Implementer-Led Evaluation and Learning (IMPEL)

Evaluation of SPIR II RFSA - Midline Survey Report

February 2024

IMPEL | Implementer-Led Evaluation & Learning Associate Award

USAID FROM THE AMERICAN PEOPLE

Save the Children

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

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 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 Interntational Food Policy Research Institute.

RECOMMENDED CITATION

Gilligan, D., Hirvonen, K., Leight, Tesfaye, H. (2024). Evaluation of SPIR-II RFSA - Midline Survey Report. Washington, DC: The Implementer-Led Evaluation & Learning (IMPEL) Associate Award.

PHOTO CREDITS

Save the Children/Talitha Brauer

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:

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

USAID
FROM THE AMERICAN PEOPLE

Save the Children

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

ACKNOWLEDGMENTS

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, Benita O'Colmain 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 workplan and coordination with the project donor team. Finally, we thank Getahun Tigabu and the EconInsight field team for their excellent work conducting interviews and collecting the data.

TABLE OF CONTENTS

Acknowledgments i

Acronyms iv

Executive Summary iv

Overview and study design iv
Data, Outcomes, and Sample iv
Exposure to SPIR interventions v
Exposure to recent shocks v
Nutrition outcomes vi
Livelihood outcomes vii
Summing up vii

1. Introduction 1

2. Methodology 3

2.1 Evaluation design 3
2.2 Interventions 4
2.3 Study arms 5
2.3.1 Surveys 6
2.3.2 Econometric analyses 11
2.3.3 Robustness checks and additional analyses 12
2.3.4 Limitations 13
2.4 Primary and secondary outcomes of interest 13

3. The Midline survey 16

4. Exposure to SPIR interventions 21

5. Experience of recent shocks 28

6. Primary Findings: Nutrition outcomes 30

6.1 Caregivers' nutrition knowledge 30
6.2 Infant and young child feeding practices 32
6.3 Child diarrhea risk 36
6.4 Child anthropometry 37
6.5 Early childhood development 42
6.6 Women's dietary diversity 43

7. Primary Findings: Livelihood outcomes 48

7.1 Food security indicators 48

7.1.1 Food Insecurity Experience Scale (FIES) 48
7.1.2 Food Consumption Score (FCS) 51
7.1.3 Food gap 53
7.2 Household consumption and poverty 53

8. Heterogeneity analyses 59
8.1 Livelihood grant eligibility 59
8.2 Livelihood outcomes, by treatment arm 63

9. Conclusion 66

References 68

Appendix A. Additional Tables and Figures 71

ACRONYMS

BHA	USAID's Bureau for Humanitarian Assistance
BL/ML/EL	Baseline/midline/endline
cRCT	Cluster randomized control trial
DFAP	Development Food Assistance Project
DFSA	Development Food Security Activity
FCS	Food Consumption Score
FFP	USAID's Office of Food for Peace
FIES	Food Insecurity Experience Scale
HAZ	Height for age Z-score
ICC	Intra-cluster correlation
IMPEL	Implementer-Led Evaluation and Learning
IP	Implementing partner
IYCF	Infant and young child feeding
M&E	Monitoring and evaluation
MDD-C	Minimum dietary diversity (MDD) score for children 6-23 months old
MDE	Minimum detectable effect
NCG	Nurturing Care Groups
NG	Neighbor Group
PPP	Purchasing power parity
PSNP	Productive Safety Net Programme
RCT	Randomized control trial
SBCC	Social and behavior change communication
SD	Standard deviation
SE	Standard error
SPIR	Strengthen PSNP Institutions and Resilience
TOC	Theory of change
VESA	Village economic and savings association
WASH	Water, sanitation, and hygiene
WHZ	Weight for height Z-score
WFP	World Food Program

EXECUTIVE SUMMARY

Overview and study design

The objective of this report is to present results from the midline survey conducted as part of the IMPEL evaluation of SPIR-II, a randomized controlled trial launched in 2022. The second phase of the Strengthen PSNP Institutions and Resilience (SPIR-II) project aims to enhance livelihoods, increase resilience to shocks, and improve food security and nutrition for rural households vulnerable to food insecurity in Ethiopia. The project 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 the kebele level in 234 kebeles. In the first arm (T1; 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 (T2), SPIR-II programming is rolled out to PSNP beneficiary households in conjunction with nurturing care groups targeting enhanced infant and young child nutritional practices. In the third arm (T3), PSNP beneficiary households receive SPIR-II programming and nurturing care groups (NCGs), supplemented with additional targeted cash transfers 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, that entailed interviews with both the primary female respondent (pregnant or lactating woman) and her spouse. The midline survey was conducted between August 2023 and January 2024 with households in the Amhara region interviewed later due to the ongoing conflict in the region. The endline survey is scheduled for August—October 2025, roughly 24 months after the midline and 36 months after the baseline.

The primary outcomes for the evaluation were defined in a pre-analysis plan prior to the baseline survey. For livelihood 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 (MDD-C) at midline, and infant and young child feeding (IYCF) 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 nine months of age (along with his / her mother or primary caretaker),

or a pregnant woman. Statistical tests indicate that the randomization achieved balance: the baseline data showed no evidence of any significant differences in observable characteristics across the three study arms.

The midline survey aimed to recontact all 3,015 households interviewed at the baseline. The survey commenced in Oromia in mid-August, concluding in late October. The midline survey was launched later in Amhara due to widespread conflict in the region. In Amhara, the survey teams were able to launch the survey at the beginning of October, completing interviews in 68 out of 80 selected kebeles by mid-November. The remaining 12 kebeles were surveyed in December 2023.

Out of the 3,015 households interviewed at baseline, 2,954 were successfully interviewed at midline, resulting in an attrition rate of a mere two percent. We see no meaningful differences in recontact rates across the three study arms.

The sample size used in the nutrition analyses is smaller, primarily due to a higher-than-expected number of unsuccessful pregnancies in Oromia. Among households initially enrolled with an infant at baseline and successfully contacted at midline, nearly 98 percent reported the child was still present. However, in Oromia, only 59 percent of pregnant woman-enrolled households had an infant or young child at midline, compared to 83 percent in Amhara. Additionally, we excluded 167 households due to child age (either younger than 6 months or older than 23 months), resulting in a sample of 2,233 children that we use in the nutrition analyses.

Exposure to SPIR interventions

Analysis of households' exposure to SPIR interventions reveals a high degree of implementation fidelity. At midline, households residing in SPIR intervention kebeles were more than four times more likely to belong to a village economic and savings association (VESA) than households in non-intervention kebeles. As intended, roughly one-third of the households in the treatment kebeles report receiving the livelihood grant, and they predominantly allocated this grant to livestock-related activities such as fattening or purchasing more animals. Meanwhile, only four percent of the control households (T1 arm) report receipt of a livelihood grant.

As for the core nutrition interventions, participation in nurturing care groups was high in both treatment arms, ranging between 75 and 89 percent of the households in T2 and T3 arms, respectively. Twenty-two percent of the households in the control arms also reported to have taken part in NG session. However, there seemed to be confusion among some control households in this regard, with many reporting that these group sessions were led by a health extension worker, and not a peer lead mother. Over 92 percent of the households in T3 reported receiving maternal cash transfers, aligning with the program design, while the proportions in the other study arms were less than 2 percent. Caregivers who received maternal cash transfers primarily reported that the funds were directed to child feeding-related expenses.

Exposure to recent shocks

The midline survey asked about households' experiences with shocks over the past year (since the baseline survey). In Amhara, the most frequently reported shock type was increasing food prices,

reported by 92 percent of households, followed by rising input prices at 65 percent. Around half of the households in Amhara faced challenges related to crop diseases and delays in receiving food or cash assistance. Thirty-nine percent of the households reported insufficient rainfall, while 40 percent experienced unpredictable rainfall. Fifteen percent reported instances of violence. In Oromia, similarly, increasing food prices affected most households, reported by 91 percent. Insufficient or unpredictable rainfall was reported by 75 percent and 65 percent of the households, respectively, while 21 percent encountered locust-related issues, and only 2 percent reported incidents of violence.

Nutrition outcomes

The project's theory of change posits that combining intensive behavior change communication to enhance caregiver knowledge with the provision of maternal cash transfers, specifically aimed at alleviating economic constraints to child feeding, will lead to improved diets and feeding practices among infants and young children, preventing linear growth faltering. The midline evaluation primarily focuses on assessing the initial impacts at the initial stages of this theory, particularly on caregiver knowledge and infant and young child feeding (IYCF) practices. The endline, to be administered in 2025 when the children are 30-47 months, will be more appropriate for assessing the impact on indicators of linear growth faltering: height for age Z-score and child stunting prevalence.

At midline, caregivers in SPIR treatment areas showed better nutrition knowledge compared to those in control areas. However, while statistically significant, these differences were not substantial in terms of magnitude, possibly due to the simplicity of the knowledge quiz.

Evaluating IYCF practices for children aged 6-23 months revealed that children in the T3 arm, receiving maternal cash transfers, had better outcomes compared to T1 and T2. Around 30 percent of children in T3 achieved a diet characterized by minimum diversity, significantly higher than T1 (13%) and T2 (14%). Children in T3 were more likely to consume eggs, dairy, legumes and nuts as well as Vitamin A rich fruits and vegetables compared to other study arms; they were also more likely to achieve the criteria for minimum meal frequency (MMF). Additionally, 22 percent of the children in T3 arm met the threshold for a minimum acceptable diet (MAD) while only 9 percent of children in T1 and 11 percent in T2 did so. These differences in IYCF practices between T3 and the other study arms are all statistically significant and attributable to the maternal cash transfers combined with other SPIR programming.

As expected, these significant improvements in IYCF feeding practices in T3 arm have not yet translated into substantially better anthropometric outcomes. At midline, 37 percent of T1 children and 38 percent of T2 children were classified as stunted, while in T3, this was slightly lower at 35 percent; this difference is narrowly statistically significant at the ten percent level. Wasting prevalence ranged between 6-8 percent, with no statistically distinct differences between the study arms.

We also measured early childhood development using the CREDI instrument. Consistent with the findings around IYCF practices and anthropometrics, we observe evidence of a large and precisely estimated increase in measured early childhood development scores in T3 only, where CREDI scores are around 0.2 standard deviations higher compared to the control arm (T1). No comparable effect is observed in T2.

Livelihood outcomes

Anticipating limited impacts on livelihood outcomes in the initial year of implementation, the midline survey focused on a small subset of livelihood indicators. According to the Food Insecurity Experience Scale, severe food insecurity decreased from 58 percent at baseline to 43 percent at midline, while moderately food-insecure households increased from 32 percent to 46 percent at midline. The trends in control and SPIR households were similar without statistically significant differences. By midline, nearly 90 percent of households reported an acceptable Food Consumption Score (FCS), up from 75 percent at baseline. Treatment estimates indicate a slightly more improved FCS among SPIR households compared to the control group.

Extreme poverty remains high. In control (T1) households, the mean per capita consumption fell from $1.78 to $1.48 between baseline and midline while in SPIR (T2 and T3) households it fell from $1.80 to $1.58. The percentage of individuals below the $1.90 poverty line increased from 70 percent to 82 percent in T1 households and from 70 percent to 78 percent among households in the pooled treatment arms (T2 and T3). A regional breakdown of the poverty numbers reveals that poverty headcount for the full sample increased more in Oromia (from 68.2% to 79.0%) than in Amhara (from 77.5% to 79.2%). The differences in consumption and poverty trends are small between control and SPIR households, indicating that the SPIR livelihood programming did not meaningfully impact these outcomes between baseline and midline. Furthermore, additional analysis that separates the T2 and T3 arms indicates that the positive effects on livelihood outcomes are predominantly attributed to households in the T3 arm, which received the maternal cash transfers.

Summing up

The SPIR project displays a strong adherence to its planned implementation. However, the extremely high poverty prevalence combined with a highly volatile context characterized by multiple types of shocks poses ongoing challenges. Nevertheless, this midline analysis highlights significant improvements in IYCF practices among households in areas where maternal cash transfers, in combination with other SPIR interventions, are provided. Considering the theory of change, this emphasizes the necessity of addressing both caregiver knowledge and resource constraints to enhance child diets and feeding practices in this setting. Overall, these findings present promising prospects for detecting positive effects on children's anthropometric outcomes at the endline.

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, and Monsalve 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, Coudouel, and Monsalve 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, and Ralston 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 health (WASH) training and other interventions to better address the underlying causes of maternal and child malnutrition (Ruel and Alderman 2013).

Against this backdrop, the second phase of the Strengthen PSNP Institutions and Resilience (SPIR-II) project aims to enhance livelihoods, increase resilience to shocks, and improve food security and nutrition for rural households vulnerable to food insecurity in Ethiopia. The project 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 the Amhara and Oromia regions of Ethiopia.

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 (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 transfers to pregnant and lactating women.

This impact evaluation will contribute to current evidence on the effectiveness of graduation model programs, particularly by adding new evidence on graduation models including substantial nutrition programming, and on whether the effectiveness of such a model is enhanced by a supplemental maternal cash transfer. There is considerable evidence from an evaluation of a six-country study of programs designed like BRAC's Targeting the Ultra-Poor (TUP) 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. 2022, Banerjee et al. 2016, Banerjee et al. 2015, Bandiera et al. 2017). However, an important evidence gap persists around the question of 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 (Leight et al. 2023). The impact evaluation of the first phase of the SPIR project (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. 2023, Alderman et al. 2022). The findings suggested that an important factor limiting program impact on child nutrition was the strategy employed for nutrition behavior change communication, relying on community health facilitators who were unable to achieve sufficiently intensive interaction with SPIR households.

The innovations in nutrition-related programming that are a focus of this evaluation center around enhancing 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 transfers 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 transfer.

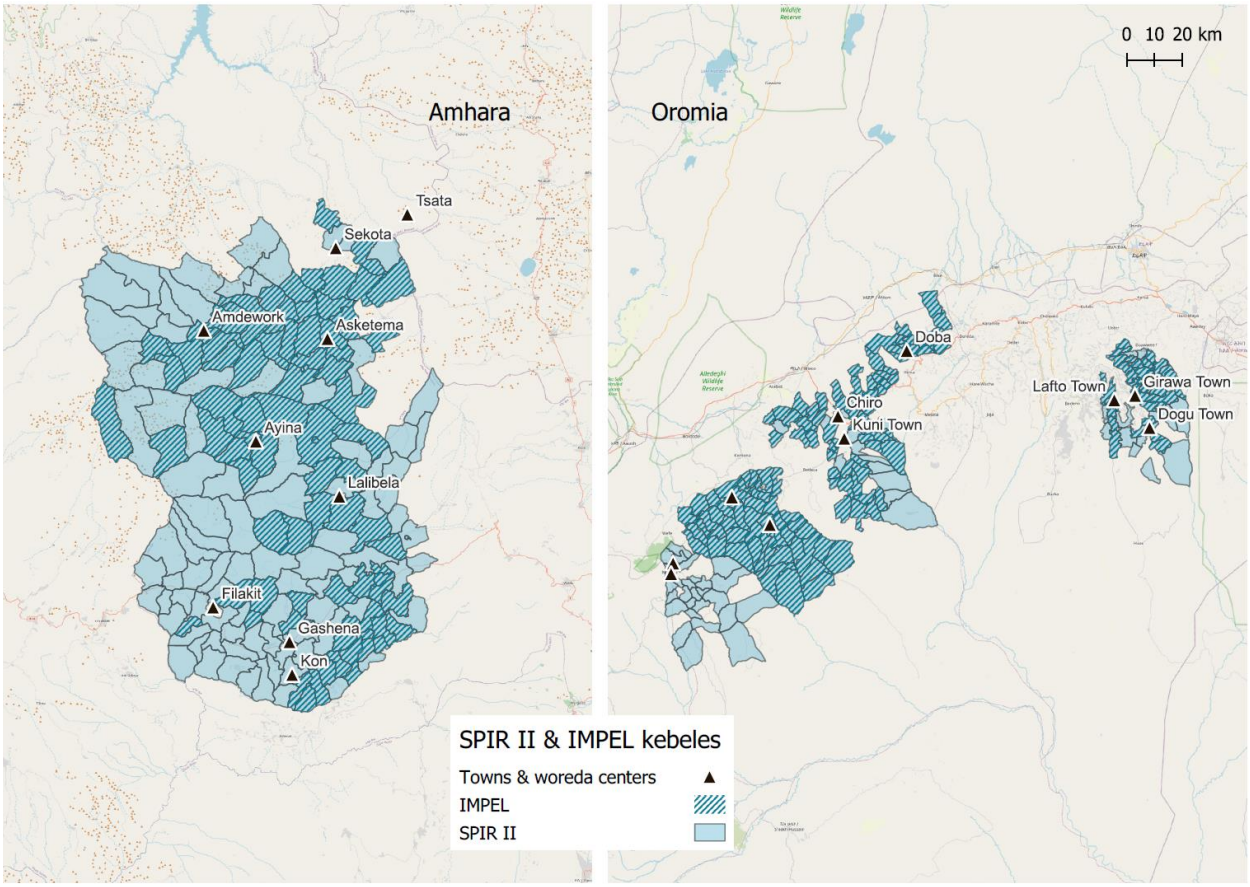
The baseline survey for this impact evaluation was conducted in August—September 2022 with results published in (Gilligan et al. 2023). The midline survey was conducted between August and November 2023 (with a final, short phase of data collection in conflict-affected areas of Amhara in December 2023 – January 2024). The objective of this document is to provide an overview of the results from the midline assessment that focuses on assessing implementation fidelity and impacts on infant and young child feeding practices, as specified in the pre-analysis plan conducted prior to the midline survey (Leight et al. 2023). The endline assessment is scheduled for 2025, and the focus will be quantifying impacts on child anthropometry and livelihood outcomes.

2. METHODOLOGY

2.1 Evaluation design

This evaluation is a cluster randomized control trial (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 IFPRI and World Vision under SPIR: this 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 eligible kebeles (three) were excluded from the study due to security concerns at the time of the baseline survey in 2022. Figure 2.1 shows the sample kebeles, and the IMPEL evaluation kebeles.[1]

Figure 2.1. SPIR-II and IMPEL evaluation kebeles



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

[1] The full operational area served by SPIR-II includes 17 woredas and 465 kebeles.

2.2 Interventions

The IMPEL impact evaluation is centered around three study arms where all households at the baseline benefitted from the PSNP. Households in T2 and T3 arms are subject to SPIR graduation and nutrition related programming whereas households in T1 arm do not.

Core SPIR graduation programming includes the organization of village economic and social associations (VESAs), used as a platform for trainings and other project 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 thirty percent of households in T2 and T3 arms is also targeted for a one-time $300 livelihood grant; these are the households identified as the poorest using an asset-based welfare index constructed by the IFPRI team from the baseline survey data.[2]

The first major goal of this evaluation is to assess the effectiveness of integrated SPIR 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[3] is largely unavailable, rendering this evaluation a meaningful contribution.

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 (T3) introduces maternal cash transfers of $20 per month during the child's first 24 months of life to relax possible financial constraints to child feeding. These cash transfers 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% increase in consumption. This is consistent with evidence in the

[2] At the kebele level, only around 10% of households outside of the IMPEL sample are eligible for livelihood grant. This eligibility process was determined separate from the IMPEL study and is not described here.

[3] 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 counseling to pregnant women and mothers of young children. In addition, the PSNP recently added explicit nutrition objectives and complementary nutrition program components, including temporary direct support transfers for 12 months beginning at conception, as well as nutrition trainings to mothers of children under 2, sometimes provided as fulfilment of their public works requirement.

literature that the most successful extant conditional cash transfer programs generally provide transfers of between 10% and 20% of household consumption (Fiszbein and Schady 2009).[4]

Table 2.1 summarizes the interventions across each treatment arm.

Table 2.1: Interventions by treatment arm

Treatment arm	Core interventions	Supplemental nutrition interventions
T1	PSNP transfers implemented with SPIR II support	N/A
T2	PSNP transfers implemented with SPIR II support VESAs (used as a platform for other project activities around financial literacy, promotion of savings and credit use, agriculture, livestock value chain development, improving social capital, and catalyzing women's empowerment) Livelihood grants for poorest 30% of households (identified using asset index)	Nurturing care groups
T3	PSNP transfers implemented with SPIR II support VESAs (used as a platform for other project activities around financial literacy, promotion of savings and credit use, agriculture, livestock value chain development, improving social capital, and catalyzing women's empowerment) Livelihood grants for poorest 30% of households (identified using asset index)	Nurturing care groups and maternal cash transfers

2.3 Study arms

To summarize, the impact evaluation includes three study arms:

- T1. PSNP transfers implemented by the government with SPIR-II support (no supplemental programming) (79 kebeles)
- T2. PSNP + SPIR-II + NCG model (79 kebeles)
- T3. PSNP + SPIR-II + NCG + maternal cash transfers (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 maternal cash transfers.

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

[4] The relevant evidence can be found in Table 2 of the cited report.

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 cash transfers to reduce cost constraints to improving these outcomes.

2.3.1 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.

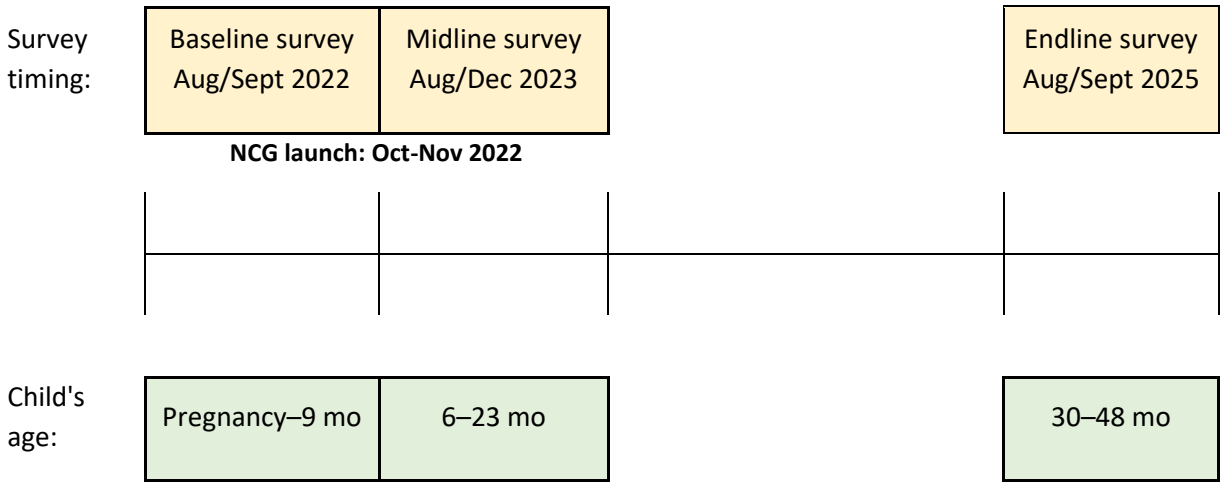
ii) The midline survey was originally scheduled for August—September 2023, around 12 months following the baseline, but due to unforeseen events took place between August and December 2023 (see Section3). Interviews were conducted with the primary female respondent only. Child anthropometric measurements were also 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. Child anthropometric measurements will also be collected.

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 2.2). Prior to the survey, we randomly allocated kebeles into the three study arms.

We planned to administer a midline quantitative survey to the entire baseline sample one year after the baseline survey when the children are between 6—23 months of age, permitting 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 had 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 program objectives.

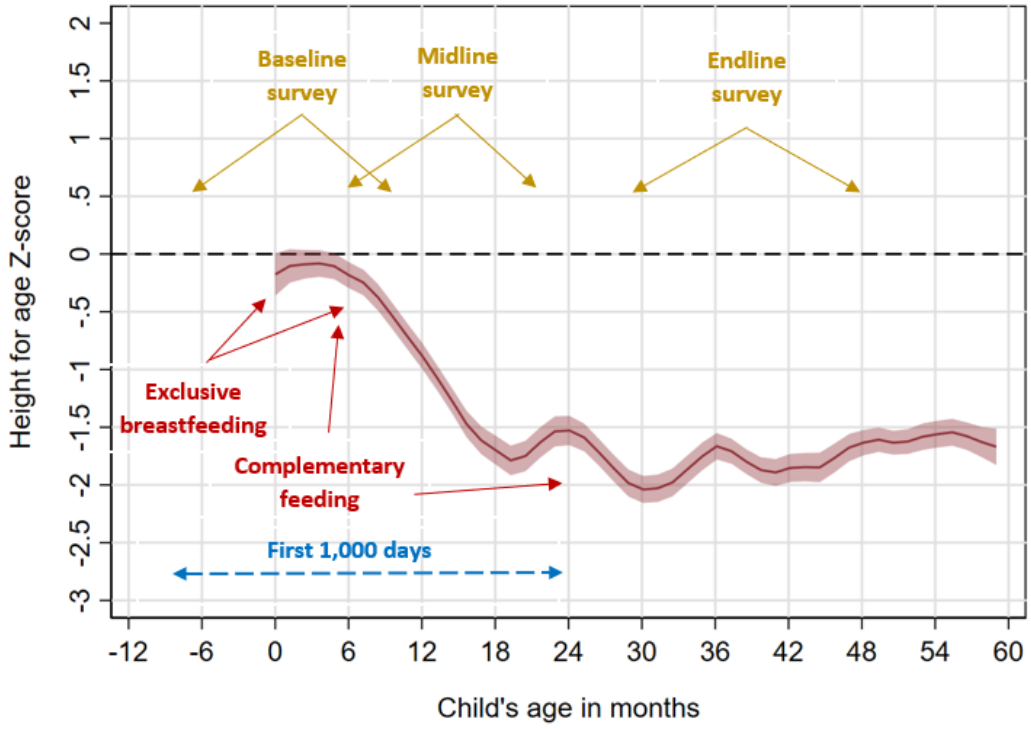
Figure 2.2: Timing of the surveys



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) 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 2.3 based on the data from the 2015/16 Ethiopia Demographic and Health Survey (DHS), during the first four to five 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.[5] 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.

[5] 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 2, 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.

Figure 2.3: 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 2.2 below summarizes the key questionnaire modules to be included in each wave of data collection. The planned modules for d endline, however, may require revision based on resource constraints.

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 percent 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 program in follow-up surveys.

Table 2.2: 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, 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 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 cash transfers	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
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

Module	Respondent	Baseline	Midline	Endline
Part 2: Child anthropometrics			X	X
Part 3: Early childhood development (CREDI at midline, TBD at endline)[6]			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.

[6] 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.

2.3.2 Econometric analyses

All regressions are estimated using Ordinary Least Squares (OLS). Since geography and environmental factors are strong predictors of poverty and food security (including child growth outcomes) in low-income settings (Karra, Subramanian, and Fink 2016, Kraay and 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. 2022), the standard errors in our regressions are clustered at the kebele level to account for the randomized design. The cluster-robust standard errors are computed using Stata's *vce(cluster)* command that adjusts the standard errors (Liang and Zeger 1986). All statistical analyses are conducted using Stata, version 17 or higher.

Randomization balance was established by comparing baseline data for households in the treatment and control arms. To test for statistical balance, we estimated 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 were reported in the baseline report (Gilligan et al. 2023).

