Impact Evaluation of Cash-Based Transfers on Food Security and Gender Equality in Rwanda

Inception Report

February 2022

Contents

Key Personnel for the Evaluation

OFFICE OF EVALUATION

Director of Evaluation, Andrea Cook

Evaluation Officer, Jonas Heirman

Evaluation Officer, Felipe Dunsch

Evaluation Officer, Hanna Paulose

IMPACT EVALUATION TEAM

Principal Investigator, Florence Kondylis

Principal Investigator, John Ashton Loeser

Principal Investigator, Dahyeon Jeong

Research Analyst, Tanay Balantrapu

Research Analyst, Eric Jospe

Research Analyst, Marc-Andrea Fiorina

Research Assistant, Julia Ashikbayeva

Field Coordinator, Guillaume Gatera

WFP RWANDA COUNTRY OFFICE

Strategic Outcome Manager, Tiina Honkanen

Programme Policy Officer, Seonghee Choi

Evaluation Manager, Sarah Cruz

Programme Policy Officer, Laurent Ulimubenshi

1. Introduction

1. Gender inequality and food insecurity are important issues in developing countries today. While social protection programmes directed at supporting both gender equality and food security outcomes are often implemented in different developing country contexts, the causality behind the impact of social protection programmes on these outcomes is still unclear. The World Economic Forum's (WEF) Global Gender Report for 2021 ranks Rwanda as 48th on the gender gap for economic participation and opportunity index, suggesting this is an area for improvement.[1] At the same time, one fifth of the population of Rwanda is considered to be food insecure.[2] This general low level of food security combined with the effects of climate change such as erratic rainfall have resulted in growing vulnerability, particularly among rural populations.[3]

2. The World Food Programme (WFP) Office of Evaluation's (OEV), Cash-Based Transfers (CBT) Division, Gender Office, Asset Creation and Livelihood Unit (OSZPR), and the Climate and Disaster Risk Reduction Unit (OSZIR) all partnered with the World Bank's Development Impact Evaluation Unit (DIME) to create the cash-based transfers and gender impact evaluation (IE) window and the climate and resilience (C&R) impact evaluation window.

3. The cash-based transfers and gender impact evaluation window aims to assess the impacts on women's social and economic empowerment of increasing women's participation in work outside the household and of women directly receiving a wage. The climate and resilience impact evaluation window aims to estimate the impacts of asset creation and livelihoods programmes on the resilience outcomes and welfare trajectories of beneficiary households. The impact evaluation in Rwanda contributes evidence for both windows, as they measure the impact of the same intervention, the food assistance for assets (FFA) programme, on two different types of outcomes – food security and gender equality. Given that the impact evaluation is falling under two separate windows with independent objectives, each section in the inception report will be subdivided to provide window-appropriate information wherever it may apply. Thus, each section will have information on cash-based transfers and gender and climate and resilience under their respective subheadings. The main direct outcomes of the intervention are: (i) improving the financial status of the household; (ii) improving the food and nutrition status of the household; (iii) increasing women's earnings; and (iv) supporting women in altering their time-use. The theory of change conjects that these outcomes (in the medium term) impact perceptions of gender norms, attitudes, agency, consumption patterns, and well-being (physical, social, and psychological).

[1] World Economic Forum. 2021. Global Gender Gap Report. Geneva: Switzerland. https://www3.weforum.org/docs/WEF_GGGR_2021.pdf

[2] World Food Programme. 2018. *Rwanda : Comprehensive food security and vulnerability analysis.*

[3] Clay, N. & King, B. 2019. Smallholders' uneven capacities to adapt to climate change amid Africa's 'green revolution': Case study of Rwanda's crop intensification program. *World Developmen*t, 116, 1-14.

2. Evaluation Context

BACKGROUND AND CONTEXT

4. Located in central Africa, Rwanda has one of the highest population densities in Africa. Rwanda is highly prone to natural hazards, including droughts, landslides, floods, earthquakes and windstorms, which have negative economic and social impacts on its development, with 40 percent of the country's population exposed to recurrent risks. As a result of climate change and low-incomes, vulnerable households are increasingly exposed to natural hazards that disrupt people's lives and livelihoods, resulting in an increase in vulnerability and a high rate of food insecurity and malnutrition, as well as, at a larger scale, hampering Rwanda's effort to eradicate extreme poverty.

5. Countrywide, 53 percent of the households use livelihood coping strategies to face food shortages during the months before the harvest, whereby half of households engage in crisis coping strategies such as harvesting immature crops, consuming seed stocks, or decreasing expenditure on productive assets, all of which may seriously impact a household's livelihood and resilience to future shocks.

6. Gender dynamics are an important factor in household vulnerability, with a much higher proportion of households headed by women categorized with *Ubudehe* status (households who are categorized in the lower national social and economic vulnerability categories). For instance, around 31 percent of the households headed by women are classified in *Ubudehe* 1 (lowest household category), against 11 percent of households headed by men. In rural areas, food insecure households mainly depend on agriculture, either through production on their own land (average land size below 0.5 ha) or through the provision of unskilled daily labour. Smallholder farmers generally cultivate a few crops (2-3) and do not have a vegetable garden, resulting in unbalanced diets and high levels of malnutrition.

7. WFP has played an important role in supporting vulnerable populations in Rwanda from climate change-related food insecurity through the implementation of food assistance for assets programmes. In many cases the programmes also take into consideration gender-specific concerns by ensuring that men and women work together on food assistance for assets programmes so as to strengthen their sense of self-worth. Under the 2019-2023 country strategic plan (CSP)[4] for Rwanda, WFP implements a portfolio of resilience and social protection activities (Strategic Outcome 2) that focuses on ensuring vulnerable populations in food-insecure areas have improved access to adequate and nutritious food all year. Under the outcome, specifically under Output 2.3: "Food-insecure people in vulnerable communities benefit from improved assets and skills to increase their resilience to climate-related shocks," WFP is launching the Sustainable Market Alliance and Asset Creation for Resilient Communities and Gender Transformation (SMART) project. The project aims to contribute to community resilience through a package of support, including a stronger soil and water management asset base, livelihood strengthening and diversification, farmer organization capacity strengthening and access to inputs and markets, and social cohesion and gender transformation activities.

8. SMART selectively targets communities with households who are categorized in the lower national social and economic vulnerability categories. In contrast to other programmes that target the "ultra-poor" with unconditional cash or asset transfers, SMART engages vulnerable households who are paid a wage to engage in activities linked to the creation of productive assets (for example, irrigation systems, terrace and marshland restoration etc.), with monitoring to ensure compliance. The typical wage is enough to purchase a standard food basket for a family of four and in line with the daily wage provided by the Government of Rwanda through the national Vision 2020 *Umurenge* Programme's public work.[5] Upon completion, households may continue to benefit from the restored and more climate-resilient assets.

9. The Rwanda impact evaluation aims to estimate the impacts of food assistance for assets programming targeting women on gender equality, household decision-making, women's social and

[4] "Rwanda Country Strategic Plan (2019-2023)" (WFP/EB.2/2018/8-A/8).

[5] Socialprotection.org, 2017. *Vision 2020 Umurenge Programme (VUP),* https://socialprotection.org/discover/programmes/vision-2020-umurenge-programme-vup, (accessed on 14 February 2022).

economic empowerment, and food security. The theory of change posits that these interventions impact perceptions of gender norms, attitudes, agency, consumption patterns, and well-being (physical, social, and psychological). Simultaneously, it is also important to understand how the food assistance for assets programme impacts households' and communities' resilience outcomes, including sustaining food security throughout the year, and their ability to withstand seasonal stressors and idiosyncratic shocks.

PROGRAMME DESCRIPTION

10. SMART selectively targets communities identified as particularly vulnerable, as evidenced by their *Ubudehe* status. The project aims to enhance food security and resilience to shocks, strengthen smallholder farmer production and market access, and build community and government capacities related to nutrition-, gender- and climate-sensitive social protection. Equal opportunities for women and men, which is in line with SDG 5, will be achieved through the introduction of mobile crèches at the food assistance for assets sites where women and men can securely leave their infants while working. The crèches will also serve as a powerful vector for improved nutrition through the provision of nutritious food for the infants, while their parents will be invited for cooking and nutrition education sessions, which will be organized regularly during the food assistance for assets activity period. Moreover, gender equality awareness training will be provided through communities, as well as specific activities aimed at promoting women's decision-making and leadership status within farmer organizations.

11. Funding for the SMART project comes from the Government of the Republic of Korea, providing a budget of 8 million United States dollars (USD) for the period July 2020 to December 2023. It will be managed and overseen by WFP Rwanda, with implementation support from the non-governmental organizations Good Neighbours International and Duhamic ADRI, and engagement of local governments at the district and sector levels. Consultation will be sought with the Ministry of Local Government , the Ministry of Agriculture and Animal Resources, District Authorities from the Government of Rwanda, and other United Nations agencies, as appropriate.

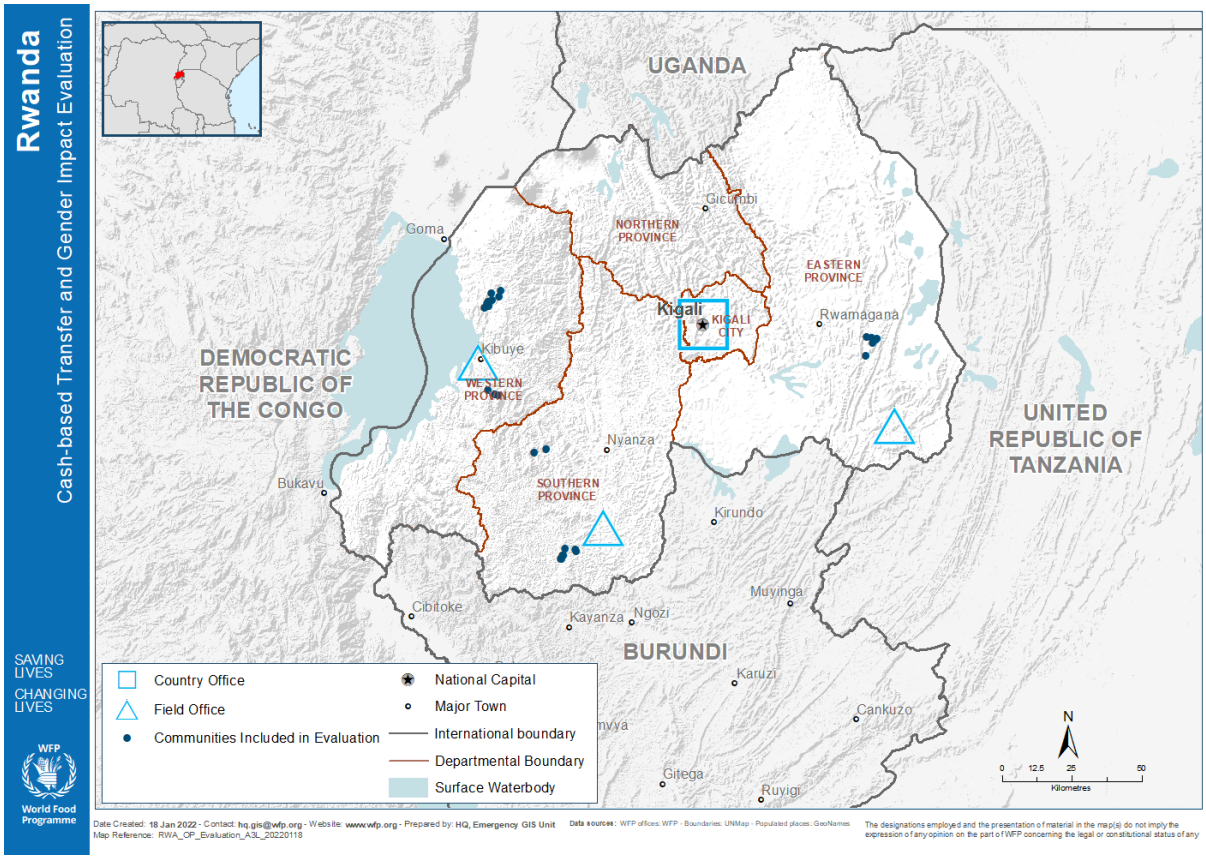
12. This SMART project is building on the positive experience and the results of the *Saemaul* Zero Hunger Communities project (Phases I and II), which has been generously funded by the Government of the Republic of Korea since 2012 and has successfully supported approximately 70,000 people to date with strengthened community resilience to climate-related shocks, improved agricultural production, diversified income-generating activities and enhanced food security. The SMART project, while a continuation of the *Saemaul* Zero Hunger Communities project, is complemented with strengthened components in terms of support to smallholder farmers, access to markets and capacity strengthening of national institutions.

13. The proposed SMART project will be implemented in eight sectors across five different districts. These districts have the highest levels of food insecurity as well as vulnerability to climate shocks:

- Western Province: Rwankuba and Ruganda sectors (Karongi district) and Ruhango and Mukura sectors (Rutsiro district) are characterized by the highest prevalence of food insecure households (up to 62 percent in Rutsiro) as well as exposure to flood and land degradation
- Southern Province: Kaduha and Kamegeri sectors (Nyamagabe district) and Rusenge sector (Nyaruguru district) are characterized by the highest proportion of households adopting crisis and emergency coping strategies to respond to shocks
- Eastern Province: Murama sector (Kayonza district), where 78 percent of the households are affected by drought.

