



INTERNATIONAL
FOOD POLICY
RESEARCH
INSTITUTE



CGIAR
FOOD FRONTIERS
AND SECURITY



CGIAR
POLICY
INNOVATIONS

IFPRI Discussion Paper 02391

December 2025

**The Impact of a Nutrition-Sensitive Graduation
Model Program on Child Nutrition**

Experimental Evidence from Ethiopia

Kalle Hirvonen

Jessica Leight

Daniel O. Gilligan

Hiwot Mekonnen Mesfin

Michael Mulford

Haleluya Tesfaye

Poverty, Gender, and Inclusion Unit
Development Strategies and Governance Unit

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Kalle Hirvonen (k.hirvonen@cgiar.org) is a Senior Research Fellow in the Poverty, Gender, and Inclusion (PGI) Unit of the International Food Policy Research Institute (IFPRI), Washington, DC.

Jessica Leight (j.leight@cgiar.org) is a Senior Research Fellow in IFPRI's PGI Unit, Washington, DC.

Daniel O. Gilligan (d.gilligan@cgiar.org) is Director of IFPRI's PGI Unit, Washington, DC.

Hiwot Mekonnen Mesfin is an Economist at the World Bank's Gender Innovation Lab.

Michael Mulford (michael.mulford@kellogg.northwestern.edu) is the Senior Director of the Global Poverty Research Laboratory at the Kellogg School of Management at Northwestern University, Evanston, IL.

Haleluya Tesfaye (haleluya.tesfaye@cgiar.org) is a Research Analyst in IFPRI's Development Strategies and Governance (DSG) Unit, Addis Ababa, Ethiopia.

Notices

¹IFPRI Discussion Papers contain preliminary material and research results and are circulated in order to stimulate discussion and critical comment. They have not been subject to a formal external review via IFPRI's Publications Review Committee. Any opinions stated herein are those of the author(s) and are not necessarily representative of or endorsed by IFPRI.

²The boundaries and names shown and the designations used on the map(s) herein do not imply official endorsement or acceptance by the International Food Policy Research Institute (IFPRI) or its partners and contributors.

³Copyright remains with the authors. The authors are free to proceed, without further IFPRI permission, to publish this paper, or any revised version of it, in outlets such as journals, books, and other publications.

Abstract

Multifaceted graduation models are a promising strategy to sustainably reduce poverty, yet evidence on their effects on child undernutrition remains limited. This randomized controlled trial evaluated a nutrition-sensitive graduation model combining village economic and savings associations, peer-led behavior change communication, and maternal cash transfers (and for a subset, lump-sum livelihoods transfers) implemented among ultra-poor households in rural Ethiopia. The model without maternal cash transfers improved maternal nutrition knowledge and financial inclusion but did not generate meaningful changes in children's diets or growth. Supplementing the pro-gram with maternal cash transfers produced at least moderate improvements in child diet quality, early childhood development, household consumption, and assets. The largest improvements in child growth occurred among households receiving both the livelihoods grant and maternal cash transfers. Overall, the results suggest that coupling behavior change communication and livelihoods support with sufficient financial support is critical for achieving meaningful progress in both economic well-being and child nutrition.

Keywords: child nutrition, livelihoods, poverty, graduation, Africa

Acknowledgments

This study was funded by the U.S. government, award No. 72DFFP19LA00003 and sub-award No. 999004127 and the Gender Innovation Lab of the World Bank. It was also supported by the donors who fund the CGIAR's Science Program on Food Frontiers and Security and the Science Program on Policy Innovations through their contributions to the CGIAR Trust Fund. We thank Sarah Baird, John Hoddinott, Reimar Macaranas, Girum Abebe Tefera, and Dean Yang for their helpful feedback on the study protocol. We have also benefited from discussions with Harold Alderman, Mara Mordini, Benita O'Colmain, Takawira Kapikinyu, Raphael Makonnen, Andrea Warren, Mohamoud Ahmed and Tom Spangler as well as the team at World Vision, particularly Sarah Hurlburt, and Tigist Mamo. Finally, we thank Tigabu Getahun and the EconInsight field team for their excellent work collecting the data. The study was reviewed and conditionally accepted for publication in advance of data analysis under the Journal of Development Economics Pre-Results Review track. The study is registered in the AEA RCT registry: <https://www.socialscienceregistry.org/trials/9923>. Ethical review and approval for the trial was provided by the International Food Policy Research Institute (IFPRI), Ethiopian Public Health Association (EPHA) and Ethiopian Society of Sociologists, Social Workers and Anthropologists (ESSWA).

1 Introduction

Over the past two decades, social safety net programs that provide food or cash transfers to support household consumption have become widespread (Fiszbein et al., 2014; Gentilini, 2022) and evidence has shown they are effective in improving food security and reducing extreme poverty in low- and middle-income countries (Andrews et al., 2018; Beegle et al., 2018). However, effects on child undernutrition are limited (Manley et al., 2020, 2022), and evidence on the persistence of poverty reduction following the conclusion of the transfers is mixed (Leight et al., 2024). Multifaceted “graduation model” programs were developed to address this latter limitation by combining temporary consumption support with livelihoods interventions that target multiple constraints simultaneously. While these programs have shown impressive economic gains both in the short and long-term (Balboni et al., 2022; Bandiera et al., 2017; Banerjee et al., 2015, 2021, 2022; Bedoya et al., 2019; Bossuroy et al., 2022; Brune et al., 2022; Barker et al., 2024), their high cost raises questions about scalability, and their effects on child nutrition are largely unknown.¹

This gap in evidence is troubling given that undernutrition 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 reduce 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 interventions, the literature suggests that reducing undernutrition 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.² Here, the evidence remains scant.

To address this evidence gap — and to explore whether a more scalable, lower-cost

¹The 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. (2015) presents evidence of positive effects on an index of adult health, and (Bedoya et al., 2019) shows positive effects in reducing the incidence of child diarrhea. 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.

²Nutrition-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).

graduation model can deliver nutritional gains — we conducted a cluster randomized controlled trial (RCT) to assess the effects of a nutrition-sensitive graduation model embedded into Ethiopia’s flagship safety net program, the Productive Safety Net Program (PSNP). SPIR II is a graduation model program implemented in the Amhara and Oromia regions of Ethiopia that incorporates targeted nutrition-related components to address high rates of stunting among infants and young children (57 % of the children 24-48 months of age in the control arm in this study were stunted at follow-up).³ The SPIR II nutrition interventions centered 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 received a bundled set of light-touch livelihoods interventions (including village savings groups and training), and a subset of households received one-time livelihoods grants.⁴ We describe this as a light-touch graduation model since the livelihoods support is notably less intensive when compared to BRAC-designed models that combine larger asset transfers with more intensive coaching and training (Bandiera et al., 2017; Banerjee et al., 2015).

The design draws on evidence from the earlier phase of SPIR. In SPIR I, an integrated package—including village savings associations, training, one-time grants of \$200, and nutrition counseling delivered by health extension workers—generated medium-term gains in savings and livestock income and modest improvements in assets, but had no consistent effects on consumption, food security, or child nutrition (Leight et al., 2026; Alderman et al., 2025a).⁵ These findings motivated two core adjustments in SPIR II: replacing the HEW-led model with a nurturing care group model—where neighborhood volunteers are trained by SPIR health facilitators to deliver content to their peers—and adding maternal cash transfers during infancy.

Our study includes a sample of 3,015 households in 234 kebeles (sub-districts), all of whom are PSNP beneficiaries enrolled including either a pregnant woman or an infant under six months of age at baseline. The kebeles are then assigned to three arms: the control arm receives the PSNP only, the second arm receives SPIR graduation programming that

³SPIR II is an acronym for Strengthen PSNP Institutions and Resilience, and was funded by the American government and implemented by World Vision International (lead), CARE, and ORDA; more details are provided below in Section 2.1.

⁴We will consistently use the terminology maternal cash transfers to refer to the monthly transfers that are labeled for infant feeding and health expenses (though the cash is unconditional); and use the terminology livelihoods grants to refer to one-time \$300 grants that are labeled for investment in livelihoods activities (though again, they are unconditional.)

⁵One arm combining a one-time cash transfer with more intensive BCC reduced stunting (Alderman et al., 2025a), though behavior change communication (BCC) contact rates were low. Other analyses documented improvements in gender equity (Alderman et al., 2025b) and protection against drought (Hirvonen et al., 2023).

includes both livelihoods components (as described above) and the nurturing care groups, and the third arm receives the same interventions as the second arm, along with maternal cash transfers. We conducted large-scale surveys at baseline (2022) and in follow-up surveys one year later (2023) and three years later (2025), achieving an extremely low attrition rate. In addition, the timing of the follow-up surveys, in conjunction with the cohort sampling strategy, was designed to ensure that we could measure nutritional outcomes at key growth milestones. In the one year follow-up, the analytic sample comprises children between six and 23 months of age, an ideal period to measure nutritional practices linked to age-appropriate complementary feeding; in the three year follow-up, the analytic sample includes children between 24 and 48 months of age, the optimal point at which to quantify the program’s impact on linear growth indicators (e.g., height-for-age).⁶ We analyze all outcomes of interest in an intent-to-treat framework, consistent with the preregistered analysis plan.

Our findings suggest that the light-touch graduation model generally did not have a significant effect on linear growth indicators, including stunting; there were, however, significant positive effects on diet quality and also children’s development, but these effects were observed only in the arm also including maternal cash transfers. More specifically, for primary outcomes we see evidence of a positive but small effect (3%) on an index of maternal nutrition knowledge in the short-term follow-up in both treatment arms, together with a 18% percent increase in the number of food groups consumed only in the arm including maternal cash transfers (both prespecified as primary outcomes). There is no statistically significant effect on height-for-age (also prespecified as primary), or on a range of other anthropometric indicators including stunting, weight-for-age, and wasting, as measured in both the short- and longer-term. The only other notable effect, observed for secondary outcomes, is a 0.1 - 0.2 standard deviation increase in an index of early child development (based on the Caregiver Reported Early Development Instruments, or CREDI), as observed in both the one and three year surveys.

We also assess the effects of the interventions implemented on livelihoods outcomes; we follow the prespecified design in pooling across the two treatment arms, given that there is no cross-arm variation in livelihoods interventions. Here, our evidence builds on, and in fact is substantially consistent with, an earlier study assessing the effects of a prior light-touch graduation model (SPIR I) implemented by the same consortium in the same context in Ethiopia (Leight et al., 2026). We see evidence of some weak positive effects on asset accumulation driven by livestock, large positive effects on savings and credit access, and increased income from livestock. Per capita consumption rises by 4% after one year and

⁶The process of growth faltering is unlikely to be ongoing at this point, and there is limited scope for catch-up growth in this age range (Leroy et al., 2014).

by 6% after three years relative to households in the control arm, where consumption (and assets) are in fact weakly declining over time. Supplementary analysis disaggregating these effects by treatment arm suggests these positive effects are also primarily observed in the treatment arm also including maternal cash transfers: in this arm, consumption increases by 7% at one year and 10% at three years.⁷

We then conduct an additional prespecified heterogeneity analysis analyzing effects on the poorest sample households that received a targeted one-time livelihoods transfer of \$300. Here, we do in fact observe substantial and statistically significant increases in height-for-age in both the short-and long-term, particularly for households that also received the maternal cash transfer; the effects on livelihoods outcomes are qualitatively similar, but larger. We interpret these findings as suggestive that both a substantial one-time livelihoods investment (to facilitate a growing stream of income) and ongoing cash support (above and beyond the PSNP, which is in itself is often inadequate) is required to substantially shift child nutrition outcomes.

Our study contributes new evidence on how a scalable, light-touch graduation model — augmented with cash and peer-led BCC—affects early childhood nutrition and household economic outcomes, and speaks to two strands of existing research. First, this trial adds to a growing set of RCTs assessing the effects of cash transfers combined with behavior change communication on child nutrition outcomes (Ahmed et al., 2025; Carneiro et al., 2021; Field and Maffioli, 2025; Levere et al., 2024; Premand and Barry, 2022; Weaver et al., 2024). A distinctive feature of our design is that both cash and BCC are layered on top of a graduation-style livelihoods program—an integration not previously evaluated experimentally. The most closely related paper is Bouguen and Dillon (2026), who study a package combining cash and BCC with a livestock transfer, though their intervention does not include the broader livelihoods and financial inclusion components present here. We also evaluate a novel approach to delivering BCC through nurturing care groups, a peer-led model that reduces reliance on government health extension workers, who are often overburdened (Mangham-Jefferies et al., 2014; Berhane et al., 2020).

Later in the paper, we will also present a more detailed synthesis of experimental findings to situate our findings within this broader cash-plus evidence base. We document that the improvements we observe in children’s dietary diversity fall near the upper end of the distribution of effects reported in comparable cash and BCC trials, while our height-for-age estimates align with the modest impacts typically found in this literature—aside from

⁷Positive effects on consumption were not observed in the prior trial of SPIR I, and the interventions implemented there also did not include any supplementary cash transfer (above and beyond the base PSNP, received by all sampled households including the control arm). Accordingly, the fact that these effects are observed only in the arm including the additional transfer is internally consistent.

the larger gains among households who also received livelihoods transfers, which are among the largest effect sizes documented to date. These patterns mirror cross-study gradients in program intensity: mid-intensity designs, such as our cash and BCC arm, tend to yield moderate improvements, whereas more comprehensive packages that combine cash, BCC, and livelihoods support generate larger effects.

Second, our broader analysis of livelihoods outcomes contributes to the literature evaluating graduation model programs (Banerjee et al., 2015; Bandiera et al., 2017; Balboni et al., 2022; Barker et al., 2024; Bedoya et al., 2019; Bossuroy et al., 2022; Brune et al., 2022; Leight et al., 2026, 2025). Relative to the distribution of consumption effects reported in Leight et al. (2025), the estimated impact in the pooled treatment arm (6% after three years) lies toward the lower end of observed effects, whereas the corresponding estimate for the T3 arm receiving maternal cash transfers (10% after three years) is in the center of the distribution. This positioning is consistent with the design of SPIR II: the core SPIR II package delivered to both treatment arms is less intensive than the multi-component ultra-poor models widely evaluated in other contexts, and the additional maternal cash transfers in T3 increase the total transfer disbursement closer to the level provided in higher-intensity graduation programs.

The findings of this trial also have important policy implications, in that we evaluate the impacts of a graduation model intervention within the context of an existing large-scale government-led safety net program. These findings are highly relevant for future social safety net programming both in Ethiopia and in other contexts.

2 Experimental design

2.1 Context and intervention

We report on a trial of multiple interventions implemented in Amhara and Oromia regions within Ethiopia, delivered to households who are beneficiaries of the Productive Safety Net Program (PSNP). 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). The program is structured around the provision of six months of payments in food or cash to rural households as payments for labor or unconditional transfers during the agricultural off-season (generally, January to June), and the median annual transfer per household in SPIR areas is estimated to be about \$440 in 2011 purchasing power parity (PPP) terms, corresponding to roughly 15% of the total annualized household consumption at baseline.⁸ Extensive evidence suggests

⁸These estimates are from the baseline survey in 2022.

the PSNP has been successful in improving household food security and asset levels (Gilligan et al., 2009; Hoddinott et al., 2012, 2024), but observational and quasi-experimental evidence suggest negligible effects on nutritional outcomes to date (Berhane et al., 2015, 2017).

The evaluation was also conducted during a period of considerable instability in rural Ethiopia, during which safety nets such as the PSNP faced even greater demands. Particularly salient has been the steady increase in conflict and insecurity. The Amhara region experienced a period of acute violence linked to the war in Tigray (with the northern part of the region occupied by Tigrayan forces) in 2021; following their withdrawal and a period of relative quiet, there has been renewed insecurity since 2023 linked to conflict between central government and regional militia. The Oromia region, while less acutely affected, has also been characterized by ongoing unrest, and periodic violence.⁹ Pronounced weather variability in the region has also continued, and intensified (Kimutai et al., 2025), placing heavy strain on farm households.

In this volatile context, the SPIR II program was implemented by World Vision International (lead), CARE, and ORDA with funding provided by the American government (2019 – 2026), and our study focuses on innovations in nutrition-related programming 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; Hirvonen et al., 2024).

A core component of SPIR II’s nutrition interventions was the use of the Nurturing Care Group (NCG) model — an adaptation of the Care Group approach originally developed by World Relief in Mozambique in 1990s (Perry et al., 2015). The model entails a cascaded structure in which groups of 10–15 volunteer “neighborhood leaders” meet regularly with SPIR community Health facilitators (and, in some cases, government health extension workers) for training using a structured 16-lesson, picture-based curriculum. These volunteers then convene biweekly meetings with small groups of pregnant and lactating women and caregivers of young children (the latter known as neighborhood groups) and conduct home visits to caregivers who miss sessions or require additional support.¹⁰ Although experimental evidence on the effects of NCGs remains non-experimental studies suggest that this model can substantially increase exposure to behavior change communication and improve IYCF practices and child growth (Davis et al., 2013).

⁹See for example an overview here: <https://www.cfr.org/global-conflict-tracker/conflict/conflict-ethiopia>

¹⁰Quality was monitored using standardized verification checklists, with facilitator and volunteer performance typically exceeding 80 percent and attendance rates above 85 percent. In total, roughly 6,000 volunteers completed the full curriculum with more than 74,000 caregivers across all SPIR II kebeles: however, this evaluation focuses on a much smaller sample included in the target kebeles.

Yet improving caregiver knowledge may not be sufficient to improve complementary feeding practices if households cannot afford to purchase nutritious foods. 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., 2023). Accordingly, SPIR II included the introduction of maternal transfers of \$20 monthly (or 38 \$PPP) for a period of 24 months to relax possible financial constraints to child feeding. Considering the baseline consumption estimates of around \$160 monthly per household (or 272 \$PPP), described in more detail below, the transfers correspond to around 13 percent of monthly consumption. These maternal cash transfers were rolled out immediately following the baseline survey, and thus given the sampling strategy (described in more detail below), households receive their first payment either in the final months of pregnancy or during early infancy. The cash transfers were unconditional, but were weakly labeled as targeted for child feeding and health-related expenses.

