Journal of Development Economics

The impact of a reduced graduation model program on child nutrition in rural Ethiopia: Evidence from a randomized controlled trial
--Manuscript Draft--

Manuscript Number:	DEVEC-D-23-00381R3
Article Type:	Registered Report Stage 1: Proposal
Section/Category:	Health, Education, gender, poverty
Keywords:	Ultra-poor; graduation model; nutrition; maternal grants; child stunting
Corresponding Author:	Jessica Leight International Food Policy Research Institute Washington, DC UNITED STATES
First Author:	Jessica Leight
Order of Authors:	Jessica Leight
	Kalle Hirvonen
	Daniel O. Gilligan
	Michael Mulford
	Haleluya Tesfaye
Abstract:	We conduct a cluster randomized controlled trial evaluating a reduced graduation model intervention addressing the high rates of chronic undernutrition among young children in the poorest households (beneficiaries of the Productive Safety Net Program) in rural Ethiopia. We hypothesize that a set of limited livelihood interventions coupled with behavioral change communication delivered via nurturing care groups and maternal cash grants will improve child nutritional status through improved child feeding practices, reduced growth faltering, and enhanced household livelihoods. We use a cohort-focused sampling strategy (sampling pregnant women and infants at baseline) in conjunction with carefully timed surveys (a midline survey 12 months post-baseline and endline 36 months post-baseline) to capture the complex age dynamics of child feeding and linear growth faltering. Primary outcome variables include caregiver nutrition knowledge, child's dietary diversity and height-for-age, and household savings and consumption. Secondary outcomes include other child feeding and anthropometric indicators and household livelihood outcomes.
Response to Reviewers:	

Highlights

Multifaceted graduation models are a promising tool to reduce poverty.

However, there is little evidence around effects on child malnutrition.

We conduct a randomized trial to assess a reduced graduation model.

The intervention is embedded into Ethiopia's flagship safety net program.

Journal of Development Economics Registered Report Stage 1: Proposal

The impact of a reduced graduation model program on child nutrition in rural Ethiopia: Evidence from a randomized controlled trial

Jessica Leight [a], Kalle Hirvonen [a,b], Daniel O. Gilligan [a], Michael Mulford [c] and Haleluya Tesfaye [a]

[a] International Food Policy Research Institute (IFPRI), [b] United Nations University World Institute for Development Economics Research (UNU-WIDER), [c] World Vision

Date of latest draft: 9/11/2023

Abstract

We conduct a cluster randomized controlled trial evaluating a reduced graduation model intervention addressing the high rates of chronic undernutrition among young children in the poorest households (beneficiaries of the Productive Safety Net Program) in rural Ethiopia. We hypothesize that a set of limited livelihood interventions coupled with behavioral change communication delivered via nurturing care groups and maternal cash grants will improve child nutritional status through improved child feeding practices, reduced growth faltering, and enhanced household livelihoods. We use a cohort-focused sampling strategy (sampling pregnant women and infants at baseline) in conjunction with carefully timed surveys (a midline survey 12 months post-baseline and endline 36 months post-baseline) to capture the complex age dynamics of child feeding and linear growth faltering. Primary outcome variables include caregiver nutrition knowledge, child's dietary diversity and height-for-age. Secondary outcomes include other child feeding and anthropometric indicators and household livelihood outcomes.

Keywords: Ultra-poor, graduation model, nutrition, maternal grants, child stunting.

JEL codes: I15, I32, I38.

Study pre-registration: https://www.socialscienceregistry.org/trials/9923.

Proposed timeline

The study was launched in 2022. The study clusters were randomized by the research team into three study arms in June 2022. The baseline survey took place between August and September in 2022. The program implementation began in November 2022. The midline survey is scheduled for August-September 2023 and the endline survey for August-September 2025. We plan to complete the research paper within 8 months of the completion of the endline survey.

Reporting checklist for Stage 1 submissions

Section	Item	Description and details to report	Reported?	Page(s)
Cover page *(required)*	*Title*	Informative title specifying the study design, population, and interventions	Yes	1
	Date of latest draft	Date of when the prospective review article was last edited.	Yes	1
	Study pre-registration status	Link, registration identifier and registry name (or intended registry if not yet registered)	Yes	1
	Keywords	Up to six keywords, to be used for indexing purposes.	Yes	1
	JEL codes	Up to six codes.	Yes	1
Abstract (*required*)	*Abstract*	Summarize research question, outcome variables, methodological framework and contribution in less than 150 words.	Yes	1
Timeline (*required*)	*Expected completion date*	Expected date for completion of the pre-specified research design.	Yes	2
Introduction	*Background and relevance of the study*	Brief overview of previous research, and relevance of the research question(s) for the field of economic development	Yes	4-6
	Research question(s)		Yes	4-6
Research design	*Basic methodological framework*	Outline of the identification strategy in your study (experimental/non-experimental)	Yes	14
	Hypotheses	Pre-specified hypotheses to be tested in the study and reported as primary findings in the Stage 2 full manuscript	Yes	8-10
	Outcome variable(s)	Definition of the main outcome variable(s) and (if applicable) secondary outcome variable(s)	Yes	10-14
		Specification of how outcome(s) will be constructed from the dataset	Yes	10-14
	Intervention(s)	Details of the intervention (when, where, how, by whom)	Yes	7-8
		Number of treatment arms and whether they are exclusive or overlapping	Yes	7-8
		Randomization strategy	Yes	18
		Blinding strategy (if applicable)	N/A	
		Instructions and supporting materials for administering the intervention	N/A	
		Source(s) of exogenous variation	Yes	14
	Theory of change	How and why the intervention is predicted to lead to certain effects	Yes	8-10
	Sample	Specification of unit of analysis (individuals, organizations, countries, etc.)	Yes	14
		Data source(s)	Yes	18

Section	Item	Description and details to report	Reported?	Page(s)
		Projected sample size and statistical power calculations	Yes	14-17
	Variations from the intended sample	Specification of the degree of attrition that may threaten the robustness of the study	Yes	13
		Strategies to deal with attrition, non-compliance with the assigned treatment, etc.	Yes	14
	Data collection and processing	Type of data, collection method/data source(s), and timeline for collection	Yes	18
		Rule for terminating data collection / stopping rule	Yes	19
		Data management plan	Yes	19
		Pilot data and experiments run in preparation of the Stage 1 submission	N/A	
Empirical analysis	*Statistical method(s)*	Main evaluation method(s) and underlying assumptions	Yes	21
		Rules for handling missing values	Yes	21
		Definition and rules for handling outliers	Yes	21
	Multiple hypothesis testing	Strategies to prevent false positives	Yes	21
	Heterogeneous effects	Anticipated heterogeneous effects and theoretical justification	Yes	22
	Statistical model	A functional (mathematical) form of the causal mechanism explored in the study	Yes	21
		Specification if regression model is linear, generalized linear, or other	Yes	21
		How will standard errors be calculated	Yes	21
Limitations and challenges	*Challenges in the study implementation*	Potential objective circumstances that might jeopardize the implementation of the proposed study design	Yes	23
Administrative information *(required)*	*Ethics approval*	Statement confirming that all necessary ethics approvals are in place.	Yes	29
	Funding	Funding sources in the suggested format	Yes	29
	Acknowledgments	List of (non-author) individuals who provided help to the research project.	Yes	29
Bibliography	*Bibliography*	References can be in any style or format as long as the style is consistent.	Yes	24-27
Other items	*Appendices*	Tables and figures	Yes	29

1.   Introduction

Over the past two decades, social safety net programs that provide food or cash transfers to support household consumption have become an important policy tool to enhance food insecurity and reduce extreme poverty in low- and middle-income countries (Andrews et al., 2018; Beegle et al., 2018; Hidrobo et al., 2018; World Bank, 2018), but there is little evidence that these interventions can effectively reduce persistently high rates of child malnutrition (Manley et al., 2022; Manley et al., 2020). Similarly, multifaceted "graduation model" interventions are highly promising in terms of their persistent effects on poverty (Balboni et al., 2022; Bandiera et al., 2017; Banerjee et al., 2015; Banerjee et al., 2021; Banerjee et al., 2022; Bedoya et al., 2019; Bossuroy et al., 2022; Brune et al., 2022) and thus can plausibly facilitate households' exit from long-term consumption support. However, their high cost raises some questions about scalability, and their effects on child nutrition remain largely unknown.[1]

An absence of evidence around the effects of scalable antipoverty interventions on child malnutrition is troubling given that malnutrition remains a widespread and persistent challenge in low- and middle-income countries. In 2021, nearly 150 million children under five years of age were estimated to suffer from chronic under-nutrition (FAO et al., 2022). Adequate nutrition and good health during infancy and early childhood form the foundation for lifelong wellbeing (Currie and Vogl, 2013) and short-term nutritional deprivations during this period have been found to lower final educational attainment and increase the risk of poor health and poverty in adulthood (Alderman et al., 2006; Dercon and Porter, 2014; Maccini and Yang, 2009; Maluccio, 2010). While some progress can be made via more nutrition-specific[2] interventions, reducing malnutrition widely requires the incorporation of nutrition-sensitive dimensions into anti-poverty interventions implemented at scale (Ruel and Alderman, 2013): for example, within existing government-led social protection programs. Here, the evidence remains scant.