We measure the impact on our primary and secondary outcomes (listed below) using an analysis of covariance (ANCOVA) estimation approach (McKenzie 2012). In our analysis, we estimate two primary specifications. For livelihood and gender-related outcomes, we are primarily interested in the pooled effect of any treatment (T2 and T3) relative to the control arm of PSNP only (T1). The regression of interest is 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 strata fixed effects to account for stratification during randomization.

For nutrition-related outcomes, we 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 transfers. 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 treatment effect estimates capture intention-to-treat effects (ITT).

Each analysis is 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 2.3 and Table 2.4.

In addition to reporting standard p-values, we 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). This multiple hypothesis testing will be conducted in the endline report.

2.3.3 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 team was in fact less than 10 percent over an evaluation period of four 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, in order 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 will be carried out in the endline report, and it should be considered to be exploratory as we may not be adequately powered to conduct these sub-group analyses.

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.

2.3.4 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.[7] 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 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.

Fourth, while we place the findings from this trial within the broader literature analyzing graduation model programs, it is important to note that SPIR II has a particularly broad set of objectives compared to many other graduation interventions. Because of these broader objectives, interpreting the findings in relation to the graduation model intervention may not be an exact comparison.

2.4 Primary and secondary outcomes of interest

Table 2.3 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 2.4 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).

[7] 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.

Table 2.3: 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
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 2.4: 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
Percent 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 (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

Detailed power calculations are reported in Leight et al. (2023). For the livelihoods analysis, the evaluation is able to detect a 20 percent increase in consumption; a 21 percent increase in the total value of household assets; a 25 percent 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 livelihood related outcomes, the main analysis will pool data from the two treatment arms to compare against the control arm. Given this design, the evaluation is able to detect a 0.34-unit change in the total asset index (0.22 SD) and a 9-percentage point increase in the probability that a household reports any savings (relative to an estimated mean probability of 47 percent in the control arm). For per capita consumption, revised power calculations based on baseline data suggest we can detect a 7-percent increase in per capita consumption. Relative to the effect sizes observed in the previous trial conducted as part of SPIR I, the evaluation is adequately powered to detect effects on savings: the increase in the probability of any savings was observed to be between 25 and 30 points. For consumption, no significant effect of the underlying reduced graduation model on consumption was observed in SPIR I.

3. THE MIDLINE SURVEY

At the baseline, the sample was formed of PSNP households with a pregnant woman or a child less than nine months of age. In the evaluation, we follow this cohort of children through the study period, and the survey rounds are carefully timed to capture the adherence to age-appropriate IYCF practices, and the growth-faltering dynamics as described in the pre-analysis plan. The baseline survey included 3,015 households in 234 kebeles. Roughly half of the households were enrolled in the study because of a pregnant woman in the household and another half because they had an infant (child less than 9 months of age) in the household.

The midline survey attempted to contact and re-interview all 3,015 households interviewed at the baseline. The midline survey was conducted during a period of widespread conflict in the Amhara region. Therefore, the survey began in the study kebeles located in the Oromia region. The survey teams began their work on 14 August 2023 with the final interviews in the region conducted on 23 September. In Amhara, the situation in the areas selected for the study sufficiently improved so that the survey teams were able to begin their work on 5 October 2023. Around mid-November, the team returned to Addis Ababa after having completed interviews in 68 kebeles out of the total of 80 Amhara kebeles. The remaining 12 kebeles were at that point actively insecure, rendering them inaccessible. Security conditions in these communities improved at the end of December and the survey teams were able to visit the remaining kebeles in the Amhara region between the end of December 2023 and mid-January 2024.

Table 3.1 provides a summary of the number of interviews conducted by region and by survey round. The household level attrition rate at midline is extremely low: 98 percent of households that were interviewed at the baseline were also interviewed at midline. About 2 percent of the baseline households were not interviewed at midline, either because they had relocated out of the study area or because they refused to be interviewed. The attrition rates across study arms range between 0.8 percent in T3 to 2.6 percent in T1.

Table 3.1. Household interviews and attrition, by region

Region	Interviewed	Not interviewed	Total	Attrition rate
Amhara	1,019	19	1,038	1.8%
Oromia	1,935	42	1,977	2.1%
Total	**2,954**	**61**	**3,015**	**2.0%**

Table 3.2 provides an overview of the attrition rate at index child level. The total attrition rate at the index child level is 20 percent. This high attrition rate is primarily due to the unexpectedly large number of unsuccessful pregnancies in the Oromia region, where the index child level attrition rate is almost 25 percent. By contrast, the index level attrition rates are considerably lower among households that were enrolled with an infant at the baseline (Table 3.2).

In Oromia, out of the households that were enrolled as a pregnant woman and successfully interviewed at midline, only 59 percent did in fact report the presence of an infant or a young child in their household at midline (Table 3.3). In Amhara, 83 percent out of the households that were enrolled as a pregnant woman and successfully interviewed at midline had an infant or a young child in their household at midline.

There are two primary reasons that could explain why households in which a woman self-reported pregnancy at baseline do not have a child present at midline. One is that the pregnancy did not conclude in a live birth (or the child subsequently passed away following the birth but prior to the survey date). The second is that the woman was not in fact pregnant at baseline but reported that she was (either believing that she was, or feigning pregnancy). Both survey and intervention staff members shared that women may be motivated to feign pregnancy in a survey in the hope of subsequently receiving some benefit targeted to pregnant women, even though the survey exercise was not explicitly linked to any programmatic benefit provided by either SPIR or any government agency. Among those women who did not have a child present, the survey directly posed a question as to what had happened to their child and 75 percent reported that they had miscarried, and 21 percent said that their child had passed away after delivery; 4 percent admitted they had not been pregnant at baseline. However, these reports may still be inaccurate if women who had previously misreported their pregnancy status are unwilling to admit to this misreporting.

Large-scale evidence on miscarriages is extremely limited in rural Ethiopia, as in many low- and middle-income countries, and major surveys such as the Demographic and Health Surveys generally do not collect data on miscarriages (only stillbirths). However, available evidence from a demographic surveillance site in Eastern Ethiopia suggests a miscarriage rate of around 23 percent (Assefa, Berhane, and Worku 2013) , very close to the rate observed in this sample; however, these data are more than a decade old. We also worked with the survey firm to further cross-check the cases of reported incomplete pregnancies in one woreda (Habro) with the local community health facilitator (CHF, employed by World Vision). In this woreda, 17 women who were pregnant at baseline did not report the presence of an infant at midline; the CHFs reported that based on the information available to them, 12 of these cases corresponded to true miscarriages, and in 5 cases the women were never in fact pregnant. We were not able to conduct this exercise on a large scale due to the limited time availability of the CHFs, but this small case study is consistent with the hypothesis of a high underlying miscarriage rate. (It is also important to note, however, that the CHFs' information may also be imperfect.)

Table 3.2. Number of index children found, by region and baseline enrollment status

	Found	Not found	Total	Attrition rate
By region:				
Amhara	917	121	1,038	11.7
Oromia	1,493	484	1,977	24.5
By baseline enrollment status:				
Pregnant	1,065	541	1,606	33.7
Infant	1,345	64	1,409	4.5
Total	**2,410**	**605**	**3,015**	**20.1**

Table 3.3. Percent of households with an index children found conditional on midline household interview, by region and baseline enrollment status

	Amhara	Oromia
Pregnant woman (%)	83.2	58.9
Infant (%)	98.1	97.3
Total (%)	**90.0**	**76.9**

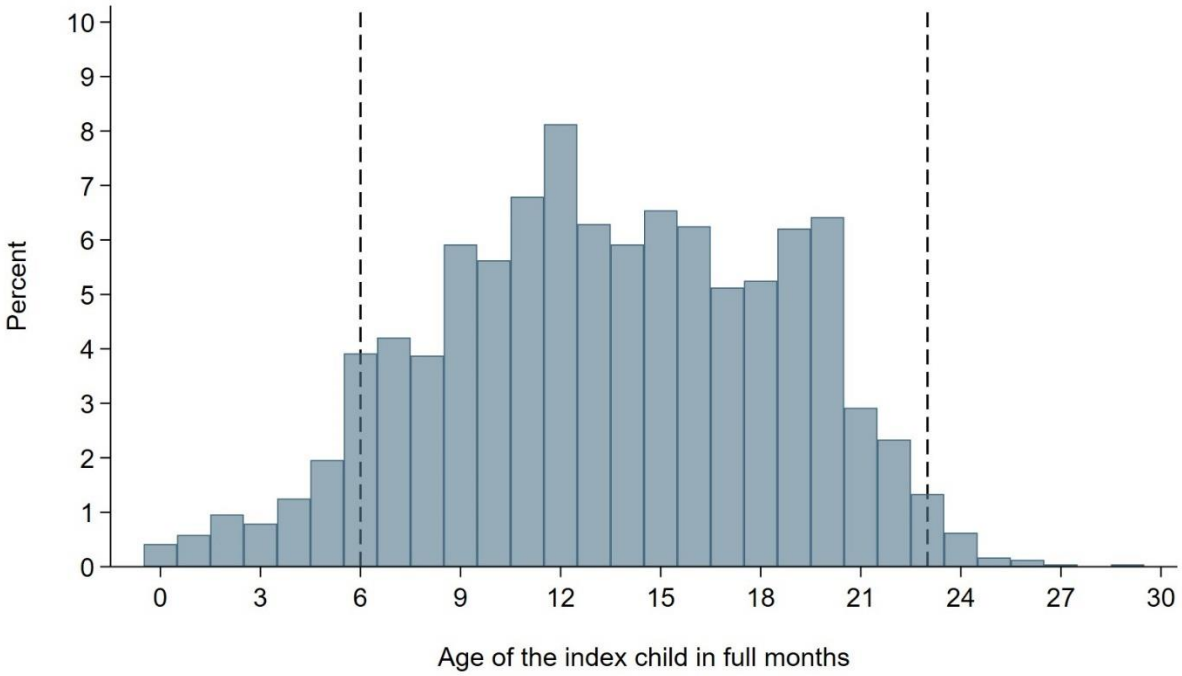
N = 2,954 households.

Finally, the IYCF indicators considered in the midline evaluation are validated for children 6-23 months of age. Figure 3.1 shows the distribution of the ages of the index children found at midline. 143 index children are less than 6 months of age, and 24 children are more than 23 months of age. We drop these 167 children from the sample for all nutrition indicators. The resulting sample used in the nutrition evaluation is formed of 2,233 children.

Our initial power calculations took into account a 10 percent attrition rate between surveys (Leight et al. 2023). The household-level attrition rate at midline (2%) falls well within this range. However, the higher index-level attrition rate (20%) raises concerns about whether our nutrition analyses remain adequately powered. To address this, we revised our power calculation, taking into consideration that the average cluster (kebele) in our midline sample includes 9 index children, rather than the 12 initially assumed. Fortunately, this reduced number of children per cluster has a negligible impact on the minimum detectable effect sizes, as detailed in the last two rows of Table 3.4.[8]

[8] This is because the statistical power in a cRCT is typically influenced more by the number of clusters than by the number of individuals within each cluster.

Figure 3.1. Distribution of the index child ages at midline



N = 2,40 index children. The area between the vertical dashed lines are children in the target age range, 6-23 months.

Table 3.4. Power calculations for the nutrition outcomes based on SPIR-I endline data, original and revised estimates

	Original power calculations			Revised power calculations		
	Caregiver IYCF knowledge	Number of food groups consumed (children 6–23 months)	Height-for-age (children 30-48 months)	Caregiver IYCF knowledge	Number of food groups consumed (children 6–23 months)	Height-for-age (children 30-48 months)
Index children per cluster	12	12	12	9	9	9
Number of clusters	79	79	79	79	79	79
Baseline mean	3.9	1.9	-2.07	3.9	1.9	-2.07
Baseline standard deviation (SD)	1.27	0.96	1.59	1.27	0.96	1.59
Adjusted SD [1)]	1.17	0.92	1.55	1.17	0.92	1.55
Intra-cluster correlation (ICC)	0.24	0.001	0.02	0.24	0.001	0.02
Adjusted minimum detectable effect (MDE) size	0.264	0.119	0.222	0.275	0.137	0.250
MDE relative to SD	0.23	0.13	0.14	0.24	0.15	0.16

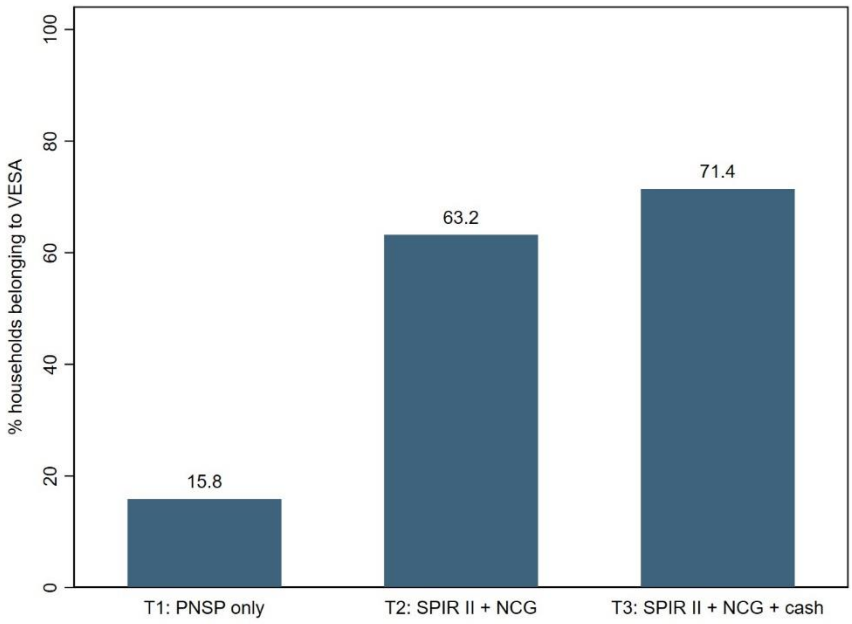
Note: Baseline means, SDs and ICCs estimated using the SPIR-1 endline data. [1)] Adjusted for the reduced variance in outcomes when including controls for baseline levels (where applicable) and woreda fixed effects.

4. EXPOSURE TO SPIR INTERVENTIONS

In this section we analyze and report household exposure to the core SPIR interventions.

The livelihood interventions in SPIR center around the formation of village economic and savings associations (VESAs). Figure 4.1 shows the share of households who are members a VESA, by survey round and study arm. We see that 63 percent of households in T2 arm report membership in a VESA while the corresponding share is 71 percent in T3 arm. Meanwhile, only 16 percent of households in the T1 arm (where there is no supplemental SPIR programming) report having joined a VESA. Figure A1 in the Appendix disaggregates these data by region. The share of T1 and T2 households reporting to be part of a VESA is slightly higher in Amhara than in Oromia. More than 70 percent of the households that reported joining a VESA did so in 2023.

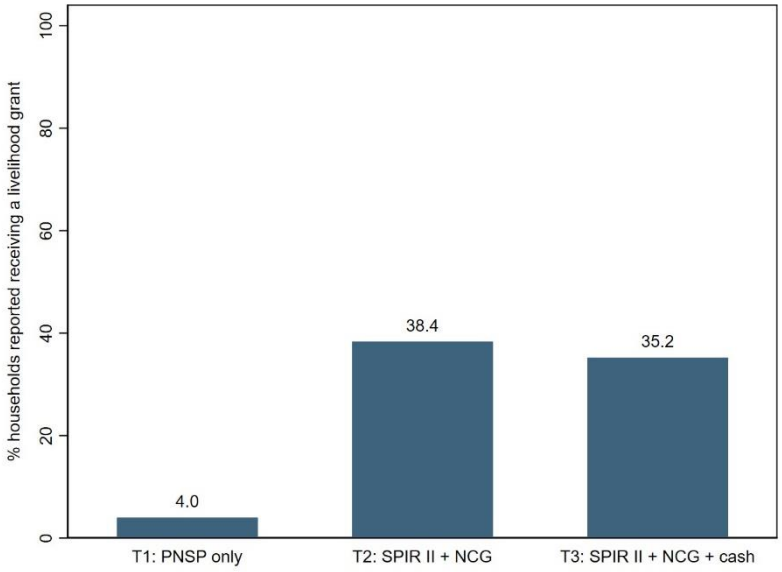
Figure 4.1. Percent of households who are members of a village economic and savings association (VESA), by study arm



N = 2,954 households.

In both the T2 and T3 arms, households eligible for a $300 livelihood grant were determined by selecting the poorest 30 percent within each kebele. Household poverty status was assessed using an asset-based welfare index measured using the baseline data. In T2 arm, 38 percent of households report receipt of such a grant, and 35 percent in T3 arm (Figure 4.2). A small number of households (4%) in the T1 arm report receipt of a livelihood grant. In Amhara, about one-third of the households in T2 and T3 arms report receipt of a livelihood grant and 11 percent of the households in T1 arm (Panel A in Figure A2 in the Appendix). In Oromia, less than 0.5 percent of the households in T1 arm report receiving a livelihood grant, and about 40 percent of the T2 and T3 report receiving the grant (Panel B in Figure A2 in the Appendix). The mean livelihood grant received was 15,518 birr (median 15,800 birr).

Figure 4.2. Percent of households reporting to have received a livelihood grant, by study arm

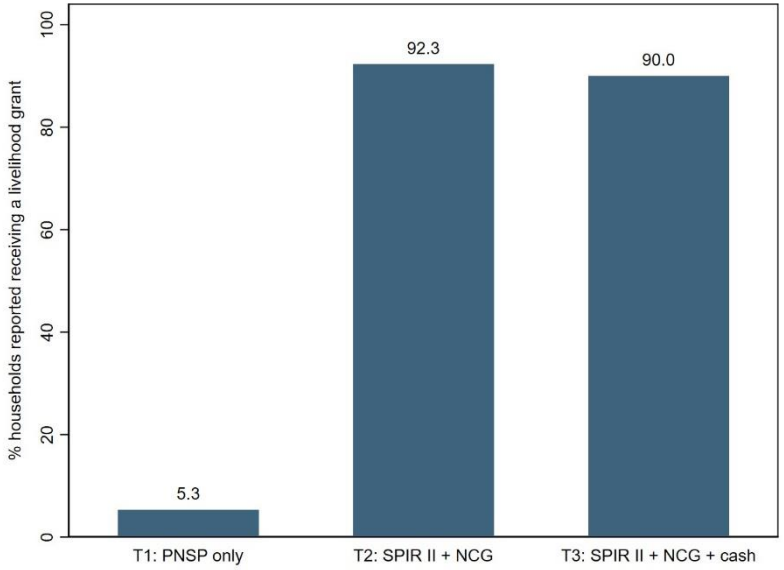


N = 2,954 households.

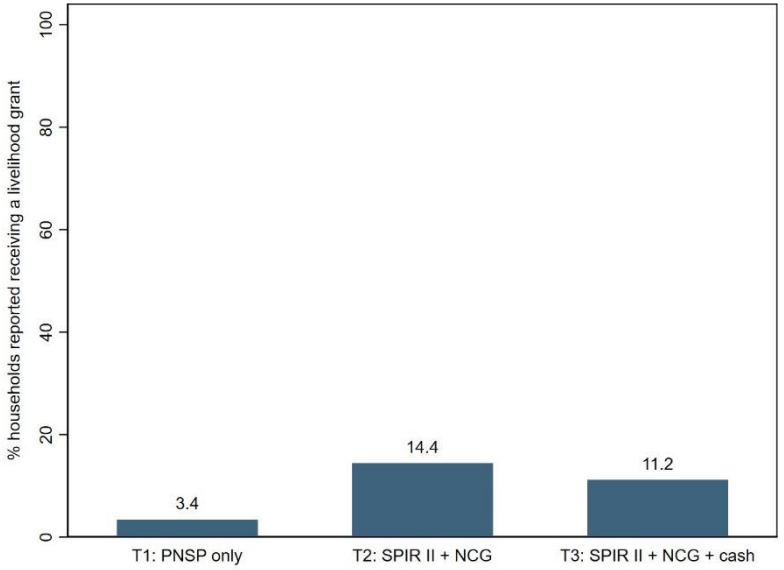
In Figure 4.3, we compare these reports based on household asset-based poverty status at baseline. We identify households eligible for the livelihood grant as 'extremely poor' and those ineligible for the transfer as 'less poor'. In Panel A of Figure 4.3, over 90 percent of 'extremely poor' households in both T2 and T3 reported receiving the livelihood grant between the baseline and midline surveys, while only 5 percent of T1 households did the same. However, Panel B of Figure 4.3 reveals some non-compliance, with 14 percent of 'less poor' households in the T2 arm and 11 percent in the T3 arm reporting receipt of the livelihood grant despite being ineligible.

Figure 4.3. Percent of households reporting to have received a livelihood grant, by study arm and baseline asset poverty status

A. 'Extremely poor' (N = 897 households)

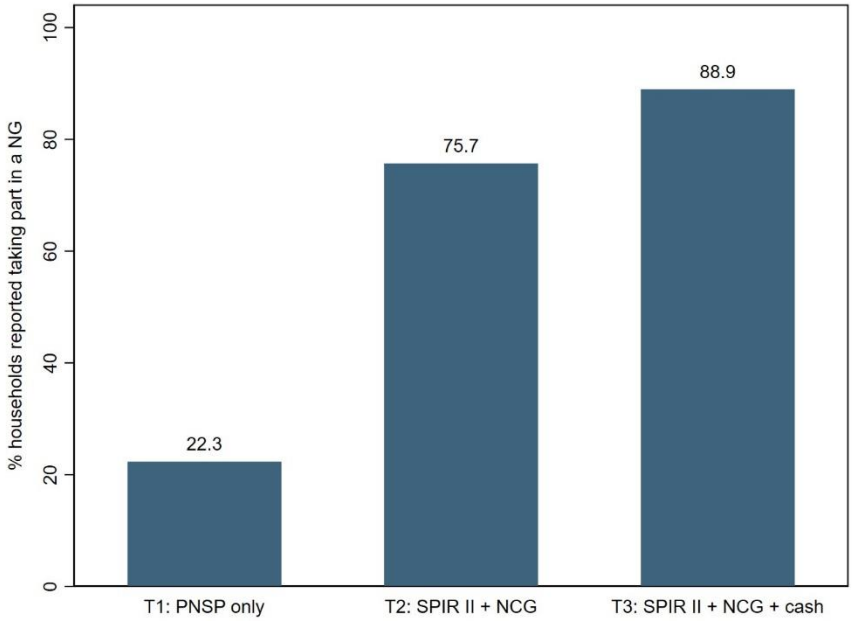


B. 'Less poor' (N = 2,057 households)



The behavioral change communication delivered as a part of the SPIR nutrition programming is delivered through neighbor groups (NGs) facilitated by nurturing care group volunteers. The midline data show that the participation in these groups among caregivers in the two treatment arms is very high (76—89%), again indicating high implementation fidelity (Figure 4.4). About 22 percent of the T1 households report taking part in NGs (Figure 3.3), with a considerably higher share among T1 households located in Amhara (34%) than in Oromia (15%) as shown by Figure A3 in the Appendix. The NGs are primarily facilitated by NG leaders or peer mothers. In the survey instrument we asked the respondents who facilitated these NG sessions. Figure 4.5 tabulates the responses by treatment status. Panel A shows that about half of the T1 respondents that reported taking part in the NGs said that these sessions were led by a health extension worker (HEW), indicating that many respondents may have confused these groups with the government health extension program operating in their kebele or to some other activity led by HEWs. Figure A4 in the Appendix disaggregates Panel A in Figure 4.3. by region.[9]

Figure 4.4. Percent of households reporting to have taken part in NGs, by study arm

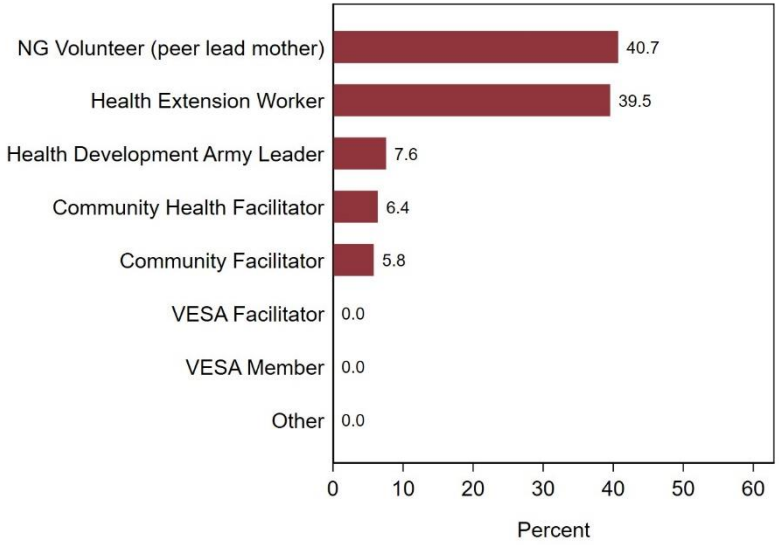


N = 2,233 households with an index child 6-23 months of age.

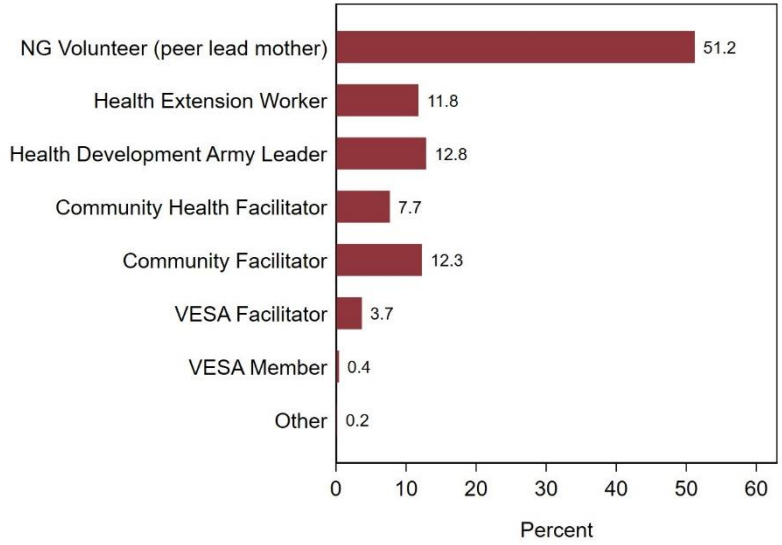
[9] The implementing partner later verified with the staff on the ground that NCGs had not been rolled out in the T1 kebeles.