14. Approximately 180,000 people across these sectors will benefit from the SMART project (36,000 direct beneficiaries, 144,000 indirect beneficiaries), including approximately 4,500 refugees. The impact evaluation focuses on five sectors and 1,173 households within the SMART project. The map below illustrates the locations of the communities involved in the evaluation.

Figure 1: Map of communities included in impact evaluation



3. Evaluation Approach and Questions

APPROACH

15. This impact evaluation will employ a clustered randomized controlled trial (RCT) design, in which 78 communities across the country are randomly assigned into one of three groups containing 26 communities each. In each group, approximately 390 households will participate, for a total sample of 1,173. All participants are expected to receive USD 90 by the end of the project.

16. The impact evaluation will involve three rounds of data collection, allowing the team to separately estimate short-term and medium-term impacts (timeline presented in Section 8). Baseline data collection will take place before programme implementation begins. Each food assistance for assets activity is expected to last on average three months, with midline data collection taking place one and a half months into activity implementation. The impact evaluation endline data collection will occur after final intervention activities for the treatment group. Additionally, the project also plans to collect food security data on a more frequent basis (every two months) starting after the baseline survey and continuing for a period of a year, using high frequency phone surveys. These phone surveys combined with midline and endline data will allow the evaluation team to observe changes in food security over shorter periods of time more frequently, providing a more nuanced picture of the fluctuations in food security over the different agricultural seasons.

17. Impact evaluation results will feed into the design of upcoming food assistance for assets programming in Rwanda and inform the next country strategic plan by WFP Rwanda, which focuses on strengthening institutions and filling gaps in the coverage of government food security and nutrition programmes.

HYPOTHESES

18. The hypotheses tested, and questions answered by the Rwanda impact evaluation are aligned with two WFP impact evaluation windows as outlined below.

Cash-based transfer and gender window

19. The impact evaluation is designed to test the hypothesis that cash-based transfer programming targeting women increases gender equality and women's economic empowerment, by increasing time spent by women in paid labour outside the household and thus increasing their earned income.

20. The first hypothesis is that involving women in work (asset creation through the food assistance for assets programme) directly impacts their time use (shifts towards paid work outside the home), as well as their earnings as they are paid directly for their work.

21. The second (following) hypothesis is that – in the medium term – these combined shifts in time use and earnings will impact women's:

- Perceptions of gender norms
- Attitudes
- Agency
- Consumption patterns
- Well-being (physical, social, and psychological).

22. Thus, in the longer run, the evaluation team hypothesize that supporting women to work outside the home can initiate a "virtuous cycle" where a change in women's perceptions of norms, attitudes, and agency further boosts their participation in paid work outside the home (time use). This in turn positively impacts their earnings, which could amplify (control over) consumption and well-being, even after the food assistance for assets intervention ends. While the programme is targeted at women, it is possible the

programming will also impact men's perceptions of gender norms (and those of the wider community) and attitudes in a way that further contributes to improvements in gender equality.

23. This theory of change is consistent with a body of literature that examines the impacts of providing women opportunities to work outside the household: "Female employment has been shown to delay marriage, increase female work aspirations, improve child health, and reduce the male: female sex ratio (...). In the United States, rapid growth in female labor force participation preceded important changes in norms regarding gender roles in both the economy and the household". [6] Recent experimental work has demonstrated attitudes[7] and norms[8] shape women's agency and, in turn, women's labour supply. While food assistance for assets programmes have demonstrated that they are an effective tool for economic development through increased earnings,[9] there is less evidence on the impacts of participant gender and there is also a lack of evidence on projects with a short duration.

24. More details on the theory of change are presented in Annex 2.

25. Related to the cash-based transfer and gender window, the evaluation answers the following questions:

1. What is the impact of women's participation in a food assistance for assets programme (working outside the household and receiving cash in return) on their social and economic empowerment?[10]
2. What is the impact of conditional cash transfer for work on women's social and economic empowerment, as well as on household income and welfare?

26. Each question will be evaluated using the same outcome indicators (detailed further in Table 1 below):

- Time use
- Earnings
- Perception of norms
- Attitudes
- Agency
- Consumption patterns
- Social, physical, and psychological well-being.

27. These evaluation questions are derived directly from the theory of change, and are intended to isolate the impact of increasing women's income and time spent working outside the household on gender equality and women's empowerment, while controlling for the "income effect" of the cash transfer generally (comparing to the second treatment arm), and understanding the overall impact of the WFP programming (comparing to the control).

---

[6] Field, E. M., Pande, R., Rigol, N., Schaner, S. G., & Moore, C. T. 2021. On Her Own Account: How Strengthening Women's Financial Control Affects Labor Supply and Gender Norms. *American Economic Review* 111(70), 2342-75.

[7] Dhar, D., Jain, T., & Jayachandran, S**.** 2018. *Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India*. Working Paper 25331, National Bureau of Economic Research.

[8] Beaman, L., Chattopadhyay, R., Duo, E., Pande, R., & Topalova, P. 2009. Powerful women: does exposure reduce bias? *The Quarterly Journal of Economics*, 124(4), 1497-1540. Bursztyn, L., Gonzalez, A. L., & Yanagizawa-Drott, D. 2018. *Misperceived social norms: Female labor force participation in Saudi Arabia*. Working Paper 24736, National Bureau of Economic Research.

[9] Imbert, C., & Papp, J. 2015. Labor market effects of social programs: Evidence from India's employment guarantee. *American Economic Journal: Applied Economics*, *7*(2), 233-63.
Gazeaud, J., Mvukiyehe, E., & Sterck, O. 2021. Cash transfers and migration: Theory and evidence from a randomized controlled trial. *The Review of Economics and Statistics,* 1-45.
and Adjognon, G. S., van Soest, D., & Gutho, J. 2020. Reducing hunger with payments for environmental services (pes): Experimental evidence from Burkina Faso. *American Journal of Agricultural Economics* 103(3), 831-857

[10] This can also include negative unintended effects.

Climate and resilience window

28. The primary hypothesis being tested in Rwanda related to the climate and resilience window is that participation in the food assistance for assets programme will result in increased short-term and long-term capacities to absorb and cope with stressors and shocks. Through a series of activities that create productive assets, support access to markets, and procure crops from farmers, the project should lead to increased household and community well-being and adaptive capacities to shocks and stressors. Concretely this should translate into improved households' and communities' food security and diversified livelihoods, improved community capacities to plan, prepare, and implement actions to reduce their vulnerabilities to shocks and stressors, and to increase household income and well-being over time.

29. This evaluation complements a growing literature on the impacts of multifaceted programmes that aim to generate long-term changes in household wellbeing.[11] The main focus will be on documenting impacts on household food security and welfare over time (through the indicators and data collection methodology detailed below). The evaluation will also directly assess how the programme affects households' ability to mitigate the effects of shocks on their food security and welfare.

30. The main climate and resilience evaluation questions are as follows:

1) Can the SMART programme increase the overall resilience of households?
2) How does the SMART programme affect resilience over time throughout the year?
3) When is the best time of the year to provide cash payments and the best time to involve participants in food assistance for assets activities?

31. The first set of questions from the climate and resilience window will be evaluated using the following outcome indicators:

- Food security
- Financial outcomes
- Shocks and coping strategies
- Earnings
- Consumption patterns.

32. In the case of the climate and resilience window, the evaluation is intended to isolate the impact of SMART programme on the overall resilience of its beneficiaries, as well as resilience over time – both post-harvest and in lean seasons. Measuring resilience requires a two-pronged approach. Firstly, it includes multi-dimensional indices at baseline, midline, and endline covering various outcomes. Secondly, it uses higher frequency measures of food security and shocks in order to assess trajectories of welfare and vulnerability over time, taking into account fluctuations due to seasonality, climatic stressors, and idiosyncratic shocks. More detailed information on how the window conceptualized the definition and measurement of resilience can be found in Annex 5.

OUTCOMES OF INTEREST

33. The outcomes were developed in close collaboration with the Rwanda country office (CO) to ensure operationally relevant indicators are captured. The outcomes are selected based on a review of relevant literature and previous studies that aimed to capture similar outcomes.

34. Inherent in the design of the evaluation is the measurement of progress on gender equality. As both men and women are asked questions on time use, agency, attitudes, perceptions, and well-being separately, the evaluation will be able to identify whether (and how much) inequalities still exist in these areas, and whether the programme contributed to decreasing the gender equality gap.

[11] Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienťe, W., Shapiro, J., Thuysbaert, B., & Udry, C. 2015. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236), 126-799.

Table 1: Main outcomes of interest

Outcome type	Outcome name	Definition	Measurement level
Primary	Consumption	Expenditures over reference period on 10 goods	Household
Primary	Earnings	Total earnings from WFP plus total earnings from other paid permanent and temporary work	Household
Primary	Time use	List of activities from 24-hour recall over past two days; asked separately of men and women	Individual
Primary	Agency	How much the woman's opinion would be considered in a series of decisions	Individual
Primary	Attitudes	The woman's belief of how much time she should spend on productive activities, relative to men	Individual
Primary	Perceptions of norms	The woman's perception of the time use, agency, and attitudes of women in her community	Individual
Primary	Well-being	Psychosocial well-being, life satisfaction, mental health, and intimate partner violence	Individual
Primary	Food security	Food Consumption Score, Food Insecurity Experience Scale	Household
Primary	Shocks & coping	Shocks experienced by the household; livelihood coping strategies used; reduced Coping Strategy Index	Household
Primary	Financial	Savings, loans, financial and in-kind transfers given and received	Household

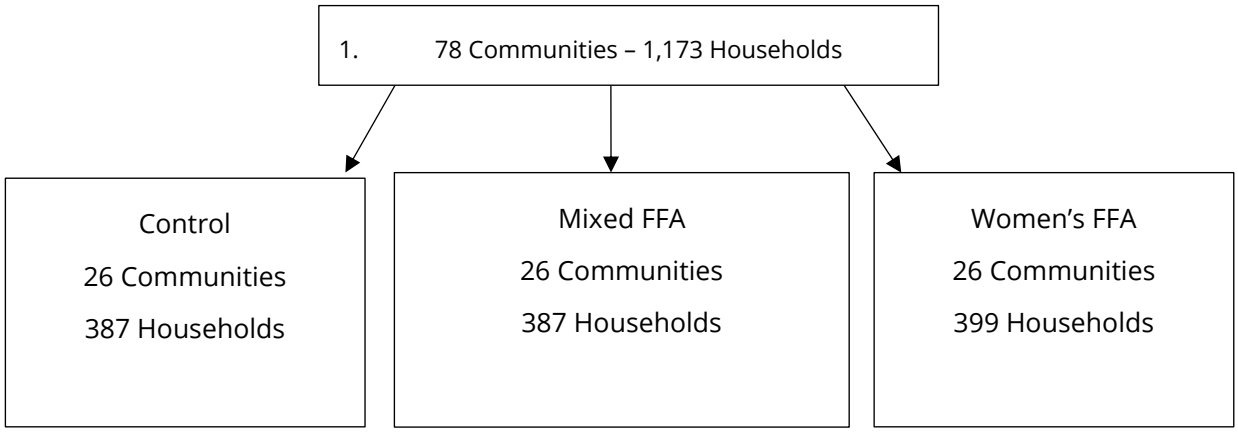
4. Evaluation Design and Sampling Strategy

35. To identify the causal impacts of WFP interventions, the impact evaluation will employ a clustered randomized control trial design. The clustered randomized control trial approach follows from the programme's implementation modality of intervening at the community level, which would not have allowed for a household-level randomization. To start, DIME and the WFP Rwanda country office selected 78 communities for inclusion in the evaluation using the following criteria:

- They do not expect a WFP transfer this year
- They rank in the *Ubudehe* category 1 and 2 in the country office's strategy

36. In a second step, the 78 communities will be randomly assigned either into one of the two treatment groups or into the control group (see Figure 2), producing a clustered randomized design.

Figure 2: WFP Rwanda impact evaluation design



37. WFP will work with local community leaders and government officials to identify 15 of the most vulnerable households within each community for a total sample size of 1,173 households (see the next section on power and sample size calculations). A feature of the clustered randomized controlled trial design is that all selected beneficiary households within a community will receive the same treatment to avoid any "spillover" concerns that might arise from a within-community household randomization approach. The household identification process in all 78 communities will be the same regardless of "treatment" assignment in order to avoid any biases.

TREATMENT AND CONTROL GROUPS

38. Details of the two treatment arms and control group are as follows:

39. Treatment 1: Beneficiaries in this treatment group receive a conditional cash transfer of approximately USD 90 disbursed over three months, provided they work on an asset – where the primary woman or man decision-maker is registered to receive the transfer and work on the asset.

40. Treatment 2: Beneficiaries in this treatment group receive a conditional cash transfer (USD 90) disbursed over three months, provided they work on an asset – where the primary woman decision-maker is registered to receive the transfer and work on the asset.

41. Control Group: Beneficiaries in this treatment group do not receive a transfer in the first phase. They might receive a conditional cash transfer (USD 90) disbursed over three months – where the primary woman or man decision-maker is registered to receive the transfer and work on the asset after the impact evaluation endline survey is completed.

42. By including a control group, the impacts of the "standard" food assistance for assets programme, which usually targets men, can be measured and compared with the impacts of not participating in the

food assistance for assets programme. The modified "women's food assistance for assets" treatment arm additionally allows for comparisons with the "standard" food assistance for assets arm, measuring impacts on women's social and economic empowerment when they are directly targeted by the food assistance for assets programme.