All households included in NCGs are also 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 (or 563 \$PPP) livelihoods grants; targeting was based on an asset index constructed (by the research team) from data on asset ownership collected at baseline. (Again, we will consistently use the terminology maternal cash transfers to refer to the ongoing, monthly cash transfers disbursed as part of T3; and will use the term livelihoods grants to refer to the one-time grants disbursed to extremely poor households in both T2 and T3.)

In sum, the most intensive intervention package — received by a subset of households in the T3 arm — combines the livelihoods interventions and the NCG-based BCC with two distinct types of transfers: the monthly maternal cash transfer (\$456 PPP per year), and the one-time livelihoods grant (\$563 PPP). All households (in both treatment and control) also receive the regular PSNP cash or food payments (about \$440 PPP per year on average); there is no experimental variation in receipt of the core PSNP.

2.2 Sample and randomization

The sample for this evaluation includes 234 kebeles (subdistricts) in 15 woredas (districts), purposively chosen from within the SPIR II operational area of more than 450 kebeles.¹¹

¹¹The purposive selection was designed to prioritize the selection of kebeles that were included in the control arm of previous trials conducted as part of the first phase of SPIR. The original evaluation design

Kebeles were randomly assigned to one of three arms (see Table A1): T1 serves as a control arm, including households receiving PSNP only.¹² PSNP households in arm T2 receive SPIR II graduation programming and the NCG intervention. PSNP households in arm T3 receive maternal cash transfers in addition to the same intervention package as households in T2.¹³

At the household level, the sampling criteria for the evaluation were as follows. The household had to be enrolled as a PSNP beneficiary, and also have either a pregnant woman present self-reporting pregnancy in the second or third trimester, or an infant present (under nine months of age).¹⁴ Households were identified and screened using a randomly ordered list of PSNP beneficiaries in order to constitute a sample of 13 households per kebele, as described in more detail in the preregistered report; the realized sample was 3,015 households in 234 kebeles, closely matching the target sample. More details about sample composition are provided in Table A2 in the Appendix.

A central feature of the design is the cohort-based sampling strategy. By enrolling households with either a pregnant woman or an infant under nine months at baseline, the evaluation follows a single birth cohort through the key developmental window from pregnancy through early childhood. The timing of the two follow-up surveys was selected to align with recommended ages for assessing nutrition outcomes (Alderman and Headey, 2018; Leroy et al., 2016): the one-year survey measures feeding practices among children 6–23 months old, while the three-year survey measures height-for-age among children 24–48 months old, when linear growth reflects cumulative nutritional conditions.

We conducted power calculations using data from both previous surveys conducted with similar samples (in the SPIR I evaluation) and from the baseline survey conducted as part of this evaluation.¹⁵ For the nutrition-related outcomes, we estimate statistical power corresponding to the pairwise 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

included 237 kebeles, but three were dropped as they were inaccessible for surveys due to conflict at baseline.

¹²Randomization was conducted by the research team in Stata. 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).

¹³Note that the core graduation model interventions were also rolled out to all other PSNP households in the broader SPIR operational area, outside of the kebeles included in this trial; however, the NCGs and maternal cash transfers were implemented only in kebeles in this trial.

¹⁴No pregnancy tests were conducted.

¹⁵We set the significance level at five percent and power at 80 percent, and allow for 10 percent attrition between baseline and follow-up surveys. Additional power calculations using SPIR I evaluation data are also included in the preregistered report; 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.

(0.23 SD). These minimum detectable effect (MDE) sizes are biologically meaningful and in line with findings in previous trials. For livelihoods-related outcomes, we prespecified combining the two treatment arms and comparing this pooled group to the control arm. Given this design, the study has an MDE of roughly 0.2 standard deviations for livelihoods outcomes.

While kebele-level randomization minimizes contamination across arms, two potential spillover channels warrant consideration. First, one-time livelihoods grants and monthly maternal cash transfers could, in principle, generate local price inflation affecting non-beneficiaries (Cunha et al., 2019; Egger et al., 2022; Filmer et al., 2021). In practice, program saturation was low—on average, only about five percent of households in a kebele were eligible for livelihoods grants, and the maternal grants were provided exclusively to households enrolled in this evaluation—so the risk of such market-level effects is negligible. Second, informational spillovers through caregiver interactions across kebeles may be possible, but these are likely limited given the cluster-randomized design and the geographic dispersion of villages within kebeles. Data from one year follow-up survey further suggest such cross-kebele interactions were rare, with only about six percent of NCG participants reporting discussing meeting topics with others outside their kebele.

3 Data and econometric specification

The study design, hypotheses, and analysis plan were pre-specified in a pre-registration report that underwent peer review and was conditionally accepted by the *Journal of Development Economics* prior to midline data collection in 2023. Deviations from the pre-registration report are discussed below and summarized in Table A3 in the appendix.

3.1 Data

We employ three large-scale household surveys in this analysis: the sample was surveyed at baseline (2022), in a short-run follow-up conducted in 2023, and in a long-run follow-up conducted in 2025. The first follow-up survey was largely implemented as planned 12 months post-baseline, though there was some variation in survey timing driven by conflict-related inaccessibility in Amhara.¹⁶ The long-run follow-up survey was conducted around three months earlier than planned due to an anticipated widespread rollout of livelihoods

¹⁶Survey work in 2022 was conducted in August - September; survey work in 2023 was conducted in August - September in Oromia, but in October - November in Amhara, with 12 additional kebeles surveyed in December.

grants by the Ethiopian government, described in more detail below in the section on program exposure. In general, all three surveys included questionnaires administered to the primary female (the pregnant woman or mother of the target infant at baseline), and her spouse (though men were not interviewed in the one-year follow-up). Anthropometric data collection was conducted in the follow-up surveys only. Ethical review and approval for the trial was provided by the International Food Policy Research Institute (IFPRI), Ethiopian Public Health Association (EPHA) and Ethiopian Society of Sociologists, Social Workers and Anthropologists (ESSWA).

The outcomes of interest were also prespecified in the registered report, and are described in detail in Table A4 in the Appendix. The primary outcomes in this trial include caregiver infant and young child feeding knowledge (measured in the one- and three-year follow-up); the number of food groups consumed by children 6–23 months (measured in the one year follow-up); and height-for-age for children 30–48 months (measured in the three year follow-up). Secondary outcomes in the nutrition domain include a range of other variables linked to infant and young child feeding practices, anthropometric status (including binary variables for stunting and wasting, measured at multiple time points) as well as early childhood development index based on Caregiver Reported Early Development Instruments or CREDI (McCoy et al., 2018), an index that captures age-specific developmental milestones that differ across early childhood. The secondary outcomes in the livelihoods domain include household per capita consumption and food security (measured at one and three years), and assets, financial inclusion, and income (all measured in the three year follow-up).

3.2 Econometric specification

Our primary specifications can be written as follows

$$Y_{ikd,t=1} = \beta_1 T_{kd}^2 + \beta_2 T_{kd}^3 + \gamma Y_{ikd,t=0} + X'_{ikd,t=0} \theta + \xi_{kd} + \epsilon_{ikd} \quad (1)$$

$$Y_{ikd,t=1} = \beta T_{kd} + \gamma Y_{ikd,t=0} + X'_{ikd,t=0} \theta + \xi_{kd} + \epsilon_{ikd} \quad (2)$$

Here, $Y_{ikd,t=1}$ captures the outcome of interest for household i in kebele (sub-district) k and woreda (district) d at time t . Variables T_{kd}^2 and T_{kd}^3 are indicator variables for arms T2 or T3, respectively; T_{kd} is a pooled treatment indicator; ξ_{kd} are strata fixed effects. All specifications are estimated conditional on baseline controls $X_{ikd,t=0}$, including baseline household size, the age and education level of the primary caregiver, and the age and sex of the child; we also include the baseline value of the outcome $Y_{ikd,t=0}$, if available. These controls were prespecified in the registered report. However, because including the age and

sex of the child would require excluding households without eligible index children, we omit these variables in the livelihoods regressions to avoid unnecessary sample loss.

For nutrition-related outcomes, we estimate equation (1) and report the p-value for the null hypothesis that the treatment effects are consistent across treatment arms, $\beta_1 = \beta_2$. For livelihoods-related outcomes, we will first analyze equation (2) pooling data across the two treatment arms; as a robustness check, we will estimate equation (1) to test whether there is any differential effect of the maternal grants on livelihoods outcomes. Standard errors are clustered at the level of treatment, i.e., the kebele. In addition to reporting standard p-values, we will also report p-values corrected for multiple hypothesis testing based on sharpened false discovery rate q-values (Benjamini et al., 2006) based on the procedure devised by Anderson (2008). This correction is implemented within the set of primary outcomes, and within the set of secondary outcomes.

4 Empirical findings

4.1 Baseline balance and loss to follow-up

Table 1 reports demographic characteristics of the sample households, as well as the balance in observable characteristics across arms. The average household size is five; more than half of the household heads report some formal education, and 87 percent report that their primary economic activity is crop production. The primary female in the sample households (defined as the pregnant or lactating mother) is on average 29 years of age, and only 43% report any formal education. The estimated prevalence of extreme poverty (consumption under \$1.90 a day) is 71 percent. We conducted a joint test across all outcomes to test the hypothesis that the observable characteristics are generally balanced across experimental arms T1, T2, and T3, and fail to reject this hypothesis.

Across the multiple survey rounds, attrition at the household level was extremely low: 2% of households were lost to follow-up at one year, and 4% at three years. However, there was substantial sample loss at the child level between baseline and both follow-up surveys, corresponding to women who were enrolled reporting pregnancy at baseline, but did not have an infant present in the one year follow-up: 34% of pregnant women did not have a child.¹⁷ Women self-reported this was largely due to pregnancy loss, though there were anecdotal

¹⁷The absence of an infant in households enrolled with a pregnant woman was particularly notable in Oromia, where only 59% of women enrolled based on pregnancy had an infant at the one year follow-up; the corresponding rate in Amhara was 83%. There is also some child-level attrition among infants enrolled at baseline, primarily due to deaths or departure of the mother and infant from the household, though this is rare: only 3% of the infants enrolled at baseline could not be surveyed in the one year follow-up, and 2.5% of the infants surveyed at the one year follow-up could not be surveyed in the three-year follow-up.

(and non-verifiable) reports that some women may have feigned pregnancy at baseline in the hope of accessing some benefit targeted to pregnant or lactating women. As a result, the sample of eligible children meeting the age criteria was 2,234 in the one-year follow-up (rather than the approximately 3,000 children originally projected), and 2,203 in the three year follow-up.¹⁸ The implications for statistical power are, however, not large.¹⁹ Below we further explore any potential bias induced by non-random attrition at both the household and child level below.

4.2 Program exposure

Table 2 reports patterns of program exposure across experimental arms, focusing on key indicators of participation in SPIR. The intervention was implemented with relatively high fidelity. Participation in the neighborhood groups implemented as part of the nurturing care group intervention is 74% in T2 and 87% in T3 in the one-year follow-up, and recall is still high at three years.²⁰ The level of participation is consistently higher in the arm also receiving cash grants, suggesting some potential of higher interest in groups when grants are also received. Nearly 90% of women targeted for grants also receive them; within those receiving grants, 83% reported that the grants were primarily used for child feeding.

Around 20–25% of women in the control arm also report participating in neighborhood groups; however, in follow-up questions, these women primarily identify the group leader as a health extension worker. While health extension workers may conduct different forms of group-based behavior change communication, they do not lead neighbor groups, suggesting that these respondents are misattributing a different intervention. Contamination in reported receipt of maternal cash transfers is minimal.

Participation in VESAs, the core SPIR livelihoods-related intervention, is around 60–70% in the short-run follow-up, and rises to around 80% in the long-run follow-up. (Again, there is some reported contamination, a pattern that most likely reflects misattribution of other interventions implemented in these kebeles.) Receipt of livelihoods grants is right around 33% in both treatment arms as intended, with minimal evidence of contamination.

¹⁸We also excluded 192 children who were surveyed in the one-year follow-up but who were outside the target age range of 6–23 months (they were older); and we excluded 91 children who were outside the target age range in the three year follow-up (younger than 24 months). This margin of error corresponds to imprecision in reported infant age at baseline (these children should not have been sampled, given their true age).

¹⁹For the primary variable of height-for-age, the minimum detectable effect increases from 0.14 to 0.18 standard deviations given the reduced cluster size.

²⁰Given that these are questions about any participation in the past, responses that are consistent over time would imply that the level of participation should be weakly higher in the long-run follow-up, but it seems that some respondents no longer remember or identify their previous participation when asked again at three years.

As previously noted, the timing of the three year follow-up survey was slightly altered given a planned large-scale rollout of one-time livelihoods grants disbursed directly by the government of Ethiopia, outside the framework of SPIR programming (and funded separately).²¹ We find that in general, the shift in survey timing was effective in avoiding any contamination of the new livelihoods grants: while 55% of households were aware of the new livelihoods-related initiative, only 22% had joined a household livelihoods group (designed to be a precursor to grant receipt), and only 5% of households had in fact received a grant, concentrated entirely in three woredas in Amhara.²² Accordingly, we do not assess that there is any meaningful risk of contamination to the experimental design linked to this new initiative.

4.3 Main findings

Table 3 presents the findings for the nutrition-related outcomes. Panel A presents findings from the one-year follow-up, when the sample children are aged 6–23 months, and Panel B presents findings from the long-run follow-up when the sample children are aged 24–48 months. Primary outcomes are denoted with a dagger. Our discussion will also draw on some supplemental findings reported in Table A5; these variables were not prespecified.²³

In Panel A of Table 3, we can see that the SPIR intervention did effectively increase IYCF knowledge in the short-term, though the effects on the knowledge score are not large: an increase of 3% in T2 including NCGs only, and 5% in T3 including both NCGs and grants (the difference between the two arms is marginally significant). There is also an increase in the number of food groups consumed by the index children that is dramatically larger in T3 (17.5%) versus T2 (1.2% and statistically significant only at the 10-% level), and an increase in the probability of achieving minimum acceptable diet (MAD) in T3 only: 11 percentage points relative to a mean in the control arm of only nine percent, roughly a 120% relative gain. This estimated relative effect far exceeds the average 17% improvement reported in a recent meta-analysis of similar interventions, which found a modest and statistically insignificant pooled effect on achieving MAD (Mahumud et al., 2022).²⁴ Findings around infant and young child consumption by food group reported in Panel A of Table A5 indicate that this

²¹A description of this shift can be found in programmatic documents, e.g., in a World Bank report that notes “GoE is prioritizing livelihoods activities with a five fold increase in targeted beneficiaries” (World Bank, 2025).

²²In general, rollout of the program seems to have been more rapid in Amhara, where more than 80% of households were aware of the program and a third had joined a livelihoods group.

²³All estimates reported in this section are conditional on the baseline controls described earlier. Re-estimating the specifications without these controls yields near-identical results; see Tables A20, A21, and A22 in the Appendix.

²⁴More specifically, the pooled prevalence ratio = 1.17, 95% CI [0.61, 1.73].

increased diet diversity was primarily driven by increases in consumption of eggs, dairy, and fruits and vegetables (including vitamin-rich fruits and vegetables).

Anthropometric outcomes are reported in Columns (4) through (6) of Panel A of Table 3. We find weak evidence of gains in anthropometric status in T3 only: height-for-age Z-scores increase by 0.09, and the probability of stunting is reduced by three percentage points. These effects are small in magnitude and not statistically significant based on conventional p-values (0.317 and 0.253, respectively), though they remain significant at the 10% level after controlling for multiple inference using the q-value correction ($q = 0.069$ and 0.082).²⁵ The corresponding coefficients in T2 are weakly negative or close to zero, and there is no evidence of any shift in weight-for-age in either arm. We also observed in exploratory analysis of short-term follow-up data a large shift (0.19 standard deviations) in the CREDI index of child development in T3 only, capturing attainment of early developmental milestones, as well a substantial reduction in diarrhea risk (a reduction of six percentage points relative to 37% in the control arm), reported in Columns (1) and (2) of Panel B of Table A5.

Shifting to the main findings reported in Panel B of Table 3 for the three-year follow-up, the effects on knowledge seem to have been persistent over time, though slightly smaller in percentage terms in T3 (now a 3% increase relative to the control mean in both arms).²⁶ There is, again, no robust evidence of any shift in anthropometric status, despite the fact that this should be the optimal period in which to detect any effect of the interventions of interest on child growth: the same weakly positive effect on height-for-age is evident, but it is smaller relative to the mean in the control arm, and not statistically significant. The pattern of growth faltering generally observed among children in this age range in Ethiopia (Golan et al., 2019) is also clearly evident here: among children in the control arm, height-for-age declines from -1.6 to -2.2 while the stunting rate increases from 37% to 57% over the two-year gap between the follow-up surveys.

Despite the insignificant effects on anthropometric outcomes, the positive treatment effects on the CREDI child development indicator — now capturing attainment of later-age developmental milestones — are also observed in the T3 arm at three years follow-up, with a magnitude of 0.12 standard deviations.²⁷ Together, these findings suggest that improve-

²⁵The fact that these coefficients are marginally significant after adjusting for multiple testing using q-values suggests that they are relatively robust within the broader set of outcomes examined.

²⁶The knowledge questions administered were identical in both survey rounds; despite this, there is a slight decline in average scores in the control arm, suggestive of perhaps reduced salience of questions linked to IYCF for the sample women now that their children are slightly older. The generally declining level of knowledge in the control arm also again suggests there have been no meaningful informational spillovers from treated areas.

²⁷Note that one risk in the use of the CREDI in the second follow-up round is that some of the children have aged past 36 months, the recommended CREDI age; however, we find that still substantial numbers of children have not achieved the age-appropriate developmental milestones that are the focus of the instrument,

ments in early child development can emerge even in the absence of measurable gains in linear growth—a pattern consistent with broader evidence showing that nurturing-care and behavior-change interventions often shift developmental outcomes without affecting height-for-age (Prado et al., 2019).

Supplementary findings again reported in Panel B of Table A5 allow us to better understand the patterns underlying some of these outcomes. Consistent with the reported increase in maternal knowledge, we see an increase in paternal IYCF knowledge at three years of roughly the same magnitude (0.2).²⁸ We also see an increase in the reported incidence of child stimulation practices (playing, reading, and singing with the child) in T3 only (around 10% relative to the control mean), consistent with the observed shifts in the CREDI. We also see a substantial decline in the probability of open defecation in T3 only (measured at the three year follow-up), consistent with the continued reduction in diarrhea risk (again, observed in T3 only). This could be interpreted as suggestive that the cash payments are used to defray the costs of some additional investment in sanitation; or alternatively, the receipt of cash allows the mother (as primary caregiver) to forgo some income-earning activity (e.g., petty trade or day labor) and her greater investment of time is also evident in better household sanitation conditions.