We seek to address this evidence gap by conducting a cluster randomized controlled trial (RCT) to assess the nutritional impacts of a reduced graduation model intervention embedded into Ethiopia's flagship safety net program, the Productive Safety Net Program (PSNP). SPIR II is a reduced graduation model program implemented in the Amhara and Oromia regions of Ethiopia that focuses on addressing the poor nutritional status and high rates of stunting among infants and young children.[3] The SPIR II interventions target PSNP households and include both livelihoods and nutrition-focused components; however, the main objective of this trial is to evaluate a novel set of interventions specifically designed to target poor nutritional outcomes.

---

[1]The existing papers in the graduation model literature generally do not measure or report any effects on child nutrition or anthropometric status, though Banerjee et al. presents evidence of positive effects on an index of adult health, and Bedoya et al. shows positive effects in reducing the incidence of child diarrhea. The positive effects on consumption and food security are promising, but positive effects on household-level indices do not necessarily generate positive effects on more specific variables capturing child malnutrition, as demonstrated by the meta-analyses studying the impacts of cash transfers on children's nutritional outcomes (Manley et al., 2022; Manley et al., 2020). Raza et al. (2018) is the only relevant paper here, and uses the Targeting the Ultra-Poor (TUP) evaluation data (Bandiera et al., 2017) to estimate the program's impact on child nutrition. The analysis does document positive effects on indicators capturing acute undernutrition but the impacts on chronic undernutrition outcomes are not significantly different from zero, possibly due to the relatively small number of young children in the data.

[2] *Nutrition-specific interventions* target immediate determinants of fetal and child nutrition, emphasizing factors such as nutrient intake and disease management. In contrast, *nutrition-sensitive interventions* address underlying factors such as food security, caregiving resources, and access to healthcare, incorporating specific nutrition goals within these contexts (Ruel and Alderman, 2013).

[3] SPIR II is an acronym for Strengthen PSNP Institutions and Resilience, and is funded by USAID's Bureau for Humanitarian Assistance (BHA) and implemented by World Vision International (lead), CARE, and ORDA. The first phase of the program (SPIR I) took place in 2016-2021 and the evaluation results are presented in Alderman et al. (2021) and summarized below in Section 2.

The SPIR II nutrition interventions center around the formation of nurturing care groups (NCGs), peer caregivers' groups led by community-based trained volunteer agents; and the provision of monthly maternal grants of $20 over a period of 24 months in infancy to a subset of households. All treated households also receive a bundled set of light-touch livelihoods interventions (including village savings groups and training), and households identified as scoring in the lowest 33 percent of a composite asset index at baseline will receive one-time $300 grants.

The cluster RCT has three study arms in which all participating households benefit from the PSNP. The first study arm serves as a control group and is formed of PSNP households that will not receive additional SPIR programming. PSNP households in the second study arm receive SPIR graduation programming that includes both livelihoods components (as described above) and the nurturing care groups. PSNP households in the third study arm receive the same set of interventions as the second arm, but in addition receive the monthly maternal grants.

The study design focuses on early childhood nutritional outcomes and is optimized to capture the effects of targeted nutrition interventions at key points in the early childhood growth trajectory. We hypothesize that the SPIR interventions will improve infant and young child feeding practices, prevent child growth faltering in the first 1,000 days (i.e., the period between conception and the child's second birthday), and enhance a specific set of livelihood outcomes. The experimental design allows us to assess the impact of SPIR (inclusive of nurturing care groups) on infant and child feeding and anthropometric outcomes, and further test the role of monthly cash transfers in targeting liquidity constraints that may constrain households in improving child nutrition outcomes. At the same time, we will evaluate the effect of the bundled livelihood interventions (provided in both treatment arms) on targeted outcomes of interest, particularly consumption and savings.

This trial adds to a growing literature analyzing the effects of nutrition-sensitive graduation model interventions and cash transfer programs in developing countries. Together with the recent pre-registered cluster-RCT by Bouguen and Dillon (2019)[4], our study is one of the first trials to assess the impact of a nutrition-sensitive graduation program.[5] We argue that our trial makes two major contributions, one in the area of the chosen interventions, and one in the area of measurement.

Programmatically, the SPIR nutrition interventions focus on improving caregiver knowledge through intensive behavior change communication to promote age-appropriate diets and child-feeding practices. We build on Bouguen and Dillon (2019) by including targeted behavioral change communication and information provision, an element excluded from their graduation model; nutritionists and public health experts consider the inclusion of an intensive behavior change communication component as essential to achieving meaningful impacts in nutrition-sensitive programs (Leroy et al., 2016; Ruel et al., 2018). We also build on a previous evaluation of SPIR I (Alderman et al. 2021) by analyzing the nurturing care groups, a novel method to deliver behavioral change communication (BCC) that does not rely on government health

---

[4] Bouguen and Dillon (2019) assesses the nutritional impacts of a multifaceted graduation program in Burkina Faso. Partnering with international NGOs, the authors assign poor households with a pregnant or lactating women and/or malnourished young child into three study arms: unconditional cash transfer, cash+asset transfer and cash+asset+nutrition arms. The households in the latter nutrition arm received a kitchen garden kit and iron fortified flour.

[5] Another recent paper by part of this author team analyzed the short-term effects of a light-touch graduation model program including either a cash or poultry transfer on the consumption of eggs by women and young children and a select set of other nutritional outcomes measured in the short-term, nine months post-transfer (Alderman et al., 2022). The effects were generally small in magnitude, and the paper did not report on a full set of outcomes related to child health and anthropometrics.

extension workers that are burdened with multiple activities and heavy client caseloads (Berhane et al., 2020). As a result, the HEWs are often unable to take on intensive BCC activities that are required to improve infant and child feeding practices.

In addition, we evaluate the marginal impact of maternal cash transfers in conjunction with NCGs relative to NCGs alone. Both the base level of poverty and food insecurity in this sample (nearly 75% of households are below the extreme poverty line as measured at baseline) and a detailed qualitative study conducted as part of the formative work for this trial highlights that households are severely cash-constrained in providing a high-quality nutritious diet, with fruits and vegetables and animal-source foods largely unattainable (Leight et al. 2022).[6] Our experimental design will allow us to test the hypothesis – generated by our qualitative work – that relaxing this cash constraint can have a significant positive effect on infant growth.

In measurement, we employ a careful cohort-focused sampling strategy that is linked to targeted survey timing and a selection of outcome indicators designed to capture the complex age dynamics of child feeding and linear growth faltering in this context. Extensive evidence from the nutrition literature shows that linear growth faltering largely occurs during the first 24 months of life (Shrimpton et al., 2001; Victora et al., 2010) with little evidence of a systematic reversal in the first five years of life (Leroy et al., 2014). Methodologically, these same insights imply that the timing of survey measurements plays a critical role; evaluating program impacts too early increases the risk of finding a null result (King and Behrman, 2009; Linnemayr and Alderman, 2011).

While previous trials evaluate a pooled sample of households with children up to age five and define outcome variables for broader periods of child growth up to three years (Bouguen and Dillon 2022), our trial targets a sample of pregnant women and infants in a very narrow age range (under nine months) at baseline to ensure that the entire sample will be reaching key developmental milestones in conjunction with the major survey waves. At midline, all sampled children will be between the ages of six and 21 months, an ideal period to measure nutritional practices linked to age-appropriate complementary feeding. At endline, all sampled children will be between the ages of 30 and 45 months, the optimal point at which to quantify the program's impact on linear growth indicators (e.g., height-for-age) as the process of growth faltering is unlikely to be ongoing at that point, and there is limited scope for catch-up growth in this age range (Leroy et al., 2014). As such, our study design adheres to the recommended best practices in evaluating nutrition interventions: the impact of nutrition interventions on linear growth outcomes should be measured after 24 months of age (given that height captures cumulative effects), while the impact on feeding practices and health behaviors should be measured before this age (Alderman and Headey, 2018; Leroy et al., 2016).

Finally, the findings of this trial also have significant policy implications, contributing to social protection programming in Ethiopia and other low-income countries. We evaluate the impacts of a graduation model intervention within the context of an existing large-scale government-led safety net program. Launched in 2005 and providing benefits to eight million individuals, the PSNP is one of the largest safety net programs in Africa (Beegle et al., 2018). Meanwhile, SPIR itself currently serves more than 400,000 PSNP beneficiaries. While the PSNP has been successful in improving household food security and asset levels (Gilligan et al., 2009; Hoddinott et al., 2012), observational and quasi-experimental evidence suggests that its impacts on child diets and anthropometric outcomes have been negligible (Berhane et al., 2017; Berhane et al., 2015). In addition, to date, the PSNP has demonstrated limited ability to achieve sustained graduation

---

[6] Earlier work in the literature also suggests that providing cash or increasing nutritional knowledge alone may not be sufficient to improve child diets and anthropometric outcomes in low-income settings (Ahmed et al., 2019; Field and Maffioli, 2021; Levere et al., 2022).

out of long-term support (Sabates-Wheeler et al., 2021). Consequently, the lessons from this trial have the potential to inform the future design of the PSNP and, if needed, can be rapidly scaled.