SAMPLE SIZE CALCULATIONS

43. A sufficient sample size in an impact evaluation assures that individual characteristics balance across treatment and control groups, so that these groups are the same on average and are representative for the population they were drawn from. The power calculations applied for this evaluation follow standard research norms to estimate the minimum sample size needed to detect a reasonable impact and minimize the risk of biased estimates.

44. The power calculations were implemented separately for the cash-based transfer and gender outcomes and the climate and resilience outcomes, with cash-based transfers and gender focusing on low frequency measures (baseline, midline, endline) and the climate and resilience outcomes focusing on high frequency measures (high frequency surveys). Given the differences in the objectives and the rounds of data that will be used to estimate impact in both cases, it is important to ensure the sample size is commensurate with power calculations for analysis that contributes to both windows.

Cash-based transfers and gender window

45. The country office's budget and implementation capacities allow for the impact evaluation to be conducted in 78 communities. For the first power calculations based on these parameters, we use women's preferred consumption as an outcome, as it can be calculated in any household survey. For the second power calculation, we use predicted household consumption. We use the 2015/2016 Kenya Integrated Household Budget Survey for these calculations, restricting it to rural poor households, consistent with the typical households targeted by WFP cash assistance for assets (CFA) programmes. We apply the standard formula for the minimum detectable effect (MDE):

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

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.

46. To calculate the expected effect size for each analysis, we focus on effects during the midline survey. For household consumption as an outcome of pooled treatment, we first apply a marginal propensity to consume from cash transfers of 0.67.[12] We then multiply this by the share of households anticipated to take up the intervention, and the monthly transfer size relative to average monthly household consumption. For women's income as an outcome of women's food assistance for assets conditional on being treated, we continue to apply a marginal propensity to consume of 0.67. We then multiply this by take-up, which is now the share of participating households who shift from men to women participants in response to women's food assistance for assets, and the monthly transfer size relative to average monthly household consumption.

47. We find the following minimum detectable effects for Rwanda, which are reasonable (as determined in the literature):

48.	Number of observations	49.	1,173 households
50.	Number of clusters	51.	78 communities
52.	Transfer size	53.	USD 90 (approximately)

[12] This is estimated based on Haushofer, J., Ringdal, C., Shapiro, J. P., & Wang, X. Y. 2019. *Income Changes and Intimate Partner Violence: Evidence from Unconditional Cash Transfers in Kenya*. Working Paper 25627, National Bureau of Economic Research.

54. MDE for consumption	55. 0.104 standard deviations

Climate and resilience window

56. The power calculations for the climate and resilience window rely on high frequency surveys to measure resilience dynamics over time. To make recommendations for sample size and power, we used data collected by the Catholic Relief Services under the Measurement Indicators for Resilience Analysis initiative in Madagascar. The sample consists of 601 households from 32 communities in Madagascar surveyed every month for a period of 18 months. We use the first twelve of these 18 months so that we are consistent in using one full year as the relevant period. While Madagascar is a different context, this data provides for a unique opportunity to test statistical power needed to measure resilience dynamics because it collects three common food security indicators: Household Hunger Scale (HHS), Food Consumption Score (FCS), and Household Dietary Diversity Score (HDDS).[13] We take this as our starting point and assess the role of survey frequency on power to compute changes in these measures over time. We conduct power calculations on parameters provided by the Rwandan context: 75 clusters, one third of which are control and two thirds of which are treatment (standard food assistance for assets and women's food assistance for assets). All households in treated communities experience one of three treatment effects:

1. Increases the mean of high frequency measures by X percent of the control mean holding other parameters constant
2. Decreases the standard deviation of food security measures for a household over time by X percent of baseline control standard deviation, keeping other parameters constant
3. Decreases 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).

57. This allows us to estimate power for detecting effects of programmes that may make households less food insecure on average but not change variability around that mean or vice versa. For each of these effects, we conduct power calculations in using traditional estimates of alpha (0.05) and power (0.8). We are able to show the minimum sample size and number of clusters required to detect a statistically significant effect for each treatment. These results give guidance for how frequently we need to collect food security data in order to identify impacts on these measures of household resilience. The outcomes are as follows for a 15 percent effect size on Food Consumption Score and Household Dietary Diversity Score.

Table 2: Frequency of data collection by outcome

Food Consumption Score	Frequency	Sample size	Number of clusters
Average FCS	Semi-annually	222	28
Average FCS	Quarterly	147	18
Average FCS	Bimonthly	126	16
Variability in FCS	Semi-annually	1764	221
Variability in FCS	Quarterly	795	99
Variability in FCS	Bimonthly	492	62
Share of FCS below threshold	Semi-annually	2835	354
Share of FCS below threshold	Quarterly	1707	213
Share of FCS below threshold	Bimonthly	1329	166
Household Dietary Diversity Score	**Frequency**	**Sample size**	**Number of clusters**
Average HDDS	Semi-annually	213	27
Average HDDS	Quarterly	144	18
Average HDDS	Bimonthly	120	15

[13] While food security is only one measure of resilience, it provides a robust foundation for the statistical estimation of power. It is responsive to shocks and stressors and widely available in existing surveys for validation of our power calculations.

Variability in HDDS	Semi-annually	2025	253
Variability in HDDS	Quarterly	804	101
Variability in HDDS	Bimonthly	495	62
Share of HDDS below threshold	Semi-annually	711	89
Share of HDDS below threshold	Quarterly	465	58
Share of HDDS below threshold	Bimonthly	381	48

58. These results show that it would be possible to detect effects in average levels of food security with a smaller sample, but not in the variation or share below the poverty line over time. In order to be powered to find programme effects on these measures throughout the year, a bimonthly survey is recommended. Based on these calculations, a sample of eight households per site surveyed bimonthly was identified in the 75 clusters that are part of the evaluation.

5. Data Collection

59. The timeline of surveys and implementation is presented in Section 8. All data will be collected using computer assisted personal interviewing techniques, utilizing Android tablets running SurveyCTO data collection software. Surveys are approximately two and a half hours in length.

60. Baseline surveys will take place just prior to the start of the intervention. A midline survey will take place during the implementation of cash transfers, and its reference period will lie entirely during the three-month food assistance for assets activity period during which cash transfers are being made. This is necessary so that all questions during the midline, particularly time use and income, can be used to estimate the direct impacts of food assistance for assets and women's food assistance for assets. An endline survey will occur just after the end of the intervention – sufficiently after it so that the reference period for the endline survey will exclude the period of the intervention. This is necessary so that all questions during the endline can be used to estimate the persistent indirect impacts of food assistance for assets and women's food assistance for assets.

61. These rounds of data collection will be complemented with high frequency phone surveys that ask a smaller set of questions at quarterly intervals (every three months), in order to capture resilience dynamics over time. These surveys will be conducted entirely over the phone, with a sub-sample of respondents with working cell phones. Whenever a phone survey is implemented at the same time as a midline, respondents who are part of the most recent midline sample will not be included. Details of the survey are presented below, and details of the sample are presented above.

62. By virtue of the impact evaluation design, data collected will be disaggregated by gender of the respondent. Importantly, the impact evaluation does not consider a "household" to be one unit, but rather individuals are within a household for the gender outcomes. As such, the survey is repeated to both men and women respondents in the same household for all of the key outcomes described in Section 3. One exception is the module measuring intimate partner violence (IPV) – for ethics and protection, this module is only asked of women responding without the men respondents being present or aware of the module.

63. While the relevant survey modules are relatively standard across all impact evaluations in each window, the survey will be piloted prior to data collection with local communities in Rwanda to ensure questions are relevant to the context. In addition, the consumption module is specifically tailored to context (described below), and the power calculations for the impact evaluation use data from the El Salvador Income and Expenditure Survey.

64. Finally, as an attempt to capture the range of ways that agency, attitudes, and norms can manifest in everyday decision-making, each of these outcomes is measured using multiple questions along three separate productive assets. Additionally, four separate activities are used to understand women's decision-making power and perception of norms. These variations on the key outcomes are described below.

QUANTITATIVE INSTRUMENTS

65. The baseline, midline, and endline surveys include ten main outcome categories, measured as follows:

Cash-based transfers and gender window

66. Time use: The woman respondent is asked for a 24-hour recall of her activities over the past two days, following the approach of the American Time Use Survey. When the primary decision-maker in the household who is a man is available, he is asked about his activities over the past two days; when he is not, the woman respondent is asked about his activities.

67. Earnings: Earnings for each household member are collected as follows: (i) for the baseline survey: income from the six months preceding the survey; (ii) for the midline survey: income from the time between baseline and midline survey; (iii) for the endline survey: income from the minimum of the six months preceding the survey *or* two weeks after the completion of the intervention. Earnings are measured as total earnings from WFP plus total earnings from other paid permanent and temporary work. They include wage labour, non-farm business, agriculture, and livestock.

68. Perceptions of norms: The woman respondent is asked how much time she believes women, relative to men, in her community spend on three productive activities. Next, the woman respondent is asked how much the opinion of women in her community would be considered, relative to primary decision-makers who are men in their households, on the same set of decisions as the questions about agency. Finally, the woman respondent is asked about the attitudes of people in her community. These questions mirror the above questions on attitudes towards time use and attitudes towards agency.

69. Attitudes: The woman respondent is asked how much time she should spend, relative to the primary decision-maker in the household who is a man, on the three productive activities listed above.

70. Agency: The woman respondent is asked, relative to the primary decision-maker in the household who is a man, how much her opinion would be considered in a series of decisions. These questions follow the Demographic and Health Survey's module on consumption ("major household purchases", "purchases from the primary male decision-maker's income", "purchases from the female respondent's income", "the female respondent's health care"), and include additional questions on decision-making over both men's and women's time in three productive activities ("work in self-employment", "work for a salary", "work on household chores").

71. Consumption: Expenditures over a standard reference period for up to ten goods are asked. Five goods are selected as the goods that most strongly predict household consumption in a household survey from the same context. Five goods are selected as the goods that most strongly predict women's income, controlling for total household consumption, in a household survey from the same context. Expenditures on education, men's clothing, and women's clothing will be included.

72. Well-being: Modules to measure locus of control, psychosocial well-being, life satisfaction, intimate partner violence, and depression using Patient Health Questionnaire-9 (PHQ 9) will be administered. Two modules will be used to assess any unintended consequences of the intervention on women. First, the time use module will reveal whether the programme has contributed to a "second shift" for women; as they pick up more work outside the home, this may not be accompanied by reduced domestic labour burdens. The share of time spent on domestic and carework duties between men and women is an important indicator of gender equality in the analysis. Second, questions from the Demographic and Health Survey module on intimate partner violence are included.[14] This will allow the impact evaluation to identify any unintended consequences of a direct transfer of cash to women on the intra-household dynamics and her experience of intimate partner violence.

Climate and resilience window

73. The baseline, midline, and endline surveys include five categories of resilience outcome variables:

74. Consumption: Expenditures over a standard reference period for up to ten goods are asked. Five goods are selected as the goods that most strongly predict household consumption in a household survey from the same context. Five goods are selected as the goods that most strongly predict women's income, controlling for total household consumption, in a household survey from the same context. Expenditures on education, men's clothing, and women's clothing will be included.

75. Food security: The woman respondent is asked about the household's food consumed over the past seven days, based on categories from WFP's standard Food Consumption Score Nutritional Quality Analysis indicator. She is also asked about her household's experience with food insecurity based on the standard Food Insecurity Experience Scale, which consists of eight questions capturing a range of food insecurity severity, with yes/no responses.

76. Earnings: Earnings for each household member are collected as follows: (i) for the baseline survey: income from the six months preceding the survey; (ii) for the midline survey: income from the time between baseline and midline survey; (iii) for the endline survey: income from the minimum of the six months preceding the survey *or* two weeks after the completion of the intervention. Earnings are measured

[14] These include adapted questions based on Haushofer, J. & Shapiro, J. 2016. The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *The Quarterly Journal of Economics*, 131(4), 1973–2042.

as total earnings from WFP plus total earnings from other paid permanent and temporary work. They include wage labour, non-farm business, agriculture, and livestock.

77. Shocks and coping strategies: The woman respondent is asked what shocks (drought, flood, family death, asset loss, job loss, etc.) the household has suffered over the previous 12 months and the severity of each shock. In response to any of the shocks identified, she is asked which mechanisms the household used to cope with shocks over the previous 12 months. Examples of coping mechanisms are selling assets for cash, reducing consumption, increasing labour supply, and access to safety nets.

78. Finance: The woman respondent is asked about four financial outcomes: their current savings levels, whether they have taken a loan and their current outstanding debt, any insurance products they currently own, and if they receive any cash transfers (from non-governmental organizations (NGOs), friends, or family members) over the past month.

79. These outcomes are measured during the baseline data collection, the midline (approximately one and a half months after the start of the project), and at endline (after the completion of the three-month project cycle).