We also see evidence in Table A6 that SPIR expanded linkages to local health services and strengthened health-seeking behaviors, another channel that is consistent with the observed shifts in child health (and also consistent with one of the stated goals of the NCG model). Exploratory analysis based on the one year follow-up shows that both treatment arms substantially increased contact with frontline health workers over the preceding three months: relative to the control group, women in T2 were 10–15 percentage points more likely to have had contact with a health extension worker (HEW) or health development army (HDA) member²⁹ in the preceding three months, with slightly larger increases in T3. Participation in food demonstrations and community conversations also rose sharply, with increases of 19–23 percentage points in T2 and 30–38 percentage points in T3.³⁰

To sum up the findings related to nutritional outcomes, the nurturing care groups alone produced modest improvements in caregiver knowledge and small increases in diet diversity, as well as some gains in health-seeking behaviors. When paired with maternal cash transfers, the program generated substantially larger improvements in children’s diets and greater

as seen in Figure A1. Accordingly, we see little risk of topcoding in the index.

²⁸At the three year follow-up, 77% of households with an index child present had a primary male who was available to be interviewed. The smaller sample size in this analysis therefore reflects missing data for households in which the primary male was either unavailable or absent.

²⁹Health development army (HDA) members are volunteer women under Ethiopia’s Health Extension Programme who support community health promotion and nutrition messaging.

³⁰These could be organized either by HEWs (though in practice this seems to be rare) or by SPIR itself.

increases in health-seeking behaviors. For anthropometric outcomes, estimated effects in the T3 arm were small at both the one-year and three-year follow-ups, with only weak statistical significance at first follow-up and similarly sized but imprecisely estimated effects in the second. Overall, the pattern of results suggests that behavior-change communication alone is insufficient to shift growth outcomes at scale, and that even with the substantial improvements in diets generated by pairing BCC with maternal cash transfers, these changes may not have been large enough or sustained long enough to translate into measurable gains in anthropometry. The intervention also generated positive impacts on child development, with statistically significant improvements observed at both follow-ups.

Tables 4 and 5 now present findings around livelihoods, and some supplemental outcomes that were not prespecified are reported in Table A7. In Table 4, we can observe that there is a positive treatment effect in per capita consumption in both the short- and long-term that slightly grows over time, 4% in the short-term and 6% in the long-term. There is, however, little evidence of any effect on either binary or continuous measures of food insecurity in either period, possibly reflecting the relatively coarse nature of the FIES indicator. The findings in Panel A of Table 5 suggest that this relatively small increase in consumption is observed in conjunction with a large increase in assets (driven by livestock) and a substantial increase in both savings and credit access. The total asset index increases by around 0.15 standard deviations, normalized relative to the baseline standard deviation in the control arm, and this is primarily driven by an increase in tropical livestock units (0.12 standard deviations).³¹ Detailed findings reported in Panel A of Table A7 unpack the increase by livestock type: significant increases are observed for cattle, sheep and goats, pack animals, and poultry, though largest for poultry.

Returning to Table 5 to examine effects on financial inclusion, we also observe a 22 percentage point increase in the probability of any savings and a 75% increase in the amount of savings in Columns (3) and (4). There is a seven percentage point increase in the probability of reporting any credit access and a 40% increase in the reported outstanding credit balance as of the survey date.³²

Panel B of the same table then probes the effects on income to understand how households' livelihoods portfolio has shifted. Unsurprisingly given the evidence of increased livestock ownership, we see a significant increase in reported past-year income from livestock along both the extensive and intensive margin: a 10 percentage point increase in the probability of any net income, relative to a mean of 77% in the control arm, and a 50% increase in

³¹The negative mean in the control arm can be interpreted as declining overall assets in the control arm, relative to baseline.

³²This increase is primarily driven by increased utilization of formal credit, as reported in Column (1) of Panel B, Table A7: the probability of accessing formal credit more than doubles, relative to the control arm.

the amount of income from livestock. However, there is some weak evidence that households are earning somewhat less income (16%) from non-agricultural production, though this effect is only weakly significant using q-values, and there is no effect on the overall probability of reporting a non-agricultural business. This would be consistent with some constraints in overall household labor and/or capital, such that households investing more in livestock are unable to sustain similar returns to other livelihoods activities.³³ More detailed, exploratory findings reported in Panel B of Table A7 also demonstrate that treatment households are less likely to engage in informal wage employment (often a coping mechanism in these contexts), and show no shift in their reported income from crop harvesting.

Figure 1 also captures some key treatment effects (for consumption, food security, and assets) using longitudinal data: we can see that the general trend is weak declines in consumption and assets over time for households in the control arm, with treatment households buffered from these effects. (In this context, the positive treatment effect on consumption that slightly grows over time, as described above, can be interpreted as buffering households from deteriorating conditions.) These temporal patterns are consistent with the challenging conditions in which the study took place: repeated covariate shocks—including conflict disruptions and erratic rainfall—likely contributed to the downward trajectory in household welfare observed among control households. If we examine treatment effects on graduation from the PSNP itself, we see a pattern that is consistent with the adverse conditions faced by control households: at the three year follow-up, 97% of households in the control arm remained PSNP beneficiaries, and the estimated treatment effect is small and statistically indistinguishable from zero (Column 5 in Panel B of Table A7). In a context in which conditions are generally stagnant or weakly declining and the intervention buffers households against some adverse effects of shocks, no substantial effect on program exit would be expected, and that is in fact exactly what we see.

We then report in Tables A8 and A9 the findings disaggregated by treatment arm. Here, the positive effects observed on consumption and assets are driven almost entirely by the arm including maternal cash transfers (T3), though effects on financial inclusion are consistent across arms and positive effects on livestock income are observed in both arms (though significantly larger in T3). Focusing first on consumption and food security, Table A8 shows that the positive effect on consumption and reduced severe food insecurity are observed only in T3, and increase in magnitude to 10% relative to the control mean in the long-run

³³The relatively high rate of reported non-agricultural enterprises, compared with other surveys in similar contexts, likely reflects differences in measurement. The enterprise module required enumerators to systematically probe for all income-generating activities, leading to more complete reporting of very small-scale enterprises—often petty trading of agricultural goods or small retail activities—that may have gone unreported in earlier surveys.

follow-up; the corresponding coefficients for consumption for T2 are opposite in sign and not statistically significant, and the difference between the two arms is highly significant.

In Table A9, the positive effects on assets in Columns (1) and (2) are again concentrated in 3, with insignificant coefficients in T2: the positive effect on assets now rises to 0.2 standard deviations in T3. For variables linked to savings and credit by contrast, the hypothesis that the effects are parallel across treatment arms generally cannot be rejected with the exception of the continuous amount of savings, where the treatment effect in T3 is again larger. For income, there are some increases in livestock income in both arms, but the increase along both the extensive and the intensive margin is double the magnitude in the arm including cash transfers (and again, the difference across arms is statistically significant).

To sum up: despite our *ex ante* hypothesis that the effects on livelihoods outcomes would be consistent across treatment arms, the treatment effects are notably larger in the arm including maternal cash transfers (in some cases, positive only in this arm). It can be reasonably inferred that part of the substantial cash transfer directed to mothers (nearly \$500 over two years in T3) was also directed toward sustaining a stock of household assets, enabling a more persistent increase in consumption over time; in fact, asset stocks are weakly declining over time in the control arm and stable in the treatment arm, generating the positive treatment effect. It is also possible that this is part of the reason for the absence of any meaningful effects on the index child’s anthropometric status (if resources were more broadly shared within the household to maintain subsistence consumption and ensure protection of assets), though this is speculative.

The pattern of livelihoods effects in T2, the arm without maternal cash transfers, is strikingly similar to the findings observed in a previous trial of the livelihoods interventions implemented in SPIR I (without any cash transfer): substantial increases in financial inclusion, linked to high participation in VESAs, and small increases in livestock income without any shift in household consumption (Leight et al., 2026). Here, we observe that the addition of another substantial cash injection seems to lead to larger shifts in livelihoods-related outcomes including consumption, consistent with the hypothesis that PSNP transfers themselves are simply not sufficient to meet households’ subsistence needs, a point also made in previous literature (Hoddinott et al., 2024; Leight et al., 2026).

A similar narrative emerges in Carneiro et al. (2021), where cash transfers paired with BCC in Northern Nigeria enabled households to accumulate livestock and expand earnings. Their results underscore how predictable maternal cash transfers can relax liquidity constraints, support productive investment, and contribute to improved child outcomes. Our T3 findings are consistent with this mechanism: shifts in consumption, asset holdings, and income arise only when households receive the additional maternal transfer, whereas the T2

arm exhibits the more limited pattern characteristic of livelihoods-only interventions.

4.4 Additional analyses

We prespecified two additional analyses in this trial. The first is an analysis of heterogeneous effects with respect to baseline male (paternal) knowledge around 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. To test this, we constructed an indicator variable identifying households in which the father’s nutrition knowledge score at baseline was above the sample median, and interacted this indicator with the treatment status. The coefficient on the interaction term captures whether treatment effects differ among households where fathers demonstrated higher baseline knowledge.

However, the findings presented in Table A10 show no meaningful evidence of heterogeneity, and if anything, the pattern is suggestive of slightly smaller effects on households characterized by higher paternal knowledge at baseline.³⁴ This finding is consistent with another recent study found that the effects of targeting fathers for nutrition-related behavior change communication as well as mothers did not lead to any differentially larger shifts in child feeding outcomes (Han et al., 2022).

The second analysis assesses variation in experimental effects with respect to experimentally assigned varying 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 and analyze effects on both subsamples.

For nutritional outcomes, the findings are presented in Table A11 for households exposed to livelihoods transfers, and in Table A12 for households that were not exposed to transfers. We can observe that for the transfer-exposed households, the effects on nutritional outcomes are meaningfully larger, and in particular, there is a significant effect on height-for-age in T3 in the one-year follow-up (0.18 in magnitude or 11% relative to the control mean), and at least weakly significant effects in both arms in the three-year follow-up (0.17 and 0.23 in

³⁴The sample is smaller in this analysis because father-level data were only collected when a primary male respondent was present and available at baseline. Overall, 66.7% of households had an interviewed male respondent, 24.8% had a primary male who was not available during the day of the interview, and 8.5% did not have a primary male.

magnitude, or 7–10% in magnitude relative to the control mean). A reduced prevalence of stunting of around seven percentage points is observed in both treatment arms at three years (weakly significant using q-values). Importantly, the hypothesis that the effects on height-for-age are equal across arm cannot be rejected (though the effects on diet diversity in the one-year follow-up remain larger in the arm that also received maternal cash transfers). For households that do not receive livelihoods transfers as reported in Table A12, the effects on knowledge and diet diversity remain robust (the latter again in T3 only), but there is no evidence of any increase in height-for-age.³⁵

The corresponding effects for livelihoods are presented in Tables A15 and A16 for households eligible for transfers, and in Tables A17 and A18 for households that are ineligible. For eligible households, the effects are qualitatively the same as those observed in the pooled sample but larger: the effect on consumption rises to 7% of the control mean at both one and three years, the increase in assets is now 0.27 standard deviations at three years, and the effect on net livestock income is now fully 130% of the control mean (the effects on financial inclusion variables, however, are roughly consistent). Conversely, the effects on livelihoods outcomes for non-transfer-eligible households are again qualitatively similar but reduced in magnitude: a 3% increase in consumption, a 0.1 standard deviation increase in the asset index, and a 38% increase in livestock income.

Overall, these findings strongly suggest that it is the conjunction of the large one-time livelihoods transfer and some supplemental cash support that leads to more pronounced effects on both livelihoods outcomes and a significant shift in children’s anthropometric status. The absence of statistically significant impacts of anthropometric outcomes at the three-year follow-up in T3 (Table 3) contrasts with the larger and persistent impacts observed in the same arm that combined maternal cash transfers with a one-off livelihoods grant (Table A11). This pattern is consistent with evidence that livelihoods transfers capable of generating a more sustained increase in household income are more likely to translate into improvements in children’s linear growth at longer horizons. At the same time, it is also plausible that the livelihoods grant—by supporting asset accumulation and household consumption—allows a greater share of maternal cash transfers to be consistently allocated toward improved infant feeding. In the absence of the lump-sum transfer, the program still generates meaningful, though smaller, improvements in livelihoods outcomes, but these do not appear to be large enough to produce measurable gains in children’s linear growth.

³⁵Height-for-age is in fact weakly decreasing in the T2 arm in the three year follow-up, but this effect is not statistically significant using conventional p-values.

4.5 Attrition

As previously noted, attrition in the conventional sense (at the household level) is extremely low in this trial: 2% in the one-year follow-up, and 4% in the three-year follow-up. However, attrition at the child level—defined as the absence of the index child from the household at follow-up—is much more common (19% in the one year follow-up and 26% in the three year follow-up). This is primarily due to high rates of miscarriage or feigned pregnancy in the sample enrolled at pregnant at baseline, but also reflecting some mortality among the sample over time.

To assess whether there is any bias due to attrition in the primary findings, we first generate binary variables for attrition at the household- and child-level at the one- and three-year follow-up and regress these variables on treatment arm binary variables, selected baseline covariates and the interaction between the two; we report the findings in Tables A13 and A14.³⁶ The first and fourth columns of each table report unconditional attrition rates across study arms: the constant represents attrition in the control group, while the coefficients on T2 and T3 show differences relative to control. At both follow-ups, these differences are relatively small — on the order of one to two percentage points — and not statistically significant in most cases.

For household attrition (Table A13), we find that baseline covariates, treatment status, and the interaction generally are not predictive of the minimal observed variation in attrition, other than a decrease in the probability of attrition in T3 at the one-year follow-up; the p-value on the joint test across all covariates and interaction terms reported is 0.172 for the one-year follow-up, and 0.818 at the three-year follow-up.

For child-level attrition (Table A14), unsurprisingly the most meaningful predictor of child-level attrition is enrollment status in pregnancy at baseline, as very few children attrited who were enrolled as infants. Households characterized by high food insecurity, low assets, or an older primary female at baseline are also somewhat more likely to report non-appearance of a child in a follow-up round, suggestive of positive correlations between poverty and maternal age and miscarriage and/or child mortality; however, again there is relatively limited evidence of any variation in significant predictors of attrition across arms.³⁷ Again, we fail to reject the hypothesis of a null effect for the joint test across all covariates and interaction effects reported (p=0.707 in the one-year follow-up, and p=0.743 in the three year follow-up).

³⁶This analysis was not prespecified in the registered report.

³⁷Another interpretation of this correlation would be that households experiencing more food security were more likely to feign pregnancy at baseline, anticipating some later benefit; but again, we have no direct evidence of this.

To assess any bias due to the higher level of attrition within the child-level sample, we also estimate Lee bounds for the primary outcomes of interest for child nutritional status. We do not estimate analogous bounds for household-level livelihoods outcomes, given the extremely low level of attrition at the household level. Given the multiarm trial design, we estimate bounds separately for pairwise comparisons of each treatment arm to the control arm. This differs from the pooled specification used in the main analysis, so the naïve (unbounded) estimates underlying the bounds may not align exactly with the previously reported estimates.³⁸ The findings in Table A19 suggest that the previously reported positive effects on IYCF knowledge and diet diversity can be bounded away from zero in both T2 and T3.³⁹

4.6 Understanding magnitudes

We can further interpret the magnitude of our findings relative to the existing literature including cash transfers and BCC (summarized in Table A23),⁴⁰ noting that these interventions can vary considerably despite the fact that they share a common theory of change. A typical BCC package involves monthly group sessions delivered over a two-year period, while the program evaluated in Ahmed et al. (2025) implemented a more intensive weekly schedule. Lower-intensity interventions include only two sessions of BCC in Bouguen and Dillon (2026), and monthly automated calls analyzed in Weaver et al. (2024) effectively reached fewer than half of recipients. Total cash transfer amounts also vary substantially, from roughly 114 \$PPP in Levere et al. (2024) to about 1,371 \$PPP in Ahmed et al. (2025), with the maternal cash transfers provided in T3 in this trial (451 \$PPP) near the middle of the distribution and the T3 and livelihoods grant package (1,014 \$PPP) among the largest. Only our study and Bouguen and Dillon (2026) add a livelihoods component: here through VESA formation, light-touch livelihoods and financial training, and (within a subset) a one-time livelihoods grant, and in Bouguen and Dillon (2026), a livestock transfer.

Figures 2, 3, and 4 summarize the estimated treatment effects on child height-for-age Z-scores (HAZ), stunting, and dietary diversity from this literature.⁴¹ For dietary diversity,

³⁸A similar bounding strategy in a multiarm trial design is used, for example, in Bhanot et al. (2021).

³⁹Given the limited differential attrition across arms (see columns 1 and 4 in Table A14), the trimming fractions used in constructing the Lee bounds are small, and the resulting bounds are correspondingly narrow.

⁴⁰Little et al. (2021) provide a broader review of BCC and cash-plus programs targeting early childhood development. Our review differs because we focus only on RCTs, incorporate studies published since 2021, and exclude non-academic evaluation reports.

⁴¹One study summarized in Table A23, Premand and Barry (2022), is not included in these cross-study figures. Their outcomes are standardized relative to the control group, and the comparable effects cannot be recovered from the published results. Nonetheless, their findings broadly align with patterns in the rest of the literature: they report sizable gains in children’s dietary diversity, coupled with precisely estimated

nearly all studies report positive effects with effect sizes between 10 and 20 percent (Figure 2) relative to the control group mean; we report percentage effects as some studies employ food group definitions that are different from the eight-group definition provided by WHO and UNICEF (2021).⁴² Our estimates for T3 and the T3 and livelihoods grant combination (18%) are in the upper part of this distribution, while Ahmed et al. (2025) reports a substantially larger estimate of 48%.