Figure 4.5. NG facilitators as reported by households reporting to have taken part in NGs, by treatment status

C. T1 (N = 172 households)

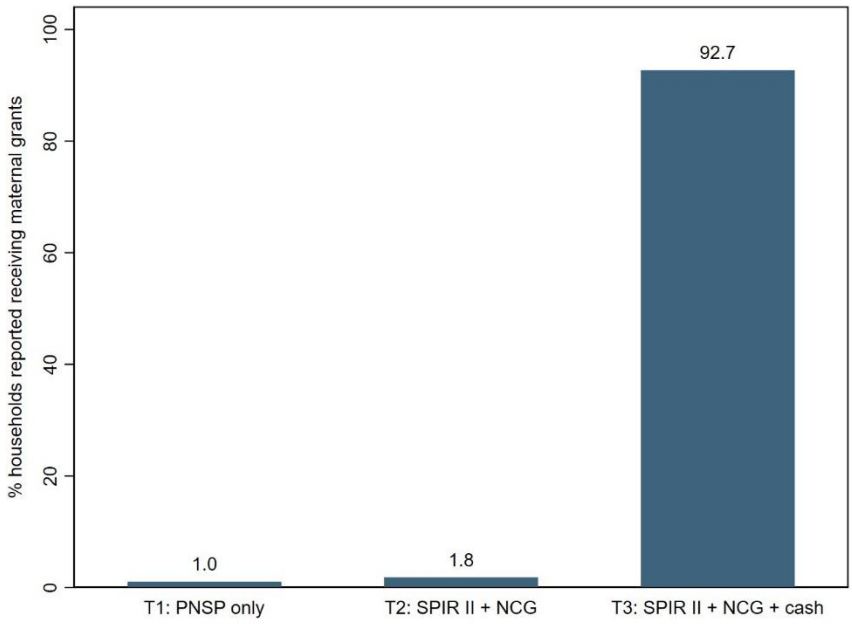


D. T2 and T3 (N = 1,199 households)



Finally, nearly all households in the maternal cash transfer arm (T3) report receiving maternal cash transfers while the opposite is true for households in T1 and T2 arms (Figure 4.6), with little differences across regions (Figure A5 in the Appendix). The respondents reported receiving their first maternal cash transfers largely between December 2022 and March 2023. The average maternal cash transfers recipient reported receiving seven transfers with the mean amount of the latest transfer being 1,524 birr (median 1,100 birr).

Figure 4.6. Percent of households reporting to have received maternal cash transfers, by study arm



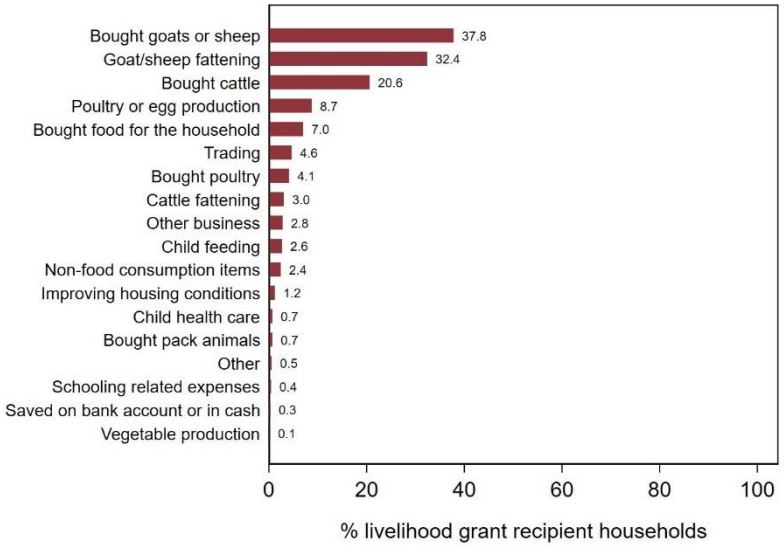
N = 2,233 households with an index child 6-23 months of age.

The livelihood and maternal cash transfers were deposited into bank accounts with the implementation teams assisting recipient households in establishing these accounts. At midline, respondents were asked whether their household operated a bank account. In the T1 arm, 60 percent of households confirmed having a bank account. For the T2 group, 99 percent of those receiving a livelihood grant had a bank account, while 68 percent of those without the grant had one. Over 98 percent of T3 households reported operating a bank account.

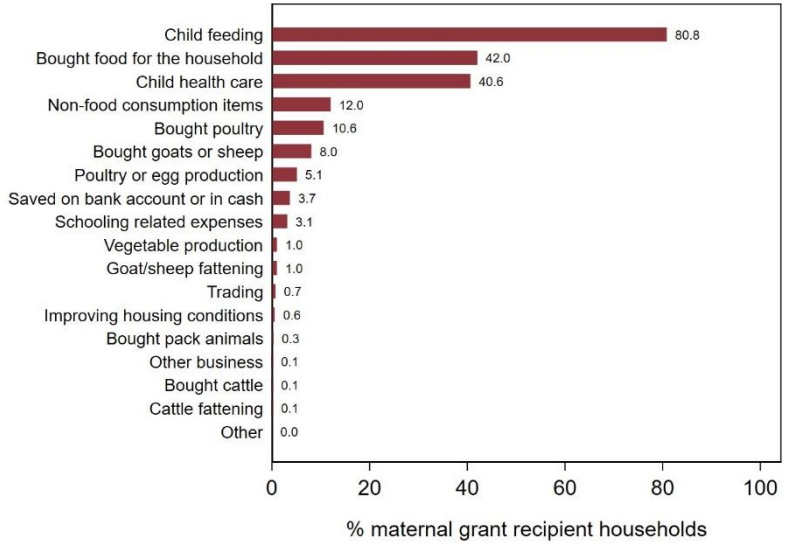
Finally, we asked households how they used the grants or transfers they received, allowing for multiple responses. Panel A in Figure 4.7 summarizes the responses for the livelihood grant and Panel B the responses regarding the maternal cash transfer. Those who received the livelihood grant predominantly directed it towards activities involving livestock: 60 percent of grant-receiving households mentioned buying or fattening goats or sheep, 20 percent invested in cattle, and 15 percent focused on purchasing poultry or making investments to poultry or egg production. Conversely, most households receiving maternal cash transfers indicated that they used the money primarily for feeding their children.

Figure 4.7. Household utilization of livelihood and maternal cash transfers

A. Livelihood grant (N = 757 households)



B. Maternal cash transfers (N = 709 households)



5. EXPERIENCE OF RECENT SHOCKS

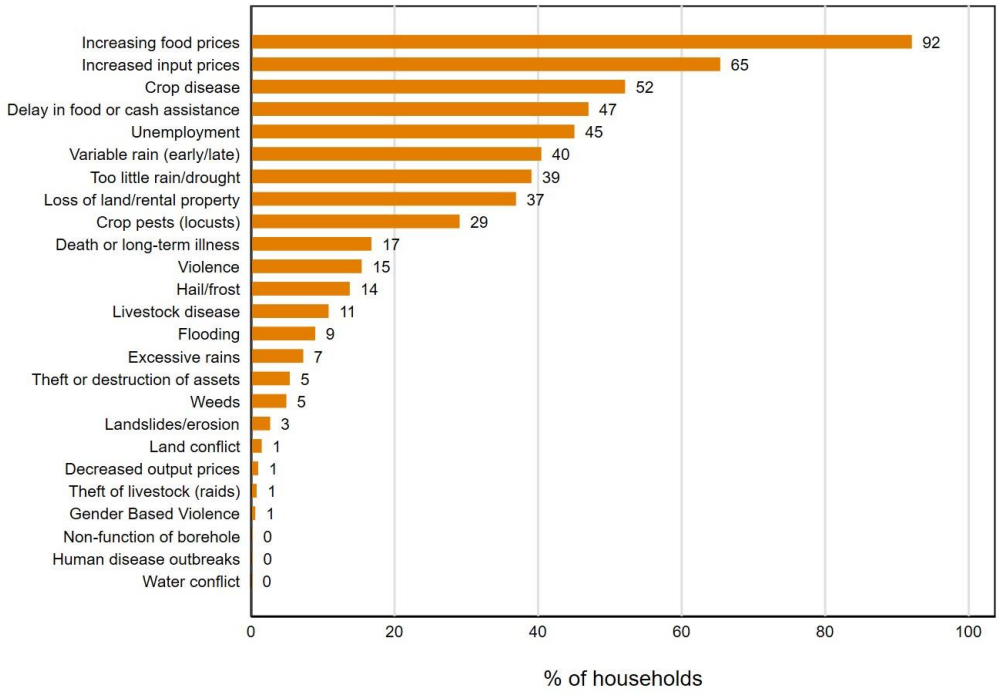
The baseline and midline survey instruments asked a series of questions about recent shocks that the households had experienced in the past 12 months. In total, the shocks module included 25 different types of shocks. At baseline, the average household in the Amhara sample reported exposure to 5.7 different shocks in the previous 12 months, while in the Oromia sample, this number was 6.5. These numbers are slightly lower at midline; the average household in the Amhara sample reported exposure to 5.3 different shocks in the previous 12 months, compared to 5.6 in Oromia.

Figures 5.1 and 5.2 show the prevalence of specific shocks reported at midline, for Amhara and Oromia, respectively. In Amhara, increasing food and input prices were the most prevalent shock types, reported by 92 percent and 65 percent of the households, respectively. Crop diseases and delays in food or cash assistance were reported by roughly half of the sampled households.[10] Violence was reported by 15 percent of the households.

In Oromia, increasing food prices were also the most prevalent shock, reported by 91 percent of the households. Too little or unpredictability of rainfall were the next most prevalent shocks, reported by 75 and 65 percent of the households, respectively. Delays in food and cash assistance were reported by 39 percent of the households in Oromia, locusts by 21 percent and violence by 2 percent.

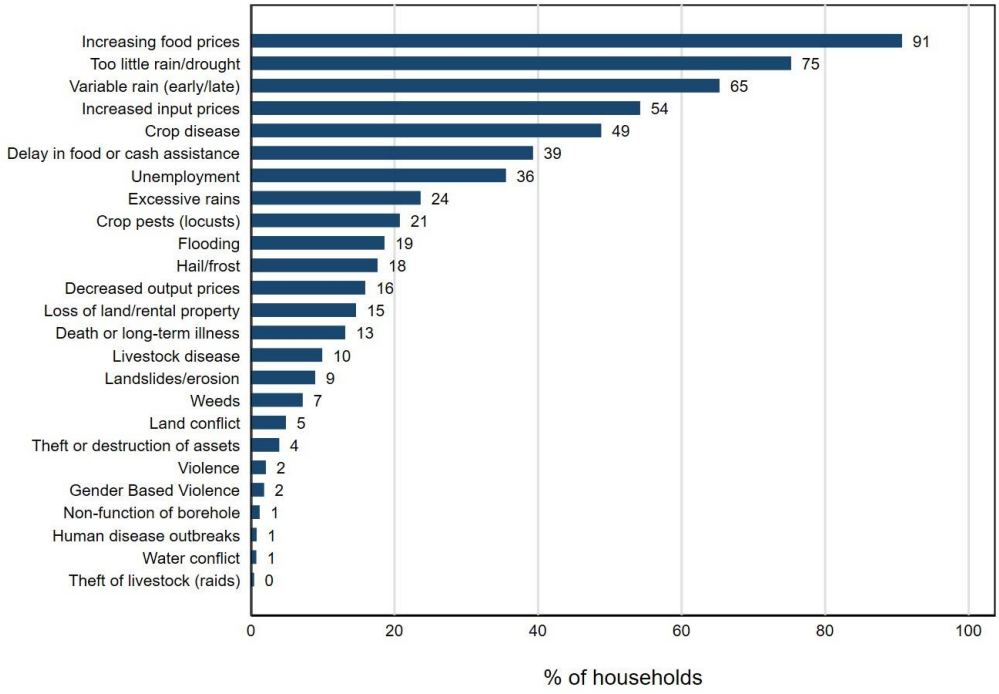
[10] This finding aligns with results from past PSNP evaluations, which demonstrate that delays in PSNP payments are both recurring and widespread, leading to irregular lump-sum payments (Berhane, Hirvonen, and Hoddinott 2016, Berhane et al. 2014, Sabates-Wheeler et al. 2022). In particular, the cash payments delivered by the PSNP are not managed by SPIR or its operating partners.

Figure 5.1. Prevalence of shocks in the past 12 months, Amhara



N = 1,038 households.

Figure 5.2. Prevalence of shocks in the past 12 months, Oromia



N = 1,977 households.

6. PRIMARY FINDINGS: NUTRITION OUTCOMES

The project's theory of change posits that combining intensive behavior change communication to enhance caregiver knowledge with the provision of maternal cash transfers, specifically aimed at alleviating economic constraints, will lead to improved diets and feeding practices among infants and young children, preventing linear growth faltering. At midline, the focus of the evaluation lies in evaluating impacts at the upstream of the theory of change: caregiver knowledge and child diets and feeding practices. The endline, to be administered in 2025 when the sample children are 30-47 months, is appropriate for assessing the impact on linear growth faltering.

6.1 Caregivers' nutrition knowledge

Starting with the first step in the theory of chain, we examine the impact of SPIR programming on caregivers' nutrition knowledge, measured through responses to a nutrition knowledge quiz administered as a part of both baseline and midline surveys. The quiz had 11 questions focusing on recommended breastfeeding and complementary feeding practices (Table 6.1). The nutrition knowledge score was then calculated as the total number of correct responses, ranging from a score with a minimum value of zero and a maximum value of 11. If multiple correct answers existed, a point was awarded if the respondent identified at least one of them.

Table 6.1 summarizes the questions in the nutrition knowledge quiz and shows the share of mothers responding correctly to them. At baseline, the average knowledge score ranged between 8.2 and 8.5 across the three study arms, indicating that the quiz may not have been very challenging for the majority of caregivers. Table 6.1 additionally shows the proportion of correct responses from women for each question, highlighting that mothers exhibit a somewhat better understanding of the recommended breastfeeding practices compared to complementary feeding practices, although the differences are marginal.

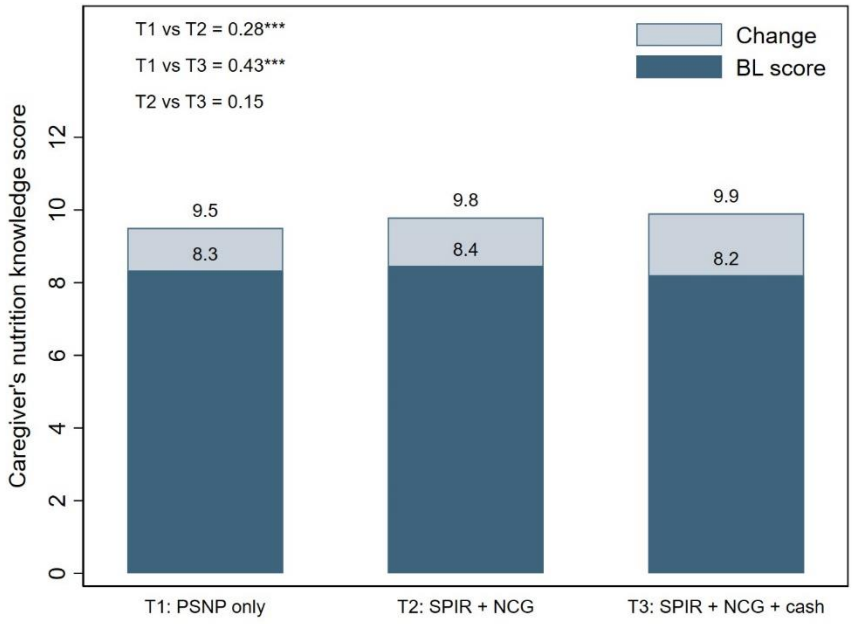
At midline, nutrition knowledge across all study arms improves further, including in T1 (control) arm, presumably as caregivers acquire relevant information while their infants grow. The improvement was larger in study arms T2 and T3 that were exposed to the BCC provided through the NGs than in the T1 group that received now SPIR programming (Figure 6.1). While statistically significant, the differences in average knowledge scores across the study arms are relatively small in terms of magnitude. Row 1b in Table A3 in the Appendix shows the conditional treatment effect estimates based on a regression that controls for pre-specified household, caregiver, and index child characteristics as well as strata fixed effects. The conditional treatment effect estimates are similar to the unconditional estimates reported in Figure 6.1.

If we adjust the strategy employed to grade questions with multiple correct answers, awarding one point for each right response, this substantially reduces the average respondent score compared to the maximum possible (see Table A5 in the Appendix). In addition, this adjustment creates somewhat larger gaps between the study groups at the midline (Figure A6 in the Appendix). For example, the 0.43 treatment effect for 'T1 vs T3' reported in Figure 6.1 translates into a 5.2 percent increase when we compare it to the baseline mean among the control group. The corresponding treatment effect reported in Figure A6 is 12.3 percent.

Table 6.1. Percent of caregivers responding correctly to the nutrition knowledge questions, by study arm

Survey round:	Baseline			Midline		
Treatment status:	T1	T2	T3	T1	T2	T3
Percent of caregivers responding correctly:						
How long after birth should a baby start breastfeeding (1p)	88.3	87.5	86.6	90.4	92.9	91.1
What should a mother do with the 'first milk' or colostrum (1p)	72.0	73.6	72.2	80.0	84.1	84.8
Until what age should a baby by exclusively breastfed (1p)	87.7	88.3	90.7	95.2	97.2	97.0
Why should a baby under 6 months be exclusively breastfed (1p)	87.0	87.1	83.5	96.1	97.6	95.8
If baby is not getting enough breast milk, what should be done (1p)	81.1	81.6	80.2	89.2	90.7	88.1
At what age should a baby first start to receive liquids (1p)	84.7	86.1	85.7	86.9	86.6	91.0
At what age should a baby first start to receive foods (1p)	72.6	74.1	70.7	78.5	78.6	78.5
What can happen to children if they do not get enough iron (1p)	68.0	70.4	64.5	89.1	92.1	94.5
What are some foods that are rich in iron (1p)	55.1	56.9	53.3	79.4	83.4	86.1
What can happen if child does not eat enough vitamin A-rich foods (1p)	69.9	71.5	67.5	85.5	90.4	92.8
What are some foods that contain vitamin A (1p)	63.6	65.5	62.1	80.4	85.7	90.8
Mean nutrition knowledge score (max 11 points):	8.3	8.4	8.2	9.5	9.8	9.9

Figure 6.1. Caregiver's nutrition knowledge score, by study arm



Note: N = 2,231 female caregivers. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

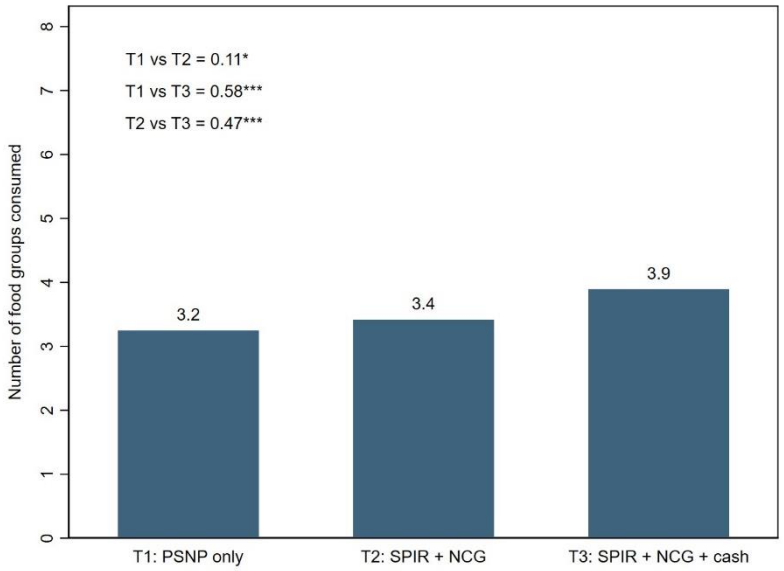
6.2 Infant and young child feeding practices

Next, we assess the impact of the treatments on children's diets and feeding practices. Since about half of the index children in our sample were not born at baseline, we focus on the data collected at the midline. The survey instrument asks caregivers a series of Yes/No questions about children's consumption of different foods and liquids in the 24 hours before the interview. Following recently revised WHO and UNICEF guidelines (WHO and UNICEF 2021), we grouped foods into eight groups: breastmilk; grains, roots, and tubers; legumes and nuts; dairy products; flesh foods; eggs; vitamin A rich fruits and vegetables; and other fruits and vegetables. Children's dietary quality is measured with a count variable capturing the number of food groups consumed by the child (i.e., with a minimum value of zero and a maximum value of eight). A child achieves a diet of minimum diversity (MDD-C) if she or he consumes from 5 or more food groups in the previous 24 hours prior to the interview.

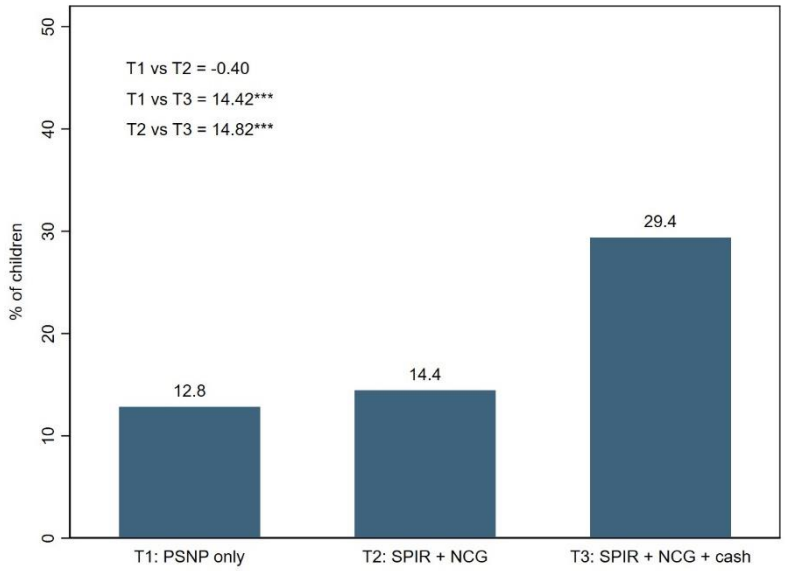
Panel A of Figure 6.2 shows the average child in the T1 arm consumed from 3.2 groups (min = 1, median = 3, max = 7), while the average child in the T2 and T3 arms who consumed from 3.4 (min = 1, median = 3, max = 7) and 3.9 food groups (min = 1, median = 4, max = 7), respectively. The conditional treatment effect estimates indicate that the mean differences across study arms are statistically different at least at the 10-percent level. Panel B shows that 13 percent of the children in the T1 arm achieved MDD-C. In the T2 arm that were exposed to SPIR programming, including NGs but not maternal cash transfers, 14 percent of the children achieved MDD-C. The difference between T1 and T2 arms is not statistically significant. However, in T3, 23 percent of children achieved MDD-C, a level considerably higher than what is observed in the other two study arms. These differences are statistically highly significant. The conditional treatment effect estimates are similar to the unconditional estimates (Row 2a in Table A3 in the Appendix).

Figure 6.2. Diet diversity of children 6–23 months, by study arm

Panel A. Number of food groups consumed



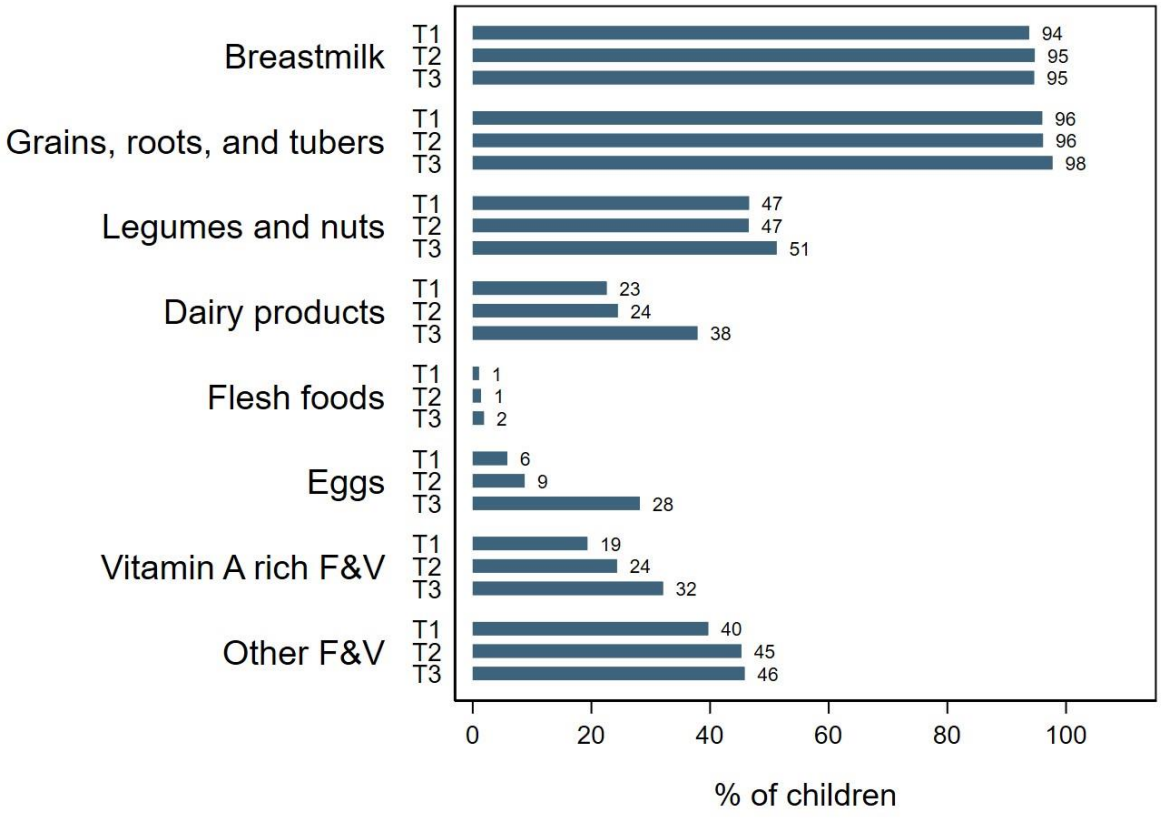
Panel B. % consuming a diet of minimum diversity (MDD-C)



Note: N = 2,221 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 6.3 provides an overview of the dietary patterns. Almost all children received breastmilk and grains, roots, and tubers. Legumes and nuts as well as fruits and vegetables not rich in vitamin A were fed to more than 40 percent of the children. Children in T3 were considerably more likely to receive eggs and dairy products. We also observe that children in this arm are more likely to consume legume and nuts as well as Vitamin A rich fruits and vegetables, compared to the children in the other two study arms (T1 and T2).