2.   Research Design

Interventions

The SPIR program is implemented by World Vision International (lead), CARE, and ORDA.[7] The focus of this study is innovations in nutrition-related programming that will be implemented as part of SPIR II; these innovations 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 Nurturing Care Group (NCG) model, pioneered by World Vision in a wide range of other contexts, is based on groups of 10–15 community-based trained volunteer agents who cascade down BCC messages and activities to caregiver groups at the community level. Non-experimental studies conducted in other contexts suggest that the model can significantly increase BCC contact rates and improve IYCF practices and child growth outcomes (Davis et al., 2013). However, large-scale experimental evidence on the effectiveness of this strategy relative to standard government-led nutrition programming is largely unavailable.[8]

In addition, improving caregiver knowledge may not be sufficient to improve complementary feeding practices if households cannot afford to purchase nutritious foods. Therefore, one study arm supplements the NCGs with the introduction of maternal grants of $20 monthly for a period of 24 months to relax possible financial constraints to child feeding; the 24-month period is initiated when the NCG groups are launched. (Considering the baseline consumption estimates of around $160 monthly per household, the transfers correspond to around 13 percent of monthly consumption.) Given the sampling strategy, households in the sample may be in their final months of pregnancy or have an infant up to 9 months upon receipt of the first cash payment.

All households included in NCGs and receiving grants will also be exposed to core SPIR graduation programming. This includes the organization of village economic and social associations (VESAs), used as a platform for general financial trainings, and access to targeted value chain trainings for households entering new productive sectors. A subset of eligible households (33 percent) is also targeted for one-time $300 livelihoods grant. In addition, nutrition programming centers around the provision of integrated nutrition BCC as well as water, sanitation, and health (WASH) activities.

Evidence from a previous trial conducted as part of the first phase of SPIR shows that an integrated set of interventions including village savings' associations, training, one-time grants of $200, and nutrition behavioral change counseling delivered by community health extension workers[9] had significant medium-term effects on household savings and livestock income and some small effects on households' asset stocks over the first three years, but did not generate any statistically significant effects on household consumption, food security, or nutritional outcomes (Alderman et al., 2021). This suggests that a more

---

[7] World Vision and CARE (Cooperative for Assistance and Relief Everywhere) are international non-governmental organizations (NGO) with headquarters in the UK and Switzerland, respectively. ORDA is an Ethiopia-based NGO.
[8] Standard government-led nutrition programming in Ethiopia is delivered primarily by health extension workers and the health development army through relatively infrequent interactions to provide nutritional information and counselling to pregnant women and mothers of young children (Abate et al., 2020; Berhane et al., 2020).
[9] More specifically, the behavioral change counseling was intended to be delivered by health extension workers and health development army workers, supported by SPIR-engaged community health facilitators.

robust and explicit focus on challenges linked to infant and young child nutrition may be required to effectively shift these outcomes.

The study clusters (kebeles) are randomly assigned into three (non-overlapping) study arms. Table 1 maps the study arms against the interventions.

Table 1. Study arms and intervention packages

Study arm:	T1	T2	T3
PSNP:	X	X	X
SPIR II Interventions:			
Graduation programming		X	X
Nurturing Care Group (NCG)		X	X
Maternal grants			X

All households in the evaluation, in both the treatment and control arms, receive the core set of food or cash transfers provided under the PSNP. The PSNP is structured around the provision of six months of payments (in food or cash) to rural households as payments for labor or unconditional transfers[10] during the agricultural off-season (generally, January to June). Median annual transfers per household were around $124 during the previous phase of PSNP programming, phase 4, as completed in 2021 (Berhane et al., 2020). Apart from small-scale and geographically focused pilots, the standard PSNP programming does not include graduation programming in which households receive larger one-off cash or asset transfers coupled with livelihood and financial trainings. As such, the main role of the PSNP is to provide consumption support to beneficiary households and to improve or rehabilitate community assets via public works.

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 graduation programming and will be exposed to the NCG intervention.[11] The study arm T3 receives the same intervention package as households in T2, but also benefit from the maternal grants. In terms of broader scale, the SPIR graduation programming targets all PSNP households residing in selected PSNP districts (woredas) in the Amhara and Oromia regions. However, the NCG interventions and maternal grants are rolled out only to women and households enrolled in this evaluation.

Hypotheses

The first 1,000 days form a critical period in child growth during which growth faltering accelerates and many children in LMICs become short for their age, or stunted (Victora et al., 2010). To support growth and development during this period, WHO and UNICEF (2003) recommend caregivers follow age-appropriate infant and young child feeding (IYCF) practices. In the first six months after birth, it is recommended that children are exclusively breastfed. At around six months of age, breastmilk is no longer sufficient to fully support children's growth and development, and the introduction of complementary foods is required.

---

[10] PSNP households with limited labor capacity receive (unconditional) permanent direct support while female PSNP participants who are pregnant, or lactating are temporarily moved to temporary direct support.
[11] Note that the nurturing care groups are a novel intervention and were not present in SPIR I. Nutrition programming in SPIR I consisted of more traditional nutritional counseling provided one-on-one to households at home by health extension workers, part of the government health system.

Because of infants' and young children's limited gastric capacity, these foods need to be highly nutritious and provided frequently, and some households can be income-constrained in providing appropriate foods.

In Ethiopia, 38 percent of children under 5 years of age are stunted (CSA and ICF, 2016). Adherence to exclusive breastfeeding during the first six months of life is relatively high, particularly among children under three months of age (CSA and ICF, 2016), and as shown in Figure 1, during this period the height of the average Ethiopian child is in fact similar to the height of the median child in the WHO-2006 growth reference. Rapid growth faltering begins at around six months of age when children should be introduced to complementary foods, and continues until about 18 months of age. The overarching hypothesis evaluated in this study is that intensive nutrition programming can prevent growth faltering during the first two years of life; therefore, it is hypothesized that the HAZ curve of the average child in the treatment arms will lie above the HAZ curve of the average child in the control arm when measured after the age of two.

At the baseline, the sample is formed of PSNP households with a pregnant woman or a child less than nine months of age. We will 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 apparent in Figure 1.

Figure 1. Timing of the surveys in relation to typical linear growth faltering in Ethiopia
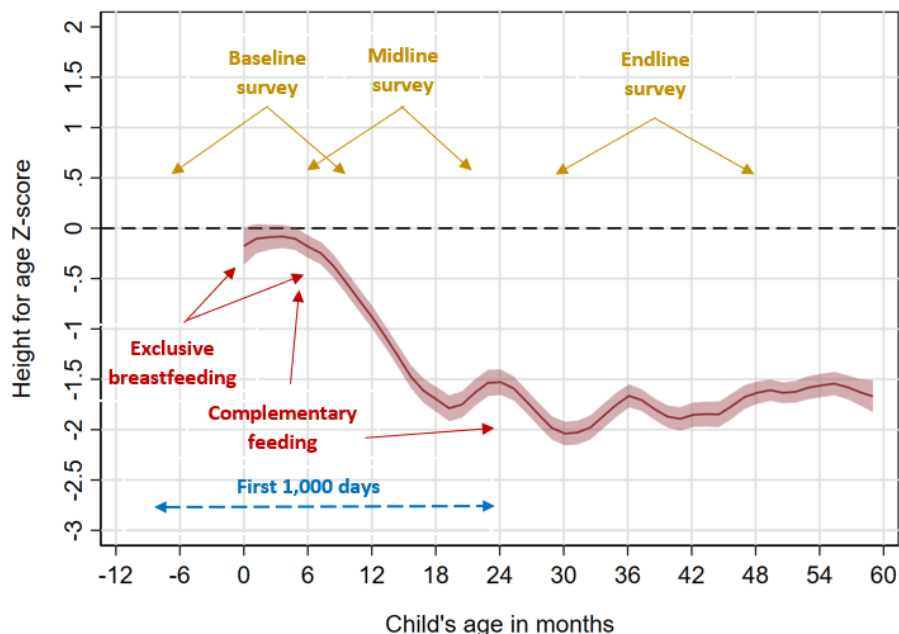
Note: Local polynomial regression based on Ethiopia 2015/16 Demographic and Health Survey (DHS). The shaded areas represent 95-% confidence intervals. N = 8,771 children 0-59 months of age. HAZ measures the height difference to the median child in the WHO-2006 growth reference sample. This difference is measured in terms of standard deviations. Thus, the HAZ of the median child in the growth reference is 0. In the graph, this is marked with the dashed horizontal line (HAZ=0).

At the same time, previous evidence from multiple evaluations of more intensive graduation models (entailing universal livelihoods transfers valued at $500 or even up to $1000) has suggested these interventions can have substantial and persistent effects on a range of livelihoods outcomes, including income (and patterns of income diversification), savings, assets, and consumption (Balboni et al., 2022; Banerjee et al., 2021). Previous evidence from the randomized controlled trial of SPIR I suggests that the reduced set of interventions implemented at scale here in both phases of programming has substantial effects on diversification into livestock production and savings, but smaller effects on assets and no detectable effect on household consumption (Alderman et al., 2021); however, this previous model did not include the nurturing care groups or maternal cash grants. The design of this new experiment will also allow us to measure the effects of the interventions of interest on these key livelihood outcomes. Given that previous evidence around effects of the core SPIR interventions on livelihood outcomes already exists in a similar context, however, we motivate our trial primarily by a focus on nutritional outcomes where our evidence will be novel; and as detailed below, the primary outcomes for the trial are all nutritional outcomes.

