Baseline Survey Report of the Strengthen PSNP Institutions and Resilience Phase II (SPIR II) Resilience Food Security Activity in Ethiopia

March 2023

IMPEL | Implementer-Led Evaluation & Learning Associate Award

USAID FROM THE AMERICAN PEOPLE

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE IFPRI

Save the Children

ABOUT IMPEL

The Implementer-Led Evaluation & Learning Associate Award (IMPEL) works to improve the design and implementation of Bureau for Humanitarian Assistance (BHA)-funded resilience food security activities (RFSAs) through implementer-led evaluations and knowledge sharing. Funded by the USAID Bureau for Humanitarian Assistance (BHA), IMPEL will gather information and knowledge in order to measure the performance of RFSAs, strengthen accountability, and improve guidance and policy. This information will help the food security community of practice and USAID to design projects and modify existing projects in ways that bolster performance, efficiency, and effectiveness. IMPEL is a seven-year activity (2019-2026) implemented by Save the Children (lead), TANGO International, Tulane University, Causal Design, Innovations for Poverty Action, and the International Food Policy Research Institute.

RECOMMENDED CITATION

Gilligan, D., Hirvonen, K., Leight, J. Tambet, H., Tesfaye, H., IMPEL. (2023). *Baseline Survey Report of the Strengthen PSNP Institutions and Resilience Phase II (SPIR II) Resilience Food Security Activity in Ethiopia.*. Washington, DC: The Implementer-Led Evaluation & Learning Associate Award

PHOTO CREDITS

Save the Children/Jordi Matas

DISCLAIMER

This report is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the Implementer-Led Evaluation & Learning (IMPEL) award and do not necessarily reflect the views of USAID or the United States Government.

CONTACT INFORMATION

IDEAL Activity
c/o Save the Children
899 North Capitol Street NE, Suite #900
Washington, DC 20002
www.fsnnetwork.org
info@fsnnetwork.org

PREPARED BY:

ACKNOWLEDGMENTS

This report was prepared by Daniel Gilligan, Jessica Leight, and Heleene Tambet of IFPRI's Poverty Health and Nutrition Division and by Kalle Hirvonen and Haleluya Tesfaye of IFPRI's Development Strategy and Governance Division.

This study was made possible by the generous support of USAID Bureau of Humanitarian Assistance, award No. 72DFFP19LA00003 and sub-award No. 999004127. We thank the study expert advisory panel Sarah Baird, John Hoddinott, Girum Abebe Tefera, and Reimar Macaranas for their helpful feedback on the study protocol. We thank Mara Mordini, Adam Trowbridge, Hilary Cook, and Takawira Kapikinyu for their guidance and feedback on the study design and implementation. We thank Mohamoud Ahmed and Tom Spangler for managing the project work plan and coordinating with the project donor team. Thank you to Michael Mulford, Retta Gudisa, and the full SPIR II implementation team for their guidance and comments on this report. Finally, we thank Tigabu Getahun and the EconInsight field team for their work conducting interviews and collecting baseline data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

BHA	Bureau for Humanitarian Assistance
CAPI	Computer Assisted Personalized Interview
CPI	Consumer Price Index
cRCT	Cluster Randomized Control Trial
CREDI	Caregiver Reported Early Development Instruments
F&M	Female and Male
FCS	Food Consumption Score
FIES	Food Insecurity Experience Scale
FNM	Female No Male
HDA	Health Development Army
HAZ	Height for Age Z score
HCES	Household Consumption Expenditure Survey
HDDS	Household Diet Diversity Score
HEW	Health Extension Worker
HFA	Humanitarian Food Assistance
IFPRI	International Food Policy Research Institute
IMPEL	Implementer-Led Evaluation and Learning
IYCF	Infant and Young Child Feeding
MAD	Minimally Acceptable Diet
MDD-C	Minimum Diet Diversity - Child
MDD-W	Minimum Diet Diversity - Woman
MDE	Minimum Detectable Effect
MMF	Minimum Meal Frequency
NCG	Nurturing Care Group
PPP	Purchasing Power Parity
PSNP	Productive Safety Net Program
RFSA	Resilience Food Security Activity
SBCC	Social Behavior Change Communication
SPIR	Strengthen PSNP Institutions and Resilience
TUP	Targeting the Ultra-Poor
UNICEF	United Nations Children's Fund
UN-OCHA	United Nations Office for the Coordination of Humanitarian Affairs
USAID	United States Agency for International Development
USD	United States Dollars
USG	Unites States Government
VESA	Village Economic and Social Association
WASH	Water, Sanitation, and Health
WHO	World Health Organization
WHZ	Weight-for-height Z-score

EXECUTIVE SUMMARY

Overview and Study Design

The objective of this report is to present results from the baseline survey conducted as part of the Implementer-Led Evaluation and Learning (IMPEL) evaluation of SPIR II, a randomized controlled trial launched in 2022. The second phase of the Strengthen PSNP Institutions and Resilience (SPIR) Resilience Food Security Activity (RFSA) aims to enhance livelihoods, increase resilience to shocks, and improve food security and nutrition for rural households vulnerable to food insecurity in Ethiopia. The RFSA is situated within Ethiopia's Productive Safety Net Program (PSNP), one of the largest safety net programs in Africa. Funded by USAID's Bureau for Humanitarian Assistance (BHA), SPIR II is implemented by World Vision International (lead), CARE, and ORDA in the Amhara and Oromia regions of Ethiopia.

The IMPEL SPIR II impact evaluation employs an experimental design with three arms, comparing two treatment combinations of livelihood and nutrition graduation model programming provided to PSNP beneficiaries relative to a control group receiving only PSNP transfers. The treatment assignment is randomized at kebele level in 234 kebeles. In the first arm (the control group), PSNP is implemented by the government with SPIR II support for the provision of cash and food transfers only (no supplemental programming). In the second arm, SPIR II programming is rolled out to PSNP beneficiary households in conjunction with nurturing care groups (NCGs) targeting enhanced infant and young child nutritional practices. In the third arm, PSNP beneficiary households receive SPIR II programming and NCGs, supplemented with additional targeted cash grants to pregnant and lactating women.

Data, Outcomes, and Sample

The evaluation includes three primary rounds of data collection. The baseline survey, conducted in August–September 2022, entailed interviews with both the primary female respondent (pregnant or lactating woman) and her spouse. The midline survey is scheduled for August–September 2023, around 12 months following the baseline. All households will be revisited and interviews will be conducted with the primary female respondent only. The endline survey is scheduled for August–September 2025, around twenty-four months after the midline. All households will be revisited and interviews will be conducted with both the primary female respondent (pregnant or lactating woman at baseline) and her spouse.

The primary outcomes for the evaluation were defined in a pre-analysis plan before the baseline survey (Appendix C). For livelihoods and gender-related outcomes, the primary outcomes are daily household per-capita consumption expenditure, the total value of productive assets, the total value of livestock assets, and savings (both binary and continuous variables). For nutrition-related outcomes, the primary outcomes are height-for-age (measured among children 30–48 months at endline), the prevalence of children 6–23 months consuming a diet of minimum diversity at midline, and infant and young child feeding knowledge.

The baseline survey included a sample of 3,015 households in 234 kebeles from 15 woredas in Amhara and Oromia regions. The inclusion criteria required that households were PSNP beneficiaries reporting

the presence of an infant under 9 months of age (along with his/her mother or primary caretaker) or a pregnant woman. The realized sample corresponds to 98% of the target sample. Three kebeles were excluded from due to insecurity or other challenges that rendered it impossible to conduct surveys. Summary statistics for the baseline survey suggest the sample for this evaluation is characterized by a high level of food insecurity and poverty and generally poor nutrition for women and young children. Here, we provide a brief overview of the key conclusions from the most important sections of the survey.

Household Demographic Characteristics

Sample households are characterized by an average size of five individual.; 52% of household heads report some formal education and 87% report that their primary economic activity is crop production. The primary female in the sample households is on average 29 years of age and 93% are married. Only 43% of primary females report any formal education and, on average, the level of education is low: only 26% report 3 to 7 years of education and only 7% report more than 8 years of education. The low level of maternal education is an important contextual factor for the interventions targeting enhanced child nutritional status given the evidence from the literature that maternal education is an important determinant of this variable.

Food Security and Poverty

Households in this sample face significant food security challenges, although the extent of food insecurity depends on the metric applied. Using the Food Insecurity Experience Scale, 31% of households are moderately food insecure, while 57% are severely food insecure. Using the Food Consumption Score, 75% report an acceptable food consumption score. Extreme poverty is high, with more than 71% of sampled individuals originating from households for which the daily per capita consumption is below the international poverty line ($1.90 PPP per capita per day). The mean daily household per capita food consumption-expenditures in this sample is 50.73 birr and the mean daily per capita non-food expenditures is 7.80 birr. Therefore, the average household in this sample spends 86% of its total budget on food, indicative of a high level of deprivation. The significant challenges related to food security may be related to high exposure to recent shocks, particularly conflict-related shocks in Amhara, as summarized in the subsection below.

Children's Nutritional Status and Women's Well-Being

Consistent with evidence from other surveys conducted among similar populations in Ethiopia, levels of dietary diversity are extremely low, though adherence to exclusive breastfeeding is generally high. Among children less than 24 months of age, 95% were breastfed at some point in their life. About 79% of the children less than 6 months of age were exclusively breastfed. More than 93% of the children 0–1 months of age in our sample were exclusively breastfed at the time of the interview with the share falling to 81% and 61% for children in the 2- to 3-month and 4- to 5-month age groups, respectively. As also documented in other recent research in Ethiopia, the introduction of complementary foods in these areas is delayed: 24% of the children 6–8 months old did not consume solid, semi-solid, or soft foods on the day before the interview. Only 3% of children 6–23 months are identified as receiving a minimum

acceptable diet. Among women, only 7% of women of reproductive age are identified as consuming a diet characterized by minimum dietary diversity.

Cash Earnings, Decision-Making, and Credit Access

With respect to reported cash earnings, slightly more than half (54%) of men and women report earning cash, but this rate is dramatically higher among men than women. Among women who report earning cash, 68% report participation in decisions about its use; 61% report that they can participate in decisions about the use of a spouse's self-earned cash. Among men, 80% report that their spouse/partner can participate in decisions around the use of cash. Fifty-two percent of women and men in a union report access to credit and 67% of those who report access to credit report that they can make decisions about credit. Seventy-five percent of men and women report they are members of a community group.

Exposure to Recent Shocks

There is also evidence that this population has been heavily affected by recent adverse shocks, including conflict and drought. Weather-related shocks were common in both regions, but were especially prevalent in Oromia, affecting 95% of households, while weather shocks affected 55% of households in Amhara. Conflict was a major event in Amhara, affecting more than three out of every four households. Biological shocks, including crop and livestock diseases and pests, as well as human diseases and infections, afflicted 42% of households in Amhara and more than 51% in Oromia.

Summing Up

In addition to reporting general summary statistics, the baseline survey can be used to assess balance in covariates across treatment arms. In general, the randomization achieved balance. There is no evidence of any significant differences in observable characteristics across arms.

This baseline survey will serve as the foundation for the randomized controlled trial. It will be followed by a midline survey (2023) and an endline survey (2025) that will be used to assess the impact of the interventions of interest on primary and secondary outcomes. Details about the planned analysis of follow-up outcomes are also provided in this report.

This page is left intentionally blank.

1. INTRODUCTION

Over the past two decades, social safety net programs have become a mainstream policy tool across sub-Saharan Africa to address food insecurity and extreme poverty (Beegle et al., 2018). Since the turn of the millennium, the number of social safety net programs has doubled (Hickey et al., 2018) and today, each country in the region operates at least one major safety net program (Beegle et al., 2018).

There is now strong evidence from a wide variety of contexts showing that these programs can be effectively used to improve food security and increase asset accumulation (Andrews et al., 2018; Hidrobo et al., 2018). Building on this evidence, there is a growing interest in using social safety net programs as a platform to achieve broader objectives over and above food security, including the reduction of poverty and the enhancement of resilience. One such objective relates to graduation: moving households away from long-term support and enabling them to build resilient and self-reliant livelihoods (Sabates-Wheeler et al., 2021). A second objective centers around rendering existing programs nutrition-sensitive by integrating transfers with nutrition and water, sanitation, and hygiene (WASH) training and other interventions to better address the underlying causes of maternal and child malnutrition (Ruel & Alderman, 2013).

Against this backdrop, the second phase of the Strengthen PSNP Institutions and Resilience (SPIR) Resilience Food Security Activity (RFSA) aims to enhance livelihoods, increase resilience to shocks, and improve food security and nutrition for rural households vulnerable to food insecurity in Ethiopia. The RFSA is situated within Ethiopia's Productive Safety Net Program (PSNP) program, one of the largest safety net programs in Africa. Funded by USAID's Bureau for Humanitarian Assistance (BHA), SPIR II is implemented by World Vision International (lead), CARE, and ORDA in the Amhara and Oromia regions of Ethiopia.

The Implementer-Led Evaluation and Learning (IMPEL) SPIR II impact evaluation employs an experimental design with multiple treatment arms comparing combinations of livelihood and nutrition graduation model programming provided to PSNP beneficiaries relative to a control group receiving only PSNP transfers. The design includes 234 kebeles assigned to three treatment arms. In the first arm (the control group), PSNP is implemented by the government with SPIR II support for the provision of cash and food transfers only (no supplemental programming). In the second arm, SPIR II programming is rolled out to PSNP beneficiary households in conjunction with nurturing care groups (NCG) targeting enhanced infant and young child nutritional practices. In the third arm, PSNP beneficiary households receive SPIR II programming and NCGs, supplemented with additional targeted cash grants to pregnant and lactating women.

This impact evaluation will contribute to current evidence on the effectiveness of graduation model programs by adding new evidence on graduation models, including substantial nutrition programming and whether the effectiveness of such a model is enhanced by a supplemental maternal cash grant. There is considerable evidence—from an evaluation of a six-country study of programs designed like BRAC's Targeting the Ultra-Poor program and related studies—that graduation models improve household economic outcomes, including consumption, food security, assets, financial inclusion, labor supply, and income, as well as some measures of mental health related to stress (Banerjee et al. 2015; Banerjee et al. 2016; Bandiera et al. 2017; Banerjee et al. 2019). A gap in this literature is whether

graduation models can also be designed to address maternal and child malnutrition and thereby potentially expand the long-term benefits of the program through investments in human capital. The impact evaluation of the first phase of the SPIR activity (2016–2021), which included integrated livelihood and nutrition programming, found impacts on women's nutrition knowledge and mother's diets, but limited impacts on child diets and virtually no impacts on child anthropometry (Alderman et al 2022; Alderman et al. 2021). Results of the evaluation suggested that an important factor limiting activity impact on child nutrition was the approach for nutrition behavior change communication, which relied on community health facilitators who were unable to achieve sufficiently intensive interaction with SPIR I households.

The innovations in nutrition-related programming that are a focus of this evaluation involve improving infant and young child feeding (IYCF) practices, particularly suboptimal complementary feeding practices that have been widely speculated to be inhibiting child growth and development in Ethiopia (Golan, et al., 2019). The NCG model, pioneered by World Vision in a wide range of other contexts, is based on groups of 10–15 community-based trained volunteer agents who cascade down social behavior change communication (SBCC) messages and activities to caregiver groups at the community level. Non-experimental studies conducted in other contexts suggest that the model can significantly increase SBCC contact rates and improve IYCF practices and child growth outcomes (Davis et al., 2013). However, large-scale experimental evidence on the relative effectiveness of this strategy relative to standard government-led nutrition programming (provided in T1) is largely unavailable, rendering this evaluation a meaningful contribution. In addition, improving caregiver knowledge may not be sufficient to improve complementary feeding practices if households cannot afford to purchase nutritious foods. Therefore, the third study arm introduces maternal grants of $20 per month during the child's first 24 months of life to relax possible financial constraints to child feeding.

This multi-arm cluster randomized control trial (cRCT) design permits the research team to evaluate the causal impact of both livelihood and nutrition graduation programming in SPIR II, while also testing the effectiveness of the NCG model and an experimental maternal cash grant.

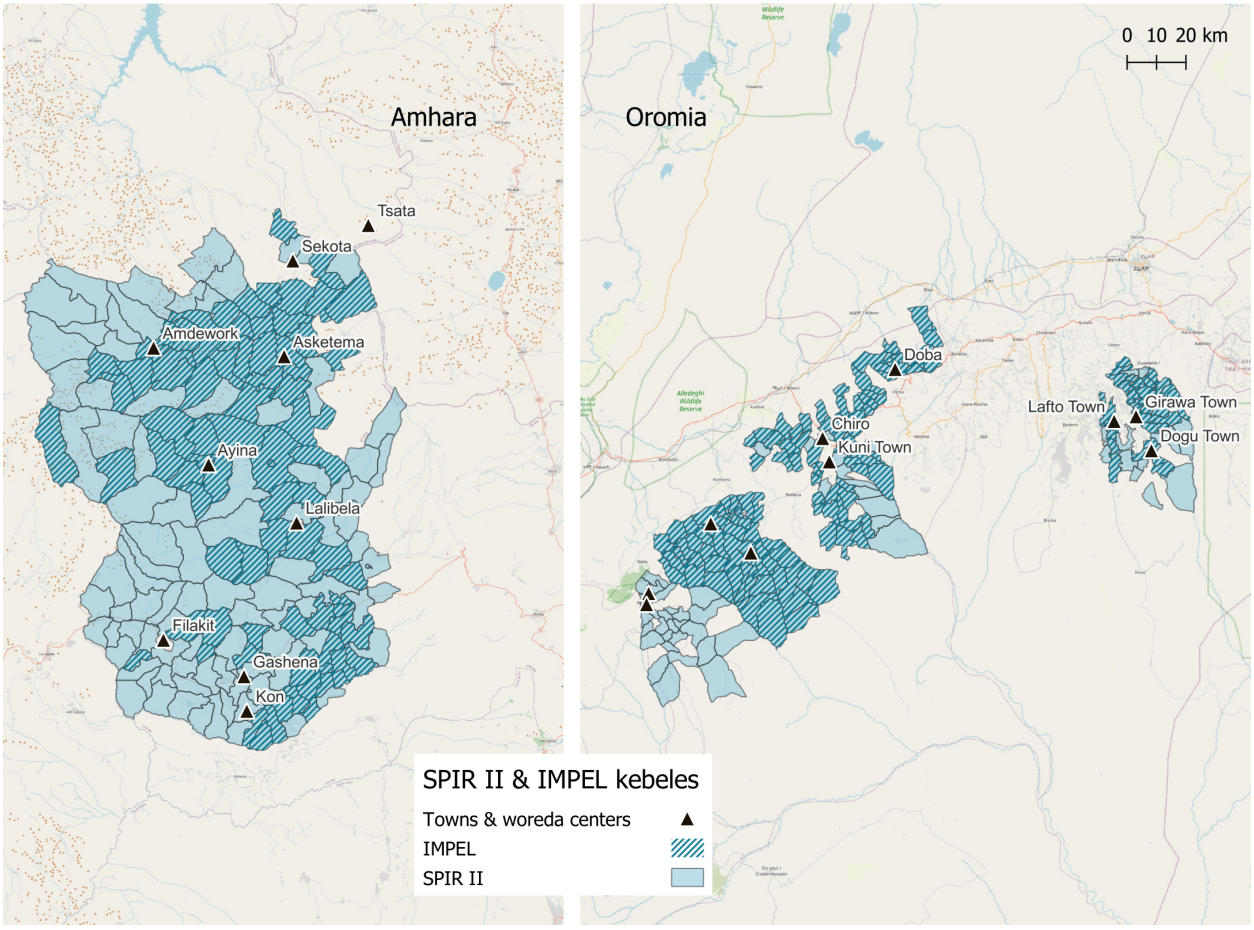
The baseline survey for this impact evaluation was conducted in August–September 2022. This baseline report will provide an overview of the evaluation design, the survey conducted, and the key findings. These findings allow us to characterize the sample and verify the balance on observable characteristics across experimental arms.

2. METHODOLOGY

2.1 Evaluation Design

This evaluation is a cRCT where the cluster is defined as the *kebele* (the lowest administrative level in Ethiopia). The target sample included 237 kebeles in Amhara and Oromia in which SPIR II is operational. This sample comprised kebeles that were not included in the previous impact evaluations conducted by the International Food Policy Research Institute (IFPRI) and World Vision under SPIR, which includes kebeles that were served by SPIR I but were not included in the SPIR I impact evaluation (generally because programming had already been initiated) and new kebeles. In addition, a small number of eligible kebeles were excluded from the study due to security concerns at the time of the baseline survey in 2022. Figure 1 shows the SPIR II kebeles and the IMPEL sample kebeles. The full operational area served by SPIR II includes 17 woredas and 465 kebeles.

Figure 1: SPIR II and IMPEL sample kebeles



Note: The implementation area map is incomplete because the shapefiles are missing 47 SPIR II kebeles, of which six are part of the IMPEL study.

2.1.1 Interventions

Core SPIR II graduation programming includes the organization of village economic and social associations (VESAs), used as a platform for trainings and other RFSA activities around financial literacy, promotion of savings and credit use, agriculture, livestock value chain development (e.g., developing business skills and production skills), improving social capital, and catalyzing women's empowerment. A subset of households (33% of IMPEL households) is also targeted for a one-time $300 livelihoods grant. These are the households identified as the poorest using an asset-based welfare index constructed by the IFPRI team from baseline survey data.[1]

The nutrition programming centers around the provision of integrated nutrition SBCC as well as WASH activities. The first major goal of this evaluation is to assess the effectiveness of integrated SPIR II graduation model programming—the bundle of interventions described above—on a range of outcomes at the household level, including livelihoods-related outcomes and graduation from poverty.

In addition, a particular focus of this evaluation is innovations in nutrition-related programming centered around enhanced IYCF practices, particularly suboptimal complementary feeding practices that have been widely speculated to be inhibiting child growth and development in Ethiopia (Golan et al., 2019). The NCG model, pioneered by World Vision in a wide range of other contexts, is based on groups of 10–15 community-based trained volunteer agents who cascade down SBCC messages and activities to caregiver groups at the community level. Non-experimental studies conducted in other contexts suggest that the model can significantly increase SBCC contact rates and improve IYCF practices and child growth outcomes (Davis et al., 2013). However, large-scale experimental evidence on the effectiveness of this strategy relative to standard government-led nutrition programming is largely unavailable, rendering this evaluation a meaningful contribution. (Standard government-led nutrition programming in Ethiopia is delivered primarily by health extension workers and the Health Development Army (HDA) through relatively infrequent interactions to provide nutritional information and counseling to pregnant women and mothers of young children.)

Improving caregiver knowledge may not be sufficient to improve complementary feeding practices if households cannot afford to purchase nutritious foods. Therefore, the third study arm introduces maternal grants of $20 per month during the child's first 24 months of life to relax possible financial constraints to child feeding. These grants are benchmarked relative to household-level consumption as observed in the SPIR I endline survey conducted among a sample of PSNP beneficiary households in early 2021. Consumption was around $100 per month per household and thus the transfer represents around a 20% bump in consumption. This is consistent with evidence in the literature that the most successful extant conditional cash transfer programs generally provide transfers of between 10% and 20% of household consumption (Fiszbein et al., 2009).[2]

Table 1 summarizes the interventions across each treatment arm.

[1] At the kebele level, only around 10% of households outside of the IMPEL sample are eligible for livelihoods grant. This eligibility process was determined separate from the IMPEL study and is not described here.
[2] The relevant evidence can be found in Table 2 of the cited report.

Table 1: Interventions by treatment arm

Treatment Arm		Targeted Nutritional Interventions
T1	PSNP only	N/A
T2	PSNP transfers VESAs (used as a platform for other RFSA activities around financial literacy, promotion of savings and credit use, agriculture, livestock value chain development, improving social capital, and catalyzing women's empowerment) Nutrition SBCC Livelihoods grants for poorest 33% of households (identified using asset index)	Nurturing care groups
T3	PSNP transfers VESAs (used as a platform for other RFSA activities around financial literacy, promotion of savings and credit use, agriculture, livestock value chain development, improving social capital, and catalyzing women's empowerment) Nutrition SBCC Livelihoods grants for poorest 33% of households (identified using asset index)	Nurturing care groups and maternal grants

2.1.2 Study Arms

To summarize, the impact evaluation includes three study arms:

- T1. PSNP implemented by the government with SPIR II support for the provision of cash and food transfers only (no supplemental programming) (79 kebeles)
- T2. PSNP + SPIR II + NCG model (79 kebeles)
- T3. PSNP + SPIR II + NCG + maternal grants (79 kebeles)

The study arm T1 serves as a control group of PSNP households against which the impacts of SPIR II programming will be measured. PSNP households in arm T2 benefit from SPIR II resilience programming and will be exposed to the NCG intervention. The study arm T3 receives the same intervention package as households in T2 but also benefits from the maternal grants.

This multi-arm cRCT design permits us to evaluate the causal impact of both livelihood and nutrition graduation programming in SPIR II. First, comparing outcomes in T1 to T2 and T3 permits us to quantify the causal impact of the SPIR II livelihood graduation programming on outcomes such as financial inclusion, assets, consumption, resilience, and poverty. Second, by experimentally varying the nutrition interventions, the study will provide valuable information on nutrition-sensitive programming within the PSNP. More specifically, by comparing outcomes across all three treatment arms, we can assess the

relative effectiveness of the NCG intervention on IYCF practices and child growth outcomes (e.g., child stunting prevalence) with and without an added maternal grant to reduce cost constraints to improving these outcomes.

2.1.3 Sampling

The target sample for this evaluation before the baseline launch was 3,081 households, drawn from 237 kebeles in 15 woredas. The inclusion criteria for households to be included in the evaluation were defined as follows.

i) The household must be enrolled as a PSNP beneficiary in a target kebele.
ii) The household must meet one of the following criteria:
    a) There is a pregnant woman present who self-reports pregnancy, with an estimated gestational age that is at least 3 months (i.e., completion of the first trimester).
    b) There is an infant present aged less than 9 months as of the date of the interview and the infant's mother or primary caretaker is also a resident in the household.

The target for sampling was 13 households per kebele: seven households including a pregnant woman, and six households including a child under 9 months. The sampling targets were informed by available resources for data collection as well as the goal of achieving adequate statistical power for the key comparisons of interest (more details about power calculations are provided below). In case the target sample for one of the two subgroups (pregnant women or households with a child less than 9 months) was not reached, the plan was to substitute a household from the other subgroup, if available. The target kebeles reported an average of 261 PSNP beneficiary households based on updated sample lists. Accordingly, the evaluation targeted around 5% of beneficiary households.

Randomization was conducted by the research team using Stata, in June 2022. The randomization process proceeded as follows. We constructed strata based on the interaction of the following characteristics: woreda; a binary variable for whether a kebele is above or below the woreda-level median in the percentage of households eligible for the PSNP; and a binary variable for whether the kebele is above or below the woreda-level median in distance from the woreda capital. The first binary variable is interpreted as a proxy for the local poverty level and the second binary variable is interpreted as a proxy for the relative remoteness of the kebele. Constructing these strata allows us to identify the effects of the interventions of interest within strata that share common characteristics in terms of geography (woreda), poverty, and remoteness.

We also evaluate balance across treatment arms for a broad set of household covariates reported in this baseline survey report (see Section 3.10), including baseline levels of all primary outcomes of interest.

2.1.4 Surveys

The evaluation includes three rounds of primary data collection.

i) The baseline survey was conducted in August–September 2022. The baseline survey entailed interviews with both the primary female respondent (pregnant or lactating woman) and her spouse.

