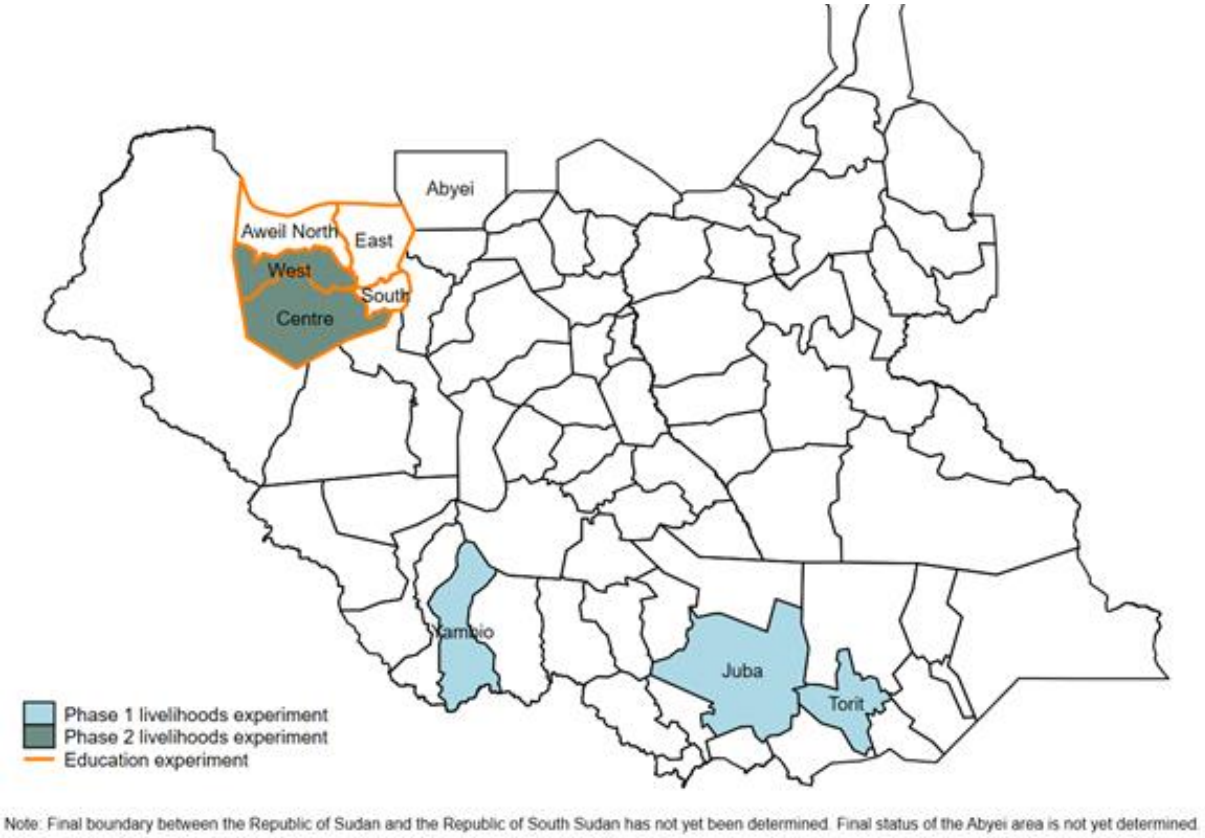
[11] UNICEF. 2018. South Sudan Country Programme Document (2019-2022), extended to 2022 [Link]

[12] UNICEF has submitted the next South Sudan Country Programme Document to the executive board for the September 2022 session [Link]

17. The UNICEF-WFP Joint Programme is aligned with the objectives set out in the Programme of Cooperation and includes a range of interventions that support communities to absorb shocks, adapt to risks, and transform livelihoods to move out of poverty. It includes a package of interventions that fall under three broad categories: i) livelihood activities – including Food Assistance for Assets (FFA) or Food Assistance for Training activities; (ii) access to education – including child protection, school feeding, and other school-based support mechanisms; and (iii) health and nutrition – including water, sanitation, and health (WASH) activities and nutrition interventions. The livelihood interventions are intended to catalyse growth, while the schooling and health facilities ensure access to basic services. The planning of these interventions is supported through the community-based participatory planning (CBPP) process.

18. The programme commitment covers four years (2020-2023) and focuses on urban and peri-urban communities in South Sudan. The specific geographic areas supported through the programme was identified based on the work of the broader Partnership for Recovery and Resilience (PfRR) initiative in South Sudan, launched in 2018. Initial PfRR efforts are focused on seven geographic areas in South Sudan that represent the diversity of ethnicities, livelihoods, political groups, and institutions found in and characterizing South Sudan: Yambio, Torit, Aweil, Wau, Rumbek, Bor and Yei. The partnership comprises peacebuilding, humanitarian, and development partners committed to working together to reduce vulnerability and build the resilience of citizens, communities, and institutions. The PfRR recognizes that some locations in South Sudan are conducive to resilience-focused programming. The UNICEF-WFP Joint Programme focuses on a subset of the "Candidate Partnership Areas" identified by the PfRR. Within the programme areas, the impact evaluation will focus on communities supported through the programme in Juba, Torit, Yambio, and Aweil counties.[13]

Figure 1: Counties covered by the impact evaluation



Note: Final boundary between the Republic of Sudan and the Republic of South Sudan has not yet been determined. Final status of the Abyei area is not yet determined.

[13] Aweil counties include Aweil West, Aweil East, Aweil North, Aweil South, and Aweil Centre.

19. The joint resilience programme in South Sudan includes the following components:

- *The livelihood component* uses villages as entry points for interventions. It is intended to meet the immediate food needs of households while restoring degraded landscapes, improving water harvesting, reducing the risk of environmental disasters, and creating community assets to secure ecosystems. The central activity within the component, Food Assistance for Assets (FFA), is carried out with a focus on increasing agricultural and pastoral productivity and yields, supporting economic development, and strengthening social ties between community members and villages. Assets are selected using a community-wide participatory process[14] and are built through FFA activities. Some of the assets are built jointly by a set of villages. Villages that may benefit from the assets are identified within the sites. Typical examples of related assets include land clearing and planting, vegetable gardening, community access road construction, pond or dyke construction, and tree seedling production. Beneficiary households within an asset site are targeted using a household-targeting exercise. Households are categorized into four socio-economic groups: very poor, poor, medium to well-off, and better-off. Households in the very poor and poor categories are eligible to participate in FFA activities. They are paid approximately $30 a month to work on the assets, typically between April and September, with several exceptions that are context-specific (for example, the construction of solar driers can take place in October, November, or even December).
- *The education* component uses schools as entry points for interventions. These activities are intended to create safe, healthy, and productive learning spaces for children. The programme specifically includes three broad activities: (i) ensuring food security and nutrition through school feeding; (ii) educational activities focusing on technical, psychosocial, and vocational capacity building; and (iii) behavioural interventions to improve the well-being and resilience of children and youth. The programme had the objective of identifying 85 schools to be supported through activities under the education component.
- *The health and nutrition component* uses health facilities as entry points for interventions. The component includes: (i) activities for improving health facilities; (ii) awareness activities to support households and caregivers to adopt best practices in child feeding, childcare, hygiene and sanitation, and health and nutrition; (iii) increasing household access to clean water resources; and (iv) improving access to immunization and priority health services. The programme had the objective of identifying 50 health facilities through which health and nutrition activities would be implemented.

2021–2023 Top-up funding and implications for extending the evaluation

20. In 2021, the programme was extended for another year (to 2023) and provided with top-up funds for programme expansion. Due to the relatively small number of villages in the first phase of the livelihoods RCT, this expansion provided an opportunity to ensure sufficient statistical power to detect impacts (detailed below). Discussions were held with the programme teams and donor agencies to scale up the impact evaluation into new areas.

Following these discussions, 51 villages in Aweil West and Aweil Centre were identified as eligible for participating in the livelihoods programme and RCT, using the same design as that in Juba, Yambio, and

[14] To select sites, WFP, government technical services, local authorities, and the community draw up an inventory of the potential and constraints for natural resource development based on a geographical approach and community-based participatory planning (CBPP). Once sites have been identified, a group of surrounding villages participates in the CBPP process to identify which priority interventions, including the livelihood assets, the communities will build. This provides a platform for inclusive community engagement, in which the most vulnerable, marginalized, and disempowered have a voice in community decisions. Participatory planning facilitates agreements for access to land and water resources for women's groups, youth, refugees, internally displaced persons, returnees, and the very poor.

Torit. To the extent that health, nutrition, and WASH project components are also being implemented in these areas, the scaling up of the evaluation can also generate evidence for the evaluation questions around the layering of these programmes.

THEORY OF CHANGE & HYPOTHESES

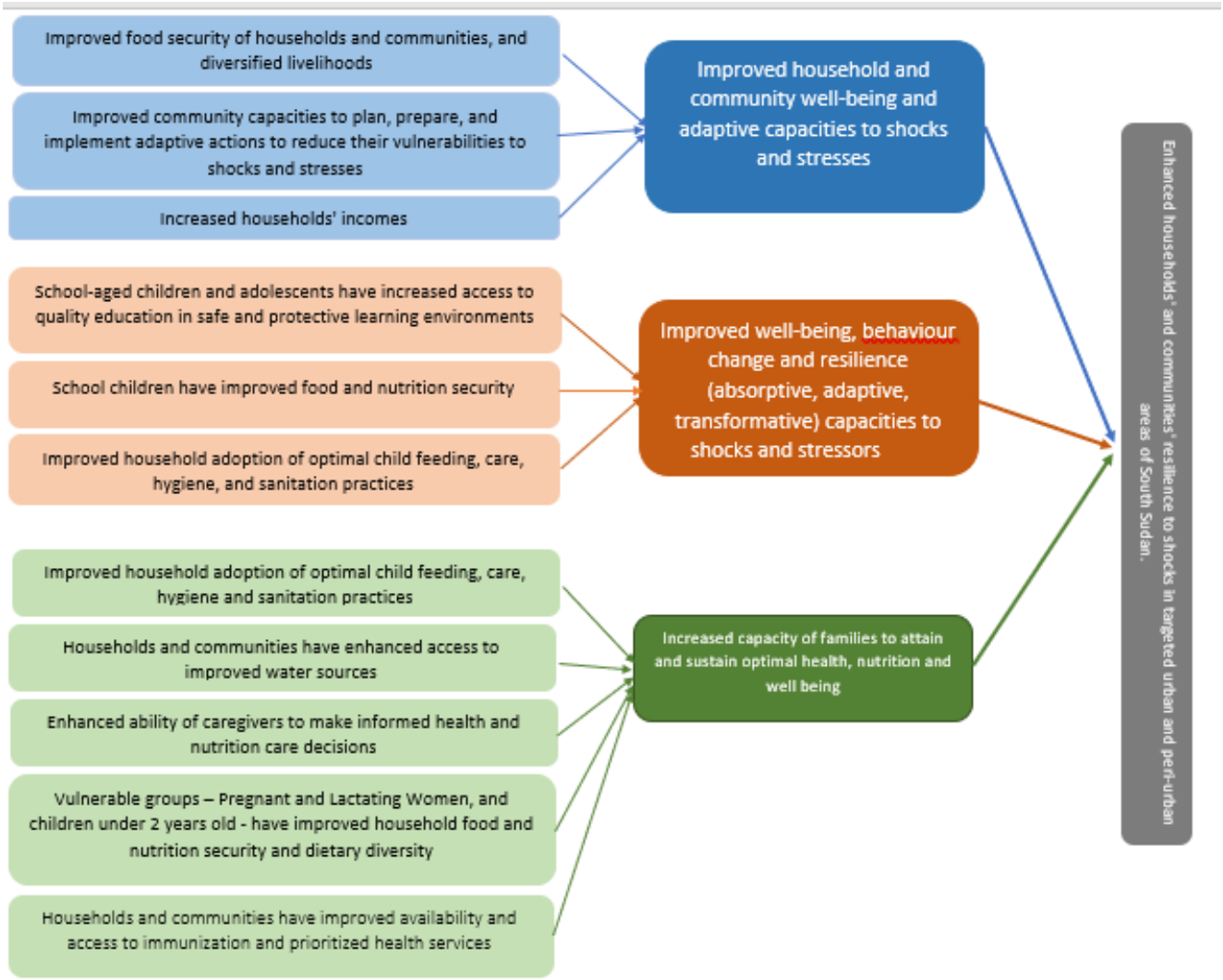
21. The theory of change for the joint-programme assumes that supporting communities through multiple activities focusing on various outcomes will: 1) support people to ensure their short-term well-being, and 2) enhance people's capacity to maintain and improve well-being while facing shocks and stressors. Figure 2 summarizes the key outcomes the programme focuses on. This is a simplified version of the full programme Theory of change.