Despite these gains in diets, it is clear in Figure 3 and 4 that this set of interventions has generally struggled to shift child anthropometry — and thus the modest findings reported in this paper are far from unusual. Across studies, impacts on height-for-age are modest and statistically significant improvements remain uncommon: the small treatment effect estimate for T3 is nonetheless similar in magnitude to effects reported in comparable trials, whereas the larger impact in the T3 and livelihoods grant group falls toward the upper end of the distribution, comparable to the estimates reported in Ahmed et al. (2025) and Carneiro et al. (2021). For stunting (Figure 4), the pattern is broadly similar: similar to our treatment effect estimate in T3, most studies report small negative but statistically insignificant effects. The T3 and livelihoods grant estimate is larger and again comparable to the larger effects reported in Ahmed et al. (2025) and Carneiro et al. (2021), though it is less precisely estimated, likely due to the smaller sample size in this sub-sample.

Across studies, dietary diversity shows the most consistent relationship with program intensity. Interventions that combine more frequent BCC engagement, higher cash transfers, or additional livelihoods support generally deliver larger relative gains. This includes the weekly BCC model in Ahmed et al. (2025), the multi-component T3 and livelihoods grant in our study, and mid-intensity programs such as Carneiro et al. (2021). Lower-intensity designs—those with minimal BCC content or small transfer amounts—typically produce more modest improvements. Overall, the cross-study pattern suggests that richer cash-plus packages are more effective at shifting dietary practices. The same pattern is generally evident for child anthropometry, with the notable exception of Bouguen and Dillon (2026), where an intervention including minimal BCC nonetheless reports relatively large reductions in stunting.

5 Conclusion

We present new evidence on the effects of a nutrition-sensitive graduation model embedded within Ethiopia’s large-scale Productive Safety Net Programme. SPIR combines a pack-

null effects on height-for-age.

⁴²This indicator is not reported in Bouguen and Dillon (2026).

age of livelihoods interventions—including village economic and social associations, training, and livelihoods grants directed to a subsample of recipients—with peer-led behavior change communication delivered through neighborhood groups and, in one arm, supplementary maternal cash transfers. The trial followed nearly 3,000 PSNP beneficiary households over three years.

This experimental design enables us to assess the effects of the combined SPIR package—including livelihoods interventions and nurturing care groups—on infant and child feeding practices and anthropometric outcomes, and to isolate the additional role of monthly maternal cash transfers in easing liquidity constraints that may limit improvements in child nutrition. It also allows us to evaluate the influence of the bundled livelihoods interventions—which were implemented in both treatment arms—on core economic outcomes, including consumption, savings, and asset accumulation.

We find that across nutrition outcomes, the model without maternal cash transfers increased maternal knowledge of IYCF practices but did not translate into improvements in children’s diets. In contrast, adding the maternal cash transfers produced sizable gains in diet quality: the probability that a child met the minimum acceptable diet increased by 11 percentage points (about a 120% increase relative to the control mean). These gains were accompanied by meaningful improvements in early childhood development, with development scores (measured using the caregiver-reported CREDI tool) increasing by 0.12 standard deviations at three years. Despite these advances in child development and diet diversity, effects on child growth remain limited. Small and weakly significant improvements in height-for-age appear only in the arm receiving cash transfers, and larger, more precisely estimated gains are found only among the subset of the poorest households that received the additional \$300 livelihoods grant.

Turning to livelihoods outcomes, the interventions bundled into SPIR—village savings groups, training, and one-time livelihoods grants—generated meaningful economic gains, though again with differences by treatment arm. Households in both arms increased their participation in savings groups and expanded their access to savings and credit. There is also evidence of modest improvements in per-capita consumption and a persistent increase in livestock holdings and livestock income. However, these gains are generally larger, and in some cases only present, in the arm that also received the maternal cash transfers. This pattern is consistent with the hypothesis that while the core livelihoods package yielded some more minor benefits, only the additional cash resources enabled more sustained investments—particularly in livestock—yielding stronger long-run improvements in economic well-being. This mechanism echoes findings from Northern Nigeria, where predictable transfers paired with BCC supported productive investment and subsequent improvements in

household earnings (Carneiro et al., 2021). More broadly, our results align with recent evidence from Ethiopia showing that lighter-touch graduation approaches may enhance financial inclusion and resilience but rarely generate transformative gains in overall household well-being (Hirvonen et al., 2025; Leight et al., 2026).

Our findings collectively suggest that a nutrition-sensitive graduation model can improve child diets and strengthen household livelihoods, but that meaningful reductions in child undernutrition likely require more intensive models that include substantial cash support beyond the base PSNP. Models implemented without these larger transfers were insufficient to shift child nutrition outcomes. Overall, the results underscore the importance of coupling behavior change communication and livelihoods support with adequate household resources if graduation-oriented social protection programs are to achieve sustained improvements in both economic and nutrition outcomes.

References

- Ahmed, A., Hoddinott, J., and Roy, S. (2025). Food transfers, cash transfers, behavior change communication and child nutrition: evidence from bangladesh. *The World Bank Economic Review*, 39(2):439–472.
- Alderman, H., Friedman, J., Ganga, P., Kak, M., and Rubio-Codina, M. (2021). Assessing the performance of the caregiver reported early development instruments (credi) in rural india. *Annals of the New York Academy of Sciences*, 1492(1):58–72.
- Alderman, H., Gilligan, D. O., Leight, J., Mamo, T., Mulford, M., and Tabet, H. (2025a). Scalable nutrition interventions in a graduation model program: Experimental evidence from ethiopia. *Economic Development and Cultural Change*, 73(3):1559–1604.
- Alderman, H., Gilligan, D. O., Leight, J., Mulford, M., and Tabet, H. (2025b). Men can cook: Effectiveness of a men’s engagement intervention to change attitudes and behaviors in rural ethiopia. *World Development*, 185:106781.
- Alderman, H. and Headey, D. (2018). The timing of growth faltering has important implications for observational analyses of the underlying determinants of nutrition outcomes. *PloS ONE*, 13:e0195904.
- Alderman, H., Hoddinott, J., and Kinsey, B. (2006). Long term consequences of early childhood malnutrition. *Oxford Economic Papers*, 58:450–474.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, 103:1481–1495.
- Andrews, C., Hsiao, A., and Ralston, L. (2018). Social safety nets promote poverty reduction, increase resilience, and expand opportunities. In Beegle, K., Coudouel, A., and Monsalve, E., editors, *Realizing the Full Potential of Social Safety Nets in Africa*. The World Bank, Washington D.C.
- Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., and Heil, A. (2022). Why do people stay poor? *Quarterly Journal of Economics*, 137:785–844.
- Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. (2017). Labor markets and poverty in village economies. *Quarterly Journal of Economics*, 132:811–870.

- Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., and Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348:1260799.
- Banerjee, A., Duflo, E., and Sharma, G. (2021). Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights*, 3:471–486.
- Banerjee, A., Karlan, D., Osei, R., Trachtman, H., and Udry, C. (2022). Unpacking a multifaceted program to build sustainable income for the very poor. *Journal of Development Economics*, 155:102781.
- Barker, N., Karlan, D., Udry, C., and Wright, K. (2024). The fading treatment effects of a multifaceted asset-transfer program in ethiopia. *American Economic Review: Insights*, 6(2):277–294.
- Bedoya, G., Coville, A., Haushofer, J., Isaqzadeh, M., and Shapiro, J. P. (2019). No household left behind: Afghanistan targeting the ultra poor impact evaluation. Technical report, National Bureau of Economic Research.
- Beegle, K., Coudouel, A., and Monsalve, E. (2018). *Realizing the Full Potential of Social Safety Nets in Africa*. The World Bank, Washington D.C.
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.
- Berhane, G., Golan, J., Hirvonen, K., Hoddinott, J. F., Kim, S. S., Taffesse, A. S., Abay, K., Assefa, T. W., Habte, Y., Abay, M. H., Koru, B., Tadesse, F., and Yimer, F. (2020). Evaluation of the nutrition-sensitive features of the fourth phase of ethiopia’s productive safety net programme. IFPRI-ESSP Working Paper 140, International Food Policy Research Institute (IFPRI), Ethiopia Strategy Support Program (ESSP), Addis Ababa.
- Berhane, G., Hirvonen, K., Hoddinott, J., Kim, S., Taffesse, A. S., Abay, K., Hiluf, M. H., Koru, B., Assefa, T., and Yimer, F. (2017). Evaluation of the nutrition sensitive features of the productive safety nets programme iv: Baseline survey report. Unpublished report.
- Berhane, G., Hoddinott, J., Kumar, N., and Margolies, A. (2015). The impact of ethiopia’s productive safety net programme on the nutritional status of children: 2008–2012. IFPRI-ESSP Working Paper 99, International Food Policy Research Institute (IFPRI), Washington D.C.

- Bhanot, S. P., Crost, B., Leight, J., Mvukiyehe, E., and Yedgenov, B. (2021). Can community service grants foster social and economic integration for youth? a randomized trial in kazakhstan. *Journal of Development Economics*, 153:102718.
- Bossuroy, T., Goldstein, M., Karimou, B., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C. C., Udry, C., and Vaillant, J. (2022). Tackling psychosocial and capital constraints to alleviate poverty. *Nature*, 605:291–297.
- Bouguen, A. and Dillon, A. (2026). Beyond poverty reduction: Evidence from a multifaceted program on poverty, nutrition and child development. *Journal of Development Economics*, 178:103596.
- Brune, L., Karlan, D., Kurdi, S., and Udry, C. (2022). Social protection amidst social upheaval: Examining the impact of a multi-faceted program for ultra-poor households in yemen. *Journal of Development Economics*, 155:102780.
- Carneiro, P., Kraftman, L., Mason, G., Moore, L., Rasul, I., and Scott, M. (2021). The impacts of a multifaceted prenatal intervention on human capital accumulation in early life. *American Economic Review*, 111(8):2506–2549.
- Cunha, J. M., De Giorgi, G., and Jayachandran, S. (2019). The price effects of cash versus in-kind transfers. *The Review of Economic Studies*, 86:240–281.
- Currie, J. and Vogl, T. (2013). Early-life health and adult circumstance in developing countries. *Annual Review of Economics*, 5:1–36.
- Davis, T. P., Wetzell, C., Avilan, E. H., de Mendoza Lopes, C., Chase, R. P., Winch, P. J., and Perry, H. B. (2013). Reducing child global undernutrition at scale in sofala province, mozambique, using care group volunteers to communicate health messages to mothers. *Global Health: Science and Practice*, 1:35–51.
- Dercon, S. and Porter, C. (2014). Live aid revisited: Long-term impacts of the 1984 ethiopian famine on children. *Journal of the European Economic Association*, 12:927–948.
- Egger, D., Haushofer, J., Miguel, E., Niehaus, P., and Walker, M. W. (2022). General equilibrium effects of cash transfers: experimental evidence from kenya. *Econometrica*, 90.
- FAO (2023). *Shiny RIMA – Preparing your dataset for a successful Shiny RIMA analysis*. Rome. Available at: <https://openknowledge.fao.org/server/api/core/bitstreams/9e7e8406-3323-491f-bf2d-0a44776a5f79/content>.

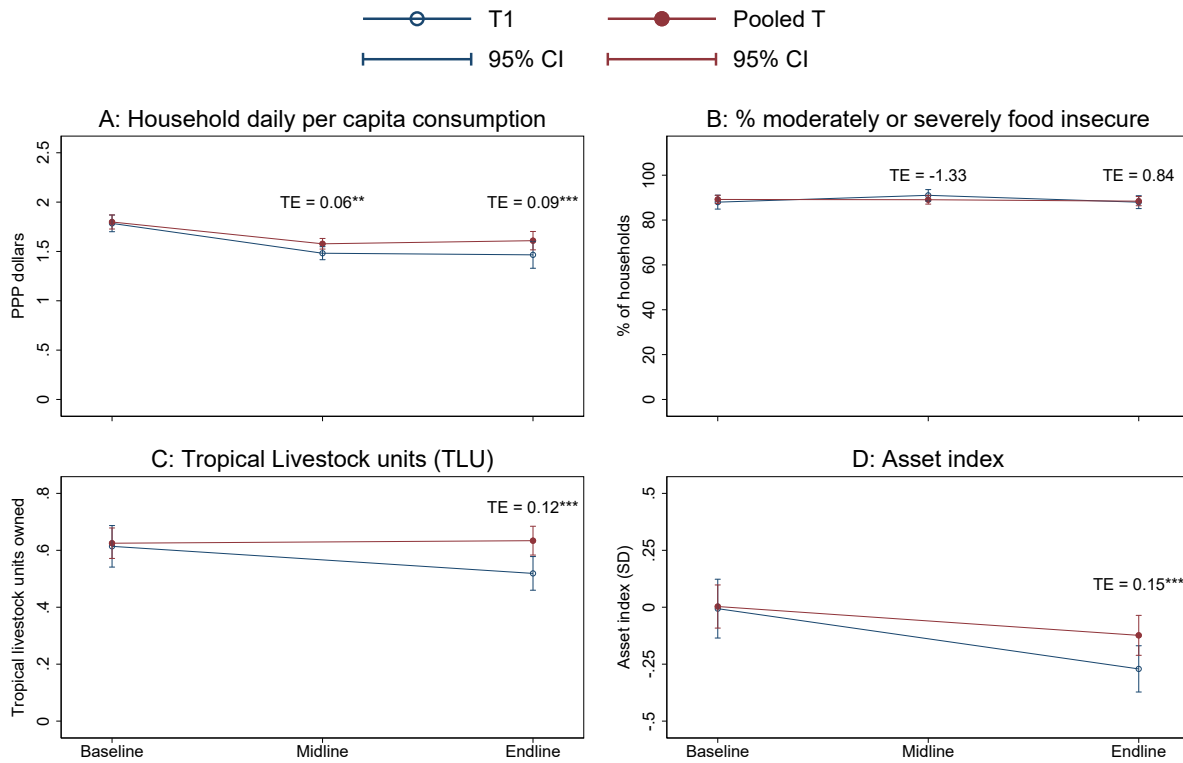
- FAO, IFAD, UNICEF, WFP, and WHO (2022). *The State of Food Security and Nutrition in the World 2022: Repurposing food and agricultural policies to make healthy diets more affordable*. FAO, IFAD, UNICEF, WFP and WHO, Rome.
- Field, E. and Maffioli, E. M. (2025). Are behavioral change interventions needed to make cash transfer programs work for children? experimental evidence from myanmar. *Economic Development and Cultural Change*, 73(3):1187–1220.
- Filmer, D., Friedman, J., Kandpal, E., and Onishi, J. (2021). Cash transfers, food prices, and nutrition impacts on ineligible children. *The Review of Economics and Statistics*, pages 1–45.
- Fiszbein, A., Kanbur, R., and Yemtsov, R. (2014). Social protection and poverty reduction: Global patterns and some targets. *World Development*, 61:167–177.
- Gentilini, U. (2022). Cash transfers in pandemic times. *Evidence, practices, and implications from the largest scale up in history*.
- Gilligan, D. O., Hoddinott, J., and Taffesse, A. (2009). The impact of ethiopia’s productive safety net programme and its linkages. *The Journal of Development Studies*, 45:1684–1706.
- Golan, J., Headey, D., Hirvonen, K., and Hoddinott, J. (2019). Changes in child undernutrition rates in ethiopia, 2000-16. In Cramer, C., Cheru, F., and Oqubay, A., editors, *The Oxford Handbook of the Ethiopian Economy*. Oxford University Press, Oxford.
- Han, Y., Park, S., Kim, J., and Hoddinott, J. (2022). 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*. In press.
- Hirvonen, K., Abate, G. T., Berhane, G., Gilligan, D. O., Hidrobo, M., Hoddinott, J. F., Leight, J., and Taffesse, A. S. (2025). Graduating from ethiopia’s productive safety net programme: What have we learned? Technical report, International Food Policy Research Institute.
- Hirvonen, K., Gilligan, D. O., Leight, J., Tabet, H., and Villa, V. (2023). *Do ultra-poor graduation programs build resilience against droughts? Evidence from rural Ethiopia*. Intl Food Policy Res Inst.
- Hirvonen, K., Wolle, A., Laillou, A., Vinci, V., Chitekwe, S., and Baye, K. (2024). Understanding delays in the introduction of complementary foods in rural ethiopia. *Maternal & child nutrition*, 20:e13247.

- Hoddinott, J., Berhane, G., Gilligan, D. O., Hirvonen, K., Kumar, N., Lind, J., Sabates-Wheeler, R., and Taffesse, A. S. (2024). Securing food, building livelihoods? a 15-year appraisal of ethiopia’s productive safety net programme. Technical Report 2024/76, UNU-WIDER Working Paper.
- Hoddinott, J., Berhane, G., Gilligan, D. O., Kumar, N., and Taffesse, A. (2012). The impact of ethiopia’s productive safety net programme and related transfers on agricultural productivity. *Journal of African Economies*, 21:761–786.
- Kimutai, J., Barnes, C., Zachariah, M., Philip, S. Y., Kew, S. F., Pinto, I., Wolski, P., Koren, G., Vecchi, G., Yang, W., et al. (2025). Human-induced climate change increased 2021–2022 drought severity in horn of africa. *Weather and Climate Extremes*, 47:100745.
- Leight, J., Alderman, H., Gilligan, D., Hidrobo, M., and Mulford, M. (2026). Can a light-touch graduation model enhance livelihood outcomes? evidence from ethiopia. *Journal of Development Economics*, 179.
- Leight, J., Alderman, H., Gilligan, D. O., Hidrobo, M., Mulford, M., and Tadesse, E. (2023). Barriers to adoption of optimal complementary feeding practices in ethiopia: A formative qualitative investigation: Evidence from spir ii. SPIR learning brief 7, International Food Policy Research Institute (IFPRI), Washington D.C.
- Leight, J., Hirvonen, K., Karachiwalla, N., and Rakshit, D. (2025). Displacement and development: Evidence from a graduation program for somalia’s ultra-poor. Technical report, International Food Policy Research Institute.
- Leight, J., Hirvonen, K., and Zafar, S. (2024). The effectiveness of cash and cash plus interventions on livelihoods outcomes: Evidence from a systematic review and meta-analysis. Technical report, International Food Policy Research Institute (IFPRI), Washington D.C.
- Leroy, J. L., Olney, D. K., and Ruel, M. (2016). Evaluating nutrition-sensitive programs: challenges, methods, and opportunities. In Covic, N. and Hendriks, S., editors, *Achieving a nutrition revolution for Africa: the road to healthier diets and optimal nutrition*, pages 130–146. International Food Policy Research Institute, Washington D.C.
- Leroy, J. L., Ruel, M., Habichet, J.-P., and Frongillo, E. A. (2014). 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*, 144:1460–1466.