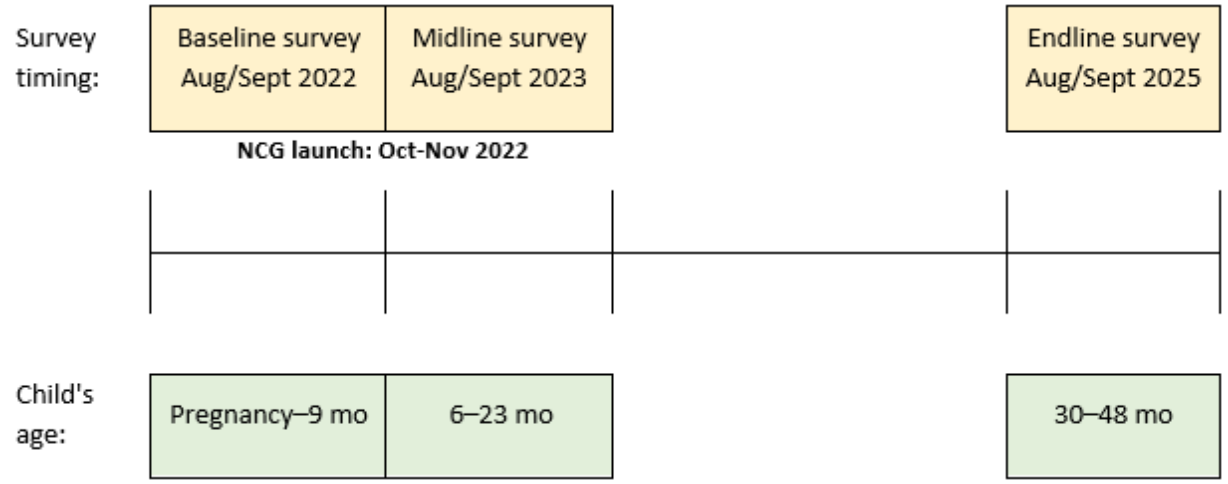
ii) The midline survey is scheduled for August–September 2023, around 12 months following the baseline. All households will be visited and interviews will be conducted with the primary female respondent only. Child anthropometric measurements will also be collected.

iii) The endline survey is scheduled for August–September 2025, around 24 months following the midline. All households will be visited and interviews will be conducted with both the primary female respondent (pregnant or lactating woman) and her spouse.

The timing of the survey rounds plays a critical role in the evaluation of the nutrition interventions. The baseline survey in August–September 2022 sampled PSNP households with a pregnant woman or a child less than 9 months of age (Figure 2). Before the survey, we randomly allocated kebeles into the three study arms.

We will administer a midline quantitative survey to the entire baseline sample 1 year after the baseline survey, in August–September 2023, when the children are between 6–23 months of age, allowing us to assess both contact rates and participation in NCGs (process indicators) and acquired knowledge about and adherence to recommended IYCF practices (outcome indicators) (WHO and UNICEF, 2021). The timing of the midline survey has several advantages, including (i) allowing timely measurement of the impact of SPIR II on child diets and IYCF practices, (ii) inclusion of process monitoring questions for the full sample around household participation in SPIR II activities and access to SPIR II components, and (iii) measurement of intermediate outcomes related to food security to examine progress against RFSA objectives.

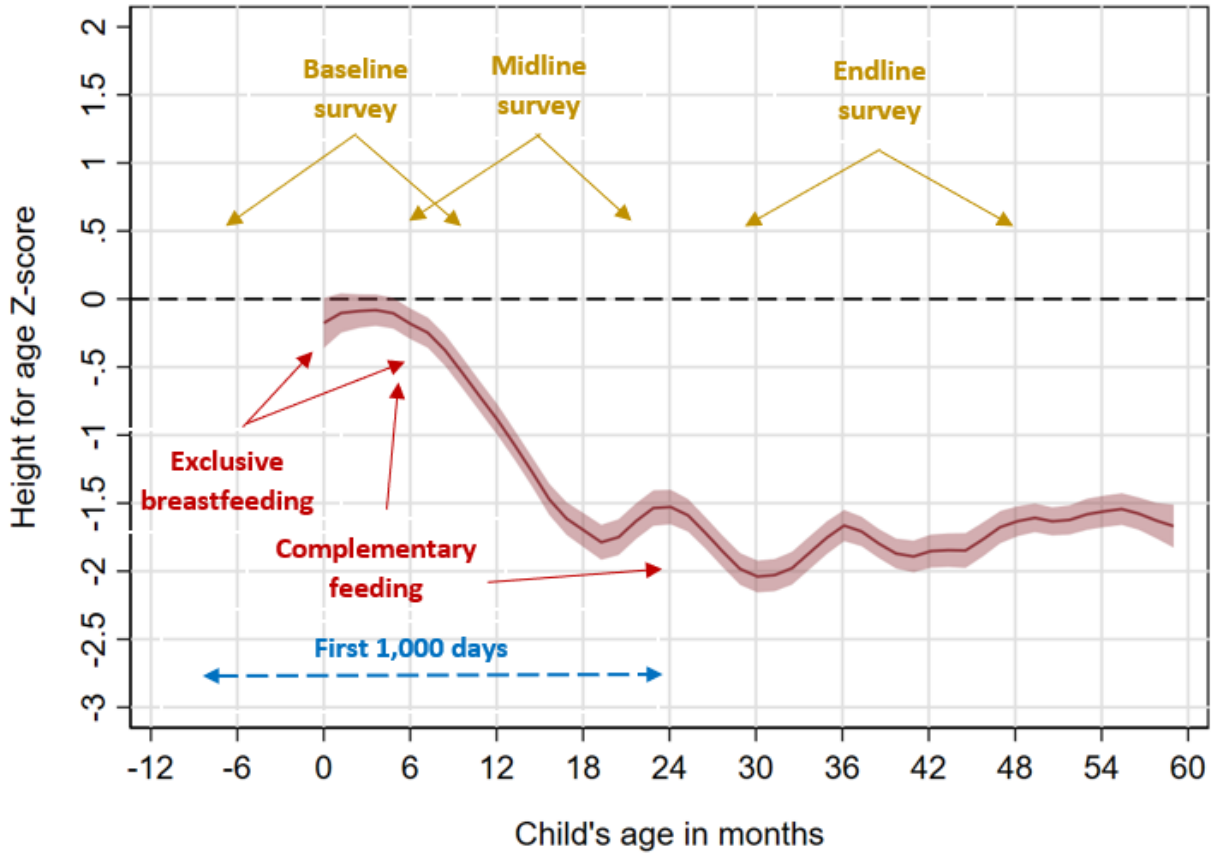
Figure 2: Timing of the surveys



The endline survey is planned for 2025, when the children will be 30–47 months. This survey will be the primary round in which we can measure effects on livelihood outcomes and is an ideal time to assess impacts on child growth outcomes (both outcomes specified in more detail below) (Alderman & Headey, 2018). Child growth faltering (measured using child height-for-age Z-scores (HAZ)) in Ethiopia and other low-income country settings largely occurs during the first 1000 days of life (Golan et al., 2019; Victora et al., 2010). As shown in Figure 3 based on the data from the 2015/16 Ethiopia Demographic and Health Survey (DHS), during the first 4 to 5 months of life, the height of the average Ethiopian child is similar to the height of the median child in the World Health Organization (WHO)-2006 growth

reference.[3] Golan et al. (2019) hypothesized that child growth during this period is supported by the relatively high adherence to exclusive breastfeeding in Ethiopia. Rapid growth faltering begins at around 6 months of age when children should be introduced to complementary foods and continues until about 18 months of age. The endline occurs after this period of rapid growth faltering. By that point, children and caregivers in study arms T2 and T3 will have been exposed to the intensive SPIR II nutrition programming for much of the critical first 1000-day period. The study hypothesis is that this nutrition programming prevents growth faltering during the first 1000 days and as a result, at the endline the HAZ curve of the average child in the treatment arms lie above the HAZ curve of the average child in the control arm.

Figure 3: Timing of the surveys in relation to typical linear growth faltering in Ethiopia



Note: Local polynomial regression based on Ethiopia 2015/16 Demographic and Health Survey (DHS). The shaded areas represent 95-percent confidence intervals. N = 8,771 children 0–59 months of age.

Table 2 below summarizes the key questionnaire modules to be included in each wave of data collection. The planned modules for midline and endline, however, may require revision based on resource constraints.

[3] HAZ measures the height difference to the median child in the WHO-2006 growth reference sample. This difference is measured in terms of standard deviations. Thus, the HAZ of the median child in the growth reference is 0. In Figure 3, this is marked with the dashed horizontal line (HAZ=0). Child is defined as stunted if her HAZ<-2 and severely stunted if HAZ<-3.

It is important to note that all survey rounds will be targeted at the full set of households sampled at baseline. Attrition will be minimized by multiple follow-ups with target households at midline and endline and tracking as needed. However, we anticipate very low levels of attrition given our previous experience with panel surveys of PSNP households. In particular, PSNP households are characterized by generally low levels of mobility given their low levels of income and receipt of PSNP benefits. In the SPIR I evaluation, we observed attrition of only 7% over a four-year period.

We cannot rule out that some households may be excluded from the PSNP during retargeting processes conducted during the sampling period, though our understanding based on discussions from World Vision is that the annual retargeting process would be minimal and would not be expected to shift the status of more than 5% of households. Given that our sample households may be particularly vulnerable with the presence of a pregnant and lactating woman and infant, they may be less likely to exit the PSNP. That being said, we will document households' participation in the PSNP and in particular dimensions of SPIR II in both follow-up surveys.

Table 2: Questionnaire modules, by survey round

Module	Respondent	Baseline	Midline	Endline
Module A. Household Identification and Consent	Household head (primary female at midline)	X	X	X
Module B. Household Roster		X	X	X
Module C1. Access to PSNP, Humanitarian Food Assistance (HFA), VESAs, & Savings Institutions				
Part 1: Past and Current Access to PSNP		X		X
Part 2: PSNP Payments since January		X		X
Part 3: Livelihoods Component		X		X
Part 4: Other Public Transfers (HFA)		X		X
Part 5: Participation in VESAs and SPIR II Activities		X		X
Part 6: Savings and Access to Savings Institutions		X		X
Part 7: Access to Health Insurance		X		X
Module C2. Access to Nurturing Care Groups and Maternal Grants	Primary female		X	X
Module D. Paternal IYCF Knowledge and Perceptions	Primary male	X		X
Module E. Agriculture	Primary male	X		X
Module F. Household Assets	Primary male			
Part 1: Productive Assets		X		X
Part 2: Consumer Durables		X		X
Part 3: Livestock Ownership		X		X
Module G. Gender (Cash)	Primary male	X		X
Module H. Gender Access to Credit and Group Participation	Primary male	X		X
Module I. Poverty Measurement (including FCS and FIES)				
Part 1: Durables and Services (annual)	Most knowledgeable	X		X
Part 2: Household Non-Food Consumables (monthly)	Most knowledgeable	X		X
Part 3: Food Consumption and Expenditure	Most knowledgeable	X		X

Module	Respondent	Baseline	Midline	Endline
Part 4: Food Security in the last 12 Months	Primary female	X	X	X
Part 5: Food Prices in the Locality	Most knowledgeable	X		X
Module J. Water, Sanitation and Hygiene (WASH)	Primary female	X		X
Module K. Children's Nutritional Status and Feeding Practices	Primary female			
Part 1: Infant and Young Child Feeding		X	X	X
Part 2: Child Anthropometrics			X	X
Part 3: Early childhood development (Caregiver Reported Early Development Instruments (CREDI) at midline, TBD at endline)[4]			X	
Module L. Women's Health, Nutritional Status, Dietary Diversity and Nutrition knowledge	Primary female			
Part 1: Women's Dietary Diversity and Access to Antenatal Care		X	X	X
Part 2: Maternal IYCF Knowledge and Perceptions		X	X	X
Part 3: Exposure to Health and Nutrition Services		X	X	X
Part 4: PSNP during Pregnancy and Lactation		X	X	X
Part 5: Maternal Anthropometrics			X	X
Part 6: Early Childhood Development			X	X
Module M. Gender (Cash)	Primary female	X		X
Module N. Gender Access to Credit and Group Participation	Primary female	X		X
Module O. Resilience	Primary female			
Part 1. Shocks and Stressors		X		X

Note: The table notes the target respondent. Enumerators can substitute another knowledgeable individual if the target respondent is not available or declines to answer the section. The modules are presented in the order that they will be administered in the baseline survey.

[4] The CREDI is designed for children under 30 months and thus would not be appropriate at endline. We may explore the feasibility of using the Malawi Developmental Assessment Tool at endline, given that it's valid for children up to age six.

2.1.5 Econometric Analyses

All regressions will be estimated using Ordinary Least Squares. Since geography and environmental factors are strong predictors of poverty and food security (including child growth outcomes) in low-income settings (Karra et al., 2017; Kraay & McKenzie, 2014), it is likely that our primary outcomes (see below) for households residing in the same kebele will be highly correlated. The computed standard errors need to be adjusted for this within-cluster correlation. Following recommendations in the literature (Abadie et al., 2017), the standard errors in our regressions will be clustered at the kebele level to account for the randomized design. The cluster-robust standard errors will be computed using Stata's *vce(cluster)* command that adjusts the standard errors (Liang & Zeger, 1986). All statistical analyses will be conducted using Stata, version 17 or higher.

Randomization balance is established by comparing baseline data for households in the treatment and control arms. To test for statistical balance, we run a series of regressions of household characteristics on an indicator variable characterizing the treatment assignment and an indicator variable for randomization strata. Standard errors are clustered at the kebele level, and an F-test is used to determine whether we can statistically reject the null hypothesis of balance. Variables in the balance tests include simple demographic characteristics and the baseline levels of the primary outcomes of interest. These results are reported in Section 3.10 of this report.

We will measure the impact on our primary and secondary outcomes (listed below) using an analysis of covariance estimation approach (McKenzie, 2012). In our analysis, we will estimate two primary specifications. For livelihood and gender-related outcomes, we are primarily interested in the pooled effect of any treatment (T2 and T3) vis-à-vis the control arm of PSNP only. The regression of interest will be estimated as follows:

$$Y_{id,t=1} = \beta T_{id} + \gamma Y_{id,t=0} + X'_{id,t=0}\vartheta + \chi_d + \varepsilon_{id}, \quad (1)$$

where $Y_{id,t=1}$ captures the outcome of interest in household *i* residing in woreda *d* at midline/endline *t* and $Y_{i,t=0}$ at baseline. Variable T_{id} is a time-invariant indicator variable, receiving a value of 1 if the household is randomly selected to study arm T2 or T3, and zero otherwise. The average impact of the pooled SPIR II interventions relative to the control group (T1) is quantified by β. To assess whether our estimates are sensitive to the inclusion of basic household-level controls, we estimate the equation (1) with and without baseline controls (captured in vector $X'_{id,t=0}$), including household size, age, and education level of both parents. The term χ_d represent woreda fixed effects, given that we are conducting stratification at the woreda level.

For nutrition-related outcomes, we will estimate the following specification for variables for which a baseline value is available.

$$Y_{id,t=1} = \beta_1 T_{id}^2 + \beta_2 T_{id}^3 + \gamma Y_{id,t=0} + X'_{id,t=0}\vartheta + \chi_d + \varepsilon_{id} \quad (2)$$

where variables T_{id}^2 and T_{id}^3 are time-invariant indicator variables, receiving a value of one if the household is randomly selected to receive the T2 or T3 treatment package, respectively, and zero otherwise. We will also report the p-value for the hypothesis that the treatment effects are consistent across treatment arms, $\beta_1 = \beta_2$, to enable us to test whether there is a differential effect of

supplementing the NCGs with cash grants. As before, we estimate equation (2) with and without baseline controls ($X'_{id,t=0}$), including household size, age and education level of the primary caregiver, and age and sex of the child.

In addition, for outcomes for which baseline value is not available (e.g., anthropometric measures), equation (3) will be estimated without $Y_{id,t=0}$ as in the following specification.

$$Y_{id,t=1} = \beta_1 T_{id}^2 + \beta_2 T_{id}^3 + X'_{id,t=0}\vartheta + \chi_d + \varepsilon_{id} \quad (3)$$

In all regressions, our treatment variables are defined based on the initial treatment assignment, and not based on actual compliance. Consequently, our impact estimates capture intention-to-treat effects .

Each analysis will be conducted at both midline and endline for the outcome variable as measured in that round of data collection. The outcomes measured in each round of data collection are summarized in Table 2, with additional details provided in Table 3 and Table 4.

In addition to reporting standard p-values, we will also report p-values corrected for multiple hypothesis testing. This correction will be implemented across the set of primary and secondary outcomes in each domain (livelihoods and gender, and nutrition).

2.1.6 Robustness Checks and Additional Analyses

By collecting contact information for each household at baseline and by engaging in multiple follow-up visits in cases in which respondents are not initially reached by enumerators, we hope to minimize survey attrition. Attrition in a previous evaluation conducted in the same region by the SPIR I team was in fact less than 10% over an evaluation period of 4 years. Nonetheless, we will test for differential attrition by treatment assignment at the time of endline analysis and will present estimates using an appropriate bounding procedure if differential attrition is detected. We will also report additional specifications in which we regress a binary variable for attrition on the interaction of baseline characteristics and treatment binary variables, to assess whether there is differential attrition with respect to baseline characteristics.

Unless explicitly stated above, there will be no imputation for missing data due to item non-response at the endline. Missing data on baseline variables will be dummied out of the relevant specifications.

In addition to the analysis of pooled treatment effects, we will report heterogeneous treatment effects along certain pre-specified dimensions. This analysis should be considered exploratory.

The first is child gender (Medhin et al., 2010). Given that nutritional practices and outcomes can significantly differ for boys and girls, assessing the differential effect of the proposed interventions by child gender may be important.

The second dimension of heterogeneity that will be assessed is baseline male (paternal) knowledge around and engagement in infant feeding practices. Our hypothesis is that households in which men are more knowledgeable about infant feeding practices at baseline or more engaged in feeding and caretaking activities, may be more responsive to the interventions and show larger shifts in behavior and outcomes than households in which men show a low baseline level of knowledge and engagement.

2.1.7 Limitations

The evaluation has a number of limitations. First, while the randomized controlled trial will generate unbiased estimates of treatment effects within the evaluation sample—a characteristic we generally refer to as internal validity—the study kebeles are not a random subsample of the SPIR II evaluation area.[5] Accordingly, the findings may not be externally valid for other parts of the SPIR II operational area or other areas served by the PSNP.

Second, the selection of indicators for this evaluation was primarily motivated by the BHA set of required indicators; the research team has then designated some indicators as primary or secondary based on their salience in the intervention theory of change. Due to resource constraints, not all indicators can be measured at both midline and endline, rendering it more challenging to ascertain the full set of treatment effects.

Third, we will generally report average effects of the interventions on the full sample of interest. We do not anticipate having sufficient statistical power to analyze effects for targeted subgroups of interest (i.e., previous PSNP beneficiaries or households that are newly qualified for the PSNP).

2.2 Primary Outcomes of Interest

Table 2 summarizes the primary and secondary outcomes for the livelihoods and gender analyses. The primary outcomes focus on per capita consumption-expenditures and levels of asset and cash savings. While we will report all primary and secondary outcomes in the relevant evaluation reports, the academic output(s) will focus on a sub-set of the secondary outcomes (see the table below).

Table 3 summarizes the primary and secondary outcomes for the nutrition analyses. Here, the primary outcomes focus on indicators of chronic child undernutrition and on meeting IYCF-related targets. As before, the academic output(s) will focus on a sub-set of secondary outcomes (see the table below).

Table 3: Primary and secondary outcomes: livelihoods and gender

	Reported in the evaluation reports?	Reported in the academic article?
Primary outcomes:		
Daily per-capita consumption-expenditure (BL40)	X	X
Total value of productive assets	X	X
Total value of livestock assets	X	X
Savings (binary and continuous variable)	X	X
Secondary outcomes:		
Food security (BL06 and BL10)	X	X
Prevalence of poverty (BL01)	X	X

[5] In particular, the selection of kebeles was driven by the identification of kebeles that were not previously included in the randomized controlled trial conducted as part of SPIR I.

	Reported in the evaluation reports?	Reported in the academic article?
Depth of poverty of the poor (BL02)	X	
Net income from livestock production (binary and continuous variable)	X	X
Net income from any non-agricultural production (binary and continuous variable)	X	X
Credit access (binary and continuous variable) (BL42)	X	X
Cash-earning indicators (BL32, BL33, BL34, BL35)	X	

Table 4: Primary and secondary outcomes: nutrition

	Reported in the evaluation reports?	Reported in the academic article?
Primary outcomes:		
Height-for-age (continuous variable, children 30-48 months at endline)	X	X
Prevalence of children 6–23 months consuming a diet of minimum diversity (MDD-C) (at midline) BL39	X	X
IYCF knowledge	X	X
Secondary outcomes:		
Early childhood development score (at midline and endline)	X	X
Percentage of children 6–23 months receiving a minimum acceptable diet (at midline) BL12	X	X
Height-for-age (continuous variable, children 6–23 months at midline)	X	X
Stunting (binary variable, children 6–23 months at midline) BL04	X	
Stunting (binary variable, children 30–48 months at endline) BL04	X	X
Wasting (binary variable, children 30–48 months at endline) BL04	X	X
Weight-for-height Z-score (WHZ) (continuous variable, children 6–23 months at midline)	X	X
Prevalence of healthy weight (WHZ ≤ 2 and ≥ -2) (binary variable, children 30–48 months at endline) BL05	X	X

We conducted power calculations using the specified sample size, setting the significance level at 5% and power at 80%, and allowing for 10% attrition between baseline and endline surveys. Power

calculations were conducted in Stata using the command *clustersampsi*. Note that given the symmetric design of the study, the minimum detectable effect is identical for any pairwise comparison of arms: the minimum detectable effect is the same for comparing T1 or T2 vis-à-vis the control arm, as well as comparing T1 and T2.

We report power calculations for the primary outcomes of interest only. All variables of interest were measured using the data from the SPIR I endline survey conducted by IFPRI in 2021, focusing on households in the control arm.

For the livelihoods analysis, the evaluation is able to detect a 20% increase in consumption; a 21% increase in the total value of household assets; a 25% increase in the total value of household livestock assets; and a 12-percentage point increase in the probability that households report any savings.

For the nutrition analysis, the evaluation can detect a 0.15-food group (8%) improvement in children's dietary diversity, a 0.29-unit change in height-for-age z-score, and an 8% improvement in IYCF knowledge score (constructed on a scale from 1 to 7).

Table 5: Power calculations

	Estimated level in the control arm (measured in SPIR I endline)	Minimum detectable effect	Estimated level in the treatment arm
Consumption (aggregate monthly consumption per adult equivalent)	1016 birr	20%	1219 birr
Value of household assets	19076 birr	21%	23164 birr
Value of household livestock assets	16457 birr	25%	20518 birr
Probability households report any savings	47%	12 percentage points	59%
Children's dietary diversity	1.9	8%	2.05
Height-for-age z-score	-1.6	18%	-1.31
IYCF knowledge	3.92	7%	4.22

2.3 Survey Procedures

The fieldwork for the baseline survey was conducted by EconInsight under the supervision of Dr. Tigabu Getahun, working in close collaboration with IFPRI. EconInsight provided all equipment (data collection was performed on Android Samsung tablets) and led on the organization of transportation and logistical services.

In the survey preparation phase, EconInsight staff translated the survey into both Amharic and Afaan Oromo, coded in SurveyCTO to be conducted as a computer assisted personalized interview (CAPI), and led the testing and refinement of the program. The activity manager, the quality control specialist, and the two field coordinators also revised the translation and phrasing of the questionnaires and

supporting implementation materials from English into the local languages (Amharic and Afan Oromo) based on the feedback received during training and piloting.

The survey firm also led on the process of obtaining appropriate approvals for survey work from local stakeholders. A formal introductory letter was sent to the implementing partners and the head government officials at the region, zone, and woreda level requesting an endorsement of the survey from local authorities. EconInsight hired highly experienced, knowledgeable, socially accepted, and well-connected guides in each target sample kebeles to maximize the participation rate of households.

2.3.1 Training and Piloting

On July 15 and 16, 2022, a group of select experienced supervisors, senior field coordinators, quality control specialists, and the project manager received a two-day training from the SPIR II RFSA principal investigators. The purpose was to review the approved questionnaire and to provide feedback to IFPRI to enable any further refinement of the questionnaire.

Following this two-day training, the activity manager at EconInsight provided intensive training from July 18 to 29 for all enumerators, senior and field coordinators, quality control specialists, programmers/ data managers, and translators. The training ensured that all enumerators interpreted the survey questions consistently and asked the questions in the prescribed manner. Field coordinators were trained to fully explain the purposes and importance of the baseline household survey in simple terms and to reassure the respondent about the confidentiality of information. Quality control specialists and database managers were trained to run high-frequency consistency checks on the data from the office using Stata, throughout the data collection process.

The training process consisted of a combination of classroom training, mock interviews, pilot interviews, and debriefing. During the training, the activity managers interpreted and explained the meaning and content of each question. They defined the code of conduct, the composition, and organization of the team, and the communication channels to be employed within the team. There was also a two-day CAPI-based mock interview session for all the survey team members at the end of the conceptual training session.

Before launching the main data collection, the survey team then conducted 3 days of pilot interviews in a sample of households outside the main evaluation sample, in two nearby rural villages in the Amhara and Oromia regions. During the pilot survey, EconInsight interviewed about 52 households, though in some cases the male respondent was not available to participate. The pilot survey was conducted in kebeles that are not part of the main study.

The objective of the pilot was to test the programmed version of the survey instrument and glean feedback on the length of the interview; the phrasing, context, and framing of the questionnaire; the quality of the translation; and strategies for approaching rural households. Pilot testing also provided the team with an opportunity to practice appropriate interviewing behavior, appropriate use of the CAPI program and tablets, and participation in the team's field routines (review of the data by supervisors, completion of field control sheets by supervisors, and distribution of work assignments and coordination by supervisors).

Following the pilot interview, a half-day debriefing session was conducted with the supervisors, senior field coordinators, data quality control specialists, the programmer, the IFPRI research assistant, and the activity manager (co-investigator). The team identified potential revisions that would enhance the questionnaire, and required changes were implemented in the CAPI. The senior survey team members or the IFPRI research assistant addressed other questions or challenges the team encountered and clarified any concepts or questions that had not been well-understood.

2.3.2 Sampling Frame and Sample Structure

The sampling frame for this survey was constituted by the updated PSNP5 beneficiary list, a complete list of all PSNP beneficiary households from the target sample kebeles. The target sample included 3,081 households in 237 kebeles. Using the full beneficiary list, IFPRI randomly selected and ordered the selected sample households in each sample *gott* (village) before the start of fieldwork. Households were first divided into a list by *gott* and then ordered within that list.

EconInsight then provided each team's supervisor with the list of selected households and location information. The list included the household ID, the full name of the household head, the marital status of the head, the full name of the spouse if married, the birth date of the household head, and the household's location. Using identifying information included in the tracking sheet and with the assistance of local guides, the supervisor and the enumerator located the structures/dwelling of the pre-selected sampled households in each of the sampled kebeles. The pre-selected eligible households were visited strictly following the sampling order and the enumerators administered the sample verification questions to verify whether the pre-selected household was eligible to participate in the baseline survey. Again, the eligibility criteria were as follows.

1. The household must be enrolled under SPIR II as a PSNP beneficiary in a target kebele.
2. The household must meet one of the following characteristics:
    a. There is a pregnant woman present who self-reports pregnancy, with an estimated gestational age that is at least 3 months (i.e., following the first trimester).
    b. There is an infant present aged less than 9 months as of the date of the survey; and the infant's mother or primary caretaker is also a resident in the household.

Households that were identified as ineligible (households with no pregnant women or children below the age of 9 months); households that refused to participate in the survey; and households that could not be located or surveyed (due to death, mental impairment, or outmigration) were replaced with households in the replacement list strictly following the original order of households. Sample replacement was implemented only after the senior field coordinators and the quality control specialist confirmed that the selected sample household was ineligible or could not be located, or the household refused to participate following multiple attempts. In those cases, the survey team carefully recorded the reason for replacement. A CAPI survey form was submitted for each screening to record the reason that a household was determined to be ineligible.

Once a sampled household was contacted and verified to be eligible, the enumerators further identified the eligible household members (primary female, primary male, or household head/knowledgeable person) who would participate in the survey. The primary female respondent for this study was defined as the mother or primary female caregiver of the index child (the child under the age of 9 months) or the

adult female who was pregnant at the time of the interview. Whenever multiple children were present below the age of 9 months, the index child was identified as the youngest of all the children in the household. The primary male respondent for this study was defined as the spouse or partner of the primary female respondent. In households where neither the primary female respondent nor the male respondent was the household head, the household head was the third eligible respondent. The household head was the individual identified to be most knowledgeable about the household's activities and who plays a leading role in household decision-making, particularly concerning farming, household economic activity, and expenditures. Generally, the person identified by the household as the household head was accepted in this role for the survey. Surveys were administered separately to each eligible respondent (primary female respondent, primary male respondent, and household head, if applicable).