The theory of change for the SPIR interventions suggests a set of four statistically testable null hypotheses:

1.   Null hypothesis 1: SPIR has no effect on caregiver nutrition knowledge;

2.   Null hypothesis 2: SPIR does not lead to changes in children's diet quality; and

3.   Null hypothesis 3: SPIR does not lead to changes in children's long-term nutritional status.

For each of the above hypotheses, we also plan to evaluate a secondary null (1b, 2b, 3b): there is no additional effect of the maternal cash transfers implemented in T3, vis-à-vis the core interventions and nurturing care groups implemented in T2.

4.   Null hypothesis 4: SPIR has no effect on livelihood outcomes including savings and consumption.

For this final hypothesis, the sample in the two treatment arms will be pooled vis-à-vis the control arm in the main specification, given that there is no variation in the core livelihoods interventions across treatment arms. (A second specification will analyze the effects of each treatment arm separately.) We hypothesize that positive impacts on household consumption can occur via two channels. First, the maternal grants may result in temporary increases in food consumption levels that form most of the households' total consumption in this context. If so, these impacts may be visible at the midline (when the maternal grants are ongoing) but not at the endline (following the conclusion of the maternal cash grant payments). Second, the SPIR livelihood interventions may lead to more persistent increases in consumption levels that are visible at the endline, although, as noted above, the evaluation of SPIR-I did not detect improvements in household consumption levels.

All primary outcomes are listed in Table 2. The outcome for null hypothesis 1 is caregiver IYCF knowledge, measured through responses to a nutrition knowledge quiz administered as a part of all three household surveys. The quiz has 11 questions focusing on recommended breastfeeding and complementary feeding practices. The nutrition knowledge score is calculated as the total number of correct responses (i.e., with a minimum value of zero and a maximum value of 11). We will measure program impacts on nutrition knowledge at midline and endline.

Experimental effects on children's dietary quality (null hypothesis 2) are measured at midline using the number of food groups consumed by the child in the 24-hour period prior to the survey. 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 will group these foods into eight food 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. The variable used in the analysis to measure children's dietary quality is 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). This relatively simple indicator has been shown to be a good proxy for children's dietary quality. Various studies demonstrate how children's dietary diversity score is highly correlated with calorie and micronutrient intakes based on more comprehensive and complex food intake measures such as the quantitative 24-hour recall (Moursi et al., 2008; Ruel, 2003; Steyn et al., 2006).

Experimental effects on children's long-term nutritional status (null hypothesis 3) are measured using height for age z-scores (HAZ). In the midline and endline surveys we will collect anthropometric measures – length/height and weight – from all index children.[12] Lengths/heights and weights will then be converted to z-scores using the 2006 WHO growth standards (de Onis et al., 2007; WHO, 2006). 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 length-for-age z-score below -2.0, and this chronic undernutrition reflects the negative effects of continued inadequate food intake together with repeated infection. To account for the fact that linear growth faltering is a long-term process, as a primary outcome we will consider the HAZ as measured at endline.

Impacts on livelihoods (null hypothesis 4) are assessed using secondary outcomes, described below.

Table 2. Primary outcomes

Outcome	Variable type	Timing of measurement
Caregiver IYCF knowledge	Continuous	Midline, endline
Number of food groups consumed by a child 6–23 months	Continuous	Midline
Height-for-age (children 30-48 months)	Continuous	Endline

We also measure the impact on a range of secondary outcomes that are listed in Table 3. For nutritional outcomes, we consider child stunting (a binary indicator obtaining value 1 if HAZ<-2 and zero otherwise) as a complementary measure. We prefer to use a continuous measure (HAZ) over a binary (stunting) one because of gains in statistical power (Royston et al., 2006) and because the biological grounds for the HAZ < -2 threshold are not well justified (Perumal et al., 2018). Moreover, we measure the impacts on HAZ also at midline. However, at this point, children are between 6-23 months of age and therefore, the process of linear growth faltering is still likely to be ongoing for many children in our sample.

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, 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

---

[12] We did not collect anthropometric outcomes in the baseline survey because a sizable share of the cohort of children that we follow had not yet been born. As discussed below, the sampling frame for the baseline survey consisted of households with a pregnant woman or a child under 9 months of age.

to detect impacts on this outcome. Considering these factors, WHZ as measured at the midline and endline is considered as a secondary outcome.

As a complementary measure of diet quality, we also assess the impact on minimum acceptable diet (MAD), a combination of sufficient meal frequency and diet diversity. We will construct this indicator following the WHO and UNICEF (2021) guidelines. In this context, very few children receive MAD (Berhane et al., 2020; CSA and ICF, 2016). Since the main constraint relates to dietary diversity (CSA and ICF, 2016), we focus on dietary diversity as our primary outcome measure instead of MAD.

The early childhood development (ECD) score will allow us to capture the effects of the interventions of interest on early childhood development. Recent evidence from Niger suggests a behavioral change communication intervention with a broad focus on nutrition and parenting practices did significantly enhance child development (without any shifts in anthropometrics) (Premand and Barry, 2022). In this trial, the NCG intervention includes relatively minimal content explicitly focused on early childhood development, though there is some; it is also possible that feeding and caretaking practices promoted by the NCGs (i.e., more active parental engagement in early feeding) could have spillover effects in broader parent-child interactions and thus shape ECD. We plan to measure ECD at endline using the Strengths and Difficulties' Questionnaire (SDQ), also used by Premand and Barry; this is an instrument that has previously been used in Ethiopia (Escueta et al., 2014; Hoosen et al., 2018), but has not to our knowledge be explicitly validated for this context.

We also include a number of secondary livelihoods-related outcomes. The first is household consumption (measured following the guidelines in Deaton and Grosh, 2000; Deaton and Zaidi, 2002). Other outcomes of interest include a total asset index;[13] the level of household savings (both a binary variable for reporting any savings, and a continuous variable); food security (measured using the Food Insecurity Experience Scale; FIES); net income from livestock and non-agricultural production; an index of household livestock assets (constructed using Tropical Livestock Units); and credit access. The measurement of net income from livestock and non-agricultural production will allow us to assess the effects of the bundled livelihoods interventions in stimulating household entry into new productive activities, particularly livestock production (a preferred livelihood activity in this sample); and the measurement of credit access will allow us to capture the effects of VESA membership on use of credit.

---

[13] Based on 22 durable goods, 15 productive assets and 19 types of livestock, aggregated into single value using a principal components analysis method (Sahn and Stifel, 2003).

Table 3. Secondary outcomes

Outcome	Variable type	Timing of measurement
Nutrition and child development		
Stunting (children 30-45 months)	Binary	Midline, endline
Height-for-age (children 6-23 months)	Continuous	Midline
Weight-for-height Z-score (children 6-23 months)	Continuous	Midline
Weight-for-height Z-score (children 30-48 months)	Continuous	Endline
Percent of children 6–23 months receiving a minimum acceptable diet (MAD)	Binary	Midline
Early childhood development score	Continuous	Endline
Livelihoods		
Household per capita consumption	Continuous	Midline, endline
Total asset index	Continuous	Endline
Household savings	Binary and continuous	Endline
Food security (FIES)	Binary and continuous	Midline, Endline
Net income from livestock production	Binary and continuous	Endline
Net income from any non-agricultural production	Binary and continuous	Endline
Household livestock asset index	Continuous	Endline
Credit access	Binary and continuous	Endline

In general, we plan to interpret the analysis primarily as assessing the effect of the nutritional interventions (nurturing care groups and maternal grants) on nutritional outcomes, while assessing the effect of the livelihood interventions implemented in the same experimental arms on livelihoods outcomes. However, it is also important to note that there could be an effect of the livelihood interventions directly on nutrition interventions: i.e., a household that has a more stable income stream and more consumption expenditure available may direct some of this consumption to an infant or young child in the form of a more diverse diet.

While our evaluation design will allow us to shed some light on these effects, previous evidence from a randomized trial conducted in a similar population suggests that effects on livelihood outcomes are not likely to be large (will not exceed 0.1 standard deviations) and thus their contribution to IYCF outcomes may be unlikely.[14] In particular, previous evidence from the SPIR I trial suggests there was no significant effect of bundled livelihoods interventions on household-level consumption even in the subsample where all households received $200 transfers; accordingly, given that only a 33 percent subsample of our pooled sample will receive lump-sum transfers (as described in more detail below), the probability of a significant effect on consumption may be low (Alderman et al. 2021). In addition, in the previous trial, the only positive

---

[14] In the SPIR I trial, the effects of the pooled set of livelihoods interventions in a sample of households that all received $200 transfers ranged between 0.1 and 0.2 standard deviations for assets, savings, and livestock income outcomes. Given that only 33% of households in this sample will receive transfers (albeit of a larger size), it is plausible to expect that the previously estimated effects correspond to an upper bound.

effect observed for the subsample of households that did not receive transfers was increased household savings.

To further substantiate the hypothesis that the primary channel for any positive effects on nutrition-related outcomes is the effects of the nurturing care groups (as distinct from any livelihoods-related interventions), we will conduct two additional tests. First, we will examine the effects of the bundled interventions on IYCF and anthropometric outcomes within the subsample (66 percent of households) that does not receive any livelihood transfers. As noted above, these households are not likely to show any meaningful effects on livelihoods outcomes other than savings, and thus the scope for any ancillary effects on IYCF outcomes is minimal. Second, we will test the hypothesis that the effects of SPIR I (a pooled treatment indicator) and SPIR II (T2) on primary nutritional outcomes (child diet diversity and height-for-age) are equal, stacking both datasets, interacting with a trial dummy (SPIR-I vs SPIR-II) and testing if the treatment coefficients are identical.[15] If this hypothesis is rejected, this is suggestive evidence that any positive effects on nutritional outcomes are attributable to the nurturing care groups (novel in SPIR II) as opposed to the broader package of livelihoods-related interventions (present in both trials).