80. High frequency phone surveys occur every three months and cover a smaller subset of questions relevant to resilience dynamics over time, including food security, shocks and coping strategies, time spent in productive activities, assets, and reservation wages (that is, the hypothetical wage that respondents would be willing to accept for labour). The following highlights the main resilience outcomes across survey rounds:

Table 3: Resilience outcomes by survey round

	Baseline	High Frequency	Endline
Consumption			
Food Consumption Score	X	X	X
Food Insecurity Experience Scale	X	X	X
Consumption Expenditure	X		X
Financial Outcomes			
Savings	X		X
Borrowing	X		X
Cash Transfers	X		X
Insurance	X		X
Other Outcomes			
Assets	X	X	X
Income Activities	X	X	X
Shocks and Coping	X	X	X
Reservation Wages		X	X
Dietary Diversity	X		X

MANAGEMENT OF DATA QUALITY

81. The evaluation team is taking multiple steps to ensure it collects high-quality data. This begins by hiring a set of 30–40 experienced enumerators. These enumerators have worked with WFP in the past and are hired on short-term contracts for the number of days required. The hiring process takes place through a third party, a recruitment and human resources management company with which WFP contracts. The

evaluation team then trains the enumerators in best practices, checks incoming data, and communicates any data issues regularly to the enumerators. Each one is described in turn:

Enumerator training

82. The training is divided into four stages and will take approximately one week to complete:

83. Review the survey's content: the evaluation team will guide enumerators through each section of the survey, eliciting their feedback about the content and answering any questions they may have about how to administer the questions to respondents. This process ensures that any ambiguities about the questionnaire are resolved ahead of time.

84. Mock surveys: once the survey has been reviewed, the evaluation team will ask the 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.

85. Review best practices: once the mock surveys are complete, the evaluation 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
- How to ensure smooth transition in telephone surveys, especially when the survey will be broken up into several telephone calls.

Ensuring beneficiary and enumerator protection

86. The survey asks about sensitive topics, including intimate partner violence and mental health, that could be distressing for respondents and elicit responses that enumerators may find emotionally difficult to discuss. To address these concerns, the evaluation will follow WFP guidelines on collecting sensitive data for impact evaluations, and seek support from the gender and/or protection officer to establish the proper protection infrastructure. This includes hiring a gender specialist from WFP to provide enumerator training. This training will instruct enumerators on how to conduct the more sensitive questionnaire modules, and on when and how to use referral pathways if a beneficiary reports an incident of violence.

Data quality protocols

87. The computer assisted personal interviewing survey will ensure the number of logical inconsistencies in the data is reduced to a minimum. Additionally, the team will carry out high frequency checks (HFCs) during the entire data collection period. High frequency checks are a data quality assurance process meant to detect any anomalies in the data collected. They are run daily so the evaluation team can make any necessary adjustments to data collection processes in the field. High frequency checks look out for the following instances:

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

88. Any anomalies detected through this process will be flagged to the data collection team immediately. In addition, the evaluation team will also perform a set of back-checks. This refers to drawing a random 10–20 percent sample of households and calling them back to validate some of their answers.

Communication strategies

89. The evaluation team has developed an innovative data tracking dashboard. Specifically, the evaluation team developed code that downloads the raw data from the server and computes the various statistics used for the high frequency checks, as well as completion status of all surveys. This information is then stored in a Google Sheet for different evaluation team members to consult. In particular, enumerators can log on to check how many surveys they have completed, and which surveys are still pending. This ensures the evaluation team is actively tracking survey progression and data quality.

QUALITATIVE INSTRUMENTS

90. Given limited data collection budgets, the evaluation team chose to focus data collection on household surveys that capture outcomes at the household and individual level. The data collected is both quantitative and qualitative, with significant opportunity for respondents to elaborate on responses through text fields and for enumerators to record "other" responses. An additional barrier to focus group discussions was the requirement for the institutional review board (IRB) to limit "research activities" that increase the risk of COVID-19 group-based spread.

91. However, in addition to measuring the impact of the WFP programme in Rwanda, the impact evaluation will collect limited qualitative data to examine important process-related questions:

1) 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?

2) How did intended beneficiaries supported by the programme experience participation in selected interventions? And, how do they perceive the positive or negative consequences of any measured outcomes?

92. If COVID-19 continues to prevent in-person interviews or focus groups, qualitative information will be collected remotely, through phone-based interviews or qualitative surveys.

EVALUATION IMPLEMENTATION MONITORING SYSTEM

93. WFP and DIME are working together to ensure beneficiaries receive the scheduled WFP programming on time. WFP regularly tracks when transfers are made to programme recipients, and also tracks whether work requirements are met. DIME is complementing these efforts by ensuring that the programme variations introduced are properly followed. In other words, DIME is monitoring treatment compliance in the following way:

94. Treatment 1: The household's primary woman decision-maker is registered as the primary beneficiary in the WFP SCOPE database. She will receive cash transfers in a timely fashion.

95. Treatment 2: The household's primary man or woman decision-maker will be registered as the primary beneficiary in WFP SCOPE. She/he will receive cash transfers in a timely fashion. She/he will also be invited to work on a community asset and will be asked to attend any necessary meetings or trainings for this work. Attendance at all meetings will be recorded and digitized.

96. Control group: Households should not receive cash transfers until after endline, nor should they be assigned an asset to work on. They should not attend asset trainings or meetings.

97. The impact evaluation field coordinator will routinely run a code that will flag any discrepancies with the aforementioned treatment compliance indicators. If any discrepancies are flagged, the field coordinator will notify WFP and/or the cooperating partner responsible for implementing field activities.

6. Data Processing and Analysis

DATA CODING, ENTRY, AND EDITING

98. All data used will be collected via tablets. The data will be stored on SurveyCTO servers. As soon as a surveyor marks a filled-out form as "finalized", the form's contents are encrypted. Whenever form data is transmitted via 3G or other Internet network, it is encrypted in transit using "Secure Sockets Layer" (SSL) technology as well. Finally, any data that is downloaded from the server will either be encrypted or be purged of any personal identifiers before analysis. A series of back-checks will be performed on the data collected. Any mistakes that are detected will be recorded and changed. This will avoid missing data systematically across treatments (if there is missing data they will be random across treatments, and therefore do not impact the analysis).

PROGRAMME-SPECIFIC QUANTITATIVE AND QUALITATIVE DATA ANALYSIS

Cash-based transfers and gender window

99. The main objective of the analysis, as per the window's design, is to estimate the impacts of women's participation in the programme on the main outcomes of interest (Section 3). In this case, comparisons across contexts are particularly complicated because we are interested in the impact of one intervention, women's food assistance for assets, controlling for an endogenous variable (programme participation). However, food assistance for assets is a plausible instrument for programme participation, suggesting an instrumental variable estimator. In addition, it is also likely the case that the extent of participation in the programme, in both arms, varies across contexts; as we are interested in assessing the impacts of programme participation, rather than of the randomly assigned arms themselves, this suggests further using women's food assistance for assets as an instrument for women's participation in the programme. The model for these estimates is provided in detail in Annex 4 and summarized here. Standard errors will be clustered at the community level, in accordance with the clustered randomization design.

100. We estimate the following instrumental variables model in each survey wave *t*. Letting Y_{ht} be outcome *Y* for household *h* in survey wave *t* (0 for baseline, 1 for midline, and 2 for endline), we estimate:

$$Y_{ht} = \beta_{1t} WomanParticipant_{hl} + \beta_{2t} AnyParticipant_{hl} + X_h' \gamma_t^Y + \varepsilon_{ht}^Y (1)$$

$$WomanParticipant_{hl} = n_{1t}^T CashforWork_h + n_{2t}^T CashforWomen'swork_{hc} + X_{h0}' \gamma_t^T + \varepsilon_{ht}^T$$

$$Anyparticipant_{hl} = n_{1t}^T CashforWork_h + n_{2t}^I CashforWomen'swork_h + X_{h0}' \gamma_t^I + \varepsilon_{ht}^I$$

where X_h is a vector of controls that includes the value of the outcome of interest at baseline and any stratifying variables used for randomization (in Rwanda the stratifying variables includes the municipality). The primary coefficient of interest is β_{it} – the estimated impact of shifting all of a household's income from men to women.

101. For inference, we will control false discovery rate across outcomes.[15] We will test balance along all outcomes of interest and key demographic variables, including household size.

102. By virtue of the impact evaluation design, outcomes will be analysed by gender to detect inequalities between household members. A feature of the clustered randomized controlled trial design is that all selected beneficiary households within a community will receive the same treatment to avoid any "spillover" concerns that might arise from a within-community household randomization approach. As non-beneficiary households within treated communities are not surveyed, the evaluation will not be able to detect any positive (or negative) spillovers within communities.

[15] The outcomes used randomization inference following guidelines from Anderson, M. L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects." *Journal of the American Statistical Association*, 103(484), 1481–1495.
Anderson (2008)

103. For each regression, we will test for differential attrition and, for questions where men respond when present, differential attrition of respondents who are men. When statistically significant attrition is present for a given outcome, we will estimate Lee bounds for that outcome and report the average upper bound and average lower bound in robustness.

Climate and resilience window

104. Estimating impact on resilience

105. In order to measure the impact of the food assistance for assets package (standard and women's groups pooled together) against the control group, the primary means of analysis is a simple regression of resilience outcomes on treatment status. A dummy variable (1/0) will be used for randomized treatment at the community level (community receives food assistance for assets or is assigned to the control group):

$$Y_{hct} = \beta_0 + \beta_{1CCT} + \varepsilon_{hct}$$

where Y_{hct} is the mean or intra-annual standard deviation of the resilience outcome of interest (food security, shocks, earnings etc.); β_{CCT} is an indicator for being assigned to the food assistance for assets programme (defined here as a conditional cash transfer). The Rwanda data will be pooled together with that of other countries in the climate and resilience window, all conducting experiments comparing food assistance for assets to a control group: Niger, Mali, South Sudan, and the Democratic Republic of the Congo. This regression will include country fixed effects. See Annex 5 for more details on how the window conceptualizes resilience through the measurement of dynamic outcomes.

Heterogeneity to shocks

106. A feature underlying a household's resilience is its ability to cope with shocks. Many programmes are designed to help households mitigate the impacts of shocks but evaluating the ability to smooth shocks can be difficult. Typically, assessing the ability of a programme to buffer against shocks is done by interacting a treatment effect with a variable measuring exposure to a shock.[16] However, evaluations that measure impact through only a baseline and 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. 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 run a regression interacting programme participation with a list of pre-specific context-specific shocks that will include both natural events (for example, droughts as defined by rainfall during main cultivation months falling below a defined threshold) and conflict (for example, as defined by a recorded conflict in standardized data such as the Armed Conflict Location and Event Data Project (ACLED)) and economic shocks:

$$Y_{hct} = \beta_0 + \beta_{1CCT} + \beta_{2CCT} \times \text{Shock} + \beta_5 \text{Shock} + \varepsilon_{hct}$$

107. For Rwanda, that list includes:

- Drought / Irregular rain
- Floods
- Landslides / erosion
- Hail / frost
- Crop pests / diseases
- Animal diseases
- Rise in agricultural input prices
- Lower prices for agricultural products
- Rising food prices

[16] 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. Premand, P.; Stoeffler, Q. 2020. Do Cash Transfers Foster Resilience? Evidence from Rural Niger. Policy Research Working Paper; No. 9473. World Bank, Washington, DC. 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. Working Paper 25969, National Bureau of Economic Research.

- Serious illness or accident of a household member
- Death of a household member
- Divorce, separation
- Theft of money, property or harvest
- Land conflict
- Militia group activity
- Religious conflict
- Ethnic conflict
- Significant loss of non-farm household income (not related to any other shock).

Sampling and specification

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

PROGRAMME-SPECIFIC QUALITATIVE DATA ANALYSIS

109. The evaluation team will be collecting qualitative information relating the implementation process as described in Section 5. The evaluation team will be asking the beneficiaries if, in their perception, the programme has had a positive or negative impact on outcomes.

7. Quality, Risks, and Ethics

ETHICAL CONSIDERATIONS

110. WFP evaluations must conform to 2020 United Nations Evaluation Group ethical guidelines. Accordingly, the Office of Evaluation and DIME are responsible for safeguarding and ensuring ethics at all stages of the evaluation cycle. This includes, but is not limited to, ensuring informed consent, protecting privacy, confidentiality and anonymity of participants, ensuring cultural sensitivity, respecting the autonomy of participants, ensuring fair recruitment of participants (including women and socially excluded groups) and ensuring that the evaluation results in no harm to participants or their communities. During the inception phase, the following ethical issues, related risks, safeguards and measures have been considered:

Institutional Review Board (IRB)

111. The impact evaluation window design, as well as the specifics of the Rwanda evaluation, received ethical approval on 10/03/2020 by Solutions IRB,[17] which is a private commercial company fully accredited by the Association for the Accreditation of Human Research Protection Programs (AAHRPP) and the Institutional Review Board.

Programme exclusion

112. Every impact evaluation participant will be a WFP beneficiary, selected through a rigorous process that consults the community in order to identify the most vulnerable households. All three treatment groups are eligible for cash payments – the control group will just receive this transfer after the impact evaluation has concluded.

Informed consent

113. Every household enrolled in this impact evaluation must consent first to being part of the WFP programme as per WFP guidelines, and then provide informed consent to be surveyed. Refusal to respond to the survey does not preclude participation in the WFP programming. Informed consent will be collected for each survey round separately (baseline, midline, and endline).

Privacy during interviews

114. A woman selected as eligible to participate in the food assistance for assets programme is the primary respondent for the survey. While most survey questions are addressed to the woman, there are a few questions directed to the primary decision-maker who is a man – who is also eligible to participate in the food assistance for assets programme. Therefore, the man who is also eligible to participate (referred to as “primary male decision-maker") will be surveyed on a reduced set of questions.