The final, realized sample was 3,015 households in 234 kebeles (80 in the Amhara and 154 in the Oromia region). Three kebeles were excluded due to inaccessibility.[6] The realized sample corresponds to 98% of our original target sample. Table 6 summarizes the number of household interviews completed in each kebele.

Table 6: Completed household surveys disaggregated by kebele

Region	Woreda	Number of kebeles targeted	Number of kebeles surveyed	Total number of household interviews completed	Average PSNP households per kebele in this woreda	Total PSNP households in woreda
Amhara	Bugna	11	10	130	1299	14294
	Dehana	16	16	207	1024	16384
	Gazgibla	13	13	169	828	10764
	Gazo	10	10	130	1213	12133
	Lasta	10	10	129	1370	13701
	Meket	6	6	78	1146	6878
	Sekota	8	6	78	1078	8624
	Wadla	9	9	117	1077	9692
Regional total		83	80	1038	1129	92470
Oromia	Boke	22	22	285	1047	23054
	Chiro	23	23	299	795	18275
	Doba	25	25	305	524	13090
	Gemechis	20	20	260	773	15457
	Grawa	18	18	233	738	13292
	Habro	32	32	416	759	24278

[6] Two kebeles in Sekota woreda were dropped due to conflict. One kebele in Bugna woreda was dropped as it was inaccessible by vehicle to the survey team and there were also some local security-related conflicts (unrelated to the conflict in Tigray).

Region	Woreda	Number of kebeles targeted	Number of kebeles surveyed	Total number of household interviews completed	Average PSNP households per kebele in this woreda	Total PSNP households in woreda
	Kurfa Chelle	14	14	179	804	11253
Regional total		154	154	1977	777	118699
Total: full sample		237	234	3015	965	211169

2.3.3 Quality Control During the Survey

During the baseline survey, EconInsight regularly updated the IFPRI impact evaluation team regarding data collection activities and reported the progress of the data collection and challenges encountered on a regular basis.

To aid in real-time data visualization, EconInsight employed a data monitoring dashboard embedded in Google Sheets. A number of additional strategies were also employed to ensure the collection of high-quality data.

- EconInsight recruited highly qualified and motivated enumerators and supervisors and provided an intensive two-week training.
- The senior field coordinators conducted a random set of call-backs to ensure that data had not been omitted or falsified and that the survey protocol was strictly observed at all times.
- The survey management team instituted routine checks on data quality in parallel with data collection to enable mistakes to be rectified during the course of the survey. A do-file written by our quality control specialist in close collaboration with the activity manager and the IFPRI research analyst was run regularly to ensure that entered data was complete, reliable, internally consistent, did not include any outliers, and was of acceptable quality. Whenever the EconInsight quality control specialist and the IFPRI research assistant flagged errors, omissions, mistakes, or data anomalies, the quality control specialist sent back the consolidated flagged potential errors to the senior field coordinators and supervisors with detailed comments on the variables that required corrective action.
- Data programmers provided another layer of data quality checking. While the data programmer was primarily tasked with designing the electronic version of the survey instrument, he was also managing data flow processes in real-time.
- The senior field coordinators and quality control specialists conducted random field visits to ensure that completed interviews were actually conducted.

The use of electronic data collection [CAPI] also provided additional opportunities for quality control. During the design stage, EconInsight exploited the capabilities of the SurveyCTO program to incorporate automatic skip patterns and constrain responses so that the enumerators were not able to continue recording responses when the data was clearly incorrect, invalid, or inconsistent.

3. PRIMARY FINDINGS

3.1 Summary of Household Demographic Characteristics

Table 7 summarizes the basic demographic characteristics of the households included in the sample. Out of the realized sample of 3,015 households, 66% of these households are in Oromia and 34% are in Amhara, in line with the sample projected before the initiation of the survey.

As noted above, the baseline survey targeted PSNP households with a pregnant woman or an infant aged less than 9 months. We call this child the index child. The bottom part of Table 7 breaks down the sample based on these eligibility criteria. About 53% of the households are characterized by a woman who is currently pregnant and 47% are characterized by a child under 9 months. Of those households reporting a young child, 20% report a child under 3 months old, 15% report a child 4 to 6 months old, and 13% report a child 7 to 9 months old. Given that our original target was that seven out of 13 sampled households in each kebele (or 54%) would be characterized by the presence of a pregnant woman, our sample composition ratios are exactly on target.

Table 7: Sample composition

	N	Percentage
Region		
Amhara	3,015	34.43
Oromia	3,015	65.57
Breakdown based on eligibility criteria		
Currently pregnant	3,015	53.33
Index child is 0–3 months old	3,015	19.97
Index child is 4–6 months old	3,015	14.49
Index child is 7–9 months old	3,015	13.10

Table 8 reports the basic demographic characteristics of the sample households. The average household size is five and 89% of households report that the head of the household is male, characterized by an average age of 35. Fifty-two percent of household heads report some formal education and 87% report that their primary economic activity is crop production. Ninety-three percent of household heads are married.

The primary female in the sample households is on average 29 years of age and 93% are married. Only 43% of primary females report any formal education and, on average, the level of education is low: only 26% report 3 to 7 years of education and only 7% report more than 8 years of education. Given that maternal education is often identified as an important predictor of child nutritional status, including in previous analyses implemented in Ethiopia (Le & Nguyen, 2020; Li et al., 2020; Tekile et al., 2019), the low level of maternal education provides important context for the planned interventions.

While all households are PSNP beneficiaries in 2022 given the sample criteria, 12% report that they are receiving temporary direct support for some part of the year; 96% report that they are public works beneficiaries. These two categories are not mutually exclusive, given that some households may receive direct support for part of the year, typically during pregnancy and a child's first year of life, but otherwise receive public works benefits. It is important to note that due to the substantial retargeting process observed at the initiation of the PSNP5 period, the majority of these households are new to the PSNP and did not previously receive any benefits; only 33% report previously receiving PSNP benefits in 2021, the final year of PSNP4 programming. (41% are previous beneficiaries in Amhara and 26% in Oromia).

Table 37 in Appendix B reports some summary demographic statistics comparing across previous PSNP beneficiaries and new PSNP beneficiaries; the observed pattern is heterogeneous. New PSNP beneficiary households have a lower level of assets, but are also less likely to be identified as extremely poor; regional differences may drive some of these variable patterns. Table 38 in Appendix B reports the full set of all required BHA indicators. These indicators are reported and discussed in more detail in the subsequent subsections, organized by outcome family.

Table 8: Household demographics

	N	Mean
Household size	3,015	5.39
Male headed household (percentage)	3,015	88.7
Household head's age	3,015	35.40
Household head has some formal education (percentage)	3,015	52.3
Household head's main occupation is crop production (percentage)	3,015	86.8
Household head is married (percentage)	3,015	93.2
Primary female's age (percentage)	3,015	28.75
Primary female has some formal education (percentage)	3,015	43.2
Primary female has 1 to 3 years of education (percentage)	3,015	17.8
Primary female has 4 to 7 years of education (percentage)	3,015	18.3
Primary female has 8 or more years of education (percentage)	3,015	7.1
Primary female's main occupation is crop production (percentage)	3,015	37.6
Primary female is married (percentage)	3,015	93.3
Public works beneficiary in 2022 (percentage)	3,015	96.1
Direct support beneficiary in 2022 (percentage)	3,015	12.1
Public works beneficiary in 2021 (percentage)	3,015	29.8
Direct support beneficiary in 2021 (percentage)	3,015	3.3

Notes: Estimates from the IMPEL baseline survey sample.

3.2 Summary of Household Demographic Characteristics

The baseline survey instrument included several household food security measures: the Food Insecurity Experience Scale (FIES), Food Gap, Food Consumption Score (FCS), and Household Diet Diversity Score (HDDS). Previous research in Ethiopia shows how the food insecurity prevalence estimates can substantially vary depending on which food insecurity indicator is used (Maxwell et al., 2013). For example, the FIES tends to predict considerably higher food insecurity than the FCS. Table 9 summarizes the key indicators and their 95% confidence intervals.

Table 9: Key indicators for household food access

	Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL06. Prevalence of moderate food insecurity in the household based on the Food Insecurity Experience Scale (FIES)	31.38	29.22	33.54	3,015
BL06. Prevalence of severe food insecurity in the household based on the Food Insecurity Experience Scale (FIES)	57.41	54.61	60.22	3,015
BL10. Percentage of households with poor Food Consumption Score (FCS)	5.47	4.24	6.70	3,015
BL10. Percentage of households with borderline Food Consumption Score (FCS)	19.30	17.27	21.34	3,015
BL10. Percentage of households with acceptable Food Consumption Score (FCS)	75.22	72.71	77.74	3,015

Notes: Estimates from the IMPEL baseline survey sample.

Developed by the Food and Agriculture Organization, the FIES is a subjective household food insecurity indicator capturing households' perceived food insecurity situation (Ballard et al., 2013). The FIES survey module contains eight 'Yes/No' questions about the household's food security situation (listed in Table 10). Following the BHA guidelines, we used a 12-month recall period for these questions. Each 'Yes' response scores a point, with higher overall scores indicating a worsening food insecurity situation.

Table 10 reports the percentage of households responding to each FIES question affirmatively. The severity of food insecurity increases as one moves down the list of questions. It is therefore expected that the percentage of households responding positively to the question decreases as we move toward questions about severe food insecurity (Ballard et al., 2013). Building on this expected data structure, we can test the data quality using the Rasch model developed by the Food and Agriculture Organization. The Rasch reliability score for the SPIR II baseline survey was estimated as 0.85, which implies a very good fit (i.e., households respond to these questions in a way that is expected). Looking at the fit at the FIES question (or item) level, the *infit* scores in our data range in value between 0.80 and 1.21, thus all falling within the acceptable range of 0.7–1.3 (Cafiero et al., 2018).

Using raw FIES scores, we categorized households as severely food insecure if they responded 'Yes' seven or eight times out of the eight questions, moderately food insecure if the number of 'Yes' responses was between four and six inclusive, and mildly food insecure if the number of 'Yes' responses was between one and three inclusive. The household is categorized as food secure if they responded 'No' to all eight questions. Figure 4 shows that according to FIES, nearly 90% of the households in our sample are categorized as moderate or severe food insecurity. Only 2.7% are categorized as being food secure.[7]

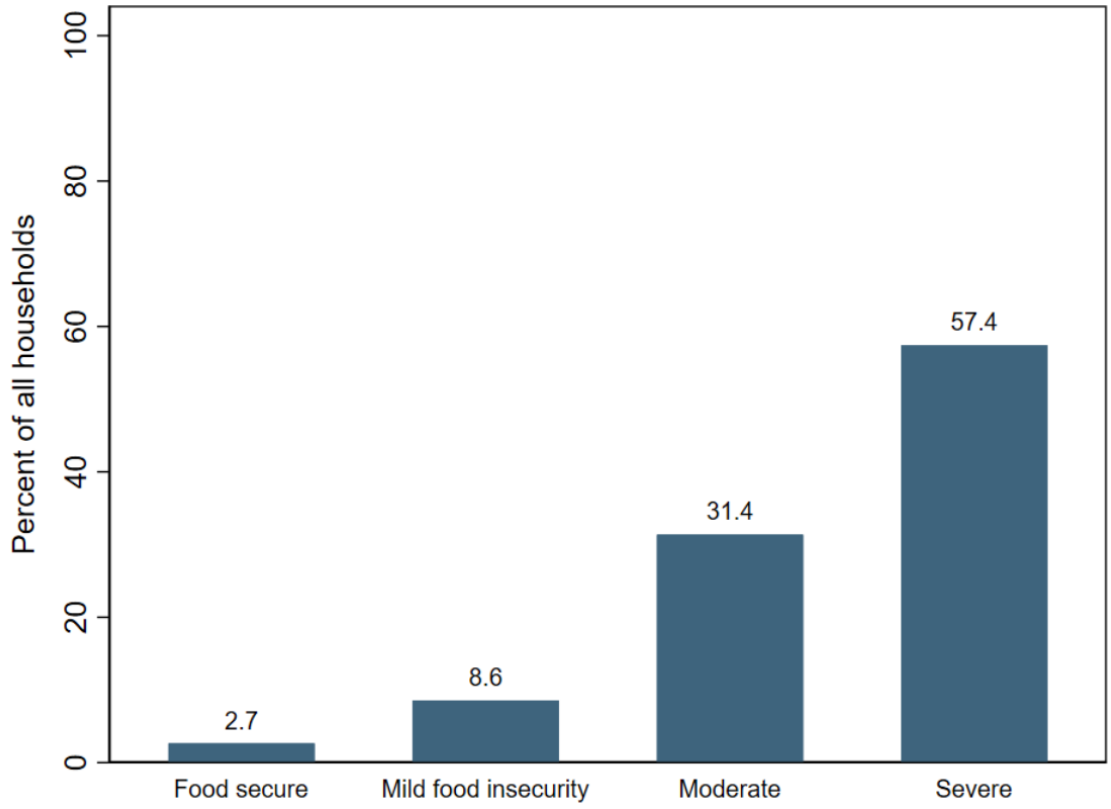
Table 10: Percentage of households responding affirmatively to Food Insecurity Experience Scale (FIES) questions

FIES Question	Yes (percentage)
In the last 12 months…	
…was there a time when you or others in your household were worried you would not have enough food to eat because of a lack of money or other resources?	96.1
…was there a time when you or others in your household were unable to eat healthy and nutritious food because of a lack of money or other resources?	95.3
…was there a time when you or others in your household ate only a few kinds of foods because of a lack of money or other resources?	94.0
…was there a time when you or others in your household had to skip a meal because there was not enough money or other resources to get food?	78.9
…was there a time when you or others in your household ate less than you thought you should because of a lack of money or other resources?	86.9
…was there a time when your household did not have food because of a lack of money or other resources?	69.0
…was there a time when you or others in your household were hungry but did not eat because there was not enough money or other resources for food?	61.8
…was there a time when you or others in your household went without eating for a whole day because of a lack of money or other resources?	31.6

N = 3,015 households.

[7] Using internationally comparable FIES food security thresholds that are based on probabilistic assignments, we estimate that 87.9 percent of the households in our sample are moderately or severely food insecure.

Figure 4: Household food security status based on the Food Insecurity Experience Scale (FIES)

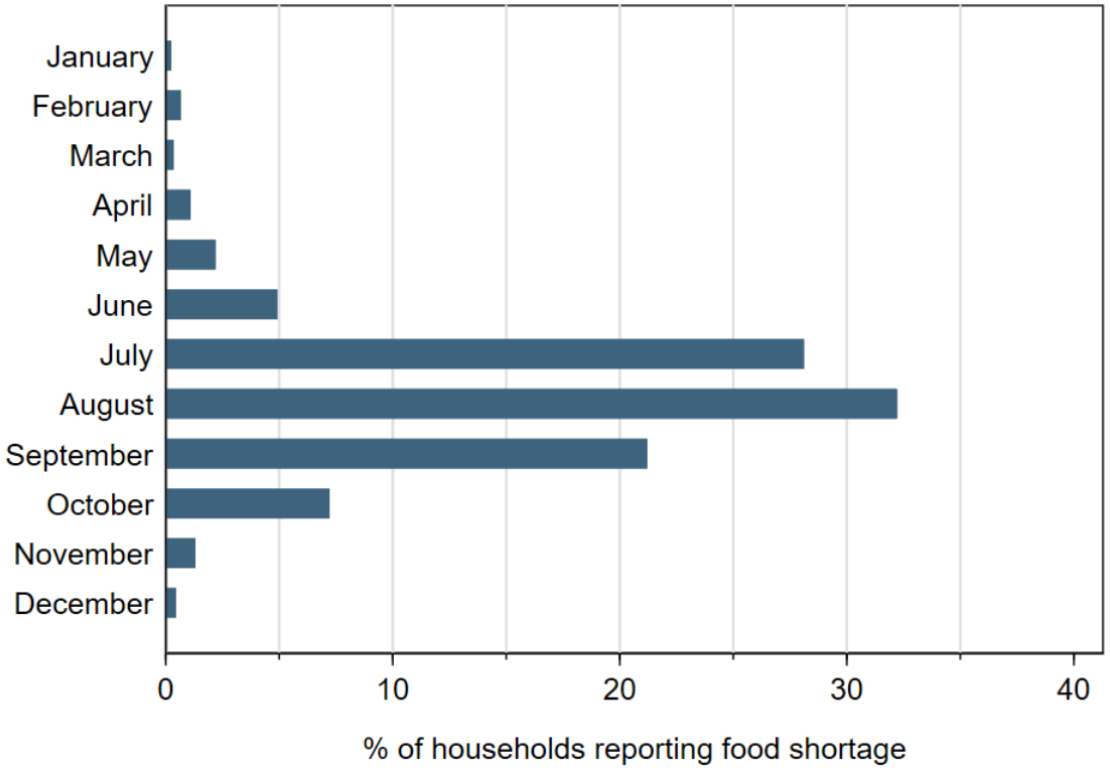


N = 3,015 households.

The food gap indicator captures the number of months that the household was unable to satisfy its food needs in the past 12 months. This indicator is based on households' subjective assessment and has been used as the primary food security indicator in the PSNP evaluations since the onset of the program. The mean food gap reported by the households in our sample is 3.5 months (median of 3).

Nearly 94% of all households reported having experienced food shortage in the last 12 months. According to these households, the most food acute food shortage months are July, August, and September (Figure 5), i.e., the months during which the survey took place or immediately before.

Figure 5: Most acute food shortage month



N= 2,828 households reported having experienced food shortage in the past 12 months.

Moving on to diet-based food security measures, we next report on the Food Consumption Score (FCS), an indicator developed by the World Food Program.[8] The FCS is a weighted index that combines dietary diversity and consumption frequency. The index is based on the household consumption of nine food groups (see Table 11 for the list of food groups and the weights attached to each group). The weighted index ranges between 0 and 112, with higher scores indicating better food security. Household diets are categorized as 'Poor' if the FCS is below 21, 'Borderline' if the score is above 21 but below 35, and 'Acceptable' if above 35.

Table 11 shows the number of days households consumed from each food group in the last 7 days before the interview. We see that the household diets are focused on starchy staples (cereals and tubers), pulses, and vegetables, while the consumption of fruit and animal-sourced foods (meat products, eggs, and dairy) is infrequent in this sample. The mean food consumption score in the sample is 44.8 and the median is 45.5. Consequently, for more than two-thirds of the households, the food consumption status is categorized as acceptable (Figure 6). Nearly 20% of the households have a food consumption score that is borderline, while 5.5% of the households have a score that classifies their food consumption situation as poor. As also found by previous work, the FCS predicts considerably lower food insecurity prevalence than the FIES (Maxwell et al., 2013). One possible reason for this

[8] FCS was incorporated into the food consumption module. For each food item in the food consumption module, respondents were asked if they consumed the item in the past 7 days, on how many days they consumed the item and the quantity consumed. We grouped the food items into the FCS food groups and computed the FCS using the responses to the consumption frequency questions.

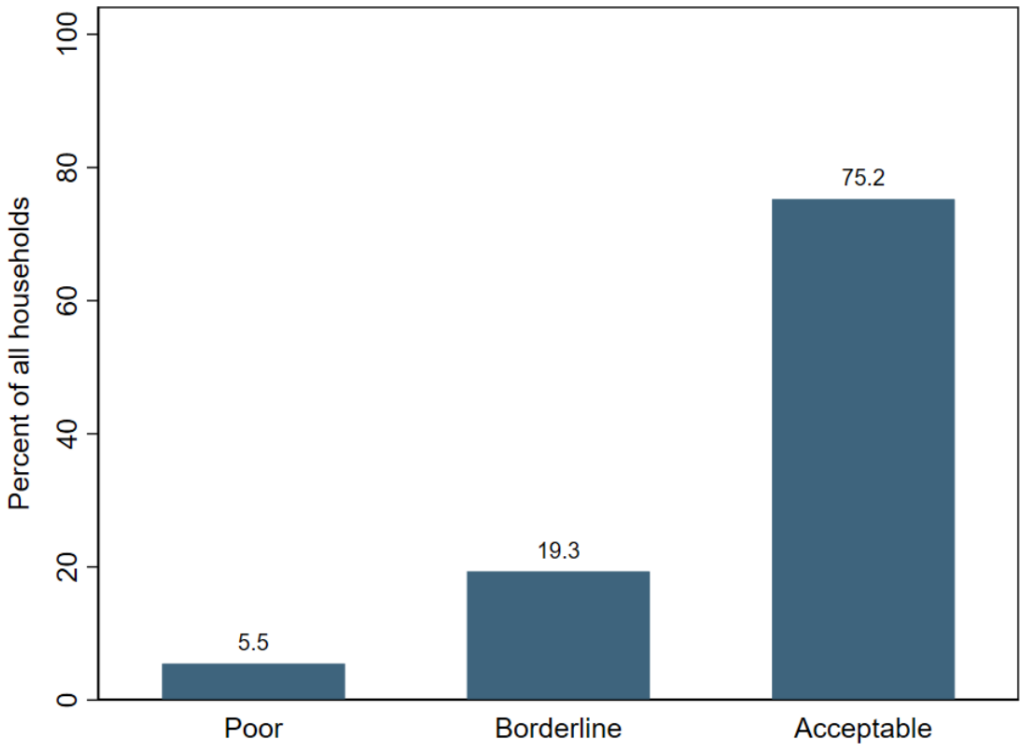
discrepancy is the thresholds that categorize households with borderline or acceptable food consumption may have been set too low (Lovon & Mathiassen, 2014). For example, an Ethiopian household that consumes *shiro* (a traditional dish made of chickpea and butter or oil and eaten with injera) each day of the week would obtain an FCS of 38.5, which is above the 'acceptable' threshold (35).

Table 11: Mean number of days households consumed from the Food Consumption Score food groups

FCS Food Group	FCS Weight	Number of Days
Cereals & tubers	2	7.0
Pulses	3	5.0
Vegetables	1	5.9
Fruit	1	0.2
Meat, eggs & fish	4	0.4
Dairy	4	0.9
Sugar products	0.5	3.1
Oils, fats & butter	0.5	5.8
Spices, condiments, etc.	0	7.0

N = 3,015 households.

Figure 6: Household food consumption status based on Food Consumption Score



N = 3,015 households

Next, we look at the Household Diet Diversity Score (HDDS), which groups food consumption into 12 food groups (Swindale & Bilinsky, 2006).[9] The 12 food groups are listed in the first column of Table 12. The second column reports the percentage of households that consumed from each food group in the 7 days before the interview. As before, the consumption of fruits and animal-sourced foods is rare among the sampled households. The mean HDDS in this sample is 6.3 and the median is 6.

Table 12: Percentage of households consuming each of the Household Diet Diversity Score food groups

HDDS Food Group	Percentage
Cereals	99.9
Root and tubers	58.3
Vegetables	93.2
Fruits	10.5
Meat, poultry, offal	6.4
Eggs	14.8
Fish and seafood	0.0
Pulses, legumes, nuts	82.9
Milk and milk products	19.6
Oil, fats	87.3
Sugar, honey	56.6
Miscellaneous	99.9

N = 3,015 households

Table 13 summarizes the key food security indicators for the full sample and separately for each region. The regional differences are relatively small. In Amhara, FIES classifies 87% of the households as moderately or severely food insecure, while in Oromia the corresponding percentage is 90%. The mean food gap in Amhara is 3.3 and in Oromia, 3.6. FCS classifies 16.2% of the households in Amhara as characterized by borderline or poor food consumption status and 29% in Oromia. The mean HDDS is 5.9 in Amhara and 6.5 in Oromia.

Table 13: Food security indicators, by region

Food Security Indicator	Full Sample N = 3,015	Amhara N = 1,977	Oromia N = 1,038
Food Insecurity Experience Scale (FIES):			
Food secure (percentage)	2.7	1.9	3.0

[9] The HDDS was computed from the responses in the household food consumption module. For each food item in the food consumption module, respondents were asked if they consumed the item in the past 7 days, on how many days they consumed the item, and the quantity consumed. For HDDS, we only considered the responses to the yes/no questions and grouped the food items into the HHDS food groups.

Food Security Indicator	Full Sample N = 3,015	Amhara N = 1,977	Oromia N = 1,038
Mild food insecurity (percentage)	8.6	11.5	7.0
Moderate food insecurity (percentage)	31.4	44.4	24.5
Severe food insecurity (percentage)	57.4	42.2	65.4
Food Gap (months)	3.5	3.3	3.6
Food Gap, Food Consumption Score (FCS):			
Acceptable (percentage)	75.2	83.7	70.8
Borderline (percentage)	19.3	11.8	23.2
Poor (percentage)	5.5	4.4	6.0
Household Diet Diversity Score (HDDS)	6.3	5.9	6.5

N: Number of households

3.3 Household Consumption and Poverty

The baseline household survey questionnaire asked respondents to report on their consumption of food and non-food goods. Analyzing these data allows us to report on household per capita consumption expenditures and poverty prevalence. Appendix A provides a detailed overview of household consumption and poverty measurement methods.

Household consumption and poverty estimates are summarized in Table 14, along with their 95% confidence intervals; this includes the three key required indicators (BL40, BL01, and BL02). The mean daily household per capita food consumption-expenditures in this sample is 50.73 birr and the mean daily per capita non-food expenditures is 7.80 birr. Therefore, the average household in our sample spends 86% of its total budget on food, indicative of the high level of deprivation in this sample of households. Summing the mean food and non-food consumption expenditures amounts to a mean total per capita consumption expenditure of 58.78 birr per day. This then translates into 1.72 in 2010 USD or 1.78 in 2011 $PPP. Note that consistent with BHA guidance for indicator construction, BL01 and BL02 are reported at the individual level.

Table 14: Household consumption and poverty indicators

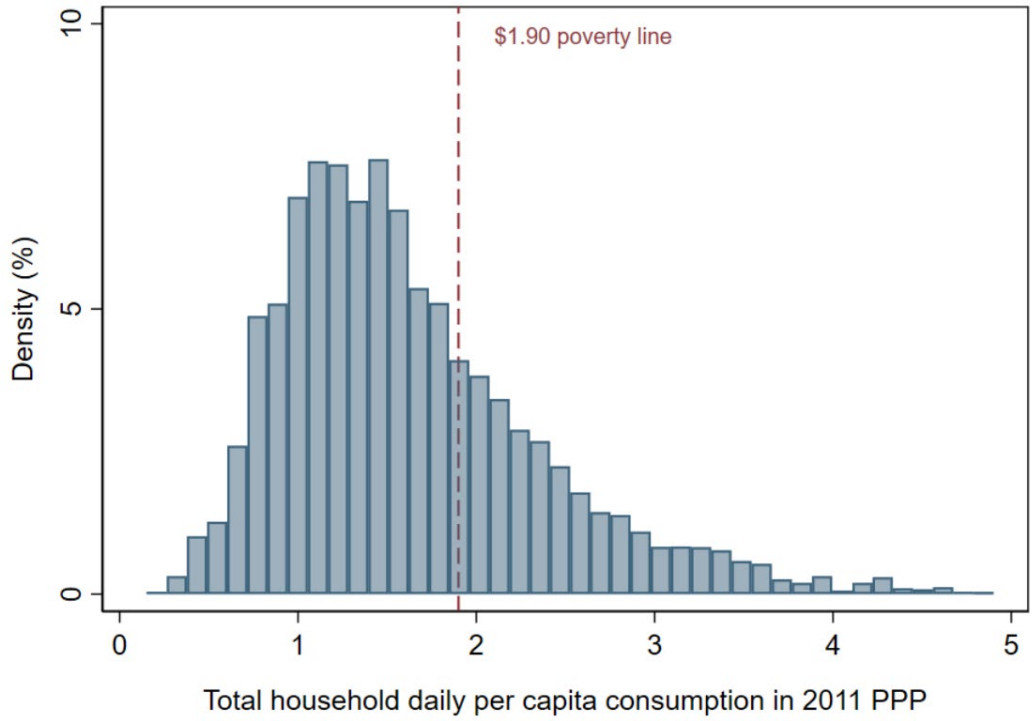
Outcome	N	Mean	CI Lower	CI Upper
Daily per capita food consumption-expenditures in birr	3,015	50.73	49.07	52.39
Daily per capita non-food consumption-expenditures in birr	3,015	7.80	7.45	8.16
Daily per capita total expenditures in birr	3,015	58.78	56.98	60.58
BL40. Daily per capita total expenditures in 2010 USD	3,015	1.72	1.67	1.77
Daily per capita total expenditures in 2011 $PPP	3,015	1.78	1.72	1.83

Outcome	N	Mean	CI Lower	CI Upper
BL01. Prevalence of poverty: Percentage of people living on less than $1.90/day 2011 PPP	16,264	70.96	68.75	73.18
BL02. Depth of poverty of the poor (percentage), based on 1.90 $PPP (*)	11,541	34.21	33.00	35.41

Notes: All confidence intervals based on standard errors clustered at the kebele level. CI = Confidence Interval. (*) Individual level measure; means and CIs estimated using household size as the frequency weight.