Identification strategy

The evaluation is a cluster RCT where the cluster is defined as the *kebele* (subdistrict, the lowest administrative level in Ethiopia). We follow a cohort of PSNP households and their infants for a three-year period, encompassing the critical period during which linear growth faltering is common in Ethiopia. Kebeles will be randomly allocated to different study arms, permitting us to identify the causal impact of SPIR programming along the theory of change on caregiver knowledge, child diets, anthropometric outcomes, and livelihood outcomes.

Sample and statistical power

The sampling criteria for the evaluation were as follows.

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

The sample was constructed using PSNP beneficiary lists at the household level; these lists were randomly ordered at the level of the gotte (village) within kebeles, and households were screened following this random order to identify whether they met the criteria above. The target sample was 13 households per kebele, including seven households including a pregnant woman and six households including an infant.

The target sample included 3,081 households in 237 kebeles. The realized sample at baseline included 3,015 households in 234 kebeles (see Table A1 in the appendix). Three kebeles were excluded from the sample due to ongoing conflict or civil unrest that rendered them inaccessible during the baseline survey.

---

[15] Both outcomes, child dietary diversity and height-for-age, were measured using comparable methodologies in both trials.

Within the remaining kebeles, the sampling targets were achieved, as summarized in the sample composition table provided in Table A1 in the appendix.

Using this sample, the unit of analysis is either a young child or his/her mother or their household observed in the baseline, midline and/or endline.

We conducted power calculations using the Stata command *power twomeans*, setting the significance level at 5 percent and power at 80 percent, and allowing for 10 percent attrition between baseline and endline surveys. Power calculations are reported only for the primary (nutrition-related) outcomes and for a subset of the livelihoods-related outcomes that have been designated as secondary.

Using data from the control arm in the SPIR-1 endline survey conducted in 2021, we estimated the mean, standard deviation, and intra-cluster correlation for each outcome; this information is summarized in Table 4a. We also use estimates derived from the SPIR I trial of the autocorrelation in outcomes and the predictive power of woreda-level fixed effects to adjust for the reduced variance in outcomes when including controls for baseline levels and woreda fixed effects.[16]

For the nutrition-related outcomes, we estimate statistical power corresponding to the comparison of each treatment arm to the control arm; the evaluation is able to detect a 0.12–food group improvement in children's dietary diversity, a 0.23–unit change in height-for-age z-score (0.14 SD), and a 0.26-unit improvement in IYCF knowledge (0.23 SD).[17] These minimum detectable effect (MDE) sizes are biologically meaningful and in line with findings in previous trials. First, our MDEs for children's dietary diversity and caregiver nutrition knowledge are theoretically meaningful. Effects smaller than these MDEs are unlikely to have any positive effects on downstream measures of children's health (most importantly, anthropometric measurements) and thus unlikely to be of interest from a welfare perspective; our study is thus adequately powered to detect any effect that is meaningful from the perspective of children's long-term health. Second, the MDE of 0.14 SD for height-for-age z-score is near-identical to the one reported in Bouguen and Dillon (2019),[18] and is similar to the impacts of cash transfers in previous research (see Table 4 in Bouguen and Dillon, 2022).

For the livelihood related outcomes, the main analysis will pool data from the two treatment arms vis-à-vis 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). As for savings and consumption, we estimate the following MDEs: 9-percentage point increase in the probability that households report any savings (relative to an estimated mean probability of 47 percent in the control arm), and a 14-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; however, in this evaluation, there could be a direct effect of the maternal cash grants on consumption as observed at midline.

---

[16] We follow the guidance provided by McKenzie in point 4 in this recent blog post in adjusting for the reduced variance of the outcomes of interest conditional on controls:
https://blogs.worldbank.org/impactevaluations/six-questions-about-doing-power-calculations.
[17] Given that our power calculations are informed by very recent data from a very similar population, we also conjecture they may be more precise than general power calculations that are estimated with typical population parameters rather than sample-specific estimates as in Bouguen and Dillon (2022).
[18] The MDE of height-for-age in Bouguen and Dillon (2019) is 15.2% of SD.

The power calculations reported in Table 4a were conducted prior to the baseline survey. In Table 4b, we update these power calculations using the 2022 SPIR-II baseline data for the control arm, with the caveat that the baseline did not (yet) measure child dietary diversity or anthropometric outcomes, nor can we control for baseline levels when estimating the adjusted SD. The MDEs relative to SDs are generally similar to those reported in Table 4a. However, for per capita consumption, the updated power calculations suggest that we are considerably better powered: the adjusted MDE indicates that we can detect a 7-percent increase in per capita consumption (instead of 14-percent as reported in Table 4a).

Table 4a. Power calculations based on SPIR-I endline data

	Nutrition outcomes			Livelihood outcomes		
	Caregiver IYCF knowledge	**Number of food groups consumed (children 6–23 months)**	**Height-for-age (children 30-48 months)**	**Total asset index**	**Household savings, binary**	**Log per capita consumption**
Control clusters	79	79	79	79	79	79
Treatment clusters [1]	79	79	79	158	158	158
Baseline mean	3.9	1.9	-2.07	-0.14	0.45	0.71
Baseline standard deviation (SD)	1.27	0.96	1.59	1.89	0.50	0.75
Adjusted SD [2]	1.17	0.92	1.55	1.56	0.42	0.68
Intra-cluster correlation (ICC)	0.24	0.001	0.02	0.26	0.27	0.22
Adjusted MDES	0.264	0.119	0.222	0.342	0.093	0.142
MDE relative to SD	0.23	0.13	0.14	0.22	0.22	0.21

Note: Baseline means, SDs and ICCs estimated using the SPIR-1 endline data. [1] Pooled treatment arms for livelihood outcomes [2] Adjusted for the reduced variance in outcomes when including controls for baseline levels (where applicable) and woreda fixed effects.

Table 4b. Power calculations based on the SPIR-II baseline data

	Nutrition outcomes			Livelihood outcomes		
	Caregiver IYCF knowledge	**Number of food groups consumed (children 6–23 months)**	**Height-for-age (children 30-48 months)**	**Total asset index**	**Household savings, binary**	**Log per capita consumption**
Control clusters	79	n/a	n/a	79	79	79
Treatment clusters [1]	79	n/a	n/a	158	158	158
Baseline mean	8.5	n/a	n/a	-0.02	0.37	0.45
Baseline standard deviation (SD)	2.26	n/a	n/a	1.74	0.48	0.48
Adjusted SD [2]	1.88	n/a	n/a	1.60	0.44	0.46
Intra-cluster correlation (ICC)	0.32	n/a	n/a	0.19	0.23	0.09
Adjusted MDES	0.511	n/a	n/a	0.314	0.093	0.072
MDE relative to SD	0.27	n/a	n/a	0.20	0.21	0.16

Note: Baseline means, SDs and ICCs estimated using the (cross-sectional) baseline data. [1] Pooled treatment arms for livelihood outcomes; [2] Adjusted for the reduced variance in outcomes when including woreda fixed effects. n/a = these outcomes were not yet measured at the baseline.

5. Data

Surveys

Table 5 summarizes the timeline, particularly for the three primary rounds of data collection. Randomization was conducted by the research team using Stata in June 2022.[19] The baseline survey was conducted between August and October 2022, and the NCGs were launched shortly after in November, 2022. The midline and the endline surveys will take place exactly one and three years after the baseline survey, respectively. In these follow-up surveys, all households interviewed at the baseline will be visited.[20] All data will be collected using Computer-Assisted Personal Interviews (CAPI) methods.

Table 5. Timing of randomization, surveys, and intervention

#	Action	Month, Year
1	Randomization of kebeles into study arms	June, 2022
2	Baseline survey	August-October, 2022
4	Intervention begins	November, 2022
3	Midline survey	August-October, 2023
4	Endline survey	August-October, 2025

The timing of the survey rounds plays a critical role in assessing the nutritional impacts of the interventions of interest (Figure 2). The midline survey is optimized to allow for (i) timely measurement of the impact of SPIR II on child diets and IYCF practices, (ii) inclusion of process monitoring questions around household participation in the intervention (iii) measurement of consumption while the maternal cash grant is ongoing. The endline survey is planned for 2025, when the children are 30-47 months, to measure child growth outcomes after the main period of growth faltering has passed.