Figure 6.3, Percent of children 6-23 months consuming from different food groups, by study arm

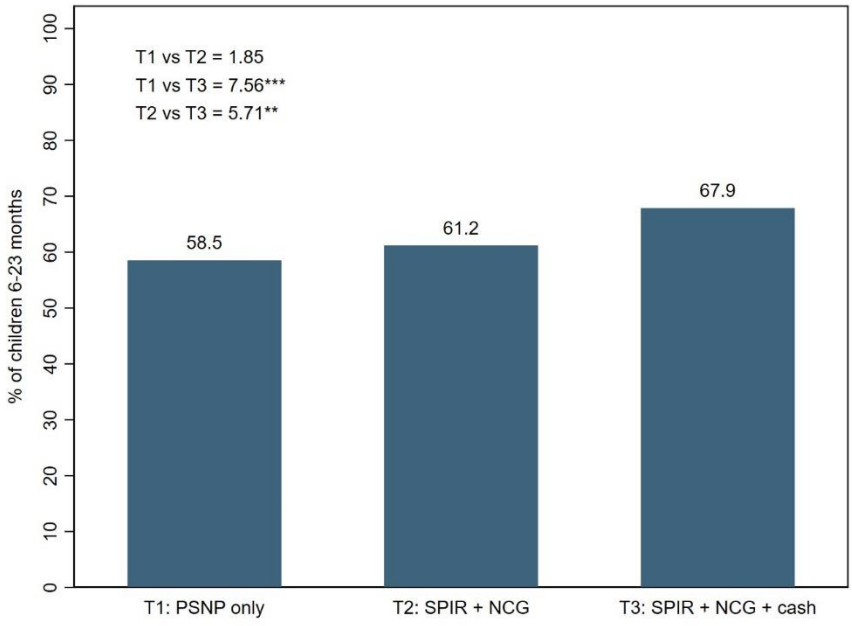


Note: N = 2,221 children.

Figure 6.4 shows the prevalence of children achieving minimum meal frequency (MMF), which 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. Breastfed children aged 6-8 months should receive at least two feedings of solid, semi-solid or soft foods and those aged 9-23 at least three such feedings. Non-breastfed children should receive four feedings of solid, semi-solid or soft foods or milk feeds out of which at least one must be a solid, semi-solid or soft feed. Using this definition, 59 and 61 percent of the children in T1 and T2 study arm achieved MMF (the difference between the two arms is not statistically significant). The corresponding share was higher in the T3 arm, where 68 percent of children 6-23 achieved MMF, respectively. The differences between T3 and the T1 and T2 arms are statistically

significant. As before, the conditional treatment effect estimates are similar to the unconditional estimates (Row 3a in Table A3 in the Appendix).

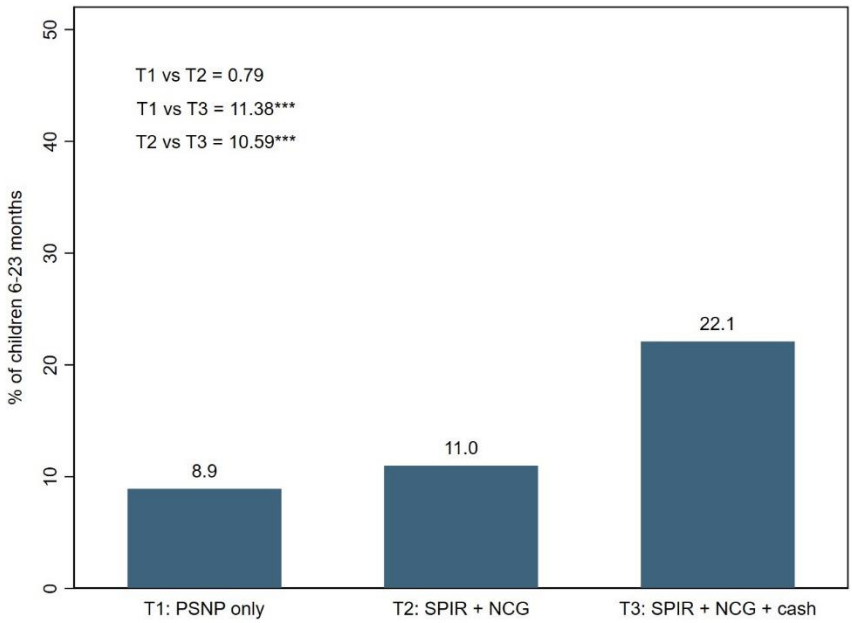
Figure 6.4. Percent of children 6–23 months achieving minimum meal frequency, by study arm



Note: N = 2,221 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 6.5 reports the impacts on the prevalence of children 6-23 months meeting the minimum acceptable diet (MAD), achieved if the child receives both MMF and MDD-C. In T1 and T2, 9 and 11 percent of children meet MAD. The share is considerably higher in T3 arm where 22 percent of the children meet MAD. The differences between T3 and the other two arms are statistically significant. The conditional treatment effect estimates are very similar to the unconditional estimates (Row 4a in Table A3 in the Appendix).

Figure 6.5. Percent of children 6–23 months receiving a minimum acceptable diet, by study arm



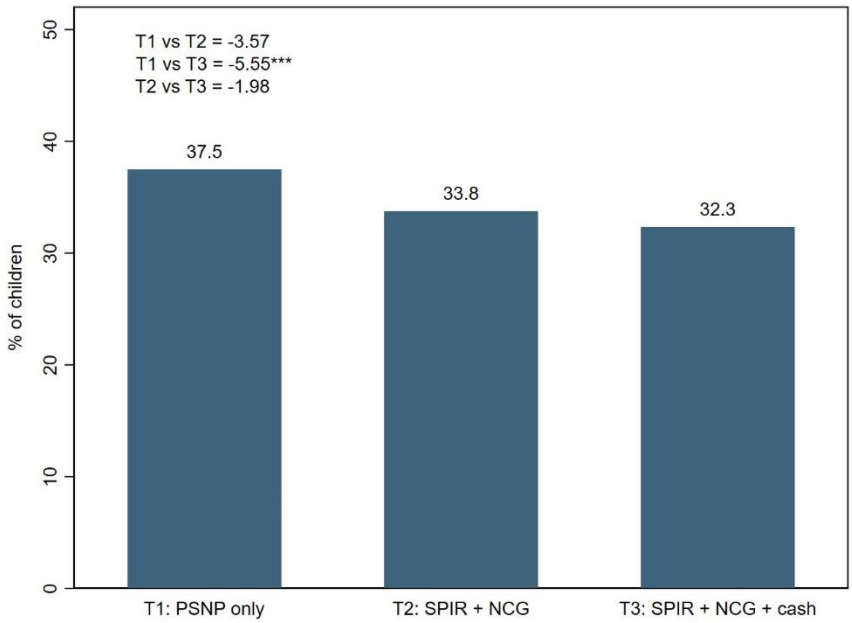
Note: N = 2,221 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

6.3 Child diarrhea risk

The midline survey asked caregivers whether the index child had diarrhea in the 2 weeks prior to the interview. The responses to this question are summarized in Figure 6.6 below.[11] Nearly 38 percent of the children in the T1 arm and 34 percent in the T2 arm experienced diarrhea during the reference period. The lowest incidence, at 32 percent, was observed among children in the T3 arm. The difference in diarrhea prevalence between the T1 and T3 arms is highly statistically significant, with a p-value of less than 0.01. The difference in diarrhea prevalence between the T1 and T2 arms, having a p-value of 0.11, is not statistically significant at conventional levels.

[11] This indicator is not among the primary or secondary outcomes tracked in this study but we report it here given its association with child undernutrition in this and similar contexts (Hirvonen et al. 2021).

Figure 6.6. Prevalence of diarrhea among children 6-23 months, by study arm



Note: N = 2,233 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

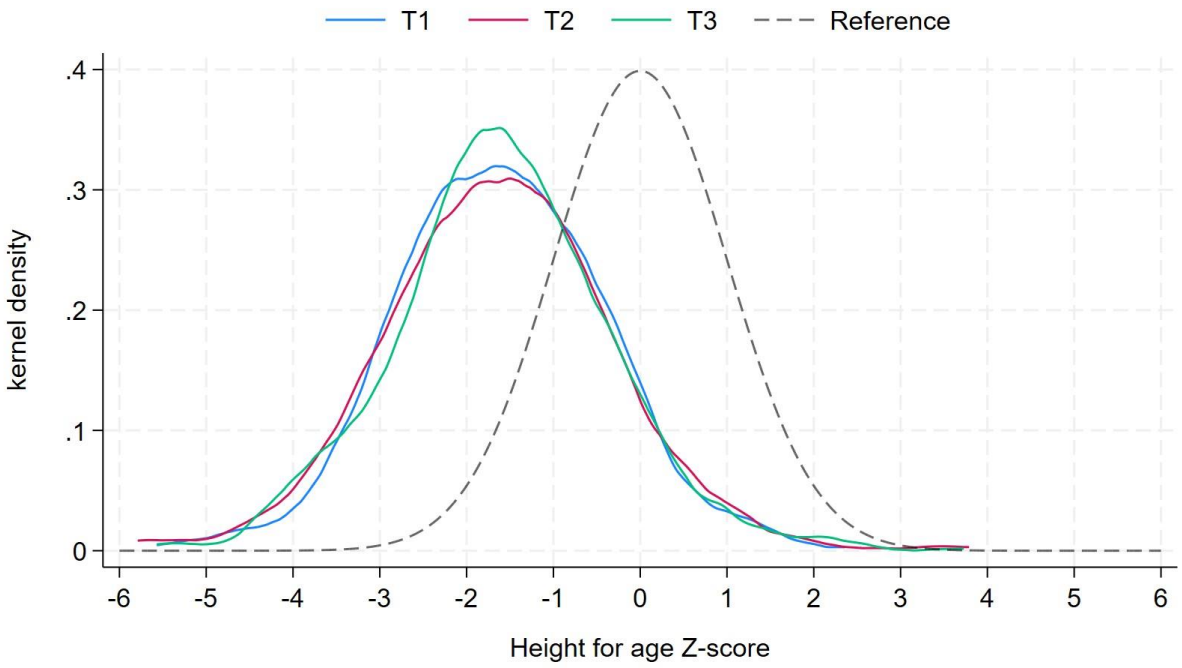
6.4 Child anthropometry

Next, we move onto anthropometric outcomes that were collected at midline, but not at baseline (when many of the sample children were still in utero). Lengths/heights and weights were converted to z-scores using the 2006 WHO growth standards (WHO 2006, de Onis et al. 2007). These standards allow us to assess child length and weight relative to well-nourished children of the same age and sex, and a Z-score expresses these measures in terms of standard deviations. Children are considered to be stunted (chronically undernourished) if they have a height for age z-score (HAZ) below -2.0, and this chronic undernutrition reflects the negative effects of continued inadequate food intake together with repeated infection. In a healthy and well-nourished population less than 2.5 percent of the children have HAZ<-2.

Linear growth faltering is a long-term process and therefore we do not expect to see impacts on HAZ or stunting prevalence in this midline assessment. We begin by reporting full HAZ distributions for each treatment arm in Figure 6.7. All three distributions lie at the left-hand side of the HAZ distribution of the reference group, indicating widespread growth faltering among the children in our sample. The distributions mostly lie on top of each other, though the T3 distribution is slightly further to the right than the other two, suggesting slightly less growth faltering among children in T3 arm. Similarly, the mean HAZ scores are slightly lower in T3 (Figure 6.8), though the differences to the other arms are at best only weakly statistically significant ($p<0.10$). The unconditional treatment effect estimates reported

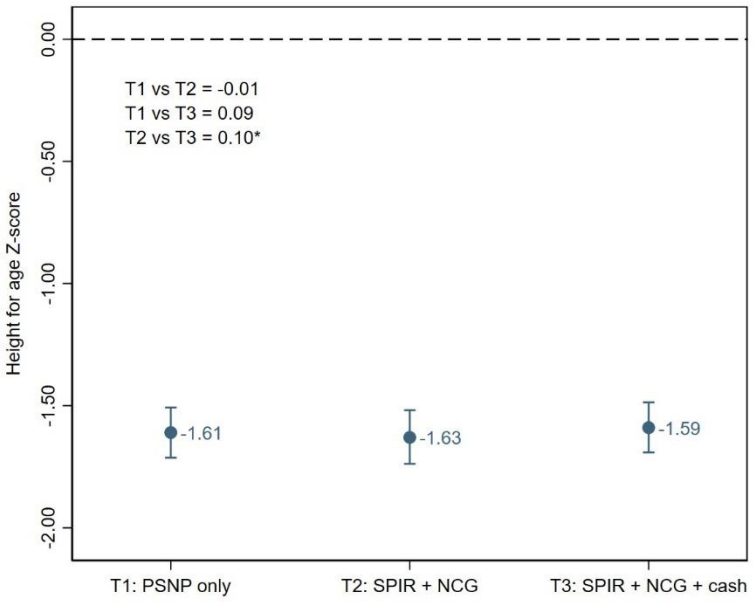
in Row 5b in Table A3 in the Appendix indicate a slightly smaller difference between T3 and T1 as well as T3 and T1 arms. As a result, all (unconditional) differences turn insignificant.

Figure 6.7. Height for age Z-score distributions of children 6-23 months, by study arm



Note: N = 2,226 children. 'Reference' is a normal distribution with a zero mean and standard deviation of one.

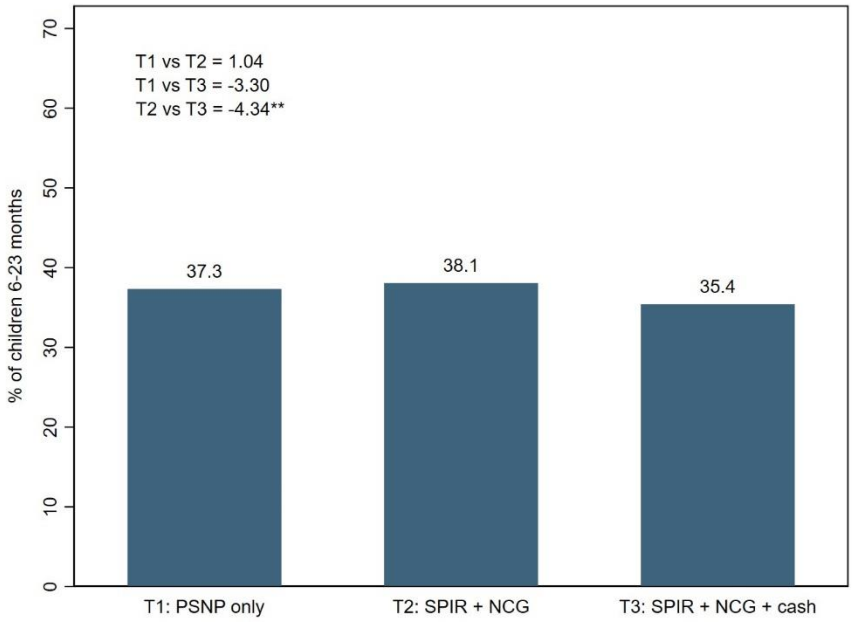
Figure 6.8. Mean height for age Z-score of children 6–23 months, by study arm



Note: N = 2,226 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 6.9 shows the stunting prevalences by study arm. The child stunting prevalence is 37 percent in T1 arm and 38 in T2 arm. Meanwhile, 35 percent of the children in T3 arm are classified as stunted. The conditional treatment effect estimates indicate that the difference between T1 and T3 arms is not statistically different from zero (p = 0.154). However, the difference between T2 and T3 arms is statistically different from zero (p < 0.05). The unconditional treatment effect estimates are slightly smaller and thus the differences across study arms are less likely to be significant.

Figure 6.9. Percent of children 6–23 months stunted (HAZ<-2), by study arm



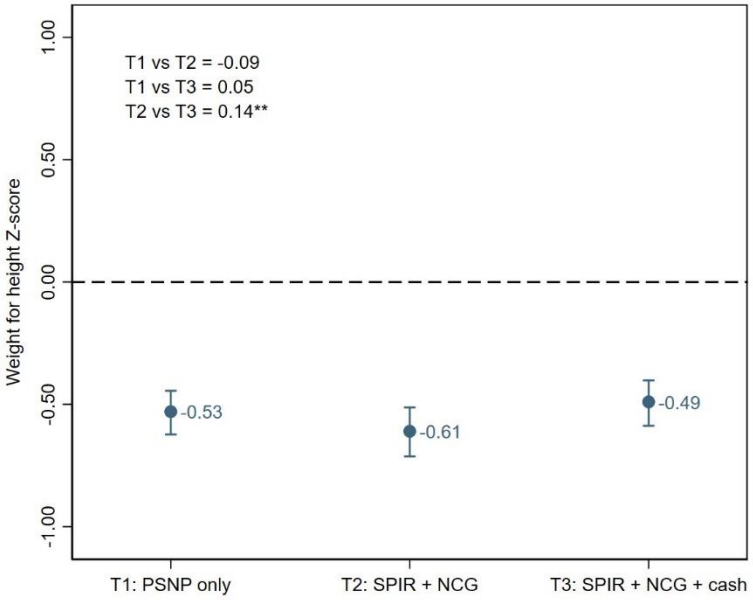
Note: N = 2,226 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Weight for height Z-score (WHZ) is an indicator reflecting short-term changes in children's nutritional status. Children are considered wasted if WHZ < -2. While SPIR may reduce the risk of child wasting, it is not the primary programmatic objective. In addition, our surveys are too infrequent to be able to adequately measure changes in acute malnutrition. Since WHZ and wasting prevalence are highly sensitive to seasonal fluctuations (Wells et al. 2019), more frequent surveys would be needed to maximize the ability to detect impacts on this outcome. With these caveats in mind, Figure 6.10 shows the mean WHZ scores for each study arm. The means in T1 and T2 are -0.53 and -0.61 and the difference between these two arms is not statistically significant. The mean WHZ in T3 is -0.49 and when compared to the mean recorded for T2, the conditional treatment effect estimate indicates that the difference is statistically significant (p<0.05). The unconditional treatment effect estimates reported in Row 7a in Table A3 in the Appendix are similar, though slightly smaller in terms of magnitude with larger standard errors.

In line with previous anthropometric surveys with PSNP households (Berhane et al. 2020), between 6 and 8 percent of the children are categorized as wasted (Figure 6.10). The differences in wasting

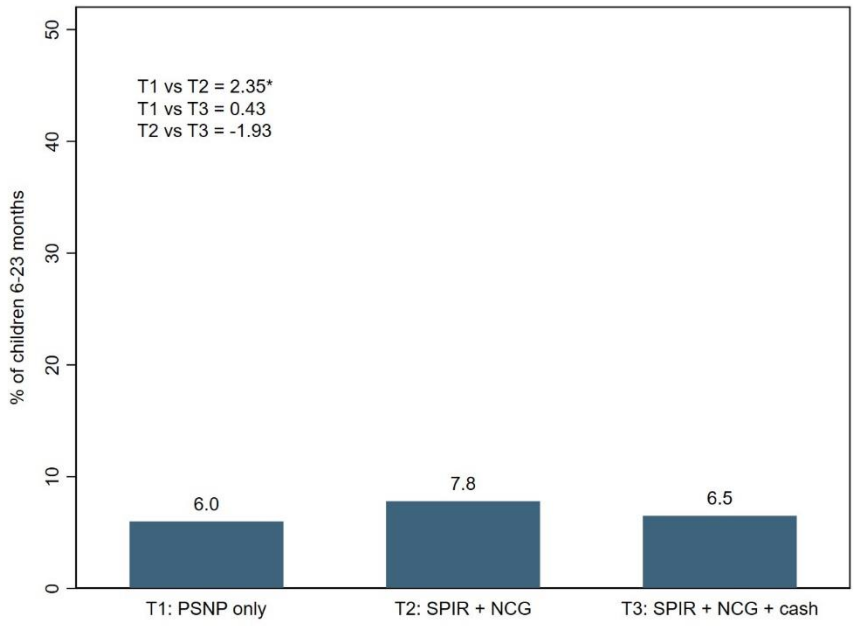
prevalence between the study arm are not statistically significant (Figure 6.11), except for T1 vs T2 that is weakly statistically significant (p = 0.063). The unconditional treatment effect estimates are similar to the conditional estimates (Row 8a in Table A3 in the Appendix).

Figure 6.10. Mean weight for height Z-score of children 6–23 months, by study arm



Note: N = 2,226 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 6.11. Percent of children 6–23 months wasted (WHZ<-2), by study arm



Note: N = 2,226 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

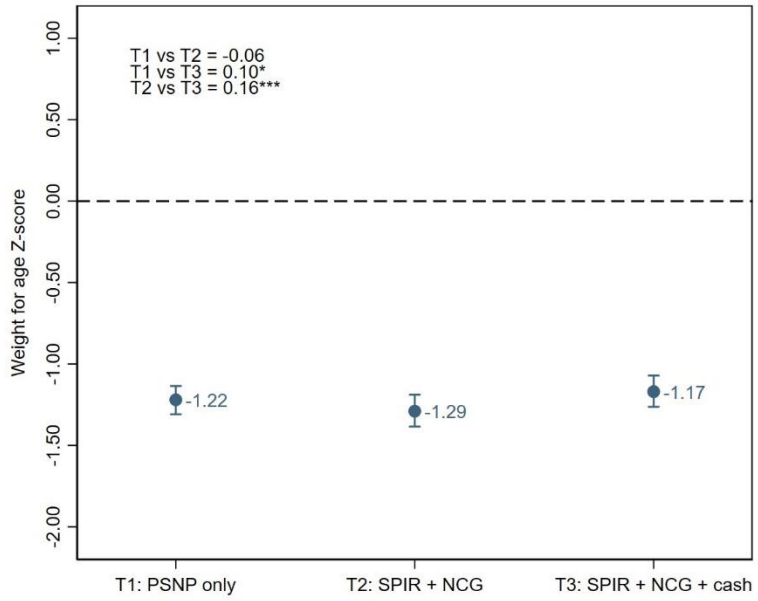
We conclude this subsection by discussing weight-for-age, an amalgam of HAZ and WHZ. A child is considered underweight if their weight for age Z-score (WAZ) falls below -2.0. While WAZ (or underweight status) is not a primary or secondary indicator tracked in our study, we include it in our report because it is commonly reported alongside HAZ (stunting) and WHZ (wasting) indicators.

Figure 6.12 presents the mean WAZ at midline for the three study arms. The average WAZ in the T1 arm is -1.22, compared to -1.29 in the T2 arm, with the difference not being statistically significant. In the T3 arm, the mean WAZ is somewhat higher at -1.17. According to the conditional treatment estimator, the differences between T3 and the T1 and T2 arms are statistically significant with $p < 0.10$ and $p < 0.01$, respectively.

With 25 percent of the children in the T2 arm characterized by a WAZ less than -2, the prevalence of underweight is highest in this group (Figure 6.13). In the other two arms, the prevalence of underweight is approximately 22 percent. The differences in underweight prevalence between the T2 arm and the other two arms are statistically significant, at least at the 10-percent level.

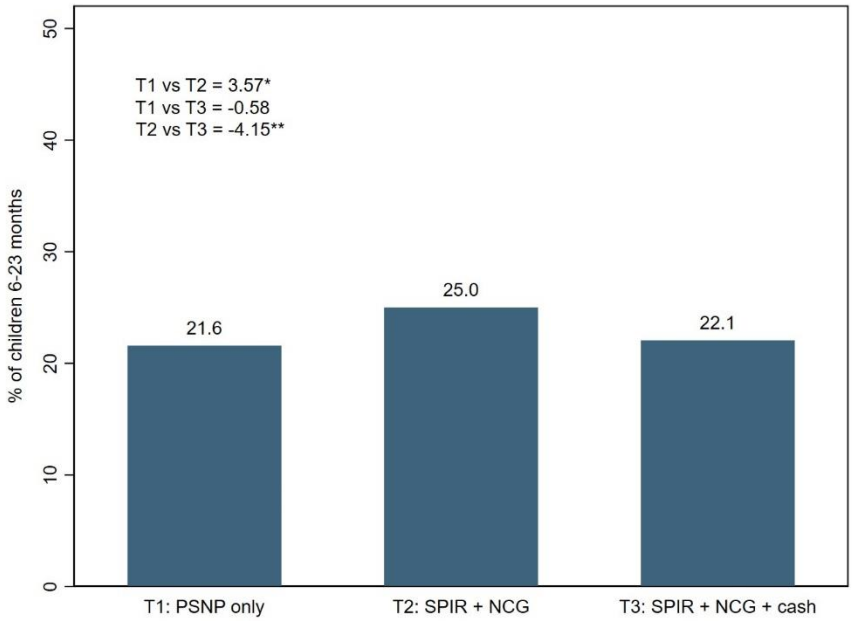
It is important to note that the differences between T1 and T2 with respect to child anthropometrics are consistently extremely small and generally statistically insignificant; the only outcome for which the difference between child outcomes in T2 and the control arm is significant is wasting, and only at the ten percent level. (Differences between T2 and T3, when observed, also reflect the additional positive effects of the cash transfers.) Though some indicators appear slightly worse in T2 when compared to the control arm, this is most plausibly interpreted as statistical noise, and is not suggestive of any meaningful difference across arms.

Figure 6.12. Mean weight for age Z-score of children 6–23 months, by study arm



Note: N = 2,226 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 6.13. Percent of children 6–23 months underweight (WAZ<-2), by study arm



Note: N = 2,226 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

6.5 Early childhood development

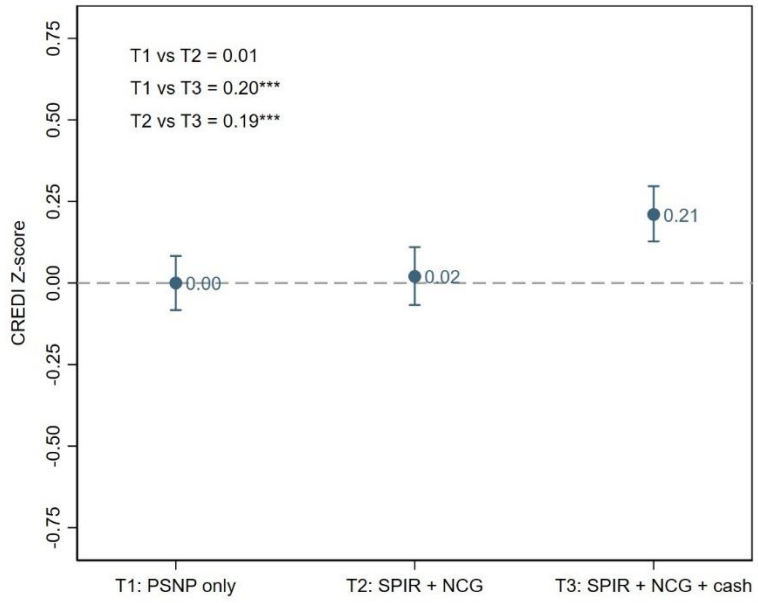
We assessed early childhood development (ECD) at midline using the short form of the Caregiver-Reported Early Development Instruments (CREDI) (McCoy et al. 2018). CREDI were designed to serve as a population-level measure of ECD for children from birth to age three. The instruments rely exclusively on caregiver reports (typically a parent of the child), and primarily focus on milestones and behaviors that are easy for caregivers to understand, observe, and describe using a yes/no response scale. There are two versions of CREDI. The short form (employed in the present study) creates a summary score for children's overall developmental status, whereas the long form creates domain-specific developmental scores. The short form includes 20 items specific to each six-month age group (0–5 months, 6–11 months, 12–17 months, 18–23 months, 24–29 months, and 30–35 months), and the administration time is on average five minutes.[12]

We used the CREDI Scoring Application to generate raw scores and the norm-referenced Z-scores for overall development for each age band. However, we chose to internally standardize (vis-à-vis the mean for age group in the control arm), given that we are interested in the treatment effect; in addition, it is not clear whether the international reference group is an appropriate comparison for our sample.