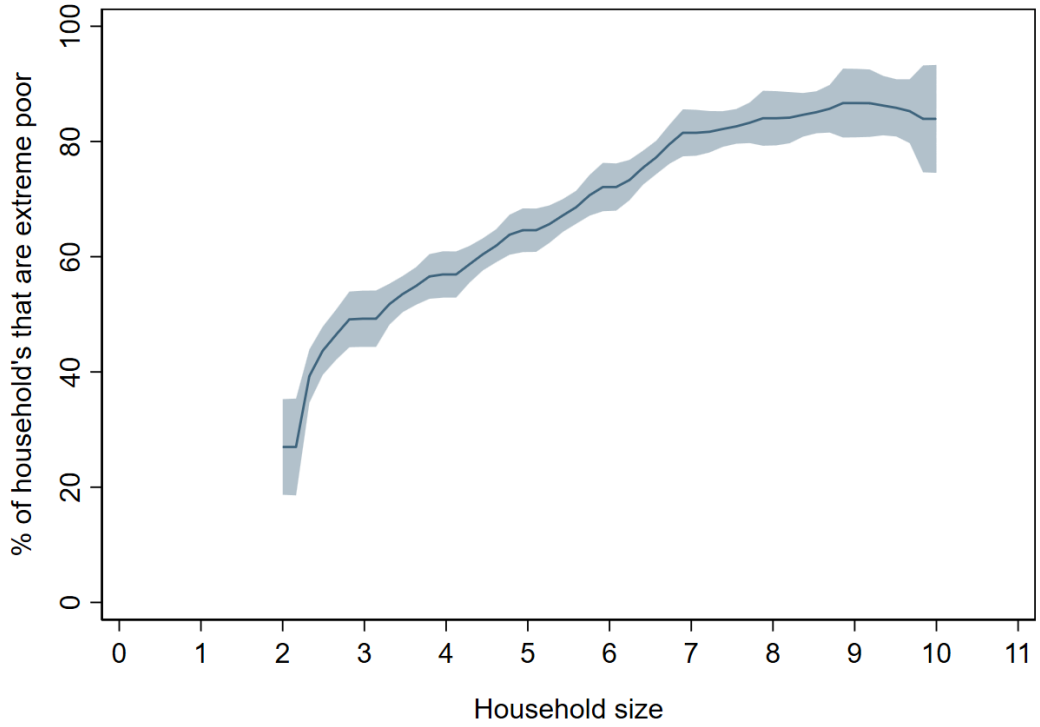
About 71% of sampled individuals originate from households for which the daily per capita consumption is below 1.90 $PPP. Figure 7 overlays the distribution of household daily per capita consumption on the 1.90 $PPP poverty line. We see that many individuals are located just below or just above the poverty line. The prevalence of extreme poverty is higher among households in Amhara (79.1%) than in Oromia (67.8%). The risk of falling below the poverty line increases with household size, with larger households at a greater risk of falling below the poverty line (Figure 8).

Figure 7: Distribution of household daily per capita consumption in 2011 PPP



Note: Household consumption estimates are weighted by household size. The horizontal axis is truncated at the 99th percentile of the consumption distribution

Figure 8: Risk of extreme poverty, by household size



Note: Local polynomial regression. Shaded lines represent 95 percent-confidence intervals. Horizontal axis truncated at 1st and 99th percentile of the household size distribution; N= 2,974 households.

3.4 Water, Sanitation, and Hygiene

The baseline survey also collected key indicators linked to water, sanitation, and hygiene, summarized in Table 15. As per the BHA guidelines, households using basic drinking water services (BL16) are defined as those that:

1) report using an improved water source (piped water, tube well/borehole, protected dug or well, rainwater collection, tanker truck, cart with a small tank or bottled water);
2) have water source in the premises or obtainable from the source in 30 minutes or less roundtrip including waiting time;
3) have water available from the source year around;
4) have water unavailable from the source in the last 2 weeks; and
5) have the water source producing 20 liters per day for each person.

Note that due to an error in the programming of the CAPI survey, the last indicator (the quantity of water collected by the household) was not collected. Accordingly, the BL16 indicator is calculated based on these four indicators only.

In general, the level of adoption of recommended WASH practices is relatively low. Nearly 60% of the households reported an improved water source and 62% had the water source in the premises or at least obtainable within a 30-minute round trip from the premises. Nearly 70% of the households said that the water from their source is consistently available. About 75% reported year around access to drinking water and 18% said that the water was not available from the source in the last 2 weeks before

the interview. Based on these four out of five criteria then, 20% of households are using basic drinking water services. Only 2% of households report access to a handwashing station and only 11% report correct use of recommended water treatment technologies. Half of the households report that they practice open defecation and only 17% have access to a basic sanitation service.

Table 15: Water, sanitation, and hygiene (WASH) indicators

	Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
Household has an improved water source	58.01	54.04	61.98	3,015
Water can be obtained in less than 30 minutes	61.89	58.79	64.99	3,015
Water is consistently available	68.60	65.59	71.62	3,013
Water is available from the source year around	75.14	72.48	77.80	3,013
Water was unavailable from the source in the last 2 weeks	18.08	15.48	20.69	3,014
BL16. Percentage of households using basic drinking water services [10]	20.48	18.12	22.84	3,013
BL17. Percentage of households with soap and water at a handwashing station on premises	1.76	1.22	2.29	3,015
BL18. Percentage of households practicing correct use of recommended household water treatment technologies	11.00	9.40	12.60	3,009
BL19. Percentage of households practicing open defecation	50.35	46.89	53.81	3,015
BL27. Percentage of households with access to a basic sanitation service	17.08	15.15	19.01	3,015

Notes: Estimates from the IMPEL baseline survey sample.

3.5 Infant and Young Child Feeding Practices and Diarrhea Risk

The first 2 years of life are a critical period in child growth and development during which growth faltering accelerates and many children in Ethiopia become short for their age or stunted (Golan et al., 2019). In the first 6 months after birth, the WHO recommends that children be exclusively breastfed. At 6 months of age, breastmilk is no longer sufficient to support their growth and development, and, thus, children need to be introduced to complementary foods. Because of limited gastric capacity, these foods need to be highly nutritious and provided frequently. Moreover, because children's immune systems are still developing during this age, they need to live in a safe and supportive environment characterized by improved water, sanitation, and hygiene and good access to health services for timely treatment.

[10] This indicator shows all households that meet four of the five criteria for basic drinking water services (excluding per person per day criterion). The quantity of water collected by the household was not collected and therefore not considered in the calculation.

Against this backdrop, this section reports on IYCF practices and child diarrhea risk. Table 16 summarizes the key required IYCF indicators for this sample, with more details provided in the discussion to follow.

Table 16: Children's nutritional status and feeding practices

	Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL12. Percentage of children 6–23 months receiving a minimum acceptable diet *	3.10	1.54	4.67	741
BL13. Prevalence of exclusive breastfeeding of children under six months	79.20	76.30	82.09	870
BL14. Percentage of children under five (0–59 months) who had diarrhea in the prior 2 weeks *	12.11	10.40	13.82	1,627
BL15. Percentage of children under five (0–59months) with diarrhea treated with Oral Rehydration Therapy *	63.83	57.09	70.57	188
BL39. Prevalence of children 6–23 months consuming a diet of minimum diversity (MDD-C) *	3.10	1.54	4.66	743

*Notes: Estimates from the IMPEL baseline survey sample. Data on these indicators were collected from all children 0–59 months residing in the household. Given the sampling frame, younger children (within each age-bracket) tend to be over-represented in these statistics. * Due to the sampling strategy, about 90% the children in this sample are less than 10 months of age.*

Information on IYCF practices was collected for all children under 24 months of age. However, due to the sampling design (see Section 2.1.3), the age profile of the children in our sample within this age group is focused on children 0–9 months of age, with nearly 85% of the children under 24 months of age in the sample being less than 10 months of age (Table 17). This skewed age profile should be considered when we interpret the statistics on IYCF practices. For example, the diets of younger children are typically less diverse than the diets of older children and the morbidity risks also vary considerably across age groups (Baye & Hirvonen, 2020).

Table 17: Age distribution of the children in the sample about whom IYCF questions were asked

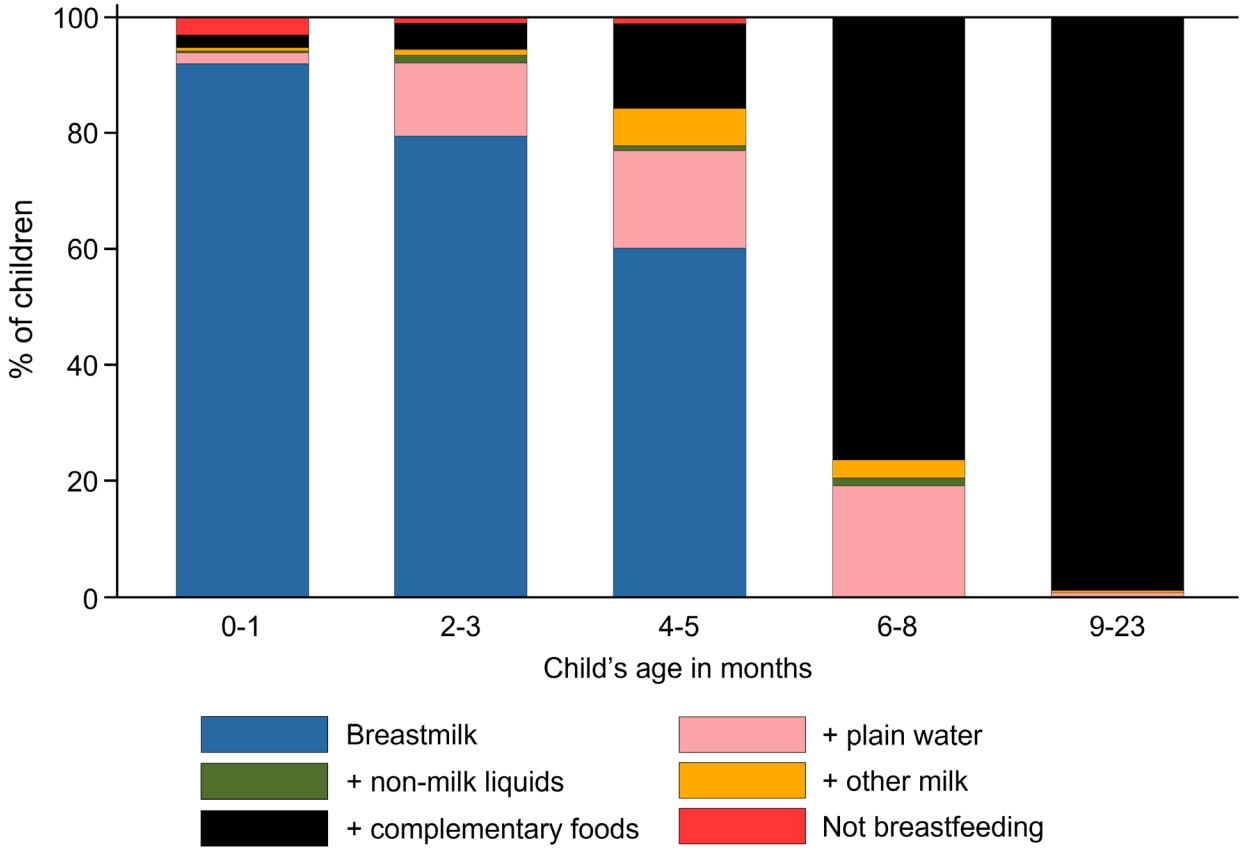
Age category	N	Percentage	Percentage, cumulative
0–1 months	301	17.7	17.7
2–3 months	306	18.0	35.6
4–5 months	279	16.4	52.0
6–9 months	549	32.2	84.2
10–23 months	269	15.8	100.0
Total	1,704	100.0	

N: number of children.

Figure 9 shows IYCF patterns by age cohort. Among children less than 24 months of age, 95% were breastfed at some point in their life. About 79% of the children under 6 months of age were exclusively

breastfed. The exclusive breastfeeding rate is somewhat higher in the Amhara sample (85% of the children) than in the Oromia sample (76% of the children).[11] More than 93% of the children 0–1 months of age in our sample were exclusively breastfed at the time of the interview with the share falling to 81% and 61% for children in the 2–3 month and 4–5 month age groups, respectively. As also documented in previous research in Ethiopia (Hirvonen et al., 2021), introduction of complementary foods in these areas is delayed: 24% of the children 6–8 months old did not consume solid, semi-solid, or soft foods in the day before the interview. Though this large-scale survey was not tailored to identify barriers to the introduction of complementary foods, a separate formative qualitative study conducted by the IFPRI team in 2022 suggested that financial constraints, limited maternal time, and low levels of interaction with the formal health system are all barriers to the adoption of appropriate complementary feeding practices (Leight et al., 2022).

Figure 9: Infant and young child feeding patterns, by age group



N = 1,704 children 0-23 months of age.

In line with previous survey evidence from the PSNP woredas (Berhane et al., 2020), child diets are extremely monotonous. The survey instrument asked caregivers a series of yes/no questions about children's consumption of different foods in the 24 hours before the interview. Following WHO and UNICEF guidelines (WHO and UNICEF, 2021), we grouped these foods into eight food groups out of which one is breastmilk (see Table 18). The diet of minimum diversity (MDD-C) is achieved if the child

[11] This difference is statistically highly significant; p-value<0.01.

consumed from five or more food groups. On average, a child 6–23 months of age in our sample consumed only from 2.3 food groups and consequently, only 3% achieved MDD-C. Less than 8% of the children consumed eggs or flesh foods and 79% of children did not consume any vegetables or fruits. Compared to the sample of children from Oromia, children's diets are less diverse in the Amhara region where none of the children received MDD-C.

Table 18: Diet diversity of children 6–23 months of age

Food Group / Indicator	Full Sample N=743	Amhara N=231	Oromia N=512
Breastmilk, percentage	83.7	93.5	79.3
Grains, roots, and tubers, percentage	71.5	45.0	83.4
Legumes and nuts, percentage	27.5	33.3	24.8
Dairy products, percentage	12.8	1.3	18.0
Flesh foods, percentage	0.9	0.0	1.4
Eggs, percentage	6.5	6.5	6.4
Vitamin A-rich fruits and vegetables, percentage	19.7	2.6	27.3
Other fruits and vegetables, percentage	2.7	0.0	3.9
Diet diversity, number of food groups	2.3	1.8	2.4
Child received Diet of Minimum Diversity (MDD-C), percentage	3.1	0.0	4.5
Child did not consume any vegetables or fruits, percentage	79.0	97.4	70.7
Child consumed egg and/or flesh food, percentage	6.9	6.5	7.0

N: number of children, 6–23 months of age.

We also estimated the prevalence of children meeting the minimum acceptable diet (MAD), which is composed of minimum meal frequency (MMF) and minimum diet diversity (MDD). MMF is calculated as the proportion of children fed a minimum number of times where the minimum number depends on the child's breastfeeding status and age (WHO and UNICEF, 2021). MDD is based on the MDD-C described above and captures the proportion of children who received a minimum number of food groups in the previous day. For breastfed children, MAD is defined as the proportion of children who receive at least the minimum dietary diversity and minimum meal frequency for their age during the previous days. For non-breastfed children, MAD is defined as the proportion receiving the minimum dietary diversity and minimum meal frequency for their age during the previous day as well as at least two milk feeds (WHO and UNICEF, 2021).

Table 19 shows that 52% of the children received MMF while only 3% received MDD. As a result, very few children (3%) received MAD. The percentages are somewhat lower in the Amhara sample where none of the children received MDD or MAD.

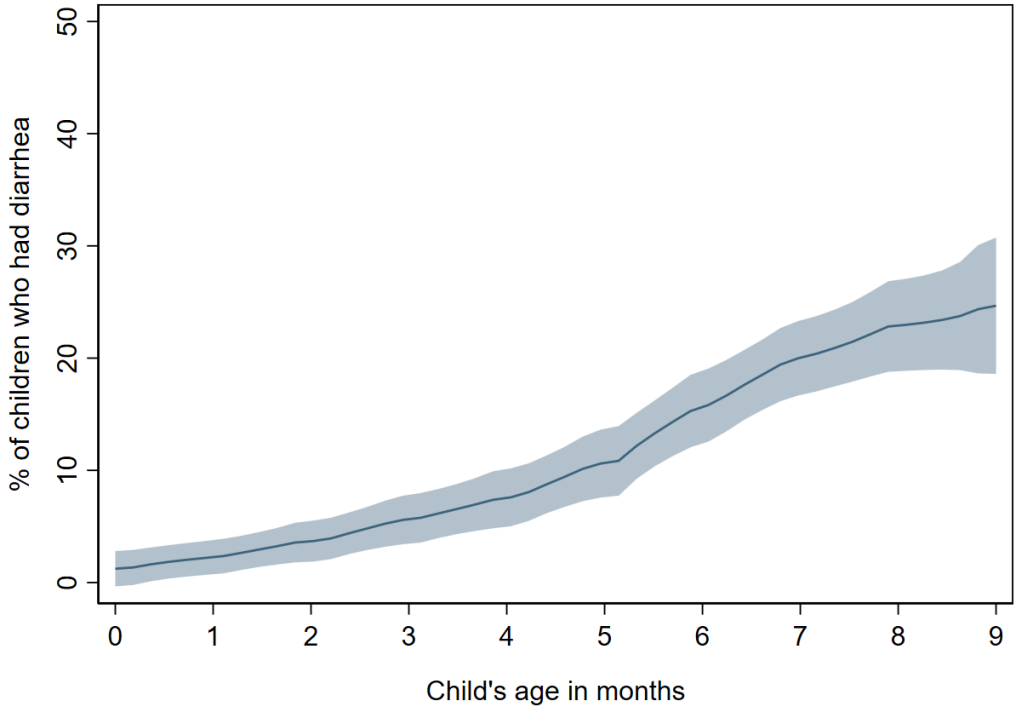
Table 19: Percentage of children 6–23 months receiving minimum meal frequency (MMF), minimum diet diversity (MDD), and minimum acceptable diet (MAD)

Indicator	Full Sample	Amhara	Oromia
Minimum meal frequency (MMF), percentage *	51.8	34.6	59.6
Minimum diet diversity (MDD), percentage *	3.1	0.0	4.5
Minimum acceptable diet (MAD), percentage *	3.1	0.0	4.5

*Note: N = 741 children, 6–23 months of age; MMF could not be calculated for two children. * Due to the sampling strategy, about 90 % the children in this sample are less than 10 months of age.*

Finally, the survey instrument asked caregivers if their child had had diarrhea in the 2 weeks before the interview. Due to the sampling strategy, virtually all children (99%) about whom this question was asked were less than 24 months of age. About 12% of the children had diarrhea during this interval. The diarrhea incidence was similar across the two regions: 10% in the Amhara and 13% in the Oromia sample. Restricting the data to children under 10 months of age that represent 90% of the sample, we see that the diarrhea risk increases linearly with age (Figure 10). About 64% of children who had diarrhea were treated with Oral Rehydration Therapy.

Figure 10: Diarrhea risk, by child's group



Note: Local polynomial regression. Shaded lines represent 95-percent confidence intervals. Vertical axis measures the share of children (in percentage) who had diarrhea in the 2 weeks before the interview. N= 1,428 children 0–9 months of age.

3.6 Women's Well-Being

Table 20 reports key required indicators around women's well-being. To measure women's dietary diversity, women report the number of food groups consumed over the previous 24 hours and a woman is identified as consuming a diet of the minimum required diversity if she consumed five or more food

groups. (The food groups are grains, white roots, tubers, and plantains; pulses such as beans, peas, and lentils; nuts and seeds, including groundnut; dairy; meat, poultry, and fish; eggs; dark green leafy vegetables; other vitamin A-rich fruits and vegetables; other vegetables; and other fruits.) In this sample, only 7% of women of reproductive age report consuming a diet of minimum dietary diversity, an extremely low rate that is consistent with evidence from other contexts, and also consistent with the evidence of low dietary diversity for infants and young children reported above.

We also measured access to antenatal care (a receipt of at least four antenatal care visits) during pregnancy. This indicator is defined only for women who report a completed pregnancy; given that roughly half of our sample is currently pregnant, this indicator is only reported for those women who were sampled with an infant under 9 months. Within this subsample, 42% of women reported that they received at least four antenatal care visits during pregnancy.

In relation to contraceptive use, women were asked if they have knowledge of modern family planning methods. Ninety percent of women reported knowing at least three of the modern family planning methods enumerated. We also further asked women living in a union if they have used any of these contraceptives in the past 12 months; 19% reported that they were using contraception. Out of those who stated they were using contraceptives, 73.7% reported they participated in the decision made around its use.

Table 20: Women's wellbeing indicators

	Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL11. Prevalence of women of reproductive age consuming a diet of minimum diversity (MDD-W)	6.87	5.54	8.21	3,011
BL26. Percentage of births receiving at least four antenatal care visits during pregnancy	41.75	38.64	44.86	1,406
BL36. Percentage of women in a union who have knowledge of modern family planning methods	89.91	88.78	91.04	2,736
BL37. Percentage of women in a union who made decisions to use modern family planning methods in the past 12 months	73.76	69.99	77.54	526
Percentage of women who report using contraceptives over the past year	19.26	17.78	20.74	2,736

Note: BL26 is reported only for the subsample of respondents who report a completed pregnancy and does not include the respondents who are currently pregnant. BL36 is reported only for women who are married, and BL37 is reported for those who are married and who reported they were using contraceptive in the past 12 month.

3.7 Asset Ownership

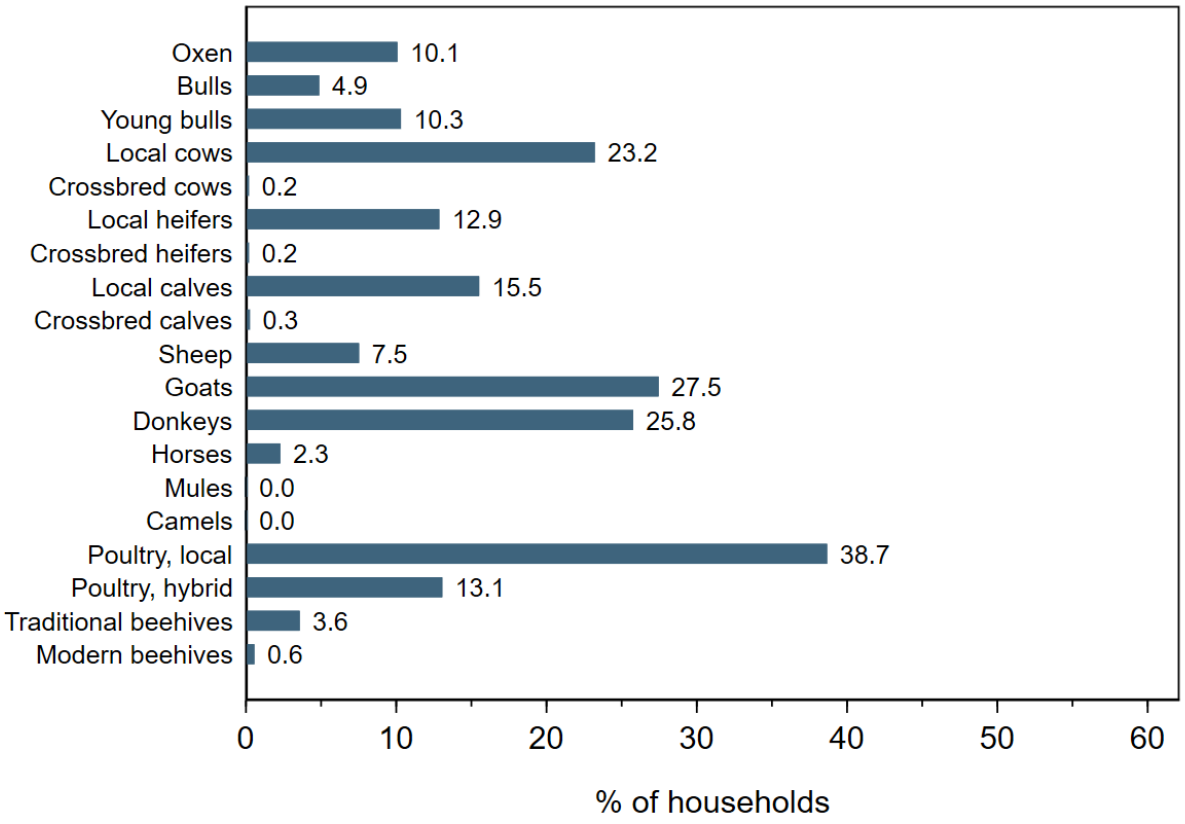
Livestock is one of the most important assets in rural Ethiopia, often serving both as a form of savings and a source of income. Figure 11 shows the share of households owning different types of livestock. Nearly 40% of the household report owning poultry. Cows, goats, and donkeys are also relatively

common with at least 20% of the households owning each livestock type. Livestock in this context is often of indigenous varieties that are typically less productive in producing milk or eggs than cross-bred or hybrid varieties.

To provide more context around the household's economic status and ownership of key productive assets, Table 21 reports the rates of ownership for key household assets. Ownership of simple agricultural assets is non-trivial, though in some cases, these assets are still not owned by a majority: around half of the households own a plow (wooden or metal), 77% own a sickle, 37% own a pickaxe, and about half own an axe. Ownership of spades is also very common.

With respect to household durable goods, ownership is extremely low: only 11% own a bed frame, almost no one owns tables, chairs, or stools, and only 12% report ownership of any type of improved stove. Approximately half of the households own a mobile phone and approximately half own a flashlight or a torch. Other types of consumer durables are very infrequently owned, with only 13% reporting ownership of a radio and fewer than 1% reporting ownership of a television.

Figure 11: Household livestock ownership



N = 3,015 households.

Table 21: Household asset ownership

	N	Mean
Household owns a plow (wood)	3,015	14.1
Household owns a plow (metal)	3,015	42.4

	N	Mean
Household owns a sickle	3,015	77.2
Household owns a pickaxe	3,015	36.5
Household owns an axe	3,015	49.9
Household owns a pruning/cutting shears	3,015	1.2
Household owns a gesso (a tool used to plough land)	3,015	32.6
Household owns a mencha (a machete-like tool to cut crass)	3,015	44.4
Household owns a hoe	3,015	22.6
Household owns a spade or shovel	3,015	72.5
Household owns a knapsack chemical sprayer	3,015	4.6
Household owns a water pump	3,015	0.5
Household owns a stone grain mill	3,015	20.1
Household owns a wheelbarrow	3,015	0.5
Household owns a cart	3,015	0.0
Household owns a bed frame (leather, wood, metal)	3,015	10.9
Household owns a mattress	3,015	56.7
Household owns a blanket/gabi	3,015	87.8
Household owns a wardrobe (for storing clothes)	3,015	1.2
Household owns a modern table	3,015	0.6
Household owns a modern chair	3,015	1.2
Household owns a wooden stool	3,015	7.7
Household owns a kerosene stove or an improved charcoal/wood stove	3,015	12.1
Household owns a cupboard (for storing food, dishes, cooking utensils)	3,015	6.9
Household owns an iron (for pressing clothes)	3,015	0.0
Household owns a radio	3,015	12.5
Household owns a mobile phone	3,015	49.4
Household owns a solar panel	3,015	84.6
Household owns a gold or gold jewelry (grams)	3,015	1.1
Household owns a silver or silver jewelry (grams)	3,015	6.0
Household owns a kerosene or other type of pressure lantern	3,015	4.4
Household owns a flashlight/torch	3,015	48.3
Household owns clocks	3,015	0.3

	N	Mean
Household owns wristwatches	3,015	2.2
Household owns a bicycle	3,015	0.1
Household owns a radio/tape recorder	3,015	2.3
Household owns a television	3,015	0.9

Notes: Estimates from the IMPEL baseline survey sample. All indicators are reported as percentages.