- Levere, M., Acharya, G., and Bharadwaj, P. (2024). The role of information and cash transfers on early-childhood development: Short-and long-run evidence from nepal. *Economic Development and Cultural Change*, 72(3):1267–1293.
- Little, M. T., Roelen, K., Lange, B. C., Steinert, J. I., Yakubovich, A. R., Cluver, L., and Humphreys, D. K. (2021). Effectiveness of cash-plus programmes on early childhood outcomes compared to cash transfers alone: A systematic review and meta-analysis in low-and middle-income countries. *PLoS medicine*, 18(9):e1003698.
- Maccini, S. and Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review*, 99:1006–1026.
- Mahumud, R. A., Uprety, S., Wali, N., Renzaho, A. M., and Chitekwe, S. (2022). The effectiveness of interventions on nutrition social behaviour change communication in improving child nutritional status within the first 1000 days: Evidence from a systematic review and meta-analysis. *Maternal & child nutrition*, 18(1):e13286.
- Maluccio, J. A. (2010). The impact of conditional cash transfers on consumption and investment in nicaragua. *The Journal of Development Studies*, 46:14–38.
- Mangham-Jefferies, L., Mathewos, B., Russell, J., and Bekele, A. (2014). How do health extension workers in ethiopia allocate their time? *Human resources for health*, 12(1):61.
- Manley, J., Alderman, H., and Gentilini, U. (2022). More evidence on cash transfers and child nutritional outcomes: a systematic review and meta-analysis. *BMJ Global Health*, 7:e008233.
- Manley, J., Balarajan, Y., Malm, S., Harman, L., Owens, J., Murthy, S., Stewart, D., Winder-Rossi, N., and Khurshid, A. (2020). Cash transfers and child nutritional outcomes: a systematic review and meta-analysis. *BMJ Global Health*, 5:e003621.
- McCoy, D. C., Waldman, M., Team, C. F., and Fink, G. (2018). Measuring early childhood development at a global scale: Evidence from the caregiver-reported early development instruments. *Early childhood research quarterly*, 45:58–68.
- Perry, H., Morrow, M., Borger, S., Weiss, J., DeCoster, M., Davis, T., and Ernst, P. (2015). Care groups i: an innovative community-based strategy for improving maternal, neonatal, and child health in resource-constrained settings. *Global Health: Science and Practice*, 3(3):358–369.

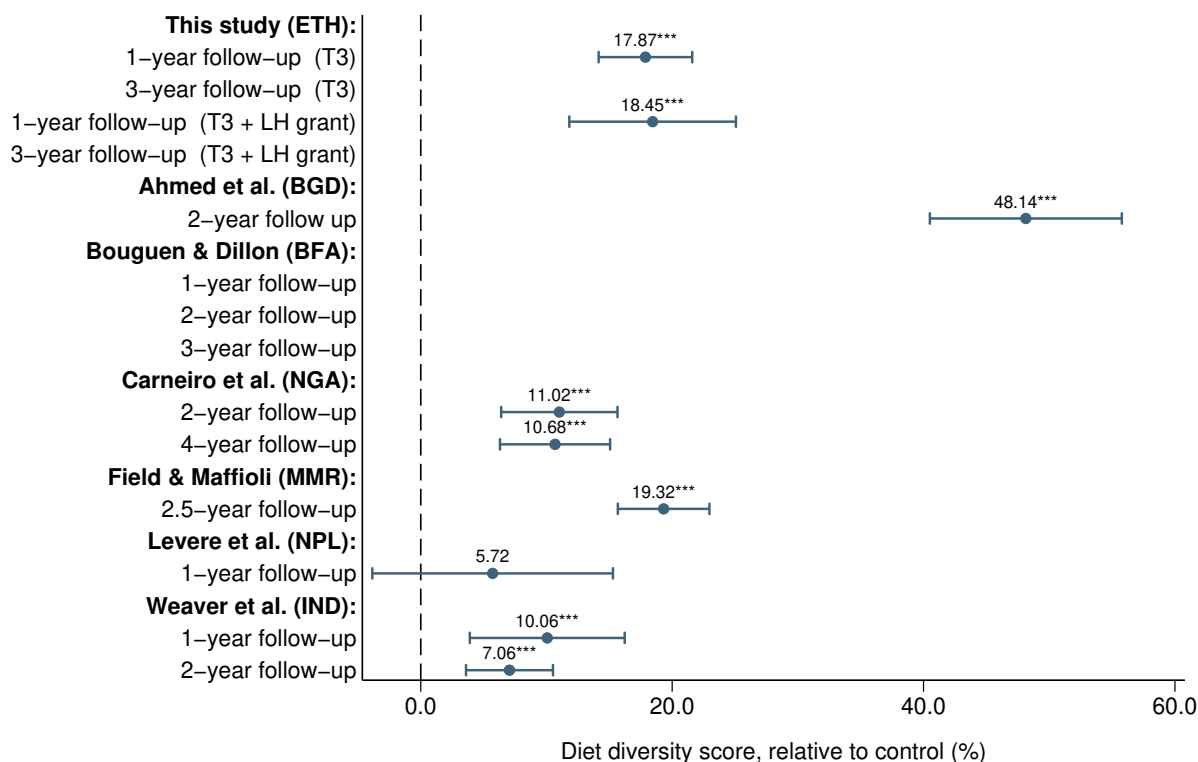
- Prado, E. L., Larson, L. M., Cox, K., Bettencourt, K., Kubes, J. N., and Shankar, A. H. (2019). Do effects of early life interventions on linear growth correspond to effects on neurobehavioural development? a systematic review and meta-analysis. *The Lancet Global Health*, 7(10):e1398–e1413.
- Premand, P. and Barry, O. (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.
- Raza, W., Van de Poel, E., and Van Ourti, T. (2018). 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*, 62:105–120.
- Ruel, M. T. and Alderman, H. (2013). Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? *The Lancet*, 382:536–551.
- Victora, C., de Onis, M., Hallal, P., Blossner, M., and Shrimpton, R. (2010). Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics*, 125:e473–480.
- Weaver, J., Sukhtankar, S., Niehaus, P., and Muralidharan, K. (2024). Cash transfers for child development: Experimental evidence from india. Technical report, National Bureau of Economic Research.
- WHO and UNICEF (2021). *Indicators for assessing infant and young child feeding practices: definitions and measurement methods*. World Health Organization and United Nations Children’s Fund, Geneva.
- World Bank (2025). Restructuring paper on a proposed project restructuring of strengthen ethiopia’s adaptive safety net. <https://documents1.worldbank.org/curated/en/099061225160018785/pdf/P172479-b55d9e81-4faf-4491-bb77-8de15099a137.pdf>, accessed October 22, 2025.
- World Health Organization (2006). Who child growth standards based on length/height, weight and age. *Acta Paediatrica*, Suppl 450:76–85.

Figure 1: Longitudinal trends in select livelihoods outcomes



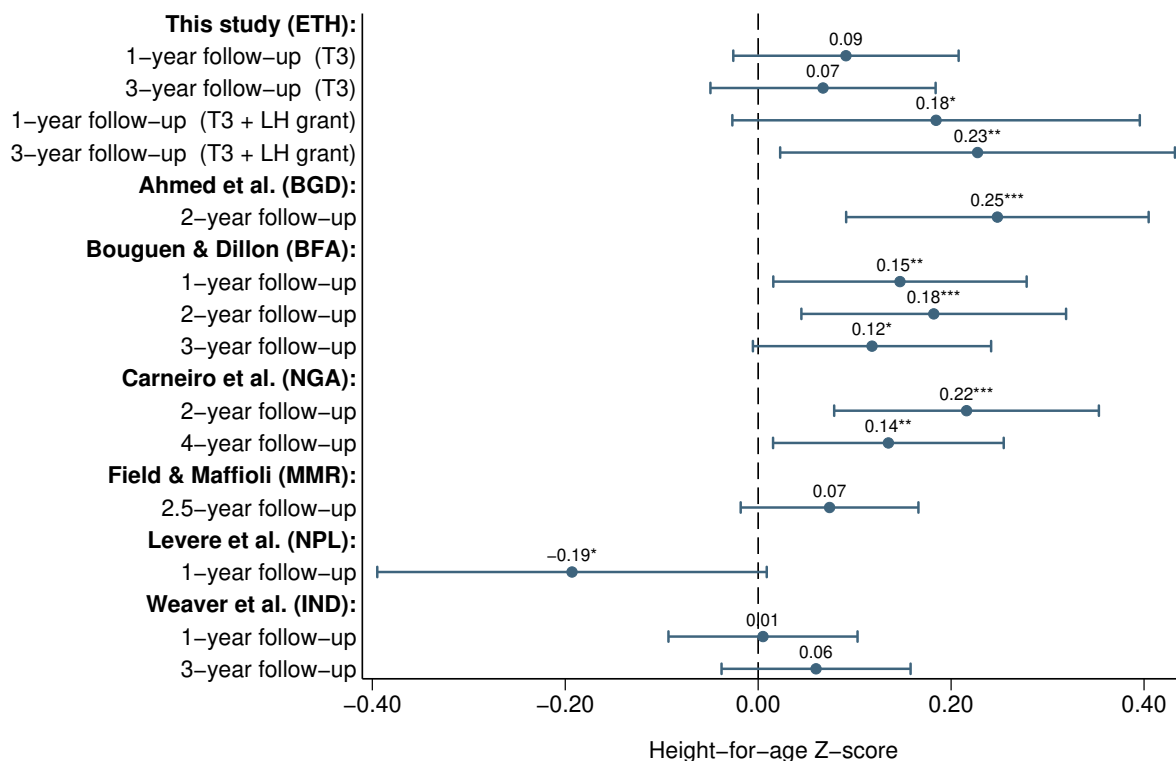
Notes: These graphs report treatment effects in both the one year follow-up (when measured) and the three year follow-up. Hollow and solid dots indicate the control and pooled treatment group means, respectively, and capped bars represent 95% confidence intervals. Mean values reflect unadjusted group averages. Treatment effects (TE) are estimated from regressions controlling for baseline consumption, household covariates, and strata fixed effects; hence, TE magnitudes may not exactly equal the raw mean differences shown. Statistical significance is denoted as * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Tropical livestock units and assets were not measured at midline.

Figure 2: Relative impacts on child dietary diversity across cash-plus-BCC evaluations



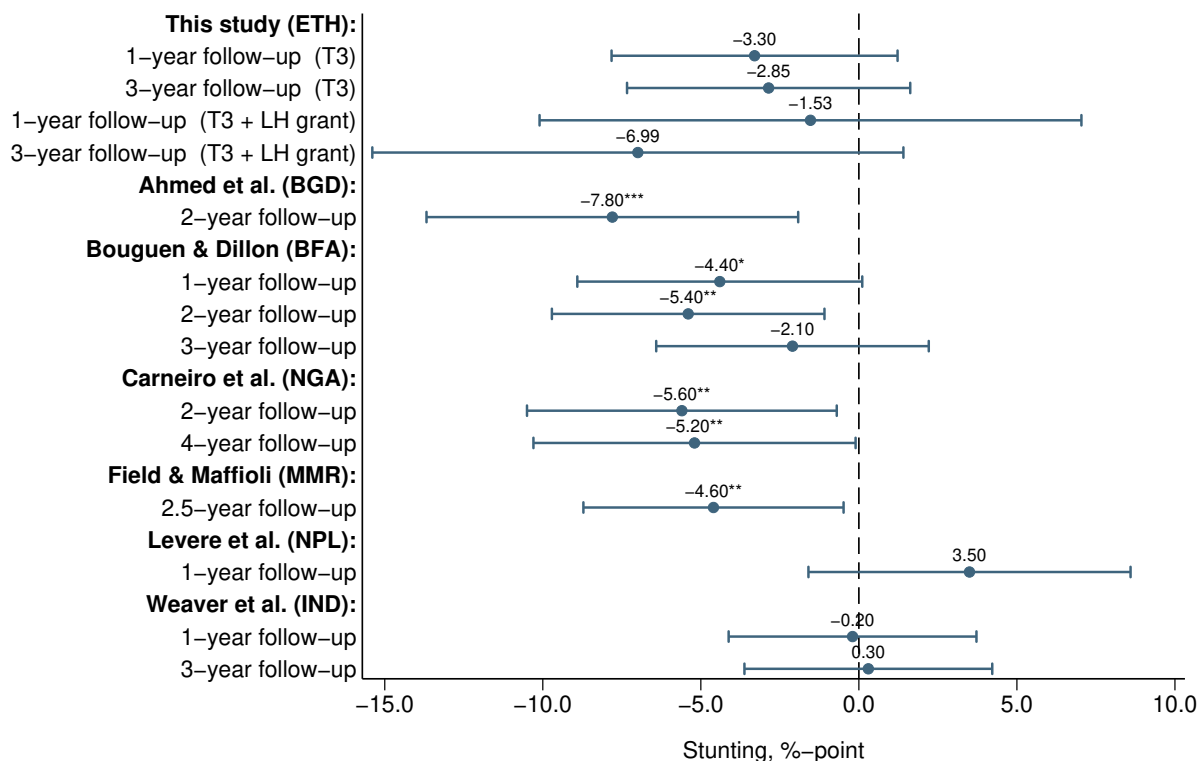
Notes: BCC = Behavior change communication; LH grant = Livelihoods grant; ETH = Ethiopia; BGD = Bangladesh; BFA = Burkina Faso; NGA = Nigeria; NPL = Nepal; IND = India. Several studies employed multi-arm designs; in such cases, we report estimates from the most expansive treatment arm relative to the control group. For Ahmed et al., the estimate corresponds to the cash + BCC arm in North Bangladesh. For Bouguen & Dillon, the estimate reflects the arm receiving cash transfers plus an animal transfer and the full nutrition package. For Field & Maffioli, the estimate is for a cash + BCC arm, and for Levere et al. BCC + cash arm. The effects are expressed as percentage changes relative to the control group to account for heterogeneous food-group definitions across studies. Levere et al. include a three-year follow-up, but outcomes are reported only as a weighted index rather than separate measures. Statistical significance is denoted as * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Confidence intervals and significance levels for all studies shown in the figure are computed using the point estimates and standard errors reported in the respective papers. In some cases, these may differ slightly from the published values because authors often calculate p-values using unrounded estimates, whereas we rely on the rounded coefficients and standard errors presented in the tables.

Figure 3: Impacts on child height-for-age (HAZ) relative to prior cash-plus-BCC evaluations



Notes: BCC = Behavior change communication; LH grant = Livelihoods grant; ETH = Ethiopia; BGD = Bangladesh; BFA = Burkina Faso; NGA = Nigeria; NPL = Nepal; IND = India. Several studies employed multi-arm designs; in such cases, we report estimates from the most expansive treatment arm relative to the control group. For Ahmed et al., the estimate corresponds to the cash + BCC arm in North Bangladesh. For Bouguen & Dillon, the estimate reflects the arm receiving cash transfers plus an animal transfer and the full nutrition package. For Field & Maffioli, the estimate is for a cash + BCC arm, and for Levere et al. BCC + cash arm. Levere et al. include a three-year follow-up, but outcomes are reported only as a weighted index rather than separate measures. Statistical significance is denoted as * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Confidence intervals and significance levels for all studies shown in the figure are computed using the point estimates and standard errors reported in the respective papers. In some cases, these may differ slightly from the published values because authors often calculate p-values using unrounded estimates, whereas we rely on the rounded coefficients and standard errors presented in the tables.

Figure 4: Impacts on child stunting relative to prior cash-plus-BCC evaluations



Notes: BCC = Behavior change communication; LH grant = Livelihoods grant; ETH = Ethiopia; BGD = Bangladesh; BFA = Burkina Faso; NGA = Nigeria; NPL = Nepal; IND = India. Several studies employed multi-arm designs; in such cases, we report estimates from the most expansive treatment arm relative to the control group. For Ahmed et al., the estimate corresponds to the cash + BCC arm in North Bangladesh. For Bouguen & Dillon, the estimate reflects the arm receiving cash transfers plus an animal transfer and the full nutrition package. For Field & Maffioli, the estimate is for a cash + BCC arm, and for Leveré et al. BCC + cash arm. Leveré et al. include a three-year follow-up, but outcomes are reported only as a weighted index rather than separate measures. Statistical significance is denoted as * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Confidence intervals and significance levels for all studies shown in the figure are computed using the point estimates and standard errors reported in the respective papers. In some cases, these may differ slightly from the published values because authors often calculate p-values using unrounded estimates, whereas we rely on the rounded coefficients and standard errors presented in the tables.