Figure 2. Timing of the surveys

---
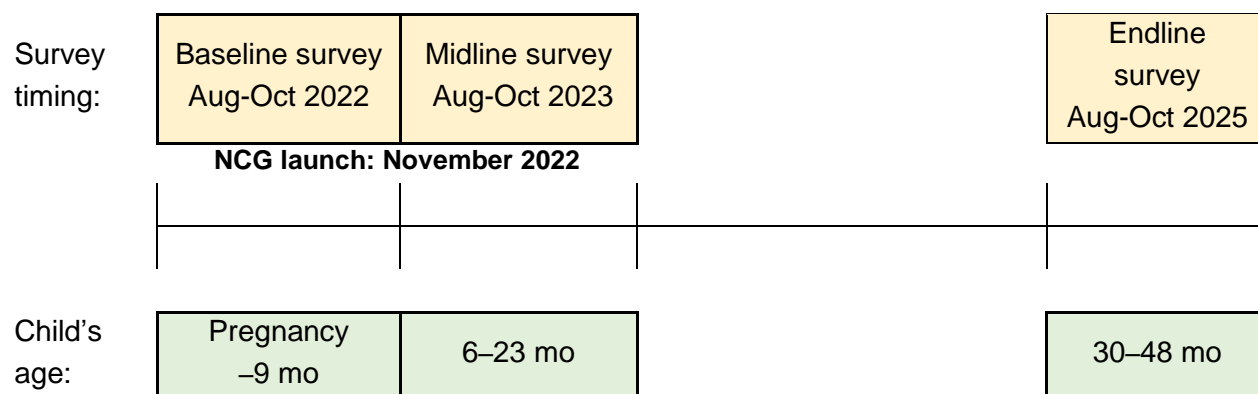
[19] Strata were constructed based on the interaction of the following characteristics: woreda; a binary variable for whether a kebele is above or below the woreda-level median in the percentage of households eligible for the PSNP; and a binary variable for whether the kebele is above or below the woreda-level median in distance from the woreda capital).
[20] Relative to the baseline and endline surveys, the midline survey will be less extensive in scope. In the baseline and endline surveys, the interviews will be conducted with both the primary female respondent (pregnant or lactating woman at the baseline) and her spouse. In the midline survey, the primary respondent will be the primary female only.

19

Baseline characteristics and balance

Table A2 in the appendix reports basic demographic characteristics of the sample households included in our baseline survey. The average household size is five, and 89 percent of households report that the head of the household is male, characterized by an average age of 35. More than half of the household heads report some formal education, and 87 percent report that their primary economic activity is crop production. 93 percent of household heads are married.

The primary female in the sample households (defined as the pregnant or lactating mother) is on average 29 years of age, and 93 percent are married. Only 43 percent of primary females report any formal education, and on average the level of education is low: only 26 percent report three to seven years of education, and only seven percent report more than eight years of education.

While all households are PSNP beneficiaries in 2022 given the sample criteria, 12 percent report that they are receiving temporary direct support for some part of the year (i.e., receiving transfers without a labor requirement); 96 percent report that they are public works beneficiaries (subject to a labor requirement).[21] It is important to note that due to the substantial retargeting process observed at the initiation of the PSNP5 period, the majority of these households are new to the PSNP and did not previously receive any benefits; only 33 percent report previously receiving PNSP benefits in 2021, the final year prior to the recent retargeting. The estimated prevalence of extreme poverty (consumption under $1.90 a day) is 71 percent.

Table 6 reports the baseline balance of key household characteristics across treatment arms. Out of the 36 t-tests comparing differences across treatment arms, only one is statistically significant at the five percent level. We conducted a joint test across all these outcomes to test the hypothesis that the observable characteristics are generally balanced across experimental arms T1, T2, and T3 across all characteristics reported. The p-values for these joint tests are between 0.665 and 0.888, meaning that the hypothesis that the characteristics are jointly balanced cannot be rejected. We therefore conclude that the randomization was successful in generating balance across key household characteristics.

Data management plan

Confidentiality of the data is protected by recording survey interview responses using CAPI, and thus no hard copy versions of survey questionnaires will be available. All files containing raw and analyzed data are securely stored in password-protected and encrypted databases. Access to the complete data will be restricted to the research team. A unique household ID is assigned to each household. The name, contact information, and geographic location of the respondent will be kept in a separate data file to which only the key members of research team will have access (the Research Analyst, and Principal Investigators). Anonymized versions of the data sets that exclude these personal identifiers will be used for analysis and made available for public access after the evaluation has ended. PII will be deleted one year following the closure of the evaluation and all related grant activities, scheduled to conclude in 2026. No stopping rule is specified for this trial.

---

[21] These two categories are not mutually exclusive, given that some households may receive direct support for part of the year, typically during pregnancy and a child's first year of life, but otherwise receive public works benefits.

Table 6. Balance of household characteristics across treatment arms

Variable	(1) T1 Mean/[SE]	(2) T2 Mean/[SE]	(3) T3 Mean/[SE]	t-test p-value (1)-(2)	t-test p-value (1)-(3)	t-test p-value (2)-(3)
Pregnancy / infant status:						
Currently pregnant (0/1)	0.529	0.538	0.533	0.161	0.571	0.487
	[0.005]	[0.005]	[0.007]			
Child is 0-3 months old (0/1)	0.214	0.180	0.192	0.047**	0.250	0.472
	[0.011]	[0.010]	[0.014]			
Child is 4-6 months old (0/1)	0.142	0.145	0.140	0.970	0.853	0.803
	[0.011]	[0.011]	[0.010]			
Child is 7-9 months old (0/1)	0.113	0.136	0.133	0.157	0.186	0.876
	[0.010]	[0.012]	[0.010]			
Household size	5.407	5.375	5.401	0.581	0.538	0.287
	[0.118]	[0.116]	[0.122]			
Primary female has some formal education (0/1)	0.444	0.423	0.428	0.633	0.527	0.878
	[0.021]	[0.019]	[0.021]			
Mother's IYCF knowledge score (max 11 points)	8.454	8.522	8.327	0.412	0.565	0.889
	[0.154]	[0.157]	[0.167]			
Food Insecurity Experience Scale (FIES), raw score	6.151	6.219	6.034	0.919	0.427	0.203
	[0.112]	[0.095]	[0.110]			
Household daily per capita consumption in 2011 PPP-USD	1.767	1.766	1.794	0.509	0.457	0.802
	[0.043]	[0.046]	[0.055]			
Tropical Livestock Units owned by the household	11.666	11.510	12.238	0.471	0.190	0.461
	[0.696]	[0.714]	[0.744]			
Durable asset index based Principal Components Analysis (PCA)	0.000	-0.079	0.079	0.640	0.410	0.203
	[0.062]	[0.063]	[0.100]			
Household reports formal savings (0/1)	0.366	0.378	0.390	0.840	0.585	0.835
	[0.029]	[0.031]	[0.029]			
F-test for joint significance of all balance variables (p-value)				0.665	0.804	0.888
N	1,033	992	990			
Clusters	80	77	77			

Notes: Estimates from the IMPEL baseline survey sample. The value displayed for t-tests are p-values. Standard errors were clustered at the kebele level. Fixed effects at the woreda level were included in all estimation regressions for the t-tests. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. 0/1 = binary indicator. Savings in rural savings and credit cooperatives (RuSACCOs), Micro-Finance Institutions, Village Savings and Loan Associations (VSLA) or bank accounts were considered as formal savings.

6.   Analysis

We will analyze the primary and secondary outcomes (summarized in Tables 2 and 3 above) using an analysis of covariance (ANCOVA) estimation strategy (McKenzie, 2012). The specifications estimated for any variables that have a reported baseline value can be written as follows:

(1) $$Y_{ikd,t=1} = \beta_1 T_{kd}^2 + \beta_2 T_{kd}^3 + \gamma Y_{ikd,t=0} + X'_{ikd,t=0}\vartheta + \chi_{kd} + \varepsilon_{ikd};$$ and

(2) $$Y_{ikd,t=1} = \beta T_{kd} + \gamma Y_{ikd,t=0} + X'_{ikd,t=0}\vartheta + \chi_{kd} + \varepsilon_{ikd},$$

$Y_{ikd,t=1}$ captures the outcome of interest in household *i* residing in kebele (sub-district) *k* and woreda (district) *d* at midline/endline *t* and $Y_{ikd,t=0}$ at baseline. Variables T_{kd}^2 and T_{kd}^3 are time-invariant indicator variables, receiving value of 1 if the kebele is randomly assigned to T2 or T3, respectively, and zero otherwise; T_{kd} is a pooled treatment indicator; χ_{kd} are strata fixed effects. Both equations will be estimated with and without baseline controls ($X'_{ikd,t=0}$), including household size, age and education level of the primary caregiver, and age and sex of the child.

We will estimate specification (1) for nutrition-related variables (assuming a baseline value is available), and report the p-value for the null hypothesis that the treatment effects are consistent across treatment arms, β₁ = β₂. For livelihoods-related variables, we will first analyze specification (2) pooling data across the two treatment arms; as a robustness check, we will estimate specification (1) to test whether there is any differential effect of the maternal grants on livelihoods outcomes.

In addition, for outcomes for which baseline value is not available (e.g., anthropometric measures), equation (1) will be estimated without $Y_{ikd,t=0}$ as in the following specification.

(3) $$Y_{ikd,t=1} = \beta_1 T_{kd}^2 + \beta_2 T_{kd}^3 + X'_{ikd,t=0}\vartheta + \chi_{kd} + \varepsilon_{ikd}.$$

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

All regressions will be estimated using the Ordinary Least Squares (OLS) method. Following the recommendation by Abadie et al. (2022), the standard errors in our regressions will be clustered at the kebele level to account for the randomized design. The cluster-robust standard errors will be computed using Stata's *vce(cluster)* command that adjusts the standard errors based on the Liang and Zeger (1986) approach. All statistical analyses will be conducted using Stata, version 17 or higher.