3.8 Agriculture, Gender Decision-Making, and Group Membership

This section summarizes findings around agriculture, gendered decision-making, and group membership. Table 22 reports key contextual indicators summarizing agricultural production in the two regions. The percentage of respondents who reported owning agricultural land is higher in Oromia (99.2%) than in Amhara (86.6%). Among households that do own land, virtually all households in both regions report cultivating crops over the past year. Ownership of animals is relatively high in both regions (70% in Amhara and 80% in Oromia.)

The percentage of farmers accessing agricultural credit is higher in Amhara, where 23.5% of the respondents reported taking agricultural credit, while in Oromia, only 9.1% of the respondents report accessing agricultural credit. Reported rates of cash savings are around 50% in both regions, while reported rates of accessing agricultural insurance are very low (1%) in both regions. At the same time, the average number of crop production practices adopted by farmers (out of 18 practices enumerated) is slightly higher in Oromia.[12]

Table 22: Summary indicators around agriculture

Indicator	N	Full Sample	Amhara	Oromia
Household owns land	3015	94.9	86.6	99.2
Cultivated crops over the past 12 months	2861	99.3	98.9	99.5
Household owns animals	3015	76.8	70.2	80.3
Accessed agricultural credit	3013	14.0	23.5	9.1
Saved cash in the past 12 month	3015	47.7	51.6	45.7
Accessed agricultural insurance over the past 12 months	2904	1.2	0.6	1.5
Number of crop production practices adopted	3015	5.1	4.0	5.7

Notes: The sample for the question around crop cultivation is restricted to the 2861 households who report owning land. All other questions are asked to the full sample; however, the question about agricultural insurance is posed to households who report owning land or animals.

[12] The enumerated practices include micro dosing, dry planting, weed control, crop rotation, intercropping, etc.

Table 23 then reports required BHA indicators linked to agriculture, access to cash and credit, and group membership. The percentage of farmers who report using any type of financial services (savings, credit, or insurance) over the past 12 months is around 53%. With respect to reported cash earnings, slightly more than half (54%) of men and women report earning cash. Among women who report earning cash, 68% report participation in decisions about its use; 61% report that they can participate in decisions about the use of a spouse's self-earned cash. Among men, 80% report that their spouse/partner can participate in decisions around the use of cash.

Membership in community groups is extremely high for both women and men, though higher for men: 71% of women report they are members of a community group and 80% of men. Fifty-two percent of women and 51% of men report access to credit; 65% of women who report access to credit report that they make decisions about credit and 71% of men report they make decisions about credit.

Table 23: Agriculture, cash and decision-making, credit, and group membership

	Mean (%)	95% Confidence Interval		N
		CI lower	CI upper	
BL32. Percentage of women and men in a union who earned cash in the past 12 months	53.91	51.50	56.32	4,823
Women	30.0	28.3	31.7	2,814
Men	87.4	86.0	88.9	2,009
BL33. Percentage of women in a union and earning cash who report participation in decisions about the use of self-earned cash	68.36	62.02	74.71	844
BL34. Percentage of women in a union and earning cash who report participation in decisions about the use of spouse/partner's self-earned cash	60.62	55.21	66.03	843
BL35. Percentage of men in a union and earning cash who report spouse/partner participation in decisions about the use of self-earned cash	79.67	77.33	82.01	1,756
BL41. Percentage of women in a union who are members of a community group	70.92	67.39	74.45	2,813
BL41. Percentage of men in a union who are members of a community group	79.89	77.11	82.67	2,009
BL42. Percentage of women in a union with access to credit	52.45	49.99	54.92	2,814
BL42. Percentage of men in a union with access to credit	50.92	48.13	53.71	2,009
BL43. Percentage of women in a union who make decisions about credit	64.88	60.89	68.87	1,475

	Mean (%)	95% Confidence Interval		N
		CI lower	CI upper	
BL43. Percentage of men in a union who make decisions about credit	70.77	66.24	75.30	1,023

Notes: Estimates from the IMPEL baseline survey sample.

Table 24 reports membership by type of community group; in this module, respondents are first asked to identify whether this type of group exists in their community and if yes, whether they are a member. The number of observations thus reflects the number of men and women reporting the presence of a certain type of community group. In general, male membership in community groups is higher across a range of organizational types; female membership in trade and business associations, in particular, is non-existent in Amhara. Participation in agriculture or livestock production groups is much higher in Oromia than in Amhara for both men and women. Membership in traditional saving and credit groups (equb), mutual help and insurance (idir), and religious organizations are common for men and women in both regions.

Table 24: Membership in community groups by gender and region

Group Membership	Female Group Membership				Male Group Membership			
	Amhara	Oromia	All	N	Amhara	Oromia	All	N
Agriculture / livestock producers' groups	10.4	51.1	46.7	443	12.7	66.2	57.9	354
Water users' group	39.7	52.8	47.0	791	46.4	62.4	57.0	567
Forest users' group	26.1	62.1	49.3	495	46.2	82.5	69.6	408
Microfinance group (equb)	40.0	63.2	55.0	860	44.6	56.3	52.4	607
Mutual help or insurance group (iddir)	83.1	77.1	79.0	992	89.1	75.8	79.5	660
Trade and business association	0.0	37.7	34.8	207	31.3	31.5	31.5	162
Civic or charitable group	56.3	50.2	50.8	317	63.3	55.6	56.4	280
Local government	11.9	37.4	28.6	3010	24.2	50.0	42.8	2012
Religious group	73.5	55.6	62.0	2428	86.1	58.7	66.7	1650

Notes: the number of observations N varies in this table given that data structure: each individual is first asked if such group exists in their village, and if reporting yes, then they are asked if they are a member of this group.

3.9 Interactions with HDAs, HEWs, and Health Services

Health Extension Workers (HEWs) are frontline health staff in rural Ethiopia whose role is to provide community-level services like antenatal and postnatal care, SBCC and WASH sessions, nutrition counseling, vaccination of children, growth monitoring and promotion, regular health check-ups, and more. Given the nutrition-related goals of this evaluation, understanding baseline patterns of interactions with HEWs in our sample is extremely useful. This information is summarized in Table 25.

In the baseline survey, women report contacts with HEWs to be by far the most common out of all health service interactions, with 86% of women reporting having any past contact with a HEW and 60% reporting interactions in the past 3 months. Home visits by a HEW were much less common, however: 19% of women were visited by a HEW at home in the past 3 months.

When asked about what services HEWs provide, an overwhelming majority—88%—of female respondents report immunizations. Antenatal care was identified by 38%, while neonatal and postnatal care was mentioned by only 5% and 14%, respectively. Counseling on child feeding was not a well-known service area, with 15% of women naming counseling on breastfeeding as a service provided by HEWs and 20% naming counseling on complementary feeding. Visits to health posts demonstrated a similar prevalence to interactions with HEWs: 84% of women have ever visited a post and 58% did so in the past 3 months.

This is not entirely aligned with the topics that women who had contact with a HEW recall discussing the last time they met. Immunizations were still the most common topic (72%), but breastfeeding was the second most recalled topic (29%). Around 20% of women reported previous discussions around both family planning and antenatal care, while all other topics are recalled by 10% of women or less.

HEWs also support and supervise Health Development Army (HDA) volunteers, who are trained health care providers without any professional certification. The HDAs provide additional services around curative and preventive healthcare, as well as the promotion of healthy behavior adoption. Only 22% of the women in the baseline survey report ever having contact with an HDA, although the majority of these contacts—14%—were reported to have occurred in the last 3 months. Home visits by HDAs are relatively more common than those by HEWs, as two-thirds of those who had had contact were visited at home.

Similar to HEWs, the most widely recalled topic discussed with HDAs during the last contact—and the only topic recalled by over half of the women—was immunizations. On the other hand, the second most remembered topics discussed with HDAs were sanitation, latrine use, and hygiene, recalled by 24% of those who reported previous interactions. This was not a topic identified as addressed by HEWs.

Women were also asked if they had ever attended a food demonstration, if they had attended a community conversation/gathering about breastfeeding, child feeding, or nutrition, and whether they listen to the radio. For all three of these information sources, roughly 20% of women reported any previous exposure. Interactions in the past 3 months were much less common: both food demonstrations and community conversations were attended by 7% of women, and nutrition information was heard on the radio by 6% in the past 3 months.

Table 25: Interactions with health services

	N	Mean
Female has had contact with a Health Extension Worker (HEW)	3,014	0.860
Female has had contact with HEW in the past 3 months	3,014	0.597
Female was visited by HEW at home in the past 3 months	3,014	0.193
Female has had contact with a Health Development Army (HDA) member *	2,835	0.221

	N	Mean
Female has had contact with an HDA in the past 3 months	2,835	0.139
Female was visited by HDA at home in the past 3 months	2,835	0.090
Female has visited a health post	3,015	0.843
Female has visited a health post related to her or her child in the past 3 months	3,015	0.584
Female has attended a food demonstration	3,015	0.199
Female has attended a food demonstration in the past 3 months	3,015	0.072
Female has attended a community conversation/gathering to talk about breastfeeding, child feeding, or nutrition	3,015	0.198
Female has attended a community conversation in the past 3 months	3,015	0.074
Female listens to the radio	3,015	0.225
Female has heard information about breastfeeding, child feeding or nutrition on the radio in the past 3 months	3,015	0.058

*Notes: * 180 respondents reported that they did not know whether they had previously had contact with a HDA member, leading to the lower number of observations for HDA-related variables.*

3.10 Balance of Baseline Characteristics across Treatment Arms

Table 26 reports the balance of characteristics across treatment arms for key indicators that are of interest to the evaluation. More specifically, we report the pooled mean, the mean in each treatment arm, and a p-value testing equality of the means in each pair-wise comparison. (For concision, we do not report the full set of BHA indicators; assessing balance in a subset of important indicators should be sufficient to assess balance.) Given the randomized design, we expect that, in general, both observable and unobservable characteristics should be balanced across treatment arms. This is what we observe for the observable characteristics: the p-values generally suggest that we cannot reject the hypothesis that all covariates are consistent across treatment arms.

We can also estimate a joint test across all characteristics to test the hypothesis that the observable characteristics are generally balanced across experimental arms T1, T2, and T3 across all characteristics reported. Given the large number of characteristics analyzed, it is unsurprising that two characteristics show minor imbalance that is significant at the 5% level, but the joint p-value allows us to assess the overall balance across arms. The p-value for this joint test is 0.167. Again, this suggests that the hypothesis that the characteristics are balanced cannot be rejected.

Table 26: Balance of characteristics across treatment arms

	Pooled Sample			Group Means			Tests of Equality		
	N	Mean	SD	T1: PSNP only	T2: SPIR II + NCG	T3: SPIR II + NCG + grants	T1 vs T2	T1 vs T3	T2 vs T3
BL06. Prevalence of moderate food insecurity in the household based on the Food Insecurity Experience Scale (FIES)	3,015	0.31	0.46	0.29	0.32	0.34	0.23	0.06	0.43
BL06. Prevalence of severe food insecurity in the household based on the Food Insecurity Experience Scale (FIES)	3,015	0.57	0.49	0.59	0.59	0.54	0.82	0.15	0.20
BL10. Percentage of households with poor food consumption score (FCS)	3,015	0.05	0.23	0.05	0.05	0.06	0.94	0.54	0.62
BL10. Percentage of households with borderline food consumption score (FCS)	3,015	0.19	0.39	0.19	0.20	0.18	0.72	0.67	0.45
BL10. Percentage of households with acceptable food consumption score (FCS)	3,015	0.75	0.43	0.76	0.74	0.76	0.74	0.96	0.72
BL16. Percentage of households using basic drinking water services based on 4 out of 5 criteria[13]	3,013	0.20	0.40	0.21	0.23	0.18	0.51	0.28	0.10
BL17. Percentage of households with soap and water at a handwashing station on premises	3,015	0.02	0.13	0.01	0.02	0.03	0.68	0.05	0.17
BL18. Percentage of households practicing correct use of recommended household water treatment technologies	3,009	0.11	0.31	0.13	0.11	0.09	0.55	0.05	0.24
BL19. Percentage of households practicing open defecation	3,015	0.50	0.50	0.48	0.51	0.52	0.54	0.39	0.82
BL27. Percentage of households with access to a basic sanitation service	3,015	0.17	0.38	0.19	0.16	0.16	0.23	0.19	0.88
BL12. Percentage of children 6–-23 months receiving a	741	0.03	0.17	0.03	0.05	0.02	0.38	0.62	0.24

[13] This indicator shows all households that meet four of the five criteria for basic drinking water services (excluding per person per day criterion). Quantity of water collected by the household was not collected and therefore not considered in the calculation. * Due to the sampling strategy, about 90 % the children in this sample are less than 10 months of age.

	Pooled Sample			Group Means			Tests of Equality		
	N	Mean	SD	T1: PSNP only	T2: SPIR II + NCG	T3: SPIR II + NCG + grants	T1 vs T2	T1 vs T3	T2 vs T3
minimum acceptable diet *									
BL13. Prevalence of exclusive breastfeeding of children under 6 months	870	0.79	0.41	0.77	0.78	0.83	0.85	0.08	0.10
BL14. Percentage of children under five (0–59 months) who had diarrhea in the prior 2 weeks *	1,627	0.12	0.33	0.12	0.11	0.13	0.72	0.90	0.59
BL15. Percentage of children under five (0–59months) with diarrhea treated with Oral Rehydration Therapy *	188	0.64	0.48	0.58	0.75	0.60	0.05	0.86	0.06
BL39. Prevalence of children 6–23 months consuming a diet of minimum diversity (MDD-C) *	743	0.03	0.17	0.03	0.05	0.02	0.38	0.62	0.24
BL11. Prevalence of women of reproductive age consuming a diet of minimum diversity (MDD-W)	3,011	0.07	0.25	0.07	0.07	0.07	0.87	0.99	0.86
BL26. Percentage of births receiving at least four antenatal care visits during pregnancy	1,406	0.42	0.49	0.42	0.38	0.45	0.26	0.58	0.10
BL29. Percentage of farmers who used financial services (savings, agricultural credit, and/or insurance) in the past 12 months	2,902	0.53	0.50	0.52	0.52	0.55	0.81	0.39	0.55
BL40. Daily per capita expenditures (as a proxy for income) in USG-assisted areas, in 2010 USD	3,015	1.72	0.92	1.71	1.71	1.74	0.98	0.70	0.69
BL01. Prevalence of poverty: Percentage of people living on less than $1.90/day 2011 PPP	16,264	0.71	0.45	0.71	0.72	0.70	0.56	0.75	0.39
BL02. Depth of poverty of the poor: Mean percentage shortfall of the poor relative to the $1.90/day 2011 PPP poverty line	11,541	34.21	19.03	34.59	34.76	33.23	0.90	0.39	0.32

Notes: Estimates from the IMPEL baseline survey sample. Columns 7–9 report p-values from tests of difference of means between the treatment arms.

3.11 Baseline Characteristics Disaggregated by Demographic Groups

Table 27 to Table 34 report baseline characteristics disaggregated by the specified demographic characteristics. For household structure, it is important to note that by construction, there are no households in this sample characterized by the presence of an adult male no female and no households characterized by the presence of children only. One of the inclusion criteria for entry into this sample is the presence of a woman of reproductive age.

In comparing households characterized by types female and male adults (F&M) and adult female and no adult male (FNM) with respect to household food access (Table 27), WASH indicators (Table 28), children's nutritional status and feeding practices (Table 29), agricultural indicators (Table 31), and poverty indicators (Table 32), there is generally little evidence of meaningful differences comparing across these household types. The difference in means is small in magnitude and varying in sign (i.e., for some indicators F&M households show evidence of slightly better status, while for some indicators FNM shows evidence of slightly better status). Table 30 reports indicators of women's well-being disaggregated by age and again there is very limited evidence of heterogeneity with respect to age.

Table 33 and Table 34 report indicators related to gender decision-making and group membership disaggregated by gender and age. The probability of reporting owning cash is dramatically higher for men (87.4%) versus women (30%). However, conditional on gender, there is limited variation concerning age. There is also relatively little variation in decision-making patterns around cash with respect to age.

For credit access, the gender gap in reported access is quite small: 53% for women and 51% for men. The gender gap in reported community group membership is also relatively small: 71% for women and 80% for men. Variation concerning age for credit access and community group membership is also minor.

Table 27: Household food access, disaggregated

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL06. Prevalence of moderate food insecurity in the household based on the Food Insecurity Experience Scale (FIES)				
Female and Male Adults (F&M)	30.9	29.2	32.7	2,805
Adult Female no Adult Male (FNM)	37.1	30.6	43.7	210
BL06. Prevalence of severe food insecurity in the household based on the Food Insecurity Experience Scale (FIES)				
Female and Male Adults (F&M)	57.7	55.9	59.5	2,805
Adult Female no Adult Male (FNM)	53.8	47.0	60.6	210
BL10. Percentage of households with poor food consumption score (FCS)				
Female and Male Adults (F&M)	5.5	4.6	6.3	2,805
Adult Female no Adult Male (FNM)	5.7	2.5	8.9	210
BL10. Percentage of households with borderline food consumption score (FCS)				
Female and Male Adults (F&M)	19.9	18.4	21.4	2,805
Adult Female no Adult Male (FNM)	11.0	6.7	15.2	210
BL10. Percentage of households with acceptable food consumption score (FCS)				
Female and Male Adults (F&M)	74.6	73.0	76.2	2,805
Adult Female no Adult Male (FNM)	83.3	78.3	88.4	210

Notes: Estimates from the IMPEL baseline survey sample.

Table 28: Water, sanitation and hygiene (WASH) indicators, disaggregated

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL16. Percentage of households using basic drinking water services[14]				
Female and Male Adults (F&M)	20.6	19.1	22.1	2,803
Adult Female no Adult Male (FNM)	19.0	13.7	24.4	210
BL17. Percentage of households with soap and water at a handwashing station on premises				
Female and Male Adults (F&M)	1.6	1.1	2.0	2,805
Adult Female no Adult Male (FNM)	4.3	1.5	7.0	210
BL19. Percentage of households practicing open defecation				
Female and Male Adults (F&M)	49.4	47.5	51.2	2,805
Adult Female no Adult Male (FNM)	63.3	56.8	69.9	210
BL27. Percentage of households with access to a basic sanitation service				
Female and Male Adults (F&M)	17.3	15.9	18.7	2,805
Adult Female no Adult Male (FNM)	14.3	9.5	19.1	210
BL18. Percentage of households practicing correct use of recommended household water treatment technologies				
Chlorination	2.7	2.1	3.3	3,009
Flocculant/disinfectant	2.9	2.3	3.5	3,009
Filtration (physical removal)	4.0	3.3	4.7	3,009
Boiling	3.4	2.7	4.0	3,009

Notes: Estimates from the IMPEL baseline survey sample.

[14] This shows all households that meet four of the five criteria for basic drinking water services (excluding per person per day criterion). Quantity of water collected by the household was not collected and therefore not considered in the calculation.

Table 29: Children's nutritional status and feeding practices, disaggregated[15]

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL12. Percentage of children 6–23 months receiving a minimum acceptable diet				
Male child	3.5	1.6	5.4	371
Female child	2.7	1.0	4.4	370
BL13. Prevalence of exclusive breastfeeding of children under 6 months				
Male child	76.2	72.3	80.1	454
Female child	82.5	78.8	86.1	416
BL14. Percentage of children under five (0–59 months) who had diarrhea in the prior 2 weeks				
Male child	12.6	10.4	14.9	831
Female child	11.6	9.3	13.8	796
BL15. Percentage of children under five (0–59months) with diarrhea treated with Oral Rehydration Therapy				
Male child	61.8	52.2	71.4	102
Female child	66.3	56.1	76.5	86
BL39. Prevalence of children 6–23 months consuming a diet of minimum diversity (MDD-C)				
Male child	3.5	1.6	5.4	372
Female child	2.7	1.0	4.4	371

Notes: Estimates from the IMPEL baseline survey sample.

[15] Data on these indicators were collected from all children 0-59 months residing in the household. Given the sampling frame, younger (< 10 months) children (within each age-bracket) tend to be over-represented in these statistics.

Table 30: Women's wellbeing indicators, disaggregated

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL11. Percentage of women of reproductive age consuming a diet of minimum diversity (MDD-W)				
Younger than 19 years of age	4.9	0.1	9.8	81
19 years of age or older	6.9	6.0	7.8	2,930
BL36. Percentage of women in a union who have knowledge of modern family planning methods that can be used to delay or avoid pregnancy				
15–19 years of age	16.0	9.3	22.6	119
20–29 years of age	20.4	18.3	22.6	1,384
30–49 years of age	18.3	16.1	20.5	1,230

Notes: Estimates from the IMPEL baseline survey sample.

Table 31: Agricultural indicators, disaggregated

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL29. Percentage of farmers who used financial services (savings, agricultural credit, insurance) in the past 12 months				
Female and Male Adults (F&M)	53.1	51.2	54.9	2,764
Adult Female no Adult Male (FNM)	48.6	40.1	57.0	138

Notes: Estimates from the IMPEL baseline survey sample.

Table 32: Poverty indicators, disaggregated

	Indicator Mean (%)[16]	95% Confidence Interval		N
		CI Lower	CI Upper	
BL40. Daily per capita expenditures (as a proxy for income) in USG-assisted areas, in 2010 USD				
Female and Male Adults (F&M)	1.7	1.7	1.8	2,805
Adult Female no Adult Male (FNM)	1.7	1.5	1.8	210
BL01. Prevalence of poverty: Percentage of people living on less than $1.90/day 2011 PPP				
Female and Male Adults (F&M)	65.7	63.9	67.4	2,805
Adult Female no Adult Male (FNM)	67.6	61.2	74.0	210
BL02. Depth of poverty of the poor: Mean percentage shortfall of the poor relative to the $1.90/day 2011 PPP poverty line				
Female and Male Adults (F&M)	32.9	32.1	33.8	1,842
Adult Female no Adult Male (FNM)	35.6	32.3	38.8	142

Notes: Estimates from the IMPEL baseline survey sample.

Table 33: Gender decision-making and group membership, disaggregated (1st level)

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL32. Percentage of women and men in a union who earned cash in the past 12 months				
Women	30.0	28.3	31.7	2,814
Men	87.4	86.0	88.9	2,009
BL33. Percentage of women in a union and earning cash who report participation in decisions about the use of self-earned cash				
15–19 years of age	73.1	54.8	91.3	26

[16] Except from daily expenditures that are expressed in USD.

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
20–29 years of age	71.7	67.3	76.1	410
30–49 years of age	64.7	60.0	69.4	408
>=50 years of age				0
BL34. Percentage of women in a union and earning cash who report participation in decisions about the use of spouse/partner's self-earned cash				
15–19 years of age	66.7	47.7	85.7	27
20–29 years of age	64.2	59.6	68.7	424
30–49 years of age	56.4	51.4	61.3	392
>=50 years of age				0
BL35. Percentage of men in a union and earning cash who report spouse/partner participation in decisions about the use of self-earned cash				
15–19 years of age	66.2	54.9	77.5	71
20–29 years of age	78.6	75.9	81.3	903
30–49 years of age	82.2	79.5	84.9	776
>=50 years of age	66.7	12.5	120.9	6
BL43. Percentage of women/men in a union who make decisions about credit				
Decisions made alone	26.6	24.9	28.3	2,595
Decisions made jointly	41.6	39.7	43.5	2,595
Decisions made by spouse/partner	6.7	5.7	7.7	2,595

Notes: Estimates from the IMPEL baseline survey sample.

Table 34: Gender decision-making and group membership, disaggregated (2nd level)

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL32. Percentage of women and men in a union who earned cash in the past 12 months				
Women: 15–19 years	22.0	14.4	29.6	118
Women: 20–29 years	28.6	26.2	30.9	1,435
Women: 30–49 years	32.4	29.8	35.0	1,259
Women: >=50 years	0.0	0.0	0.0	2
Men: 15–19 years	89.9	83.1	96.7	79
Men: 20–29 years	86.7	84.7	88.8	1,041
Men: >=30 years	88.0	85.8	90.1	882
Men: >=50 years	85.7	50.8	120.7	7
BL42. Percentage of women/men in a union with access to credit				
Women: 15–19 years	51.7	42.5	60.8	118
Women: 20–29 years	51.3	48.7	53.9	1,435
Women: 30–49 years	53.9	51.1	56.6	1,259
Women: >=50 years	50.0	-585.3	685.3	2
Men: 15–19 years	53.2	41.9	64.4	79
Men: 20–29 years	50.3	47.3	53.4	1,041
Men: 30–49 years	51.6	48.3	54.9	882
Men: >=50 years	28.6	-16.6	73.7	7
BL43. Percentage of women/men in a union who make decisions about credit				
Women: 15–19 years	54.1	41.2	67.0	61

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
Women: 20–29 years	68.1	64.7	71.4	736
Women: 30–49 years	62.5	58.8	66.1	677
Women: >=50 years	0.0			1
Men: 15–19 years	83.3	71.6	95.1	42
Men: 20–29 years	74.8	71.1	78.5	524
Men: 30–49 years	65.1	60.7	69.5	455
Men: >=50 years	50.0	-585.3	685.3	2
BL41. Percentage of women/men in a union who are members of a community group				
Women: 15–19 years	65.3	56.5	74.0	118
Women: 20–29 years	71.7	69.4	74.0	1,434
Women: 30–49 years	70.5	68.0	73.1	1,259
Women: >=50 years	100.0	100.0	100.0	2
Men: 15–19 years	73.4	63.5	83.4	79
Men: 20–29 years	80.0	77.6	82.5	1,041
Men: 30–49 years	80.2	77.5	82.8	882
Men: >=50 years	100.0	100.0	100.0	7
BL37. Percentage of women in a union who made decisions about modern family planning methods in the past 12 months				
Age 15–19				
Decisions made alone	25.0	4.2	45.8	20
Decisions made jointly	60.0	36.5	83.5	20
Decisions made by spouse/partner	15.0	-2.1	32.1	20

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
Age 20–29				
Decisions made alone	22.3	17.4	27.2	282
Decisions made jointly	54.6	48.8	60.5	282
Decisions made by spouse/partner	23.0	18.1	28.0	282
Age 20–49				
Decisions made alone	19.6	14.4	24.9	224
Decisions made jointly	49.1	42.5	55.7	224
Decisions made by spouse/partner	31.3	25.1	37.4	224

Notes: Estimates from the IMPEL baseline survey sample.

Table 35: Gender decision-making and group membership, disaggregated (3rd level)

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL43. Percentage of women/men in a union who make decisions about credit				
Women 15–19 years: decisions made alone	25.0	14.4	35.6	68
Women 15–19 years: decisions made jointly	29.4	18.3	40.5	68
Women 15–19 years: decisions made by spouse/partner	13.2	5.0	21.5	68
Women 20–29 years: decisions made alone	31.6	28.3	34.9	778
Women 20–29 years: decisions made jointly	37.7	34.2	41.1	778
Women 20–29 years: decisions made by spouse/partner	9.9	7.8	12.0	778
Women 30–49 years: decisions made alone	33.2	29.7	36.6	720

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
Women 30–49 years: decisions made jointly	31.5	28.1	34.9	720
Women 30–49 years: decisions made by spouse/partner	6.7	4.8	8.5	720
Women >=50 years: decisions made alone	33.3	-110.1	176.8	3
Women >=50 years: decisions made jointly	0.0	0.0	0.0	3
Women >=50 years: decisions made by spouse/partner	0.0	0.0	0.0	3
Men 15–19 years: decisions made alone	38.6	23.7	53.6	44
Men 15–19 years: decisions made jointly	45.5	30.1	60.8	44
Men 15–19 years: decisions made by spouse/partner	0.0	0.0	0.0	44
Men 20–29 years: decisions made alone	19.0	15.7	22.4	525
Men 20–29 years: decisions made jointly	55.8	51.5	60.1	525
Men 20–29 years: decisions made by spouse/partner	4.4	2.6	6.1	525
Men 30–49 years: decisions made alone	15.4	12.1	18.7	455
Men 30–49 years: decisions made jointly	49.7	45.1	54.3	455
Men 30–49 years: decisions made by spouse/partner	3.7	2.0	5.5	455
Men >=50 years: decisions made alone	0.0	0.0	0.0	2
Men >=50 years: decisions made jointly	50.0	-585.3	685.3	2
Men >=50 years: decisions made by spouse/partner	0.0	0.0	0.0	2

Notes: Estimates from the IMPEL baseline survey sample.