Hypotheses

22. The resilience impact evaluation in South Sudan aims to test the following hypotheses:

    - Hypothesis 1: The programme will support people to maintain their food security by meeting a household's immediate needs through food or cash transfers. The effects of activities focused on meeting immediate needs are reflected mainly in:
        - Household-level food consumption
        - Household-level health and nutrition outcomes
        - The coping strategies of households
    - Hypothesis 2: The resilience programme will support households experiencing multiple and/or recurring shocks and stressors by improving capacities that are associated with maintaining and/or improving food security and health/nutrition. These capacities ultimately drive the effect on household consumption, nutrition and health detailed above. They include, but are not limited to:
        - Livelihood activities (how income-generating opportunities and labour market opportunities change)
        - Time use (how households allocate their time across productive activities)
        - Educational outcomes (how educational needs within the household are met)
        - Household assets (how assets are accumulated)
        - Financial outcomes (how loans, savings, and expenses fluctuate)
        - Variations in food consumption over time

Figure 1: Programme theory of change



3. Evaluation Approach and Questions

23. Impact evaluations aim to 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. UNICEF's and 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.

24. Based on regional discussions, in-country consultations, and subsequent conversations with the programme, and monitoring and evaluation (M&E) teams, a list of priority evaluation questions was developed addressing different programme components.

25. Through an extensive feasibility assessment, the teams agreed on an approach that is suitable for South Sudan's context and UNICEF's and WFP's programme implementation plans. The questions, and agreed methodological approaches are summarized in Table 1 below.

26. The main focus of the impact evaluation will be on measuring impacts on education, health, and food security, as well as related changes in well-being associated with a household's resilience capacities. These indicators are detailed further in section 5 (subsection 'OUTCOME MEASURED'), and include:
    - Health outcomes
    - Educational outcomes
    - Consumption and food security outcomes
    - Nutrition outcomes
    - Financial outcomes
    - Asset and livelihood outcomes

27. 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 bi-monthly surveys:
    - Food consumption
    - Coping strategies
    - Shocks and stressors experienced

28. The impact evaluation uses a mixed-methods evaluation design, using quantitative and qualitative data withtwo complementary quantitative components:

    - *Cluster Randomised Controlled Trial (RCT)*: The RCT aims to provide causal evidence on the relationship between programme activities and resilience capacities, as well as wellbeing outcomes (see Section 4). The RCT uses randomised livelihoods and asset creation activities, cash-transfers, and education interventions to construct credible counterfactuals and identify the impact of the programme on resilience outcomes.
    - *Heterogeneity analysis*: Not all the resilience programme activities are suitable for randomised assessment. Questions related to health and nutrition interventions will therefore be examined using heterogeneity analysis. Heterogeneity analysis allows us to examine sub-groups within the treatment communities to understand the effect of the activities on each sub-group. For example, comparison of health outcomes between households at different distances from health facilities will be carried out using quasi-experimental analysis. Relevant measures (such as distance from health facilities) will be collected during the baseline and programme monitoring to conduct the necessary analysis.

29. *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, qualitative data will be used to generate additional insights about the patterns observed in the quantitative data in particular to understand 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 s, and interviews with select village leaders. As described in Table 1, the questions and design of the impact evaluation are aligned with the implementation modalities and timelines of the joint resilience programme.

30. All 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 Food Assistance for Asset activities and school-based activities began. The endline data collection will occur after at least one year of implementation of activities in the treatment groups.

31. 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 will households require support the most to maintain or improve their food security.

32. 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 wellbeing.

Table 1: Overview of Evaluation Questions and Methods

Questions answered through a Randomized Control Trial (RCT)	Details of evaluation methodology
Do integrated education programming, and school feeding programming lead to better resilience outcomes when implemented jointly?	This question is answered by comparing households and children in schools receiving the school feeding programme with those receiving school feeding and integrated education programming.
What is the added value of livelihood activities for households beyond the impact of cash and food transfers alone?	This question is answered by comparing households who participate in the FFA programme with those who receive unconditional cash or food support, and those who are not supported by the programme.
Can de-linking the timing of FFA cash and food transfers and asset building activities improve the benefit for communities?	This question is examined by comparing villages with flexible FFA asset creation timing (working when demand for labour is low) with villages that link labour to transfer timings.
If communities have the opportunity to update lists of targeted beneficiaries, how do newly selected households compare to those selected through the current targeting process?	This question examines how the timing of targeting will determine which households are vulnerable. Reserve programme participants are identified during the initial targeting for FFA activities. After the first year, the community is asked to confirm if these reserve households still meet the vulnerability criteria to participate in the programme.
Questions answered through Heterogeneity Analysis	**Details of evaluation methodology**
To what extent are households able to cope with shocks over time, with and without livelihood interventions?	High frequency surveys (every two months) capture shocks that happen during the impact evaluation. The evaluation will examine how household characteristics and their participation in different activities correlates with their ability to cope with shocks.
How do multi-dimensional interventions build absorptive, adaptive, and transformative resilience capabilities at the community and household levels?	The high frequency data allows us to examine whether household food security improves after encountering a shock (absorptive capacity); whether households are able to maintain acceptable levels of food security over time (adaptive capacity); and whether household food security is improving over time (transformative capacity).
How did the education intervention work in conjunction with other interventions (health, nutrition, WASH, psychosocial support (PSS), school feeding (SF), communication for development (C4D) to achieve retention and safe learning outcomes of children?	The impact evaluation will collect data of household's participation in multiple programme interventions. By examining wellbeing outcomes and resilience capacities of households that receive different packages of interventions (e.g., Education vs. education & WASH vs. Education, wash and C4D) the evaluation will assess how different intervention packages contribute to overall household resilience.
How does distance from health, WASH and nutrition facilities influence take up and use of the assistance provided?	The surveys will collect data on the distance from the households to various basic services facilities such as health, nutrition or WASH centres. This information is important to understand the impact of the distribution of basic services within the community on household resilience.

4. Evaluation Methodology

33. The evaluation is taking place between 2021 and 2023. In 2021, 41 villages and 30 schools were included in the impact evaluation. Later, the evaluation leveraged the 2022 expansion of activities to include an additional 51 villages in the design.

34. As outlined in Section 3, the impact evaluation design relies on two quantitative components, cluster RCTs and heterogeneity analysis, which are complemented by qualitative analysis.

CLUSTER RANDOMISED CONTROLLED TRIAL (RCT) DESIGNS:

35. In a cluster RCT design, communities are randomly assigned to one of the comparison groups. The approach is depicted in **Figure 3**. The RCT analysis will compare treatment and control[15] groups to estimate the credible and unbiased treatment effects of the resilience package.

36. In South Sudan, communities (e.g., villages) are an important entry point for programme targeting and implementation. Therefore, to identify the causal impact of the resilience programme on different comparison groups, the impact evaluation utilizes a cluster randomized control trial (RCT) that allows for village-level randomisation.

37. The cluster RCT utilizes two primary programme entry points as units of randomization: 1) villages (for livelihoods interventions) and 2) schools (for education interventions). UNICEF and WFP preselected villages and schools that meet the eligibility criteria for multiple activities under the joint resilience programme.

38. The eligibility of these villages and schools was determined by the resilience programme's selection criteria, which includes a vulnerability assessment and a technical assessment. This ensures that the villages in the different comparison groups are all similar in core characteristics.

39. After the UNICEF and WFP South Sudan Country Offices identify eligible villages and schools for the resilience programme using standard targeting criteria, equally eligible villages and schools are randomly assigned to comparison groups. As the sample size is sufficient, the randomization eliminates any systematic differences between the treatment and control group and thus creates a valid counterfactual. The experimental designs for livelihoods (RCT Design 1) and education (RCT Design 2) are summarized below.

Design 1: RCT Examining the impact of livelihood activities:

40. Two complementary experimental designs (1.1 and 1.2) examine the impact of different livelihood activities and their implementation modalities. In addition, the impact evaluation includes an experimental pilot (design 1.3) of a labour intervention. These experiments closely follow the designs specified in the pre-analysis plan.

<u>1.1 Examining the impact of asset creation activities beyond the direct impact of cash transfers:</u>

41. This design is intended to generate understanding of the impact of assets created in the livelihood component of the programme through FFA activities, beyond the direct cash transfers provided to households in return for working on the assets. To examine this, 76 villages are randomly assigned to the following three groups:

    - Group A: Control group – villages that will not be supported through FFA activities in 2021 [24 villages].
    - Group B: Unconditional cash transfer (UCT) group – villages that will receive cash transfers in 2021 but will not implement any asset creation activities [23 villages].

[15] Control groups are used to construct credible counterfactuals, estimating what would have happened in the absence of the intervention, thereby establishing what outcomes would not be present.

- Group C: FFA group – villages that will implement regular FFA activities including asset creation and cash transfers [29 villages].

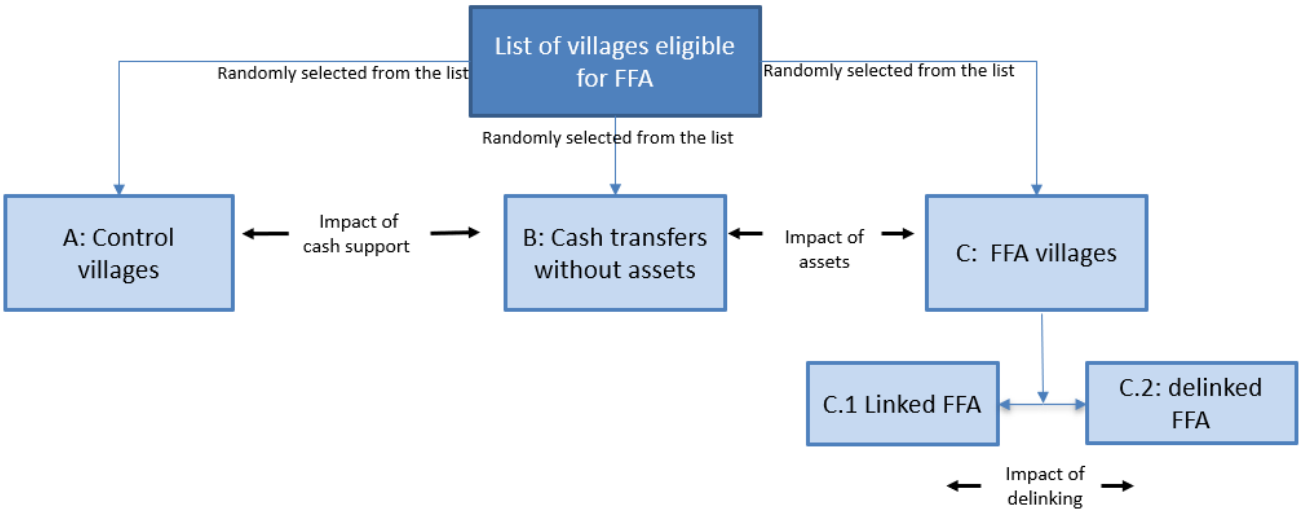
42. Comparing the FFA group to the control villages will enable us to isolate the benefits of the livelihood programme on resilience. Meanwhile, comparing the UCT and FFA groups will help us understand whether the positive returns from the asset outweigh the costs of having to invest in additional labour. In doing so, we can isolate the benefit of assets on their own.