[12] Importantly, the CREDI cannot be used to generate cutoffs to identify children who are on track developmentally, and is not effective for screening any particular child for developmental delay (see this document: https://credi.gse.harvard.edu/frequently-asked-questions). Those goals require more detailed instruments that are typically administered by more specialized personnel with expertise in child development.

Figure 6.14 presents the mean CREDI Z-score for the three study arms. By construction, the mean Z-score in the control (T1) arm is zero. The CREDI Z-score in T2 arm is 0.02 SD and the difference to T1 arm is not statistically significant. The mean CREDI Z-score in T3 arm is 0.21 SD and the differences compared to the T1 and T2 arms are statistically highly significant ($p < 0.01$). These effects are large relative to several other recent papers that have analyzed the effects of cash transfers and a parental stimulation intervention in Niger, and childcare in Malawi and Burkina Faso, and found effects in the range of .1 - .15 standard deviations (Premand and Barry 2022, Özler et al. 2018, Ajayi, Dao, and Koussoubé 2022).

Figure 6.14. Mean CREDI Z-score of children 0 – 23 months, by study arm



Note: N = 2,349 children. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

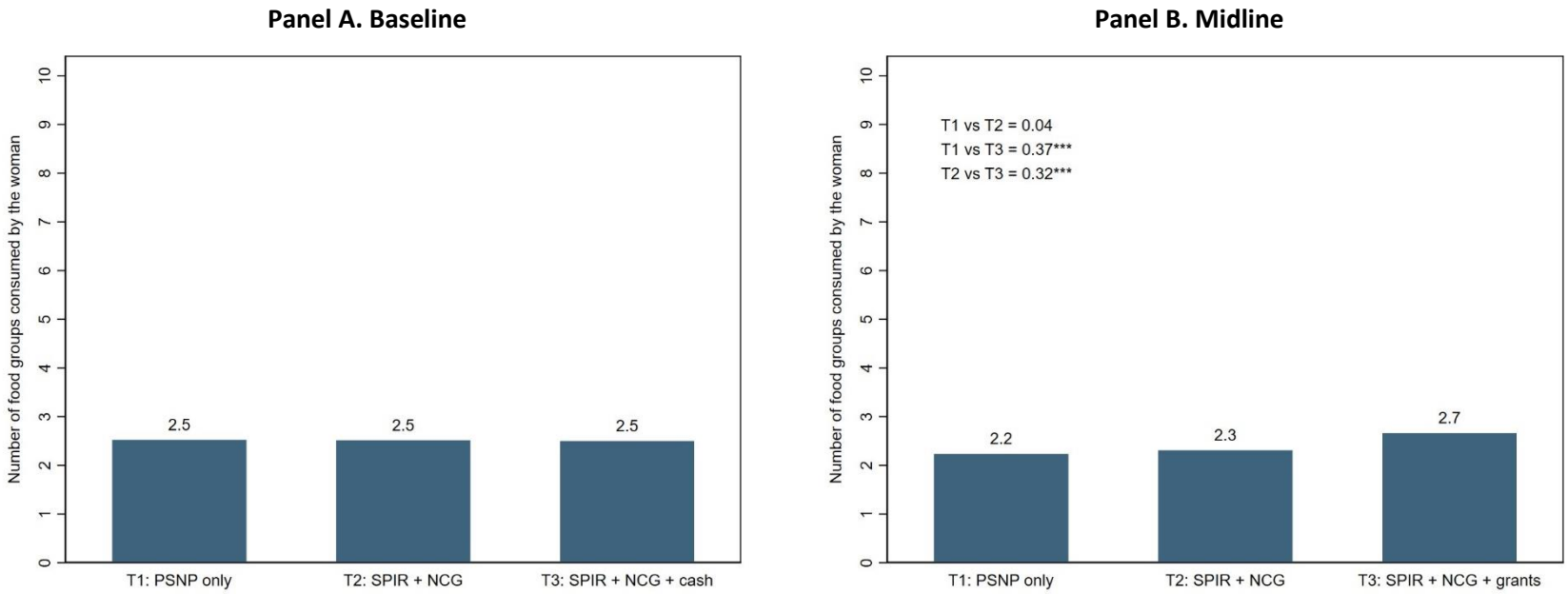
6.6 Women's dietary diversity

Both baseline and midline surveys included a module capturing the number of food groups consumed over the previous 24 hours by the female caregiver of the index child. We used these data to construct the MDD-W indicator: the prevalence of women of reproductive age (15-49 years) consuming a diet of minimum diversity. A woman is identified as consuming a diet of minimum required diversity if she consumed five or more food groups from a total of 10. [13] MDD-W is not among the primary or secondary outcome indicators in this evaluation, but we will briefly report this indicator since these data were collected in the surveys and there is considerable interest in this outcome among the policymakers and other stakeholders in Ethiopia.

[13] The food groups are 1. grains, white roots, tubers, and plantains; 2. pulses such as beans, peas, and lentils; 3. nuts and seeds, including groundnut; 4. dairy; 5. meat, poultry, and fish; 6. eggs; 7. dark green leafy vegetables; 8. other vitamin A-rich fruits and vegetables; 9. other vegetables; and 10. other fruits.

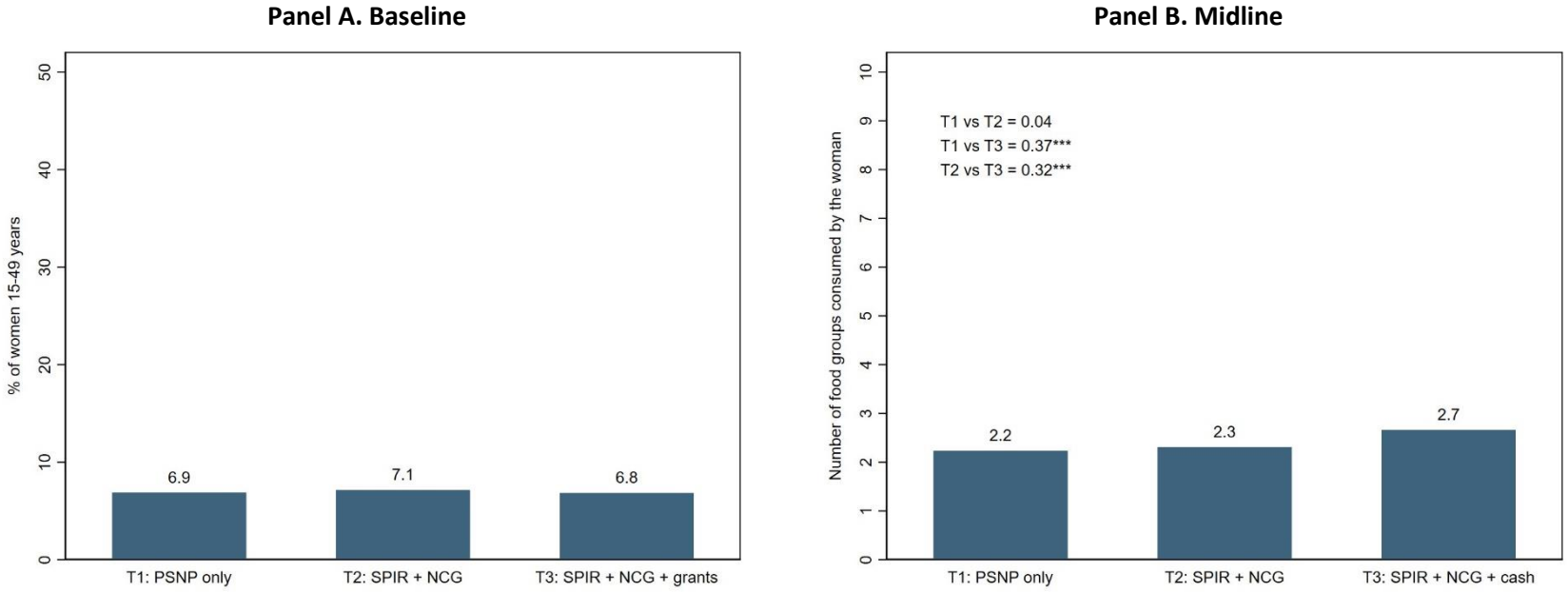
At baseline, the average female caregiver in our sample consumed from 2.5 food with negligible differences across the study arms groups (Panel A in Figure 6.15). By midline, the mean number of food groups is 2.2 and 2.3 in T1 and T2 arms, respectively (Panel B in Figure 6.15). Women in the T3 arm consumed from 2.7 food groups, and the difference compared to both the T1 and T2 arms is statistically highly significant ($p < 0.01$). At baseline, 7 percent of the women achieved MDD-W with limited differences across the three study arms (Panel A in Figure 6.16). By midline, this share has fallen to 2 percent in T1 and T2 arms while in T3 arms, 6 percent of the women achieved MDD-W. This difference between T3 and the other two arms is again highly significant ($p < 0.01$) (Panel B in Figure 6.16).

Figure 6.15. Diet diversity of female caregivers, by study arm and survey round



Note: N = 2,231 women in reproductive age (15-49 years). ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

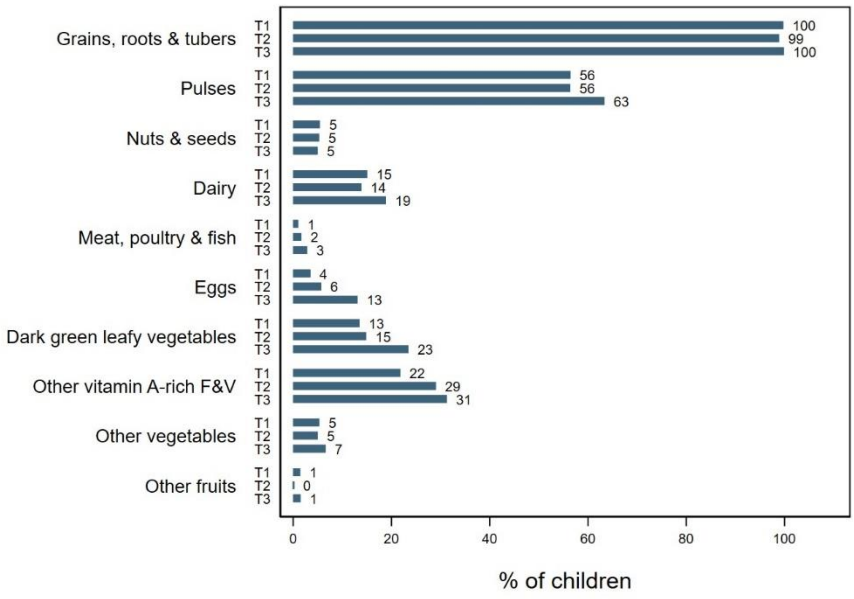
Figure 6.16. Percent of female caregivers consuming a diet of minimum diversity (MDD-W), by study arm and survey round



Note: N = 2,231 women in reproductive age (15-49 years). ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 6.17 provides a summary of the dietary patterns among the female caregivers. Virtually all women consumed grains, roots, and tubers. The other key food groups in are pulses (consumed by more than half of the women) and vitamin A-rich fruit and vegetables, other than dark green leafy vegetables (consumed by more than 20 percent of the women). Women in T3 arm were more likely to consume pulses, dairy, eggs and dark green leafy vegetables than women in T1 and T2 arms.

Figure 6.17. Percent of women consuming from different food groups, by study arm



Note: N = 2,231 women in reproductive age (15-49 years). F&V = Fruits and vegetables.

7. PRIMARY FINDINGS: LIVELIHOOD OUTCOMES

Core SPIR livelihood programming includes the organization of VESAs, used as a platform for trainings and other project 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 (30 percent of IMPEL households) is also targeted for a one-time $300 livelihood grant; these are the households identified as the poorest using an asset-based welfare index constructed by the IFPRI team from baseline survey data.

We do not expect to see meaningful impacts on livelihood outcomes within the first year of implementation. Consequently, the impact of the livelihood interventions will primarily be evaluated at endline, in 2025. Considering this, the midline survey instrument collected data on limited set livelihood indicators, primarily household food security indicators: the Food Insecurity Experience Scale (FIES), Food Gap, Food Consumption Score (FCS), and Household Diet Diversity Score (HDDS) as well as household consumption that permit us to compute per capital consumption and poverty related outcomes. Here we report the treatment effect estimates on FIES, FCS, household per capita consumption, extreme poverty, and depth of poverty, the five pre-specified livelihood outcomes to be reported at midline (see Table A2 in the Appendix). It was also pre-specified that we analyze the impact on all livelihood outcomes by pooling the two treatment arms: T2 and T3. We follow this plan here in this section, but in Section 8.1 below, we present the treatment effects for key livelihood outcomes separately for T2 and T3.

The comparisons between baseline and midline are based on households that were interviewed in both rounds. It is worth noting that this specific sample restriction may account for variations in baseline values when compared to the corresponding values reported in the baseline report.

7.1 Food security indicators

We begin by reporting on food security indicators: FIES, FCS and the food gap. As we noted in the baseline report, it is important to keep in mind that these indicators capture different dimensions of food (in)security and therefore, they are not always in agreement with each other. Maxwell, Vaitla, and Coates (2014) show how the food insecurity prevalence estimates in Ethiopia substantially vary depending on which food insecurity indicator is used. For example, the FIES tends to predict considerably higher food insecurity than the FCS (Maxwell, Vaitla, and Coates 2014).

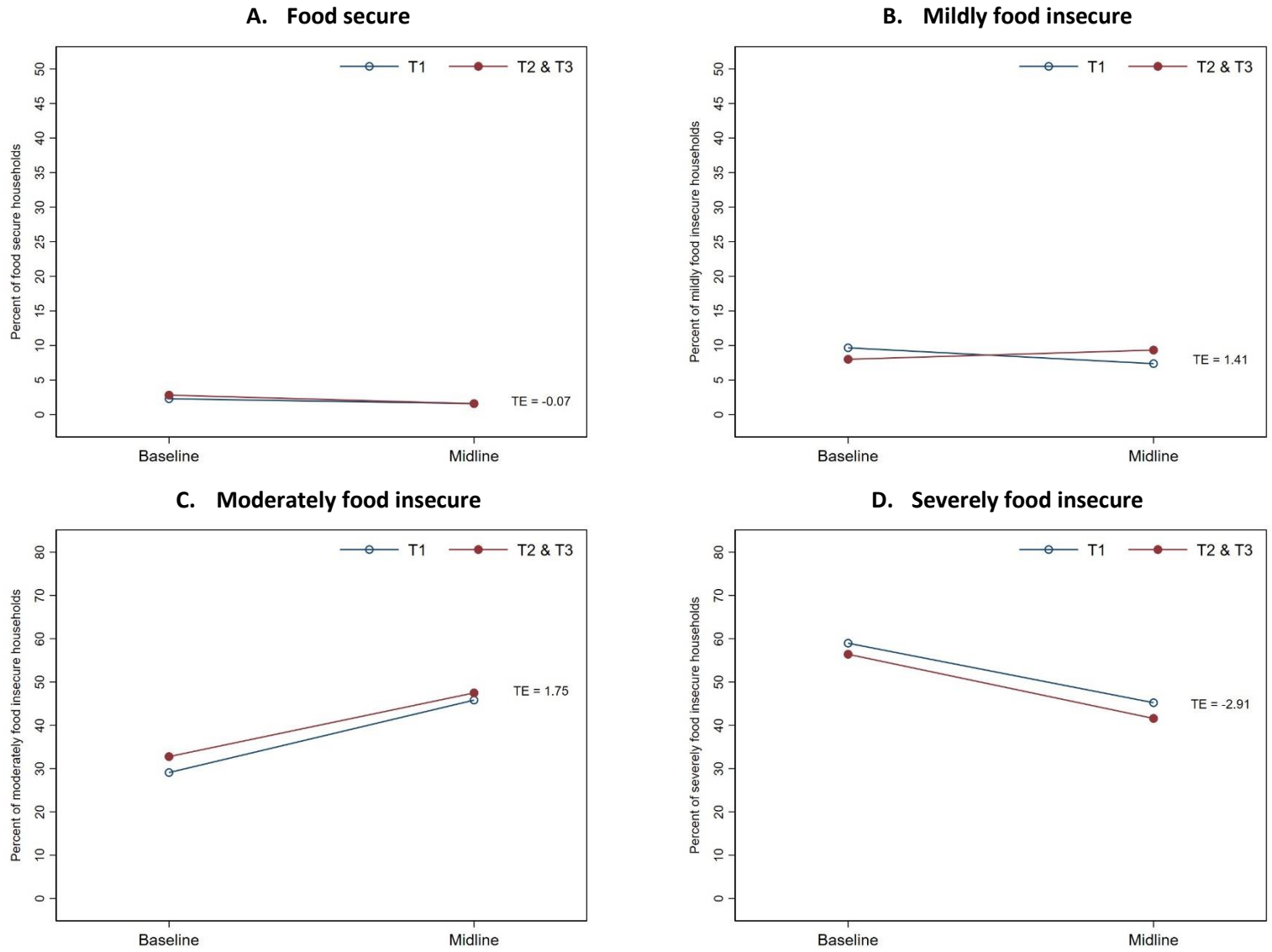
7.1.1 Food Insecurity Experience Scale (FIES)

Developed by the FAO, the FIES is a subjective household food insecurity indicator capturing households' perceived food insecurity situation (Ballard, Kepple, and Cafiero 2013). The FIES survey module administered at baseline and endline contained eight 'Yes/No' questions about the household's food security situation. 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.

Using the raw FIES scores, we then 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 7.1 shows the share of households in each FIES category by survey round and study arm. A considerably smaller share of households is categorized as food secure or mildly food insecure in each round. There is also less movement into or out of these two categories. Compared to the baseline, fewer households are categorized as being severely food insecure at midline. It seems that many households in this category have shifted to being moderately food insecure, a group that has increased since the baseline. All conditional treatment effect estimates are small in magnitude and statistically not different from zero.

Figure 7.1. FIES food security status, by treatment status and survey round



Note: N = 2,954 households. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

7.1.2 Food Consumption Score (FCS)

Next, we report on the Food Consumption Score (FCS), an indicator developed by the World Food Program (WFP).[14] 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. 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.

As also found by previous work, the FCS predicts considerably lower food insecurity prevalence than the FIES (Maxwell, Vaitla, and Coates 2014).[15] In both survey rounds, the mean FCS in the sample is comfortably above 35, the 'acceptable' threshold (Figure 7.3, Panel A). By midline, we see that the mean FCS improved in both control (T1) and treated (T2 and T3) groups. However, the improvement was larger among the treated households. We estimate an unconditional treatment effect of 4.3 FCS units, showing that SPIR treatments led to an improvement in food security as measured by FCS. Compared to the baseline mean in T1 households, this treatment effect represents approximately a 10 percent increase in FCS.

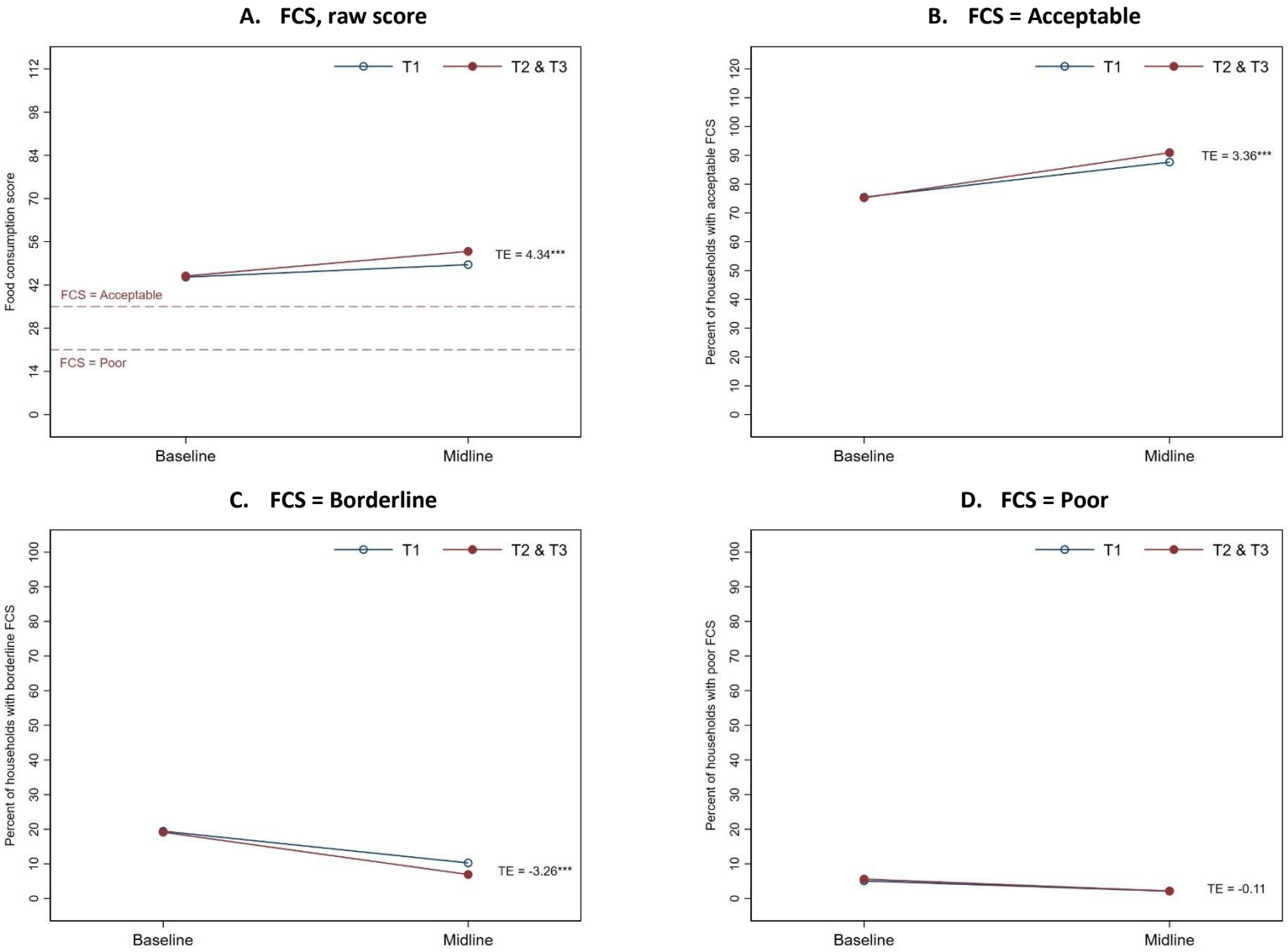
Regarding FCS, we observe that the positive effect reported in Panel A is primarily attributed to households in the T3 arm that received maternal cash transfers. The treatment effect estimate for T2 is nearly zero and not statistically significant, while for T3, it is approximately 8 FCS units and highly significant.

At midline, nearly 90 percent of the households report acceptable FCS (Figure 7.3, Panel B). For both groups, the share of households with acceptable FCS has increased. At the same time, the share of households with a borderline FCS has decreased. However, as indicated by the positive (negative) and statistically significant treatment effect estimate for acceptable (borderline) FCS, the increase (decrease) has been slightly larger among households exposed to the SPIR treatments. The unconditional treatment effect estimates are similar to the conditional ones (Rows 4 to 7 in Table A4 in the Appendix).

[14] 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 seven 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.

[15] One possible reason for this discrepancy is the thresholds that categorize households with borderline or acceptable food consumption may have been set too low (Lovon and 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).

Figure 7.2. Food consumption score (FCS), by treatment status and survey round

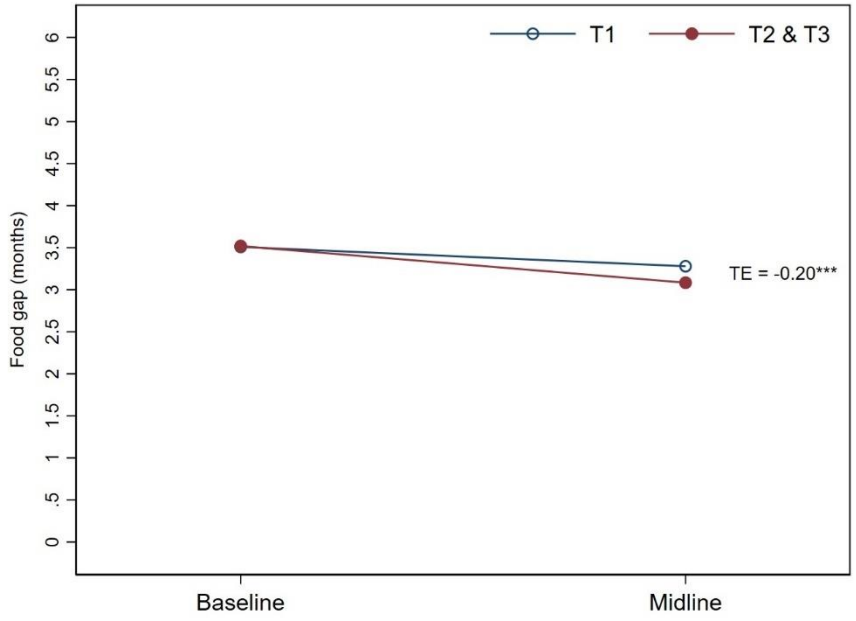


Note: N = 2,954 households. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

7.1.3 Food gap

Both baseline and midline survey instruments asked households about the number of months that the household was unable to satisfy its food needs in the past 12 months. This 'food gap' 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 food gap is not among the primary or secondary outcome indicators in this evaluation, but we will briefly report it here due to its central role in PSNP evaluations. At baseline, the average (mean) food gap among the control households (T1) was 3.5 months, and it had reduced to 3.3 months at midline (Figure 7.3). This difference is statistically significant. For the treated households (T2 and T3), the baseline average food gap was 3.5 months, reducing to 3.1 months at midline, also a statistically significant difference. The ANCOVA estimator with controls suggests a treatment effect of -0.20 (statistically significant at the one percent level), demonstrating that SPIR treatments marginally decreased the food gap by approximately 0.2 months.

Figure 7.3. Food gap by, by treatment status and survey round



Note: N = 2,954 households. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

7.2 Household consumption and poverty

Next, we report on daily per-capita consumption-expenditure (BL40), prevalence of poverty (BL01) and the depth of poverty of the poor (BL02). Constructing these indicators also requires data on household consumption of food and non-food goods, which were collected both at the baseline and midline surveys. Detailed overview of household consumption and poverty measurement methods were provided in Appendix A of the baseline report (Gilligan et al. 2023).