4. EXPERIENCE OF RECENT SHOCKS AND TRENDS IN FOOD INSECURITY

In this section, we summarize the prevalence of shocks faced by PSNP5 households participating in SPIR II and compare their food insecurity in 2022 to that of PSNP4 households in the 2021 SPIR I endline sample. We also explore the association between household food insecurity in the SPIR II baseline and shocks experienced in the past 12 months. This will help inform how the current context and changes in PSNP targeting shape the indicators reported in this report.

4.1 Prevalence of Shocks

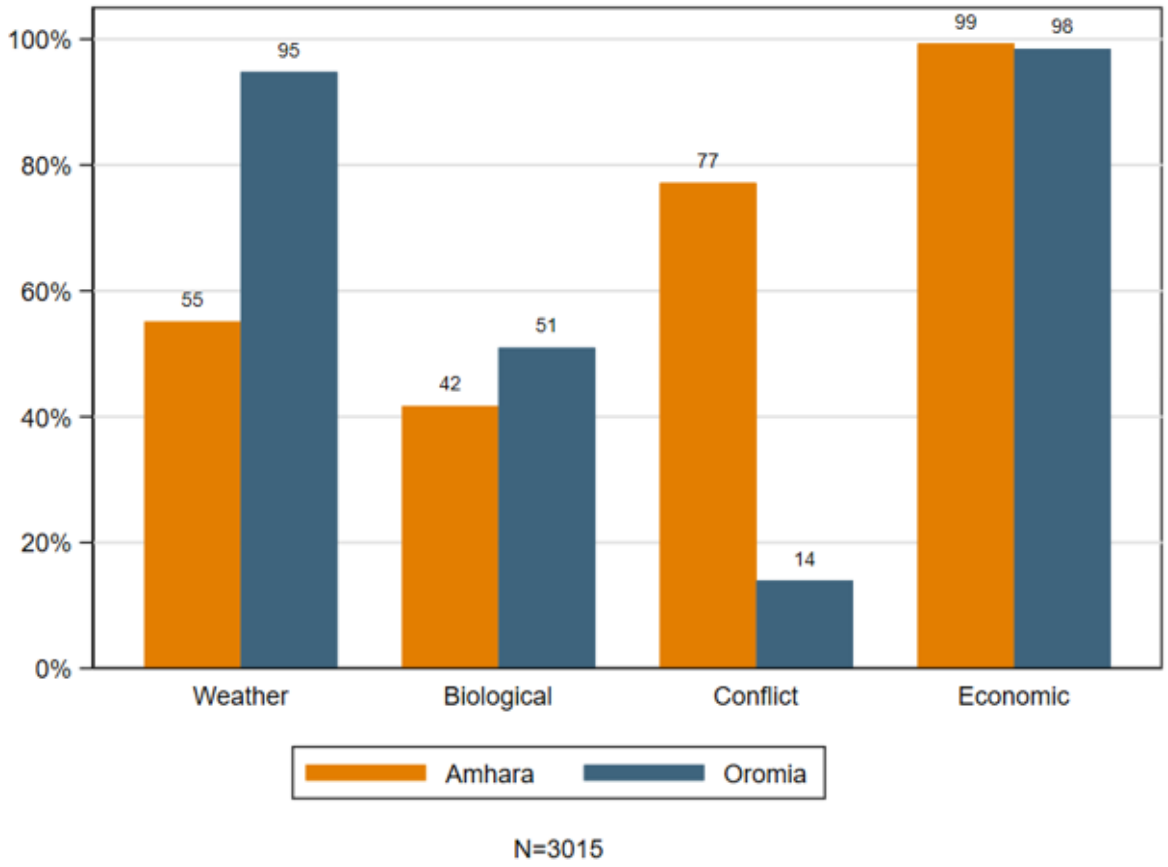
Between late 2021 and the baseline survey in August–October 2022, several significant shocks affected households living in Amhara and Oromia. The conflict in Tigray became a major source of violence and caused temporary displacement in some woredas in Amhara that are close to the Tigray border. Parts of Oromia were affected by a drought that UN-OCHA noted was among the worst to affect this region in 40 years. In addition, increases in international prices of fuel, food, and fertilizer caused initially by the COVID-19 pandemic and exacerbated by the war in Ukraine and drought in the United States and Canada put pressure on input and food prices in Ethiopia. These multiple shocks contributed to a growing food crisis in Ethiopia in 2022. Here, we present evidence on the prevalence of these shocks in the study sample.

In the baseline survey, respondents were asked to report whether their household had experienced each of several shocks in the past 12 months. We grouped these shocks into the following categories:

- **Weather:** excessive rains, flooding, too little rain/drought, variable rain (early/late), hail/frost, landslides/erosion;
- **Biological:** crop disease (rust on wheat, sorghum), crop pests (locusts), weeds (e.g., associated with striga), livestock disease, human disease outbreaks (from contaminated water);
- **Conflict:** theft or destruction of assets, theft of livestock (raids), land conflict, water conflict, violence, gender-based violence; and
- **Economic:** delay in food or cash assistance, increasing food prices, increased prices of agricultural or livestock inputs, decreased prices for agricultural or livestock products, loss of land/rental property, unemployment, death or a long-term illness of household member, and non-functioning boreholes.

Figure 12 summarizes the prevalence of these shocks in the Amhara and Oromia regions. Virtually all households in both regions reported experiencing an economic shock in the past 12 months. Weather-related shocks were common in both regions, but were especially prevalent in Oromia, affecting 95% of households, while weather affected 55% of households in Amhara. Conflict was a major event in Amhara, affecting more than three out of every four households. Biological shocks, including crop and livestock diseases and pests, as well as human diseases and infections, afflicted 42% of households in Amhara and more than 51% in Oromia.

Figure 12: Prevalence of shocks by type and region



The prevalence of specific shocks that comprise these aggregate indices are reported, by order of prevalence, for Amhara and Oromia in Figure 13 and Figure 14, respectively. In Amhara, increasing food prices were the most prevalent shock, affecting 98% of households. Seventy-five percent of households reported experiencing violence, which underscores the widespread effects of the war for households living in these northern woredas in Amhara. The next most prevalent shocks included unemployment and increased input prices (nearly 50% each), followed by variability in the timing of rains or too little rain (each affecting 33% of households). Thirty-two percent of households reported having difficulty with delays in receipt of food or cash assistance and 30% reported unusually bad problems with weeds. Loss of land or rental property, representing a loss of a significant asset and source of income, affected 22% of households.

In Oromia, increasing food prices were also the most prevalent shock reported, affecting 97% of households. Drought was the next most prevalent shock, which affected more than 86% of households. Variability in the timing of rains was reported by 80% of households, putting it fourth in shock prevalence in Oromia. Increasing input prices affected 84% of households. Nearly half of all households in Oromia (48%) reported decreased output prices. After that, the next most prominent shocks were unemployment (38%), delays in food or cash assistance (37%), crop disease (36%), and crop pests like locusts (26%). Twenty-three percent of households lost land or rental property.

These shock patterns indicate a period of significant crisis for these households in which they are facing multiple shocks either concurrently or in rapid succession. The data also provide detailed evidence

around the extent of exposure to violence experienced by households in the North Wollo and Wag Hemra zones of northern Amhara, confirming that the vast majority of households have been affected by the violence. The data also show that a large proportion of households in the zones of East and West Hararghe and West Arsi in Oromia were affected by drought.

At the same time that households faced conflict and weather shocks that pose a threat to safety and stability or a risk to agricultural and livestock earnings, they also had to cope with the challenges of the food and input price crisis. The data show that the well-documented increases in international prices for food and fertilizer in 2022 (Food Security Information Network, 2022) also affect poor households in rural Ethiopia, although the price increases reported in the data are likely also exacerbated by local factors, including the conflict and weather shocks.

Figure 13: Prevalence of shocks in the past 12 months, Amhara

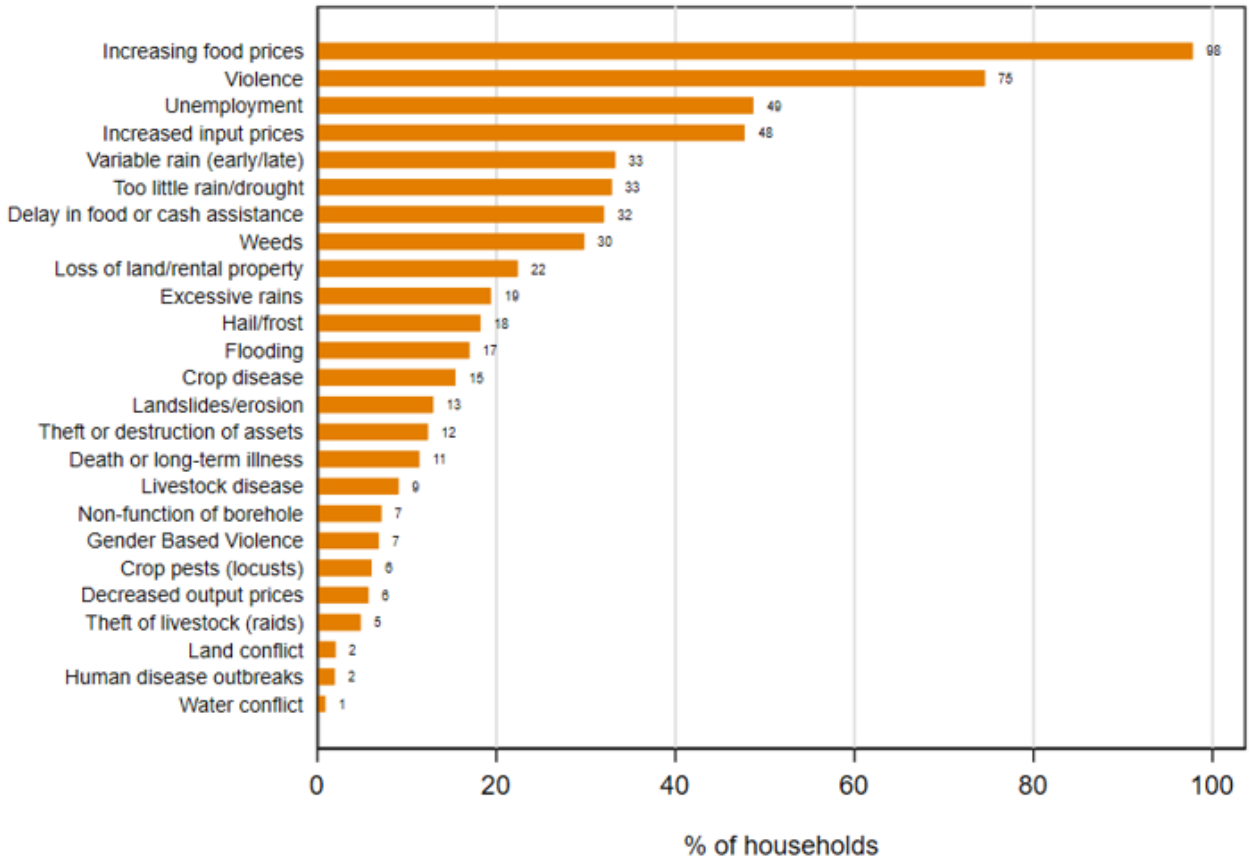
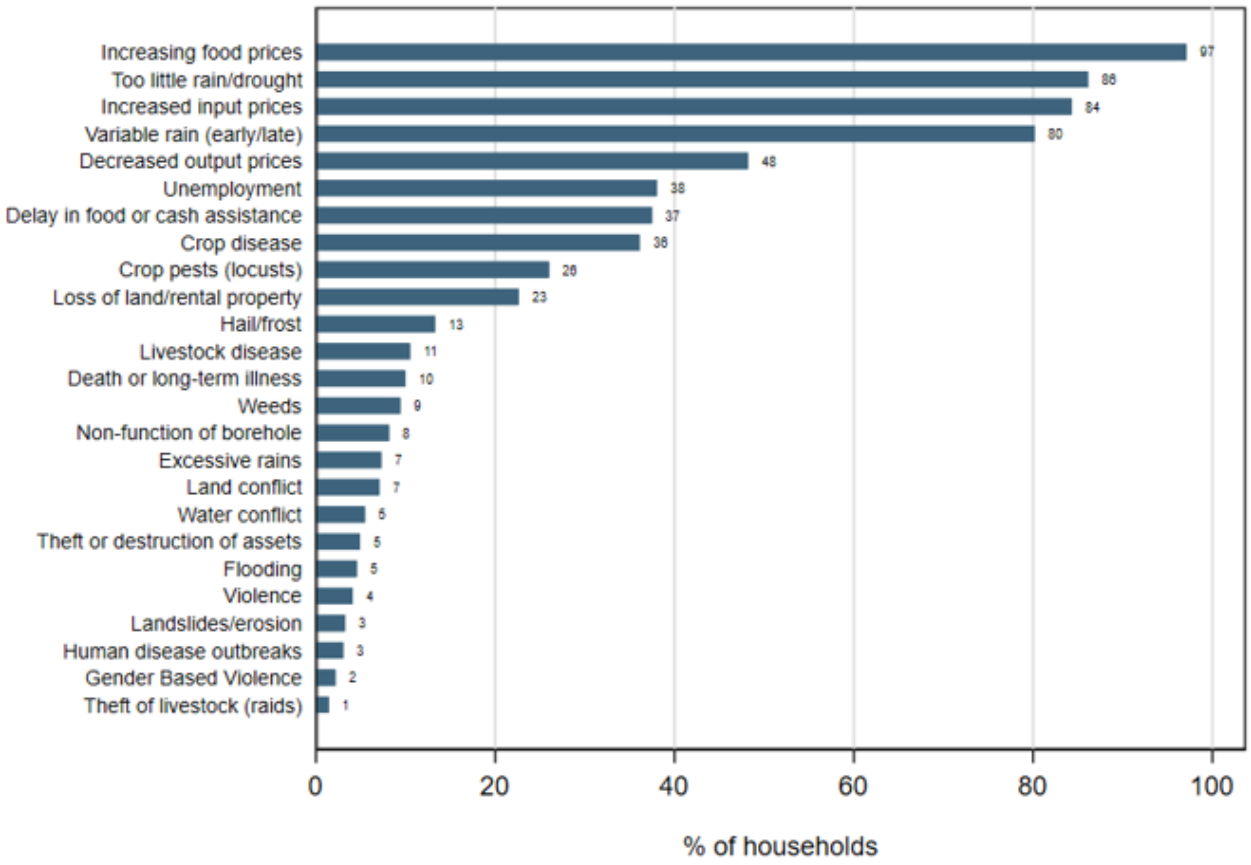


Figure 14: Prevalence of shocks in the past 12 months, Oromia

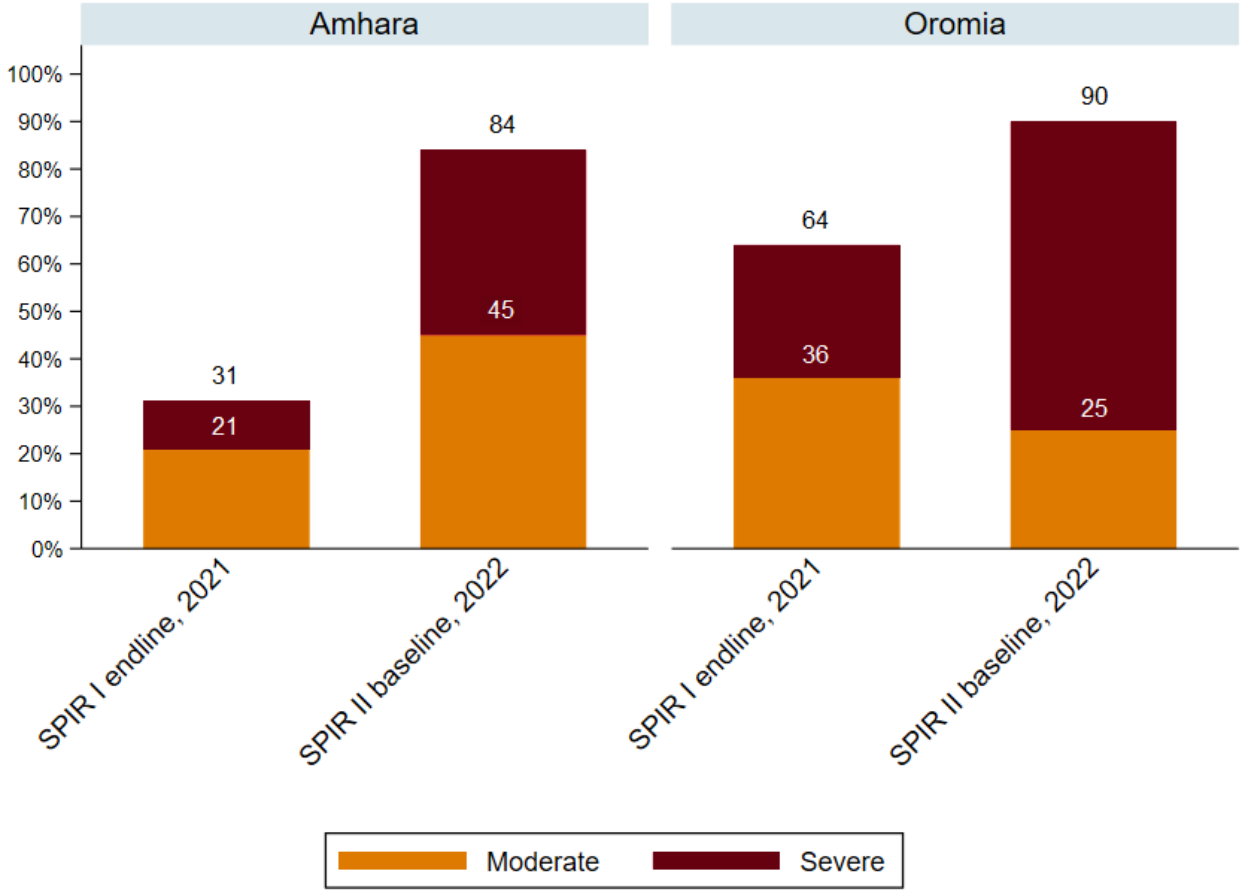


4.2 Differences in Food Insecurity between the 2022 SPIR II Baseline and 2021 SPIR I Endline

To further place the baseline survey results in context, we also briefly summarize differences in household food insecurity between the SPIR II baseline survey in August-October 2022 and the SPIR I endline survey in February-April 2021. This comparison is based on the Food Insecurity Experience Scale (FIES) described in Section 3.

Figure 15 shows that, both in Amhara and in Oromia, food insecurity measured by the FIES was substantially higher in the SPIR II baseline survey than in the SPIR I endline survey. In the SPIR I endline, 31% of households in Amhara experienced moderate-to-severe food insecurity, while in the SPIR II baseline, this figure was 84%, with 45% of households being moderately food insecure and 39% of households being severely food insecure. In Oromia, 36% of households were moderately food insecure and 28% were severely food insecure in the SPIR I endline survey. In the SPIR II baseline, moderate food insecurity in Oromia was lower, at 25%, but severe food insecurity was higher, at 65%. In all, 90% of households in Oromia in the SPIR II baseline experienced moderate-to-severe food insecurity.

Figure 15: Prevalence of moderate-to-severe food insecurity by region in the SPIR I endline and SPIR II baseline surveys



The much higher food insecurity in the SPIR II baseline survey reflects several factors: (i) differences in seasons between the SPIR I endline and SPIR II baseline, (ii) changes in targeting between PSNP4 and PSNP5, and (iii) the widespread exposure to the conflict, drought, and price shocks documented above. We now consider each of these factors in turn.

Regarding the role of seasonality, the SPIR I endline took place in the post-harvest season, when the food security situation is relatively better. The SPIR II baseline took place at the height of the hunger season. This mainly affects the comparison of indicators with a short recall period (24 hours, 7 days, 1 month), but recall may also be influenced more by recent events over long recall periods, like the 12-month recall used for the FIES, due to a form of recall bias known as recall decay (Beegle et al. 2012).

In addition, new targeting criteria implemented at the start of PSNP5 in September 2021 led to the selection of households that were poorer on average than under PSNP4 as described in the previous section. In particular, PSNP5 households are more likely to be landless and have more limited livestock holdings.

Finally, although data on shocks from the SPIR I endline and SPIR II baseline capture the prevalence of shocks rather than their severity, the large number of shocks and the type of shocks that were most prevalent suggest that exposure to shocks may have contributed more to food insecurity in the period

leading up to the SPIR II baseline than before the SPIR I endline. The SPIR I endline report (Alderman et al. 2021) documented that households had experienced significant shocks, especially drought and the effect of pests (fall armyworm and desert locusts) on crops and grazing lands. However, the data suggest that the severity of these shocks was lower on average than the shocks experienced in the 12 months before the SPIR II baseline survey.

To learn more about the effect of recent shocks on household food insecurity in the SPIR II baseline survey, we ran a regression of household food insecurity (the FIES) against household indicators for each of the seven most prominent shocks reported in the baseline survey. We included kebele fixed effects in the regression model to control for unobserved factors driving average differences in food insecurity at the local level. The results of this regression model are presented in Table 36. The average FIES in the data is 5.8 on a scale from 0 to 8.

In this table, each coefficient can be interpreted as follows: the coefficient captures the increase in the total FIES score observed for households that report experiencing a particular shock, relative to households who do not report experiencing this shock. Below the coefficient, the standard error is reported in parentheses; the standard error captures the statistical uncertainty associated with the main estimate. If the estimate is statistically significant (i.e., we can conclude it is significantly different from zero), the estimate is denoted with asterisks (a single asterisk denotes significance at the 10% level, two asterisks denotes significance at the 5% level, and three asterisks denotes significant at the 1% level). For example, the first coefficient can be interpreted as follows: households who report experiencing increasing food prices in the past 12 months show evidence of a FIES score that is .06 points higher, but this difference is not statistically significant.

The results show that the shocks that were most associated with increasing food insecurity were exposure to higher input prices, loss of land or rental property, delays in food or cash assistance, and unemployment. This measure of food insecurity was not associated with the increases in food prices, but this is likely because nearly all households experienced higher food prices. Interestingly, exposure to violence and exposure to drought in the past 12 months was not associated with higher food insecurity. We also checked whether violence was associated with food insecurity only in Amhara, but there was no significant association in Amhara alone. Similarly, drought was significantly associated with food insecurity in Oromia alone, where exposure to drought was more prevalent. Also, variability in the timing of rain is associated with a reduction in food insecurity, a result that we cannot easily explain. Overall, these results show that many of the shocks experienced by households in the year before the baseline survey contributed substantially to household food insecurity, although we do not find significant effects of violence or drought on this measure of food insecurity.

Table 36: The relationship of self-reported shocks to the household Food Insecurity Experience Scale

	Total FIES Score
Experienced increasing food prices in the past 12 months	0.060
	(0.27)
Experienced increasing input prices in the past 12 months	0.349***
	(3.83)
Experienced violence in the past 12 months	0.169
	(1.37)
Experienced too little rain/drought in the past 12 months	0.026
	(0.25)
Experienced variable rain (early/late) in the past 12 months	-0.271***
	(2.72)
Experienced unemployment in the past 12 months	0.191**
	(2.36)
Experienced delay in food or cash assistance in the past 12 months	0.259***
	(3.20)
Experienced loss of land or rental property in the past 12 months	0.271***
	(2.87)
Constant	5.829***
	(10.53)
R^2	0.24
N	3,014

*Notes: * p<0.1 ** p<0.05 *** p<0.01*

5. CONCLUSION

The objective of this report was to report on the baseline survey conducted as part of the IMPEL evaluation of SPIR II, a randomized controlled trial launched in 2022. The IMPEL SPIR II impact evaluation employs an experimental design with multiple treatment arms comparing combinations of livelihood and nutrition graduation model programming provided to PSNP beneficiaries relative to a control group receiving only PSNP transfers. The design includes 234 kebeles assigned to three treatment arms. In the first arm, PSNP is implemented by the government with SPIR II support for the provision of cash and food transfers only (no supplemental programming). In the second arm, SPIR II programming is rolled out in conjunction with nurturing care groups targeting enhanced infant and young child nutritional practices. In the third arm, SPIR II programming and NCGs are supplemented with targeted cash grants to pregnant and lactating women.

The baseline survey for the evaluation was conducted in August–September 2022. A sample of 3,015 households in 234 kebeles was included in the survey; the inclusion criteria required that households were PSNP beneficiaries reporting the presence of an infant under 9 months of age (along with his / her mother or primary caretaker) or a pregnant woman. The baseline survey included a survey of the primary female (mother of an infant or pregnant woman) as well as her spouse and collected data on a wide variety of indicators.

Summary statistics for the baseline survey suggest the sample for this evaluation is characterized by a high level of food insecurity and poverty and generally poor nutritional outcomes for women and young children. There is also evidence that this population has been heavily affected by recent adverse shocks, including conflict and drought. Weather-related shocks were common in both regions but were especially prevalent in Oromia. Conflict was a major event in Amhara, affecting more than three out of every four households. Biological shocks, including crop and livestock diseases and pests, as well as human diseases and infections, affected nearly half of the households in both regions.

In addition to reporting general summary statistics, the baseline survey can be used to assess balance in covariates across treatment arms. In general, the randomization achieved balance: the hypothesis that there is no difference in observable characteristics across arms cannot be rejected. This baseline survey will serve as the foundation for the randomized controlled trial. It will be followed by a midline survey (2023) and an endline survey (2025) that will be used to assess the impact of the interventions of interest on primary and secondary outcomes.

REFERENCES

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* National Bureau of Economic Research.

Alderman, H., & Headey, D. (2018). The timing of growth faltering has important implications for observational analyses of the underlying determinants of nutrition outcomes. *PloS One*, *13*(4), e0195904.

Andrews, C., Hsiao, A., & Ralston, L. (2018). Social safety nets promote poverty reduction, increase resilience, and expand opportunities. *Realizing the Full Potential of Social Safety Nets in Africa, Washington, DC: World Bank*, 87–138.

Ballard, T. J., Kepple, A. W., & Cafiero, C. (2013). The food insecurity experience scale: development of a global standard for monitoring hunger worldwide. *Rome: FAO*, *61*.

Baye, K., & Hirvonen, K. (2020). *Accelerating progress in improving diets and nutrition in Ethiopia* (Vol. 144). Intl Food Policy Res Inst.

Beegle, K., Coudouel, A., & Monsalve, E. (2018). *Realizing the full potential of social safety nets in Africa*. World Bank Publications.

Berhane, G., Golan, J., Hirvonen, K., Hoddinott, J. F., Kim, S. S., Taffesse, A. S., Abay, K., Assefa, T. W., Habte, Y., & Abay, M. H. (2020). *Evaluation of the nutrition-sensitive features of the fourth phase of Ethiopia's Productive Safety Net Programme* (Vol. 140). Intl Food Policy Res Inst.

Cafiero, C., Viviani, S., & Nord, M. (2018). Food security measurement in a global context: The food insecurity experience scale. *Measurement: Journal of the International Measurement Confederation*, *116*. https://doi.org/10.1016/j.measurement.2017.10.065

Central Statistical Agency of Ethiopia. (2022). *Country and regional level consumer price indices (CPI)*.

CSA, NBE, & T. W. B. (2017). *LSMS - Integrated Surveys on Agriculture: Ethiopia Socioeconomic Survey (ESS) 2015/16. A report by the Central Statistic Agency (CSA) of Ethiopia in collaboration with the National Bank of Ethiopia (NBE) and the World Bank*.

Davis, T. P., Wetzel, C., Avilan, E. H., de Mendoza Lopes, C., Chase, R. P., Winch, P. J., & Perry, H. B. (2013). Reducing child global undernutrition at scale in Sofala Province, Mozambique, using Care Group Volunteers to communicate health messages to mothers. *Global Health: Science and Practice*, *1*(1), 35–51.

De Weerdt, J., Gibson, J., & Beegle, K. (2020). What can we learn from experimenting with survey methods? *Annual Review of Resource Economics*, *12*, 431–447.

Deaton, A., & Zaidi, S. (2002). *Guidelines for constructing consumption aggregates for welfare analysis* (Vol. 135). World Bank Publications.

Fiszbein, A., Schady, N., Ferreira, F. H. G., Grosh, M., Keleher, N., Olinto, P., & Skoufias, E. (2009). *Conditional Cash Transfers: Reducing Present and Future Poverty.*

Food Security Information Network. (2022). *2022 Global Report on Food Crises*.

Gaddis, I. (2016). Prices for poverty analysis in Africa. *World Bank Policy Research Working Paper*, *7652*.