115. Despite the minimal risks, the evaluation team will take a number of precautions to ensure questions addressed to respondents respect their privacy and comfort. First, interviews will be done at a central point in the village so respondents feel comfortable answering questions about their agency, time use, etc. Second, interviews will be conducted outside of earshot of other participants (including those from the same household) and enumerators. Following a first section of the survey when both woman and men respondents might be present, enumerators will request for others to step away as they interview the woman or man respondent, with the goal of providing a safe and quiet environment for the survey. Third, in contexts where particularly necessary, enumerators will be woman, to ensure the highest degree of comfort for survey respondents. Fourth, the evaluation team will coordinate with WFP and community leaders to help care for the respondents’ children (as necessary) to ensure maximum privacy during the survey. Fifth, all enumerators will go through a training that will last for 1–2 weeks and will be followed by extensive piloting in the field. The goal of the training is to ensure enumerators follow survey best practices in terms of protocols and ethics, but also that questions are asked in a uniform and contextually appropriate manner. Sixth, for the most sensitive questions related to intimate partner violence, third-party

[17] More details on this company can be found here: Solutions IRB

experts will be contracted to train enumerators on how to ask these questions, and handle/refer cases of intimate partner violence to the relevant authorities.

116. These issues will be monitored and managed during the implementation of the evaluation. If any additional ethical issues arise during the implementation of the evaluation, they will be recorded and managed in consultation with the Office of Evaluation and DIME.

LIMITATIONS AND RISKS

Limitations and risks of evaluation method

117. One of the study limitations may be that the results of a single evaluation might not be externally valid. The evaluation team will test the external validity of results across all countries where these interventions are implemented (see each window pre-analysis plan for details). As with any in-field randomized control trial, spillover across communities and differential attrition are potential risks for the evaluation. The evaluation team will work closely with the implementing partners on the ground to monitor potential spillover risks and design clear implementation protocols. The evaluation team expects differential attrition to be less common than in other contexts, since the control group is aware they will be receiving the food assistance for assets intervention in the second year of the programme.

Direct income versus work effect

118. The impacts of the treatment arm focusing on women's work are a combination of women engaging in work outside the household and receiving a direct cash transfer (as pay for their work). The impact evaluation design estimates the combined impact of both features, which makes it hard to disentangle the relative importance of either feature. However, work outside the household usually entails direct pay, which makes this combination operationally relevant to investigate. There is also already a large body of literature on the impacts of cash transfers to women alone, and the contribution of this evaluation therefore is more focused on the work component.

Risks due to COVID-19

119. As a result of COVID-19, the country office has had to implement all of its programmes with third-party non-governmental organizations, which are now responsible for all field-related activities. This creates additional monitoring challenges as the evaluation team has to make sure the non-governmental organizations 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 DIME team has developed a strong working relationship with the country office, and is in frequent communication with the country office and the non-governmental organizations to monitor these dynamics.

Risks due to instability

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

8. Impact Evaluation Management

EVALUATION TEAM AND MAIN COUNTERPARTS

Table 4: Impact evaluation team and main counterparts

Name	Role	Organization/Unit
John Loeser	Principal Investigator, Lead Researcher	DIME
Florence Kondylis	Principal Investigator, Lead Researcher	DIME
Dahyeon Jeong	Principal Investigator, IE Technical Team Leader	DIME
Paul Christian	Principal Investigator	DIME
Jonas Heirman	Principal Investigator	WFP OEV
Hanna Paulose	Window Coordinator	WFP OEV
Felipe Dunsch	Window Coordinator	WFP OEV
Tanay Balantrapu	Research Analyst	DIME
Eric Jospe	Research Analyst	DIME
Guillaume Gatera	Field Coordinator	DIME
Marc-Andrea Fiorina	Research Analyst	DIME
Sarah Cruz	Monitoring and Evaluation	WFP Rwanda
Veronica Rammala	Monitoring and Evaluation	WFP Rwanda
Seonghee Choi	Programme Policy	WFP Rwanda
Laurent Ulimubenshi	Programme Policy	WFP Rwanda
Tiina Honkanen	Programme Policy	WFP Rwanda

Table 5: Evaluation committee

Name	Role	Organization/Unit
Guillaume Gatera	Field Coordinator	DIME
Veronica Rammala	Monitoring and Evaluation	WFP Rwanda
Tiina Honkanen	Programme Policy	WFP Rwanda
Inka Himanen	Head of Programme	WFP Rwanda
Analee Pepper	Gender Consultant	WFP Regional Bureau Nairobi
Nikki Zimmerman	Regional Evaluation Officer	WFP Regional Bureau Nairobi
Ruth Musili	Evaluation Consultant	WFP Regional Bureau Nairobi

WORK PLAN AND DELIVERABLES

121. A baseline survey will take place one month before the start of the intervention to provide enough time for the enumerators to reach the entire sample. The midline will take place half-way through the intervention. The intervention in Rwanda lasts three months and the midline will therefore start in the middle of month two. The endline needs to take place after the intervention ends, with enough time so that the reference period for the relevant questions excludes the intervention period. This corresponds to one and a half months after the intervention ends in Rwanda.

Table 6: Timeline of data collection activities by cohort

	2020		2021												2022									
	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10
All Cohorts - Baseline																								
Treatment Cohort 1 - FFA																								
Treatment Cohort 2 - FFA																								
Treatment Cohort 3 - FFA																								
Control group - FFA																								
Cohort 1 - phone survey						**ML***		R2		R3		R4		R5		**EL***		R7		R8		R9		R10
Cohort 2 - phone survey						R1		R2		**ML***		R4		R5		**EL***		R7		R8		R9		R10
Cohort 3 - phone survey						R1		R2		R3		**ML***		R5		**EL***		R7		R8		R9		R10
All Cohorts - Endline **(Q4 2023)**																								
Harvest			Season A					Season B							Season A					Season B				

Notes: *ML indicates midline survey and EL indicates the endline survey. The sample is divided into three cohorts depending on when sites start FFA activities, and the first and second midline survey are going to be treated as one round of high frquency data collection, respectively.

Table 7: Milestones, deliverables, and estimated timeline

Milestones	Deliverables	Completion Date
Peer-reviewed concept note	Draft inception report	January 2022
Data collection plan and pilot	Terms of reference Questionnaires	November 2020
Data collection (baseline) completed	Cleaned data Dictionaries	January 2021
First data analysis	Presentation Data file Do-files Baseline report	May 2021 February 2022
Implementation of intervention aligned to evaluation	Roll-out plan Monitoring reports verifying treatment and control status	February 2022
Midline data collection (completed)	Cleaned data	March 2022
Follow-up data collection plan	Terms of reference Questionnaire	March 2022
Data collection (follow-up) completed	Cleaned data Dictionaries	June 2022
Final report and policy notes	Technical note Policy note Data file Do-files	November 2022
Dissemination of findings	Presentations	November 2022

QUALITY ASSURANCE AND PEER REVIEW

122. 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 consists of process guidance, quality checklists, templates, technical notes and other reference material to guide evaluation teams and partners throughout the evaluation process. Quality assurance measures will be systematically applied throughout the evaluation phases. These include preparation and selection, design, data collection,[18] consistency of programme

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

implementation with the evaluation design, analysis and reporting. Evaluation reports, including inception, baseline and final reports, are prepared by the evaluation team. Drafts are reviewed by the Evaluation Committee (see Table 5 above). Inception reports and endline reports are also reviewed by external quality support peer-reviewers, the window's steering committee and the window's technical advisory group. Reports are revised based on feedback received and reviewed by the Head of Impact Evaluation. The WFP Director of Evaluation finally approves all the reports before they are submitted for publication. In addition, all final evaluation reports will be subjected to a post hoc quality assessment by an independent entity through a process that is managed by the Office of Evaluation. The overall rating category of the reports will be made public alongside the evaluation reports.

COMMUNICATION PLAN FOR DISSEMINATION AND USE

123. DIME and WFP will ensure that the regional bureaux and the country offices are full partners in discussing and using the evidence created in impact evaluations. DIME field coordinators will 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 is collected, DIME will be responsible for analysis, which ensures a degree of independence in data analysis, but results of this analysis will be regularly shared and discussed with the country and regional teams to ensure that findings can be used for programme decisions and implementing teams' insights can be incorporated in the data analysis. This analysis will be shared with the relevant teams in the form of baseline and endline reports and accompanying presentations. In addition, the evaluation team will draft an academic paper for submission to a peer-reviewed journal and results from the impact evaluation will feed into the broader cross-country analysis being undertaken as part of the partnership.

124. DIME and WFP will communicate regularly with the respective national government and other partner agencies to provide them with updates on the impact evaluation work and results as needed. This will be done through a series of in-country and virtual seminars (as allowed based on context). As the studies are built into WFP programmes, results will feed into future phases of these programmes. 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. DIME and WFP will support the use of results from these evaluations to inform project design of other partners by ensuring easy access and promoting awareness for the evidence generated.

Window synthesis and dissemination

125. When results from individual impact evaluations are finalized, the Office of Evaluation and DIME will conduct a formal window-level synthesis to examine the effectiveness of WFP interventions across all countries included in a window. The window steering committee at WFP will then support the development of consistent, targeted policy messages corresponding to the evaluation questions, which can then be used to feed into the upcoming policies and country strategic plans.

Annexes

Annex 1: Window Summaries

126. The WFP Impact Evaluation Strategy (2019-2026)[19] aims to ensure impact evaluations contribute to global evidence as well as organizational learning. Impact evaluation windows are organized around priority evidence needs identified through literature reviews and extensive consultations. Impact evaluations included in each window are guided by window-level pre-analysis plans (PAPs), which increase the ability of WFP to conduct formal syntheses in order to understand what works across countries (for example, increase external validity).

127. Impact evaluation windows generate evidence for future WFP programming and global engagements. In 2019, all WFP country offices were invited to express interest in both the cash-based transfers and gender window and the climate and resilience window. Following extensive feasibility assessments, El Salvador, Kenya, Rwanda, and the Syrian Arab Republic country offices have been selected for the cash-based transfers and gender window, while Niger, Mali, the Democratic Republic of the Congo, South Sudan, and Rwanda country offices have been selected for the climate and resilience window. The following is a summary of the two window-level pre-analysis plans that inform the impact evaluation design in Rwanda.

Cash-based transfers and gender window

128. The cash-based transfers and gender window measures the impacts of cash transfers on gender equality and women's empowerment (GEWE) outcomes across a series of WFP country programmes. One of the hypotheses tested by the cash-based transfers and gender impact evaluation window is that providing women with opportunities to work outside the household will enhance their agency, as well as increase their control over financial resources, which in turn will lead to expanded social and economic empowerment. The ambition is to learn what works (and what does not) in a way that informs country office programming and contributes to a global evidence base.

129. The first round of programmes selected for the cash-based transfers and gender window is anchored to a version of the WFP food assistance for assets intervention modality. Food assistance for assets is one of the WFP flagship initiatives aimed at addressing the most food-insecure people's immediate food needs with cash, vouchers, or food-based transfers and improving their long-term food security and resilience. The concept is simple: people receive cash, vouchers or food-based transfers to address their immediate food needs, while they build or boost assets such as constructing a road or rehabilitating degraded land. This will improve their livelihoods by creating healthier natural environments, reducing risks and impacts of shocks, increasing productivity and strengthening resilience to natural disasters.

130. DIME and the Office of Evaluation have developed a window-level pre-analysis plan that details an overall impact evaluation design to be followed across the first round of WFP programmes selected for impact evaluations, as well as the outcomes to be measured. The country-specific impact evaluation designs are adapted to each country context but should still allow joint analysis across contexts. Within the food assistance for assets framework, the impact evaluation design for the cash-based transfers and gender window explicitly focuses on cash payments through food assistance for assets and aims to include three groups, where households or communities are assigned to one of three groups:

- Standard food assistance for assets: Cash-based programming (households deciding who would be participating, which could be men or women)
- Women's food assistance for assets: Mandating that women are the named cash recipients and participate in the asset-creation activities

[19] WFP. 2019. Impact Evaluation Strategy (2019-2026), https://www.wfp.org/publications/wfp-impact-evaluation-strategy-2019-2026.

- Control group: Not benefitting from the food assistance for assets project (in the first cycle – where the programmes are usually rolled out in multiple cycles so control group beneficiaries receive the programme at a later date).

131. By including a control group, the impacts of the standard food assistance for assets, which usually targets men, can be measured and compared with the impacts of not participating in food assistance for assets. The modified women's food assistance for assets treatment arm additionally allows for comparisons with the standard food assistance for assets arm, measuring impacts on women's social and economic empowerment when they are directly targeted by the food assistance for assets programme.

Climate and resilience window

132. The climate and resilience impact evaluation window recognizes that food insecurity has many drivers, and that climate change and extreme weather events are exacerbating food insecurity and increasing the likelihood and severity of shocks associated with food crises. The frequency and diversity of shocks require multiple interventions to support populations as they develop and maintain their resilience over time. The climate and resilience window supports resilience programme teams in designing impact evaluations to understand how the integrated packages of interventions, and activities within these packages, contribute to resilience.

133. WFP supports a range of interventions that contribute to resilience. These interventions focus on: 1) the prevention of, or early action and response, to crises; or 2) raising well-being in pre- and post-crisis periods. The climate and resilience impact evaluation window will help inform the implementation of future WFP resilience interventions by assessing:

a) Which interventions are effective in preventing acute food insecurity and in strengthening households' ability to cope with and recover from shocks and stresses
b) Whethe there are interventions or combinations of interventions that are more effective in responding to predictable versus unpredictable factors affecting food insecurity.