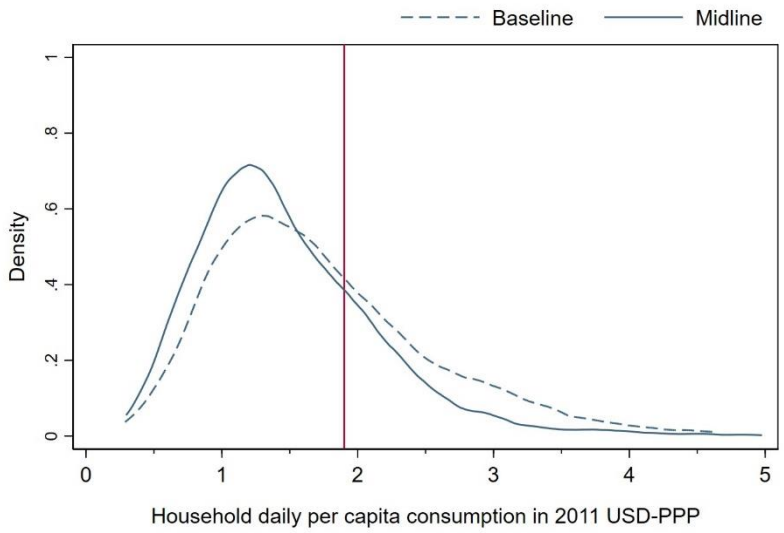
To account for inflation between the baseline and midline surveys, we utilized food price data collected in both survey rounds to construct an Ideal Fisher Index for measuring changes in the prices of representative food baskets in each woreda (see, for example, Deaton and Tarozzi 2000). According to this price index, the average food inflation between the baseline and midline survey in the study area stood at 33 percent, which is in line with the national food inflation estimated by the Ethiopian Statistical Service for the same period (ESS 2023). We applied this food price index to adjust the midline consumption values to baseline (Autumn 2023) price levels.[16] After this, we followed the same approach outlined in the baseline report to convert the consumption values into 2011 purchasing power parity (PPP) terms.

Figure 7.4 overlays the distribution of baseline household daily per capita consumption (dashed line) on the 1.90 $PPP poverty line for both groups. We see that many individuals are located just below or just above the poverty line. Consequently, even relatively small changes in average consumption levels are likely to induce shifts in poverty headcounts. By midline, the distribution for both groups (solid line) has shifted left indicating a reduction in average consumption levels and increase in the number of people falling below the poverty line.

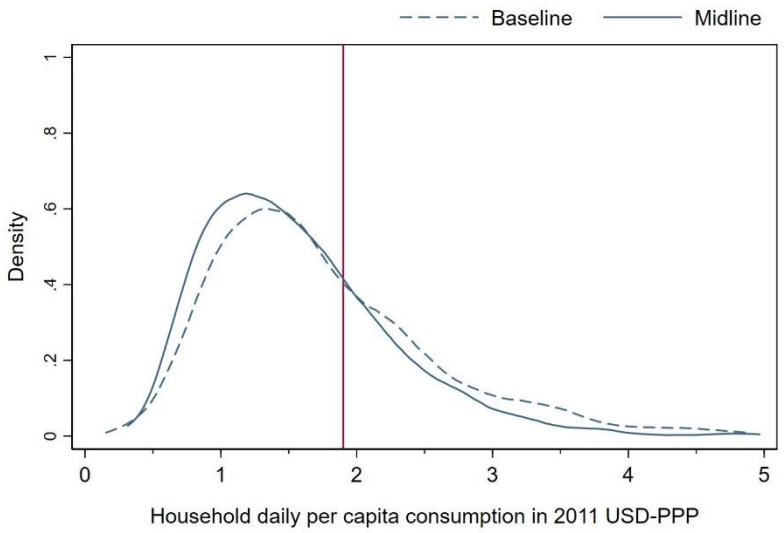
[16] Using a food price index to adjust total consumption is warranted because food represents a large portion of overall consumption in this context. At baseline, the average household in the sample allocated 86 percent of its total expenditure to food.

Figure 7.4. Household daily per capita consumption, by treatment status and survey round

A. T1 households



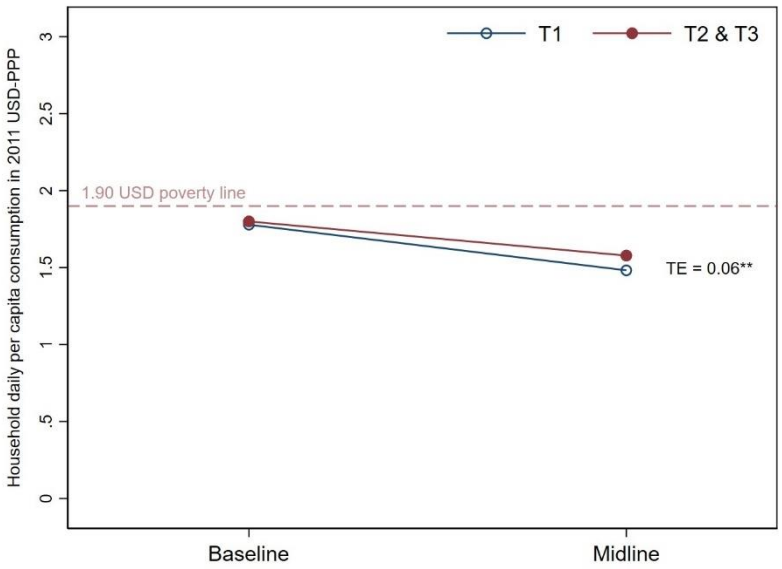
B. T2 AND T3 households



Note: Panel A = 1,004 households; Panel B = 1,950 households. Vertical red line marks the $1.90 poverty line.

The reduction in average consumption levels is confirmed in Figure 7.5 that compares the mean daily per capita consumption levels by round and treatment status. In T1 households, the mean consumption fell from $1.78 to $1.48 while in T2 AND T3 households it fell from $1.80 to $1.58. The treatment effect based on an ANCOVA estimator with controls is positive and statistically significant, but small in terms of magnitude. The unconditional treatment effect estimate is similar (Row 8a in Table A4 in the Appendix).

Figure 7.5. Household daily per capita consumption in 2011 USD-PPP, by treatment status and survey round

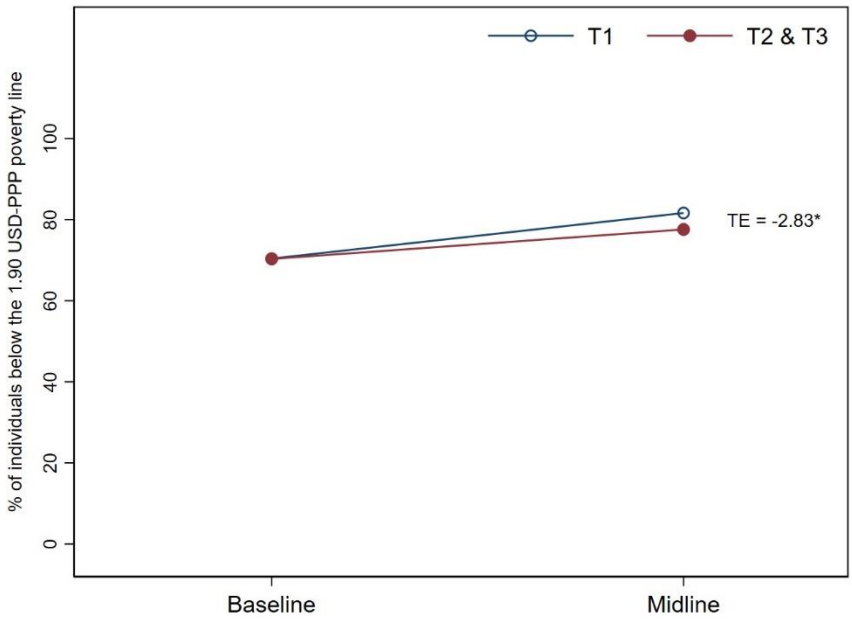


N = 2,954 households. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

This decline in mean consumption levels between baseline and midline resulted in an increase in poverty headcount (Figure 7.6). The share of individuals whose household per capita consumption level was below the $1.90 poverty line increased from 70 percent to 82 percent in T1 households and from 70 percent to 78 percent in T2 and T3 households.[17] A regional breakdown of the poverty numbers reveals that poverty headcount increased more in the Oromia sample (from 68.2% to 79.0%) than in the Amhara sample (from 77.5% to 79.2%). The treatment effect based on an ANCOVA estimator with controls is negative indicating that individuals in SPIR households have slightly lower poverty rate at midline with the difference being statistically significant at the ten percent level. The unconditional estimate reported on Row 9b in Table A4 is similar in terms of magnitude and also significant at the ten percent level.

[17] Note that the unit of analysis in poverty headcount and depth of poverty indicators is an individual, not household. Therefore, the means, standard deviations and treatment effects have been estimated using household size as the frequency weight. This also explains the larger reported number of observations in these analyses.

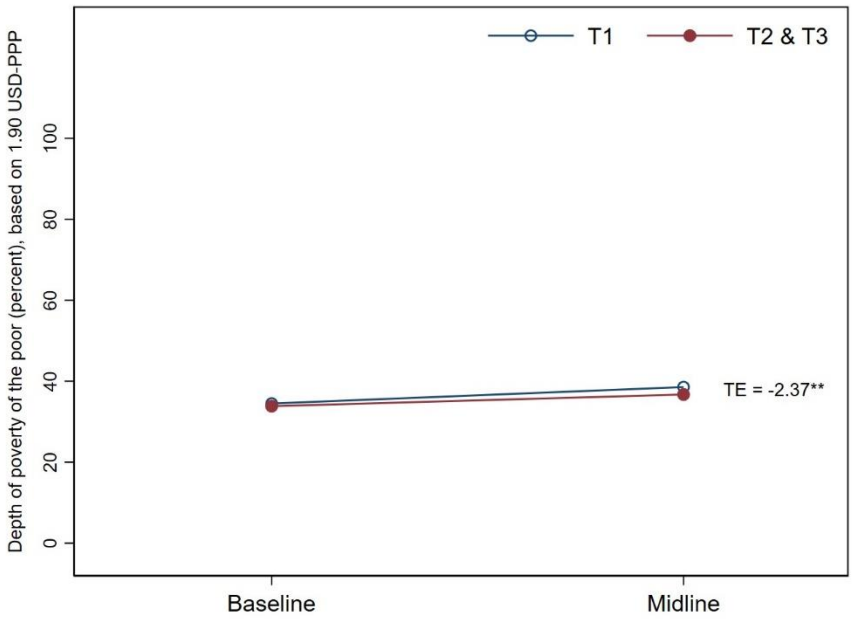
Figure 7.6. Poverty headcount based on $1.90 poverty line, by treatment status and survey round



Note: N = 17,041 individuals from 2,954 households. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

The poverty gap index tells us about the depth of poverty: the degree to which the per capita consumption of the poor falls below the poverty line. At baseline, the per capita consumption level of the average poor in households in the T1 group was 34.3 percent of the $1.90 poverty line and 33.3 percent in households in the T2 and T3 arms (Figure 7.7). At midline, these poverty gap indices have increased to 38.1 and 36.7 percent, respectively. The estimated treatment effect is negative suggesting that SPIR reduced the poverty gap by 2.4 percentage points. This treatment effect estimate is highly significant ($p < 0.01$). The unconditional estimate reported on row 10a in Table A4 in the Appendix is somewhat smaller (2.08) and not significant ($p = 0.125$).

Figure 7.7. Depth of poverty of the poor (percent) based on $1.90 poverty line, by treatment status and survey round



Note: N = 11,226 individuals from 1,921 households at baseline and 13,454 individuals from 2,204 households at midline. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

Consumption and poverty are measured per person, requiring a division of the household consumption estimate by the number of household members. Apart from changes in consumption patterns, changes in household size or its composition may therefore affect consumption and poverty estimates. This is particularly important here because a large share of the households welcomed a new baby between the baseline and midline surveys. To understand whether the declining consumption levels and rising poverty levels are explained by an increase in household size, we report consumption and poverty estimates based on adult equivalent units, adjusting consumption needs by age and sex.

Figure A7 in the Appendix reports the household size in terms of number of people (Panel A) and adult equivalent units (Panel B). The average household size at the baseline was 5.5 members if we use the number of members definition and 4.1 if we use the adult equivalent definition. The average household size has increased between baseline and midline by 0.4 member or by 0.2 adult equivalents. Irrespective of the definition, the changes in average household size are identical across study arms, indicating that any changes we observed in consumption or poverty indicators must be due to changes in consumed amounts.

Finally, through its impact on the denominator, the use of adult equivalent units will result in higher per capita consumption levels and lower poverty headcount rate than the standard 'number of members' approach. Figure A8 demonstrates this for per capita consumption. In contrast to Figure 7.5., the mean per capita consumption based on adult equivalent units is above the poverty line at both survey rounds and in both groups. Similarly, the poverty headcount rate is 20-25 percentage points lower if we used

adult equivalent units (Figure A9) versus number of members (Figure 7.6) to define *per capita* in per capita consumption. However, the trends are not affected by the definition of per capita: the mean consumption levels decreased, and poverty rate increased in both cases. Perhaps more importantly, the findings of the evaluation do not depend on the definition of *per capita*: the treatment effects remain statistically significant.

8. HETEROGENEITY ANALYSES

8.1 Livelihood grant eligibility

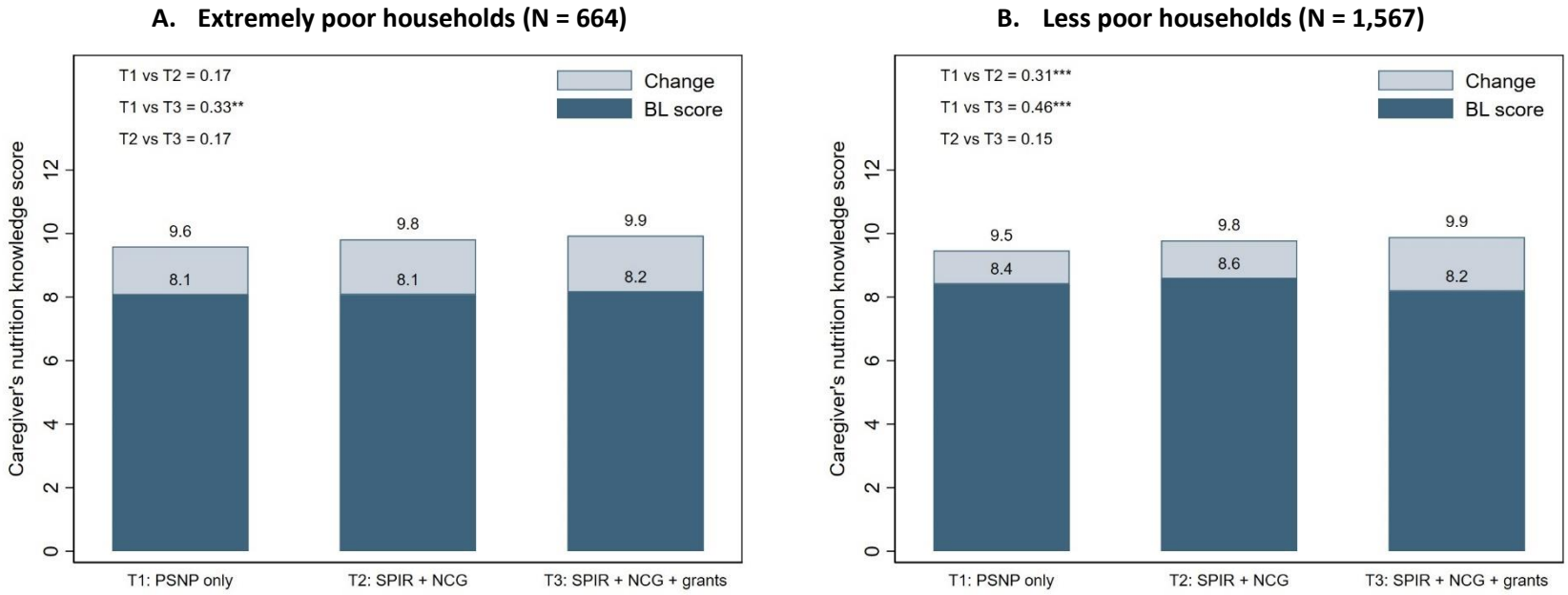
As previously mentioned, a subset of households (30 percent of IMPEL households) was selected to receive a one-time $300 livelihood grant. These households were identified as the poorest, based on an asset-based welfare index created by the IFPRI team using data from the baseline survey. This asset-based targeting process was implemented in all kebeles, including those assigned to T1, although the livelihood grants were not distributed to households in this arm. In this section, we explore heterogeneity with respect to this eligibility. To do so, we classify households eligible for the livelihood grant as 'extremely poor' and those not eligible as 'less poor'. We then conduct separate estimations for these two sub-groups to assess the extent to which the livelihood grant influences the results presented in Sections 6 and 7. To conserve space, our analysis of heterogeneity focuses only on the primary outcomes relevant to this midline assessment: Caregivers' IYCF knowledge, the prevalence of children aged 6–23 months consuming a diet of minimum diversity (MDD-C), and daily per-capita consumption-expenditure. It is important to note that the study may not have adequate statistical power to assess impacts on these sub-groups.

Figure 8.1 reports the results for caregivers' IYCF knowledge, with Panel A focusing on the 'extremely poor' sub-sample and Panel B on the 'less poor' sub-sample. At baseline, the average caregiver in 'less poor' households displays slightly better nutrition knowledge compared to their counterparts in 'extremely poor' households. However, by midline, the differences between the two sub-samples have become negligible. The estimated treatment effects are relatively similar across both sub-samples, although the difference between T1 and T2 arms is not significant for the 'extremely poor' sub-sample.

Figure 8.2 presents the results for MDD-C, with Panel A focusing on the 'extremely poor' sub-sample and Panel B on the 'less poor' sub-sample. A comparison across treatment arms reveals that children in 'less poor' households are more likely to meet MDD-C criteria than those in 'extremely poor' households. The treatment effect estimates are similar across both sub-samples.

Figure 8.3 reports on daily per-capita consumption-expenditure. At baseline, the mean consumption in households classified as 'extremely poor' based on their asset levels is somewhat lower than that of the 'less poor' households. However, this difference is small, underscoring the extremely high levels of poverty and deprivation in the entire sample of households. At midline, average consumption levels decrease in both sub-samples. The treatment effects are positive, but only statistically significant in the extreme poor sub-sample.

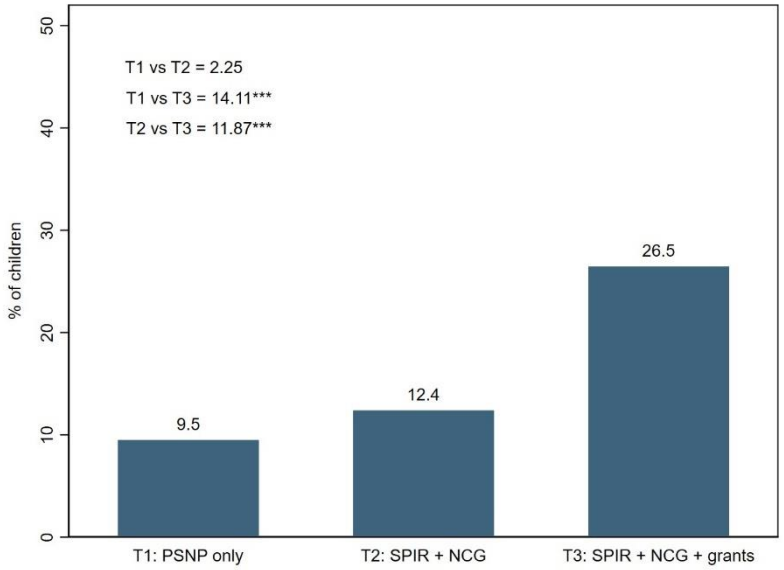
Figure 8.1. Caregiver's nutrition knowledge score, by study arm and asset-based poverty status at baseline



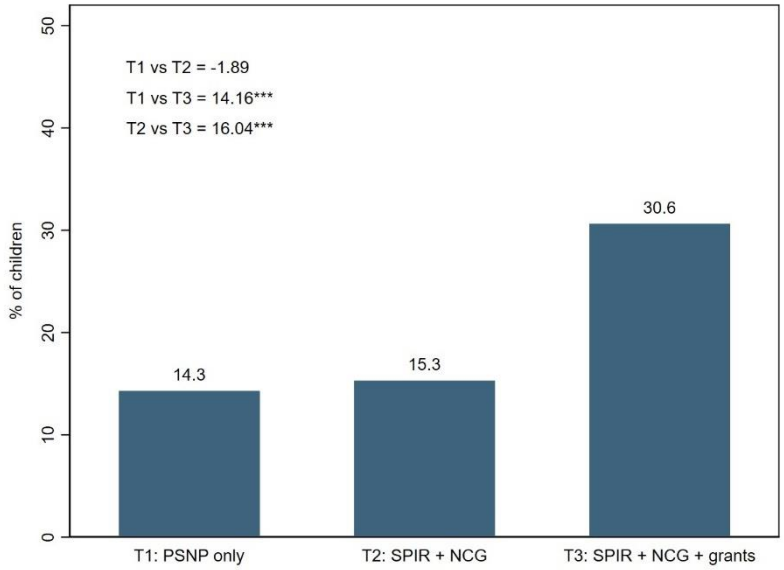
Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

Figure 8.2. % of children consuming a diet of minimum diversity (MDD-C), by study arm and asset-based poverty status at baseline

A. Extremely poor households (N = 665)



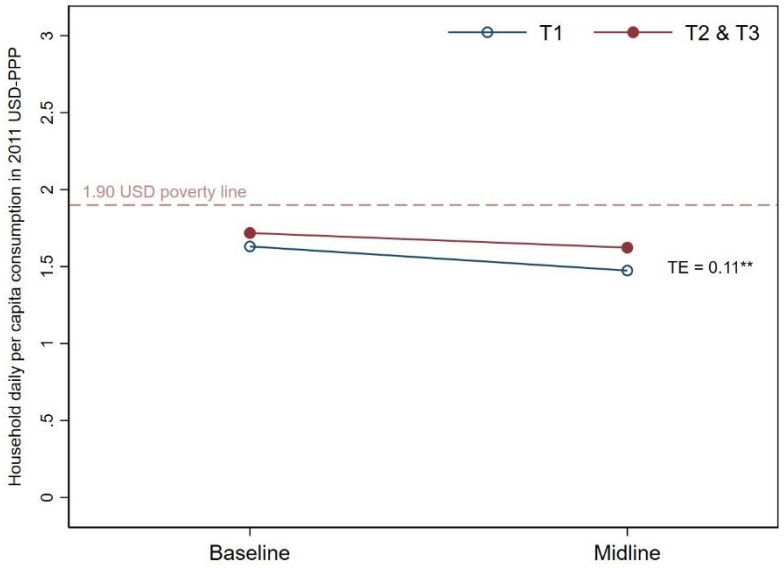
B. Less poor households (N = 1,568)



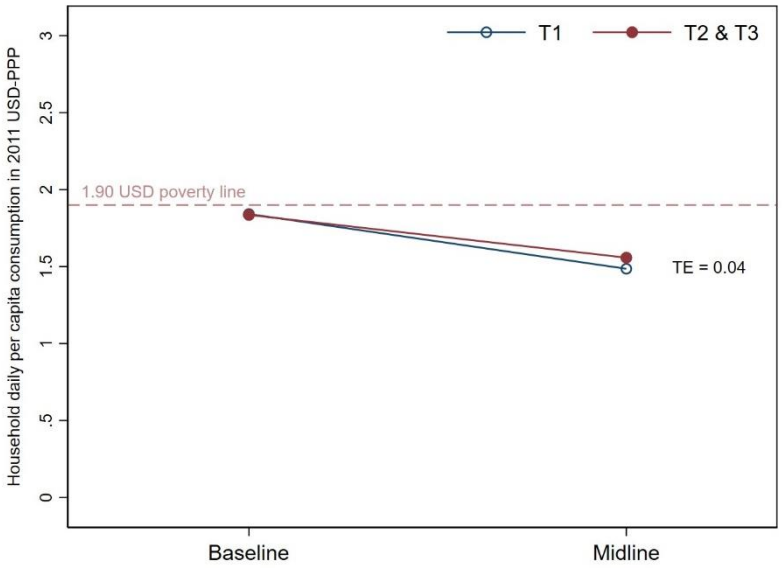
Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

Figure 8.3. Household daily per capita consumption in 2011 USD-PPP, by treatment status, survey round and asset-based poverty status at baseline

A. Extremely poor households (N = 897)



B. Less poor households (N= 2,057)



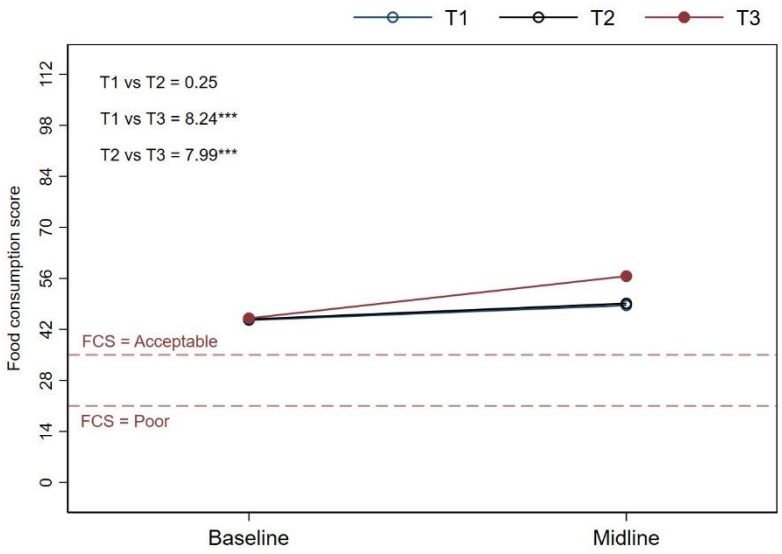
Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

8.2 Livelihood outcomes, by treatment arm

In Section 7, we followed our pre-analysis plan and reported the impact on the livelihood outcomes by pooling the two treatment arms: T2 and T3. However, it is possible that the maternal cash transfers received by the households in the T3 arm may have had a direct effect on the consumption and food security outcomes measured at the midline. To explore this, here we report the impact estimates separately for T2 and T3.

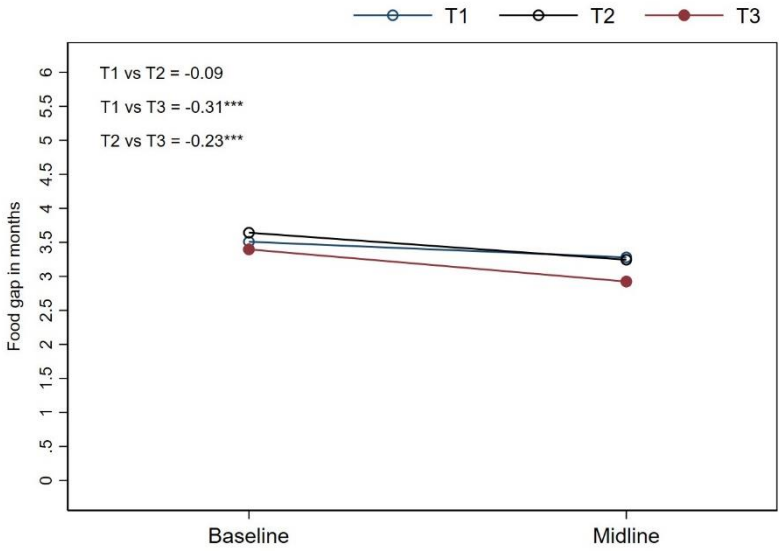
The unconditional and conditional impact estimates for all pre-specified livelihood outcomes are reported in Table A6 in the Appendix. Here we report the graphs and conditional impact estimates for selected livelihood outcomes: food consumption score (Figure 8.4), food gap (Figure 8.5), household daily per capita consumption (Figure 8.6) and poverty headcount (Figure 8.7). Across all these outcomes, we see that the T3 households perform better at midline as compared to the T1 and T2 households that did not receive maternal cash grants. These maternal cash transfers were approximately 20 USD per month while the average household size in our sample is 5.8 members. Therefore, a monthly transfer of 20 USD per household translates into 3.5 USD per member per month, or 0.11 USD per person per day. Meanwhile, the average household in T3 arm report 0.11 to 0.12 USD-PPPs higher per capita consumption compared to the average household in T1 and T2 arm, respectively. While we should be cautious when comparing USDs expressed in market exchange rates to USDs expressed in PPP exchange rates, the maternal cash transfers do seem to show up in the T3 consumption levels relative to the other two arms.