Table 1: Baseline balance

Variable	(1)	(2)	(3)	(1)-(2)	(1)-(3)	(2)-(3)
	T1 Mean/(SE)	T2 Mean/(SE)	T3 Mean/(SE)	Mean difference	Pairwise t-test Mean difference	Mean difference
Currently pregnant	0.527 (0.005)	0.536 (0.005)	0.533 (0.007)	-0.010	-0.007	0.003
Child is 0-3 months old	0.215 (0.011)	0.180 (0.010)	0.192 (0.014)	0.034**	0.023	-0.011
Child is 4-6 months old	0.143 (0.011)	0.145 (0.011)	0.140 (0.010)	-0.002	0.003	0.005
Child is 7-9 months old	0.113 (0.010)	0.136 (0.012)	0.133 (0.010)	-0.023*	-0.020	0.003
Household size	5.407 (0.118)	5.375 (0.116)	5.401 (0.122)	0.032	0.006	-0.026
Primary female has some formal education	0.444 (0.021)	0.423 (0.019)	0.428 (0.021)	0.021	0.016	-0.005
Mother's IYCF knowledge score (max 11 points)	8.454 (0.154)	8.522 (0.157)	8.327 (0.167)	-0.068	0.127	0.195
Food Insecurity Experience Scale (FIES), raw score	6.151 (0.112)	6.219 (0.095)	6.034 (0.110)	-0.068	0.117	0.184*
Household daily per capita consumption (2011 PPP)	1.785 (0.042)	1.784 (0.046)	1.813 (0.055)	0.000	-0.029	-0.029
Tropical Livestock Units	0.614 (0.037)	0.606 (0.038)	0.644 (0.039)	0.008	-0.030**	-0.038
Durable asset index (PCA)	0.023 (0.035)	0.011 (0.045)	-0.035 (0.020)	0.013	0.058	0.045
Household reports formal savings	0.520 (0.030)	0.516 (0.029)	0.541 (0.030)	0.004	-0.022	-0.025
F-test of joint significance (F-stat)				0.996	0.993	1.257
Number of observations	1033	992	990	2025	2023	1982
Number of clusters	80	77	77	157	157	154

Notes: The value displayed for t-tests are p-values, derived from regressions including strata fixed effects and standard errors clustered at the kebele 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. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table 2: Program exposure

	T1		T2		T3	
	%	(N)	%	(N)	%	(N)
Ever participated in neighborhood group (one year)	22.7	(961)	73.6	(927)	87.3	(948)
Ever participated in neighborhood group (three years)	24.1	(999)	59.6	(935)	78.5	(968)
Ever participated in neighborhood group (pooled)	36.2	(1028)	84.4	(989)	92.8	(989)
Ever received maternal cash grant (one year)	1.2	(961)	1.8	(924)	88.7	(947)
Ever received maternal cash grant (three years)	2.6	(999)	3.4	(935)	80.6	(968)
Ever received maternal cash grant (pooled)	3.5	(1028)	5.0	(989)	92.6	(989)
Ever participated in VESA (one year)	15.8	(1004)	63.2	(970)	71.4	(980)
Ever participated in VESA (three years)	32.0	(999)	78.3	(935)	82.6	(968)
Ever participated in VESA (pooled)	38.8	(1028)	87.2	(989)	91.1	(990)
Receipt of livelihoods cash transfer (one year)	4.0	(1004)	38.4	(970)	35.2	(980)

Note: This table reports on exposure to SPIR interventions across treatment arms. VESA = Village Economic and Social Association.

Table 3: Nutrition

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: One-year follow-up (children are 6-23 months)						
	IYCF knowledge [†]	Diet diversity [†]	MAD	HAZ	Stunted	WHZ
T2	0.28*** (0.09)	0.11* (0.06)	0.01 (0.02)	-0.01 (0.06)	0.01 (0.02)	-0.09 (0.06)
q-value	0.001***	0.006***	0.273	0.330	0.276	0.070*
T3	0.43*** (0.09)	0.58*** (0.06)	0.11*** (0.02)	0.09 (0.06)	-0.03 (0.02)	0.05 (0.06)
q-value	0.001***	0.001***	0.001***	0.075*	0.089*	0.194
ANCOVA?	Yes	No	No	No	No	No
T2 vs T3 (p-value)	0.052*	0.000***	0.000***	0.089*	0.048**	0.011**
Control mean	9.51	3.25	0.09	-1.61	0.37	-0.53
Observations	2231	2233	2221	2226	2226	2226
Panel B: Three-year follow-up (children are 24-48 months)						
	IYCF knowledge [†]	HAZ [†]	Stunted	WHZ	CREDI	
T2	0.29*** (0.08)	-0.03 (0.06)	0.01 (0.02)	-0.07 (0.05)	0.03 (0.04)	
q-value	0.001***	0.276	0.276	0.103	0.210	
T3	0.27*** (0.07)	0.07 (0.06)	-0.03 (0.02)	0.01 (0.05)	0.12*** (0.04)	
q-value	0.001***	0.142	0.120	0.307	0.002***	
ANCOVA?	Yes	No	No	No	No	
T2 vs T3 (p-value)	0.824	0.104	0.098*	0.123	0.042**	
Control mean	9.16	-2.21	0.57	-0.43	0.01	
Observations	2163	2203	2203	2203	2195	

Notes: [†] = Primary outcome. IYCF knowledge = maternal knowledge of infant and young child feeding practices; Diet diversity = score measuring the number of food groups from which the child consumed food in the previous day; MAD = binary indicator equal to 1 if the child achieved a minimum acceptable diet; HAZ = height-for-age Z-score; Stunted = binary indicator equal to 1 if the child's HAZ < -2; WHZ = weight-for-height Z-score; CREDI = Early childhood development score, standardized so that control group has zero mean and standard deviation of 1. All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education, a dummy for the index child being male, and the index child's age in months. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table 4: Livelihoods: Consumption and food security

	(1)	(2)	(3)
Panel A: Consumption and food security, one year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
Pooled T	0.06** (0.03)	-0.01 (0.01)	-0.10 (0.07)
q-value	0.012**	0.123	0.071*
ANCOVA?	Yes	Yes	Yes
Control mean	1.48	0.91	5.91
Observations	2954	2954	2954
Panel B: Consumption and food security, three year follow-up			
	Per capita cons.	Food security binary	Food security cont.
Pooled T	0.09*** (0.04)	0.01 (0.01)	-0.03 (0.07)
q-value	0.007***	0.214	0.273
ANCOVA?	Yes	Yes	Yes
Control mean	1.47	0.88	5.96
Observations	2902	2902	2902

Notes: All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

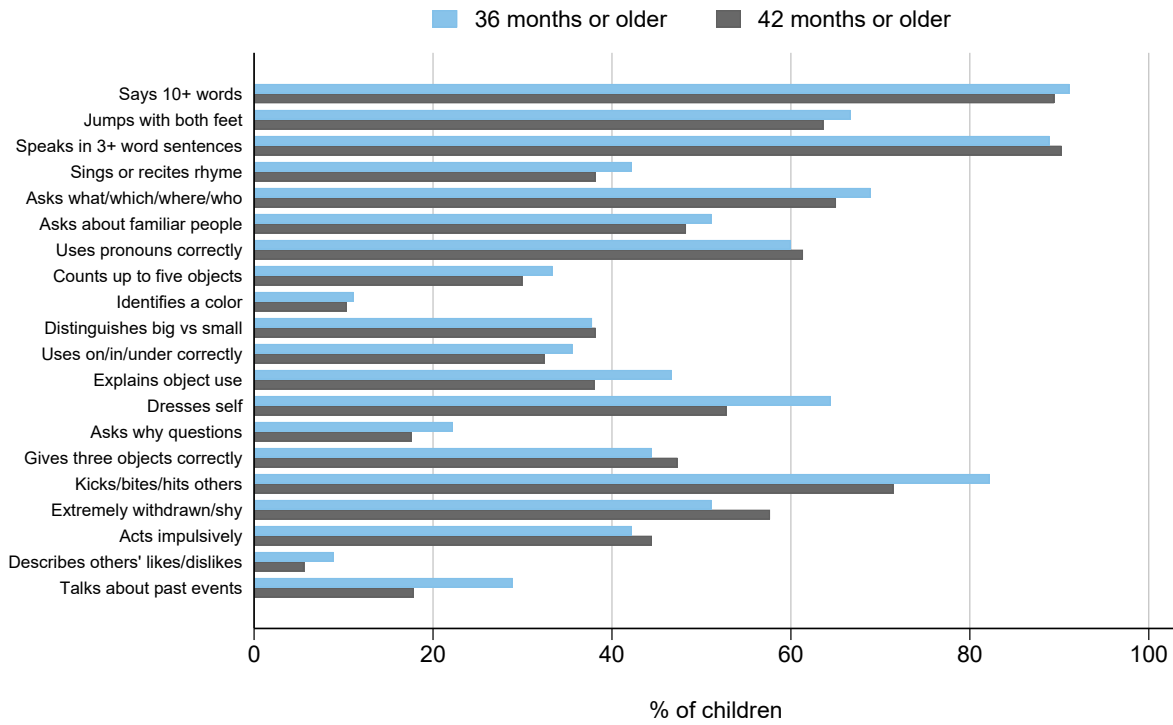
Table 5: Livelihoods: Assets, financial inclusion, and income

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Assets and financial inclusion, three year follow-up						
	Total asset index	TLU	Any savings	Amount saved	Any credit	Amount credit
Pooled T	0.15*** (0.04)	0.12*** (0.03)	0.22*** (0.02)	15.21*** (1.67)	0.07*** (0.02)	48.49*** (9.48)
q-value	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
ANCOVA?	Yes	Yes	Yes	Yes	No	No
Control mean	-0.27	0.52	0.74	20.37	0.68	113.05
Observations	2902	2902	2902	2902	2902	2902
Panel B: Income, three year follow-up						
	Any net income: livestock	Net income: livestock	Any net income: non-ag.prod.	Net income: non-ag.prod.		
Pooled T	0.10*** (0.01)	32.64*** (5.66)	0.02 (0.02)	-12.99 (8.16)		
q-value	0.001***	0.001***	0.170	0.070*		
ANCOVA?	No	No	No	No		
Control mean	0.77	58.66	0.28	76.29		
Observations	2902	2902	2902	2902		

Notes: All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education. TLU = Tropical Livestock Units Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Appendix

Figure A1: CREDI milestone attainment among children older than 36 and 42 months



Notes: Bars show the share of children achieving each CREDI developmental milestone among those aged 36 months or older and 42 months or older. The figure illustrates that many older children have not yet reached all CREDI milestones, indicating limited ceiling effects. This pattern supports the inclusion of slightly older children in the analysis.

Table A1: Summary of trial arms and intervention components

Study arm	Description	Interventions	Clusters
T1	PSNP only	Core PSNP transfers (public works or direct support).	80
T2	PSNP + SPIR II graduation programming + Nurturing Care Groups (NCGs)	<i>Livelihoods/financial inclusion:</i> village economic and savings associations (VESAs); light-touch livelihoods and financial training; <i>subset:</i> one-time livelihoods grant of \$300. <i>Nutrition-sensitive:</i> peer-led Nurturing Care Groups delivering behavior change communication on IYCF and health.	77
T3	PSNP + SPIR II graduation programming + NCGs + maternal cash transfers	All T2 components, plus a monthly maternal grant of \$20 for 24 months.	77

Notes: Randomization was conducted at the kebele (subdistrict) level across 15 woredas, stratified by woreda, PSNP coverage, and distance to the woreda capital. Three kebeles were dropped due to conflict at baseline, yielding 234 kebeles.

Table A2: Sample composition by region and pregnancy status at baseline

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

Notes: This table reports the sample composition at baseline. N = 3,015 households.

Table A3: Deviations from the pre-analysis plan (PAP)

Category	Description
Timeline	The PAP specified that the first follow-up survey would be conducted in August–September 2023 and the second follow-up survey in August–September 2025. The first survey was conducted largely as planned 12 months post-baseline, though data collection in the Amhara region was delayed until October–December 2023 due to temporary conflict-related inaccessibility. The second follow-up survey was fielded approximately three months earlier than planned to avoid potential confounding with the anticipated PSNP-wide rollout of livelihoods grants by the Government of Ethiopia. As a result, children surveyed in the three year follow-up were between 24 and 48 months of age, rather than the originally intended 30 to 48 months specified in the PAP. This modest shift expanded the lower tail of the intended age range, meaning that a subset of children were slightly younger than planned at follow-up. Given that growth faltering typically occurs in the first 24 months of life (Victora et al., 2010), this deviation is unlikely to affect interpretation of estimated impacts on child growth.
Sample composition	There was substantial sample loss at the child level between baseline and both follow-up surveys. This was driven primarily by women who were enrolled while pregnant but did not have a live infant present at the one-year follow-up: 34% of pregnant women did not have an infant at follow-up. Women self-reported that this was largely due to pregnancy loss, though there were anecdotal reports that some women may have feigned pregnancy at baseline in the expectation of accessing benefits targeted to pregnant or lactating women. As a result, the sample of eligible children meeting the age criteria was 2,234 in the one-year follow-up (rather than the approximately 3,000 originally projected) and 2,203 in the three-year follow-up. The implications for statistical power are modest: for the primary outcome of height-for-age, the minimum detectable effect increased from 0.14 to 0.18 standard deviations given the reduced cluster size.
Exploratory analyses	Several exploratory analyses were added post hoc to better interpret mechanisms and longer-run dynamics. These are reported in Appendix Tables A5 and A7 and include estimating treatment effects on diarrhea prevalence, paternal IYCF knowledge, the child stimulation index, open defecation, PSNP participation, and livelihoods-related outcomes such as formal and informal employment, formal credit access, and the value of crop harvests. We also added exploratory short-term analyses of maternal engagement with health and nutrition services (Table A6). In addition, the final study includes detailed attrition analyses that were also not pre-specified (Tables A13, A14, and A19).

Notes: This table summarizes deviations from the pre-analysis plan (PAP) that was peer-reviewed and conditionally accepted under the *Journal of Development Economics* pre-results review track.

Table A4: Description of the outcome variables

Outcome Variable	Definition	When assessed?
Primary outcomes:		
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).	Midline, Endline
Number of food groups consumed by a child 6–23 months	A count variable capturing the number of food groups consumed by the child. Following WHO and UNICEF (2021) guidelines, eight food groups were used: breastmilk; grains, roots, and tubers; legumes and nuts; dairy products; flesh foods; eggs; vitamin A rich fruits and vegetables; and other fruits and vegetables.	Midline
Height-for-age (children 24–48 months)	Each child’s measured length/height and age in months were used to compute height-for-age z-scores (HAZ) following World Health Organization (2006) Child Growth Standards. Values below –6.0 or above 6.0 were treated as outliers and set to missing.	Endline
Secondary outcomes: Nutrition and child development		
Height-for-age (children 6-23 months)	See above.	Midline
Stunting (children 24-48 months)	A binary indicator obtaining value 1 if HAZ <-2 and zero otherwise	Endline
Weight-for-height (children 6-23 months)	Each child’s measured weight and length/height were used to compute weight-for-height z-scores (WHZ) following World Health Organization (2006) Child Growth Standards. Values below –5.0 or above 5.0 were treated as outliers and set to missing.	Midline
Weight-for-height (children 24-48 months)	See above.	Endline
Receiving a minimum acceptable diet (MAD) (children 6–23 months)	A binary indicator equal to one if the child met both minimum dietary diversity (consumption of foods from at least five of the eight WHO food groups) and minimum meal frequency, following WHO and UNICEF (2021) definitions.	Midline

Continued on next page

Outcome Variable	Definition	When assessed?
Early childhood development score	Computed using the Caregiver-Reported Early Development Instruments (CREDI) short form (McCoy et al., 2018). The score summarizes caregiver-reported attainment of age-appropriate milestones across motor, cognitive, language, and socio-emotional domains. At younger ages, the instrument focuses on the emergence of foundational developmental milestones, while at older ages it captures the consolidation and increasing complexity of previously acquired skills. Because the CREDI instrument is designed for children aged 0–36 months, child age was top-coded at 36 months when computing the score; this affected about 38% of the children in our sample who were 37–48 months at endline. Prior validation work suggests that the CREDI maintains reliability and shows limited ceiling effects among slightly older children (Alderman et al., 2021), a pattern also observed among children aged 37–48 months in this sample. The final CREDI scores are standardized relative to the control group mean and standard deviation within child age categories, so that the resulting z-scores have a mean of zero and a standard deviation of one among control children within each age group (the overall control mean across groups is approximately zero).	Endline
Secondary outcomes: Livelihoods		
Per capita consumption	The sum of ‘Food consumption’ and ‘Non-food expenditures’. Food consumption: Aggregated consumption based on 57 food items with a 7-day recall, initially measured in quantities, then converted to monetary values using price estimates from a food price opinion module. Non-food expenditures: Aggregated expenditures based on 39 non-food items or services with a 1-month or 12-month recall, depending on expected purchasing frequency. Measured in 2011 purchasing power parity (PPP) dollars, using PPP exchange rates for household final consumption expenditure provided by the World Bank.	Midline and Endline
Total asset index	Based on 55 asset categories, including durable, productive, and livestock assets. Calculated using principal components analysis (PCA) on a pooled data containing both baseline and endline rounds. The index is then normalized with respect to baseline so that at baseline it gets zero mean and SD of 1.	Endline

Continued on next page

Outcome Variable	Definition	When assessed?
Household savings	Measured as both a binary and continuous variable. The binary indicator equals one if the household reported savings in any formal institution, including Rural Savings and Credit Cooperatives (RuSACCOs), Microfinance Institutions, Village Savings and Loan Associations (VSLAs), or bank accounts. The continuous measure captures the total amount of household savings (in 2011 PPP dollars) held in these formal institutions.	Endline
Food insecurity (FIES)	Measured using the Food Insecurity Experience Scale (FIES), based on eight yes/no questions capturing household experiences of food access constraints over the past 12 months. Responses were used to construct a continuous food insecurity score (0–8), with higher values indicating greater food insecurity, and a binary indicator identifying households experiencing moderate or severe food insecurity (FIES>3).	Midline and Endline
Net income from livestock production	Defined as total revenues plus change in stock minus total costs. Revenues include imputed value of live animal sales (valued at current unit values), slaughtered meat (sold + consumed), cow/goat milk, and eggs (milk and eggs valued using district-median prices). Change in stock equals the value of current holdings minus the value one year ago, both valued at current prices. Costs include hired labor, purchased feed, medicines, veterinary services, other inputs, and insurance premiums. All subcomponents were top-winsorized at the 95th percentile before aggregation.	Endline
Net income from any non-agricultural production	Calculated as the sum of household formal wage income and net income from non-agricultural enterprises. Wage income includes all payments received by household members over the past 12 months from formal wage employment. Formal wage employment is assumed to be non-agricultural, defined as work with either a written contract or a contract lasting more than two weeks. Enterprise income equals total revenues from household-operated non-agricultural businesses minus operating costs.	Endline
Household livestock asset index	Constructed using Tropical Livestock Units (TLU) to aggregate the number of animals owned across species into a single standardized measure of livestock holdings. Species-specific conversion factors follow FAO (2023) guidelines, e.g., cattle = 0.5, pack animals = 0.328, sheep/goats = 0.1, poultry = 0.01.	Endline

Continued on next page

Outcome Variable	Definition	When assessed?
Credit access	Measured as both a binary and continuous variable. The binary indicator equals one if any household member reported borrowing from any source in the past 12 months, including formal institutions (banks, Microfinance Institutions, Village Savings and Loan Associations, Savings and Credit Cooperative Organizations) and informal sources such as friends, relatives, shops, or informal lending groups. The continuous measure captures the total amount borrowed from all sources over the past 12 months (in 2011 PPP dollars).	Endline
Additional outcomes (not prespecified)		
Early childhood development score	See above.	Midline
Diarrhea prevalence	A binary indicator equal to 1 if the household practices open defecation or unsafe disposal of excreta, defined as not having or using a latrine (simple pit, ventilated improved pit [VIP], flush, or chemical toilet) and instead defecating in open spaces or fields.	Midline and Endline
Paternal IYCF knowledge	Same as caregiver IYCF knowledge described above, but the respondent is primary male.	Endline
Child stimulation index	The total number (0–6) of stimulation activities that any household member aged 15 or older engaged in with the index child during the three days preceding the interview. Activities include reading, storytelling, singing, taking the child outside, playing, and naming, counting, or drawing with the child.	Endline
Open defecation	A binary indicator equal to 1 if the household lacks access to an improved toilet facility.	Endline
Any formal credit	A binary indicator for households that reported obtaining credit from formal sources (from a bank, microfinance institution or other formal financial institution) in the past 12 months.	Endline
Formal employment	A binary indicator for households that reported formal employment in the past 12 months. Formal employment is defined as work based on either a written contract or a verbal contract lasting at least two weeks.	Endline
Informal employment	A binary indicator for households that reported informal employment in the past 12 months. Informal employment is defined as work without a written contract and with a contract lasting less than two weeks.	Endline
Crop harvest	The total estimated value of the most recent crop harvest (from up to the three most important crops), measured in 2017 PPP dollars.	Endline

Continued on next page

Outcome Variable	Definition	When assessed?
PSNP	A binary indicator for households that were part of the PSNP during the current calendar year.	Endline

Table A5: Additional nutrition-related outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Consumption by food group for 6–23 month olds (one year follow-up)								
	Breast-milk	Staples	Legumes or nuts	Dairy	Meat	Eggs	Vit. A F&V	Other F&V
T2	0.00 (0.01)	0.00 (0.01)	0.01 (0.02)	0.00 (0.02)	0.00 (0.01)	0.03 (0.02)	0.03* (0.02)	0.03 (0.02)
T3	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)	0.17*** (0.03)	0.01 (0.01)	0.22*** (0.02)	0.10*** (0.02)	0.05** (0.02)
ANCOVA?	No	No	No	No	No	No	No	No
T2 vs T3 (p-value)	0.561	0.123	0.985	0.000***	0.477	0.000***	0.003***	0.358
Control mean	0.94	0.96	0.47	0.23	0.01	0.06	0.19	0.40
Observations	2233	2233	2233	2233	2233	2233	2233	2233
Panel B: Other outcomes								
	CREDI one year	Diarrhea prevalence (one year)	Paternal IYCF knowl (three years)	Child stim. (three years)	Open def. (three years)	Diarrhea prevalence (three years)		
T2	0.03 (0.04)	-0.04 (0.02)	0.24** (0.11)	0.03 (0.06)	-0.02 (0.03)	0.00 (0.02)		
T3	0.19*** (0.05)	-0.06*** (0.02)	0.22** (0.11)	0.22*** (0.06)	-0.08*** (0.03)	-0.10*** (0.02)		
ANCOVA?	No	No	Yes	No	Yes	No		
T2 vs T3 (p-value)	0.000***	0.379	0.888	0.005***	0.044**	0.000***		
Control mean	-0.00	0.37	8.63	1.82	0.43	0.32		
Observations	2212	2233	1316	2210	2210	2209		

Notes: This table reports regressions for additional nutrition-related variables that were not pre-specified. Panel A reports binary variables for the child eating food from the specified food group in the one year follow-up. Panel B reports the CREDI child development index, a binary variable for open defecation, and a binary variable for any child experience of diarrhea in the last two weeks, in the one year follow-up; and a child stimulation index and the same variables for child defecation and experience of diarrhea in the three year follow-up. All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education, a dummy for the index child being male, and the index child’s age in months. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A6: Maternal engagement with health and nutrition services, one-year follow-up

	(1)	(2)	(3)	(4)	(5)
	Contact with HEW	Health post visit	Contact with HDA	Food demonstration	Community conversation
T2	0.100*** (0.026)	0.069** (0.027)	0.153*** (0.020)	0.231*** (0.024)	0.185*** (0.022)
T3	0.139*** (0.026)	0.065** (0.026)	0.169*** (0.019)	0.380*** (0.025)	0.301*** (0.022)
ANCOVA?	No	No	No	No	No
T2 vs T3 (<i>p</i> -value)	0.110	0.886	0.418	0.000***	0.000***
Control mean	0.551	0.603	0.124	0.084	0.082
Observations	2227	2231	2213	2231	2231

Notes: This table reports regressions estimating the effects of the interventions on maternal engagement with health and nutrition services at the one year follow-up; these outcomes were not pre-specified. All outcomes are binary, taking value 1 if: the woman had contact with a health extension worker (HEW) in the past three months (column 1); visited a health post for herself or her child in the past three months (column 2); had contact with a health development army (HDA) member in the past three months (column 3); attended a food demonstration in the past three months (column 4); or attended a community conversation in the past three months (column 5). All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education, a dummy for the index child being male, and the index child’s age in months. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table A7: Additional livelihoods-related outcomes: three year follow-up

	(1)	(2)	(3)	(4)	(5)
Panel A: Livestock outcomes (endline)					
	Num. cattle	Num. sheep or goats	Num. pack animals	Num. poultry	
Pooled T	0.14*** (0.04)	0.24*** (0.06)	0.05** (0.02)	0.71*** (0.14)	
ANCOVA?	Yes	Yes	Yes	Yes	
Control mean	0.66	0.75	0.29	1.87	
Observations	2902	2902	2902	2902	
Panel B: Credit, employment, crop income and PSNP					
	Any formal credit	Formal empl.	Informal empl.	Crop harvest PPP	PSNP
Pooled T	0.18*** (0.02)	0.00 (0.01)	-0.07*** (0.02)	11.02 (13.14)	0.01 (0.01)
ANCOVA?	No	No	No	No	No
Control mean	0.14	0.05	0.39	326.62	0.97
Observations	2902	2902	2902	2902	2902

Notes: This table reports regressions for additional livelihoods variables that were not pre-specified. Panel A reports binary variables for ownership of various livestock types. Panel B reports binary variables for any access to formal credit and formal and informal wage employment, as well as the total estimated value of the most recent crop harvest and whether the household was still part of the PSNP at endline. Note that formal and informal employment variables are not mutually exclusive, as households may engage in both types of employment. All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education, a dummy for the index child being male, and the index child’s age in months. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table A8: Livelihoods (all arms)

	(1)	(2)	(3)
Panel A: Consumption and food security, one year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
T2	0.01 (0.03)	-0.00 (0.01)	-0.01 (0.07)
T3	0.11*** (0.03)	-0.02* (0.01)	-0.19** (0.08)
ANCOVA?	Yes	Yes	Yes
T2 vs T3 (p-value)	0.003***	0.032**	0.007***
Control mean	1.48	0.91	5.91
Observations	2954	2954	2954
Panel B: Consumption and food security, three year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
T2	0.02 (0.04)	0.02 (0.01)	0.07 (0.08)
T3	0.14*** (0.04)	0.00 (0.02)	-0.12 (0.08)
ANCOVA?	Yes	Yes	Yes
T2 vs T3 (p-value)	0.006***	0.322	0.028**
Control mean	1.47	0.88	5.96
Observations	2902	2902	2902

Notes: All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table A9: Livelihoods (all arms)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Assets and financial inclusion, three year follow-up						
	Total asset index	TLU	Any savings	Amount saved	Any credit	Amount credit
T2	0.09** (0.04)	0.07*** (0.03)	0.20*** (0.02)	11.69*** (1.78)	0.08*** (0.02)	54.00*** (11.70)
T3	0.21*** (0.04)	0.16*** (0.03)	0.23*** (0.02)	18.26*** (2.17)	0.07*** (0.02)	43.49*** (10.47)
ANCOVA?	Yes	Yes	Yes	Yes	No	No
T2 vs T3 (p-value)	0.006***	0.001***	0.048**	0.004***	0.816	0.356
Control mean	-0.27	0.52	0.74	20.37	0.68	113.05
Observations	2902	2902	2902	2902	2902	2902
Panel B: Income, three year follow-up						
	Any net income: livestock	Net income: livestock	Any net income: non-ag.prod.	Net income: non-ag.prod.		
T2	0.07*** (0.02)	21.92*** (7.11)	0.01 (0.02)	-17.23* (9.42)		
T3	0.12*** (0.02)	41.30*** (6.35)	0.02 (0.02)	-11.83 (9.27)		
ANCOVA?	No	No	No	No		
T2 vs T3 (p-value)	0.003***	0.009***	0.563	0.576		
Control mean	0.77	58.66	0.28	76.29		
Observations	2902	2902	2902	2902		

Notes: All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table A10: Nutrition, heterogeneity by paternal IYCF knowledge at baseline

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: One-year follow-up (Children are 6-23 months)						
	IYCF knowledge [†]	Diet diversity [†]	MAD	HAZ	Stunted	WHZ
T2	0.36** (0.15)	0.22*** (0.08)	0.02 (0.02)	0.08 (0.11)	-0.01 (0.04)	-0.05 (0.08)
T2 X paternal IYCF knowl.	-0.01 (0.20)	-0.27* (0.14)	-0.02 (0.04)	-0.07 (0.17)	0.03 (0.07)	-0.07 (0.12)
T3	0.53*** (0.15)	0.74*** (0.10)	0.12*** (0.03)	0.10 (0.10)	-0.05 (0.04)	0.09 (0.08)
T3 X paternal IYCF knowl.	-0.16 (0.21)	-0.12 (0.15)	0.01 (0.05)	0.13 (0.17)	-0.04 (0.06)	-0.11 (0.14)
ANCOVA?	Yes	No	No	No	No	No
T2 vs T3 (p-value)	0.194	0.000***	0.003***	0.847	0.282	0.092*
Control mean	9.51	3.25	0.09	-1.61	0.37	-0.53
Observations	1478	1480	1471	1473	1473	1473
Panel B: Three-year follow-up (Children are 24-48 months)						
	IYCF knowledge [†]	HAZ [†]	Stunted	WHZ	CREDI	
T2	0.39*** (0.11)	0.01 (0.10)	0.01 (0.04)	-0.09 (0.08)	0.05 (0.07)	
T2 X paternal IYCF knowl.	-0.14 (0.18)	-0.02 (0.17)	-0.01 (0.07)	0.01 (0.12)	0.05 (0.12)	
T3	0.31*** (0.11)	0.09 (0.10)	-0.04 (0.04)	-0.02 (0.08)	0.11 (0.07)	
T3 X paternal IYCF knowl.	0.05 (0.21)	0.04 (0.17)	-0.06 (0.07)	-0.01 (0.12)	0.12 (0.13)	
ANCOVA?	Yes	No	No	No	No	
T2 vs T3 (p-value)	0.447	0.391	0.171	0.395	0.409	
Control mean	9.16	-2.21	0.57	-0.43	0.01	
Observations	1425	1449	1449	1449	1445	

Notes: [†] = Primary outcome. IYCF knowledge = maternal knowledge of infant and young child feeding practices; Diet diversity = score measuring the number of food groups from which the child consumed food in the previous day; MAD = binary indicator equal to 1 if the child achieved a minimum acceptable diet; HAZ = height-for-age Z-score; Stunted = binary indicator equal to 1 if the child's HAZ < -2; WHZ = weight-for-height Z-score. All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1-3 years, 4-7 years, or 8 or more years of education, a dummy for the index child being male, and the index child's age in months. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table A11: Nutrition, households eligible for livelihood transfer

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: One-year follow-up (children are 6-23 months)						
	IYCF knowledge [†]	Diet diversity [†]	MAD	HAZ	Stunted	WHZ
T2	0.17 (0.14)	0.23** (0.10)	0.04 (0.03)	0.08 (0.11)	0.01 (0.05)	-0.01 (0.09)
q-value	0.029**	0.010***	0.067*	0.159	0.220	0.249
T3	0.33** (0.14)	0.57*** (0.11)	0.13*** (0.03)	0.18* (0.11)	-0.02 (0.04)	0.10 (0.09)
q-value	0.010***	0.001***	0.001***	0.050*	0.220	0.122
ANCOVA?	Yes	No	No	No	No	No
T2 vs T3 (p-value)	0.185	0.003***	0.008***	0.345	0.495	0.227
Control mean	9.59	3.11	0.06	-1.65	0.36	-0.55
Observations	664	665	659	664	664	664
Panel B: Three-year follow-up (children are 24-48 months)						
	IYCF knowledge [†]	HAZ [†]	Stunted	WHZ	CREDI	
T2	0.44*** (0.16)	0.17 (0.12)	-0.06 (0.05)	0.02 (0.09)	0.10 (0.09)	
q-value	0.005***	0.077*	0.092*	0.239	0.122	
T3	0.43*** (0.14)	0.23** (0.10)	-0.07 (0.04)	0.05 (0.09)	0.24*** (0.09)	
q-value	0.003***	0.019**	0.053*	0.184	0.005***	
ANCOVA?	Yes	No	No	No	No	
T2 vs T3 (p-value)	0.982	0.586	0.817	0.679	0.112	
Control mean	8.95	-2.32	0.61	-0.48	-0.15	
Observations	648	660	660	660	657	

Notes: [†] = Primary outcome. IYCF knowledge = maternal knowledge of infant and young child feeding practices; Diet diversity = score measuring the number of food groups from which the child consumed food in the previous day; MAD = binary indicator equal to 1 if the child achieved a minimum acceptable diet; HAZ = height-for-age Z-score; Stunted = binary indicator equal to 1 if the child's HAZ < -2; WHZ = weight-for-height Z-score. All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1-3 years, 4-7 years, or 8 or more years of education, a dummy for the index child being male, and the index child's age in months. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A12: Nutrition, households not eligible for livelihood transfer

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: One-year follow-up (children are 6-23 months)						
	IYCF knowledge [†]	Diet diversity [†]	MAD	HAZ	Stunted	WHZ
T2	0.31*** (0.10)	0.04 (0.07)	-0.01 (0.02)	-0.04 (0.08)	0.01 (0.03)	-0.11* (0.06)
q-value	0.002***	0.065*	0.273	0.273	0.273	0.056*
T3	0.46*** (0.10)	0.58*** (0.07)	0.11*** (0.02)	0.07 (0.07)	-0.05* (0.03)	0.04 (0.06)
q-value	0.001***	0.001***	0.001***	0.184	0.069*	0.269
ANCOVA?	Yes	No	No	No	No	No
T2 vs T3 (p-value)	0.112	0.000***	0.000***	0.130	0.043**	0.024**
Control mean	9.47	3.31	0.10	-1.59	0.38	-0.53
Observations	1567	1568	1562	1562	1562	1562
Panel B: Three-year follow-up (children are 24-48 months)						
	IYCF knowledge [†]	HAZ [†]	Stunted	WHZ	CREDI	
T2	0.23** (0.10)	-0.11 (0.08)	0.04 (0.03)	-0.10 (0.06)	0.00 (0.05)	
q-value	0.009***	0.096*	0.097*	0.074*	0.308	
T3	0.20** (0.09)	0.01 (0.07)	-0.01 (0.03)	0.01 (0.06)	0.07 (0.05)	
q-value	0.009***	0.308	0.273	0.308	0.096*	
ANCOVA?	Yes	No	No	No	No	
T2 vs T3 (p-value)	0.767	0.075*	0.062*	0.101	0.196	
Control mean	9.25	-2.16	0.56	-0.41	0.08	
Observations	1515	1543	1543	1543	1538	

Notes: [†] = Primary outcome. IYCF knowledge = maternal knowledge of infant and young child feeding practices; Diet diversity = score measuring the number of food groups from which the child consumed food in the previous day; MAD = binary indicator equal to 1 if the child achieved a minimum acceptable diet; HAZ = height-for-age Z-score; Stunted = binary indicator equal to 1 if the child's HAZ < -2; WHZ = weight-for-height Z-score. All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1-3 years, 4-7 years, or 8 or more years of education, a dummy for the index child being male, and the index child's age in months. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A13: Household level attrition analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Attrited at one year			Attrited at three years		
T2	-0.006 (0.008)	-0.010 (0.007)	-0.001 (0.067)	0.025 (0.026)	0.039 (0.023)	-0.023 (0.060)
T3	-0.018*** (0.006)	-0.020*** (0.006)	-0.052 (0.059)	-0.011 (0.019)	0.001 (0.019)	-0.042 (0.056)
Cons. per capita		-0.001 (0.003)	0.006 (0.006)		0.004 (0.004)	-0.000 (0.004)
X T2			-0.012 (0.009)			0.021* (0.013)
X T3			-0.008 (0.006)			-0.007 (0.007)
FIES score		0.000 (0.001)	0.002 (0.003)		0.004*** (0.002)	0.002 (0.003)
X T2			-0.004 (0.004)			0.005 (0.004)
X T3			-0.002 (0.003)			0.002 (0.005)
Asset index		-0.002 (0.004)	-0.002 (0.008)		0.003 (0.004)	-0.001 (0.008)
X T2			0.007 (0.010)			0.018 (0.018)
X T3			-0.004 (0.008)			-0.004 (0.012)
IYCF knowledge		-0.003** (0.001)	-0.004 (0.003)		0.001 (0.002)	-0.001 (0.003)
X T2			0.000 (0.003)			0.003 (0.005)
X T3			0.003 (0.003)			0.003 (0.004)
Household size		-0.001 (0.003)	0.002 (0.005)		-0.002 (0.003)	0.002 (0.005)
X T2			-0.006 (0.007)			-0.013 (0.009)
X T3			-0.004 (0.006)			-0.001 (0.005)
Age prim. female		-0.002* (0.001)	-0.003* (0.002)		-0.001 (0.001)	-0.001 (0.001)
X T2			0.002 (0.002)			0.002 (0.002)
X T3			0.002 (0.002)			0.000 (0.002)
Prim. fem.: 1-3 years educ.		-0.017*** (0.007)	-0.020 (0.013)		-0.010 (0.010)	-0.005 (0.016)
X T2			-0.003 (0.017)			-0.012 (0.024)
X T3			0.011			-0.001

			(0.015)			(0.022)
Prim. fem.: 4-7		-0.006	-0.019		-0.007	-0.002
years. educ.		(0.008)	(0.017)		(0.009)	(0.013)
X T2			0.019			-0.024
			(0.022)			(0.024)
X T3			0.018			0.015
			(0.019)			(0.019)
Prim. fem.: 8+		0.006	0.017		-0.013	-0.003
years educ.		(0.012)	(0.025)		(0.012)	(0.018)
X T2			-0.020			-0.057
			(0.033)			(0.036)
X T3			-0.017			0.012
			(0.026)			(0.028)
Enrolled pregnant		-0.007	-0.006		-0.003	-0.011
		(0.006)	(0.010)		(0.005)	(0.009)
X T2			-0.006			0.000
			(0.015)			(0.015)
X T3			0.003			0.023*
			(0.013)			(0.012)
Constant	0.028***	0.114***	0.123**	0.033**	0.019	0.047
	(0.006)	(0.028)	(0.056)	(0.013)	(0.034)	(0.032)
Observations	3015	3014	3014	3015	3014	3014
p-value (T2=T3)	0.049	0.085	.	0.175	0.123	.
Joint p(T×X)	.	.	0.172	.	.	0.818

Notes: The dependent variable obtains value 1 if the household was not interviewed in the one year follow-up (Columns (1) to (3) or the three year follow-up (Columns (4) to (6)). Regressions presented in columns 2, 3, 5 and 6 control for strata fixed effects. All independent variables are measured at baseline. Standard errors are clustered at the kebele level and reported in parentheses. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table A14: Child-level attrition analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Attrited at one year			Attrited at three years		
T2	-0.001 (0.021)	-0.017 (0.015)	-0.042 (0.140)	0.024 (0.029)	0.018 (0.023)	0.044 (0.158)
T3	-0.026 (0.021)	-0.029** (0.014)	0.048 (0.126)	-0.041 (0.027)	-0.036* (0.021)	0.023 (0.142)
Cons. per capita		0.004 (0.007)	0.015 (0.012)		0.005 (0.008)	0.014 (0.013)
X T2			-0.019 (0.019)			-0.017 (0.022)
X T3			-0.015 (0.015)			-0.012 (0.017)
FIES score		0.007** (0.003)	0.011** (0.005)		0.011*** (0.004)	0.011* (0.006)
X T2			0.003 (0.008)			0.006 (0.010)
X T3			-0.013* (0.007)			-0.006 (0.009)
Asset index		-0.014* (0.007)	-0.018 (0.012)		-0.004 (0.009)	-0.014 (0.015)
X T2			0.016 (0.019)			0.036 (0.026)
X T3			0.002 (0.016)			-0.002 (0.021)
IYCF knowledge		0.003 (0.003)	0.001 (0.005)		0.003 (0.004)	0.002 (0.006)
X T2			0.003 (0.007)			0.002 (0.008)
X T3			0.003 (0.007)			0.002 (0.008)
Household size		-0.003 (0.005)	0.002 (0.008)		-0.005 (0.005)	0.005 (0.009)
X T2			-0.018 (0.011)			-0.029** (0.013)
X T3			-0.000 (0.011)			-0.003 (0.012)
Age prim. female		0.006*** (0.002)	0.005* (0.003)		0.005*** (0.002)	0.005 (0.003)
X T2			0.004 (0.004)			0.004 (0.004)
X T3			-0.000 (0.004)			-0.001 (0.004)
Prim. fem.: 1–3 years educ.		-0.018 (0.017)	-0.027 (0.029)		-0.033* (0.020)	-0.049 (0.034)
X T2			-0.005 (0.041)			-0.000 (0.046)
X T3			0.030			0.050

			(0.042)			(0.047)						
Prim. fem.: 4–7 years educ.	0.019	-0.023	(0.018)	(0.034)	0.025	(0.021)	-0.024	(0.035)				
X T2				0.074*				0.071				
				(0.045)				(0.052)				
X T3				0.048				0.073				
				(0.043)				(0.046)				
Prim. fem.: 8+ years educ.	-0.010	-0.036	(0.023)	(0.041)	-0.024	(0.025)	-0.036	(0.044)				
X T2				0.021				-0.035				
				(0.054)				(0.062)				
X T3				0.049				0.051				
				(0.052)				(0.057)				
Enrolled pregnant	0.308***	0.312***	(0.016)	(0.028)	0.269***	(0.017)	0.276***	(0.030)				
X T2				-0.004				-0.020				
				(0.040)				(0.043)				
X T3				-0.010				-0.001				
				(0.039)				(0.042)				
Constant	0.199***	-0.195***	(0.015)	(0.060)	-0.211**	(0.095)	0.243***	(0.019)	-0.129*	(0.068)	-0.159	(0.107)
Observations	3015	3014	3014	3015	3014	3014	3014	3014				
p-value (T2=T3)	0.243	0.415	.	0.026	0.026	.	.	.				
Joint p(T×X)	.	.	0.707	.	.	.	0.743	.				

Notes: Dependent variable obtains value 1 if the household was not interviewed at endline (columns 1–3) or if the household was not interviewed at endline or if the household did not have an index child at endline. All regressions control for strata fixed effects. All independent variables are measured at baseline. Standard errors are clustered at the kebele level and reported in parentheses. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table A15: Consumption and food security: households eligible for a \$300 transfer

	(1)	(2)	(3)
Panel A: Consumption and food security, one year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
Pooled T	0.11** (0.04)	-0.06*** (0.02)	-0.29*** (0.09)
q-value	0.010***	0.001***	0.002***
ANCOVA?	Yes	Yes	Yes
Control mean	1.47	0.95	6.12
Observations	897	897	897
Panel B: Consumption and food security, three year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
Pooled T	0.11* (0.07)	0.01 (0.02)	-0.20* (0.12)
q-value	0.050*	0.220	0.050*
ANCOVA?	Yes	Yes	Yes
Control mean	1.47	0.90	6.23
Observations	882	882	882

Notes: All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A16: Assets, financial inclusion, and income: households eligible for a \$300 transfer

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Assets and financial inclusion, three year follow-up						
	Total asset index	TLU	Any savings	Amount saved	Any credit	Amount credit
Pooled T	0.27*** (0.05)	0.21*** (0.04)	0.24*** (0.03)	14.42*** (2.80)	0.05* (0.03)	30.35*** (10.24)
q-value	0.001***	0.001***	0.001***	0.001***	0.050*	0.003***
ANCOVA?	Yes	Yes	Yes	Yes	No	No
Control mean	-0.69	0.30	0.73	20.18	0.68	103.39
Observations	882	882	882	882	882	882
Panel B: Income, three year follow-up						
	Any net income: livestock	Net income: livestock	Any net income: non-ag.prod.	Net income: non-ag.prod.		
Pooled T	0.15*** (0.03)	46.69*** (8.07)	0.04 (0.03)	-15.33 (20.14)		
q-value	0.001***	0.001***	0.092*	0.159		
ANCOVA?	No	No	No	No		
Control mean	0.67	35.91	0.32	98.90		
Observations	882	882	882	882		

Notes: All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A17: Consumption and food security: Households not eligible for a \$300 transfer

	(1)	(2)	(3)
Panel A: Consumption and food security, one year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
Pooled T	0.04 (0.03)	0.01 (0.01)	-0.02 (0.07)
q-value	0.096*	0.273	0.298
ANCOVA?	Yes	Yes	Yes
Control mean	1.49	0.89	5.83
Observations	2057	2057	2057
Panel B: Consumption and food security, three year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
Pooled T	0.08** (0.04)	0.01 (0.01)	0.04 (0.08)
q-value	0.017**	0.269	0.273
ANCOVA?	Yes	Yes	Yes
Control mean	1.46	0.87	5.84
Observations	2020	2020	2020

Notes: All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A18: Assets, financial inclusion, and income, households not eligible for a \$300 transfer

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Assets and financial inclusion, three year follow-up						
	Total asset index	TLU	Any savings	Amount saved	Any credit	Amount credit
Pooled T	0.10** (0.04)	0.08*** (0.03)	0.21*** (0.02)	15.52*** (1.80)	0.08*** (0.02)	56.31*** (11.66)
q-value	0.017**	0.007***	0.001***	0.001***	0.001***	0.001***
ANCOVA?	Yes	Yes	Yes	Yes	No	No
Control mean	-0.09	0.61	0.74	20.45	0.69	117.25
Observations	2020	2020	2020	2020	2020	2020
Panel B: Income, three year follow-up						
	Any net income: livestock	Net income: livestock	Any net income: non-ag.prod.	Net income: non-ag.prod.		
Pooled T	0.07*** (0.02)	26.30*** (6.79)	0.01 (0.02)	-12.14 (8.55)		
q-value	0.001***	0.001***	0.273	0.097*		
ANCOVA?	No	No	No	No		
Control mean	0.81	68.57	0.26	66.45		
Observations	2020	2020	2020	2020		

Notes: All regressions control for strata fixed effects and the following baseline covariates: household size, age of the primary female, indicators for the primary female having 1–3 years, 4–7 years, or 8 or more years of education. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A19: Lee bounds estimates for primary nutrition outcomes

Pair	Bound	Coef.	SE	p-value	95% CI	N
IYCF knowledge (one year follow-up)						
T1 vs T2	Lower	0.290***	0.084	0.001	[0.123, 0.456]	1,490
T1 vs T2	Upper	0.290***	0.084	0.001	[0.123, 0.456]	1,490
T1 vs T3	Lower	0.414***	0.088	0.000	[0.239, 0.588]	1,512
T1 vs T3	Upper	0.489***	0.085	0.000	[0.321, 0.657]	1,502
IYCF knowledge (three year follow-up)						
T1 vs T2	Lower	0.308***	0.071	0.000	[0.168, 0.449]	1,418
T1 vs T2	Upper	0.251***	0.070	0.000	[0.113, 0.390]	1,410
T1 vs T3	Lower	0.296***	0.064	0.000	[0.169, 0.424]	1,484
T1 vs T3	Upper	0.455***	0.058	0.000	[0.341, 0.569]	1,458
Diet diversity (one year follow-up)						
T1 vs T2	Lower	0.127***	0.048	0.009	[0.032, 0.222]	1,491
T1 vs T2	Upper	0.127***	0.048	0.009	[0.032, 0.222]	1,491
T1 vs T3	Lower	0.592***	0.062	0.000	[0.470, 0.714]	1,513
T1 vs T3	Upper	0.641***	0.062	0.000	[0.519, 0.763]	1,499
HAZ (three year follow-up)						
T1 vs T2	Lower	0.041	0.058	0.474	[-0.073, 0.155]	1,424
T1 vs T2	Upper	-0.107*	0.063	0.091	[-0.231, 0.017]	1,424
T1 vs T3	Lower	-0.047	0.053	0.382	[-0.151, 0.058]	1,481
T1 vs T3	Upper	0.156***	0.055	0.005	[0.048, 0.265]	1,482

Notes: Each cell reports the estimated treatment effect of T2 or T3 relative to T1 under lower and upper bound scenarios which trim the higher-retention arm so that effective retention matches the lower-retention arm. Standard errors clustered at the kebele level. All regressions include baseline controls and strata fixed effects. Significance levels: *** < 0.01, ** < 0.05, and * < 0.10.

Table A20: Nutrition, no controls

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: One-year follow-up (children are 6-23 months)						
	IYCF knowledge [†]	Diet diversity [†]	MAD	HAZ	Stunted	WHZ
T2	0.27*** (0.09)	0.09 (0.06)	0.01 (0.02)	-0.01 (0.06)	0.01 (0.02)	-0.10* (0.06)
q-value	0.001***	0.012**	0.284	0.401	0.275	0.054*
T3	0.42*** (0.09)	0.58*** (0.06)	0.11*** (0.02)	0.07 (0.06)	-0.03 (0.02)	0.04 (0.06)
q-value	0.001***	0.001***	0.001***	0.133	0.137	0.236
ANCOVA?	Yes	No	No	No	No	No
T2 vs T3 (p-value)	0.049**	0.000***	0.000***	0.191	0.092*	0.016**
Control mean	9.51	3.25	0.09	-1.61	0.37	-0.53
Observations	2231	2233	2221	2226	2226	2226
Panel B: Three-year follow-up (children are 24-48 months)						
	IYCF knowledge [†]	HAZ [†]	Stunted	WHZ	CREDI	
T2	0.28*** (0.08)	-0.04 (0.07)	0.01 (0.02)	-0.08 (0.05)	0.03 (0.04)	
q-value	0.001***	0.249	0.255	0.070*	0.240	
T3	0.27*** (0.07)	0.05 (0.06)	-0.02 (0.02)	0.00 (0.05)	0.11*** (0.04)	
q-value	0.001***	0.208	0.183	0.409	0.006***	
ANCOVA?	Yes	No	No	No	No	
T2 vs T3 (p-value)	0.884	0.134	0.132	0.110	0.054*	
Control mean	9.16	-2.21	0.57	-0.43	0.01	
Observations	2163	2203	2203	2203	2195	

Notes: [†] = Primary outcome. IYCF knowledge = maternal knowledge of infant and young child feeding practices; Diet diversity = score measuring the number of food groups from which the child consumed food in the previous day; MAD = binary indicator equal to 1 if the child achieved a minimum acceptable diet; HAZ = height-for-age Z-score; Stunted = binary indicator equal to 1 if the child's HAZ < -2; WHZ = weight-for-height Z-score. All regressions control for strata fixed effects. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A21: Livelihoods: Consumption and food security, no controls

	(1)	(2)	(3)
Panel A: Consumption and food security, one year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
Pooled T	0.06** (0.03)	-0.01 (0.01)	-0.10 (0.07)
q-value	0.024**	0.133	0.073*
ANCOVA?	Yes	Yes	Yes
Control mean	1.48	0.91	5.91
Observations	2954	2954	2954
Panel B: Consumption and food security, three year follow-up			
	Per capita cons.	Food insecurity binary	Food insecurity cont.
Pooled T	0.09** (0.04)	0.01 (0.01)	-0.03 (0.07)
q-value	0.012**	0.210	0.276
ANCOVA?	Yes	Yes	Yes
Control mean	1.47	0.88	5.96
Observations	2902	2902	2902

Notes: All regressions control for strata fixed effects. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A22: Livelihoods: Assets, financial inclusion, and income, no controls

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Assets and financial inclusion, three year follow-up						
	Total asset index	TLU	Any savings	Amount saved	Any credit	Amount credit
Pooled T	0.15*** (0.04)	0.12*** (0.03)	0.22*** (0.02)	15.13*** (1.67)	0.07*** (0.02)	48.51*** (9.50)
q-value	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
ANCOVA?	Yes	Yes	Yes	Yes	No	No
Control mean	-0.27	0.52	0.74	20.37	0.68	113.05
Observations	2902	2902	2902	2902	2902	2902
Panel B: Income, three year follow-up						
	Any net income: livestock	Net income: livestock	Any net income: non-ag.prod.	Net income: non-ag.prod.		
Pooled T	0.10*** (0.01)	32.05*** (5.72)	0.01 (0.02)	-14.41* (8.00)		
q-value	0.001***	0.001***	0.211	0.047**		
ANCOVA?	No	No	No	No		
Control mean	0.77	58.66	0.28	76.29		
Observations	2902	2902	2902	2902		

Notes: All regressions control for strata fixed effects. TLU = Tropical Livestock Units Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the endline control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10. The reported q-values are p-values adjusted for multiple inference using the false discovery rate correction procedure outlined in Anderson (2008).

Table A23: Overview of the cash+BCC literature

Study	T3 in this study	Ahmed et al. (2025)	Bouguen and Dillon (2026)	Carneiro et al. (2021)	Field and Maffioli (2025)	Levere et al. (2024)	Premand and Barry (2022)	Weaver et al. (2024)
Country	Ethiopia	Bangladesh	Burkina Faso	Nigeria	Myanmar	Nepal	Niger	India
Year launched	2022	2012	2018	2014	2016	2013	2012	2018
Years elapsed	3	2	3	4	2.5	3	2.5	3.5
Sample size (children)	2,233	2,218	4,000	3,688	2,783	2,956	2,460	2,355
BCC	Biweekly cascaded model using trained community volunteers to deliver IYCF, hygiene, care-seeking, and broader caregiving messages through group meetings and targeted home visits.	Weekly, hour-long group sessions delivered by community workers, plus twice-monthly home visits and community engagement.	Fortified-flour distribution plus two brief nutrition trainings (one aimed at husbands) and a small garden kit; no recurring contact.	Mixed-channel BCC including mass media, community/religious outreach, small-group parenting sessions, and home-visit counseling.	Two-stage BCC: initial community mobilization groups followed by intensive multisection training on IYCF, hygiene, health-seeking, and spending for mothers, and fathers, and elder.	Monthly sessions appended to existing meetings, led by trained local health workers on pregnancy care, breastfeeding, complementary feeding, and infant care.	18-month package with monthly village assemblies, small-group sessions, and home visits covering nutrition, health, hygiene, stimulation, and child protection.	Light-touch framing through nutrition messages at registration and monthly automated calls; no group meetings and limited reach.
Cash interval	monthly	monthly	monthly (lean season)	monthly	monthly	monthly	monthly	monthly
Cash duration (months)	24	24	8	30	30	5	24	24
Total cash (\$PPP)	451	1,371	576	459	1,023	114	518	517
Livelihoods	VESAs; light-touch livelihoods and financial training; one-third received a \$300 livelihoods grant (563 \$PPP).	No	Livestock voucher redeemable for poultry or small ruminants; animals monitored for two weeks post-distribution.	No	No	No	No	No

Notes: VESA = Village Economic and Social Association; BCC = Behaviour change communication; LH grant = Livelihoods grant. Several studies used multi-arm designs; we report estimates from the most expansive treatment arm relative to the control group. For Ahmed et al., this corresponds to the cash + BCC arm in North Bangladesh. For Bouguen & Dillon, the effect pertains to the arm receiving cash transfers, an animal transfer, and the full nutrition package. For Field & Maffioli, we report the cash + BCC arm, and for Levere et al., the BCC + cash arm. All monetary values have been converted to 2011 purchasing parity dollars (\$PPP).

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available [here](#)

They can be downloaded free of charge

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

www.ifpri.org

IFPRI HEADQUARTERS

1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org