In addition to reporting standard p-values, we will also report p-values corrected for multiple hypothesis testing based on sharpened FDR (false discovery rate) q-values (Anderson, 2008). This correction will be implemented within the set of primary outcomes, and within the set of secondary outcomes.

Missing data from item non-response

There will be no imputation for missing data due to item non-response at midline or endline. Missing data on baseline variables will be set to zero and dummied out in the ANCOVA specifications (i.e., equations 1 and 2). When measuring the impact on anthropometric outcomes, we will omit children with biologically implausible values: those below -6.0 and above 6.0 for HAZ and those below -5.0 and above 5.0 for WHZ.

Heterogeneous effects

In addition to the analysis of pooled treatment effects, we will report heterogeneous treatment effects along certain pre-specified dimensions.[22] These analyses should be considered to be exploratory.

The primary 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. One recent study found that the effects of targeting fathers for nutrition-related behavioral change communication as well as mothers did not lead to any differentially larger shifts in child feeding outcomes (Han et al., 2022). However, given evidence from formative work that maternal time constraints in PSNP households can be a meaningful barrier to implementing optimal feeding practices (and these constraints could be potentially alleviated by paternal caretaking), we are particularly interested in exploring male engagement in caretaking as distinct from knowledge alone (Leight et al., 2023).

The second dimension of heterogeneity is exposure to SPIR cash grants. As previously noted, the poorest third of households in treatment communities are eligible for one-time $300 livelihood grants. Eligibility is determined based on an asset score constructed using baseline data on assets, and this eligibility determination was made uniformly by the research team for all sampled communities; accordingly, we can identify households who are and are not eligible for grants in both treatment and control communities. The effects of the bundled graduation model interventions on livelihoods outcomes (other than savings) is plausibly very small for the set of households (66 percent) who do not receive grants, given previous evidence. Accordingly, assessing the effects of NCGs and maternal grants in this subsample will allow us to minimize the role of any potential spillovers from livelihoods outcomes, and more credibly estimate the effects of NCGs and maternal cash grants alone on nutritional outcomes of interest.

Spillover effects

We can identify two channels through which households that are not directly exposed to SPIR programming can be affected by the program. First, the one-off livelihood grants and the monthly maternal cash grants could theoretically lead to food and non-food price inflation in local markets resulting in harmful effects on non-beneficiaries (Cunha et al., 2019; Egger et al., 2022; Filmer et al., 2021). In our study context, the program saturation with respect to livelihood transfers is low: in each kebele only five percent of households, on average, are eligible for these transfers. As the maternal grants are only provided to households enrolled in this evaluation, the saturation levels are even lower. Consequently, we consider the risk of such negative spillover effects as negligible. Second, caregivers in the control kebeles may interact with caregivers exposed to the NGC programming. Considering our cluster-RCT design in which each kebele is formed of multiple villages, we anticipate that such learning spillover effects will be small. Nevertheless, we plan to monitor this by carefully analyzing the changes in caregiver knowledge in the control kebeles across survey

---

[22] We also considered exploring heterogeneity with respect to child age to see if longer exposure to BCC or maternal cash transfers during the 1,000-day window would result in stronger positive impacts on IYCF or child anthropometrics. However, since we do not have a staggered rollout of the intervention, it would be impossible to disentangle treatment exposure effects from age or time effects (e.g., season of birth). Therefore, we do not explore this dimension of heterogeneity here.

rounds. In the midline survey, we will also include qualitative questions about interactions between caregivers within and across kebele boundaries.

7. Limitations and Challenges

We emphasize that the broader interpretation of the treatment effects in our trial is conditional on the receipt of PSNP consumption support by households in all three treatment arms. We lack the counter-factual to what would be observed if the control group did not benefit from the PSNP. Therefore, while our findings can be informative about extending safety net programs to include similar livelihood and nutrition components, we should be cautious about exporting these lessons to contexts without an existing safety net program.

As for challenges, perhaps the primary risk for the program implementation and for this evaluation is the re-emergence of the conflict that unfolded in Northern Ethiopia in 2020-2022. The main warring parties signed a peace accord in November 2022 and the situation has remained calm since then. A possible re-emergence of the conflict would likely affect the implementation of the SPIR program in the Amhara region; moreover, we will not send survey teams to any active conflict areas. Both the study team and the implementation partners are actively monitoring the situation and will act accordingly to ensure the safety of the implementation and survey staff. However, as of March 2023, the re-emergence of the conflict in Northern Ethiopia seems unlikely.

Bibliography

Abadie A, Athey S, Imbens GW, Wooldridge J. When should you adjust standard errors for clustering? Quart. J. Econ. 2022;138; 1–35.

Abate GT, Dereje M, Hirvonen K, Minten B. Geography of public service delivery in rural Ethiopia. World Development 2020;136.

Ahmed A, Hoddinott J, Roy S. Food transfers, cash transfers, behavior change communication and child nutrition: Evidence from Bangladesh. IFPRI Discussion Paper, vol. 1868. International Food Policy Research Institute (IFPRI); 2019.

Alderman H, Billings L, Gilligan DO, Hidrobo M, Leight J, Taffesse AS, Tambet H. Impact evaluation of the strengthen PSNP4 institutions and resilience (SPIR) development food security activity (DFSA): Endline report. International Food Policy Research Institute (IFPRI): Washington D.C.; 2021.

Alderman H, Gilligan DO, Leight J, Mulford M, Tambet H. The role of poultry transfers in diet diversity: A cluster randomized intent to treat analysis. Food Policy 2022;107; 102212.

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

Alderman H, Hoddinott J, Kinsey B. Long term consequences of early childhood malnutrition. Oxford Economic Papers-New Series 2006;58; 450-474.

Anderson ML. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. J. Amer. Statistical Assoc. 2008;103; 1481-1495.

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

Balboni C, Bandiera O, Burgess R, Ghatak M, Heil A. Why do people stay poor? The Quarterly Journal of Economics 2022;137; 785-844.

Bandiera O, Burgess R, Das N, Gulesci S, Rasul I, Sulaiman M. Labor markets and poverty in village economies. The Quarterly Journal of Economics 2017;132; 811-870.

Banerjee A, Duflo E, Goldberg N, Karlan D, Osei R, Parienté W, Shapiro J, Thuysbaert B, Udry C. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. Science 2015;348; 1260799.

Banerjee A, Duflo E, Sharma G. Long-term effects of the targeting the ultra poor program. American Economic Review: Insights 2021;3; 471-486.

Banerjee A, Karlan D, Osei R, Trachtman H, Udry C. Unpacking a multi-faceted program to build sustainable income for the very poor. Journal of Development Economics 2022;155; 102781.

Bedoya G, Coville A, Haushofer J, Isaqzadeh M, Shapiro JP. No household left behind: Afghanistan targeting the ultra poor impact evaluation. National Bureau of Economic Research; 2019.

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

Berhane G, Golan J, Hirvonen K, Hoddinott J, Kim S, Taffesse AS, Abay K, Assefa T, Habte Y, Hiluf M, Koru B, Tadesse F, Yimer F. Evaluation of the Nutrition-sensitive Features of the Fourth Phase of Ethiopia's Productive Safety Net Programme. IFPRI-ESSP Working Paper, vol. 140. International Food Policy Research Institute (IFPRI), Ethiopia Strategy Support Program (ESSP): Addis Ababa; 2020.

Berhane G, Hirvonen K, Hoddinott J, Kim S, Taffesse AS, Abay K, Hiluf M, Koru B, Assefa T, Yimer F. Evaluation of the Nutrition Sensitive Features of the Productive Safety Nets Programme IV: Baseline Survey Report. Unpublished report. 2017.

Berhane G, Hoddinott J, Kumar N, Margolies A. The impact of Ethiopia's Productive Safety Net Programme on the nutritional status of children: 2008–2012. IFPRI-ESSP working paper, vol. 99. International Food Policy Research Institute (IFPRI): Washington D.C.; 2015.

Bossuroy T, Goldstein M, Karimou B, Karlan D, Kazianga H, Parienté W, Premand P, Thomas CC, Udry C, Vaillant J. Tackling psychosocial and capital constraints to alleviate poverty. Nature 2022;605; 291-297.

Bouguen A, Dillon A. The impact of a multidimensional program on nutrition and poverty in Burkina Faso. Journal of Development Economics 2019; Pre-results review.

Brune L, Karlan D, Kurdi S, Udry C. Social protection amidst social upheaval: Examining the impact of a multi-faceted program for ultra-poor households in Yemen. Journal of Development Economics 2022;155; 102780.

CSA, ICF. Ethiopia Demographic and Health Survey 2016. Central Statistical Agency (CSA) of Ethiopia and ICF: Addis Ababa, Ethiopia, and Rockville, Maryland, USA; 2016.

Cunha JM, De Giorgi G, Jayachandran S. The price effects of cash versus in-kind transfers. The Review of Economic Studies 2019;86; 240-281.

Currie J, Vogl T. Early-Life Health and Adult Circumstance in Developing Countries. Annual Review of Economics 2013;5; 1-36.

Davis TP, Wetzel C, Avilan EH, de Mendoza Lopes C, Chase RP, Winch PJ, Perry HB. 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 2013;1; 35-51.

de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull. World Health Organ. 2007;85; 660-667.