134. The first climate and resilience window pre-analysis plan is focused on estimating the impacts of livelihoods, education, health and complementary activities on resilience (for example, absorptive, adaptive, and transformative capacities). Resilience is measured by tracking household well-being (that is, food security and nutrition status) using higher frequency data, and by measuring changes in multidimensional indicators at baseline and endline. Experiments are used to test the causal impact of components of livelihoods, health and education interventions in isolation, and together. Coordinated data collection and experimental designs across six countries will allow for the pooling of impact estimates across countries included in the window. Within countries, experimental designs about the timing and targeting modalities of delivery will produce actionable evidence to optimize programme impacts during implementation.

Annex 2: Theory of Change

Cash-based transfers and gender window

135. Two key elements of the intervention will be evaluated:

a) The involvement of a household member in asset creation activities

b) The transfer of money to the household (or to a named woman recipient).

136. The impact evaluation's theory of change posits that the women's food assistance for assets treatment arm will result in greater gains for gender equality and women's empowerment outcomes than either the second treatment arm, involving a conditional transfer for work to the household, or the control arm.

137. As a first step, involving women in work (asset creation through the food assistance for assets programme) directly impacts their time use (shifts towards paid work outside the home), as well as their earnings as they are paid directly for their work. The hypothesis is that – in the medium term – these combined shifts in time use and earnings will impact women's:

- Perceptions of gender norms
- Attitudes
- Agency
- Well-being (physical, social, and psychological).

138. This theory of change is consistent with a body of literature that examines the impacts of providing women with opportunities to work outside the household.[20]

139. In Figure 3a, solid lines trace out the direct impacts of these changes, while dotted lines trace out secondary impacts. For example, suppose women's participation in public works shifted only agency, time use, earnings, and consumption (including food security). It is reasonable to conclude that impacts on agency were caused by changes in earnings and time use.

140. In the long run, the evaluation team hypothesize that including women in work outside the home can initiate a "virtuous cycle" where a change in women's perceptions of norms, attitudes, and agency further boosts their participation in paid work outside the home (time use). This then positively impacts their earnings, which could amplify consumption and well-being, even after the cash assistance for assets intervention ends.

141. In Figure 3b, solid lines trace out the direct impacts of these changes, while dotted lines trace out secondary impacts. For example, suppose we saw long term changes in attitudes and time use. The conclusion would be that the impacts on time use were driven by persistent changes in attitudes, as opposed to changes in perceptions of norms or agency.

142. While the programme is targeted at women, it is possible the programming will also impact men's perceptions of gender norms (and those of the wider community) and attitudes in a way that further contributes to improvements in gender equality.

[20] Field, E. M., Pande, R., Rigol, N., Schaner, S. G., & Moore, C. T. 2021. On Her Own Account: How Strengthening Women's Financial Control Affects Labor Supply and Gender Norms. *American Economic Review* 111(70), 2342-75.

Figure 3a: Theory of change (medium term)

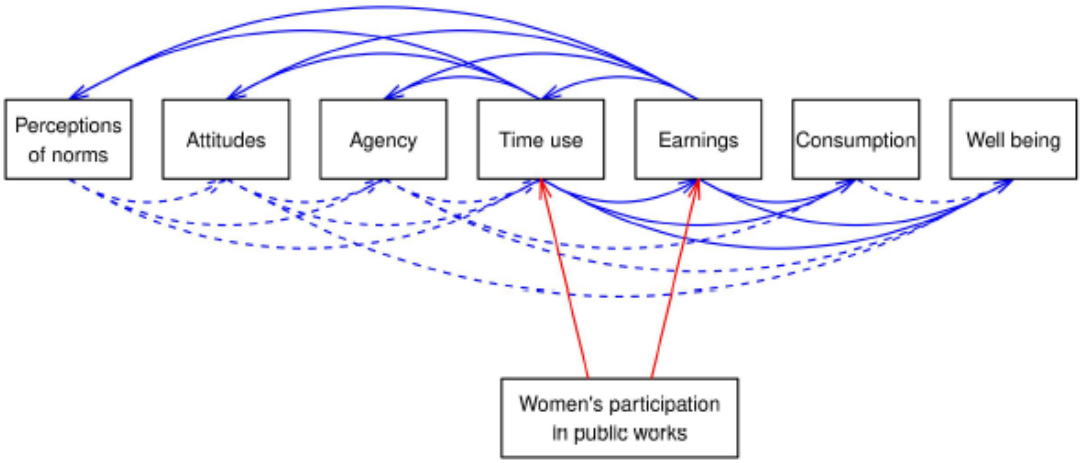


Figure 3b: Theory of change (long term)

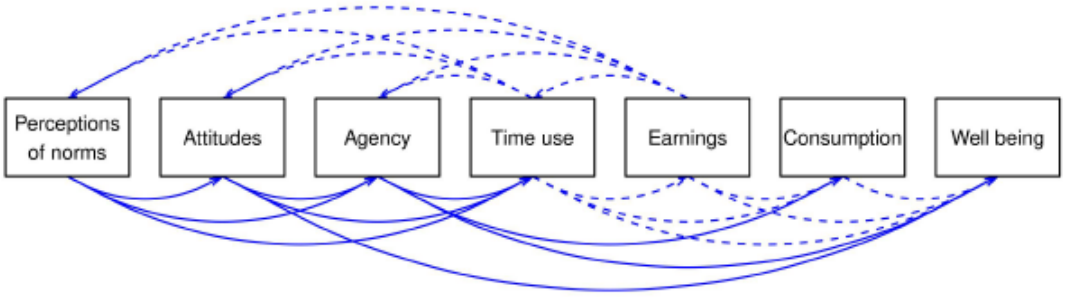
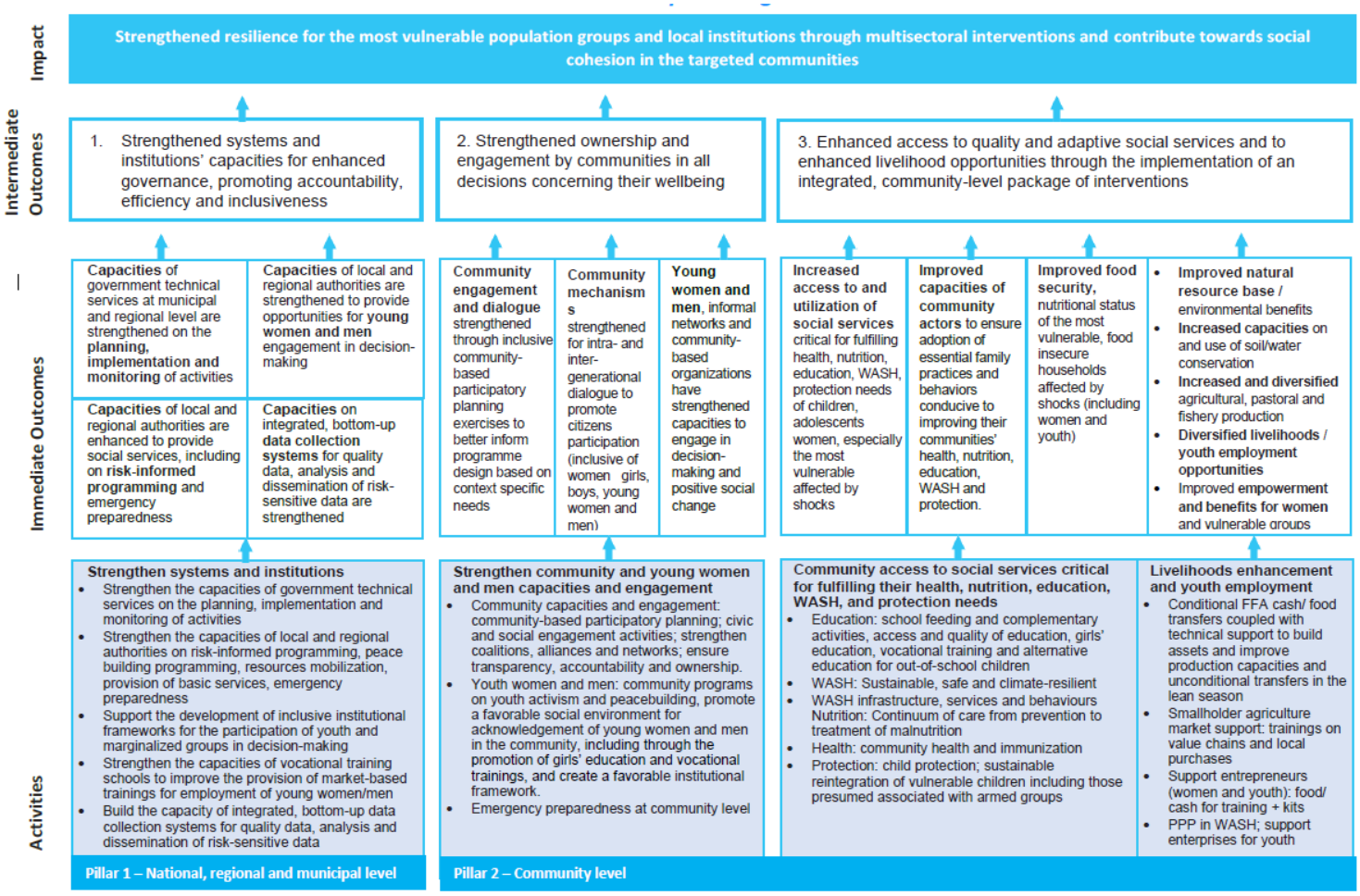


Figure 4: Climate and resilience window



Annex 3: Details on Sample Size Calculations

143. For power calculations, we estimate statistical power for the reduced form:

$$Y_{hc1} = \alpha + \delta_{1c}\text{Treated}_{hc} + \delta_{2c}\text{Cash-for-Women's work}_{hc} + \epsilon_{hc1}^{RF}$$

where for convenience we ignore the presence of controls (yielding modestly conservative power calculations), and "Treated$_{hc}$" denotes that household *h* in country *c* received either the cash assistance for assets or women's cash assistance for assets treatment. We focus on power for δ_{2c}, the effect of women's cash assistance for assets conditional on Treated at midline. For analysis of the impacts of household income, we also estimate statistical power for the reduced form:

$$Y_{hc1} = \alpha + \delta_{1c}\text{Treated}_{hc} + \epsilon_{hc1}^{RF}$$

where we pool across both food assistance for assets and women's food assistance for assets.

144. We apply the standard formula for the minimum detectable effect:

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

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.

145. To calculate σ_e for predicted household consumption, we first select via LASSO the five goods that best predict household consumption, controlling for village fixed effects and number of women, men, and children under the ages of 2, 5, 10, and 16 in the household. We assume predicted household consumption is a surrogate for household consumption.[21] We derive power under their worst-case bounds when surrogacy is violated: doing so is equivalent to scaling σ_e by 1/R2, where R2 is from a regression of residualized predicted household consumption on residualized household consumption. To construct a single measure that we can use across contexts, we normalize by average household consumption. Lastly, we replicate this exercise for women preferred consumption by assuming it is a surrogate for women's income, and we also include controls for total household consumption and total household income.

Power for high frequency measures

146. To make recommendations for sample size and power, we used data collected by the Catholic Relief Services under the MIRA initiative in Madagascar. The sample consists of 1,600 households from 90 communities in Madagascar surveyed every month for a period of 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. These data are unique, because they collect 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. We model through simulations a hypothetical experiment that assigns half of the 90 communities to treatment. All households in treated communities experience one of three treatment effects:

[21] Athey, S., Chetty, R., Imbens, G., & Kang, H. 2016. *Estimating treatment effects using multiple surrogates: The role of the surrogate score and the surrogate index*. https://arxiv.org/abs/1603.09326

1. Increases the mean of high frequency measures by X percent of the control mean holding other parameters constant
2. Decreases the standard deviation of food security measures for a household over time by X percent of baseline control SD, keeping other parameters constant
3. Decreases 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).

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

148. Table 8 below presents results of power calculations to detect a 15 percent effect size for each of the three outcome measures with varying frequencies of data. For a 15 percent effect on power gains in increasing frequency from bimonthly to monthly frequency are small, but the power losses in going from a quarterly to semi-annual schedule are large. We therefore focus on comparisons with the bimonthly and quarterly schedules and compare effect sizes needed to obtain 80 percent power to guide the decisions on whether to plan for quarterly or bimonthly data collection.

149. Table 9 repeats the power exercise for different effect sizes for bimonthly and quarterly schedules. We aim for bimonthly data collection for 1600 households in 90 communities, which is sufficient to detect a 20 percent a change in either mean or standard deviation of food security at 80 percent power for all three measures. Quarterly frequency is sufficient to detect impacts on both means and standard deviations of 20 percent of control averages with 80 percent power for only 2 of these three food security measures.

Table 8

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 9

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

150. Based on the high frequency data power calculations, we establish a sample size of X households per cluster to detect effect sizes of 0.2 standard deviation with a power of 0.8 surveyed bimonthly.