1.2 Examining the impact of flexible asset creation timing:

43. This design will examine the impact of introducing more flexible timing for participants to work on the assets, on their well-being and resilience. The design hypothesizes that the benefits of FFA activities can be improved if the timing of the asset creation activities is adjusted to accommodate seasonality and shocks. The timing of activities is especially relevant in agricultural economies, where households might be in most need of assistance when they need to put in the most effort on the farms. During the pre-harvest season, households have less disposable income and food, and less time to devote to non-farm activities. In the post-harvest season, households have additional income from selling their crops and fewer demands on their time. It follows that cash and food transfers should be provided in the pre-harvest season when the marginal utility of consumption is high, and work requirements should be reserved for the post-harvest season when the returns to alternative labour allocations are low. To test this hypothesis, the FFA villages (Group C in design 1.1) will be further divided into two groups:

    - Group C.1: Linked-FFA Group: villages where households are invited to work on the asset while they receive cash payments or food transfers [8 villages].
    - Group C.2: De-linked FFA Group: villages where cash or food transfers are provided when the need is high (pre-harvest season), and assets are created when the agricultural work requirement is low (post-harvest season) [8 villages].[16]

Figure 3: RCT Designs 1.1 and 1.2



1.3 Labour experiment (pilot):

44. In addition to the designs listed above, we will pilot a labour experiment in which community members are offered jobs supporting data collection (guiding enumerators around the village, making appointments with households) for two days in the next month. The offered jobs do not require any

[16] Following discussions with programme teams, it was decided that the de-linking experiment would be conducted with a sample of villages that were already participating in the FFA programme, rather than newly eligible villages.

specific skillsets and hence are equally open to everyone. High-frequency surveys will include a question asking if members of the household would be willing to accept this job for two days in the next month, and the minimum wage they would accept for taking such a job. Before the next high-frequency survey, two community members will be randomly selected to be offered this job for a wage of US$5. This exercise enables us to measure the opportunity cost of labour throughout the year. Moreover, we can identify the extent to which benefiting from an FFA/UCT safety net affects an individual's willingness to work on other/additional jobs. An individual's willingness may be measured by his or her decision to take up or refuse the enumerator assistant job and/or by the size of his or her preferred wage rate. Thus, we can see how the presence of FFA programmes affects the local labour market.

Design 2: RCT Examining the impact of integrated education and school feeding activities:

45. This design examines the impact of the UNICEF package of interventions and investigates the benefits that may result from having both UNICEF and WFP interventions in treated schools. Among the resilience programmes covered under the Climate and Resilience Window, the education intervention is unique to South Sudan and therefore not pre-specified in the pre-analysis plan. Given the central role of access to education and school feeding in the South Sudan resilience programme, it is nevertheless important to understand these programmes' contribution to household resilience.

46. In 2021, it was planned to expand the education component of the UNICEF-WFP Joint Programme to additional schools in Aweil counties. The impact evaluation has been able to leverage this opportunity and has developed a design that includes 30 schools. The main WFP interventions include school meals, take-home rations, deworming, and the construction of kitchens where necessary. UNICEF is also implementing a package of interventions which include, among others, the building and rehabilitation of classrooms, the building and rehabilitation of WASH facilities in schools, the provision of teaching and learning materials, awareness raising and information campaigns to promote the importance of education, and capacity building to strengthen child protection.

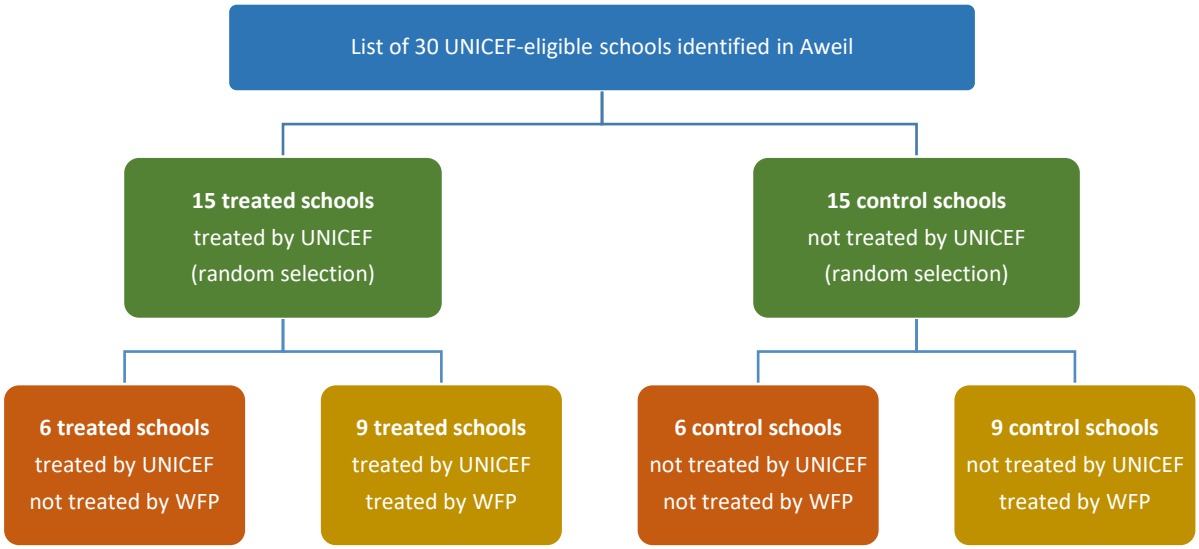
47. The integrated UNICEF education programme was randomly assigned to eligible schools – some of which had been selected to receive WFP school feeding support. As the WFP school targeting was already completed before the impact evaluation was designed, it was not possible experimentally to vary school feeding.

48. The 30 eligible schools were randomly assigned to two groups:

    - Treatment group (15 schools where UNICEF implements its package of interventions).
    - Control group (15 schools where UNICEF is not present).

49. The above two groups are stratified to take into account the presence of WFP interventions in schools. Consequently, within both the treatment and control groups presented above, there are schools where WFP interventions are present (nine of the schools in each group) and where they are absent (the remaining six schools in each group). The resulting comparison groups are presented in Figure 4. This design will enable us to compare schools where UNICEF is present with schools where there are no such interventions. Moreover, this research design also enables us to estimate the effects of the joint presence of UNICEF and WFP interventions relative to schools where UNICEF is present independently or where neither agency implements any activities.

Figure 4: RCT Design 2



50. To estimate the impact of school feeding, treatment schools will be matched with control schools using observable characteristics (for example, WFP criteria employed to target schools).

51. This approach, under which there is partial overlap between UNICEF and WFP interventions, serves to (i) reach as many schools as resources allow for implementation of interventions, and (ii) balance the call for increased UNICEF-WFP integration with the need to understand the gains of individual versus integrated programming by each agency.

HETEROGENEITY ANALYSIS

52. As it is not feasible or practical to randomise all the resilience programme activities individually in South Sudan, the impact evaluation uses heterogeneity analysis to understand how the comprehensive programme correlates with targeted outcomes.

53. The heterogeneity analysis harnesses the cluster RCT designs as a basis for identifying treatment and comparison communities. Heterogeneity analysis allows us to compare health, food security and other wellbeing outcomes of household groups with different characteristics or varying levels of access to services. The impact evaluation employs heterogeneity analysis to examine the following interactions and outcome areas.

Impact of joint programming:

54. Within the impact evaluation villages, heterogeneity analysis will be used to examine household groups receiving different intervention packages (e.g., FFA + Education vs. FFA + Health vs. FFA + Health + Education). To enable this, the impact evaluation team will map the types of interventions received by each household within the impact evaluation sample, through a combination of self-reported surveys and programme monitoring tools. This will allow us to understand 1) the jointness of the programme and how different activities of the programme overlap within a village; and 2) changes in outcomes of household sub-groups receiving different packages of support.

Overlaps and benefits of education and supplementary feeding programmes

55. Supplementary feeding programmes and education interventions, which primarily focuses on children, has an important role within the joint-resilience programme. The impact evaluation examines the value added by these components by examining child-specific outcomes within eligible sub-groups of households who do and do not receive the support. Since the nutrition and education support target

children of specific age group, the child-specific outcomes (e.g., nutrition, enrolment, health etc.) are analysed for specific age groups.

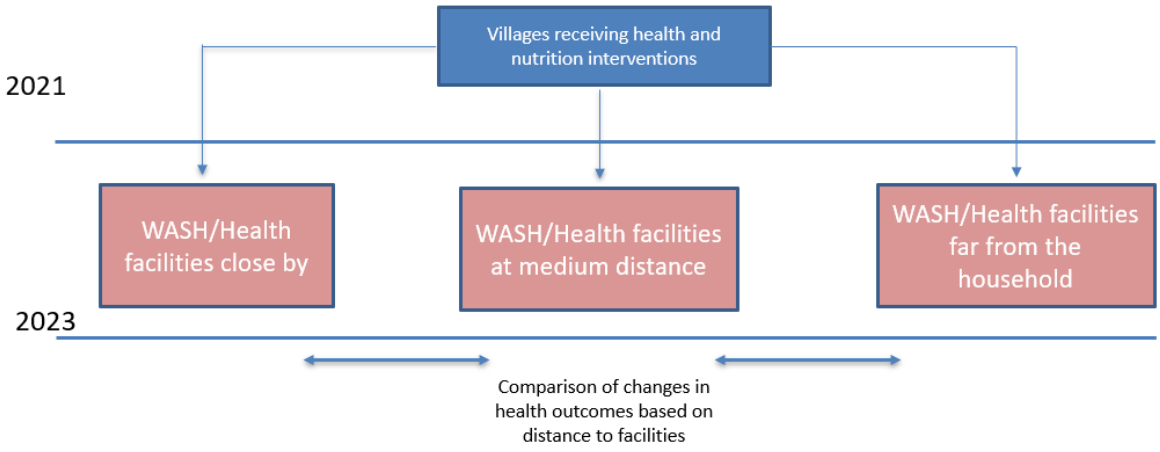
Variations in access to services

56. Some interventions under the joint-resilience programme, such as education, health, nutrition, sanitation activities etc are implemented at the community level (i.e., through public health centres, schools, or WASH facilities). Therefore, access to these facilities could influence how effectively households can benefit from these services. The heterogeneity analysis will examine how household access to services and their wellbeing outcomes vary within the impact evaluation villages. Outcomes among sub-groups of households with varying access to health and nutrition services will be analysed. Relevant measures (such as distance from health facilities, nutrition centres etc.) will be collected during the baseline and programme monitoring to conduct the necessary analysis.

57. The heterogeneity analysis will examine how improved access to water and sanitation facilities contributes to better health outcomes. The analysis compares households belonging to the following three groups.
    - Group I: Households with Schools, WASH & health facilities a short distance from the house.
    - Group II: Households with Schools, WASH & health facilities a medium distance from the house.
    - Group III: Households with Schools, WASH & health facilities a long distance from the house.

58. The distance to facilities from the household will be collected at the baseline. Short, medium, and longer distances will be defined during the analysis based on WHO guidelines on access to water and sanitation. Data will be collected during the baseline and endline surveys from all the households in the sample.

Figure 5: Design for heterogeneity analysis of variation in access to health facilities



QUALITATIVE ANALYSIS

59. 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?
    - [COVID & context permitting] How did intended beneficiaries supported by the programme experience participation in selected interventions?

60. 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.

61. 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.

62. 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 will use 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

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

DATA COLLECTED FOR THE RCTS

64. The impact evaluation employs baseline, endline and high-frequency surveys (every two months) to collect data on key outcomes relevant to the joint-programme. 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.

65. 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.

66. The livelihoods RCT (design 1) 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 5). We present our sampling strategy for both exercises below.

Baseline and endline surveys

67. In 2021, the livelihoods RCT design included 25 villages with 38 households in each village. This was later expanded, with another 51 villages with 15 households in each village. Combined, the FFA treatment group will have 540 households in 24 villages; the UCT group will include 505 households in 23 villages; and the control group will include 595 households in 29 villages. Power calculations are presented for both Phase 1 and the combined sample.

68. We use the Food Consumption Score (FCS) as the main outcome for power calculations, as it is a primary outcome for the impact evaluation and will be measured in all surveys.

69. We apply the standard formula for the minimum detectable effect (MDE):

$$MDE = \sigma_e(z_{0.8} + z_{0.975}) + \sqrt{1 + \frac{\rho(m-1)}{NP(1-P)}}$$

70. Where σ_e is the standard deviation of the outcome, $z_{0.8} + z_{0.975} = 2.80$ is the sum of the two z-scores, ρ is the intra-cluster correlation, m is the number of observations per cluster, N is the number of observations, and P is the share of observations assigned to treatment. We set $\rho = 0.05$ for all calculations. We use a power of 0.8, a significance level of 0.05, and an intra-cluster correlation of 0.16 (taken from baseline data). Focusing on new livelihoods sites, the mean FCS was 39.2, with a standard deviation of 20.4. In the first phase with 25 clusters (villages), our MDE was a 25 percent increase in FCS using a single baseline and a single follow-up survey. Following the scaling up of the RCT to 76 clusters, our MDE using the parameters from the baseline is a 15 percent increase in FCS, a much more achievable target.