Deaton A, Grosh M. Consumption. In: Grosh M, Glewwe P (Eds), Designing Household Survey Questionaires for Developing Countries: Lessons from 15 Years of Living Standards Measurement Study, vol. 1. World Bank: Washington D.C.; 2000. p. 91-133.

Deaton A, Zaidi S. Guidelines for constructing consumption aggregates for welfare analysis. World Bank Publications; 2002.

Dercon S, Porter C. Live Aid Revisited: Long-term Impacts of the 1984 Ethiopian Famine on Children. Journal of the European Economic Association 2014;12; 927-948.

Egger D, Haushofer J, Miguel E, Niehaus P, Walker MW. General equilibrium effects of cash transfers: experimental evidence from Kenya. Econometrica 2022;90.

Escueta M, Whetten K, Ostermann J, O'Donnell K. Adverse childhood experiences, psychosocial well-being and cognitive development among orphans and abandoned children in five low income countries. BMC Int. Health Hum. Rights 2014;14; 1-13.

FAO, IFAD, UNICEF, WFP, WHO. The State of Food Security and Nutrition in the World 2022: Repurposing food and agricultural policies to make healthy diets more affordable. FAO, IFAD, UNICEF, WFP and WHO: Rome; 2022.

Field EM, Maffioli EM. Are behavioral change interventions needed to make cash transfer programs work for children? Experimental evidence from Myanmar. National Bureau of Economic Research; 2021.

Filmer D, Friedman J, Kandpal E, Onishi J. Cash Transfers, Food Prices, and Nutrition Impacts on Ineligible Children. The Review of Economics and Statistics 2021; 1-45.

Gilligan DO, Hoddinott J, Taffesse AS. The impact of Ethiopia's Productive Safety Net Programme and its linkages. The Journal of Development Studies 2009;45; 1684-1706.

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

Han Y, Park S, Kim J, Hoddinott J. Engaging fathers through nutrition behavior communication change does not increase child dietary diversity in a cluster randomized control trial in rural Ethiopia. The Journal of Nutrition 2022;In press.

Hidrobo M, Hoddinott J, Kumar N, Olivier M. Social Protection, Food Security, and Asset Formation. World Development 2018;101; 88-103.

Hoddinott J, Berhane G, Gilligan DO, Kumar N, Taffesse AS. The impact of Ethiopia's Productive Safety Net Programme and related transfers on agricultural productivity. Journal of African Economies 2012;21; 761-786.

Hoosen N, Davids EL, de Vries PJ, Shung-King M. The Strengths and Difficulties Questionnaire (SDQ) in Africa: a scoping review of its application and validation. Child and adolescent psychiatry and mental health 2018;12; 1-39.

King EM, Behrman JR. Timing and duration of exposure in evaluations of social programs. The World Bank Research Observer 2009;24; 55-82.

Leight J, Alderman H, Gilligan DO, Hidrobo M, Mulford M, Tadesse E. Barriers to adoption of optimal complementary feeding practices in Ethiopia: A formative qualitative investigation: Evidence from SPIR II. SPIR learning brief, vol. 7. International Food Policy Research Institute (IFPRI): Washington D.C.; 2023.

Leroy JL, Olney DK, Ruel M. Evaluating nutrition-sensitive programs: challenges, methods, and opportunities. In: Covic N, Hendriks SL (Eds), Achieving a nutrition revolution for Africa: the road to healthier diets and optimal nutrition. International Food Policy Research Institute: Washington D.C.: ; 2016. p. 130-146.

Leroy JL, Ruel M, Habicht J-P, Frongillo EA. Linear growth deficit continues to accumulate beyond the first 1000 days in low-and middle-income countries: global evidence from 51 national surveys. The Journal of Nutrition 2014;144; 1460-1466.

Levere M, Acharya G, Bharadwaj P. The Role of Information and Cash Transfers in Early Childhood Development: Short and Long Run Evidence from Nepal. Economic Development and Cultural Change 2022;in press.

Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73; 13-22.

Linnemayr S, Alderman H. Almost random: Evaluating a large-scale randomized nutrition program in the presence of crossover. Journal of Development Economics 2011;96; 106-114.

Maccini S, Yang D. Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. Amer. Econ. Rev. 2009;99; 1006-1026.

Maluccio JA. The impact of conditional cash transfers on consumption and investment in Nicaragua. The Journal of Development Studies 2010;46; 14-38.

Manley J, Alderman H, Gentilini U. More evidence on cash transfers and child nutritional outcomes: a systematic review and meta-analysis. BMJ global health 2022;7; e008233.

Manley J, Balarajan Y, Malm S, Harman L, Owens J, Murthy S, Stewart D, Winder-Rossi NE, Khurshid A. Cash transfers and child nutritional outcomes: a systematic review and meta-analysis. BMJ Global Health 2020;5; e003621.

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

Moursi MM, Arimond M, Dewey KG, Trèche S, Ruel MT, Delpeuch F. Dietary diversity is a good predictor of the micronutrient density of the diet of 6-to 23-month-old children in Madagascar. The Journal of Nutrition 2008;138; 2448-2453.

Perumal N, Bassani DG, Roth DE. Use and misuse of stunting as a measure of child health. The Journal of Nutrition 2018;148; 311-315.

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

Raza WA, Van de Poel E, Van Ourti T. Impact and spill-over effects of an asset transfer program on child undernutrition: Evidence from a randomized control trial in Bangladesh. Journal of health economics 2018;62; 105-120.

Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in medicine 2006;25; 127-141.

Ruel MT. Operationalizing dietary diversity: a review of measurement issues and research priorities. The Journal of Nutrition 2003;133; 3911S-3926S.

Ruel MT, Alderman H. Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? The Lancet 2013;382; 536-551.

Ruel MT, Quisumbing AR, Balagamwala M. Nutrition-sensitive agriculture: What have we learned so far? Global Food Security 2018;17; 128-153.

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

Sahn DE, Stifel D. Exploring alternative measures of welfare in the absence of expenditure data. Rev. Income Wealth 2003;49; 463-489.

Shrimpton R, Victoria CG, de Onis M, Lima RC, Blossner M, Clugston G. Worldwide timing of growth faltering: Implications for nutritional interventions. Pediatrics 2001;107; art. no.-e75.

Steyn N, Nel J, Nantel G, Kennedy G, Labadarios D. Food variety and dietary diversity scores in children: are they good indicators of dietary adequacy? Public Health Nutr. 2006;9; 644-650.

Victora CG, de Onis M, Hallal PC, Blossner M, Shrimpton R. Worldwide timing of growth faltering: revisiting implications for interventions. Pediatrics 2010;125; e473-480.

Wells JC, Briend A, Boyd EM, Berkely JA, Hall A, Isanaka S, Webb P, Khara T, Dolan C. Beyond wasted and stunted—a major shift to fight child undernutrition. The Lancet Child & Adolescent Health 2019.

WHO. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr. 2006;Suppl 450; 76-85.

WHO, UNICEF. Global strategy for infant and young child feeding. World Health Organization (WHO): Geneva; 2003.

WHO, UNICEF. Indicators for assessing infant and young child feeding practices: definitions and measurement methods. World Health Organization (WHO) and the United Nations Children's Fund (UNICEF): Geneva; 2021.

World Bank. The State of Social Safety Nets 2018. World Bank: Washington D.C.; 2018.

Appendix

Table A1: Sample composition

	N	Percent
Region		
Amhara	3,015	34.43
Oromia	3,015	65.57
Pregnancy/age of child status		
Currently pregnant	3,015	53.33
Child is 0-3 months old	3,015	19.97
Child is 4-6 months old	3,015	14.49
Child is 7-9 months old	3,015	13.10

Table A2: Household demographic characteristics (N = 3,015 households)

	Mean
Household size	5.394
Male headed household	0.887
Household head's age	35.40
Household head has some formal education	0.523
Household head's main occupation is crop production	0.868
Household head is married	0.932
Primary female's age	28.75
Primary female has some formal education	0.432
Primary female has 1 to 3 years of education	0.178
Primary female has 4 to 7 years of education	0.183
Primary female has 8 or more years of education	0.071
Primary female's main occupation is crop production	0.376
Primary female is married	0.933
Public Works beneficiary in 2022	0.961
Direct Support beneficiary in 2022	0.121
Public Works beneficiary in 2021	0.298
Direct Support beneficiary in 2021	0.033
Share of people living on less than $1.90/day 2011 PPP-USD	0.710

Administrative information

Funding

Institutional Review Board (ethics approval)

Ethical approvals were obtained from the International Food Policy Research Institute (IFPRI), Ethiopian Public Health Association (EPHA) and Ethiopian Society of Sociologists, Social Workers and Anthropologists (ESSSWA).

Declaration of interest

Gilligan, Hirvonen, Leight and Tesfaye declare no conflict of interest. Mulford is a member of the program implementation team (World Vision) that designed and implemented the SPIR interventions. Mulford reviewed the protocol and provided inputs into description of the study design. However, the final decisions for content were made by the primary researchers from the evaluation team (Gilligan, Hirvonen, Leight and Tesfaye).

Acknowledgments

Gilligan, Hirvonen, Leight and Tesfaye declare no conflict of interest. Mulford is a member of the program implementation team (World Vision) that designed and implemented the SPIR interventions. Mulford reviewed the protocol and provided inputs into description of the study design. However, the final decisions for content were made by the primary researchers from the evaluation team (Gilligan, Hirvonen, Leight and Tesfaye).