Round	April	June	Aug	Oct	December	February	April	June
Sample size	600 HHs	600 HHs	600 HHs	600 HHs	600 HHs	600 HHs	600 HHs	600 HHs

Annex 4: Details on Quantitative Analysis

For the window-level analyses, we estimate the following IV model in each country *c* and survey wave *t*. Letting Y_{hct} be outcome *Y* for household *h* in country *c* in survey wave *t* (0 for baseline, 1 for midline, and 2 for endline), we estimate:

$$Y_{hct} = \beta_{1ct} Womenparticipant_{hcl} + \beta_{2ct} Anyparticipant_{hcl} + X'_{hc}\gamma^{Y}_{ct} + \varepsilon^{Y}_{hct} (1)$$

$$Womanparticipant_{hcl} = n^{T}_{1ct} CashforWork_{hc} + n^{T}_{2ct} CashforWomen'swork_{hc} + X'_{hc0}\gamma^{T}_{ct} + \varepsilon^{T}_{hct}$$

$$Anyparticipant_{hcl} = n^{I}_{1ct} CashforWork_{hc} + n^{I}_{2ct} CashforWomen'swork_{hc} + X'_{hc0}\gamma^{I}_{ct} + \varepsilon^{I}_{hct}$$

where X_{hc} is a vector of controls that includes the value of the outcome of interest at baseline and any stratifying variables used for randomization. The primary coefficient of interest is β_{ict} – the estimated impact of shifting all of a household's income from men to women.

151. Equations 2 and 3 are our first stage equations: the effect of treatment on women's participation and household participation. We expect η^{I}_{1ct} and η^{I}_{2ct} to be similar across countries – each treatment will have similar effects on household income. However, we expect $\eta^{T}_{2ct} \gg \eta^{T}_{1ct}$ – our first treatment (effectively, a programme that increases women's wage) – will increase women's participation, while our second treatment (effectively, a programme that provides conditional cash for work to the household, or a programme that increases household participation) will have limited effects on women's earned income.

152. For inference, we will control false discovery rate across outcomes, using randomization inference.[22] For each outcome we will present average coefficients across countries using inverse variance weights, and report F-tests for equality of coefficients across countries, controlling false discovery rate across these tests.[23] For some outcomes, these are placebo outcomes on which we do not expect impacts. This would reduce our power on other outcomes in a naive multiple inference correction. We have noted these outcomes as a 0 (instead of an X), but do not yet have a plan to improve upon a naive multiple inference correction.

[22] Anderson, M. L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects". *Journal of the American Statistical Association*, 103(484), 1481–1495.

[23] Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienť e, W., Shapiro, J., Thuysbaert, B., & Udry, C. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science*, 348(6236), 126-799.

Annex 5: Defining Resilience

154. This annex describes the ways 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

155. Measurement of resilience has mostly taken one of three approaches in the literature. The first is to define ex-ante characteristics of households that are expected to be associated with lower resilience, and construct a "resilience index."' This is the approach of the Food and Agriculture Organization's (FAO) RIMA index or the TANGO resilience index, as well as examples of resilience evaluations that use characteristics like diversification of livelihood strategies as a proxy for resilience.[24] 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 measure parameters of the imputed distribution.[25]

156. 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,[26] 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.[27] 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.

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

158. 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

---

[24] 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.

[25] 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.

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

[27] The model is flexible and allows for the observation unit to be an individual, a household, or a village/community, etc, 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, semi-annually, yearly, etc., as is relevant.

has a steeper δ, indicating a steeper trend in food security, meaning that this household will move above the poverty line and/or farther away from it. The blue household in panel C has a seasonal pattern with greater variability than the household with a red line. Seasonality could lead to households falling below a food security threshold in the lean season. In panel D, both the red and blue household experience a shock at the same point.

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

160. These impact evaluations will estimate welfare trajectories within a one-year period following the start of a programme. Figure TA1 shows a hypothetical consumption path for a household in a period t = 0…T. The dynamics shown could represent either a seasonal consumption path with one lean season and one peak season, or a household who experiences one positive and one negative shock.

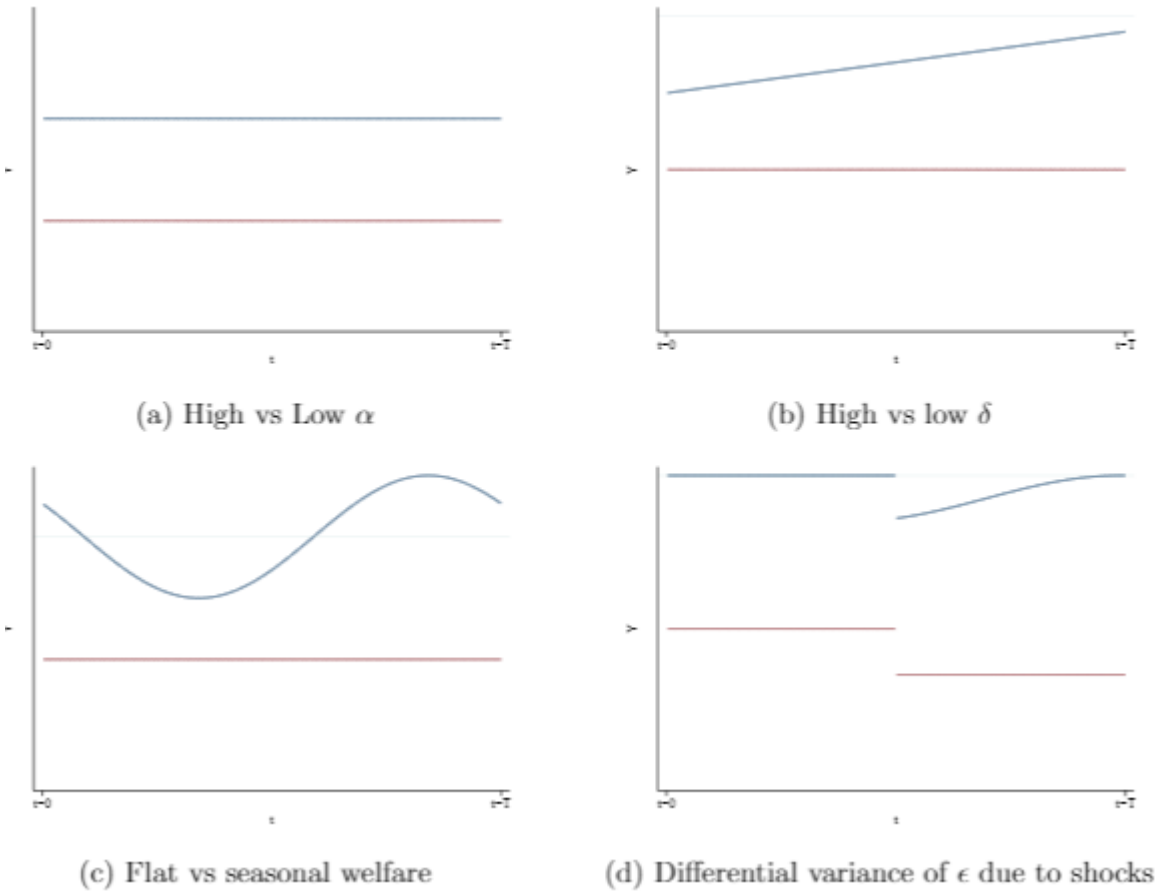
161. The first measure of the consumption equation we are concerned with is the household's intra-annual reference level of consumption -- this is α_i in the structural equation. If we observed a household's value of consumption on every day, this would be measured as a household's average food security status over the period -- as shown by m in Figure TA2 -- Panel A. Next we consider the household's intra-annual standard deviation, the average of the household's deviations from the reference mean (Figure TA2 -- Panel B). The standard deviation captures the combined influence of both $f(d)$ and (ϵ) on household welfare trajectories. This single indicator summarizes the variability associated with both seasonality and shocks within the period. The third measure is the time trend. However, by limiting the comparison within a year, we do not consider a year-on-year trend in welfare. The final measure we consider is the share of the period the household spends below a poverty line or food security range. This is the number of days covered below the poverty line divided by the total number of days in the period of interest (Figure TA2 -- 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 the likelihood of falling below a food security threshold, the share of time spent below the poverty threshold, and/or 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 who spend more time below the threshold are less resilient than households who spend less time above the line. The share 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.

162. Figure TA3 shows what the measures look like for the household with the hypothetical sinusoid function shown so far, assuming a quarterly data collection schedule in which food security status is observed at three month intervals. For this household, the reference level of consumption m (shown by the red dashed line) is simply the average of the four points. The intra-annual standard deviation estimated by calculating the standard deviation of the four points, the average of the solid red lines. The range is the difference between the highest of the four values and the lowest, the difference between the dashed black lines. And the share of the period spent below the poverty line is the number of observations that fall below the poverty line (the grey dashed line) divided that by the total number of observations (number of grey dots divided by number of blue dots).

Figure TA1: Examples of capacities over time



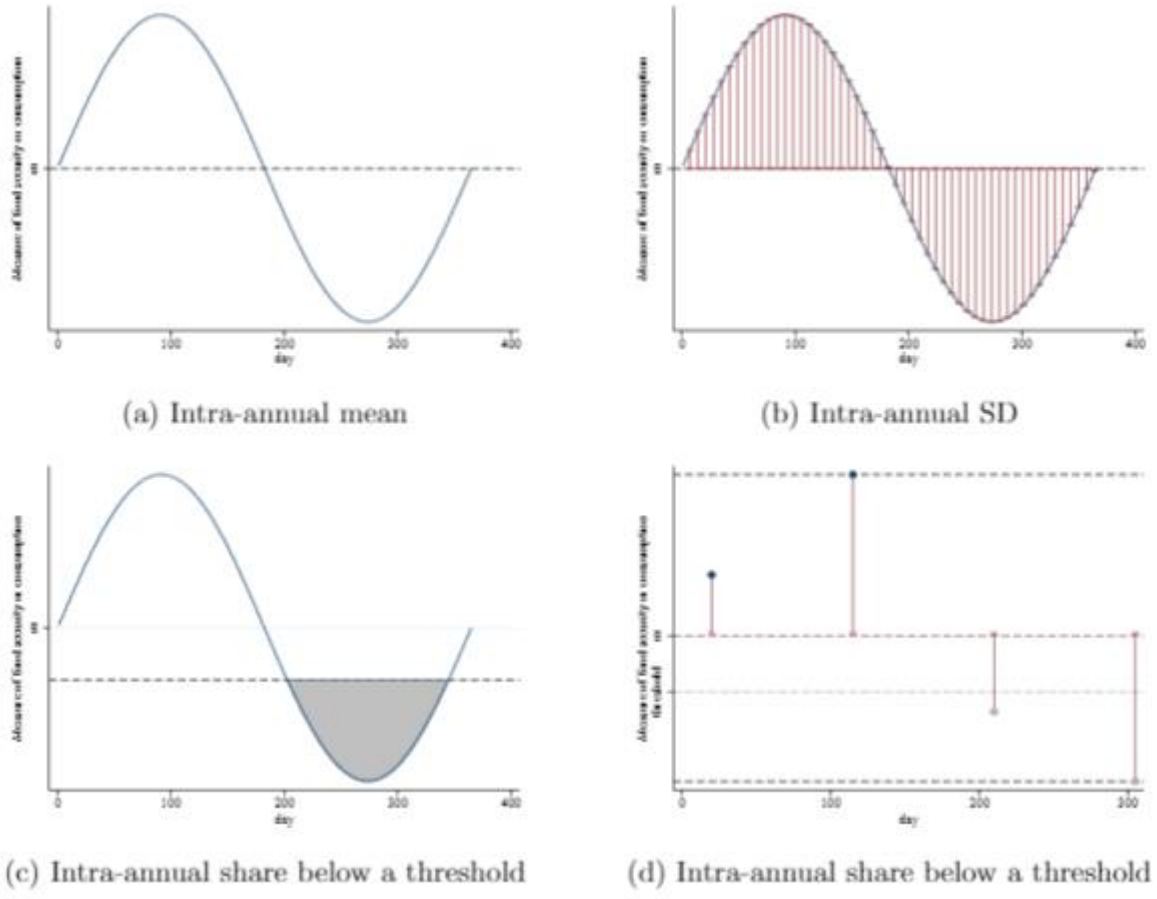
(a) High vs Low α

(b) High vs low δ

(c) Flat vs seasonal welfare

(d) Differential variance of ϵ due to shocks

Figure TA2: Measures of capacities



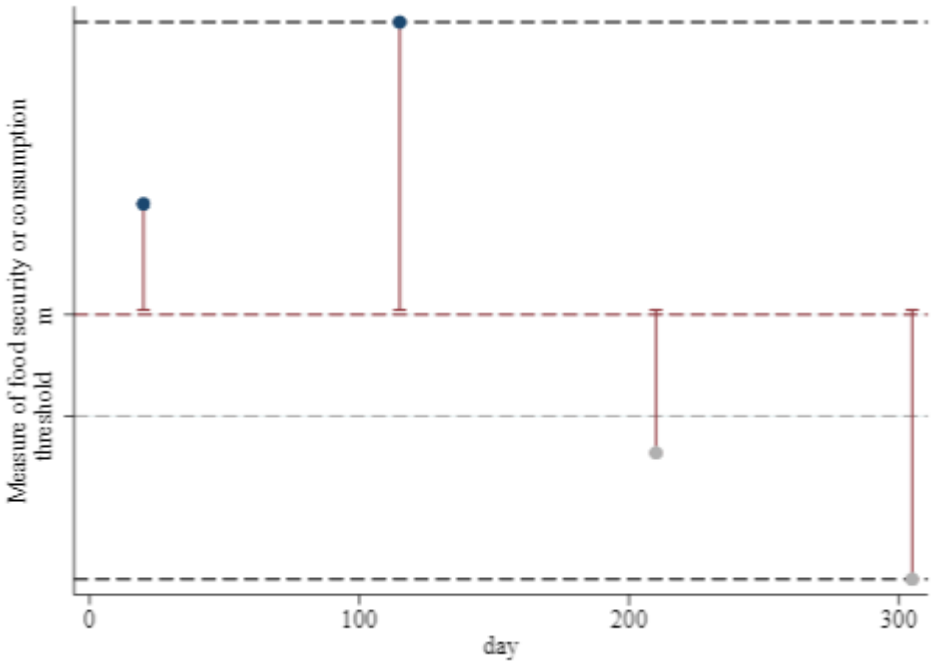
(a) Intra-annual mean

(b) Intra-annual SD

(c) Intra-annual share below a threshold

(d) Intra-annual share below a threshold

Figure TA3: Feasible measurement of capacities



Annex 6: Detailed Stakeholder Analysis

163. 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 WFP country office in Rwanda is intended to be the primary user of this evaluation. In addition, WFP Rwanda provides technical guidance at the national level to inform national policy and dialogue on social protection, and the country office has expressed interest in using the results of this evaluation to support this technical advisory capacity.