Gibson, J., & Rozelle, S. (2003). Poverty and access to roads in Papua New Guinea. *Economic Development and Cultural Change*, *52*(1), 159–185.

Golan, J., Headey, D., Hirvonen, K., & Hoddinott, J. (2019). *Changes in child under-nutrition in Ethiopia, 2000–16*.

Headey, D., Hirvonen, K., Hoddinott, J., & Stifel, D. (2019). *Rural food markets and child nutrition*. Wiley Online Library.

Hickey, S., Lavers, T., Niño-Zarazúa, M., & Seekings, J. (2018). *The negotiated politics of social protection in sub-Saharan Africa*.

Hidrobo, M., Hoddinott, J., Kumar, N., & Olivier, M. (2018). Social protection, food security, and asset formation. *World Development*, *101*, 88–103.

Hirvonen, K., Wolle, A., Laillou, A., Vinci, V., Chitekwe, S., & Baye, K. (2021). Understanding delays in the introduction of complementary foods in rural Ethiopia. *Maternal & Child Nutrition*, e13247.

IMF. (2022). *Country Indexes and Weight*. https://data.imf.org/regular.aspx?key=61015892.

Karra, M., Subramanian, S. V, & Fink, G. (2017). Height in healthy children in low-and middle-income countries: an assessment. *The American Journal of Clinical Nutrition*, *105*(1), 121–126.

Kraay, A., & McKenzie, D. (2014). Do poverty traps exist? Assessing the evidence. *Journal of Economic Perspectives*, *28*(3), 127–148.

Le, K., & Nguyen, M. (2020). Shedding light on maternal education and child health in developing countries. *World Development*, *133*, 105005. https://doi.org/https://doi.org/10.1016/j.worlddev.2020.105005

Leight, J., Alderman, H., Gilligan, D. O., Hidrobo, M., Mulford, M., & Tadesse, E. (2022). *Barriers to adoption of optimal complementary feeding practices in Ethiopia: A formative qualitative investigation: Evidence from SPIR II*.

Li, Z., Kim, R., Vollmer, S., & Subramanian, S. V. (2020). Factors Associated With Child Stunting, Wasting, and Underweight in 35 Low- and Middle-Income Countries. *JAMA Network Open*, *3*(4), e203386–e203386. https://doi.org/10.1001/jamanetworkopen.2020.3386

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Lovon, M., & Mathiassen, A. (2014). Are the World Food Programme's food consumption groups a good proxy for energy deficiency? *Food Security*, *6*(4). https://doi.org/10.1007/s12571-014-0367-z

Maxwell, D., Coates, J., & Vaitla, B. (2013). How do different indicators of household food security compare? Empirical evidence from Tigray. *Feinstein International Center*, 1–19.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, *99*(2), 210–221. https://doi.org/https://doi.org/10.1016/j.jdeveco.2012.01.002

Medhin, G., Hanlon, C., Dewey, M., Alem, A., Tesfaye, F., Worku, B., Tomlinson, M., & Prince, M. (2010). Prevalence and predictors of undernutrition among infants aged six and twelve months in Butajira, Ethiopia: the P-MaMiE Birth Cohort. *BMC Public Health*, *10*(1), 1–15.

Ruel, M. T., & Alderman, H. (2013). Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? *The Lancet*, *382*(9891), 536–551. https://doi.org/https://doi.org/10.1016/S0140-6736(13)60843-0

Sabates-Wheeler, R., Lind, J., Hoddinott, J., & Tefera Taye, M. (2021). Graduation after 10 years of Ethiopia's Productive Safety Net Programme: Surviving but still not thriving. *Development Policy Review*, *39*(4), 511–531.

Swindale, A., & Bilinsky, P. (2006). Development of a universally applicable household food insecurity measurement tool: process, current status, and outstanding issues. *The Journal of Nutrition*, *136*(5), 1449S-1452S.

Tekile, A. K., Woya, A. A., & Basha, G. W. (2019). Prevalence of malnutrition and associated factors among under-five children in Ethiopia: evidence from the 2016 Ethiopia Demographic and Health Survey. *BMC Research Notes*, *12*(1), 391. https://doi.org/10.1186/s13104-019-4444-4

USAID. (2021). *Bureau for Humanitarian Assistance Indicator Handbook. Supplement to Part I: BHA Baseline/Endline Questionnaire and Indicator Tabulations for Resilience Food Security Activities. Revised version, July 2021.*

Victora, C. G., De Onis, M., Hallal, P. C., Blössner, M., & Shrimpton, R. (2010). Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics*, *125*(3), e473–e480.

WHO and UNICEF. (2021). *Indicators for assessing infant and young child feeding practices: definitions and measurement methods*.

APPENDIX A: HOUSEHOLD CONSUMPTION AND POVERTY MEASUREMENT

Introduction

The baseline survey collected detailed information on household food and non-food consumption and expenditures. These questionnaire modules were designed based on the consumption modules used in the phase-1 SPIR evaluation surveys and the Ethiopia Living Standard Measurement Study (LSMS) surveys (CSA, NBE, 2017).

The consumption data are used to compute daily per capita consumption expenditures following the approach outlined in previous work (Deaton & Zaidi, 2002). The household consumption-expenditure estimates are then used to calculate poverty headcount and other poverty indicators.

Consumption Components

The baseline survey questionnaire asked households to report on their consumption expenditures over a specific time interval. The consumption module had four sub-modules: Household expenditures on durables and services over the past 12 months; Household expenditures on consumables over the past 1 month; Household food consumption over the past 7 days; Household expenditures on food consumed away from home in the past 7 days. Below we provide more details about each consumption component.

1. Household Expenditures on Durables and Services over the Past 12 Months

Part 1 of Module H in the questionnaire asked households to report on their expenditures on 25 non-food items and services. These expenditure items are assumed to be consumed infrequently and therefore the recall period was set to 12 months.

The reported values for each expenditure item were screened for outliers using expenditure per capita to allow variation across the household size. The reported per capita expenditures that were above 3 standard deviations among those who consumed the item were marked as outliers. Only 0.6% of the reported values were marked as outliers. These outlier values were winsorized to the 90^{th} percentile of the item-specific per capita expenditure distribution among those who reported consuming the item. The reported value was missing in 4 out of 70,000 cases because the respondent could not remember the amount spent or was not able to answer the question. In these cases, we used the median item-specific per capita expenditure value calculated from the full distribution that also included zero expenditures.

2. Household Expenditures on Consumables over the Past 1 Month

Part 2 of Module H in the questionnaire asked households to report on their expenditures on 14 non-food items. Households are expected to consume these items frequently and therefore the recall period was set to 1 month.

As before, the reported values for each expenditure item were screened for outliers using expenditure per capita to allow variation across the household size. The reported per capita expenditures that were above 3 standard deviations among those who consumed the item were marked as outliers. Less than 1% of the reported values were marked as outliers. These outlier values were winsorized to the 90th percentile of the item-specific per capita expenditure distribution among those who reported consuming the item. The reported value was missing in 4 out of about 40,000 cases because the respondent could not remember the amount spent or was not able to answer the question. In these cases, we used the median item-specific per capita expenditure value calculated from the full distribution that also included zero expenditures.

3. Household Food Consumption over the Past 7 Days

The household food consumption module in Part 3 of Module H included 62 food items commonly used in the areas in which the survey took place. The enumerators first went through the list of 62 items asking whether the household consumed the item in the past 7 days or not. The survey instrument was programmed to carry forward all food items that the household reported to have consumed in the past 7 days to the next section that asked about the consumption frequency ('on how many days was the item consumed') and quantity ('amount consumed') within the 7 days.

Households were allowed to report the consumed amounts in any unit. In the analysis stage, we converted all amounts to kilograms using the conversion factors provided by the World Bank or conversion factors gathered in previous IFPRI surveys. In very few cases (less than 0.5%), we could not find a conversion factor for the reported unit. If the price data was mostly reported in the same unit, we valued the consumed amount using the price data measured in the same unit. In the remaining cases, we used the median reported item-specific consumption value among those who reported having consumed the item. All consumed amounts were then valued in birr terms using the food price data collected as a part of the survey (for more details, see below).

The final food consumption aggregate was formed of 59 food items. None of the households in the sample reported consuming enset or pork. In addition, we were not able to get a price estimate for margarine. Since only two households had reported having consumed margarine, this item was also omitted from the final food consumption aggregate.

As before, the reported values for each food item were screened for outliers using consumption-expenditure per capita to allow variation across the household size. The reported per capita consumption expenditures that were above 3 standard deviations among those who consumed the item were marked as outliers. Less than 0.5% of the reported values were marked as outliers. These outlier values were winsorized to the 90th percentile of the item-specific per capita expenditure distribution among those who reported consuming the food item.

4. Household Expenditures on Food Consumed Away from Home in the Past 7 Days

Finally, Part 3 of Module H in the questionnaire also asked households whether they purchased any prepared food or eaten outside in the last 7 days. About 7% of the households responded Yes to this

question. These households were then asked to report the total amount in birr they spent on food away from home in the past 7 days.

We screened for outliers by looking at the per capita expenditures among those who did consume food away from home. Using the same approach as described above, only four households were classified as outliers. However, a visual inspection of the data did not reveal anything unusual and therefore, no adjustments were made.

Price Data

The price data were collected using a price opinion survey (Part 5 of Module H in the questionnaire). While this approach is not common in consumption or poverty measurement in Africa (Gaddis, 2016), a survey experiment conducted in Papua New Guinea suggests that this method yields accurate price data (Gibson & Rozelle, 2003). The standard approach is to send enumerators to the local markets to collect price data. However, a recurring problem with such market visits in Ethiopia is that often the food items are not available in the local food market. Particularly meat and milk prices are frequently missing from the market survey data because they are not on sale in rural periodic markets (Headey et al., 2019). Therefore, asking for price opinions directly from households is likely to yield fewer missing observations in a given geographical area than market surveys.

The survey instrument included a price module in which households were asked to estimate food prices in their locality. The list of food items in the price module matched the food items in the food consumption modules except for pork which was omitted from the price module. However, households were not asked to estimate the price of all food items. Instead, the food price module was divided into six segments with each segment containing 10-11 food items. In each household interview, only one segment was randomly selected to be included in the questionnaire. As a result, each household was asked to estimate the price of 10 or 11 items.

Households were allowed to report the price in any unit. We used the same approach as described above to convert the non-standard units to kilograms.

We then aggregated these household-level price estimates at the woreda level. In total, we obtained more than 16,000 price estimates across 16 woredas. Considering the 61 items in the price module, this translates into more than 250 price estimates per item.

To reduce the influence of outliers, we then took the median estimate for each item to represent the price in the woreda. For a small number of items that were not commonly consumed in the locality and as a result respondents were not able to give a price estimate. In these cases, we used a region-specific or sample-specific median price.

Consumption Aggregate

The consumption aggregate was formed of the four consumption components listed above. All consumption amounts were transformed into daily terms by dividing the reported amount by the number of days in the recall period.

Daily household consumption expenditures were reported on a per capita basis by dividing the consumption expenditure by household size. No attempts were made to adjust for household age or sex-specific demographics (e.g., by using an adult equivalence scale).

Deflation

To estimate the prevalence and depth of poverty based on the international poverty line, we needed to convert the household per capita consumption expenditures measured in birr terms to 2011 Purchasing Power Parity (PPP) dollars. To do so, we followed the BHA guidelines (USAID, 2021) (USAID, 2022) as described below.

We first converted the birr values measured at the time of the survey (August-September 2022) to birr values in 2011 prices. For the average household in our sample food expenditures make 85% of the total consumption basket. Therefore, we used food consumer price index (CPI) estimates to deflate the estimates to 2011 prices. The CPI estimates come from the International Monetary Fund (IMF) which reports monthly estimates for Ethiopia (IMF, 2022). At the time of writing, the latest IMF CPI estimate for Ethiopia was for March 2022. To extend the series to August-September 2022, we used the CPI estimates reported by the Ethiopia Central Statistical Agency (Central Statistical Agency of Ethiopia, 2022).[17] We could not access region-specific CPI estimates for the whole period of 2011-2022 and therefore, we deflated our consumption expenditure data using national-level food CPI estimates.[18]

The food CPI value for August 2022 was 352.3 and for September 2022 it was 359.6. Meanwhile, the mean food CPI for the 12 months in 2011 was 58.29.[19] Therefore, we divided the estimated consumption-expenditure amounts by the ratio of the CPI for the survey month: 6.04 if the household was interviewed in August and 6.17 if the household was interviewed in September or October.[20]

We then converted the consumption-expenditure values expressed in 2011 prices to PPP terms using the PPP conversion rate for individual consumption expenditure by households, obtained from the International Comparison Program (ICP). The 2011 estimate for Ethiopia is 5.44.

Finally, as requested by the BHA, we also express the consumption-expenditure values in 2010 constant US dollar terms by dividing the PPP consumption-expenditure values by 1.032, which is the ratio of the US CPI in 2011 to the US CPI in 2010.

[17] The CSA changed the methodology of their CPI estimation in December 2016. Perhaps as a result, we were not able to get access to consistent CPI series from the CSA sources that cover the whole period between 2011 and 2022. Therefore, we opted for using the IMF series because they cover the whole period, and the CPI estimates made comparable across time.
[18] In August 2022, the national food CPI estimate was 352.3. The corresponding estimate for Amhara region was 348.6 and for Oromia region was 357.2, indicating that the food inflation in these two most populous regions of Ethiopia closely tracks the national food inflation rate.
[19] More specifically, the 2011 mean food CPI calculated from the IMF series is 58.29 and the March 2022 value in the series is 299.00. The March 2022 value in the CSA series is also 299.00 and the August 2022 value is 352.3 and September 2022 value is 359.6.
[20] Approximately 63 percent of the households were interviewed in August, 29% in September and 8% in October. Because the CPI estimate for October was not yet available at the time of writing, we used the September CPI estimate for households interviewed in October.

Poverty Estimation

The poverty headcount ratio is defined as the share of the population in the sample with daily per capita consumption expenditures less than US$1.90 at 2011 prices.[21] We also compute the poverty gap index, which measures the depth of poverty as the mean income of the poor as a percentage of the US$1.90 poverty line.

Comparability

Our consumption and poverty estimates are not comparable to the official estimates from the Household Consumption Expenditure Survey (HCES). First, our sample is not representative of the regions in which the survey took place or even the PSNP households residing in the two regions. Our sampling frame is formed of PSNP households with a pregnant woman or a child under 9 months of age residing in the SPIR intervention woredas in the Amhara and Oromia regions. In contrast, the sample in the HCES is designed to be representative at the national and regional levels. Second, the survey was conducted at the end of the meher rainy season in August-October 2022. The last HCES survey was conducted over a 12-month period in 2015/16 to address consumption seasonality. Finally, the survey questionnaires are vastly different. The HCES covers nearly 400 food items and more than 850 non-food items. Our consumption modules are substantially shorter. Literature on survey experiments conducted on consumption measurements shows that changes in survey designs can result in sizable variation in poverty headcount estimates, of up to 20 percentage points (De Weerdt et al., 2020). Therefore, for an impact evaluation, it is important to make sure that the survey methodology remains consistent across baseline and follow-up rounds.

[21] Note that since the poverty line is defined in 2011 prices, we use the consumption values reported in 2011 $PPP (and not in 2010 USD).

APPENDIX B: ADDITIONAL TABLES

Table 37: Characteristics of previous and new PSNP beneficiaries

	Previous PSNP Beneficiary	New PSNP Beneficiary	Difference	p-value
Household size	5.757	5.229	-0.528	0
Male headed household	0.849	0.904	0.055	0
Household head has some formal education	0.491	0.537	0.046	0.018
Primary female has some formal education	0.38	0.456	0.076	0
Food gap in months	3.521	3.518	-0.003	0.975
Food Consumption Score	43.808	45.196	1.388	0.019
iBL06 - FIES status based on simple sum	3.424	3.441	0.017	0.586
Tropical Livestock Units owned by the household	12.651	11.415	-1.236	0.013
iBL01 - Prevalence of Poverty: % of people living on less than 1.90/day 2011 PPP	0.735	0.623	-0.112	0

Table 38: Summary table of all required indicators

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL06. Prevalence of moderate food insecurity in the household based on the Food Insecurity Experience Scale (FIES)	31.38	29.22	33.54	3,015
BL06. Prevalence of severe food insecurity in the household based on the Food Insecurity Experience Scale (FIES)	57.41	54.61	60.22	3,015
BL10. Percentage of households with poor food consumption score (FCS)	5.47	4.24	6.70	3,015
BL10. Percentage of households with borderline food consumption score (FCS)	19.30	17.27	21.34	3,015
BL10. Percentage of households with acceptable food consumption score (FCS)	75.22	72.71	77.74	3,015
BL16. Percentage of households using basic drinking	20.48	18.12	22.84	3,013

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
water services based on 4 out of 5 criteria[1]				
BL17. Percentage of households with soap and water at a handwashing station on premises	1.76	1.22	2.29	3,015
BL18. Percentage of households practicing correct use of recommended household water treatment technologies	11.00	9.40	12.60	3,009
BL19. Percentage of households practicing open defecation	50.35	46.89	53.81	3,015
BL27. Percentage of households with access to a basic sanitation service	17.08	15.15	19.01	3,015
BL12. Percentage of children 6-23 months receiving a minimum acceptable diet *	3.10	1.54	4.67	741
BL13. Prevalence of exclusive breastfeeding of children under 6 months *	79.20	76.30	82.09	870
BL14. Percentage of children under five (0-59 months) who had diarrhea in the prior 2 weeks *	12.11	10.40	13.82	1,627
BL15. Percentage of children under five (0-59months) with diarrhea treated with Oral Rehydration Therapy *	63.83	57.09	70.57	188
BL39. Prevalence of children 6-23 months consuming a diet of minimum diversity (MDD-C) *	3.10	1.54	4.66	743
BL11. Prevalence of women of reproductive age consuming a diet of minimum diversity (MDD-W)	6.87	5.54	8.21	3,011
BL26. Percentage of births receiving at least four antenatal care visits during pregnancy [1]	41.75	38.64	44.86	1,406
BL36. Percentage of women in a union who have knowledge of modern family planning methods	89.91	88.78	91.04	2,736
BL37. Percentage of women in a union who made decisions to use modern family planning methods in the past 12 months	73.76	69.99	77.54	526
BL29. Percentage of farmers who used financial services (savings, agricultural credit, and/or insurance) in the past 12 months	52.86	49.86	55.86	2,902
BL40. Daily per capita expenditures (as a proxy for income) in USG-assisted areas, in 2010 USD	1.72	1.67	1.77	3,015

	Indicator Mean (%)	95% Confidence Interval		N
		CI Lower	CI Upper	
BL01. Prevalence of poverty: Percentage of people living on less than $1.90/day 2011 PPP	70.96	68.75	73.18	16,264
BL02. Depth of poverty of the poor: Mean percentage shortfall of the poor relative to the $1.90/day 2011 PPP poverty line	34.21	33.00	35.41	11,541
BL32. Percentage of women and men in a union who earned cash in the past 12 months	53.91	51.50	56.32	4,823
BL33. Percentage of women in a union and earning cash who report participation in decisions about the use of self-earned cash	68.36	62.02	74.71	844
BL34. Percentage of women in a union and earning cash who report participation in decisions about the use of spouse/partner's self-earned cash	60.62	55.21	66.03	843
BL35. Percentage of men in a union and earning cash who report spouse/partner participation in decisions about the use of self-earned cash	79.67	77.33	82.01	1,756
BL41. Percentage of women in a union who are members of a community group	70.92	67.39	74.45	2,813
BL41. Percentage of men in a union who are members of a community group	79.89	77.11	82.67	2,009
BL42. Percentage of women in a union with access to credit	52.45	49.99	54.92	2,814
BL42. Percentage of men in a union with access to credit	50.92	48.13	53.71	2,009
BL43. Percentage of women in a union who make decisions about credit	64.88	60.89	68.87	1,475
BL43. Percentage of men in a union who make decisions about credit	70.77	66.24	75.30	1,023

** Due to the sampling strategy, about 90 % the children in this sample are less than 10 months of age.*

APPENDIX C: PRE-ANALYSIS PLAN

IMPEL/SPIR II RFSA Evaluation

Pre-analysis plan

International Food Policy Research Institute (IFPRI)

Date: July 25, 2022

1. Introduction

Over the past two decades, social safety net programs have become a mainstream policy tool across sub-Saharan Africa to address food insecurity and extreme poverty (Beegle, Coudouel, & Monsalve, 2018). Since the turn of the millennium, the number of social safety net programs has doubled (Hickey, Lavers, Niño-Zarazúa, & Seekings, 2018) and today, each country in the region operates at least one major safety net program (Beegle et al., 2018).

There is now strong evidence from a wide variety of contexts showing that these programs can be effectively used to improve food security and increase asset accumulation (Andrews, Hsiao, & Ralston, 2018; Hidrobo, Hoddinott, Kumar, & Olivier, 2018). Building on this evidence, there is a growing interest in using social safety net programs as a platform to achieve broader objectives over and above food security, including the reduction of poverty and the enhancement of resilience. One such objective relates to graduation: moving household away from long-term support and enabling them to build resilient and self-reliant livelihoods (Sabates-Wheeler, Lind, Hoddinott, & Tefera Taye, 2021). A second objective centers around making existing programs nutrition-sensitive by integrating transfers with nutrition and WASH trainings and other interventions to better address the underlying causes of maternal and child malnutrition (Ruel & Alderman, 2013).

Against this backdrop, the second phase of the Strengthen PSNP4 Institutions and Resilience (SPIR) RFSA aims to enhance livelihoods, increase resilience to shocks, and improve food security and nutrition for rural households vulnerable to food insecurity in Ethiopia. The RFSA is situated within Ethiopia's PSNP program, one of the largest safety net programs in Africa. Funded by USAID's Bureau for Humanitarian Assistance (BHA), SPIR II is implemented by World Vision International (lead), CARE, and ORDA in Amhara and Oromia regions of Ethiopia. The IMPEL SPIR II impact evaluation proposes to employ an experimental design with multiple treatment arms comparing combinations of livelihood and nutrition graduation model programming provided to PSNP beneficiaries relative to a control group receiving only PSNP transfers. The objective of this document is to provide a pre-analysis plan for the impact evaluation.

2. Context and the SPIR II Program

2.1 The Productive Safety Net Program (PSNP)

Launched in February 2005, the PSNP was designed as a more sustainable response mechanism than the recurring *ad hoc* humanitarian appeals that occurred in Ethiopia throughout the 1990s and early 2000s

(De Waal, 2017; Wiseman, Van Domelen, & Coll-Black, 2010). With more than eight million beneficiaries, the PSNP is one of the largest safety net programs in Africa. Beneficiary households receive food or cash payments in return for labor-intensive public works. Households with limited labor capacity (e.g., elderly, disabled, pregnant, and lactating women) receive unconditional transfers (i.e., direct support).

The PSNP has been successful in improving household food security and resilience in the areas in which it operates (Berhane, Gilligan, Hoddinott, Kumar, & Taffesse, 2014; Knippenberg & Hoddinott, 2017). However, despite multiple initiatives, the program has been less successful in promoting sustainable graduation out of long-term social assistance (Sabates-Wheeler et al., 2021). Similarly, despite the recent efforts to make the PSNP more nutrition sensitive, the impacts of the program on children's diets, complementary feeding practices, and anthropometry outcomes including stunting and wasting have remained limited (Berhane et al., 2020;).

2.2 The Strengthen PSNP Institutions and Resilience II (SPIR II) RFSA

Core SPIR II graduation programming includes the organization of village economic and social associations (VESAs), used as a platform for trainings and other RFSA activities around financial literacy, promotion of savings and credit use, agriculture, and livestock value chain development (e.g., developing business skills and production skills), improving social capital, and catalyzing women's empowerment. A subset of households (<20%) is also targeted for one-time $300 livelihoods grants. In addition, nutrition programming centers around the provision of included integrated nutrition social behavior change communication (SBCC) as well as WASH activities. The first major goal of this evaluation is to assess the effectiveness of integrated SPIR II graduation model programming—the bundle of interventions described above—on a range of outcomes at the household level, including livelihoods-related outcomes and graduation from poverty.

In addition, a particular focus of this evaluation is innovations in nutrition-related programming centered around enhanced infant and young child feeding (IYCF) practices, particularly suboptimal complementary feeding practices that have been widely speculated to be inhibiting child growth and development in Ethiopia (Golan, Headey, Hirvonen, & Hoddinott, 2019). The Nurturing Care Group (NCG) model, pioneered by World Vision in a wide range of other contexts, is based on groups of 10–15 community-based trained volunteer agents who cascade down SBCC messages and activities to caregiver groups at the community level. Non-experimental studies conducted in other contexts suggest that the model can significantly increase SBCC contact rates and improve IYCF practices and child growth outcomes (Davis et al., 2013). However, large-scale experimental evidence on the relative effectiveness of this strategy relative to standard government-led nutrition programming is largely unavailable, rendering this evaluation a meaningful contribution. (Standard government-led nutrition programming in Ethiopia is delivered primarily by health extension workers and the health development army through relatively infrequent interactions to provide nutritional information and counselling to pregnant women and mothers of young children.)

In addition, improving caregiver knowledge may not be sufficient to improve complementary feeding practices if households cannot afford to purchase nutritious foods. Therefore, the third study arm introduces maternal grants of $20/month during the child's first 24 months of life to relax possible financial constraints to child feeding.

3. Evaluation Methods

3.1 Treatment Arms

The proposed evaluation is a cluster randomized control trial (cRCT) where the cluster is defined as the *kebele* (lowest administrative level in Ethiopia). The sample will include 237 kebeles in Amhara and Oromia in which SPIR II is operational. The target sample comprises kebeles that were not included in the previous impact evaluations conducted by IFPRI and World Vision under SPIR: this IMPEL sample includes kebeles that were served by SPIR but were not included in the SPIR impact evaluation (generally because programming had already been initiated); and new kebeles. In addition, a small number of kebeles have been excluded from the study due to ongoing insecurity.

The impact evaluation will have three study arms:

- T1. PSNP implemented by SPIR II; no supplemental programming (79 kebeles)[22]
- T2. PSNP + SPIR II + NCG model (79 kebeles)
- T3. PSNP + SPIR II + NCG + maternal grants (79 kebeles)

The study arm T1 serves as a control group of PSNP households against which the impacts of SPIR II programming will be measured. PSNP households in arm T2 benefit from SPIR II resilience programming and will be exposed to the NCG intervention. The study arm T3 receive the same intervention package as households in T2 but also benefit from the maternal grants.

This multi-arm cRCT design permits the research team to evaluate the causal impact of both livelihood and nutrition graduation programming in SPIR II. First, comparing outcomes in T1 to T2 and T3 permits us to quantify the causal impact of the SPIR II livelihood graduation programming on outcomes such as financial inclusion, assets, consumption, resilience, and poverty. Second, by experimentally varying the nutrition interventions, the study will provide valuable information on nutrition sensitive programming within the PSNP. More specifically, by comparing outcomes across all three treatment arms, we can assess the relative effectiveness of the NCG intervention on IYCF practices and child growth outcomes (e.g., child stunting prevalence) with and without an added maternal grant to reduce cost constraints to improving these outcomes.

3.2 Sampling

The target sample for this evaluation includes an estimated 3,081 households in the sample, drawn from 237 kebeles in 15 woredas. This includes an estimated 1,027 households in each treatment arm. The inclusion criteria for households to be included in the evaluation are the following.

i) The household must be enrolled as a PSNP beneficiary in a target kebele.
ii) The household must meet one of the following characteristics:
    a) There is a pregnant woman present who self-reports pregnancy, with an estimated gestational age that is at least 3 months (i.e., following the first trimester).

[22] World Vision is implementing SPIR II on behalf of local governments. The caseload is provided to World Vision by the local governments; World Vision and partners provide the food transfers, while government partners provide the cash transfers.

b) There is an infant present aged less than 9 months as of the date of survey; and the infant's mother or primary caretaker is also resident in the household.

Households will be sampled following a listing exercise in each kebele; households meeting the eligibility criteria will be identified and a random subsample will be selected. Having reviewed updated data on PSNP beneficiary numbers and demographic trends in these regions, we project that in each kebele, we will identify around 40 households who will meet the above criteria. Our target for sampling is 13 households per kebele: 7 households including a pregnant woman and six households including a child under one. The sampling targets were informed by available resources for data collection as well as the goal of achieving adequate statistical power for the key comparisons of interest (more details about power calculations are provided below). If we are unable to reach the target sample of one of the two subgroups (pregnant women, or households characterized by a child under one), we will substitute a household from the other subgroup, if available. The target kebeles report an average of 261 PSNP beneficiary households based on updated sample lists; accordingly, the evaluation will target around 5% of beneficiary households.