Figure 8.4. Food consumption score, by (un-pooled) treatment status, survey round



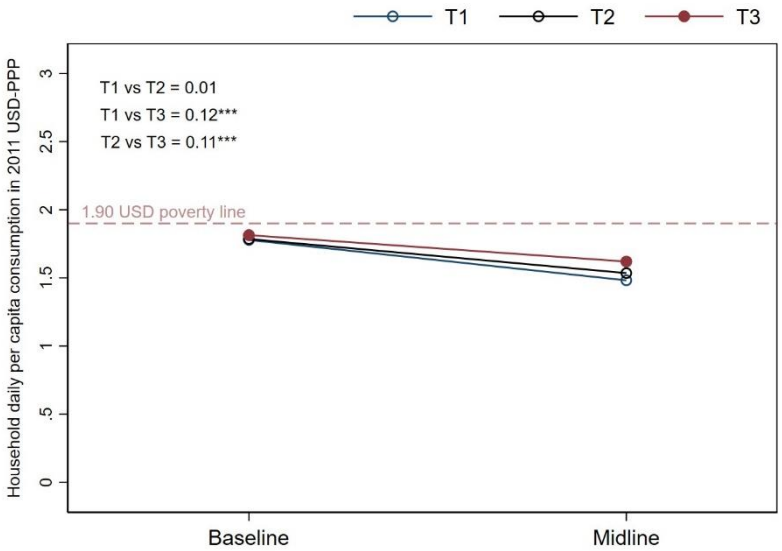
Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 8.5. Food gap, by (un-pooled) treatment status, survey round



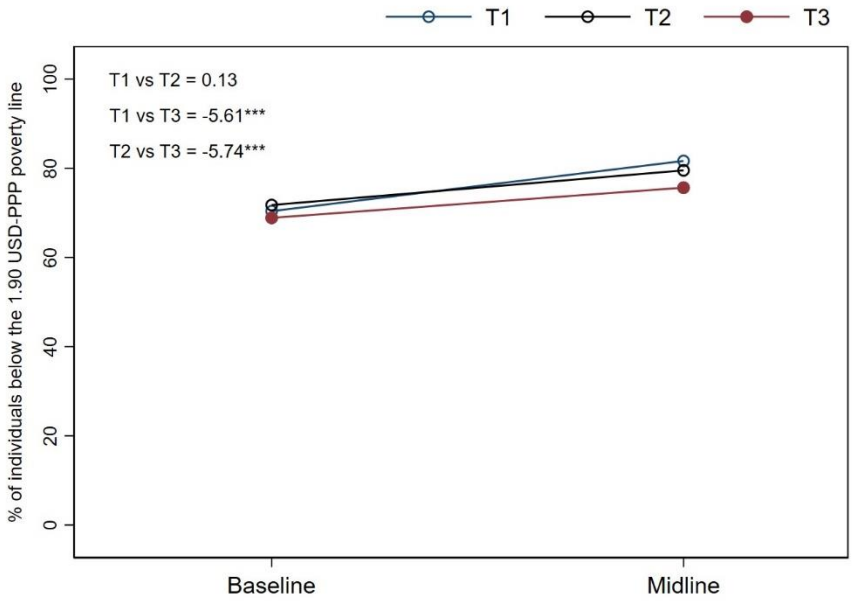
Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 8.6. Household daily per capita consumption in 2011 USD-PPP, by (un-pooled) treatment status, survey round



Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 8.7. Poverty headcount based on $1.90 poverty line, by (un-pooled) treatment status, survey round



Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

9. CONCLUSION

In this report, we present findings from the midline survey conducted as part of an ongoing randomized controlled trial analyzing the effects of SPIR II. The survey successfully re-interviewed 98 percent of the 3,015 sample households interviewed at baseline.

Our findings suggest that SPIR II has been characterized by a high level of implementation fidelity. Households in SPIR II kebeles show a high probability of membership in village economic and savings associations; the poorest households targeted for livelihoods grants have received the grants, primarily allocating them to livestock-related activities such as fattening or purchasing more animals. As for the core nutrition interventions, participation in nurturing care groups was high in both treatment arms, ranging between 75 and 89 percent of the households in the T2 and T3 arms, respectively. Over 92 percent of the households in T3 reported receiving maternal cash transfers, aligning with the program design, and caregivers who received maternal cash transfers primarily reported that the funds were directed to child feeding-related expenses.

In examining key outcomes of interest, we concluded that the implementation of SPIR II centering around nurturing care groups has led to an increase in nutrition knowledge on the part of caregivers, though the effect is small. For IYCF practices, there is evidence of significantly enhanced feeding practices for children only in the T3 arm, who are more likely to achieve minimum meal frequency and minimum adequate diet. There are no significant treatment effects observed in the T2 arm. Taken together, these findings suggest that enhancing IYCF practices in this context requires addressing constraints related to gaps in IYCF knowledge (via BCC) and the unaffordability of nutritious foods (via maternal cash transfers) simultaneously.

As expected, these significant improvements in IYCF feeding practices in T3 arm have not yet translated into substantially better anthropometric outcomes. There is some evidence of a marginally significant reduction in stunting in the T3 arm. At the same time, there is robust evidence of enhanced child development in the same arm. These findings suggest that the nurturing care groups alone may not be sufficient to catalyze a meaningful effect on early childhood health and development outcomes, but in conjunction with maternal cash transfers, they can have positive effects.

Anticipating limited impacts on livelihood outcomes in the initial year of implementation, the midline survey focused on a small subset of livelihood indicators. According to the Food Insecurity Experience Scale, severe food insecurity decreased from 58 percent at baseline to 43 percent at midline, while moderately food-insecure households increased from 32 percent to 46 percent at midline. The trends in control and SPIR households were similar without statistically significant differences. By midline, nearly 90 percent of households reported an acceptable Food Consumption Score (FCS), up from 75 percent at baseline. Treatment estimates indicate a slightly more improved FCS among SPIR households compared to the control group.

Extreme poverty remains high. In control (T1) households, the mean per capita consumption fell from $1.78 to $1.48 between baseline and midline while in SPIR (T2 and T3) households it fell from $1.80 to $1.58. The percentage of individuals below the $1.90 poverty line increased from 70 percent to 82 percent in T1 households and from 70 percent to 78 percent among households in the pooled treatment

arms (T2 and T3). A The differences in consumption and poverty trends are small between control and SPIR households, indicating that the SPIR livelihood programming did not meaningfully impact these outcomes between baseline and midline. Furthermore, additional analysis that separates the T2 and T3 arms indicates that the positive effects on livelihood outcomes are predominantly attributed to households in the T3 arm, which received the maternal cash transfers.

Further data collection and analysis at endline (scheduled in 2025) will allow us to evaluate whether the effects on IYCF practices observed at midline have generated larger effects on anthropometric outcomes at endline; and to assess the ongoing effects on livelihoods-related outcomes.

REFERENCES

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2022. "When should you adjust standard errors for clustering?" *Quarterly Journal of Economics* 138 (1):1–35.

Ajayi, Kehinde F., Aziz Dao, and Estelle Koussoubé. 2022. The Effects of Childcare on Women and Children: Evidence from a Randomized Evaluation in Burkina Faso. *CGD Working Paper 628*.

Alderman, Harold, Daniel Gilligan, Melissa Hidrobo, Jessica Leight, and Michael Mulford. 2023. Including Scalable Nutrition Interventions in a Graduation Model Program: Evidence from an Impact Evaluation in Ethiopia.

Alderman, Harold, Daniel O Gilligan, Jessica Leight, Michael Mulford, and Heleene Tambet. 2022. "The role of poultry transfers in diet diversity: A cluster randomized intent to treat analysis." *Food Policy* 107:102212.

Andrews, Colin, Allan Hsiao, and Laura Ralston. 2018. "Social Safety Nets Promote Poverty Reduction, Increase Resilience, and Expand Opportunities." In *Realizing the Full Potential of Social Safety Nets in Africa*, edited by Kathleen Beegle, Aline Coudouel and Emma Monsalve. Washington, D.C.: The World Bank.

Assefa, Nega, Yemane Berhane, and Alemayehu Worku. 2013. "Pregnancy rates and pregnancy loss in Eastern Ethiopia." *Acta obstetricia et gynecologica Scandinavica* 92 (6):642-647.

Ballard, Terri J, Anne W Kepple, and Carlo Cafiero. 2013. The food insecurity experience scale: development of a global standard for monitoring hunger worldwide. *FAO Technical Paper Version 1.1*.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor markets and poverty in village economies." *The Quarterly Journal of Economics* 132 (2):811-870.

Banerjee, Abhijit, Esther Duflo, Raghabendra Chattopadhyay, and Jeremy Shapiro. 2016. "The long term impacts of a graduation program: Evidence from West Bengal."

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science* 348 (6236):1260799.

Banerjee, Abhijit, Dean Karlan, Robert Osei, Hannah Trachtman, and Christopher Udry. 2022. "Unpacking a multi-faceted program to build sustainable income for the very poor." *Journal of Development Economics* 155:102781.

Beegle, Kathleen, Aline Coudouel, and Emma Monsalve. 2018. *Realizing the Full Potential of Social Safety Nets in Africa*. Washington D.C.: The World Bank.

Berhane, Guush, Daniel O Gilligan, John Hoddinott, Neha Kumar, and Alemayehu Seyoum Taffesse. 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, Guush, Jenna Golan, Kalle Hirvonen, John Hoddinott, Sunny Kim, Alemayehu Seyoum Taffesse, Kibrewossen Abay, Thomas Assefa, Yetmwork Habte, Mehari Hiluf, Bethelem Koru, Fanaye Tadesse, and Feiruz Yimer. 2020. Evaluation of the Nutrition-sensitive Features of the Fourth Phase of Ethiopia's Productive Safety Net Programme. *IFPRI-ESSP Working Paper* 140.

Berhane, Guush, Kalle Hirvonen, and John Hoddinott. 2016. The Implementation of the Productive Safety Nets Programme, 2014: Highlands Outcomes Report (2015). Addis Ababa: Ethiopia Strategy Support Program, International Food Policy Research Institute.

Davis, Thomas P, Carolyn Wetzel, Emma Hernandez Avilan, Cecilia de Mendoza Lopes, Rachel P Chase, Peter J Winch, and Henry B Perry. 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 Onis, M., A. W. Onyango, E. Borghi, A. Siyam, C. Nishida, and J. Siekmann. 2007. "Development of a WHO growth reference for school-aged children and adolescents." *Bulletin of the World Health Organization* 85 (9):660-7.
Deaton, Angus, and Alessandro Tarozzi. 2000. "Prices and poverty in India." *Princeton, July*.
ESS. 2023. Inflation Report on September, EFY 2016. In *Statistical Bulletin*. Addis Ababa: Ethiopian Statistical Service (ESS).
Fiszbein, Ariel, and Norbert R Schady. 2009. *Conditional cash transfers: reducing present and future poverty*: World Bank Publications.
Gilligan, Daniel O, Kalle Hirvonen, Jessica Leight, Heleene Tambet, and Haleluya Tesfaye. 2023. *Baseline survey report of the Strengthen PSNP Institutions and Resilience phase II (SPIR II) resilience food security activity in Ethiopia*, *The Implementer-Led Evaluation & Learning Associate Award*. Washington D.C.: International Food Policy Research Institute (IFPRI).
Golan, Jenna, Derek Headey, Kalle Hirvonen, and John Hoddinott. 2019. "Changes in child undernutrition rates in Ethiopia, 2000-16." In *The Oxford Handbook of the Ethiopian Economy*, edited by Christopher Cramer, Fantu Cheru and Arkebe Oqubay. Oxford: Oxford University Press.
Hickey, Sam, Tom Lavers, Miguel Niño-Zarazúa, and Jeremy Seekings. 2018. The negotiated politics of social protection in sub-Saharan Africa. *UNU-WIDER Working Paper* 2018/34.
Hidrobo, Melissa, John Hoddinott, Neha Kumar, and Meghan Olivier. 2018. "Social Protection, Food Security, and Asset Formation." *World Development* 101:88-103.
Hirvonen, Kalle, Abdulazize Wolle, Arnaud Laillou, Vincenzo Vinci, Stanley Chitekwe, and Kaleab Baye. 2021. "Child growth faltering dynamics in food insecure districts in rural Ethiopia." *Maternal & Child Nutrition*:e13262.
Karra, Mahesh, SV Subramanian, and Günther Fink. 2016. "Height in healthy children in low-and middle-income countries: an assessment." *The American Journal of Clinical Nutrition* 105 (1):121-126.
Kraay, Aart, and David McKenzie. 2014. "Do poverty traps exist? Assessing the Evidence." *The Journal of Economic Perspectives* 28 (3):127-148.
Leight, Jessica, Kalle Hirvonen, Daniel O. Gilligan, Michael Mulford, and Haleluya Tesfaye. 2023. "The impact of a reduced graduation model program on child nutrition in rural Ethiopia: Evidence from a randomized controlled trial." *Journal of Development Economics* Registered Report Stage 1: Proposal.
Liang, Kung-Yee, and Scott L Zeger. 1986. "Longitudinal data analysis using generalized linear models." *Biometrika* 73 (1):13-22.
Lovon, Margarita, and Astrid Mathiassen. 2014. "Are the World Food Programme's food consumption groups a good proxy for energy deficiency?" *Food security* 6:461-470.
Maxwell, Daniel, Bapu Vaitla, and Jennifer Coates. 2014. "How do indicators of household food insecurity measure up? An empirical comparison from Ethiopia." *Food Policy* 47 (August):107-116.
McCoy, Dana Charles, Marcus Waldman, CREDI Field Team, and Günther Fink. 2018. "Measuring early childhood development at a global scale: Evidence from the Caregiver-Reported Early Development Instruments." *Early childhood research quarterly* 45:58-68.
McKenzie, David. 2012. "Beyond baseline and follow-up: The case for more T in experiments." *Journal of Development Economics* 99 (2):210-221.
Medhin, Girmay, Charlotte Hanlon, Michael Dewey, Atalay Alem, Fikru Tesfaye, Bogale Worku, Mark Tomlinson, and Martin Prince. 2010. "Prevalence and predictors of undernutrition among infants aged six and twelve months in Butajira, Ethiopia: the P-MaMiE Birth Cohort." *BMC public health* 10:1-15.

Özler, Berk, Lia CH Fernald, Patricia Kariger, Christin McConnell, Michelle Neuman, and Eduardo Fraga. 2018. "Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial." *Journal of Development Economics* 133:448-467.

Premand, Patrick, and Oumar Barry. 2022. "Behavioral change promotion, cash transfers and early childhood development: Experimental evidence from a government program in a low-income setting." *Journal of Development Economics* 158:102921.

Ruel, Marie T, and Harold Alderman. 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, Rachel, Kalle Hirvonen, Jeremy Lind, and John Hoddinott. 2022. "Expanding social protection coverage with humanitarian aid: Lessons on targeting and transfer values from Ethiopia." *The Journal of Development Studies* 58 (10):1981-2000.

Sabates-Wheeler, Rachel, Jeremy Lind, John Hoddinott, and Mulugeta Tefera Taye. 2021. "Graduation after 10 years of Ethiopia's Productive Safety Net Programme: Surviving but still not thriving." *Development Policy Review* 39 (4):511-531.

Wells, Jonathan CK, André Briend, Erin M Boyd, James A Berkely, Andrew Hall, Sheila Isanaka, Patrick Webb, Tanya Khara, and Carmel Dolan. 2019. "Beyond wasted and stunted—a major shift to fight child undernutrition." *The Lancet Child & Adolescent Health*.

WHO. 2006. "WHO Child Growth Standards based on length/height, weight and age." *Acta paediatrica* Suppl 450:76-85.

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

APPENDIX A. ADDITIONAL TABLES AND FIGURES

Table A1. Primary and secondary outcomes: nutrition

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

Table A2. Primary and secondary outcomes: livelihoods

	Reported at midline	Reported at endline
Primary outcomes:		
Daily per-capita consumption-expenditure (BL40)	X	X
Total value of productive assets		X
Total value of livestock assets		X
Savings (binary and continuous variable)		X
Secondary outcomes:		
Food Insecurity Experience Scale (BL06)	X	X
Food consumption score (BL10)	X	
Prevalence of poverty (BL01)	X	X
Depth of poverty of the poor (BL02)	X	X
Net income from livestock production (binary and continuous variable)		X
Net income from any non-agricultural production (binary and continuous variable)		X
Credit access (binary and continuous variable) (BL42)		X
Cash-earning indicators (BL32, BL33, BL34, BL35)		X

Table A3. Unconditional and conditional treatment effect estimates, nutrition indicators

Arm	Coefficient	Standard error	p-value	N
1a) Caregiver's nutrition knowledge score, unconditional estimates (ANCOVA)				
T2-T1	0.276	0.113	0.015	2,231
T3-T1	0.410	0.106	0.000	2,231
T3-T2	0.135	0.095	0.158	2,231
1b) Caregiver's nutrition knowledge score, conditional estimates (ANCOVA)				
T2-T1	0.280	0.088	0.002	2,231
T3-T1	0.426	0.090	0.000	2,231
T3-T2	0.146	0.075	0.052	2,231
2a) % of children consuming a diet of minimum diversity, unconditional estimates				
T2-T1	1.60	2.30	0.487	2,233
T3-T1	16.54	2.95	0.000	2,233
T3-T2	14.94	2.97	0.000	2,233
2b) % of children consuming a diet of minimum diversity, conditional estimates				
T2-T1	-0.40	1.90	0.835	2,233
T3-T1	14.42	2.16	0.000	2,233
T3-T2	14.82	2.31	0.000	2,233
3a) % of children receiving a diet of minimum meal frequency, unconditional estimates				
T2-T1	2.69	3.08	0.384	2,221
T3-T1	9.38	3.06	0.002	2,221
T3-T2	6.69	2.94	0.024	2,221
3b) % of children receiving a diet of minimum meal frequency, conditional estimates				
T2-T1	1.85	2.52	0.463	2,221
T3-T1	7.56	2.22	0.001	2,221
T3-T2	5.71	2.41	0.018	2,221
4a) % of children receiving a diet of minimum acceptable diet, unconditional estimates				
T2-T1	2.09	1.92	0.279	2,221
T3-T1	13.19	2.48	0.000	2,221
T3-T2	11.10	2.58	0.000	2,221
4b) % of children receiving a diet of minimum acceptable diet, conditional estimates				
T2-T1	0.79	1.59	0.617	2,221
T3-T1	11.38	1.80	0.000	2,221
T3-T2	10.59	2.07	0.000	2,221
5a) Mean height for age Z-score, unconditional estimates				
T2-T1	-0.018	0.075	0.812	2,226
T3-T1	0.022	0.073	0.767	2,226
T3-T2	0.039	0.075	0.600	2,226
5b) Mean height for age Z-score, conditional estimates				
T2-T1	-0.006	0.061	0.921	2,226
T3-T1	0.091	0.060	0.128	2,226
T3-T2	0.097	0.057	0.089	2,226
6a) % of children stunted, unconditional estimates				
T2-T1	0.75	2.71	0.78	2,226

Arm	Coefficient	Standard error	p-value	N
T3-T1	-1.92	2.69	0.48	2,226
T3-T2	-2.67	2.77	0.34	2,226
6b) % of children stunted, conditional estimates				
T2-T1	1.04	2.34	0.66	2,226
T3-T1	-3.30	2.31	0.15	2,226
T3-T2	-4.34	2.18	0.05	2,226
7a) Mean weight for height Z-score, unconditional estimates				
T2-T1	-0.079	0.067	0.241	2,226
T3-T1	0.039	0.064	0.547	2,226
T3-T2	0.118	0.068	0.086	2,226
7b) Mean weight for height Z-score, conditional estimates				
T2-T1	-0.088	0.056	0.115	2,226
T3-T1	0.052	0.056	0.355	2,226
T3-T2	0.140	0.055	0.011	2,226
8a) % of children 6–23 months wasted, unconditional estimates				
T2-T1	1.83	1.47	0.213	2,226
T3-T1	0.50	1.22	0.678	2,226
T3-T2	-1.32	1.54	0.392	2,226
8b) % of children 6–23 months wasted, conditional estimates				
T2-T1	2.35	1.26	0.063	2,226
T3-T1	0.43	1.16	0.714	2,226
T3-T2	-1.93	1.36	0.159	2,226

Note: Conditional regression estimates control for household size at baseline, primary female's age and education level at baseline, index child's sex and age and strata fixed effects. When a baseline value is available, both the unconditional and conditional estimate is based on the ANCOVA method. Standard errors in all regressions are clustered at the level of treatment (kebele).

Table A4. Unconditional and conditional treatment effect estimates, livelihood indicators

Arm	Coefficient	Standard error	p-value	N
1a) FIES: % of food secure households, unconditional estimates (ANCOVA)				
SPIR	-0.02	0.63	0.971	2,954
1b) FIES: % of food secure households, conditional estimates (ANCOVA)				
SPIR	-0.07	0.53	0.890	2,954
2a) FIES: %of mildly food insecure households, unconditional estimates (ANCOVA)				
SPIR	2.16	1.34	0.109	2,954
2b) FIES: % of mildly food insecure households, conditional estimates (ANCOVA)				
SPIR	1.41	0.95	0.140	2,954
3a) FIES: % of moderately food insecure households, unconditional estimates (ANCOVA)				
SPIR	1.36	2.99	0.649	2,954
3b) FIES: % of moderately food insecure households, conditional estimates (ANCOVA)				
SPIR	1.75	1.94	0.367	2,954
4a) FIES: % of severely food insecure households, unconditional estimates (ANCOVA)				
SPIR	-3.15	3.55	0.376	2,954
4b) FIES: % of severely food insecure households, conditional estimates (ANCOVA)				
SPIR	-2.91	1.86	0.120	2,954
5a) FCS: % of households with acceptable diets, unconditional estimates (ANCOVA)				
SPIR	3.29	1.62	0.044	2,954
5b) FCS: % of households with acceptable diets, conditional estimates (ANCOVA)				
SPIR	3.36	1.19	0.005	2,954
6a) FCS: % of households with borderline diets, unconditional estimates (ANCOVA)				
SPIR	-3.33	1.42	0.020	2,954
6b) FCS: % of households with borderline diets, conditional estimates (ANCOVA)				
SPIR	-3.26	1.11	0.004	2,954
7a) FCS: % of households with poor diets, unconditional estimates (ANCOVA)				
SPIR	0.04	0.63	0.954	2,954
7b) FCS: % of households with poor diets, conditional estimates (ANCOVA)				
SPIR	-0.11	0.50	0.825	2,954
8a) Daily per-capita consumption-expenditure, unconditional estimates (ANCOVA)				
SPIR	0.09	0.04	0.024	2,954
8b) Daily per-capita consumption-expenditure, conditional estimates (ANCOVA) *				
SPIR	0.06	0.03	0.018	2,954
9a) Prevalence of extreme poverty, unconditional estimates (ANCOVA) *				
SPIR	-4.07	2.12	0.056	17,041
9b) Prevalence of extreme poverty, conditional estimates (ANCOVA) *				

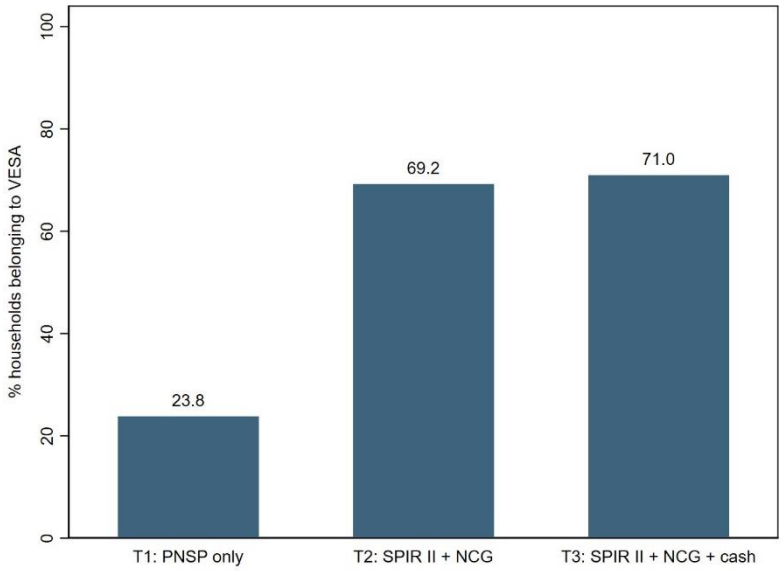
Arm	Coefficient	Standard error	p-value	N
SPIR	-2.83	1.57	0.074	17,041
10a) Poverty gap index: Depth of poverty of the poor, unconditional estimates (ANCOVA) *				
SPIR	-2.08	1.35	0.125	9,815
10b) Poverty gap index: Depth of poverty of the poor, conditional estimates (ANCOVA) *				
SPIR	-2.37	0.98	0.016	9,815

Note: Conditional regression estimates control for household size at baseline, primary female's age and education level at baseline, and strata fixed effects. Standard errors in all regressions are clustered at the level of treatment (kebele). The coefficients represent the difference between T1 and the pooled treatment arms (T2 and T3).

***** The unit of analysis in poverty headcount and depth of poverty indicators is an individual, not household. Therefore, the means, standard deviations and treatment effects have been estimated using household size as the frequency weight. This explains the larger reported number of observations in these analyses.