164. The various categories of stakeholders include:

- Internal Rwanda-based stakeholders: the Country Director and Deputy Director, the Head of Programme, and all technical and management personnel
- Internal stakeholders outside of Rwanda: the Office of Evaluation, the Regional Bureau for Eastern Africa, and the Cash-Based Transfer and Gender Divisions and Protection Unit at headquarters
- 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 Rwanda
- National stakeholders including national and subnational government actors, and non-governmental organizations.

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

166. 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.

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

Table 10: Stakeholder analysis

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

	including men, women, boys, and girls will be the primary participants of the intervention	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 stake in the results meant to shed light on recommendations for improving gender equality	primary source of data on effectiveness	selection stage	
Government at the 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	Sector-level governance provides technical advice for programme design and is involved in beneficiary selection within communities	At the targeting phase of the intervention	Medium
Government at the district level	District staff play key roles on the steering committee for programming in their jurisdiction and providing support on	District staff influence the prioritization of resources in their district; evaluation results can help to inform their prioritization	The evaluation relies on the activity programming coordination and targeting efforts of district staff	At the targeting phase of the intervention	Medium

	mobilization and targeting of beneficiary villages	efforts in the future			
Government at the central level: National Institute of Statistics Rwanda 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.	WFP has an established relationship with the national Government as a technical advisor for social protection; evaluation results will support these efforts	The evaluation receives national-level clearance before inception	At the initial scoping for the intervention & dissemination of findings	Medium
Local non-governmental organizations: Duhamic ADRI	Duhamic ADRI is an implementing partner for the programme under evaluation	Evaluation results can inform their own livelihood and gender transformation programming	As a cooperating partner, Duhamic ADRI is responsible for ensuring the programme is implemented in line with the evaluation design	At the initial scoping for the intervention & dissemination of findings	High

International non-governmental organizations: Good Neighbors	Good Neighbors is an implementing partner for the programme under evaluation	Evaluation results can inform their own livelihood and gender transformation programming	As a cooperating partner, Good Neighbors is responsible for ensuring the programme is implemented in line with the evaluation design	At the initial scoping for the intervention & dissemination of findings	High
World Bank	Development Impact Evaluation 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 Korea International Cooperation Agency (KOICA)	Primary funder of the intervention	As a user of the evaluation	KOICA is informed at key milestones in the evaluation. They have an interest to use the results as evidence for other funded projects	At the proposal stage of the intervention	Medium

Annex 7: Detailed Evaluation Process

Phase 1 – Preparation	Involved	Estimated Date
Initial discussion between country office and the Office of Evaluation to assess the feasibility	CO/OEV	June 2020
Memorandum of understanding between the Office of Evaluation and country office signed	CO/OEV	June 2020
Set up impact evaluation (IE) team and evaluation committee (EC)	OEV/DIME	July 2020
Agreement on the questions, design, implementation and timelines between country office and impact evaluation team	DIME/OEV/CO	July 2020
Targeting potential intervention sites (including both potential intervention and comparison areas)	CO/DIME	Oct 2020
Phase 2 - Inception report		
Inception report drafted by impact evaluation team, submitted for quality assurance and revisions	DIME/OEV	January 2022
Publication of the inception report	OEV	February 2022
Dissemination of the inception report with country office, regional bureau, evaluation committee, window's reference group, steering committee, online/social media as appropriate	DIME/OEV	February 2022
Phase 3 – Baseline data collection		
Preparation data collection tools, including survey questionnaire, digital devices, sampling strategy, training material, etc.	DIME	November 2020
Pilot and finalization of data collection tools	DIME/CO	November 2020
Recruitment enumerators/data collection firm	CO	November 2020
Enumerators training	DIME/CO	December 2020
Data collection process and live monitoring data quality checks	DIME/CO	December 2020
Phase 4 – Baseline report		
Data analysis and baseline report drafted by impact evaluation team, submitted for quality assurance and revisions	DIME/OEV	June 2021
Publication of the baseline report	OEV	February 2022
Dissemination of the baseline report with survey respondents, country office, regional bureau, evaluation committee (and other evaluation stakeholders), window's reference group, steering committee, online/social media as appropriate	DIME/OEV	March 2022
Phase 5 – Programme implementation		
Randomization	DIME	February 2021
Assignment intervention and comparison sites	DIME/CO	February 2021
Rollout programme activities as per randomization	CO	February 2021
Monitoring programme activities verifying treatment and control status	CO/DIME	February 2021

Phase 6 – Endline data collection		
Preparation data collection tools, including survey questionnaire, digital devices, sampling strategy, training material, etc.	DIME/CO	April 2022
Pilot and finalization of data collection tools	DIME	May 2022
Recruitment enumerators/data collection firm	CO	May 2022
Enumerators training	CO	June 2022
Data collection process and live monitoring data quality checks	DIME	June 2022
Feedback/ data sharing mechanisms, as appropriate/possible		August 2022
Phase 7 – Final evaluation reports		
Gender window		
Data analysis and final evaluation report drafted by impact evaluation team, submitted for quality assurance and revisions	DIME/OEV	November 2022
Publication of the final evaluation report	OEV	January 2023
Dissemination of the final evaluation report with survey respondents, country office, regional bureau, evaluation committee (and other evaluation stakeholders), window's reference group, steering committee, online/social media as appropriate	OEV/DIME/CO	January 2023
Final evaluation report reviewed by post-hoc quality assessment	OEV	January 2023
Resilience window		
Data analysis and final evaluation report drafted by impact evaluation team, submitted for quality assurance and revisions	DIME/OEV	April 2022
Publication of the final evaluation report	OEV	June 2023
Dissemination of the final evaluation report with survey respondents, country office, regional bureau, evaluation committee (and other evaluation stakeholders), window's reference group, steering committee, online/social media as appropriate	OEV/DIME/CO	June 2023
Final evaluation report reviewed by post-hoc quality assessment	OEV	June 2023
Phase 8 – Management response		
Gender window		
Based on findings country office to develop a management response	CO	December 2022
The Office of Evaluation to review and if needed respond to the management response	OEV	December 2022
Publication of the management response	OEV	January 2023
Resilience window		
Based on findings country office to develop a management response	CO	May 2023
The Office of Evaluation to review and if needed respond to the management response	OEV	May 2023
Publication of the management response	OEV	June 2023
Phase 9 – Dissemination and learning		
Gender window		

Webinar presenting the findings	OEV/DIME	December 2022
Blogs, summary briefs, other relevant communication products	OEV/DIME	December 2022
Considerations for academic publication	DIME/OEV	December 2022
Resilience window		
Webinar presenting the findings	OEV/DIME	May 2023
Blogs, summary briefs, other relevant communication products	OEV/DIME	May 2023
Considerations for academic publication	DIME/OEV	May 2023

Annex 8: References

Adjognon, G. S., van Soest, D., & Gutho, J. 2020. "Reducing hunger with payments for environmental services (pes): Experimental evidence from Burkina Faso." *American Journal of Agricultural Economics* 103(3), 831-857

Anderson, M. L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects." *Journal of the American Statistical Association*, 103(484), 1481–1495.

Athey, S., Chetty, R., Imbens, G., & Kang, H. 2016. Estimating treatment effects using multiple surrogates: The role of the surrogate score and the surrogate index. https://arxiv.org/abs/1603.09326

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienťe, W., Shapiro, J., Thuysbaert, B., & Udry, C. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science*, 348(6236), 126-799.

Beaman, L., Chattopadhyay, R., Duo, E., Pande, R., & Topalova, P. 2009. "Powerful women: does exposure reduce bias?" *The Quarterly Journal of Economics*, 124(4), 1497-1540.

de Brauw, A. & Peterman, A. 2020. "Can conditional cash transfers improve maternal health care? Evidence from El Salvador's Comunidades Solidarias Rurales program." *Health Economics*, 29(6), 700–715.

Bursztyn, L., Gonzalez, A. L., & Yanagizawa-Drott, D. 2018. *Misperceived social norms: Female labor force participation in Saudi Arabia*. Working Paper 24736, National Bureau of Economic Research. https://econpapers.repec.org/paper/nbrnberwo/24736.htm

Christian, P. and Dillon, B. 2018. "Growing and Learning When Consumption Is Seasonal: Long-Term Evidence From Tanzania." *Demography* 55(3), 1091-1118

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

Clay, N. & King, B. 2019. Smallholders' uneven capacities to adapt to climate change amid Africa's 'green revolution': Case study of Rwanda's crop intensification program. *World Developmen*t, 116, 1-14.

Dhar, D., Jain, T., & Jayachandran, S. 2018. *Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India*. Working Paper 25331, National Bureau of Economic Research. https://www.nber.org/papers/w25331

Field, E. M., Pande, R., Rigol, N., Schaner, S. G., & Moore, C. T. 2021. *On Her Own Account: How Strengthening Women's Financial Control Affects Labor Supply and Gender Norms. American Economic Review,* 11(7), 2342-75.

Gazeaud, J., Mvukiyehe, E., & Sterck, O. 2021. Cash transfers and migration: Theory and evidence from a randomized controlled trial. *The Review of Economics and Statistics,* 1-45.

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*. Working Paper 25969, National Bureau of Economic Research.

Haushofer, J., Ringdal, C., Shapiro, J. P., & Wang, X. Y. 2019. *Income Changes and Intimate Partner Violence: Evidence from Unconditional Cash Transfers in Kenya*. Working Paper 25627, National Bureau of Economic Research.

Haushofer, J. & Shapiro, J. 2016. "The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya." *The Quarterly Journal of Economics*, 131(4), 1973–2042.

Imbert, C., & Papp, J. 2015. "Labor market effects of social programs: Evidence from India's employment guarantee." *American Economic Journal: Applied Economics*, *7*(2), 233-63.

Jayachandran, S. 2015. "The roots of gender inequality in developing countries." *Annual Review of Economics,* 7(1), 63–88.

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

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.

McKelway, M. 2019. *Vicious and virtuous cycles: Self-efficacy and employment of women in India.* *https://www.arec.umd.edu/sites/arec.umd.edu/files/files/documents/Seminars/2020/McKelway_Dec2019_GSEEmpPaper.pdf*

Premand, P.; Stoeffler, Q. 2020. *Do Cash Transfers Foster Resilience? Evidence from Rural Niger*. Policy Research Working Paper; No. 9473. World Bank, Washington, DC.

Sakho, S., Caruso, G., & Acuña, M. 2020. "Progress without women is not possible." *World Bank Blogs: Latin America and Caribbean*. https://blogs.worldbank.org/latinamerica/progress-without-women-not-possible

World Economic Forum. 2021. Global Gender Gap Report. Geneva: Switzerland. https://www3.weforum.org/docs/WEF_GGGR_2021.pdf

World Food Programme. 2018. *Rwanda: Comprehensive food security and vulnerability analysis.*

World Food Programme. 2019. *Rwanda Country Strategic Plan (2019-2023)* (WFP/EB.2/2018/8-A/8).

World Food Programme, Gender Office. 2019. *The potential of cash-based interventions to promote gender equality and women's empowerment: A multi-country study*. https://www.wfp.org/publications/gender-and-cash-wfp-study

World Food Programme, Office of Evaluation. 2020. *El Salvador: An Evaluation of WFP's Country Strategic Plan (2017–2021).*

Acronyms

AAHRPP	Association for the Accreditation of Human Research Protection Programs
ACLED	Armed Conflict Location and Event Data
CAPI	Computer Assisted Personal Interviewing
CBT	Cash-Based Transfer
CFA	Cash Assistance For Assets
CO	Country Office
C&R	Climate and Resilience
CSP	Country Strategic Plan
DIME	Development Impact Evaluation Unit (World Bank)
FFA	Food assistance for assets
FCS	Food Consumption Score
GEWE	Gender Equality and Women's Empowerment
HDDS	Household Dietary Diversity Score
HDI	Human Development Index
HFC	High Frequency Checks
HHS	Household Hunger Scale
IE	Impact Evaluation
IEQAS	Impact Evaluation Quality Assurance System
IPC	Integrated Food Security Phase Classification
IPV	Intimate Partner Violence
IRB	Institutional Review Board
KOICA	Korean International Cooperation Agency
MDE	Minimum Detectable Effect
OEV	Office of Evaluation (World Food Programme)
PAP	Pre-Analysis Plan
PHQ9	Patient Health Questionnaire-9
PPP	Public and Private Partnerships
RCT	Randomized Controlled Trial
SMART	Sustainable Market Alliance and Asset Creation for Resilient Communities and Gender Transformation
SSL	Secure Sockets Layer
USD	United States Dollar
WASH	Water, Sanitation and Hygiene
WEF	World Economic Forum
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