High-frequency Measures

71. High-frequency data collection is a relatively new approach in South Sudan, and, therefore, previous datasets that would inform power calculations are not available in the national context. Therefore, we use a dataset from another humanitarian context in Sub-Saharan Africa to assess the size of the sample required, and how frequently data needs to be collected to detect reasonably small changes in outcomes such as Food Consumption Score.[17] The data used are unique because they include three common food security indicators: Household Hunger Scale, Food Consumption Score, and Household Dietary Diversity Score. We take this as our starting point and assess the role of survey frequency on power to compute changes in these measures over time.

72. Specifically, we used data collected in Madagascar, on 601 households (HHs) from 32 communities surveyed every month over 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.[18]

73. We model, through simulations, a hypothetical experiment that assigns half of the 32 communities to treatment. All the households in the treated communities experience one of three treatment effects:[19]

    - Increases by X percent in the mean of high-frequency measures against the control mean, holding other parameters constant.
    - Decreases in the standard deviation 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).

[17] 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 South Sudan, 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.

[18] Outcomes such as the 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.

[19] These treatment effects are estimated through the regression specifications in Section 6.1. of the Climate and Resilience Pre-analysis Plan. This includes specifications in which the mean and variance outcomes are interacted with time effects to further disaggregate dynamics between level shifts in intra-annual average outcomes and variance of these outcomes.

74. These simulations enable us to estimate power for detecting effects of programmes that may make households less food insecure on average but will 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 needed to estimate the effect. The goal of these simulations is to give guidance for how frequently countries need to collect food security data to identify impacts on food consumption scores and other measures. Power calculation tables for 32 clusters and 600 households are provided in Annex 5.

75. Table 9 in Annex 5 present the results of the power calculations needed to detect a 15 percent effect size for each of the three outcome measures with varying frequencies of data. For a 15 percent effect size, the power gains from increasing frequency from two-monthly to monthly are small, but the power losses in going from quarterly to semi-annual schedules are large. We therefore focus on comparisons with the two-monthly and quarterly schedules and compare effect sizes needed to obtain 80 percent power to guide the decisions on whether to plan for quarterly or two-monthly data collection.

76. Based on the results of these high-frequency data power calculations, we decided on a two-monthly frequency, with a sample size of 38 households per cluster. The number of clusters was established from the programme scale. For new livelihoods sites (design 1A), this was initially set at 25 sites and split between three treatment groups: 9 FFA, 8 UCT, and 8 control. The schedule and sample size for the first year of high-frequency surveys in design 1A (Phase 1) is shown below. The schedule of high frequency surveys for Phase 2 of this RCT is not shown but would follow a similar trend with the additional 51 villages.

Table 2: High-frequency data collection schedule (new livelihoods sites)

Design 1A	October	December	February	April	June	August
Sample	950 HHs	950 HHs	950 HHs	950 HHs	950 HHs	950 HHs

77. For design 1B (impact of flexible asset creation timing), 16 villages were initially part of the RCT: 8 with de-linked FFA and 8 with linked FFA. The timing and sample size for the first year of high frequency surveys in this design are as follows:

Table 3: High-frequency data collection schedule (flexible asset creation timing)

Design 1B	October	December	February	April	June	August
Sample	608 HHs	608 HHs	608 HHs	608 HHs	608 HHs	608 HHs

DATA USED FOR HETEROGENEITY ANALYSIS

78. 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 wellbeing among households with varying characteristics, and receiving varying packages of types of programme support. Additionally, 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.

79. Additional information required for creating sub-groups of households (e.g., distance to public health centres, 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 such as distance to the

public-health centres, cannot be pre-determined. 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.

OUTCOMES MEASURED

80. 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.

81. 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.

82. 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 measures are calculated from a minimum set of indicators collected at higher frequencies in each country supported.

83. A wider range of likely outcomes is considered when answering the main evaluation questions. Annex 2 summarizes and briefly defines the key outcomes of interest for the impact evaluation in South Sudan.

84. The indicators were selected in collaboration with UNICEF's and WFP's South Sudan country offices 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.

85. The primary set of outcomes are education, health, and food security indicators. For example, educational access and attendance, uptake of health behaviours, 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.

86. The outcomes are measured during the baseline data collection, high-frequency surveys (bi-monthly surveys following the baseline), and at the endline (after 24 months 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.

---

[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.

SURVEY IMPLEMENTATION

87. Data for all the research arms is being collected through baseline, high-frequency (HF), and endline surveys. The surveys will be identical in structure for all impact evaluations in the Climate and Resilience Window and will only be adapted to reflect the different country contexts:

    - Baseline and endline surveys: Estimated to take 2.5 hours, on average, to administer for the median household.
    - HF surveys: Estimated to last 35 minutes on average, for the median household. The HF survey will be implemented every two months for two years.
    - Surveys for school administrators (specific to design 2): Estimated to last 30 minutes.

88. The data collected at baseline is important for informing 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 (that is, the health, education, food consumption, food and nutrition security outcomes) are balanced, and thus to ensure 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.

89. The baseline surveys were conducted at the start of the intervention (June–September 2021). These are being followed by six rounds of short bi-monthly surveys (see Table 2 & Table 3). As the programme studied in South Sudan is multi-year, an endline survey will occur after two years of the intervention and will be aligned to the programme cycle.

90. While the baseline survey itself is relatively standard across all evaluations in the Window, it was piloted prior to data collection with local communities in South Sudan to ensure the questions are relevant to the context. In addition, specific modules, such as the food consumption module and asset lists, are adapted to the context of South Sudan.

91. The baseline survey took place after the households eligible for the programme were identified by the cooperating partners, using criteria that were set out in the programme implementation strategy. However, because the FFA/UCT implementation schedule was tight, cash and food transfers began before deployment of the baseline survey. To mitigate the risk that the impacts of FFA/UCT activities may have materialized shortly after implementation, a short pre-baseline survey was implemented to establish the pre-transfer levels of key outcomes that may have been affected by the cash transfers in the short term.

92. We will implement the endline survey after the second year of the programme 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 (please see Tables 2 and 3). Endline surveys will take place at the same time of year as baseline to avoid seasonality biasing results.

93. A key feature of the high-frequency measurements is the capturing of intra-annual dynamics of well-being through high-frequency surveys. This strategy enables us to get a better understanding of the resilience impacts of these interventions, not only by exploring the static difference between beneficiaries and non-beneficiaries, 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. This is an important measurement strategy for understanding how the joint programme contributes to resilience capacities in South Sudan.

MANAGEMENT OF DATA QUALITY

94. Multiple steps are being taken to ensure the high quality of data collected through the impact evaluation in South Sudan. The factors to consider and measures taken at each stage of the data collection are summarized below.

<u>Questionnaire development</u>

95. Data are collected using multi-module household surveys covering a range of outcomes. The planned endline survey will be identical to the baseline survey in structure and format. The high-frequency data collection also follows 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 outcomes. The data collection instruments have been piloted extensively in South Sudan to ensure context-specific details and option-sets for each question are appropriately identified.

Tracking participants over multiple survey rounds

96. The high-frequency and endline surveys involve revisiting as many of the baseline households as possible to create a panel, allowing us to control for differences in initial levels of key outcome indicators. Tracking information on these households will be collected to allow for the possibility of revisiting some, or all, of the households following the second round of the survey. Collecting identifiable data is necessary to verify the identity of respondents, to merge data across survey rounds, and to locate respondents for subsequent survey rounds. However, participants may choose to skip these questions if they're not comfortable answering them. 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 locating study participants
    - GPS coordinates of respondents' households

    Baseline and detailed follow-up surveys

    - Names, addresses, and phone numbers of study participants

Enumerator management and training

97. In parallel to the development of the questionnaire, a suitable third-party monitoring agency has been identified for carrying out data collection on the ground. The criteria for selecting these agencies in South Sudan include: i) prior experience with UNICEF, WFP, or DIME in collecting high-quality survey data; ii) experience in South Sudan; and iii) capacity for delivering multiple household surveys in the country over the study period. While the third party is responsible for hiring the enumerators and managing them in South Sudan, all data collection activities are supported by the DIME and OEV impact evaluation team. Detailed protocols have been developed to guide the data collection. These were developed by the impact evaluation team, which has also led the enumerator training.

98. The enumerator training includes classroom and field training. Enumerators have been selected based on their performance during the training. The training has been divided into three stages and has taken approximately one week to complete, including the following:

    - Reviewing the survey's content: The team guides the 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 questionnaire 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 asks enumerators to pair up and conduct "mock surveys" where they administer the 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.
    - Reviewing best practices: Once the mock surveys are complete, the team comes together to discuss best practices for engaging with respondents and recording their answers into the software. This includes a review of:
        - How to record survey responses.
        - How to provide alternative phrasing so respondents understand the question.

Confidentiality of data

99. As 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 encrypted servers protected by passwords so that individual enumerators do not have access to the data. The data will be de-identified for analysis. All analysis will therefore be carried out on de-identified data. Only the research team will have the key to link anonymized data to individually identifiable information; the Principal Investigators (PIs) will consequently be responsible for ensuring the security of this key. No individual-level results will be reported, and all the results will be aggregated to protect the identities of individual study subjects.

Data quality protocols

100. The data are collected electronically, using a Computer Assisted Personal Interviewing (CAPI) platform; CAPI surveys reduce the logical inconsistencies in the questionnaire. This also enables us to programme consistency checks into the survey and perform quality checks daily.

High-frequency checks look 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

101. Any anomalies that we detect through this process are flagged to the data collection team immediately. In addition, the team also performs 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 allows us to provide immediate feedback to the field teams in case of divergences or other problems.

Internal team coordination

102. All aspects of the evaluation process will be overseen directly by the PIs involved in the project. This includes coordination with programme counterparts, data collection, and analysis. All data collection instruments will be extensively piloted to account for the country and/or regional context. Analysis will be performed by the project Research Assistant, through close supervision from the PIs . The analysis will be set up so that it can be done in a replicable and reproducible manner.

IMPLEMENTATION MONITORING SYSTEM

103. UNICEF, WFP, and DIME are working together to monitor whether beneficiaries receive the scheduled WFP and UNICEF programming on time. The agencies regularly track when transfers are made to programme recipients, when activities are implemented, as well as whether work requirements are being met. DIME is complementing these efforts by ensuring that the programme variations we introduce are properly followed. More specifically, DIME is 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 periodically cross-checks with field teams to ensure that the initial randomization plans are being adhered to. Any deviation is recorded and systematically documented to be considered during the analysis stage.
- The evaluation team also monitors any new activities which may be introduced into the treatment or control communities, and, where possible, captures the impact of these activities through the measurement framework.

104. For the livelihoods component, the evaluation team checks whether the asset work that individuals report during data collection matches the data obtained from the attendance lists compiled by WFP in the field while the asset activities are being implemented.

6. Data processing and analysis

DATA CODING, ENTRY, AND EDITING

105. All survey data is collected via tablets. The data is stored on SurveyCTO servers. As soon as a surveyor marks a filled-out form as "finalized", the form's contents is encrypted. Whenever form data is transmitted via a 3G or other internet network, they are encrypted in transit using SSL as well. Finally, any data that is downloaded from the server are either encrypted or purged of any personal identifiers before analysis. A series of back-checks are performed on the data we collect. Any mistakes that are detected are recorded and changed.

106. Data is collected by enumerators recruited by a survey firm contracted by DIME with experience of collecting surveys in the local context and the relevant languages. The DIME field coordinator and research assistant make daily high-frequency checks on data quality, with regular reports to the impact evaluation teams and field teams. The high-frequency surveys are piloting methods for reaching respondents by phone and in person, to determine which method is more cost-effective for minimizing non-response.

PROGRAMME-SPECIFIC QUANTITATIVE DATA ANALYSIS

Sampling and specification

107. The sampling frame is the lists of project sites and households provided by the UNICEF and WFP South Sudan teams. The sample is households identified to receive benefits. Identification of recipients before implementation in all treatment arms ensures 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 by lasso to select controls for precision, and we control for baseline measures of our outcomes when they are available through an analysis of covariance (ANCOVA) specification. We cluster standard errors at the community level whenever the treatment of interest is assigned at the community level (this is the case for livelihoods and education interventions). In the event of non-random attrition, we report Lee bounds on primary impacts.

Descriptive targeting analysis

108. The study will document the profile of select beneficiaries along with a wide range of indicators collected at baseline. All outcomes of interest will be disaggregated by gender, age, and other relevant demographic characteristics, in consultation with the country offices. This will provide descriptive information to UNICEF and WFP about the efficiency of their targeting protocols, and their ability to identify the poorest households. The first rounds of high-frequency data collection includs questions about households' satisfaction with targeting to shed further light on the legitimacy of targeting among local populations.

Identifying the profile of the households that benefit most from the programme

109. Finally, after follow-up data is collected, the study will help to assess how programme targeting could be improved to select the households that benefit the most from the programme. This will be based on recent statistical methods that can identify the households that benefit the most from UNICEF's and WFP's assistance, before analysing whether their profiles correspond to the profiles of households selected by their respective communities.

Livelihoods (RCT Design 1 in Section 4 above)

110. To measure the impact of the FFA 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) is used for randomized treatment at community level (the community receives FFA or is assigned to the control group).

Livelihood components (RCT Design 1.1 in Section 4 above)

111. To determine the additional contribution of assets to resilience, we use a similar regression, comparing food security and other resilience outcomes in communities assigned to the UCT arm and communities assigned to the full FFA package. This difference will test if the productive asset confers additional benefits to households on top of the basic needs support alone.

112. Ideally, all complementary interventions would be cross randomized, allowing for the marginal impact of each programme to be evaluated separately. Unfortunately, due to constraints in programmatic implementation and statistical power, this type of design is infeasible. Therefore, the interpretation of the treatment effects will vary depending on how these additional programmes are included at each site. In cases where additional programmes are offered in both treatment and control locations, the UCT and FFA arms simply identify the additional effect of cash and food transfers on top of these base programmes.

Timing of cash and livelihood activities (RCT Design 1.2 in Section 4 above)

113. To test whether the impacts of FFA depend on the timing of the cash and work requirements, food security and other resilience measures will be regressed on assignment to the de-linked FFA, where the timing of asset-building activities is more flexible than in the standard FFA programme.

Shocks and targeting

114. After the first year, when selected communities had a chance to retarget households, we will compare impacts on resilience measures between newly targeted households and reserve households. This tells us the welfare consequences of changing the set of beneficiaries based on revealed shocks. By monitoring the outcomes in both the communities undertaking retargeting, and in the communities not undertaking retargeting, we can identify whether decisions to retarget are related to expected impacts from the programme: in particular, if communities retarget households that are more food insecure than those which would have been selected at the initial targeting stage. If this has occurred, it would suggest that communities prefer to select households for whom the programme inputs will lead to faster recovery of food security status.

Shocks

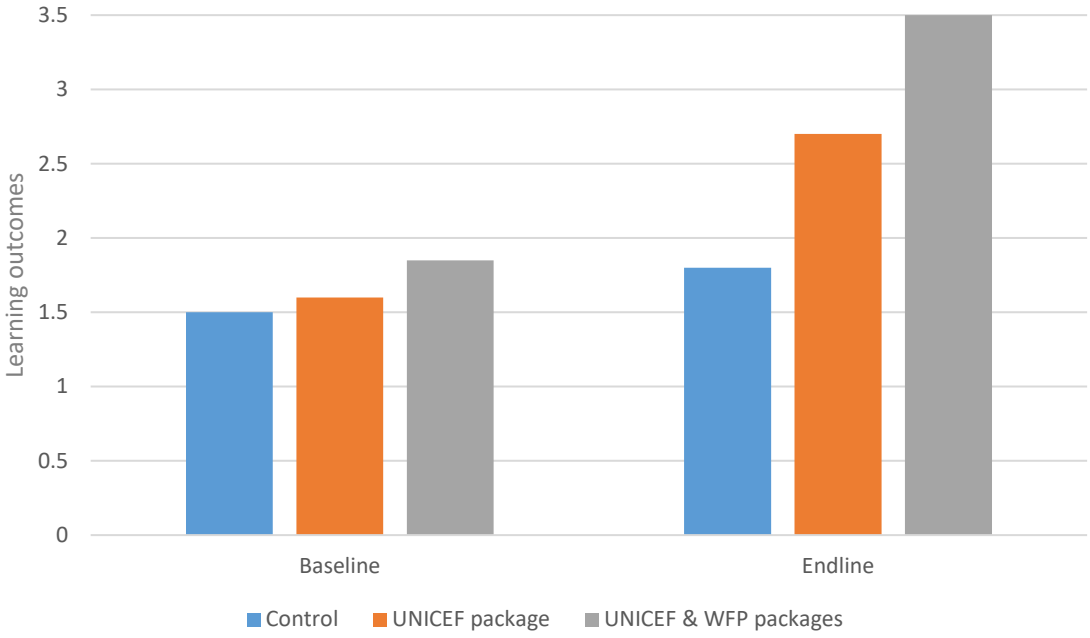
115. A feature underlying a household's resilience is their ability to anticipate, absorb, adapt, cope with, and recover from, shocks while improving their trajectory for well-being (i.e., transformative capacity). Many programmes are designed to help households mitigate the impacts of shocks, but evaluating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

117. To determine the differential impact of the programmes based on whether a household was exposed to a shock (from a pre-determined list of shocks measured in the surveys), we will estimate a regression interacting programme participation with a list of pre-specified context-specific shocks that will include both natural events (such as droughts – as defined by rainfall during main cultivation months falling below a defined threshold), conflict (as defined by a recorded conflict in standardized data such as the Armed Conflict Location & Event Data Project (ACLED), and economic shocks. The high-frequency data will enable us to estimate this regression for multiple points in the year while accounting for different types and severity of shocks.

Education (Design 2 in Section 4 above)

118. For the experimental design, our primary means of analysis is a simple regression of education outcomes on the UNICEF treatment status and its interaction with the WFP treatment package. The coefficients of interest will measure the impact of the UNICEF package and the integrated programme. An example of how this effect will be visualized is presented in the following chart.

Figure 6: Education analysis graph



Note: This graph uses hypothetical data to show how data will be analysed when they are collected. It should not be interpreted as a measured programme effect.

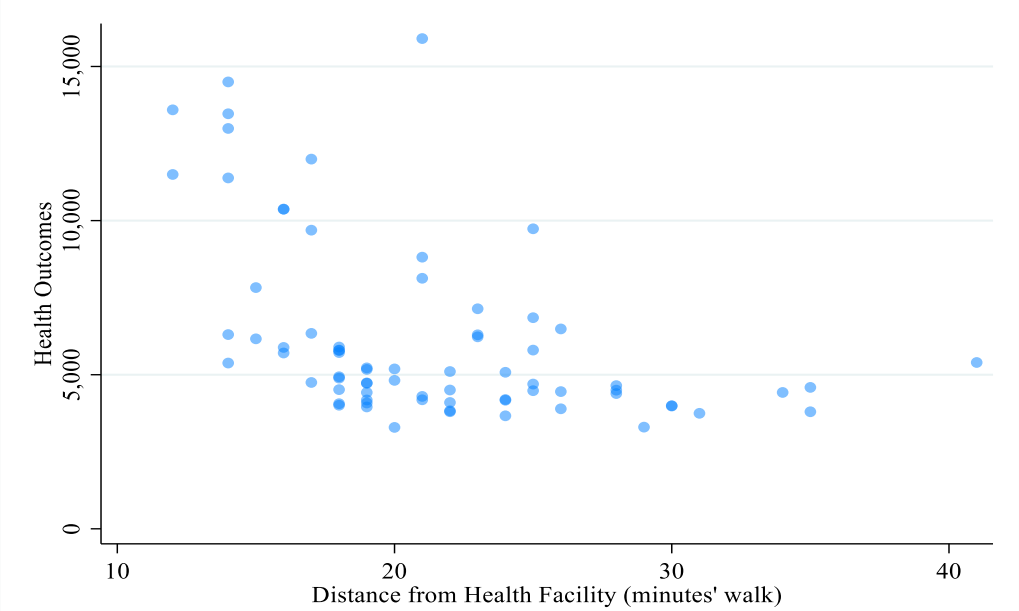
119. The same analysis strategy can be used for multiple outcomes, such as the number of days children were in school; well-being outcomes of children; and absorptive, adaptive, and transformative resilience capabilities. Demographic and programme indicators can also be incorporated into the analysis to enable more granular programme learning. For example, gender, age, ethnicity, and other relevant characteristics can be controlled for, allowing for a sub-group analysis. Access to specific programme components (WASH, PSS, SF, C4D) can help us to determine which parts of the integrated package correlated most with learning outcomes.

Health and nutrition

120. The impact of the health programmes will be analysed using detailed survey data (collected through baseline, endline, and high-frequency surveys) focusing on health, nutrition, and WASH outcome

indicators over time and distance. The health outcomes of households who live at different distances to health facilities will be measured to understand how access to services affect outcomes. As the location of health facilities is not random, this will measure the correlation of health access and outcomes rather than causal impact. However, by controlling for demographic and other relevant characteristics, this can provide useful information to programme teams. The following graph visualizes this relationship.

Figure 7: Health Analysis Graph



Note: This graph uses hypothetical data to show how data will be analysed when they are collected. It should not be interpreted as a measured programme effect.

121. In addition, nutritional data collected for children, women, and households will be analysed according to food-based dietary guidelines provided by UNICEF for various demographic groups, such as pregnant and lactating women, and children under five years of age.[24]

PROGRAMME-SPECIFIC QUALITATIVE DATA ANALYSIS

122. The impact evaluation is using semi-structured interviews with implementing partners to capture information about the process of programme implementation and the experience of programme participants. We are collecting qualitative information relating to the implementation process as described in Section 5. We are asking the beneficiaries if, in their view, the programme has had a positive or negative impact on outcomes. Within the survey tools, space is provided for specifying "other" responses if the responses listed do not adequately describe the respondent's answer. If a certain "other" response occurs with significant frequency across surveys, this will be coded and included in the analysis.

123. 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.

[24] UNICEF (2021) Food-based dietary guidelines. Link: https://www.unicef.org/media/102761/file/2021-Food-based-Dietary-Guidelines-final.pdf

7. Ethical considerations

124. A key goal of WFP's Impact Evaluation Strategy is to increase the use of rigorous evidence to inform programmes in countries where WFP works and globally. Guided by this overarching principle, the evaluation takes into account several ethical considerations, and puts in place relevant best practices.

IRB approval

125. The evaluation team has obtained international approval from an Institutional Review Board provided by Solutions IRB for the Climate and Resilience Impact Evaluation Window design, as well as for the specific design and measurement elements in South Sudan. The window approval was received on 12 November 2020 and the South Sudan amendment approval was received on 12 September 2021. In addition, the evaluation team has obtained approvals from local institutions in South Sudan. This ensures that the evaluation complies with local regulations and does not violate any laws.

Communication to the participants

126. 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 the purposes of research. To mitigate this risk, UNICEF, WFP, and DIME are working together to ensure transparent and clear communication to communities.

Informed consent

127. The team is ensuring 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 clear that refusal to respond to our survey does not come with any consequences for their participation in the joint 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, there are a few questions that may be directed to other members of the household, including women (such as questions on women's empowerment, food consumption for children aged from 6–23 months and so on). To avoid respondent discomfort during surveys, we will take several precautions to ensure that interviews take into account the respondents' privacy and comfort:

- Participants may skip any questions they do not wish to answer or withdraw from the survey at any time.
- Interviews will be conducted at participant's homes to increase the likelihood that they will be comfortable answering questions.
- Finally, all enumerators will go through 1-2 weeks of training, which will be followed by extensive piloting in the field. The goal of the training is to ensure that enumerators follow survey best practices in terms of protocols and ethics, but also that questions are asked in a uniform and contextually appropriate manner.

Confidentiality

128. The evaluation team ensures 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 about individuals will not be shared with anyone outside the evaluation team.

Transparency in evaluation designs

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

8. Risks and limitations

Risks to internal validity

130. The primary risk to an internally valid and causal estimate of programme components is the statistical power required for estimating multiple treatment arms. When sites and participants have already been selected, options to create counterfactuals are limited, and with the small number of planned expansion sites, not all possible treatment arms can be implemented simultaneously. To address this problem, we are using the most rigorous impact evaluation method available – a Randomized-Controlled Trial (RCT) – to address the questions that are feasible to answer using this. Other questions will be explored through comparisons of non-experimentally varied exposure groups, such as families with children who live further from or nearer to a health clinic or school. In addition, as with any in-field RCT, spillover across communities and differential attrition are potential risks for the evaluation. The team will work closely with the implementing partners on the ground to monitor potential spillover risks and design clear and direct implementation protocols.

Risks to external validity

131. One of the evaluation's limitations may be that the results of a single study may not be externally valid. We can provide an indication of the robustness of our findings in external contexts through the formal synthesis of findings from all the countries participating in the Climate and Resilience Window (currently ongoing impact evaluations in Mali, Niger, Rwanda, and South Sudan).

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

Risks due to instability

133. A further risk is that a crisis (conflict, political instability, or natural disaster) impedes programme progress or the ability of implementing teams to follow the planned evaluation design. To mitigate the potential consequences of unforeseen issues, the evaluation team will work with the implementing partners to proactively resolve potential delays ex-ante, including through supporting the planning and implementation of operational activities and the timely launch of procurement processes. Furthermore, field coordinators will work closely with DIME, WFP, and the implementing partners to ensure that programme activities take place 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

134. As a result of COVID-19, the country offices have had to implement all its programmes with third party NGOs who are now responsible for all field-related activities. This creates additional monitoring challenges as the evaluation team must ensure the NGOs are complying 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 offices and is in frequent communication with the country offices and the NGOs to monitor these dynamics.

135. 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 in that case 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 WFP in multiple countries.

Additional considerations

136. In South Sudan, the Joint Programme is working with a population that is somewhat more urban than in other countries in the Window. Therefore, estimated impacts from South Sudan may reflect impacts for more urban populations than the other countries.

9. Quality assurance and peer review

137. The WFP Impact Evaluation Quality Assurance System (IEQAS) sets out guidance on definitions, methods, processes, and procedures for ensuring that impact evaluation outputs provide robust and credible evidence about impact. The IEQAS comprises process guidance, quality checklists, templates, technical notes, and other reference material to guide evaluation teams and partners throughout the evaluation process. Quality assurance measures are being systematically applied throughout the evaluation phases. These include preparation and selection, design, data collection,[25] and consistency of programme implementation with the evaluation design, analysis, and reporting.

138. 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 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 will produce a window-level meta-analysis and peer reviewed journal articles. All reports and articles are reviewed by the Head of Impact Evaluation. The WFP Director of Evaluation finally approves all the reports before they are published.

139. In addition, all final evaluation reports are 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.

10. Communication plan

140. In South Sudan, the WFP integrated resilience programme is intended to address chronic food insecurity and to support communities in responding and adapting to climate shocks and conflicts. The evidence generated from the impact evaluation will inform future scale-up or expansion plans for the programme. The impact evaluation is also intended to provide insights on the most effective approaches for targeting the most vulnerable and providing support at the most effective times.

141. More broadly, the impact evaluation evidence will also contribute to the planning of UNICEF's and WFP's strategic planning in South Sudan by supporting the country offices to identify which activities or *combination of activities* have the greatest impact on resilience, and how this varies across regions.

142. Considering these objectives, the impact evaluation team developed a communication plan to ensure timely dissemination of the evidence, and to facilitate its use in programme design and delivery.

143. UNICEF, WFP, and DIME will ensure that the UNICEF and WFP regional bureaux and country offices are full partners in discussing and using the evidence created by the impact evaluation. More specifically, four complementary avenues are envisioned for dissemination:

- **Active engagement with programme teams:** This includes evaluation workshops and country-level engagements. On 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. A report will be produced by the evaluation team

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

to be shared with operational teams and policy makers in each country to summarize learning, solicit suggestions and improvements, and generate new uses for the resulting data.

- **Harnessing the global networks of UNICEF, WFP, and DIME:** 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 the workshops. In addition, we plan to make our findings broadly available to other teams within UNICEF, WFP and the World Bank, including resilience, nutrition, health and social protection teams, to emphasize the role of community targeting of public goods. UNICEF's and WFP's Offices of Evaluation and DIME host or participate in multiple workshops every year on the topic of using impact evaluations to improve learning. Each of these workshops will be an opportunity to share evaluation findings and lessons with other agriculture and rural development projects from the World Bank and outside the Bank.
- **Academic publications:** Papers designed for journal publication are planned. The impact evaluation team will engage broader academic communities to both contribute to and shape the knowledge generated by the impact evaluation. All the data collected as part of the set of evaluations will be made available online through the impact evaluation database, following the Bank's open data policy.

144. During the impact evaluation, field coordinators regularly update country teams on evaluation plans and keep track of any adjustments in field implementation plans to ensure that the evaluation plan remains aligned with field concerns. As data are collected, the impact evaluation team is responsible for analysis, which ensures a degree of independence, but the results of this analysis are regularly shared and discussed with the country and regional programme teams to ensure that findings can be used for programme decisions, and programme insights can be incorporated into the data analysis.

145. The evaluation analysis will be shared in the form of baseline and endline reports published by WFP. The results of the impact evaluation will feed into the broader cross-country analysis being undertaken as part of the partnership.

146. In addition, UNICEF, WFP, and DIME will communicate regularly with BMZ, the Government of South Sudan, and other partner agencies to provide them with updates on the impact evaluation work and results. This will take place through a series of in-country and virtual seminars (as allowed based on context).

147. Moreover, knowledge produced by the proposed impact evaluation activities will also be more broadly relevant to other actors and governments. Lessons drawn from these impact evaluation activities will also inform future policy implementation in other regions. UNICEF, WFP, and DIME will support the use of the results from these evaluations to inform the project design of other partners by ensuring easy access and promoting awareness for the evidence generated. Finally, the impact evaluation team will eventually draft academic papers for submission to a peer-reviewed journal, increasing the visibility of the evidence generated.

11. Organization of the evaluation

148. The impact evaluation will be delivered through a partnership between UNICEF, WFP and the World Bank's Development Impact Evaluation Department (DIME). DIME and WFP will deliver the impact evaluation through the existing Memorandum of Understanding between the Office of Evaluation of WFP and the World Bank, and in consultation with the UNICEF Evaluation Office Focal Point. Key governing and management structures within the partnership are outlined below.

EVALUATION TEAM

149. The evaluation team will consist of Principal Investigators, and focal points from the World Bank DIME, UNICEF and WFP. The composition of the team is summarized in table 4 below.

150. The responsibilities of the evaluation team include:

- Preparation of the impact evaluation concept note and work plan: DIME, WFP and UNICEF;
- Delivery of all activities set out in the impact evaluation workplan: DIME and WFP;
- Monitor and report the progress made in delivering the workplan to the evaluation Steering Committee. Prepare Annual Progress Reports: DIME, WFP and UNICEF.

TABLE 4: EVALUATION TEAM AND MAIN COUNTERPARTS

Name	Role	Organization/Unit
Paul Christian	Principal Investigator, Lead Researcher	DIME
Erin Kelley	Principal Investigator	DIME
Greg Lane	Principal Investigator	DIME
Marcus Holmlund	Research Manager	DIME
Jonas Heirman	Principal Investigator	WFP OEV
Hanna Paulose	Principal Investigator	WFP OEV
Eric Jospe	Research Analyst	DIME
Roxana Elena Manea	Field Coordinator	DIME
Julia Ashikbayeva	Research Analyst	DIME
Alwin Nijholt	Social Policy Advisor	UNICEF South Sudan
Wilson Kaikai	Programme Officer (Monitoring and Evaluation)	WFP South Sudan

EVALUATION COMMITTEE

151. The evaluation committee will include the evaluation field coordinator, and representatives of UNICEF and WFP Country Offices in South Sudan and the Regional Bureaux in Nairobi. 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 side-lines of the evaluation learning workshops. Key members of the evaluation committee are listed in Table 5 below.

Table 5: Evaluation committee

Name	Role	Organization/Unit
Roxana Elena Manea	Field Coordinator	DIME
Alwin Nijolt	Social Policy Advisor	UNICEF South Sudan
Wilson Kaikai	Programme Officer (M&E)	WFP South Sudan
Ernesto Gomez	Head of Programme	WFP South Sudan
Justus Kamwesingye	Regional Evaluation Officer	UNICEF Regional Office Nairobi

Nikki Zimmerman	Regional Evaluation Officer	WFP Regional Bureau Nairobi

WORK PLAN AND DELIVERABLES

Table 6: Milestones, Deliverables, and Estimated Timeline

MILESTONES	DELIVERABLES	TIMELINE
Agreement on the IE design	Methodology note	January 2021
Data collection plan and pilot	TORs Questionnaires	April–May 2021
Implementation of intervention aligned to evaluation	Rollout plan Monitoring reports verifying treatment and control status	Starting from April 2021
Data collection (baseline)	Cleaned data Dictionaries	June–September 2021
First data analysis	Presentation Data file Do files Baseline report	December 2021
High-frequency survey data collection plan	TORs Questionnaire Sampling plan	August–September 2021
Inception Report	Peer reviews & steering committee reviews Final inception report	June 2022
Baseline Report	Baseline analysis and report	December 2021 - June 2022
Data collection (high-frequency surveys)	Cleaned data dictionaries Datasets	Two-monthly November 2021 – November 2022
Follow-up data collection plan	TORs Questionnaire	Jan–Feb 2023 (to be aligned with programme timeline)
Data collection (follow up)	Cleaned data Dictionaries	September 2023 (to be aligned with programme timeline)
Final report and policy notes	Technical note Policy note Data file Do files	October 2023/aligned with programme timeline
Dissemination of findings	Presentations	December 2023–March 2024 (to be aligned with programme timeline)

Annexes

Annex 1: Window Summary

146. 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 the importance of 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.

147. 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."[26] The capacities are required before, during and after the onset of shocks and stressors 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 in the phase of shocks and stressors.

148. To strengthen resilience, WFP employs an integrated approach to programming, in which multiple forms of support are provided to the same community. These integrated packages of interventions are intended to improve food security and nutrition by smoothing and improving food consumption in the short term, while supporting livelihoods and addressing barriers to development (such as better climate information, access to markets, education, WASH, community ownership and leadership and so on) 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, and meet immediate food needs.

149. The Climate and Resilience Impact Evaluation Window will support resilience programme teams to design impact evaluations to understand how the integrated packages of interventions, and activities within the package, contribute to resilience. As of today, 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.

150. 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.

151. The Climate and Resilience Impact Evaluations contribute to the growing literature on the determinants of household resilience and the impact of integrated programming. Much of this evidence focuses on only a single welfare measure, and very few studies come from the most shock-prone and high-poverty countries. In particular, there is a lack of evidence on whether these interventions can help households cope with seasonality and shocks, and transition more permanently out of poverty. This window builds on existing measurement approaches for resilience and develops a high-frequency survey to capture variations in food security due to seasonal changes or other covariates (such as extreme weather events, conflicts, or economic downturns that affect larger populations simultaneously) and idiosyncratic shocks (such as death in the family, or loss of livestock

[26] FSIN Resilience Measurement Technical Working Group (2014) Resilience Measurement Principles: Toward an Agenda for Measurement Design. Rome, FAO & WFP.

that affect specific individuals or households).[27] This approach to measurement allows the evaluation to separately estimate the effects of livelihood programming on mean consumption over time, from impacts associated with improved capacity to smooth consumption across seasonality or shocks (that is, the standard deviation over time for a given household). Furthermore, this approach can be used to establish whether certain types of households can successfully transition out of poverty; and the times of the year in which they are most likely to do so. In addition to contributing to richer outcome measures for evaluating impacts, expanding the use of high-frequency data may also contribute to the use of the resilience measures for targeting purposes. For example, if high-frequency data enable the evaluations to separate the people, households, or communities who are food insecure only sometimes from those who are food insecure all the time, better targeting decisions could be made.

[27] Barrett et al. 2014, Cissé & Barrett 2018, Phadera et al. 2019

Annex 2: Main outcomes of interest

Table 7. Main outcomes of interest

Outcome Type	Outcome Name	Definition	Measurement Level
Primary	Consumption and food security*	Food Consumption Score-Nutrition Food Insecurity Experience Scale Food expenditure	Household/individual (head of household)
Primary	Educational access, attendance,* progress, and attainment	Grade and grade progression; educational expenditure; school attendance and dropout rates	Household/individual (head of household/children)
Primary	Knowledge and uptake of health, WASH, and health-seeking behaviours. Health and nutrition outcomes	Immunization coverage and dropout rate Dietary diversity for children Dietary diversity for women WASH & health-seeking behaviours Diarrhoea prevalence Child feeding practices	Household/individual/ (female head of household, children)
Primary	Shocks and coping mechanisms*	Households' main respondent asked which shocks (such as drought, flood, family death, asset loss, or job loss) the households have suffered over the previous 12 months and the severity of each shock. Reduced Coping Strategy Index Food Expenditure Share Livelihood Coping Strategy Index	Household/individual (head of household)
Secondary	Time use	Time spent on various activities at different points of the day by selected household members	Household/individual (head of household)
Secondary	Assets	Number of assets owned by the household and access to various public facilities/services for a contextually pre-defined list	Household (head of household)
Secondary	Income generating activities	Participation in non-farm business, agriculture and livestock, or wage employment and revenue from these activities	Household/individual (head of household)
Secondary	Financial outcomes	Current savings levels, number of loans taken and current outstanding debt, and cash transfers	Household/individual (head of household)
Secondary	Migration	Which household members have migrated	Household/individual (head of household)
Secondary	Psychosocial well-being	This measurement tool that captures incidents of depression	Household/individual (head of household)
Secondary	Women's empowerment	This measurement tool measures women's decision-making	Household/individual (female head of household)
Secondary	Social capital	Financial support index	Household/individual (head of household)
Secondary	Safety nets	Amount and source of transfers from other NGOs and government sources	Household/individual (head of household)
Secondary	Reservation wages*	Minimum hourly wage selected household members would accept to engage in short term labour and the length of time they would be willing to work.	Household/individual (head of household)

Outcome Type	Outcome Name	Definition	Measurement Level
***Note: High-frequency data collection enables these measures to be reported as the intra-annual mean, variance, and trends (both unconditional and conditional on shocks) to measure the impacts in terms of absorptive, adaptive, and transformative capacities (see Technical annex 3).**			

Annex 3: Defining resilience

152. This annex describes the ways in which we plan to conceptualize resilience through the measurement of dynamic outcomes such as food security, school attendance, and labour outcomes. The material in this appendix is closely adapted from the WFP climate and resilience pre-analysis plan.

DEFINING RESILIENCE THROUGH HIGH FREQUENCY MEASUREMENT

153. 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) Resilience Index Measurement and Analysis (RIMA) or the Technical Assistance to Non-Governmental Organizations (TANGO) resilience index, as well as examples of resilience evaluations that use characteristics such as diversification of livelihood strategies as a proxy for resilience.[28] The second is to regress outcomes on measures of shocks, in order to isolate the contribution of shocks to food security. The third is to use measurement of different households' food security at different times to impute a given household's food security path, and then to measure the parameters of the imputed distribution.[29]

154. 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,[30] 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:

$$y_{it} = \alpha_i + f_i(d) + \delta_i t + \epsilon_{it}$$

where y_{it} is a measure of wellbeing 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 wellbeing.[31] 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, α_i, 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, ϵ_{it} measures exposure to shocks not systematically correlated with survey dates. Figure TA1 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.

155. 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

[28] Macours, K., Patrick P., and Renos V. 2020. *Transfers, Diversification and* Household Risk Strategies: Can Productive Safety Nets Help Households Manage Climatic Variability? Working Paper.

[29] Cisse, J. D. and Barrett, C.B. 2018. "Estimating Development Resilience: A Conditional Moments-Based Approach." *Journal of Development Economics* 135, 272-284. And Christian, P. and Dillon, B. 2018. "Growing and Learning When Consumption Is Seasonal: Long-Term Evidence From Tanzania." *Demography* 55(3), 1091-1118.

[30] Ligon, E. and Schechter, L. 2003. Measuring Vulnerability. *The Economic Journal* 113 (486):C95-C102.

[31] The model is flexible and allows for the observation unit to be an individual, a household, or a village/community, with analysis for each main specification planned for the household level. Similarly, the length of the interval defined by the time t could be defined as daily, monthly, twice-yearly, yearly, etc., as relevant.

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.

156. However, the average food security of the household over the period (α_i) 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 further 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 the blue household experience a shock at the same point.

157. 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 are virtually impossible to collect and it may not be necessary to distinguish impacts arising from influencing different components of the well-being equation. We propose operationalizing resilience measurement by repeated sampling of the same household on different dates within a pre-defined 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

164. These impact evaluations will estimate welfare trajectories within a one-year period following the start of a programme. Figure 8 shows a hypothetical consumption path for a household in 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.

165. The first measure of the consumption equation we are concerned with is the household's intra-annual reference level of consumption -- this is α_i in the structural equation. If we observed a household's value of consumption every day, this would be measured as the household's average food security status over the period, as shown by m in Figure 9 Panel A. Next, we consider the household's intra-annual standard deviation, the average of the household's deviations from the reference mean (Figure 9 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 proportion 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 9 Panel C). Resilience is then defined as the ability of a household to avoid poverty over time, which we operationalize in the following way:

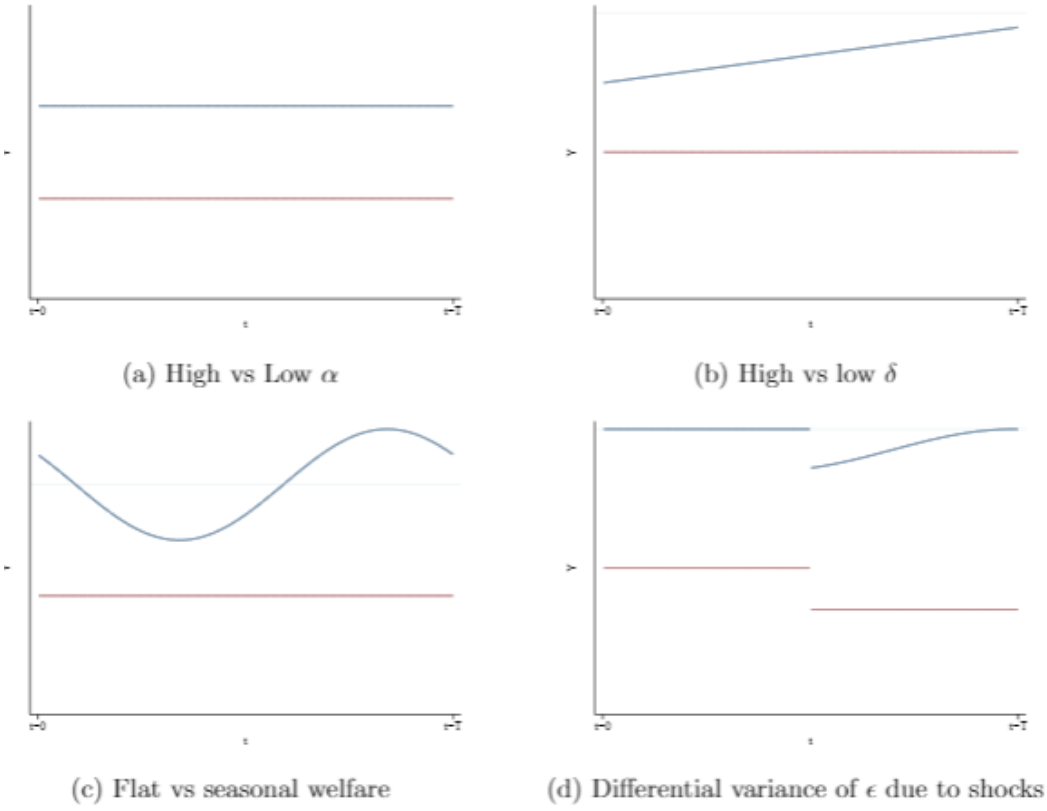
- A household with a higher m is on average higher above or less below the food security threshold. So, households with higher m are more resilient than households with lower m. The intra-annual reference mean of food security is measured by: $\widehat{m_i} = \frac{1}{n_i}\sum_{t=0}^{T} y_{it}$
- Conditional on m, having a higher standard deviation will increase (1) the likelihood of falling below a food security threshold, (2) the share of time spent below the poverty threshold, and/or (3) the number of days that are 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: $\hat{s} = \frac{1}{\sqrt{n_i}}\sqrt{\sum_{t=0}^{T}(y_{it} - m_i)^2}$
- Households that spend more time below the threshold are less resilient than households that spend less time above the line. The proportion of observations below a poverty line is measured by: $\widehat{share_i} = \frac{1}{n_i}\sum_{t=1}^{T}\backslash mathbbm1(y_{it} < \bar{y})$

  where n_i is the number of times community, household, or individual *i* is surveyed; T is the length of the period over which resilience is measured, y_{it} is a measure of household food security

status, and $\bar{y}$ 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 on these outcomes associated with interventions.

166. Figure 10 shows what the measures look like for a household with the hypothetical sinusoid function shown so far, assuming a quarterly data collection schedule in which food security status is observed at quarterly 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 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 proportion of the period spent below the poverty line is the number of observations that fall below the poverty line (the grey dashed line) divided that by the total number of observations (number of grey dots divided by number of blue dots).

Figure 8: Examples of capacities over time



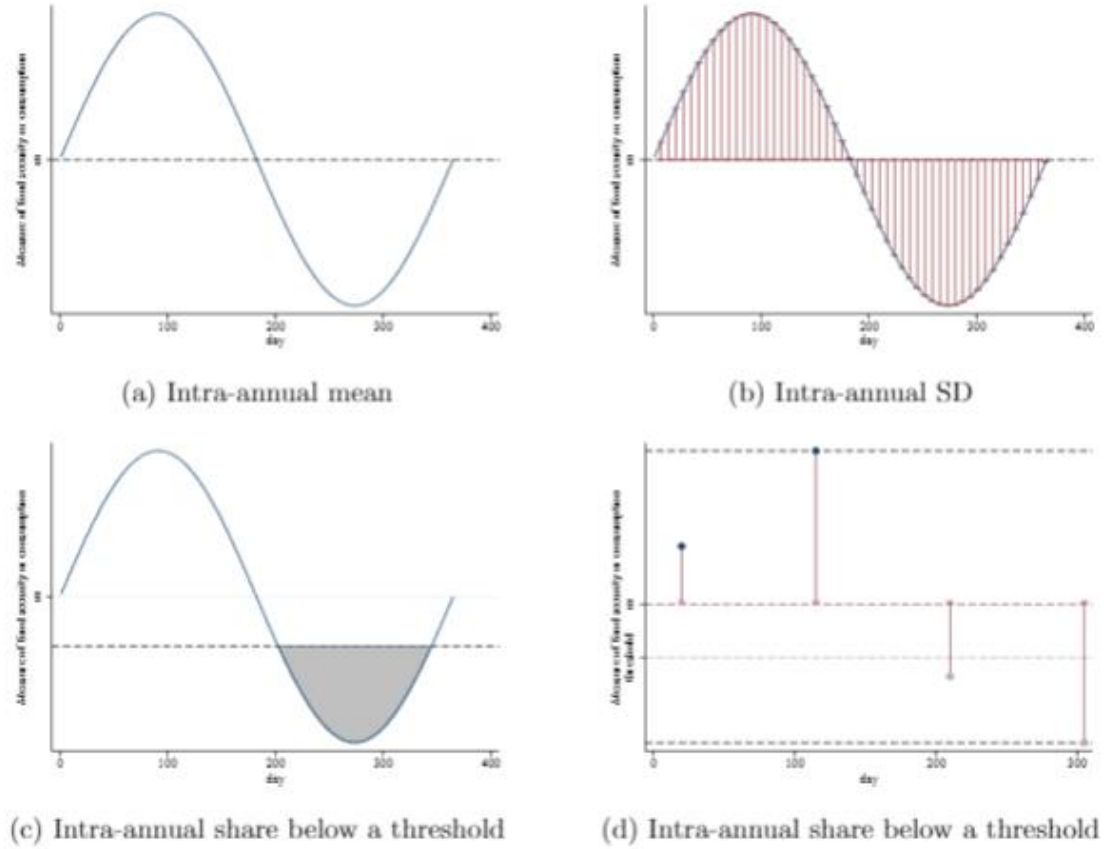
(a) High vs Low α

(b) High vs low δ

(c) Flat vs seasonal welfare

(d) Differential variance of ϵ due to shocks

Figure 9: Measures of capacities



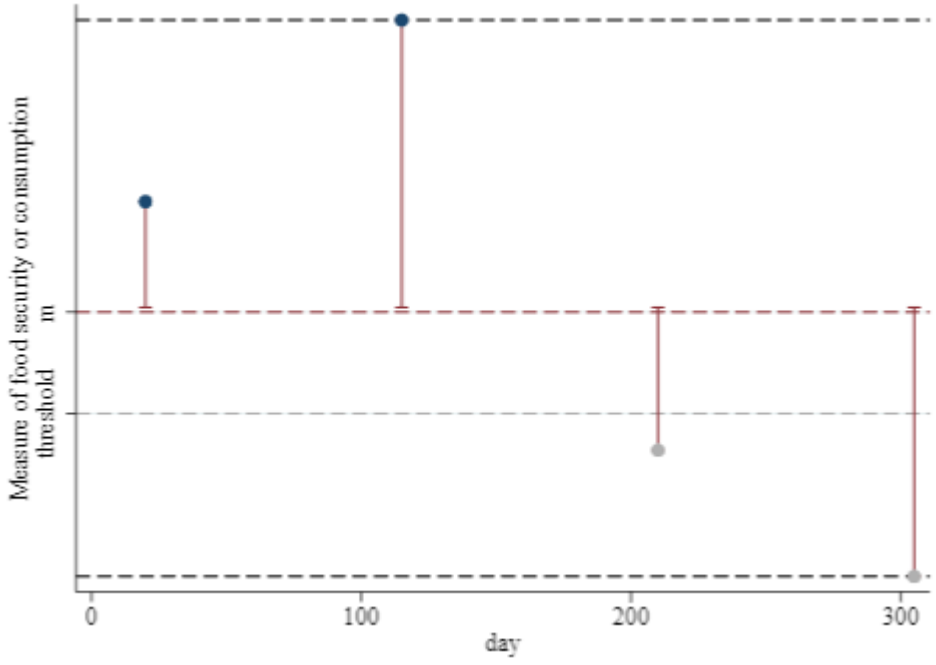
(a) Intra-annual mean

(b) Intra-annual SD

(c) Intra-annual share below a threshold

(d) Intra-annual share below a threshold

Figure 10: Feasible measurement of capacities



Annex 4: Village and household listing

Table 8: Number of households targeted and interviewed at baseline – across all treatment and control sites

Counties	# of households targeted	# of households interviewed
Juba	441	406*
Yambio	776	767
Torit	251	0*
Aweil	450	450
Total	1,918	1,623

Note: Due to security concerns, data collection was not possible in Torit and in one village in Juba.

Annex 5: Power calculations for high-frequency surveys

Table 9: Power calculations for 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 10: Power calculations for different effect sizes - two-monthly 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 6: Questionnaires

The baseline questionnaire for the resilience window is available here. The modules included in the survey are summarized in the table below.

Table 11: 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-ag business
F	Income Generating Activities - ag & livestock
G	Food Consumption Score (FCS)
H	Food Insecurity Experience Scale (FIES)
I	Consumption expenditure (food & non-food)
J	Asset Index + 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 7: Detailed stakeholder analysis

167. Stakeholders and users of this evaluation are defined as those actors that may influence the evaluation, and those that may be influenced by it. This includes internal, external and national actors and programme beneficiaries. The UNICEF and WFP country offices in South Sudan are intended to be the primary users of this evaluation. In addition, UNICEF and WFP South Sudan offices provide technical guidance at national level to inform national policy and dialogue on social protection, and the country offices have expressed interest in using the results of this evaluation to support this technical advisory capacity.

168. The various categories of stakeholders include:

- Internal South Sudan-based stakeholders: the Country Director/Representative and Deputy Director/Representative, the Head of Programme, and all technical and management personnel
- Internal stakeholders outside South Sudan: the Offices of Evaluation, the Regional Bureau for Eastern Africa (WFP) and the Eastern and Southern Africa Regional Office (UNICEF), and the Resilience and Social Protection teamss within UNICEF and 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 South Sudan
- National stakeholders, including national and subnational government actors, and non-governmental organizations.

169. 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 UNICEF and WFP assistance will also have a high stake in the results, and will be the primary providers of data for the evaluation.

170. Stakeholder engagement will vary depending on 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.

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

Table 12: 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 informants, 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)
UNICEF and WFP internal stakeholders					
UNICEF and WFP country offices	Main implementers of the programme under evaluation	To inform upcoming country strategic plan and relevant programming	The country offices are responsible for implementing the programme according to the evaluation design. They actively provide feedback on the tools and outputs of the evaluation.	From the scoping stage	High
UNICEF and WFP regional bureaux	Governance and technical advisory roles	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
UNICEF and WFP Offices of Evaluation	Coordination of impact evaluation and liaisons with country office	As coordinators of the impact evaluation and for 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	Beneficiaries will	Beneficiaries and non-	From the	High

	communities – including men, women, boys, and girls – will be the primary participants of the intervention	likely have strong interest in any changes in targeting, reach, or effectiveness of future programming as a result of the evaluation and recommendations. Women and girls have particular stakes in results meant to shed light on recommendations for improving gender equality	beneficiaries alike will be the primary sources of data on effectiveness	targeting and selection stage	
Government at local level	Sector and village-level government staff provide technical backstopping for livelihoods and agricultural programming	As local community members and technical experts, staff are interested in supporting an evaluation of the livelihood programme effectiveness	Local governance structures provide technical advice for programme design and are involved in beneficiary selection within communities	At the targeting phase of the intervention	Medium
Government at district level	District staff play key roles on the steering committee for programming in their jurisdiction and providing	District staff influence the prioritization of resources in their district; the evaluation results can help to inform	The evaluation relies on the activity programming coordination and targeting efforts of district staff	At the targeting phase of the intervention	Medium

	support for mobilization and targeting of beneficiary villages	their prioritization efforts in the future			
Government at central level: Ministry of Local Governments Ministry of Agriculture	National government structures provide ethical and administrative clearance for programming and evaluation efforts and oversee local development initiatives and national social protection programmes.	UNICEF and WFP have established relationships with the national government providing technical support on food security and nutrition; the evaluation results will support these efforts	The evaluation receives national-level clearance before inception	At the initial scoping for the intervention and dissemination of findings	Medium
Local/regional non-governmental organizations: Action Africa Help	As implementing partners for the programme being evaluated	Evaluation results can inform their own livelihood and gender transformation programming	As a cooperating partner, Action Africa Help is responsible for ensuring the programme is implemented in line with the evaluation design	At the initial scoping for the intervention and dissemination of findings	High

International non-governmental organizations: World Vision Plan International	World Vision and Plan International are implementing partners for the programme under evaluation	Evaluation results can inform their own livelihoods and gender transformation programming	As cooperating partners, World Vision and Plan International are responsible for ensuring the programme is implemented in line with the evaluation design	At the initial scoping for the intervention and dissemination of findings	High
World Bank	Development Impact Evaluation Unit	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
Donor Federal Ministry for Economic Cooperation and Development (BMZ) German Development Bank (Kreditanstalt für Wiederaufbau (KfW))	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 8: Detailed evaluation process

Table 13: Detailed evaluation process

Phase 1 – Preparation	Involved	Estimated Date
Initial discussion between country offices and the Office of Evaluation to assess the feasibility	CO/OEV	September 2019
Confirmation of the impact evaluation outline	CO/OEV	January 2020
Establishment of impact evaluation (IE) team and evaluation committee (EC)	OEV/DIME	January 2020
Agreement on the questions, design, implementation and timelines between country offices and impact evaluation team	DIME/OEV/CO	March 2020 - January 2021
Targeting potential intervention sites (including both potential intervention and comparison areas)	CO/DIME	February 2021
Phase 2 - Inception report		
Inception report drafted by impact evaluation team, submitted for quality assurance and revisions	DIME/OEV	June 2021 – March 2022
Publication of the inception report	OEV	August 2022
Dissemination of the inception report with country offices, regional bureaus, evaluation committee, window's reference group, steering committee, online/social media as appropriate	DIME/OEV	August 2022
Phase 3 – Baseline & High Frequency data collection		
Preparation of data collection tools, including survey questionnaire, digital devices, sampling strategy, training material, etc.	DIME	March 2021
Pilot and finalization of data collection tools	DIME/CO	April 2021
Recruitment enumerators/data collection firm	CO	May 2021
Training of enumerators	DIME/CO	May 2021
Data collection process and live monitoring data quality checks	DIME/OEV/CO	June-Sept 2021
Data collection through high-frequency surveys	DIME/OEV/CO	November 2021 – Nov 2022
Baseline data collection in the new impact evaluation sites	DIME	June 2022
Phase 4 – Baseline report		
Data analysis and baseline report drafted by impact evaluation team, and submitted for quality assurance and revisions	DIME/OEV	March 2022
Publication of baseline report	OEV	August 2022
Dissemination of baseline report to survey respondents, country offices, regional bureaux, evaluation committee (and other evaluation stakeholders), window's reference group, steering committee, online/social media as appropriate	DIME/OEV	June – September 2022
Phase 5 – Programme implementation		
Randomization	DIME	March 2021

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

Annex 9: References

Barrett, C. B. & Constas, M. A. 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.

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

Giri, R. *2021. Combating malnutrition in South Sudan, one child at a time*, United Nations Children's Fund, https://www.unicef.org/southsudan/stories/combating-malnutrition-south-sudan-one-child-time

Gunnsteinsson, S., Adhvaryu A., Christian P., Labrique A., Sugimoto J., Shamim A. A., and West Jr. K. P. 2019. Protecting Infants from Natural Disasters: The Case of Vitamin A Supplementation and a Tornado in Bangladesh. *NBER Working Papers*. 35

Macours, K., Premand P. & Vakis R. 2020. "Transfers, Diversification and Household Risk Strategies: Can Productive Safety Nets Help Households Manage Climatic Variability?" Working Paper.

Office for Coordination of Humanitarian Affairs. *2021. 2020 South Sudan Humanitarian Response in Review*, https://reliefweb.int/report/south-sudan/2020-south-sudan-humanitarian-response-review

Phadera, L., Michelson H., Winter-Nelson A. & Goldsmith P. 2019. Do Asset Transfers Build Household Resilience? *Journal of Development Economics*. 138:205-227.

Premand, P. & Stoeffler, Q. 2020. Do Cash Transfers Foster Resilience? Evidence from Rural Niger. *Policy Research Working Paper* No. 9473. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/34774 License: CC BY 3.0 IGO

United Nations Children's Fund. *2021. Health in South Sudan: Briefing note*, https://www.unicef.org/southsudan/media/7691/file/Health%20Briefing%20Note%202021%20Q2.pdf

United Nations Children's Fund South Sudan. *2021. All together to prevent child malnutrition,* https://www.unicef.org/southsudan/all-together-prevent-child-malnutrition

World Food Programme. 2017. South Sudan Interim Country Strategic Plan (2018–2022), extended November 2021, https://www.wfp.org/operations/ss01-south-sudan-interim-country-strategic-plan-2018-2021 Accessed 19 May 2022.

World Food Programme. *2021. Global Report on Food Crises: Joint analysis for better decisions*, https://docs.wfp.org/api/documents/WFP-0000127413/download/

World Food Programme. *2021, WFP South Sudan Situation Report #294*, https://api.godocs.wfp.org/api/documents/d9f25212a8d04385ba103b5871a6c9b1/download/

World Food Programme. *2021. WFP South Sudan Country Brief, July 2021*, https://docs.wfp.org/api/documents/WFP-0000131531/download/

Acronyms

CAPI	Computer-assisted personal interviewing
CSP	Country strategic plan
DIME	Development Impact Evaluation Unit (World Bank)
FFA	Food-for-assets
IE	Impact evaluation
IRB	Institutional review board
OEV	Office of Evaluation (World Food Programme)
PAP	Pre-analysis plan
PI	Principal Investigator
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