Figure A1. Percent of households belonging to a village economic and savings association (VESA), by study arm

A. Amhara



B. Oromia

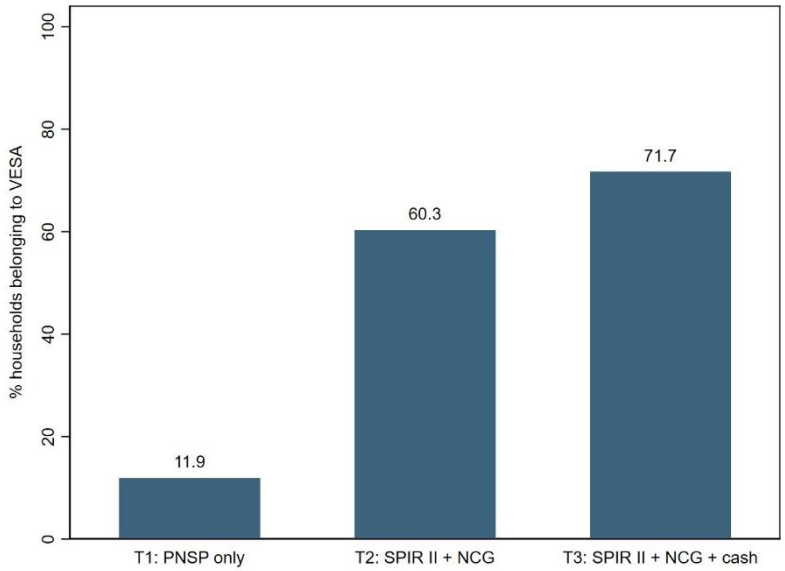
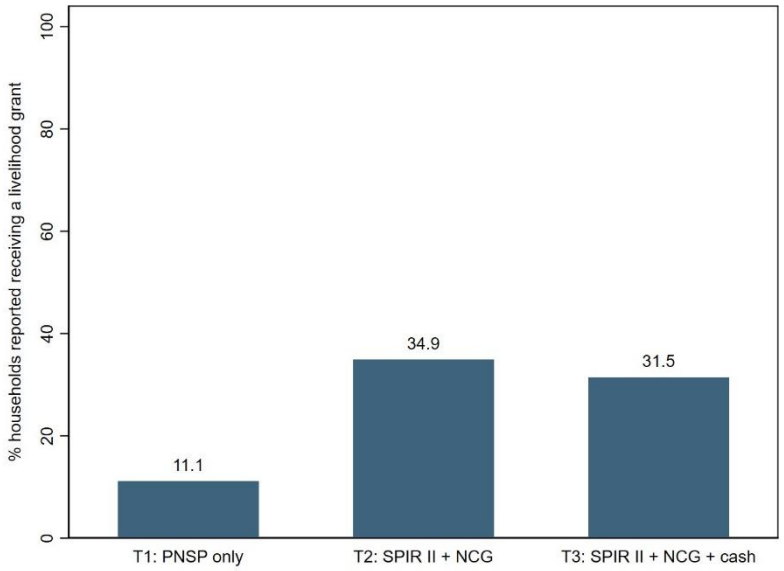


Figure A2. Percent of households reporting to have received a livelihood grant, by study arm

A. Amhara



B. Oromia

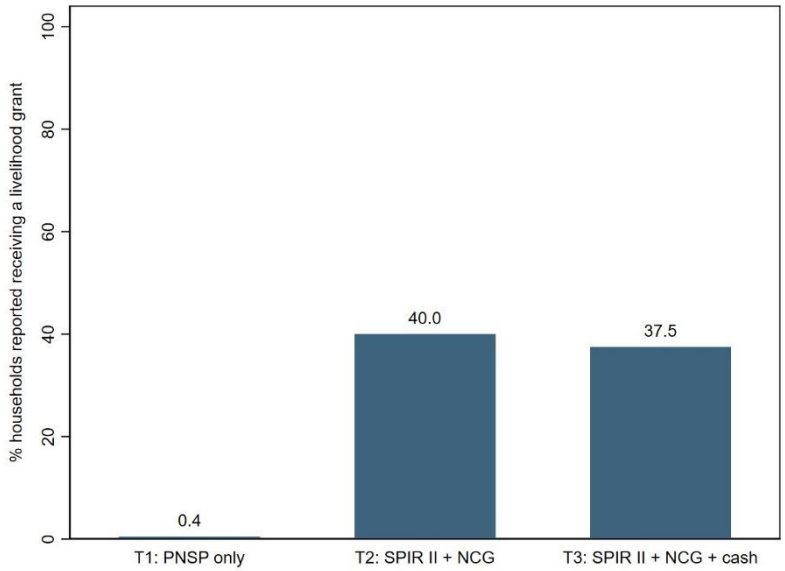
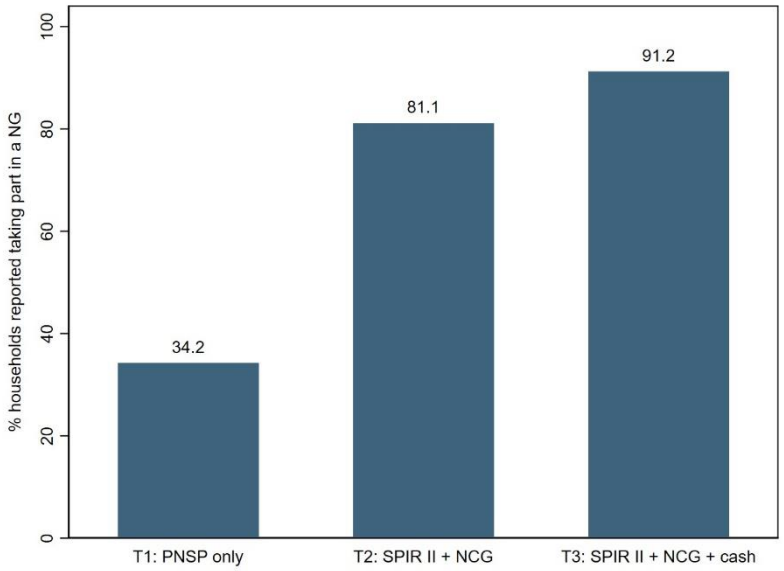


Figure A3. Percent of households reporting to have taken part in NGs, by study arm

A. Amhara



B. Oromia

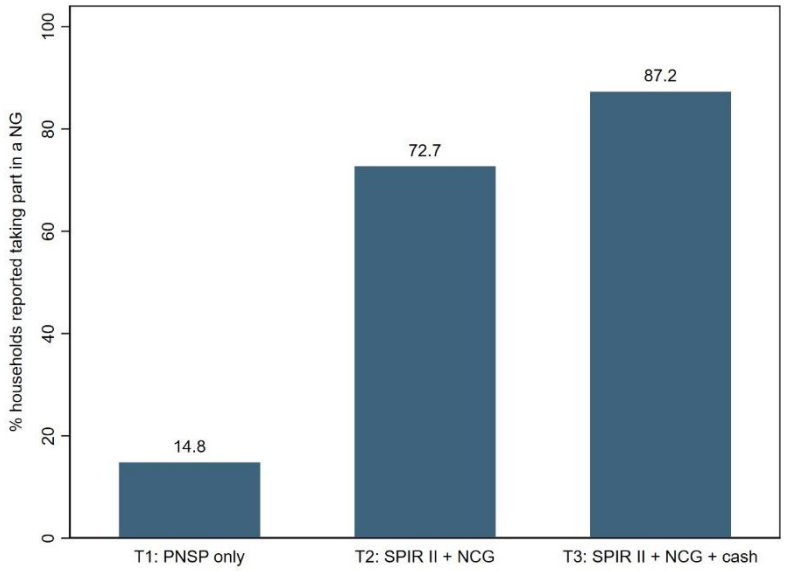
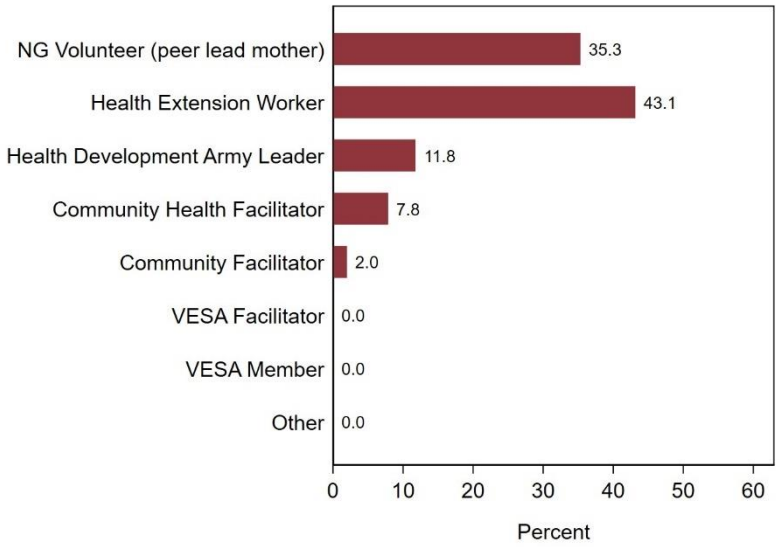


Figure A4. NG facilitators as reported by households reporting to have taken part in NGs, T1 households only

A. Amhara



B. Oromia

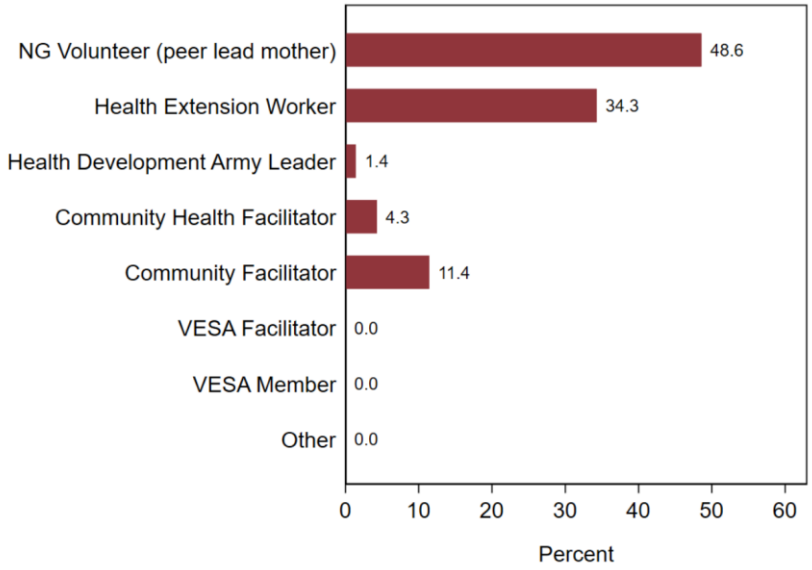
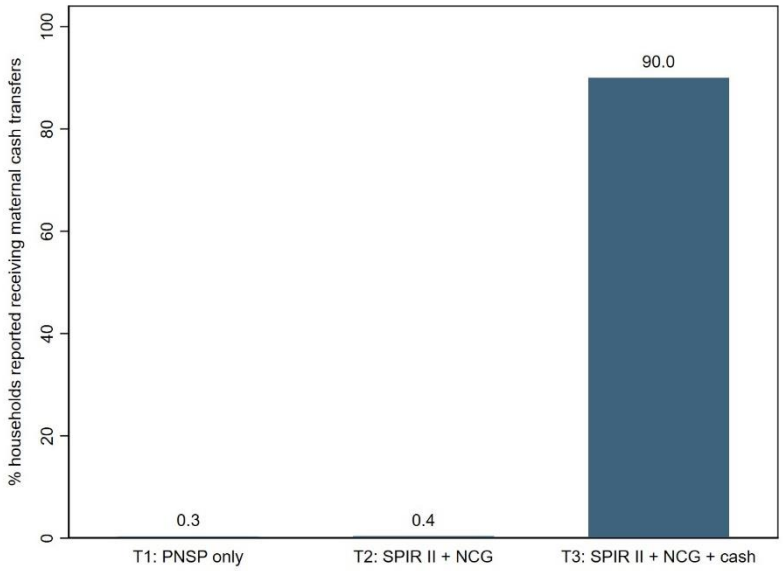


Figure A5. Percent of households reporting to have received maternal cash transfers, by study arm

A. Amhara



B. Oromia

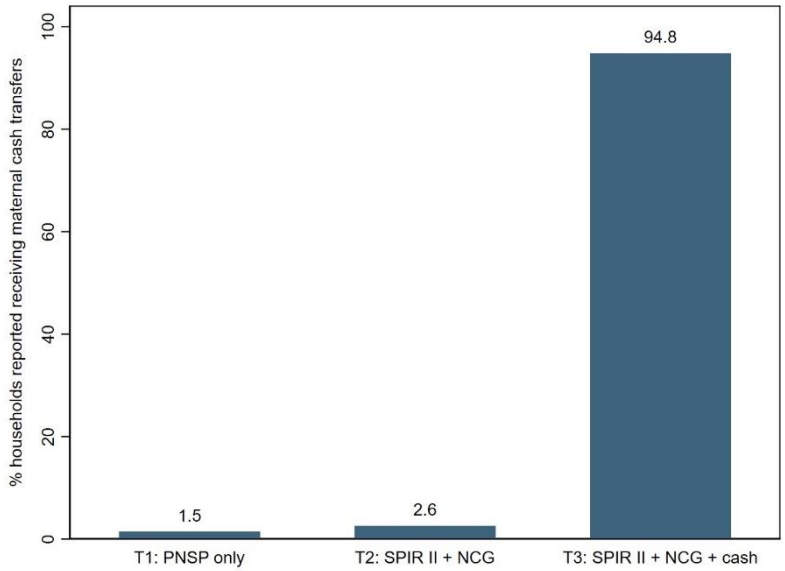
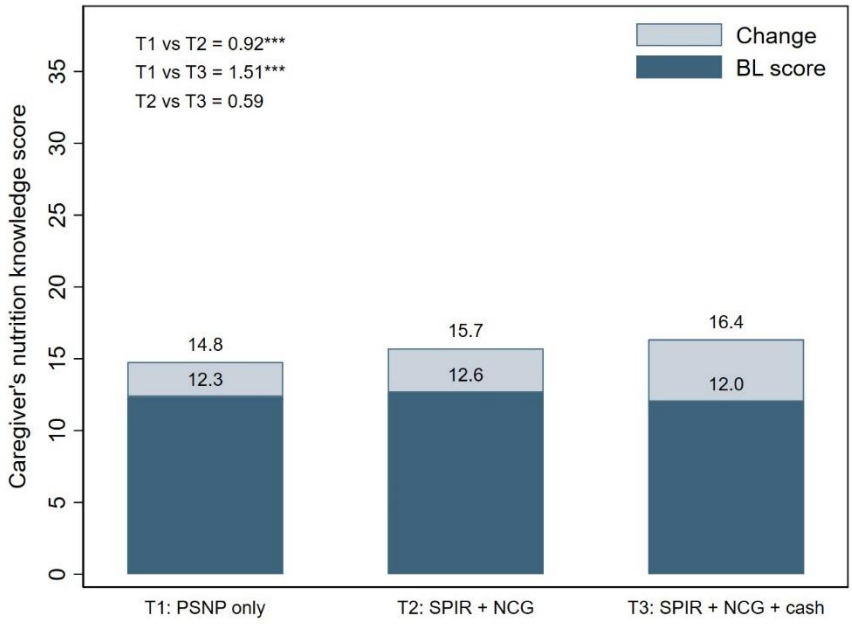


Table A5. Percent of caregivers receiving full points in the nutrition knowledge questions when alternative grading approach is used, by study arm

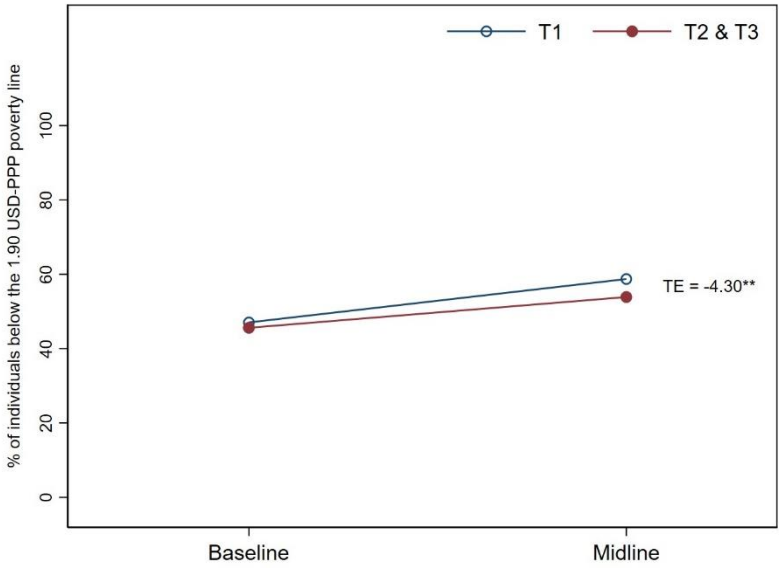
Survey round:	Baseline			Midline		
Treatment status:	T1	T2	T3	T1	T2	T3
Percent of caregivers responding correctly:						
How long after birth should a baby start breastfeeding (1p)	88.3	87.5	86.6	90.4	92.9	91.1
What should a mother do with the 'first milk' or colostrum (1p)	72.0	73.6	72.2	80.0	84.1	84.8
Until what age should a baby by exclusively breastfed (1p)	87.7	88.3	90.7	95.2	97.2	97.0
Why should a baby under 6 months be exclusively breastfed (8p)	25.5	25.9	24.4	28.7	31.3	31.5
If baby is not getting enough breast milk, what should be done (3p)	46.9	49.3	47.9	57.7	60.5	61.0
At what age should a baby first start to receive liquids (1p)	84.7	86.1	85.7	86.9	86.6	91.0
At what age should a baby first start to receive foods (1p)	72.6	74.1	70.7	78.5	78.6	78.5
What can happen to children if they do not get enough iron (6p)	24.6	25.1	22.9	33.4	36.7	40.1
What are some foods that are rich in iron (4p)	24.9	25.2	23.4	35.3	37.5	39.6
What can happen if child does not eat enough vitamin A-rich foods (2p)	40.7	41.8	39.9	56.1	59.5	62.1
What are some foods that contain vitamin A (6p)	26.1	27.6	24.4	31.8	35.4	39.3
Mean nutrition knowledge score (max 34 points):	**12.3**	**12.6**	**12.0**	**14.8**	**15.7**	**16.4**

Figure A6. Caregiver's nutrition knowledge score based on alternative grading, by study arm



Note: N = 2,231 female caregivers. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

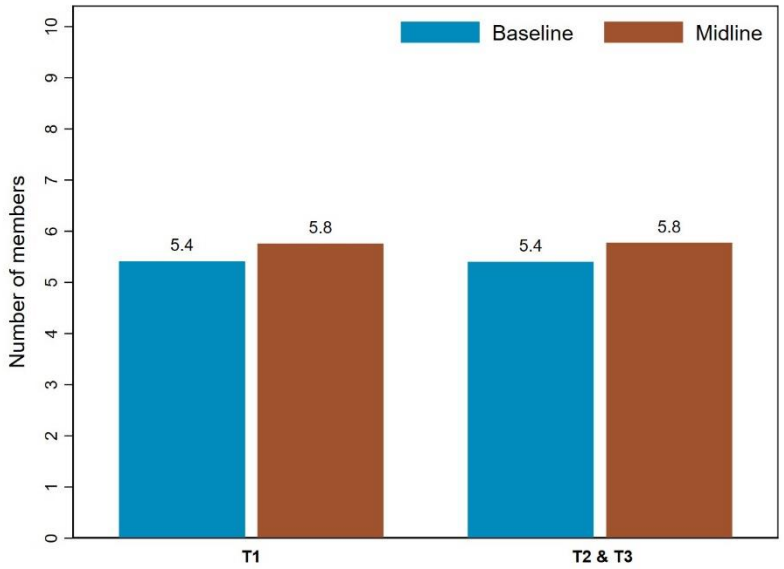
Figure A7. $1.90 poverty headcount based on adult equivalent units, by treatment status and survey round



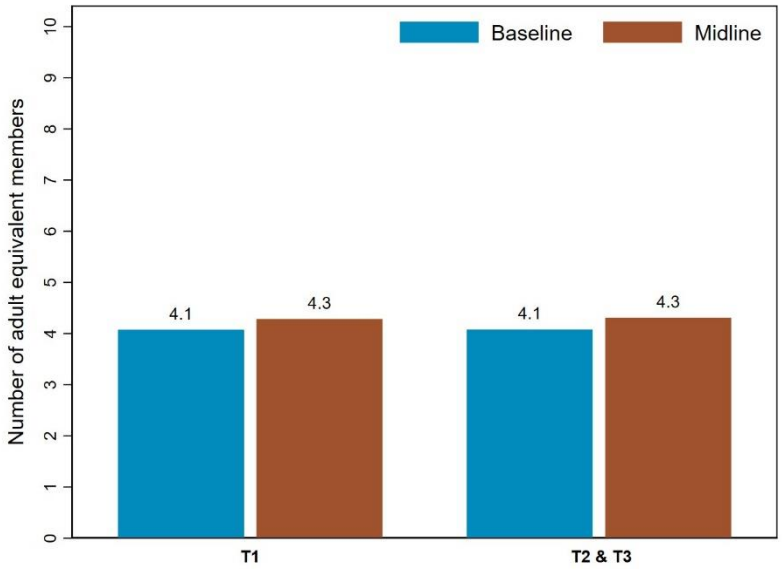
Note: N = 17,041 individuals from 2,954 households. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

Figure A8. Household size, by treatment status and survey round

A. Number of members

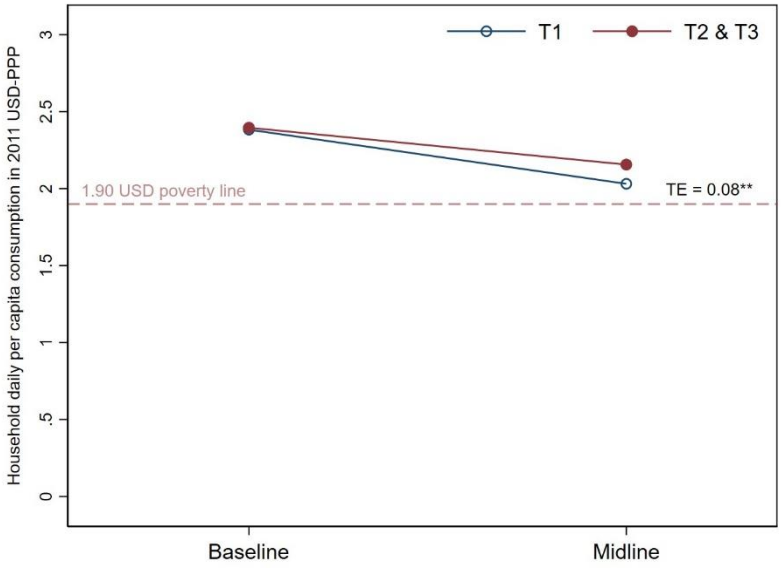


B. Number of adult equivalent members



Note: 2,954 households.

Figure A9. Household daily per capita consumption in 2011 USD-PPP in terms of adult equivalent units, by treatment status and survey round



N = 2,954 households. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. TE = Treatment effect.

Table A6. Treatment effects on key livelihood outcomes, disaggregated treatment arms

Arm	Coefficient	Standard error	p-value	N
Food consumption score (FCS, raw score), unconditional estimates (ANCOVA)				
T2	0.50	0.92	0.59	2,954
T3	7.93	1.23	0.00	2,954
T3-T2	7.44	1.22	0.00	2,954
Food consumption score (FCS, raw score), conditional estimates (ANCOVA)				
T2	0.25	0.82	0.76	2,954
T3	8.24	0.92	0.00	2,954
T3-T2	7.99	0.96	0.00	2,954
Daily per-capita consumption-expenditure, unconditional estimates (ANCOVA)				
T2	0.05	0.04	0.22	2,954
T3	0.13	0.05	0.02	2,954
T3-T2	0.08	0.05	0.14	2,954
Daily per-capita consumption-expenditure, conditional estimates (ANCOVA)				
T2	0.01	0.03	0.82	2,954
T3	0.12	0.03	0.00	2,954
T3-T2	0.11	0.03	0.00	2,954
Prevalence of extreme poverty, conditional estimates (ANCOVA) *				
T2	-2.28	2.33	0.33	17,041
T3	-5.85	2.80	0.04	17,041
T3-T2	-3.57	2.93	0.23	17,041
Prevalence of extreme poverty, conditional estimates (ANCOVA) *				
T2	0.13	1.79	0.94	17,041
T3	-5.61	1.91	0.00	17,041
T3-T2	-5.74	1.99	0.00	17,041

Note: Conditional regression estimates control for household size at baseline, primary female's age and education level at baseline, and strata fixed effects. Standard errors in all regressions are clustered at the level of treatment (kebele). The coefficients represent the difference relative to T1 arm.

***** The unit of analysis in poverty headcount and depth of poverty indicators is an individual, not household. Therefore, the means, standard deviations and treatment effects have been estimated using household size as the frequency weight. This explains the larger reported number of observations in these analyses.

Table A7. Summary table of required indicators at midline

Indicators	T1: PNSP only				T2: SPIR II + NCG				T3: SPIR II + NCG + cash transfers			
	Mean (%)	CI lower	CI upper	N	Mean (%)	CI lower	CI upper	N	Mean (%)	CI lower	CI upper	N
BL06. Prevalence of moderate food insecurity in the household based on the Food Insecurity Experience Scale (FIES)	45.82	42.73	48.90	1,004	43.81	40.69	46.94	970	51.12	47.99	54.26	980
BL06. Prevalence of severe food insecurity in the household based on the Food Insecurity Experience Scale (FIES)	45.22	42.14	48.30	1,004	47.32	44.17	50.47	970	35.92	32.91	38.93	980
BL10. Percent of households with poor food consumption score (FCS)	2.09	1.20	2.98	1,004	3.51	2.35	4.66	970	0.82	0.25	1.38	980
BL10. Percent of households with borderline food consumption score (FCS)	10.26	8.38	12.14	1,004	8.56	6.79	10.32	970	5.31	3.90	6.71	980
BL10. Percent of households with acceptable food consumption score (FCS)	87.65	85.61	89.69	1,004	87.94	85.88	89.99	970	93.88	92.37	95.38	980
BL12. Percent of children 6-23 months receiving a minimum acceptable diet	8.90	6.88	10.92	764	10.99	8.70	13.28	719	22.0	19.09	25.09	738
BL14. Percent of children under five (0-59 months) who had diarrhea in the prior two weeks	36.52	33.21	39.83	816	31.87	28.58	35.16	775	30.16	26.99	33.33	809
BL15. Percent of children under five (0-59 months) with diarrhea treated with Oral Rehydration Therapy	60.20	54.62	65.78	299	62.50	56.43	68.57	248	69.67	63.86	75.48	244
BL39. Prevalence of children 6-23 months consuming a diet of minimum diversity (MDD-C)	12.84	10.47	15.21	771	14.44	11.87	17.02	720	29.38	26.10	32.67	742
BL11. Prevalence of women of reproductive age consuming a diet of minimum diversity (MDD-W)	2.69	1.69	3.70	1,002	2.38	1.41	3.34	968	6.63	5.07	8.19	980
BL40. Daily per capita expenditures (as a proxy for income) in USG-assisted areas, in 2010 USD	1.72	1.66	1.78	1,004	1.73	1.67	1.79	970	1.76	1.70	1.82	980
BL01. Prevalence of poverty: Percent of people living on less than $1.90/day 2011 PPP *	81.64	80.64	82.63	5,783	79.52	78.46	80.58	5,596	75.64	74.53	76.76	5,662

* The unit of analysis in poverty headcount is an individual, not household. This explains the larger reported number of observations for this indicator.