Using the current beneficiary lists in the target samples and demographic estimates of the percentage of households that will have a pregnant woman or a child under 9 months, we estimate that in around 10% of kebeles (30 kebeles), it may be challenging to enroll 13 households into the sample, as the number of PSNP beneficiary households is lower and we may not identify 13 such households meeting the demographic criteria. In these 10% of kebeles, the average number of (projected) eligible households is 11; thus we estimate that the potential shortfall of respondents may be around 60 households, if projections are accurate. This is an estimate only, as demographic trends may vary; but we note in the power calculations below that this shortfall, if observed, would have minimal effects on the ability of the evaluation to detect experimental effects of interest.[23]

Randomization

Randomization will be conducted by the research team using Stata, in June 2022. The randomization process will proceed as follows. We will construct strata based on the interaction of the following characteristics: woreda; a binary variable for whether a kebele is above or below the woreda-level median in the percentage of households eligible for the PSNP; and a binary variable for whether the kebele is above or below the woreda-level median in distance from the woreda capital).

Following the baseline survey, we will also evaluate balance across treatment arms for a broad set of household covariates reported in the baseline survey, including baseline levels of all primary outcomes of interest. These balance checks will be reported in the baseline report.

3.3 Surveys

The evaluation will include three primary rounds of data collection.

i) Randomization of kebeles will be conducted in June 2022.

[23] The sampling targets reflected in field documents and manuals are unchanged, as this reflects the mandate provided to the survey firm and they will enroll eligible households up to the 3,081 target if such households can be identified.

ii) The baseline survey is scheduled for July–August 2022. All households will be visited and surveys will be conducted with both the primary female respondent (pregnant or lactating woman) and her spouse.
iii) The rollout of the NCGs and other SPIR II interventions will immediately follow the baseline survey, in September 2022.
iv) The midline survey is scheduled for July–August 2023, around twelve months following the baseline. All households will be visited and surveys will be conducted with the **primary female respondent** only. Child anthropometric measurements will also be collected.
v) The endline survey is scheduled for July–August 2025, around twenty-four months following the midline. All households will be visited and surveys will be conducted with both the primary female respondent (pregnant or lactating woman) and her spouse.

The timing of the survey rounds plays a critical role in the evaluation of the nutrition interventions. The baseline survey in July/August 2022 will sample PSNP households with pregnant woman or a child less than 9 months of age (Figure 16). At the same time, we will randomly allocate kebeles into the three study arms.

We will administer a midline quantitative survey of the entire baseline sample 1 year after the baseline survey, in July/August 2023, when the children are between 6–23 months of age, permitting us to assess both contact rates and participation in nutrition care groups (process indicators) and acquired knowledge about and adherence to recommended IYCF practices (outcome indicators) (WHO & UNICEF, 2021). The timing of the midline survey has several advantages, including (i) allowing timely measurement of the impact of SPIR II on child diets and IYCF practices, (ii) inclusion of process monitoring questions for the full sample around household participation in SPIR II activities and access to SPIR II components, (iii) measurement of intermediate outcomes related to food security to examine progress against RFSA objectives.

The endline survey is planned for 2025, when the children are 30–47 months. This survey will be the primary round in which we can measure effects on livelihoods outcomes (specified in more detail below) and is also an ideal time to assess impacts on child growth outcomes, also specified in more detail below (Alderman & Headey, 2018). Child growth faltering (measured using child height-for-age Z-scores) in Ethiopia and other low-income country settings largely occurs during the first 1000 days of life (Golan et al., 2019; Victora, de Onis, Hallal, Blossner, & Shrimpton, 2010). As shown in Figure 17 based on the data from the 2015/16 Ethiopia Demographic and Health Survey (DHS), during the first 4 to 5 months of life, the height of the average Ethiopia child is similar to the height of the median child in the WHO-2006 growth reference.[24] Golan et al., 2019 hypothesize that child growth during this period is supported by the relatively high adherence to exclusive breastfeeding in Ethiopia. Rapid growth faltering begins at around 6 months of age when children should be introduced to complementary foods and continues until about 18 months of age. The endline occurs after this period of rapid growth faltering. By now, children and caregivers in study arms T2 and T3 have been exposed to the intensive SPIR II nutrition programming throughout the critical first 1000-day period. The study hypothesis is that this nutrition programming prevents growth faltering during the first 1000 days and as a result, at the endline the HAZ

[24] HAZ measures the height difference to the median child in the WHO-2006 growth reference sample. This difference is measured in terms of standard deviations. Thus, the HAZ of the median child in the growth reference is 0. In Figure 17, this is marked with the dashed horizontal line (HAZ=0). Child is defined as stunted if her HAZ<-2 and severely stunted if HAZ<-3.

curve of the average child in the treatment arms lie above the HAZ curve of the average child in the control arm.

Figure 16: Timing of the surveys

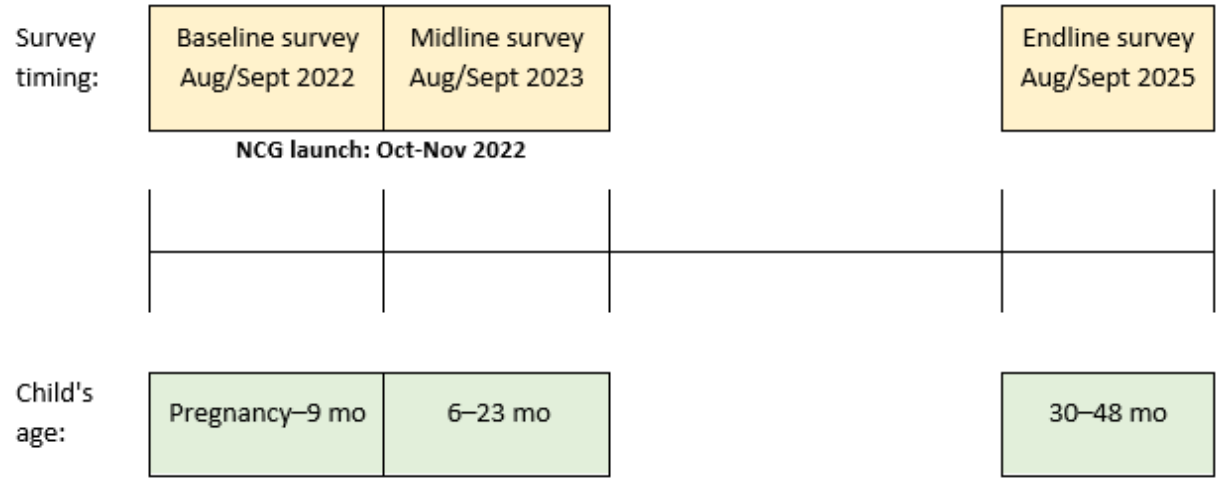
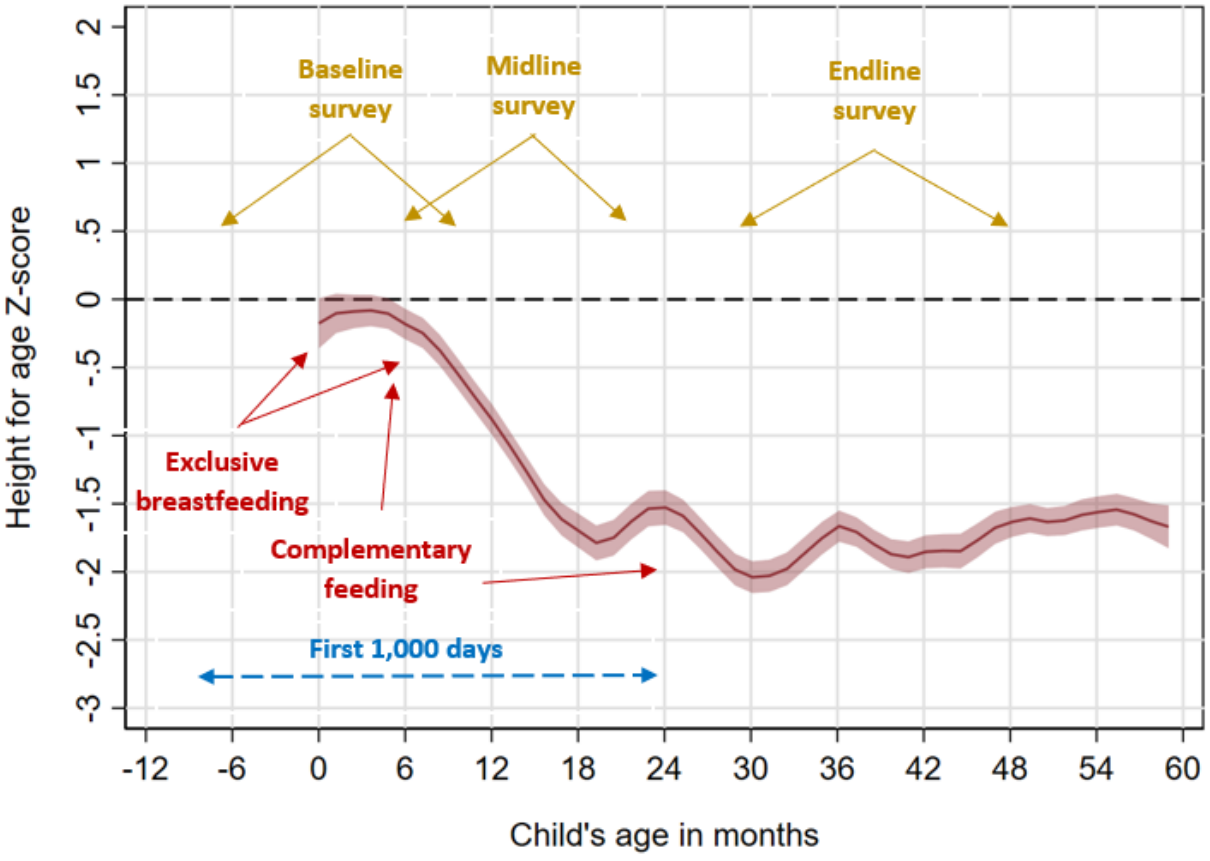


Figure 17: Timing of the surveys in relation to typical linear growth faltering in Ethiopia



Note: Local polynomial regression based on Ethiopia 2015/16 Demographic and Health Survey (DHS). The shaded areas represent 95-% confidence intervals. N = 8,771 children 0-59 months of age.

Table 39 below summarizes the key questionnaire modules to be included in each wave of data collection. The planned modules for midline and endline, however, may require revision based on resource constraints.

Table 39: Questionnaire modules, by survey round

(Note: The table notes the target respondent. Enumerators can substitute another knowledgeable individual if the target respondent is not available or declines to answer the section. The modules are presented in the order that they will be administered in the baseline survey.)

Module	Respondent	Baseline	Midline	Endline
Module A. Household Identification and Consent	Household head (primary female at midline)	X	X	X
Module B. Household Roster		X	X	X
Module C1. Access to PSNP, HFA, VESAs, & Savings Institutions				
Part 1: Past and Current Access to PSNP		X		X
Part 2: PSNP Payments since January		X		X
Part 3: Livelihoods Component		X		X
Part 4: Other Public Transfers (HFA)		X		X
Part 5: Participation in VESAs and SPIR II Activities		X		X
Part 6: Savings and Access to Savings Institutions		X		X
Part 7: Access to Health Insurance		X		X
Module C2. Access to Nurturing Care Groups and Maternal Grants	Primary female		X	X
Module D. Paternal IYCF Knowledge and Perceptions	Primary male	X		X
Module E. Agriculture	Primary male	X		X
Module F. Household Assets	Primary male			
Part 1: Productive Assets		X		X
Part 2: Consumer Durables		X		X
Part 3: Livestock Ownership		X		X
Module G. Gender (Cash)	Primary male	X		X
Module H. Gender Access to Credit and Group Participation	Primary male	X		X

Module	Respondent	Baseline	Midline	Endline
Module I. Poverty Measurement (including FCS and FIES)				
Part 1: Durables and Services (annual)	Most knowledgeable	X		X
Part 2: Household Non-Food Consumables (monthly)	Most knowledgeable	X		X
Part 3: Food Consumption and Expenditure	Most knowledgeable	X		X
Part 4: Food Security in the last 12 Months	Primary female	X	X	X
Part 5: Food Prices in the Locality	Most knowledgeable	X		X
Module J. Water, Sanitation and Hygiene (WASH)	Primary female	X		X
Module K. Children's Nutritional Status and Feeding Practices	Primary female			
Part 1: Infant and Young Child Feeding		X	X	X
Part 2: Child Anthropometrics			X	X
Part 3: Early Childhood Development (CREDI at midline, TBD at endline)[25]			X	
Module L. Women's Health, Nutritional Status, Dietary Diversity and Nutrition knowledge	Primary female			
Part 1. Women's Dietary Diversity and Access to Antenatal Care		X	X	X
Part 2: Maternal IYCF Knowledge and Perceptions		X	X	X
Part 3: Exposure to Health and Nutrition Services		X	X	X
Part 4: PSNP during Pregnancy and Lactation		X	X	X
Part 5: Maternal Anthropometrics			X	X
Part 6: Early Childhood Development			X	X

[25] The CREDI is designed for children under 30 months and thus would not be appropriate at endline. We may explore the feasibility of using the Malawi Developmental Assessment Tool at endline, given that it's valid for children up to age six.

Module	Respondent	Baseline	Midline	Endline
Module M. Gender (Cash)	Primary female	X		X
Module N. Gender Access to Credit and Group Participation	Primary female	X		X
Module O. Resilience	Primary female			
Part 1. Shocks and Stressors		X		X

4. Analyses

4.1 Estimation and Inference

All regressions will be estimated using the Ordinary Least Squares method. Since geography and environmental factors are strong predictors of poverty and food security (including child growth outcomes) in low-income settings (Karra, Subramanian, & Fink, 2016; Kraay & McKenzie, 2014), it is likely that our primary outcomes (see below) for households residing in the same kebele will be highly correlated. The computed standard errors need to be adjusted for this within-cluster correlation. Following the recommendation by Abadie, Athey, Imbens, and Wooldridge (2017), the standard errors in our regressions will be clustered at the kebele level to account for the randomized design. The cluster-robust standard errors will be computed using Stata's *vce(cluster)* command that adjusts the standard errors based on the Liang and Zeger (1986) approach. All statistical analyses will be conducted using Stata, version 17 or higher.

4.2 Randomization Balance

Randomization balance will be established by comparing baseline data for households in the treatment and control arms. To test for statistical balance, we will run a series of regressions of household characteristics on an indicator variable characterizing the treatment assignment and an indicator variable for randomization strata. Standard errors will be clustered at the kebele level and an F-test will be used to determine whether we can statistically reject the null hypothesis of balance. Variables included in the balance tests will include simple demographic characteristics and the baseline levels of the primary outcomes of interest.

4.3 Main Specifications

We measure impact on our primary and secondary outcomes (listed below) using an analysis of covariance estimation approach (McKenzie, 2012). In our analysis, we will estimate two primary specifications. For livelihood and gender-related outcomes, we are primarily interested in the pooled effect of any treatment (T2 and T3) vis-à-vis the control arm of PSNP only. The regression of interest will be estimated as follows:

$$Y_{id,t=1} = \beta \ T_{id} + \gamma Y_{id,t=0} + X'_{id,t=0}\vartheta + \chi_d + \varepsilon_{id}, \quad (1)$$

where $Y_{id,t=1}$ captures the outcome of interest in household *i* residing in woreda *d* at midline/endline *t* and $Y_{i,t=0}$ at baseline. Variable T_{id} is time-invariant indicator variable, receiving value of 1 if the household is randomly selected to study arm T2 or T3 and zero otherwise. The average impact of the pooled SPIR II interventions relative to the control group (T1) is quantified by β . To assess whether our estimates are sensitive to the inclusion of basic household level controls, we estimate the equation (1) with and without baseline controls (captured in vector $X'_{id,t=0}$), including household size and age and education level of both parents. The term χ_d represent woreda fixed effects, given that we are conducting stratification at the woreda level.

For nutrition-related outcomes, we will estimate the following specification for variables for which a baseline value is available.

$$(2) \qquad Y_{id,t=1} = \beta_1 T_{id}^2 + \beta_2 T_{id}^3 + \gamma Y_{id,t=0} + X'_{id,t=0}\vartheta + \chi_d + \varepsilon_{id}$$

where variables T_{id}^2 and T_{id}^3 are time-invariant indicator variables, receiving value of 1 if the household is randomly selected to receive the T2 or T3 treatment package, respectively, and zero otherwise. We will also report the p-value for the hypothesis that the treatment effects are consistent across treatment arms, $\beta_1 = \beta_2$, to enable us to test whether there is a differential effect of supplementing the NCGs with cash grants. As before, we estimate equation (2) with and without baseline controls ($X'_{id,t=0}$), including household size, age and education level of the primary caregiver, and age and sex of the child.

In addition, for outcomes for which baseline value is not available (e.g., anthropometric measures), equation (3) will be estimated without $Y_{id,t=0}$ as in the following specification.

$$(3) \qquad Y_{id,t=1} = \beta_1 T_{id}^2 + \beta_2 T_{id}^3 \quad + X'_{id,t=0}\vartheta + \chi_d + \varepsilon_{id}$$

In all regressions, our treatment variables are defined based on the initial treatment assignment and not based on actual compliance. Consequently, our impact estimates capture intention-to-treat effects .

In addition to reporting standard p-values, we will also report p-values corrected for multiple hypothesis testing; this correction will be implemented across the set of primary and secondary outcomes in each domain (livelihoods and gender, and nutrition).

4.4 Primary and Secondary Outcomes: Livelihoods and Gender

Table 40 summarizes the primary and secondary outcomes for the first set of analyses. The primary outcomes focus on per capita consumption-expenditures and levels of asset and cash savings. While we will report all primary and secondary outcomes in the relevant evaluation reports, the academic output(s) will focus on a sub-set of the secondary outcomes (see the table below).

Table 41 summarizes the primary and secondary outcomes for the second set of analyses. Here the primary outcomes focus on indicators of chronic child undernutrition and on meeting IYCF related targets. As before, the academic output(s) will focus on a sub-set of secondary outcomes (see the table below).

Table 40: Primary and secondary outcomes: livelihoods and gender

	Reported in the evaluation reports?	Reported in the academic article?
Primary outcomes:		
Daily per-capita consumption-expenditure (BL40)	X	X
Total value of productive assets	X	X
Total value of livestock assets	X	X
Savings (binary and continuous variable)	X	X
Secondary outcomes:		
Food security (BL06 and BL10)	X	X
Prevalence of poverty (BL01)	X	X
Depth of poverty of the poor (BL02)	X	
Net income from livestock production (binary and continuous variable)	X	X
Net income from any non-agricultural production (binary and continuous variable)	X	X
Credit access (binary and continuous variable) (BL42)	X	X
Cash-earning indicators (BL32, BL33, BL34, BL35)	X	

Table 41: Primary and secondary outcomes: nutrition

	Reported in the evaluation reports?	Reported in the academic article?
Primary outcomes:		
Height-for-age (continuous variable, children 30-48 months at endline)	X	X
Prevalence of children 6–23 months consuming a diet of minimum diversity (MDD-C) (at midline) BL39	X	X
IYCF knowledge	X	X
Secondary outcomes:		
Early childhood development score (at midline and endline)	X	X
Percentage of children 6–23 months receiving a minimum acceptable diet (at midline) BL12	X	X
Height-for-age (continuous variable, children 6–23 months at midline)	X	X

	Reported in the evaluation reports?	Reported in the academic article?
Stunting (binary variable, children 6–23 months at midline) BL04	X	
Stunting (binary variable, children 30–48 months at endline) BL04	X	X
Wasting (binary variable, children 30–48 months at endline) BL04	X	X
Weight-for-height Z-score (continuous variable, children 6–23 months at midline)	X	X
Prevalence of healthy weight (WHZ ≤ 2 and ≥ -2) (binary variable, children 30–48 months at endline) (BL05)	X	X

5. Robustness Checks and Additional Analyses

5.1 Survey Attrition

By collecting contact information for each household at baseline and by engaging in multiple follow-up visits in cases in which respondents are not initially reached by enumerators, we hope to minimize survey attrition. Attrition in a previous evaluation conducted in the same region by the SPIR team was in fact less than 10% over an evaluation period of 4 years. Nonetheless, we will test for differential attrition by treatment assignment at the time of endline analysis and will present estimates using an appropriate bounding procedure if differential attrition is detected. We will also report additional specifications in which we regress a binary variable for attrition on the interaction of baseline characteristics and treatment binary variables, assess whether there is differential attrition with respect to baseline characteristics.

5.2 Missing Data from Item Non-Response

Unless explicitly stated above, there will be no imputation for missing data due to item non-response at endline. Missing data on baseline variables will be dummied out of the relevant specifications.

Heterogeneous effects

In addition to the analysis of pooled treatment effects, we will report heterogeneous treatment effects along certain pre-specified dimensions. This analysis should be considered to be exploratory.

The first is child gender (Medhin et al. 2010); given that nutritional practices and outcomes can significantly differ for boys and girls, assessing the differential effect of the proposed interventions by child gender may be important.

The second dimension of heterogeneity that will be assessed is baseline male (paternal) knowledge around and engagement in infant feeding practices. Our hypothesis is that households in which men are more knowledgeable about infant feeding practices at baseline, or more engaged in feeding and caretaking activities, may be more responsive to the interventions and show larger shifts in behavior and outcomes than households in which men show a low baseline level of knowledge and engagement.

6. Power Calculations

We conducted power calculations using the specified sample size, setting the significance level at 5% and power at 80% and allowing for 10% attrition between baseline and endline surveys. Power calculations were conducted in Stata using the command "clustersampsi". Note that given the symmetric design of the study, the minimum detectable effect is identical for any pairwise comparison of arms: the Minimum Detectable Effect (MDE) is the same for comparing T1 or T2 vis-à-vis the control arm, as well as comparing T1 and T2.

We report power calculations for the primary outcomes of interest only. All variables of interest were measured using the data from the SPIR endline survey conducted by IFPRI in 2021, focusing on households in the control arm.[26]

For the livelihoods analysis, the evaluation is able to detect a 20% increase in consumption; a 21% increase in the total value of household assets; a 25% increase in the total value of household livestock assets; and a 12-percentage point increase in the probability that households report any savings.

For the nutrition analysis, the evaluation is able to detect a 0.15-food group improvement in children's dietary diversity, a 0.29-unit change in height-for-age z-score, and a 7% improvement in IYCF knowledge score (constructed on a scale from 1 to 7).

The above calculations are reported assuming a consistent cluster sample size of 13 (i.e., assuming that sampling targets are met in every kebele). Projections suggest that we may have around 10% of clusters characterized by a sample size of 11; this is a coefficient of variation in cluster size of .05. Updating the power calculations using this information generates no differences in the above estimates, because this coefficient of variation remains extremely low.

[26] The mean level of consumption (monthly, per adult equivalent) is 1016 birr, standard deviation 1070, intra-cluster correlation .103. The mean estimated value of total household assets is 19,076 birr, standard deviation 16,302, ICC .25. The mean estimated value of total livestock assets is 16,457 birr, standard deviation 16,194 birr, ICC .25. The probability of reporting any household savings is .47, intra-cluster correlation .23. The mean level of children's food group diversity (number of food groups consumed, aged 6-23 months) is 1.9, standard deviation .96, intra-cluster correlation .05. The mean height-for-age score (children aged 0-5 years) is -1.6, standard deviation 1.95, intra-cluster correlation .03. The mean IYCF score is 3.9, standard deviation 1.27, intra-cluster correlation .19; the minimum detectable effect is thus .29, equal to 7% of the mean.

References for Pre-Analysis Plan

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* NBER Working Paper, (24003). National Bureau of Economic Research, Cambridge, MA.

Alderman, H., & Headey, D. (2018). The timing of growth faltering has important implications for observational analyses of the underlying determinants of nutrition outcomes. *PloS one, 13*(4), e0195904.

Andrews, C., Hsiao, A., & Ralston, L. (2018). Social Safety Nets Promote Poverty Reduction, Increase Resilience, and Expand Opportunities. In K. Beegle, A. Coudouel, & E. Monsalve (Eds.), *Realizing the Full Potential of Social Safety Nets in Africa*. Washington, D.C.: The World Bank.

Beegle, K., Coudouel, A., & Monsalve, E. (2018). Realizing the Full Potential of Social Safety Nets in Africa. In. Washington D.C.: The World Bank.

Berhane, G., Gilligan, D. O., Hoddinott, J., Kumar, N., & Taffesse, A. S. (2014). Can Social Protection Work in Africa? The Impact of Ethiopia's Productive Safety Net Programme. *Economic Development and Cultural Change, 63*(1), 1–26.

Berhane, G., Golan, J., Hirvonen, K., Hoddinott, J., Kim, S., Taffesse, A. S., & Yimer, F. (2020). *Evaluation of the Nutrition-sensitive Features of the Fourth Phase of Ethiopia's Productive Safety Net Programme*. Retrieved from Addis Ababa:

Berhane, G., Hoddinott, J., Kumar, N., & Margolies, A. (2016). The impact of the Productive Safety Net Programme on schooling, child labour and the nutritional status of children in Ethiopia. *3ie Grantee Final Report.*

Davis, T. P., Wetzel, C., Avilan, E. H., de Mendoza Lopes, C., Chase, R. P., Winch, P. J., & Perry, H. B. (2013). Reducing child global undernutrition at scale in Sofala Province, Mozambique, using Care Group Volunteers to communicate health messages to mothers. *Global Health: Science and Practice, 1*(1), 35–51.

De Waal, A. (2017). *Mass starvation: The history and future of famine*. Cambridge, UK: Polity Press.

Golan, J., Headey, D., Hirvonen, K., & Hoddinott, J. (2019). Changes in child undernutrition rates in Ethiopia, 2000-16. In C. Cramer, F. Cheru, & A. Oqubay (Eds.), *The Oxford Handbook of the Ethiopian Economy*. Oxford: Oxford University Press.

Hickey, S., Lavers, T., Niño-Zarazúa, M., & Seekings, J. (2018). The negotiated politics of social protection in sub-Saharan Africa. *UNU-WIDER Working Paper, 2018/34*.

Hidrobo, M., Hoddinott, J., Kumar, N., & Olivier, M. (2018). Social Protection, Food Security, and Asset Formation. *World Development, 101*, 88–103.

Karra, M., Subramanian, S., & Fink, G. (2016). Height in healthy children in low-and middle-income countries: an assessment. *The American journal of clinical nutrition, 105*(1), 121–126.

Knippenberg, E., & Hoddinott, J. (2017). Shocks, social protection, and resilience: Evidence from Ethiopia. *ESSP working paper, 109*.

Kraay, A., & McKenzie, D. (2014). Do poverty traps exist? Assessing the Evidence. *The Journal of Economic Perspectives, 28*(3), 127-148.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13-22.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics, 99*(2), 210-221.

Medhin, G., Hanlon, C., Dewey, M., Alem, A., Tesfaye, F., Worku, B., Tomlinson, M., & Prince, M. (2010). Prevalence and predictors of undernutrition among infants aged six and twelve months in Butajira, Ethiopia: The P-MaMiE Birth Cohort. *BMC Public Health,* **10 (**27).

Ruel, M. T., & Alderman, H. (2013). Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? *The Lancet, 382*(9891), 536-551.

Sabates-Wheeler, R., Lind, J., Hoddinott, J., & Tefera Taye, M. (2021). Graduation after 10 years of Ethiopia's Productive Safety Net Programme: Surviving but still not thriving. *Development Policy Review, 39*(4), 511-531.

Victora, C. G., de Onis, M., Hallal, P. C., Blossner, M., & Shrimpton, R. (2010). Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics, 125*(3), e473-480. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20156903

WHO, & UNICEF. (2021). Indicators for assessing infant and young child feeding practices: definitions and measurement methods.

Wiseman, W., Van Domelen, J., & Coll-Black, S. (2010). *Designing and implementing a rural safety net in a low income setting: Lessons learned from Ethiopia's Productive Safety Net Program 2005-2009*. Retrieved from Washington D.C.: