Can a graduation model reduce child undernutrition? Experimental evidence from Ethiopia

Kalle Hirvonen[1], Jessica Leight[1], Daniel O. Gilligan[1], Hiwot Mekonnen Mesfin[2], Michael Mulford[3], and Haleluya Tesfaye[1]

[1]International Food Policy Research Institute (IFPRI)

[2]World Bank Gender Innovation Lab

[3]Northwestern Global Poverty Research Lab

June 25, 2026

Abstract

We experimentally evaluate a nutrition-sensitive graduation model embedded in a large-scale safety net in Ethiopia. The program combined savings groups, behavior change communication, livelihoods grants for the poorest, and maternal cash transfers in one arm. Without cash transfers, the model improved caregiver knowledge and savings but not children's diets or growth, while adding cash transfers improved diet quality and enhanced child development. The largest gains in child growth (a gain in 0.2 standard deviations in height-for-age) occurred among households receiving both livelihoods grants and maternal cash transfers. Coupling livelihoods support with sufficient financial resources appears critical for reducing child undernutrition.

Acknowledgments: Corresponding author email: J.Leight@cgiar.org. This study was funded by the government of the United States, 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 Programs on Food Frontiers and Security, and 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 and comments from Harold Alderman, Mara Mordini, Benita O'Colmain, Carlos Albuja Jijon, 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 (ESSSWA).

1 Introduction

Nearly 150 million children under five years of age suffer from chronic undernutrition in low- and middle-income countries (FAO et al., 2022), with children in the poorest households the most severely affected.[1] Early-life undernutrition has lasting consequences for human capital accumulation and is a central channel through which poverty persists across generations (Currie and Vogl, 2013; Hoddinott et al., 2008). To lift poor households out of poverty, multifaceted "graduation models" that provide a bundle of interventions have been widely deployed over the last 20 years, and they have had impressive effects in both the short and long term (Balboni et al., 2022; Banerjee et al., 2021; Barker et al., 2024). While these programs primarily target poverty traps arising from capital and livelihoods constraints, household economic gains alone may fail to close nutritional deficits accumulated during critical developmental windows (Victora et al., 2010; Grantham-McGregor et al., 2007). To date, most graduation models have neither explicitly incorporated child nutritional investments into their design nor systematically evaluated effects on nutritional outcomes (Attanasio, 2026), potentially undermining their long-term returns.[2]

To address this evidence gap, we conduct a cluster randomized controlled trial (RCT) evaluating whether a nutrition-sensitive graduation model embedded in Ethiopia's flagship safety net program – the Productive Safety Net Programme (PSNP) – can reduce child undernutrition and enhance child development while also improving household livelihoods and consumption. SPIR II was a graduation model program implemented in the Amhara and Oromia regions of Ethiopia that incorporated targeted nutrition-related interventions to address high rates of childhood stunting (close to 60%).[3] This centered on the formation of nurturing care groups (NCGs), peer caregivers' groups providing nutrition-related behavioral change counseling led by community-based trained volunteer agents, and the provision of monthly maternal cash transfers of $20 USD over a period of 24 months in infancy to a subset

[1]Brown et al. (2019) show that across 30 sub-Saharan African countries, around 40% of children in the poorest household wealth quintile are stunted — well above the WHO threshold for very high burden (≥30%; De Onis et al. (2019)).

[2]The existing papers in the graduation model literature generally do not measure or report any effects on child nutrition or anthropometric status, though Bedoya et al. (2019) shows positive effects in reducing the incidence of child diarrhea. Raza et al. (2018) 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. Another recent paper analyzed nutritional effects of an earlier iteration of the SPIR program (Alderman et al., 2025).

[3]SPIR 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. To be more precise about the cited statistic, 57% of the children 24–48 months of age in the control arm in this study were stunted at follow-up.

of households. All treated households also received a bundled set of livelihoods interventions (including village savings groups and training), and a subset of households (the poorest third at baseline) received one-time livelihoods grants.[4] SPIR differs from the canonical BRAC-style graduation model (Bandiera et al., 2017; Banerjee et al., 2015) in two important respects: the livelihoods support is notably less intensive —- with smaller one-time grants in place of larger asset transfers, and no equivalent of BRAC's hands-on coaching and training -— and the intervention is embedded within an existing national safety net program that conducts targeting and provides the administrative framework.

Our study includes a sample of 3,015 households in 234 kebeles (sub-districts), all of which are PSNP households that included either a pregnant woman or an infant under nine months of age at baseline. The kebeles were randomly assigned to three arms. The control arm received only the standard PSNP (T1). The second arm (T2) received SPIR graduation programming — village savings groups, livelihoods training and nurturing care groups — with a one-time livelihoods grant of $300 for the poorest third. The third arm (T3) received all T2 components plus monthly maternal cash transfers. We conducted large-scale surveys at baseline (2022) and follow-up surveys one year later (2023) and two and a half years later (2025), achieving a remarkably low attrition rate of 4% across the full panel. The cohort sampling strategy was designed to capture outcomes at key developmental milestones: in the short-run follow-up, children are six to 23 months of age, an optimal period for measuring age-appropriate complementary feeding practices; in the medium-run, children are 24 to 48 months, the age at which linear growth indicators such as height-for-age (HAZ) most clearly reflect cumulative early-life investments.[5]

We find that gains in child height-for-age (HAZ) are concentrated in the subsample of households exposed to the most intensive model: recipients of both the lump-sum livelihoods grant and the ongoing maternal cash transfer show increases in child HAZ of 0.23 SDs (a seven percentage point decline in stunting). These households also accumulate substantially more assets, with livestock holdings 43% higher than in the control arm. The associated increases in livestock income (82%) and consumption (13%) reflect this asset accumulation. The gains in child height are matched by higher early childhood development scores at both the short- and medium-term follow-up. For households receiving neither the livelihoods grant nor the cash transfers, effects are generally muted: these households are slightly more knowledgeable

[4]We 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).

[5]The 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).

about early childhood nutrition (a 3% shift in the key index) but show no change in child investment or growth, and no detectable change in income or consumption. Across all treatment arms, however, participation in village economic and social associations (VESAs) generates consistent gains in financial inclusion, primarily through increased savings.

For the subgroups that receive the maternal cash transfer alone or the livelihoods grant alone, we see evidence of trade-offs between production and consumption uses for the transfer — though these play out differently. Importantly, this comparison is only partially experimental, since the lump-sum livelihoods grant was targeted based on baseline asset levels rather than randomized within arms. Households receiving the lump-sum grant alone show short-term gains in consumption — reduced food insecurity, increased non-food consumption, and increased diet diversity for the index child — and accumulate livestock assets with modest associated income. Although the consumption and food security effects fade by the medium-run, HAZ at follow-up remains substantially higher, if imprecisely estimated, consistent with early-life nutritional investment in the index child.

Households receiving the stream of maternal cash transfers alone also direct some resources toward livestock accumulation, and the combination of livestock income and the transfer itself generates a small but significant consumption gain that grows over time. Unlike the grant-only arm, this subgroup shows a precisely estimated null on child HAZ in the medium-run, even as dietary diversity and early childhood development scores improve in the short-run. We interpret this pattern as follows. The lump-sum grant, though labeled for "livelihoods", coincided with a critical developmental window (close to birth or during the first year of life) and appears to have facilitated nutritional investments that translated into measurable gains in child linear growth, though its effects on broader household economic outcomes (other than livestock asset accumulation) have decayed. The maternal cash transfer stream, by contrast, generated household-level economic gains that have grown over time, but spreading the transfer across multiple objectives diluted its effectiveness in shifting child growth.

Our study contributes new evidence on how a nutrition-sensitive graduation model affects early childhood nutrition and household economic outcomes, and speaks to several strands of existing research. First, and most fundamentally, ours is one of the first studies to document significant effects of a graduation model program on child linear growth. While the graduation model literature has extensively documented impacts on assets, consumption, and food security (Banerjee et al., 2015; Bandiera et al., 2017; Balboni et al., 2022; Barker et al., 2024; Bossuroy et al., 2022; Brune et al., 2022; Bedoya et al., 2019; Leight et al., 2026, 2025), child anthropometric outcomes have been largely unmeasured or, when measured, null (Raza et al., 2018; Alderman et al., 2025). The HAZ effects we document among households receiving the most intensive intervention demonstrate that graduation models — when augmented with nutrition-sensitive components and timed to coincide with key phases in early childhood development — can affect the intergenerational human capital channel that existing graduation program evaluations have largely overlooked.

Second, 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 livelihoods program, a combination that allows us to study how these components interact in shaping child nutrition outcomes. The most closely related paper in terms of the combination of components is Bouguen and Dillon (2026), who evaluate a resilience program in Burkina Faso that combines seasonal cash transfers, livestock asset transfers, and a short-duration nutrition component.[6] They find HAZ improvements of comparable magnitude to ours (0.18 SD at two years), emerging only in their most intensive arm combining all three components — with no effects from cash alone or cash plus livestock. The parallel findings across two very different designs and contexts suggest that sufficiently intensive multi-component interventions can shift linear growth, while less intensive variants may not.

Within this literature, our findings also speak to the mechanism documented in Carneiro et al. (2021), who present evidence of a pathway from regular monthly transfers to women's livestock investment to increased consumption of protein-rich animal-source foods and improved child growth in Northern Nigeria. In our context, the maternal cash transfers generate large increases in children's consumption of dairy, eggs, and vitamin-A-rich foods — in line with the dietary channel emphasized by Carneiro et al. — while the livelihoods grant is the primary driver of livestock asset accumulation. However, significant gains in child linear growth emerge only when both components are combined, suggesting complementarity between the dietary intake and asset accumulation channels in shifting HAZ. In a synthesis of experimental findings, we show 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 HAZ estimates for the households receiving the most intensive intervention are among the largest effect sizes documented to date — consistent with the

[6] Unlike our design, their intervention is not structured as a graduation model: it lacks savings groups, livelihoods training, and the broader financial inclusion framework, and the cash transfers are seasonal rather than sustained. Our paper is also related to a previous paper assessing the effectiveness of nutrition interventions as implemented in SPIR I (Alderman et al., 2025). Our cohort design, which targets households with pregnant women or infants under nine months and measures outcomes at key developmental milestones, also differs from their broader sample of all children under five.

pattern that more intensive programs produce larger effects.

Third, we provide new evidence on a question that the cash transfer literature has not previously addressed: whether lump-sum transfers or ongoing transfer streams are more effective for child anthropometric outcomes. The cash transfer literature documents that stream transfers generally produce larger food security effects while lump-sum transfers facilitate asset accumulation (Haushofer and Shapiro, 2016; Crosta et al., 2024), but this distinction has not been examined with child nutrition as a primary outcome. The pattern across our subgroups — households receiving only the maternal cash stream show large dietary diversity gains but no HAZ improvement, while combining the stream with a livelihoods grant is required to shift child growth — suggests that sustained dietary improvements alone may not translate into measurable linear growth gains without the concurrent asset accumulation and income gains that a concentrated resource injection enables. This finding has direct implications for the design of graduation programs and cash transfer interventions targeting early childhood nutrition.

Finally, we conduct a cost-effectiveness analysis comparing SPIR II to other interventions targeting increased child linear growth, and a systematic cost-benefit analysis. Our cost-effectiveness analysis shows that the cost per 0.1 SD of height-for-age gained for the T3-G intervention ($987) is broadly comparable to the costs observed in other interventions (Ahmed et al., 2025; Bouguen and Dillon, 2026; Carneiro et al., 2021). Our cost-benefit analysis extends the standard graduation model framework in two ways. First, we harmonize assumptions across all major graduation evaluations (standardizing time horizons, decay rates, and discount rates), enabling direct comparison of benefit-cost ratios. Second, building on Carneiro et al. (2021), we estimate the benefits through two channels: increased household consumption, and an intergenerational pathway capturing the net present value of lifetime consumption gains attributable to improved HAZ. The subarms receiving maternal cash transfers generate benefit-cost ratios of 1.6 to 3.3 and internal rates of return of 16 to 26 percent, placing the full SPIR II intervention among the most cost-effective graduation programs evaluated to date. Incorporating the child HAZ pathway raises these benefit-cost ratios by roughly 10% — a modest shift that in part reflects heavy discounting of benefits under standard CBA assumptions, highlighting a tension between the economic case for investing in early childhood nutrition and the conventional treatment of intergenerational gains.

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 that are beneficiaries of the Productive Safety Net Programme (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 $734 in 2021 purchasing power parity (PPP) terms, corresponding to roughly 14% of the mean annualized household consumption at baseline.[7] Extensive evidence suggests 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, 2020).

The evaluation also unfolded during a period of considerable instability in rural Ethiopia, during which the PSNP has faced even greater demands. The Amhara region has experienced several periods of acute violence and steadily increasing insecurity; while Oromia has been less acutely affected, it has also been characterized by ongoing unrest, and periodic violence.[8]

In this volatile context, the SPIR II program was implemented by World Vision International (lead), CARE, and ORDA with funding provided by the U.S. government (2021–2026), and our study focuses on innovations in nutrition-related programming implemented as part of SPIR II. These innovations were designed around the first 1,000 days of life — from conception through age two — during which growth faltering accelerates and nutritional investments have the highest returns for long-term human capital (Victora et al., 2010). In Ethiopia, this faltering is closely linked to suboptimal complementary feeding practices from six months of age (Hirvonen et al., 2024; Golan et al., 2019).

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 (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 con-

[7]These estimates are from our baseline survey in 2022.
[8]See for example an overview here: https://www.cfr.org/global-conflict-tracker/conflict/conflict-ethiopia

vene 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.[9] Although experimental evidence on the effects of NCGs remains limited, 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). The expected pathway is that improved caregiver knowledge translates into better complementary feeding practices, particularly the introduction of diverse, nutrient-dense foods from six months of age.

The second element of the theory of change recognizes that improving caregiver knowledge may not be sufficient if households cannot afford to purchase adequate 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 52 $PPP) for a period of 24 months to relax possible financial constraints to child feeding. (The transfers correspond to around 12 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. Though unconditional, the transfers were weakly labeled as targeted for child feeding and health-related expenses.

A third element is through livelihoods support. All households included in NCGs are also exposed to core SPIR graduation programming. This includes the organization of 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 734 $PPP) livelihoods grants; targeting was based on an asset index constructed (by the research team) from data on asset ownership collected at baseline. These grants are expected to build productive assets — particularly livestock — that can generate income and, potentially, animal-source foods through own production.

In sum, the most intensive intervention package – received by a subset of households in one arm – combines the livelihoods interventions and the NCG-based BCC with two distinct types of transfers: the monthly maternal cash transfer (1,248 $PPP in total over 24 months), and the one-time livelihoods grant (734 $PPP). All households (in both treatment and control) also receive the regular PSNP cash or food payments; there is no experimental

[9] Quality 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.

variation in receipt of the core PSNP. This layered design allows us to assess the relative contribution of knowledge, purchasing power, and productive assets to child nutritional outcomes, and in particular whether improved caregiver knowledge alone is sufficient or whether complementary resource transfers are required.

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.[10] Kebeles were randomly assigned to one of three arms (see Table A1): T1 serves as a control arm, including households receiving PSNP only.[11] 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.[12]

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).[13] 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 follow-up survey conducted at one year measures feeding practices among children 6–23 months old, while the follow-up survey conducted at two and a half years

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

[11]Randomization 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).

[12]Note 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.

[13]No pregnancy tests were conducted.

measures height-for-age among children 24–48 months old, when linear growth reflects cumulative nutritional conditions.

We conducted detailed power calculations reported in the preregistered report.[14] For the nutrition-related outcomes, the evaluation is powered to detect a 0.22–unit change in height-for-age z-score (0.14 relative to SD) in a pairwise comparison across arms.

3 Data and econometric specification

The study design, hypotheses, and analysis plan were pre-specified in a detailed protocol. Deviations from the pre-registration report are discussed below and summarized in Table A3 in the appendix.

3.1 Data

The sample was interviewed at baseline in 2022, in a short-run follow-up conducted in 2023, and in a medium-run follow-up conducted in 2025. The short-run 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.[15] The second follow-up survey was conducted slightly earlier than planned (at two and a half years post-baseline) due to an anticipated widespread rollout of livelihoods grants by the Ethiopian government, described in more detail below in the section on program exposure. All three surveys included questionnaire modules administered to the primary female; the pregnant woman or mother of the target infant at baseline. Modules for her spouse were included at baseline and the second follow-up only. 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 (ESSSWA).

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 short- and medium-run follow-up); the number of food groups consumed by children 6–23 months (measured in the short-run follow-up); and height-for-age for children 24–48 months (measured in the medium-run

[14] We set the significance level at five percent and power at 80 percent, and allow for 10 percent attrition between baseline and follow-up surveys.

[15] Survey 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.

follow-up). These primary outcomes were selected based on the theory of change outlined above: knowledge captures the effect of the BCC channel, dietary diversity captures whether improved knowledge and resources translate into better feeding practices, and height-for-age captures whether these improvements reduce growth faltering.

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 measured using the 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 in both follow-up surveys), and assets, financial inclusion, and income (all measured at two and a half years only).

3.2 Econometric specification

Our primary specification can be written as follows

$$\begin{aligned} Y_{ikd,t=1} &= \beta_1 T^2_{kd} \times LG_{ikd} + \beta_2 T^2_{kd} \times NoLG_{ikd} + \beta_3 T^3_{kd} \times LG_{ikd} + \beta_4 T^3_{kd} \times NoLG_{ikd} \\ &+ (\gamma' Y_{ikd,t=0} + X^{'}_{ikd,t=0}\theta + \xi_{kd}) \times LG_{ikd} \\ &+ (\gamma' Y_{ikd,t=0} + X^{'}_{ikd,t=0}\theta + \xi_{kd}) \times NoLG_{ikd} + \beta_5 LG_{ikd} + \epsilon_{ikd}. \end{aligned} \quad (1)$$

Here, $Y_{ikd,t=1}$ captures the outcome of interest for household i in kebele (sub-district) k and woreda (district) d. The primary variables of interest are indicator variables for arms T2 and T3, interacted with binary variables for households that receive livelihoods grants (LG) or do not receive LG grants (NoLG); again, grants were targeted to the poorest 33 percent of the sampled households in each cluster. 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, as well as strata fixed effects ξ_{kd}, all in a fully saturated specification including interactions with LG_{ikd}; we also include the baseline value of the outcome $Y_{ikd,t=0}$, if available.[16]

Standard errors are clustered at the level of treatment (kebele). In addition to reporting standard p-values, we will also report sharpened q-values (Benjamini et al., 2006) following Anderson (2008), and this correction is implemented within the set of primary outcomes,

[16] 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.

and within the set of secondary outcomes.[17]

The ITT estimates β_1 and β_2 capture the effects of the T2 arm (livelihoods and BCC) for grant-eligible and non-eligible households, respectively, while β_3 and β_4 capture the corresponding effects of the T3 arm (livelihoods, BCC, and maternal cash transfers). Our pre-analysis plan specified simpler specifications as primary (pooling across grant and non-grant recipients, and pooling across the two treatment arms for livelihoods outcomes), and the pre-registered pooled specifications are reported in the Appendix B. The cluster-randomized design provides experimental identification for all comparisons across treatment arms, including comparisons within grant-eligibility strata. By contrast, comparisons between grant-eligible and non-eligible households within a treatment arm are not experimentally identified: grant eligibility is determined by baseline asset poverty, and differences in treatment effects across eligibility strata may reflect both the additional transfer and unobserved correlates of baseline poverty. We provide additional evidence on the interpretation of these comparisons in Section 4.3.1 and Appendix C.

4 Empirical findings

4.1 Baseline balance and loss to follow-up

Table 1 reports demographic characteristics of the sample households, as well as balance in observable characteristics across arms. The average household size at baseline 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 $3.00 a day in 2021 PPP) is 72 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.

We can also characterize the sample examining the characteristics of households that are eligible for grants and households that are not eligible. Again, eligibility was determined at baseline (by the research team), using a uniform criterion that is consistent across all clusters: households characterized by asset indices in the lowest third of the local distribution are eligible for transfers. Table B1 compares characteristics at baseline and shows that eligible households (unsurprisingly) have higher food insecurity, lower consumption, much lower assets, and reduced levels of savings. The difference in consumption, however, is not

[17] Exploratory outcomes are not included in the MHT-correction set.

particularly large (8%). While we cannot assess baseline gaps in anthropometric status since anthropometric outcomes were not measured, Table B2 then reports anthropometric variables among eligible and non-eligible households in the control arm only, in both follow-up surveys. Children in households that would have been grant-eligible do generally show lower height-for-age and lower CREDI scores, and the gap (between eligible and non-eligible control households) seems to be widening over time.

Across the survey rounds, household-level attrition was extremely low: 2% of households were lost to follow-up in the short-run, and 4% in the medium-run. However, a substantial share of the expected child sample did not materialize: 34% of pregnant women who were enrolled at baseline as pregnant did not have an infant present in the one year survey.[18] Women self-reported this was largely due to pregnancy loss, though there were anecdotal (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 2.5-year follow-up.[19] The implications for statistical power are, however, not large.[20] Below we further explore any potential bias induced by non-random attrition at both the household and child level.

4.2 Program exposure

Table B3 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

[18]The 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 at one year, and 2.5% of the infants surveyed at the one year follow-up could not be surveyed at two and a half years.

[19]We 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 2.5-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).

[20]For the primary variable of height-for-age, the minimum detectable effect increases by 0.04 standard deviations given the reduced cluster size. More details are provided in the report on the short-term follow-up survey (Gilligan et al., 2024).

is still high nearly two years later.[21] The level of participation is also consistently higher in the arm also receiving cash grants. 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.[22] Participation in VESAs, the core SPIR livelihoods-related intervention, is around 60–70% in the short-run follow-up, and rises to around 80% over time. (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 around 33% in both treatment arms as intended, with minimal evidence of contamination.

4.3 Main findings

We present the main findings in three parts. First, we examine nutrition outcomes — caregiver knowledge, child dietary practices, and anthropometric status — at both follow-up rounds. Second, we present livelihoods outcomes — consumption, food security, assets, and income — to understand the household-level economic effects of the intervention. Third, we synthesize the patterns across nutrition and livelihoods to assess the results against the theory of change and situate our findings within the broader cash-plus-BCC evidence base.

4.3.1 Diet and nutrition outcomes

Table 2 and 3 present the findings for the nutrition-related outcomes, first in the one-year follow-up when the sample children are aged 6–23 months, and next in the medium-run follow-up when the sample children are aged 24–48 months. The results are organized to track the causal chain outlined above: in the short-run, we focus on the primary outcomes of caregiver knowledge and child dietary practices during the critical complementary feeding window. Anthropometric outcomes at this stage are secondary, as growth faltering is still ongoing (Leroy et al., 2014). The primary outcome of height-for-age is measured at the two and a half year follow-up, when children are aged 24–48 months and linear growth reflects cumulative nutritional conditions. Throughout the main tables, primary outcomes are denoted with a dagger, and exploratory outcomes (that were not prespecified) are denoted with an *e*.

[21] Given 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 medium-run follow-up, but it seems that some respondents no longer remember or identify their previous participation when asked again.

[22] 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 routine BCC within the government health extension program or another intervention. Contamination in reported receipt of maternal cash transfers is minimal.

In Table 2, we can see that the nurturing care groups intervention did effectively increase IYCF knowledge in the short-term, though the effects are not large: the effect ranges between 2% and 5% and is consistent across both arms and across grants and non-grant recipients. There is also an increase in the number of food groups consumed by the index children that is observed among all treatment sub-arms other than the T2 - non-grant eligible (the only treatment households that received no form of additional cash support). Meal frequency (Column 3) increases significantly in T3 only, by approximately 0.25 meals per day (a 10% increase relative to the control mean), with no effect in T2. The increase in the probability of achieving a minimum acceptable diet is similarly observed in T3 only: 11 - 13 percentage points relative to a mean in the control arm of only nine percent, roughly a 120% relative gain. Detailed findings around IYCF consumption by food group reported in Table B7 in the Appendix indicate that this increased diet diversity was primarily driven by increases in consumption of eggs, dairy, and fruits and vegetables (including vitamin-rich fruits and vegetables); there is no shift in breastfeeding, either reported rates of ever-breastfeeding (which are close to 100%), or the probability of breastfeeding over the past day.[23]

Anthropometric outcomes when children are 6-23 months of age (columns 5 to 7) show a suggestive 0.18 SD increase in height-for-age among T3 grant recipients, but effects are generally imprecise at this stage, consistent with growth faltering still being in progress. There are, however, substantial positive effects of 0.2 standard deviations on the CREDI index of early childhood development (Column 8), observed in T3 only.

Shifting to the findings reported in Table 3 for the medium-run follow-up, the effects on knowledge seem to have been persistent over time but with some interesting variation across arms: among livelihoods grant recipients in both T2 and T3, the effects have grown, particularly among those who received livelihoods grants in T2 (where the knowledge effect more than doubled, though it remains only 4% relative to the control arm). Knowledge effects among non-grant recipients, by contrast, seem to have slightly decayed, particularly in T3.[24] In the medium-run follow-up, we also measured paternal IYCF knowledge, and it shows positive treatment effects of relatively consistent magnitude, relative to a slightly lower mean in the control arm: and again, the largest medium-run treatment effect (though imprecisely estimated) is observed among households in the T2 arm receiving livelihoods

[23]We also explored treatment effects on gestational age for the subsample of children enrolled in utero and found no evidence of any treatment effects, though gestational age is likely reported with considerable error.

[24]The 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.

grants.[25]

The positive effects on height-for-age are concentrated among the double-grant recipients (T3-G), where the effect on HAZ has slightly grown to 0.23 standard deviations as reported in Column (3), in conjunction with a seven percentage point decline in stunting (significant at the ten percent level using conventional clustered standard errors; q-value = 0.130). These households also show a 0.38 standard deviation increase in an index of parental stimulation behaviors (playing, reading, and singing with the child), as reported in Column (6), and a 0.24 standard deviation increase in child development (Column 7). The previous weak effects on anthropometric status in the T3 - no grant arm have, however, decayed to relatively precise zeros other than an increase in child stimulation, suggesting that the early gains observed at least suggestively in short-term follow-up did not persist.

By contrast, the T2 - grant arm now shows more positive effects: a 0.17 increase in height-for-age that is significant at the ten percent level when using the sharpened q-value, a slightly larger increase in the child stimulation index, and a 0.1 standard deviation increase in CREDI scores (albeit imprecisely estimated).[26] Effects in the T2 - G arm are in fact weakly negative, and the pattern of growth faltering generally observed among children in this age range in Ethiopia and other low-and-middle income countries (Golan et al., 2019; Victora et al., 2010) is also clearly evident: among children in the control arm, the stunting rate increases from 37% to 57% over the 18-month period between the follow-up surveys.

Supplementary findings again reported in Table B8 capture further effects on additional health behaviors and particularly use of health services. There are substantial declines in the probability of diarrhea (reported in the short- and medium-run) and open defecation, concentrated among households receiving maternal cash transfers. We also see increases in the probability of households reporting recent (past three month) engagement with health extension workers or health development army workers, visits to the health post, or attendance at food demonstrations or community conversations, generally relatively consistent across arms (though in some cases larger in T3).

One caveat in interpreting the larger effects on child growth among grant-eligible households is that these households are poorer at baseline, and the elasticity of child growth with respect to household investment may simply be larger. We exploit a feature of the study

[25] In the medium-run follow-up survey, 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.

[26] Note that one risk in the use of the CREDI in the medium-run follow-up survey is that some of the children have aged past 36 months, the recommended maximum age for CREDI use; however, we find that still substantial numbers of children have not achieved the age-appropriate developmental milestones that are the focus of the instrument, as seen in Figure B1. Accordingly, we see little risk of topcoding in the index.

design to probe this concern (see Appendix C). Grants were targeted to the poorest 33 percent of sampled households in each cluster, so the eligibility cutoff varies across clusters; a household in a wealthier cluster can be classified as eligible despite having more assets in absolute terms than a non-eligible household in a poorer cluster. This generates substantial overlap in the baseline asset distributions of the two groups (Figure C1). We use this overlap to test whether the differential treatment effects by eligibility status survive after controlling for the baseline asset level or after interacting the treatment variables with the baseline asset level. If the larger effects among eligible households reflected a mechanical relationship between poverty and growth responsiveness, the estimated treatment effects should narrow. They do not: adding a linear control for the baseline asset index and allowing treatment effects to vary with baseline assets leaves the results essentially unchanged (Table C1). This finding suggests that the larger effects among grant-eligible households reflect the additional resources provided by the grant rather than a mechanical relationship between baseline asset level and the responsiveness of child growth.

As previously noted, attrition at the household level was extremely low, while attrition at the child level is somewhat higher, mainly driven by the sample enrolled with pregnant women at baseline. We present a more detailed analysis of attrition in Appendix D, and show there is limited evidence of differential attrition across treatment arms at either the household or child level, and joint tests of the interaction between treatment status and baseline covariates fail to reject the null at both follow-up surveys (Appendix Tables D1 and D2). Lee bounds for the primary nutritional outcomes confirm that the positive effects on IYCF knowledge and dietary diversity can be bounded away from zero in both treatment arms (Appendix Table D3), and the positive effect on height-for-age for the double grant recipients can similarly be bounded away from zero.

4.3.2 Livelihoods outcomes

We now present the findings around livelihoods outcomes in Tables 4 through 7. For consumption and food security, positive short-term effects on consumption are observed primarily among maternal cash transfer recipients in the T3 arm: consumption increases by 6% relative to the control arm for non-livelihoods grants recipients, and 12% for double grant recipients (Table 4). For food insecurity, Food Insecurity Experience Scale (FIES) scores decline significantly among grant recipients at the one year follow-up, even in the T2-Grant group where there is no detectable shift in consumption. A more detailed decomposition of food insecurity indicators (Table B9) suggests that for T2-Grant households, this reflects reductions in moderate food insecurity experiences — eating fewer kinds of food and eating less than desired — rather than severe deprivation, consistent with asset accumulation reducing perceived food insecurity even before gains in consumption are observed. For T3-Grant recipients, improvements extend across the full severity spectrum, consistent with the provision of direct consumption support. T3 households also show large gains in household dietary diversity, while T2-Grant households show more modest improvements.

In the medium-run (Table 5), consumption effects for households receiving maternal cash transfers grow (now 9% and 13% relative to the control mean, with the larger effect among double grant recipients), while the T2 - G arm shows a significant increase in non-food consumption (10%) but no shift in overall consumption. The FIES-based improvements have largely faded across all subgroups.[27] However, household dietary diversity continues to improve: all four treatment subgroups now show significant gains in Household Dietary Diversity Score (HDDS) (Table B10), including households in the T2 arm where dietary diversity effects were absent or modest in the short run.

Table 6 presents findings around financial inclusion and assets. We can see there is an approximately 20 percentage point increase in the probability of any savings relative to a mean of 74% in the control arm (roughly consistent among all treatment households, though slightly larger among maternal cash transfer recipients), and an increase in the continuous amount of savings (50% in T2 and nearly 90% in T3). There is also a small increase in the probability of credit access (the magnitude varies between four and nine percentage points, relative to a control mean of 68%) and an increase in the reported outstanding credit balance as of the survey date.[28]

Moving on to assets, there are positive treatment effects on the index of total assets (Column 4) and tropical livestock units (Column 5) observed in all subarms receiving a livelihoods grant or a maternal cash transfer: the increases are largest among double grant recipients (around 0.3 standard deviations, normalized relative to the baseline control mean, and 0.2 tropical livestock units), followed by the livelihood grants only (T2 - G) recipients.[29] We also seek to unpack more detailed findings around livestock in a series of appendix tables; the consumption of animal-sourced foods is both theoretically and empirically linked to improved linear growth in low-income settings (Headey et al., 2018; Iannotti et al., 2014), and markets for these foods — particularly dairy — are thin in rural Ethiopia, creating highly localized markets and frequent non-separability between own production and consumption (Hoddinott et al., 2015). This renders understanding the treatment effects on livelihoods-

[27]The disjuncture between these findings for a subjective variable such as the FIES and the patterns observed for consumption suggests that households buffeted by multiple shocks, including conflict, perceive their food security deteriorating.

[28]This increase is primarily driven by increased utilization of formal credit, as reported in Column (1) of Table B11.

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

related outcomes particularly important.

We can see in Table B12 that there are significant positive treatment effects for ownership of poultry, goats, calves and heifers, and cows: the treatment effect for goats and cows (and to some extent heifers) is largest for livelihood grants recipients (who received a larger lump-sum), while the treatment effect for poultry is much larger for maternal grant recipients (who received more frequent but smaller payments, potentially rendering acquisition of poultry more feasible). The increased asset accumulation leads to increased production of both eggs and milk as observed in Table B13, with larger increases in egg production in T3 (though positive in both treatment arms), and larger increases in milk production among livelihood grants recipients, and particularly among the T3 - G households. In Table B14, we can observe there is also increased allocation of eggs to consumption particularly in T3: the share of egg producers primarily consuming rather than selling rises from 42% to around 60%. This would be consistent with the hypothesis that these households, receiving a regular cash transfer, have a reduced need for income generation from poultry production.[30] Given this shift in allocation, increases in household egg and milk consumption are notably larger in T3 in the one-year follow-up as shown in Table B15 – presumably also facilitated by the use of the maternal cash transfer to purchase these products – but the treatment effect on milk consumption among T2-G households has converged upward by medium-run.

Table 7 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: the increase in the probability of any net income from livestock ranges from four percentage points (one of the few significant treatment effects observed in the T2 - no grant arm) to 17 percentage points for the double grant recipients, relative to a mean of 77% in the control arm. The increase in the continuous net income is insignificant for households receiving no transfers, and ranges between 66% and 83% in the other treatment arms. There is no shift in the probability of reporting any net income from non-agricultural production, and imprecisely estimated declines in the continuous measure of net income; 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.[31] Columns (5) and (6) report effects on wage employment, for

[30]It is also possible that this is a behavioral shift induced by messages delivered via behavioral change counseling, though these messages were consistent across both treatment arms.

[31]The 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.

the extensive margin only (given low participation rates): there is evidence of substitution away from informal employment (often used as a coping mechanism, and consistent with households' new and increased access to higher-return livelihoods activities). There is also a significant increase observed in formal wage employment, only among livelihoods grants recipients in the T2 arm.[32]

Figure 1 also captures some key treatment effects (for consumption, food security, and assets) using longitudinal data. Here, it is clear that the general trend is weak declines in consumption and assets over time for households in the control arm. 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 reflect the challenging conditions in which the study took place, including intensifying conflict particularly in Amhara. 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: in the medium-run follow-up survey, 97% of households in the control arm remained PSNP beneficiaries, and the estimated treatment effect is small and statistically indistinguishable from zero (Column (4) in Table B11).

4.3.3 Synthesis

Taken together, these findings speak clearly to the program's theory of change. Improving caregiver knowledge alone is clearly insufficient, and at the alternate end of the spectrum of program intensity, households that received both livelihoods grants and maternal cash transfers showed unambiguously positive effects: more diversified child diets, enhanced child growth and development, substantial livestock accumulation, and persistent (and growing) consumption gains.

The arms receiving more partial grant support, however, show an intriguing pattern of findings, with the positive effects on height-for-age in the medium-run concentrated among households that received a livelihoods grant. The divergence in linear growth across subgroups cannot be fully explained by the measured IYCF indicators: T3-Grant and T3-No grant households show nearly identical effects on dietary diversity, meal frequency, and minimum acceptable diet – yet their effects on HAZ at two and a half years diverge sharply (0.23 SD versus essentially zero). One hypothesis is that the maternal cash was divided across competing objectives, including household consumption and accumulation of livestock, diluting the resources directed toward child feeding during the critical window to an extent

[32]SPIR did not target any shift in crop income and we can also confirm there is no significant shift in the reported value of the crop harvest, Column (2) in Table B11. We also report for completeness the treatment effects on gross livestock income in Column (3) of the same table.

not captured by the child dietary diversity measure.[33]

Our design allows us to test the relative effectiveness of both lump-sum livelihoods grants and ongoing consumption support transfers within a single trial, with the caveat that comparisons across grant-eligible and non-eligible households are not fully experimental. The within-arm pattern — in which grant-eligible households in T3 show significant HAZ gains while non-eligible households do not, and similarly in T2 — is consistent with the substantial gaps observed in household-level economic outcomes: e.g., within T3, households that receive livelihoods grants show roughly twice the consumption gain, significantly greater reductions in food insecurity, and larger livestock accumulation. These economic differences may have supported child nutrition through channels that a single 24-hour dietary recall cannot capture, including greater household economic stability, a more consistent and higher quantity of food consumed (including animal-source foods, driven by greater household production in a context of non-separability as shown above), and the day-to-day consistency of feeding over the full complementary feeding window. Initially more modest effects on parental nutrition knowledge and behavior are also growing over time among households that receive livelihoods grants, in contrast to stagnant or declining patterns among households receiving maternal cash transfers only, consistent with the hypothesis that positive income effects at the household level may be driving a more gradual, but sustained, shift toward new nutritional practices.

It is also important to note that a disconnect between diet and linear growth is not unique to our study: across the cash-plus-BCC literature (reviewed in Appendix E), improvements in child dietary outcomes frequently do not translate into gains in height-for-age (see Figures E1a and E1b). The studies that do document significant HAZ effects share a common feature: either substantially larger total transfer values (Ahmed et al., 2025), or households accumulating productive livestock assets through direct transfers (Bouguen and Dillon, 2026) or essentially, both: households using large transfers to accumulate livestock assets and thus generate predictable cash flows (Carneiro et al., 2021).

4.4 Cost-effectiveness and cost-benefit analysis

Given the substantial evidence of positive treatment effects generated by this intervention, we assess its cost-effectiveness and cost-benefit relative to other comparable papers.

[33] This is also consistent with evidence that PSNP transfers alone are generally insufficient to meet households' subsistence needs (Hoddinott et al., 2024), making it unsurprising that part of the maternal cash was redirected toward basic consumption and asset accumulation rather than child nutrition.

4.4.1 Costs

Abstracting away from the cost of the PSNP (received by all households), we estimate that the total cost of the SPIR program was 250 $PPP per household, without livelihoods grants or maternal cash transfers. The livelihood grant is valued at 734 $PPP per household, and maternal cash transfers at 1,248 $PPP. Therefore, adding livelihood grants to T2 increases the total cost to 983 $PPP per household. The total cost of implementing T3 (SPIR + maternal cash transfers) is 1,498 $PPP per household, and adding the livelihood grant to this yields a total cost of 2,231 $PPP per household.

4.4.2 Cost-effectiveness on child linear growth

Because child height-for-age is the outcome of primary policy interest for this intervention, we begin by expressing cost-effectiveness as the cost per 0.1 standard deviation gain in HAZ per beneficiary child. We restrict the cost-effectiveness analysis to the T3-Grant group, the only arm with a statistically significant treatment effect on HAZ (0.226 SD, $p < 0.05$; see Table 3, column 3).

Panel A of Table 8 summarizes the SPIR intervention costs mentioned above: the maternal cash transfer alone accounts for 56 percent of marginal cost, more than the livelihood grant (33 percent) and the base SPIR package (11 percent). Dividing the marginal cost per household by the treatment effect on HAZ, and scaling to a 0.1 SD increment, we obtain an estimated cost-effectiveness ratio of 987 $PPP per 0.1 SD HAZ gain per child for SPIR T3-Grant (Panel B of Table 8).[34]

To place this estimate in context, we compare it to three other cash-plus-BCC-style interventions that report experimental effects on HAZ: Ahmed et al. (2025) in Bangladesh, Carneiro et al. (2021) in Nigeria, and Bouguen and Dillon (2026) in Burkina Faso. A key complication is that of the three studies, full costing data are only available for Ahmed et al. (2025), as reported in Ahmed et al. (2016).[35] We report two cost-effectiveness estimates for Carneiro et al., using both the assumption applied in their original paper about (low) rates of cost for BCC and administrative costs, and a higher estimate corresponding to what is

[34] The index child followed by the SPIR surveys is the only beneficiary child explicitly measured at the household level, and we therefore express cost-effectiveness per child assuming one beneficiary child per household. If positive spillovers to other children in the household are present, the true per-child cost would be lower.

[35] Carneiro et al. (2021) assume administrative and BCC delivery costs to equal 10 percent each of the value of the cash transfer. Bouguen and Dillon (2026) report values for cash transfers and the different livelihood package components, but no BCC or administrative costs. The BCC and administrative cost shares reported by Ahmed et al. (2016) are roughly three times Carneiro et al.'s assumed value: cash delivery at 29.0 percent of cash transfer value and BCC delivery at 27.9 percent of cash transfer value.

observed in Ahmed et al.[36] Appendix F provides additional details on the construction of these estimates.

The cost-effectiveness estimate for the T3-G subarm in this trial ($987 per 0.1 SD HAZ gain per child) sits in the middle of this distribution, between Bouguen and Dillon ($606; col. 4 of Table 8) and Carneiro et al. under their published cost assumption ($701; col. 3a) on one side, and Ahmed et al. (2025) ($1,201) on the other (col. 2). When we replace the BCC and administrative cost shares assumptions in Carneiro et al with Ahmed et al.'s observed shares, Carneiro et al.'s cost-effectiveness ratio rises from $701 to $917 (col. 3b), close to SPIR T3-Grant. Two important caveats are worth emphasizing. First, all four studies have wide upper bounds on their cost-effectiveness confidence intervals, reflecting that the underlying HAZ treatment effects are relatively imprecise. Second, the BCC delivery and administration costs for Carneiro et al. and Bouguen and Dillon are based on assumptions rather than actual accounting records, and should therefore interpreted with caution.

4.4.3 Cost-benefit analysis

We complement the cost-effectiveness analysis with a full cost-benefit analysis that monetizes two benefit pathways: an immediate increase in household consumption as the primary measure of relatively short-term program benefits, and a lifetime consumption gain in the next generation from improved child HAZ. The cost-effectiveness analysis above isolates the second of these pathways; the cost-benefit analysis brings them together in a single benefit-cost ratio and internal rate of return. Panel A of Table 9 summarizes program costs, already reported above.

We innovate in the estimation of benefits within a graduation model intervention by assessing two pathways: an immediate increase in household consumption as the primary measure of relatively short-term program benefits, and increased consumption in the next generation (among currently observed children), due to improved child nutritional status. For the first pathway, the consumption ITT estimates are annualized and expressed in non-per capita terms; year one and year three benefits are based on the ITTs estimated in the short- and medium-run, respectively, while year two benefits are linearly interpolated.

We consider four scenarios of effect duration (a 10-year and a 20-year time horizon, each without and with 10 percent annual decay), assuming that in the absence of decay benefits following the conclusion of the trial remain at the level observed at two and a half years.[37] As is standard in this literature, we follow the World Bank guidelines and apply a five

[36]We do not report this second scenario for Bouguen and Dillon because their BCC component is minimal (two training sessions rather than the weekly group sessions and home visits of Ahmed et al.'s design), so applying the 27.9 percent BCC share to their cash transfer would not be a meaningful upper bound.

[37]Detailed household level benefit calculations are presented in Table F1 in the appendix.

percent annual discount rate. For the second pathway, we focus on the HAZ ITT and build on Carneiro et al. (2021), applying the long-run HAZ–consumption estimates provided by Hoddinott et al. (2013b): these estimates suggest a 1 SD increase in HAZ at age two results in a 21 percent increase in per capita consumption in adulthood. We then sum these gains over the adult life cycle, and discount back to year zero at the same discount rate. Unlike the household consumption pathway, the HAZ benefit does not vary across scenarios, as it is not subject to the time-horizon or decay assumptions applied to consumption benefits. More methodological details on these calculations are provided in Section F in the Appendix.

Panel B of Table 9 provides the estimated benefits under the two pathways. Starting with the consumption pathway, the estimated net benefit in T3-grant group ranges between 3,600 and 7,400 $PPP, depending on assumptions about benefit duration and the degree of decay. The corresponding net benefit range in the T3-non-grant group is 2,300 to 4,900 $PPP, and the consumption ITTs in the T2 arms are considerably smaller and statistically insignificant. For the child HAZ pathway, the NPV value of the gains in lifetime consumption are smaller, 389 $PPP among double grant recipients and around $289 among the T2 - grants recipients. This reflects both the relatively modest effect of stunting on estimated future consumption, and the substantial discounting of any inter-generational effects.[38]

Considering both pathways, the estimated benefit-cost ratios (Panel C) and internal rates of return (Panel D) confirm that the T3 arms generate substantial net benefits. For T3-G, the consumption-only BCR ranges from 1.61 to 3.3, with IRRs between 18 and 26 percent. Adding the child HAZ pathway raises these to 1.8–3.5 for the BCR and 24–31 percent for the IRR, reflecting the positive contribution of improved nutritional status. The T3-NG estimates are similar, and these results are robust to pessimistic assumptions about impact persistence: the benefits would need to decay at 35% to 40% annually before the program fails to break even at a long time horizon. The results for the T2 arms are less favorable, however, reflecting the initially low, or null, consumption benefits.[39]

To place these estimates in context, we also replicated and harmonized the cost-benefit

[38] The relevance of the HAZ pathway depends heavily on the discounting applied to the future consumption gains that do not materialize for roughly 16 years, as the standard cost-benefit framework mechanically devalues benefits that accrue to the next generation. A lower discount rate, reflecting a more patient social planner's perspective, would substantially increase the relative weight of the HAZ pathway. The appropriate social discount rate for health interventions remains debated (Haacker et al., 2020), but any rate below 5 percent would increase the relative contribution of the HAZ pathway. As an illustration, reducing the discount rate to 3 percent would increase the NPV of the HAZ pathway by roughly 85 percent.

[39] The consumption ITTs for T2-NG and T2-G are not statistically significant at conventional levels, nor are the HAZ effects for these arms. The point estimates for T2-NG suggest a positive return–with consumption-only BCRs of 1.5–3.6–but these should be interpreted with caution given the lack of statistical significance. T2-G, in contrast, yields BCRs below unity across all scenarios, indicating that the livelihood grant alone, without the maternal cash transfers, does not generate sufficient consumption gains to justify its cost.

calculations reported in the graduation model literature to match our four scenarios. To ensure comparability, Table 10 reports consumption-only BCRs and IRRs, as none of the earlier evaluations incorporate a child nutrition pathway. The estimated effects for both T3 subarms show this intervention to be among the most cost-effective in the literature: the BCRs (roughly 1.6–3.3 depending on the scenario) and IRRs (16–26%) are comparable to India in Banerjee et al. (2015) and Bangladesh in Bandiera et al. (2017), and not far from the BCRs (1.7–3.8) and IRRs (21–30%) reported in an evaluation in Afghanistan (Bedoya et al., 2019). The BCRs and IRRs reported by Bossuroy et al. (2022) in Niger are substantially higher than those of all other programs, driven largely by the very large effects achieved at very low cost. However, again it is important to note that despite the explicit goal of this intervention (and this trial) of identifying strategies to address challenges in undernutrition, ultimately the achieved gains in child nutritional status have relatively little effect on the estimated benefit-cost calculus here – though other methods of assessing the benefits of enhanced linear growth might lead to different conclusions.

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 package 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 and, in one arm, supplementary maternal cash transfers. Our trial tracks a large sample of PSNP households over nearly three years.

We find that across nutrition outcomes, the model without maternal cash transfers increased maternal knowledge of IYCF practices but generated limited dietary improvements. In contrast, adding the maternal cash transfers produced sizable gains in diet quality, driven by increased consumption of eggs, dairy, and vitamin-A-rich foods, as well as improvements in early child development. Despite these advances in child development and diet diversity, however, effects on linear growth are concentrated among households receiving the most intensive intervention (livelihoods grants and maternal cash transfers). Grant-eligible households in the arm without the maternal cash transfer also show a positive HAZ effect of 0.17 SD, suggesting that the livelihoods grant contributes to child growth through channels beyond measured dietary improvements — possibly through the household economic stability and sustained animal-source food production that livestock accumulation provides. By contrast, households receiving the maternal cash transfer without the livelihoods grant show no improvement in linear growth despite large dietary diversity gains, a disconnect that echoes a broader pattern in the cash-plus-BCC literature in which dietary improvements frequently fail to translate into height-for-age gains.

Turning to livelihoods outcomes, households in both arms increased their participation in savings groups and expanded their access to savings and credit. There is also evidence of improvements in per-capita consumption and a persistent increase in livestock holdings and livestock income, with consumption effects concentrated among maternal cash transfer recipients and growing over time. 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, with larger effects observed only when transfers are larger (Hirvonen et al., 2025; Leight et al., 2026). In a context where control households experienced declining consumption and asset levels over time, the positive treatment effects can be understood in part as buffering households against deteriorating conditions.

A harmonized cost-benefit analysis situates these findings within the broader graduation model literature. In particular, the model including both livelihoods grants and maternal cash transfers generates benefit-cost ratios of 1.6 to 3.3 and internal rates of return of 16 to 26 percent under a range of assumptions about impact persistence, placing this intervention among the most cost-effective graduation programs evaluated to date. Incorporating the intergenerational returns from improved child nutritional status — using the estimated relationship between early-childhood height-for-age and adult consumption — further increases these returns. 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 combining behavior change communication and livelihoods support with substantial cash support beyond what is provided by government-led social protection programs.

References

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

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.

Ahmed, A., Hoddinott, J. F., Roy, S., Sraboni, E., Quabili, W., and Margolies, A. (2016). Which kinds of social safety net transfers work best for the ultra poor in bangladesh? operation and impacts of the transfer modality research initiative.

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 Tambet, H. (2025). Scalable nutrition interventions in a graduation model program: Experimental evidence from ethiopia. *Economic Development and Cultural Change*, 73(3):1559–1604.

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.

Attanasio, O. (2026). The first 1000 days and beyond: The process of child development. *Journal of Economic Literature*, forthcoming.

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(4):471–486.

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., 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.

Brown, C., Ravallion, M., and van de Walle, D. (2019). Most of africa's nutritionally deprived women and children are not found in poor households. *The Review of Economics and Statistics*, 101(4):631–644.

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.

Carletto, G., Tiberti, M., and Zezza, A. (2022). Measure for measure: comparing survey based estimates of income and consumption for rural households. *The World Bank Research Observer*, 37(1):1–38.

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.

Cordero, M. E., D'Acuña, E., Benveniste, S., Prado, R., Nuñez, J. A., and Colombo, M. (1993). Dendritic development in neocortex of infants with early postnatal life undernutrition. *Pediatric Neurology*, 9(6):457–464.

Crosta, T., Karlan, D., Ong, F., Rüschenpöhler, J., and Udry, C. R. (2024). Unconditional cash transfers: A bayesian meta-analysis of randomized evaluations in low and middle income countries. Technical report, National Bureau of Economic Research.

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., Wetzel, 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.

De Onis, M., Borghi, E., Arimond, M., Webb, P., Croft, T., Saha, K., De-Regil, L. M., Thuita, F., Heidkamp, R., Krasevec, J., et al. (2019). Prevalence thresholds for wasting, overweight and stunting in children under 5 years. *Public health nutrition*, 22(1):175–179.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. World Bank Publications.

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.

Gilligan, D. O., Hirvonen, K., Leight, J., and Tesfaye, H. (2024). *Implementer-Led Evaluation and Learning (IMPEL) evaluation of SPIR II RFSA-Midline survey report*. Intl Food Policy Res Inst.

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.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., and Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *The lancet*, 369(9555):60–70.

Haacker, M., Hallett, T. B., and Atun, R. (2020). On discount rates for economic evaluations in global health. *Health Policy and Planning*, 35(1):107–114.

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

Headey, D., Hirvonen, K., and Hoddinott, J. (2018). Animal sourced foods and child stunting. *American Journal of Agricultural Economics*, 100(5):1302–1319.

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., 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., Alderman, H., Behrman, J. R., Haddad, L., and Horton, S. (2013a). The economic rationale for investing in stunting reduction. *Maternal & child nutrition*, 9:69–82.

Hoddinott, J., Behrman, J. R., Maluccio, J. A., Melgar, P., Quisumbing, A. R., Ramirez-Zea, M., Stein, A. D., Yount, K. M., and Martorell, R. (2013b). Adult consequences of growth failure in early childhood. *The American journal of clinical nutrition*, 98(5):1170–1178.

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.

Hoddinott, J., Headey, D., and Dereje, M. (2015). Cows, missing milk markets, and nutrition in rural ethiopia. *The Journal of Development Studies*, 51(8):958–975.

Hoddinott, J., Maluccio, J. A., Behrman, J. R., Flores, R., and Martorell, R. (2008). Effect of a nutrition intervention during early childhood on economic productivity in guatemalan adults. *The lancet*, 371(9610):411–416.

Iannotti, L. L., Lutter, C. K., Bunn, D. A., and Stewart, C. P. (2014). Eggs: the uncracked potential for improving maternal and young child nutrition among the world's poor. *Nutrition reviews*, 72(6):355–368.

Kar, B. R., Rao, S. L., and Chandramouli, B. (2008). Cognitive development in children with chronic protein energy malnutrition. *Behavioral and Brain Functions*, 4(1):31.

King, W., Miguel, E., and Walker, M. W. (2026). Why does height pay? evidence from the kenya life panel survey. Technical report, National Bureau of Economic Research.

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.

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.

Lundborg, P., Nystedt, P., and Rooth, D.-O. (2014). Height and earnings: The role of cognitive and noncognitive skills. *Journal of Human Resources*, 49(1):141–166.

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.

Persico, N., Postlewaite, A., and Silverman, D. (2004). The effect of adolescent experience on labor market outcomes: The case of height. *Journal of political Economy*, 112(5):1019–1053.

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.

UN (2024). *World Population Prospects 2024*. Population Division of the Department of Economic and Social Affairs of the United Nations (UN) Secretariat.

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.

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., and Sachdev, H. S. (2008). Maternal and child undernutrition: consequences for adult health and human capital. *The lancet*, 371(9609):340–357.

Vogl, T. S. (2014). Height, skills, and labor market outcomes in mexico. *Journal of Development Economics*, 107:84–96.

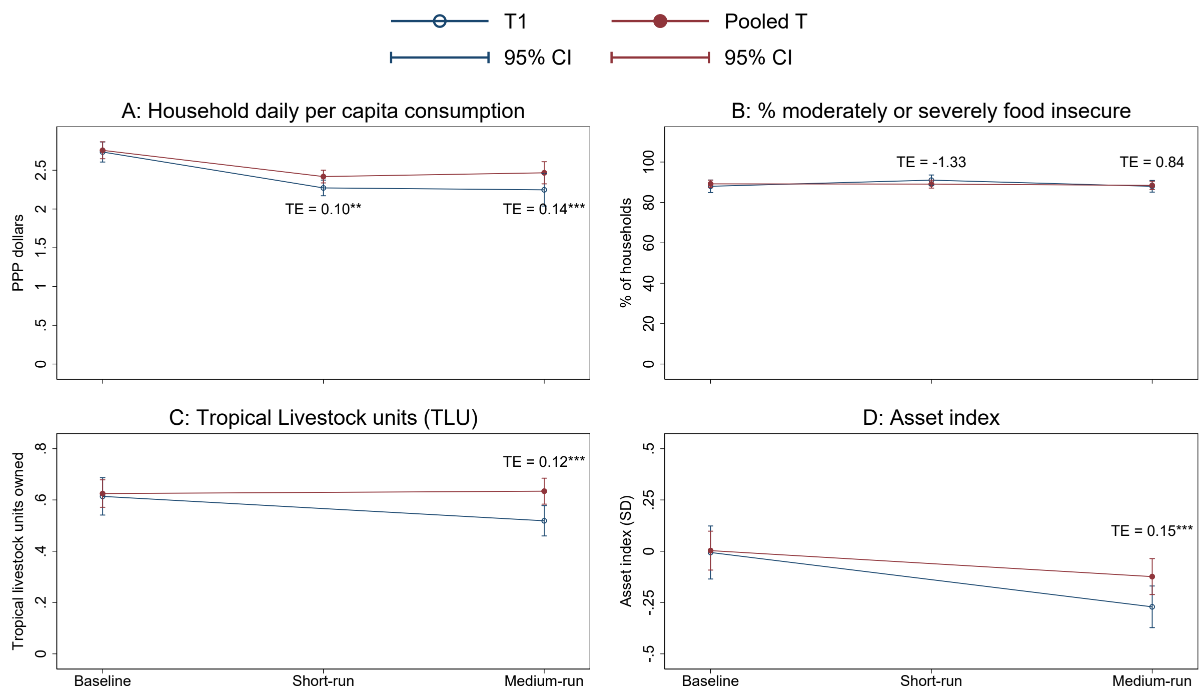
Weaver, J., Sukhtankar, S., Niehaus, P., and Muralidharan, K. (2024). Cash transfers for child development: Experimental evidence from india. Technical report, National Bureau of Economic Research.

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 one year.

Table 1: Balance

Variable	(1) T1 Mean/(SE)	(2) T2 Mean/(SE)	(3) T3 Mean/(SE)	(1)-(2) Mean difference	(1)-(3) Pairwise t-test Mean difference	(2)-(3) Mean difference
Currently pregnant	0.527	0.536	0.533	-0.010	-0.007	0.003
	(0.005)	(0.005)	(0.007)			
Child 0-3 months	0.215	0.180	0.192	0.034**	0.023	-0.011
	(0.011)	(0.010)	(0.014)			
Child 4-6 months	0.143	0.145	0.140	-0.002	0.003	0.005
	(0.011)	(0.011)	(0.010)			
Child 7-9 months	0.113	0.136	0.133	-0.023*	-0.020	0.003
	(0.010)	(0.012)	(0.010)			
Household size	5.407	5.375	5.401	0.032	0.006	-0.026
	(0.118)	(0.116)	(0.122)			
Primary female has some education	0.444	0.423	0.428	0.021	0.016	-0.005
	(0.021)	(0.019)	(0.021)			
Maternal IYCF knowledge (0-11)	8.454	8.522	8.327	-0.068	0.127	0.195
	(0.154)	(0.157)	(0.167)			
FIES raw score	6.151	6.219	6.034	-0.068	0.117	0.184*
	(0.112)	(0.095)	(0.110)			
Daily per capita consumption (PPP)	2.736	2.735	2.780	0.000	-0.044	-0.044
	(0.065)	(0.071)	(0.085)			
Tropical Livestock Units (TLU)	0.614	0.606	0.644	0.008	-0.030**	-0.038
	(0.037)	(0.038)	(0.039)			
Durable asset index (PCA)	0.001	-0.047	0.047	0.048	-0.046	-0.094*
	(0.043)	(0.044)	(0.070)			
Has formal savings	0.520	0.516	0.541	0.004	-0.022	-0.025
	(0.030)	(0.029)	(0.030)			
Eligible for livelihood grant	0.303	0.307	0.309	-0.004	-0.006*	-0.002
	(0.004)	(0.003)	(0.003)			
F-test of joint significance (F-stat)				1.017	0.987	0.848
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: Nutrition: One-year follow-up (Children 6–23 months)

	(1) IYCF knowledge†	(2) Diet diversity†	(3) Meal frequency^e	(4) MAD	(5) HAZ	(6) Stunted	(7) WHZ	(8) CREDI^e
T2 – No grant	0.311***	0.043	0.012	-0.007	-0.038	0.010	-0.114*	0.032
	(0.102)	(0.072)	(0.055)	(0.020)	(0.076)	(0.028)	(0.064)	(0.053)
q-value	[0.007]	[0.226]		[0.482]	[0.433]	[0.482]	[0.110]	
T2 – Grant	0.167	0.232**	0.111	0.039	0.081	0.011	-0.008	0.020
	(0.140)	(0.095)	(0.089)	(0.025)	(0.103)	(0.045)	(0.090)	(0.106)
q-value	[0.091]	[0.021]		[0.155]	[0.366]	[0.496]	[0.562]	
T3 – No grant	0.457***	0.579***	0.263***	0.106***	0.072	-0.046*	0.038	0.201***
	(0.104)	(0.074)	(0.051)	(0.021)	(0.072)	(0.027)	(0.063)	(0.055)
q-value	[0.001]	[0.001]		[0.001]	[0.309]	[0.130]	[0.399]	
T3 – Grant	0.334**	0.574***	0.230***	0.126***	0.183*	-0.015	0.099	0.187**
	(0.134)	(0.103)	(0.085)	(0.031)	(0.105)	(0.043)	(0.087)	(0.092)
q-value	[0.020]	[0.001]		[0.001]	[0.114]	[0.482]	[0.269]	
ANCOVA	Yes	No	No	No	No	No	No	No
p-value: T2 = T3	0.040	0.000	0.000	0.000	0.070	0.038	0.014	0.000
p-value: T2-G = T2-NG	0.380	0.128	0.326	0.139	0.347	0.998	0.309	0.921
p-value: T3-G = T3-NG	0.427	0.968	0.740	0.568	0.364	0.530	0.529	0.897
Control mean	9.51	3.25	2.51	0.09	-1.61	0.37	-0.53	-0.00
Observations	2231	2233	2221	2221	2226	2226	2226	2212

Notes: † = Primary outcome. ^e = Exploratory outcome (not pre-specified). Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). IYCF = infant and young child feeding; MAD = minimum acceptable diet; HAZ = height-for-age Z-score; Stunted = HAZ < -2; WHZ = weight-for-height Z-score; CREDI = Caregiver Reported Early Development Instrument. See Table A4 in the Appendix for full definitions of the outcome variables. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, education indicators, child sex, and child age in months. Standard errors clustered at the kebele level in parentheses. Sharpened q-values (Anderson 2008) in brackets; not reported for exploratory outcomes. *** p<0.01, ** p<0.05, * p<0.10.

Table 3: Nutrition: 2.5-year follow-up (Children 24–48 months)

	(1) IYCF knowledge†	(2) Paternal IYCF knowledge^e	(3) HAZ†	(4) Stunted	(5) WHZ	(6) Stimulation^e	(7) CREDI
T2 – No grant	0.233**	0.239*	-0.111	0.040	-0.098	-0.059	0.004
	(0.102)	(0.126)	(0.077)	(0.028)	(0.061)	(0.081)	(0.052)
q-value	[0.025]		[0.062]	[0.184]	[0.138]		
T2 – Grant	0.436***	0.429*	0.168	-0.059	0.022	0.208*	0.098
	(0.155)	(0.233)	(0.114)	(0.045)	(0.085)	(0.110)	(0.087)
q-value	[0.011]		[0.062]	[0.216]	[0.496]		
T3 – No grant	0.203**	0.244**	0.012	-0.010	0.006	0.159**	0.074
	(0.088)	(0.121)	(0.070)	(0.029)	(0.057)	(0.077)	(0.051)
q-value	[0.025]		[0.229]	[0.482]	[0.559]		
T3 – Grant	0.433***	0.232	0.226**	-0.070*	0.051	0.382***	0.243***
	(0.137)	(0.224)	(0.102)	(0.042)	(0.087)	(0.113)	(0.086)
q-value	[0.006]		[0.026]	[0.130]	[0.400]		
ANCOVA	Yes	Yes	No	No	No	No	No
p-value: T2 = T3	0.784	0.671	0.082	0.096	0.131	0.004	0.031
p-value: T2-G = T2-NG	0.307	0.449	0.035	0.062	0.205	0.053	0.373
p-value: T3-G = T3-NG	0.174	0.958	0.064	0.248	0.636	0.114	0.120
Control mean	9.16	8.63	-2.21	0.57	-0.43	1.82	0.01
Observations	2163	1316	2203	2203	2203	2210	2195

Notes: † = Primary outcome. ^{e} = Exploratory outcome (not pre-specified). Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). IYCF = infant and young child feeding; HAZ = height-for-age Z-score; Stunted = HAZ < -2; WHZ = weight-for-height Z-score; CREDI = Caregiver Reported Early Development Instrument. See Table A4 in the Appendix for full definitions of the outcome variables. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, education indicators, child sex, and child age in months. Standard errors clustered at the kebele level in parentheses. Sharpened q-values (Anderson 2008) in brackets; not reported for exploratory outcomes. *** p<0.01, ** p<0.05, * p<0.10.

Table 4: Consumption and food insecurity: One-year follow-up

	(1) Per capita cons.	(2) Food cons.[e]	(3) Nonfood cons.[e]	(4) Food insecurity binary	(5) Food insecurity cont.
T2 – No grant	-0.012	-0.005	-0.006	0.020	0.089
	(0.051)	(0.049)	(0.006)	(0.015)	(0.083)
q-value	[0.496]			[0.191]	[0.291]
T2 – Grant	0.056	0.057	0.004	-0.050***	-0.220**
	(0.076)	(0.073)	(0.008)	(0.019)	(0.108)
q-value	[0.380]			[0.019]	[0.063]
T3 – No grant	0.140**	0.129**	0.012**	-0.006	-0.120
	(0.055)	(0.053)	(0.006)	(0.016)	(0.085)
q-value	[0.022]			[0.482]	[0.185]
T3 – Grant	0.276***	0.253***	0.029***	-0.067***	-0.359***
	(0.079)	(0.076)	(0.009)	(0.019)	(0.106)
q-value	[0.002]			[0.002]	[0.003]
ANCOVA	Yes	Yes	Yes	Yes	Yes
p-value: T2 = T3	0.001	0.002	0.000	0.039	0.007
p-value: T2-G = T2-NG	0.416	0.441	0.292	0.004	0.011
p-value: T3-G = T3-NG	0.105	0.121	0.095	0.014	0.039
Control mean	2.27	2.05	0.22	0.91	5.91
Observations	2954	2954	2954	2954	2954

Notes: [e] = Exploratory outcome (not pre-specified). Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). See Table A4 in the Appendix for full definitions of the outcome variables. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, and education indicators. Standard errors clustered at the kebele level in parentheses. Sharpened q-values (Anderson 2008) in brackets; not reported for exploratory outcomes. *** p<0.01, ** p<0.05, * p<0.10.

Table 5: Consumption and food insecurity: 2.5-year follow-up

	(1) Per capita cons.	(2) Food cons.[e]	(3) Nonfood cons.[e]	(4) Food insecurity binary	(5) Food insecurity cont.
T2 – No grant	0.043	0.054	-0.005	0.025	0.136
	(0.068)	(0.063)	(0.011)	(0.016)	(0.093)
q-value	[0.395]			[0.150]	[0.173]
T2 – Grant	0.030	0.003	0.037**	-0.002	-0.111
	(0.109)	(0.097)	(0.019)	(0.020)	(0.136)
q-value	[0.496]			[0.559]	[0.363]
T3 – No grant	0.206***	0.177***	0.033***	-0.005	-0.057
	(0.064)	(0.059)	(0.012)	(0.018)	(0.099)
q-value	[0.004]			[0.496]	[0.400]
T3 – Grant	0.304***	0.251**	0.066***	0.013	-0.275**
	(0.116)	(0.103)	(0.020)	(0.021)	(0.131)
q-value	[0.019]			[0.395]	[0.056]
ANCOVA	Yes	Yes	Yes	Yes	Yes
p-value: T2 = T3	0.004	0.011	0.001	0.326	0.036
p-value: T2-G = T2-NG	0.908	0.632	0.034	0.297	0.123
p-value: T3-G = T3-NG	0.391	0.457	0.167	0.477	0.146
Control mean	2.25	1.90	0.35	0.88	5.96
Observations	2902	2902	2902	2902	2902

Notes: [e] = Exploratory outcome (not pre-specified). Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). See Table A4 in the Appendix for full definitions of the outcome variables. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, and education indicators. Standard errors clustered at the kebele level in parentheses. Sharpened q-values (Anderson 2008) in brackets; not reported for exploratory outcomes. *** p<0.01, ** p<0.05, * p<0.10.

Table 6: Financial inclusion and assets: 2.5-year follow-up

	(1) Any savings	(2) Amount saved	(3) Any credit	(4) Amount credit	(5) Total asset index	(6) TLU
T2 − No grant	0.198***	19.691***	0.094***	97.756***	0.037	0.025
	(0.024)	(3.190)	(0.025)	(22.216)	(0.049)	(0.031)
q-value	[0.001]	[0.001]	[0.001]	[0.001]	[0.380]	[0.366]
T2 − Grant	0.227***	14.549***	0.037	48.785**	0.241***	0.191***
	(0.027)	(4.451)	(0.036)	(19.783)	(0.055)	(0.037)
q-value	[0.001]	[0.003]	[0.299]	[0.026]	[0.001]	[0.001]
T3 − No grant	0.223***	27.568***	0.074***	75.799***	0.166***	0.131***
	(0.023)	(3.501)	(0.023)	(19.930)	(0.051)	(0.035)
q-value	[0.001]	[0.001]	[0.004]	[0.001]	[0.004]	[0.001]
T3 − Grant	0.256***	28.807***	0.063*	44.504**	0.295***	0.222***
	(0.027)	(5.440)	(0.033)	(19.286)	(0.057)	(0.044)
q-value	[0.001]	[0.001]	[0.090]	[0.035]	[0.001]	[0.001]
ANCOVA	Yes	Yes	No	No	Yes	Yes
p-value: T2 = T3	0.055	0.006	0.754	0.340	0.012	0.002
p-value: T2-G = T2-NG	0.328	0.314	0.171	0.060	0.002	0.000
p-value: T3-G = T3-NG	0.233	0.816	0.771	0.200	0.054	0.075
Control mean	0.74	31.23	0.68	173.30	-0.27	0.52
Observations	2902	2902	2902	2902	2902	2902

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). TLU = Tropical Livestock Units. See Table A4 in the Appendix for full definitions of the outcome variables. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, and education indicators. Standard errors clustered at the kebele level in parentheses. Sharpened q-values (Anderson 2008) in brackets; not reported for exploratory outcomes. *** p<0.01, ** p<0.05, * p<0.10.

Table 7: Income and employment: 2.5-year follow-up

	(1) Any net income: livestock	(2) Net income: livestock	(3) Any net income: non-ag.prod.	(4) Net income: non-ag.prod.	(5) Formal empl.[e]	(6) Informal empl.[e]
T2 – No grant	0.044**	20.141	0.003	-23.958	0.001	-0.069***
	(0.019)	(13.113)	(0.025)	(15.709)	(0.011)	(0.025)
q-value	[0.029]	[0.155]	[0.559]	[0.155]		
T2 – Grant	0.127***	68.597***	0.037	-23.065	0.052***	-0.066*
	(0.034)	(14.105)	(0.038)	(35.194)	(0.019)	(0.038)
q-value	[0.001]	[0.001]	[0.309]	[0.395]		
T3 – No grant	0.101***	58.906***	0.015	-13.691	-0.014	-0.058**
	(0.018)	(11.495)	(0.024)	(15.250)	(0.011)	(0.027)
q-value	[0.001]	[0.001]	[0.395]	[0.328]		
T3 – Grant	0.167***	74.213***	0.043	-23.875	0.006	-0.092**
	(0.033)	(15.521)	(0.034)	(32.652)	(0.015)	(0.037)
q-value	[0.001]	[0.001]	[0.219]	[0.380]		
ANCOVA	No	No	No	No	No	No
p-value: T2 = T3	0.004	0.012	0.623	0.646	0.019	0.983
p-value: T2-G = T2-NG	0.025	0.005	0.409	0.981	0.012	0.944
p-value: T3-G = T3-NG	0.062	0.396	0.464	0.776	0.230	0.467
Control mean	0.77	89.93	0.28	116.95	0.05	0.39
Observations	2902	2902	2902	2902	2902	2902

Notes: [e] = Exploratory outcome (not pre-specified). Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). See Table A4 in the Appendix for full definitions of the outcome variables. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, and education indicators. Standard errors clustered at the kebele level in parentheses. Sharpened q-values (Anderson 2008) in brackets; not reported for exploratory outcomes. *** p<0.01, ** p<0.05, * p<0.10.

Table 8: Cost-effectiveness: height-for-age Z-score

	SPIR T3–G	Ahmed et al.	Carneiro et al.		Bouguen & Dillon
	Ethiopia	Bangladesh	Nigeria		Burkina Faso
	(1)	(2)	(3a)	(3b)	(4)
Panel A. Program cost per household (2021 PPP)					
Cash transfers	1,248	2,087	789		743
Livelihood transfer	734	—	—		366
BCC, delivery, admin	250	1,188	158†	449†	149†
Total cost	**2,231**	**3,275**	**946**	**1,237**	**1,257**
Panel B. Cost-effectiveness on child height-for-age Z-score					
HAZ ITT (SD)	0.226**	0.248***	0.135**		0.182**
(SE)	(0.102)	(0.080)	(0.061)		(0.070)
Child age range	24–48 mo	0–48 mo	24–48 mo		0–60 mo
Cost per 0.1 SD HAZ	**987**	**1,201**	**701**	**917**	**606**
95% CI	[524, 8,555]	[736, 3,265]	[372, 6,129]	[486, 8,014]	[346, 2,462]

Notes: PPP = purchasing power parity; BCC = Behavioral communication change; SE = standard error; CI = Confidence interval, HAZ = height for age Z-score. All monetary values in 2021 PPP per household. SPIR T3-Grant costs are marginal relative to T1 control and reflect actual program accounting. † BCC, delivery and administration costs are not directly observed. The daggered cells are based on assumptions: BCC and administrative costs are 10% each of the cash transfer value in columns (3a) and (4), Ahmed et al.-observed BCC and administrative shares in column (3b). 95% confidence intervals are calculated from the reported standard errors associated with the intention-to-treat estimates. *** p<0.01, ** p<0.05, * p<0.10.

Table 9: Cost-benefit analysis of SPIR treatment arms

	T2-NG (1)	T2-G (2)	T3-NG (3)	T3-G (4)
Panel A. Program costs per household (2021 PPP)				
Base SPIR cost	249.7	249.7	249.7	249.7
Livelihood grant	—	733.6	—	733.6
Maternal cash transfers	—	—	1247.8	1247.8
Total marginal cost (Year 0)	249.7	983.3	1497.5	2231.1
Panel B. Benefits per household (PV at 5%)				
Consumption pathway				
One-year ITT (daily p.c.)	-0.012	0.056	0.140**	0.276***
2.5-year ITT (daily p.c.)	0.043	0.030	0.206***	0.304***
A: 10yr, no decay	502.0	528.6	2952.3	4549.2
B: 10yr, 10% decay	368.4	435.4	2311.9	3604.1
C: 20yr, no decay	903.8	808.9	4877.1	7389.6
D: 20yr, 10% decay	485.7	517.2	2874.1	4433.8
Child HAZ pathway				
HAZ ITT (SD)	-0.111	0.168	0.012	0.226**
NPV lifetime consumption gain	-190.9	288.9	20.6	388.7
Panel C. Benefit-cost ratio				
Consumption only				
A: 10yr, no decay	2.01	0.54	1.97	2.04
B: 10yr, 10% decay	1.47	0.44	1.54	1.62
C: 20yr, no decay	3.62	0.82	3.26	3.31
D: 20yr, 10% decay	1.94	0.53	1.92	1.99
Combined (consumption + child HAZ)				
A: 10yr, no decay	1.25	0.83	1.99	2.21
B: 10yr, 10% decay	0.71	0.74	1.56	1.79
C: 20yr, no decay	2.85	1.12	3.27	3.49
D: 20yr, 10% decay	1.18	0.82	1.93	2.16
Panel D. Internal rate of return (%)				
Consumption only				
A: 10yr, no decay	18.0	-7.0	20.9	22.8
B: 10yr, 10% decay	12.6	-11.9	16.0	18.0
C: 20yr, no decay	22.2	2.6	24.4	25.9
D: 20yr, 10% decay	15.4	-4.5	18.2	20.0
Combined (consumption + child HAZ)				
A: 10yr, no decay	7.2	-0.8	21.3	29.0
B: 10yr, 10% decay	1.8	-5.7	16.4	24.5
C: 20yr, no decay	13.5	7.0	24.7	31.4
D: 20yr, 10% decay	6.3	0.1	18.6	25.9

Notes: All monetary values in 2021 PPP per household. Costs are marginal relative to T1 control. Consumption benefits: per capita daily ITT × household size (5.4) × 365. Year 2 benefits are the simple average of Year 1 (one-year) and Year 3 (2.5-year) benefits. Post-2.5-year projected at 2.5-year level with annual decay from Year 4. HAZ pathway: NPV of lifetime consumption gain from improved HAZ, using Hoddinott et al. (2013b) estimate (1 SD HAZ → 21% higher consumption), applied over working ages 18–65 with Ethiopia survival probabilities, discounted at 5%. Combined BCR adds HAZ NPV to consumption PV. Combined IRR treats HAZ NPV as a Year-0 benefit. *** p<0.01, ** p<0.05, * p<0.10.

Table 10: Cost-benefit analysis: graduation programs comparison

		Benefit-cost ratio				Internal rate of return (%)			
Study	Paper	A	B	C	D	A	B	C	D
Ethiopia	Banerjee	0.91	0.73	1.47	0.89	3.2	-1.9	10.1	3.1
Ghana		0.52	0.41	0.85	0.51	-6.4	-11.5	3.2	-4.1
Honduras		-0.19	-0.13	-0.39	-0.19	—	—	—	—
India		1.68	1.35	2.64	1.64	18.0	13.4	21.8	15.8
Pakistan		0.76	0.62	1.19	0.74	-0.5	-5.5	7.3	0.3
Peru		0.72	0.57	1.19	0.70	-1.0	-6.2	7.1	-0.1
Bangladesh	Bandiera	1.86	1.49	3.32	1.96	17.0	13.2	21.4	16.0
Capital (Niger)	Bossuroy	4.04	2.93	6.67	3.62	46.4	39.8	47.4	40.5
Psychosocial (Niger)		10.73	7.64	18.03	9.56	94.7	88.5	94.8	88.6
Full (Niger)		6.96	5.01	11.55	6.22	73.1	66.9	73.4	67.1
Afghanistan	Bedoya	2.33	1.72	3.76	2.10	27.5	20.9	30.0	22.6
T2-NG†	SPIR	2.01	1.47	3.62	1.94	18.0	12.6	22.2	15.4
T2-G†		0.54	0.44	0.82	0.53	-7.0	-11.9	2.6	-4.5
T3-NG		1.97	1.54	3.26	1.92	20.9	16.0	24.4	18.2
T3-G		2.04	1.62	3.31	1.99	22.8	18.0	25.9	20.0

Notes: Scenario A: 10-year horizon, no decay. Scenario B: 10-year horizon, 10% annual decay. Scenario C: 20-year horizon, no decay. Scenario D: 20-year horizon, 10% annual decay. Discount rate: 5%. † Underlying consumption ITTs are not statistically significant at conventional levels.

Appendix

A Additional methodological details

A.1 Roll-out of new interventions

As noted in the main manuscript, the timing of the medium-run 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).[40] 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.[41] Accordingly, we do not assess that there is any meaningful risk of contamination to the experimental design linked to this new initiative.

A.2 Spillovers

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 cash transfers 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.

[40]A 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).

[41]In 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.

A.3 Study design

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)	*Livelihoods/financial inclusion:* village economic and savings associations (VESAs); light-touch livelihoods and financial training; *subset:* one-time livelihoods grant of $300. *Nutrition-sensitive:* 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 follow-up surveys in August–September 2023 and August–September 2025. The first survey was conducted largely as planned, though data collection in Amhara was delayed until October–December 2023 due to conflict-related inaccessibility. The second follow-up was fielded approximately three months early to avoid confounding with the anticipated PSNP-wide rollout of livelihoods grants. As a result, children at the medium-run follow-up were 24–48 months of age rather than the intended 30–48 months. 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-ups, driven primarily by women enrolled while pregnant who did not have a live infant at the one-year follow-up (34% of pregnant women). The sample of eligible children was 2,234 at the one-year follow-up and 2,203 at 2.5 years, rather than the approximately 3,000 originally projected. The minimum detectable effect for HAZ increased from 0.14 to 0.18 standard deviations.
Primary specification	The PAP specified estimating equations using T2 and T3 treatment indicators (or a pooled indicator), with strata fixed effects and ANCOVA controls. The paper instead interacts T2 and T3 with livelihood grant eligibility, yielding four treatment cells (T2–No grant, T2–Grant, T3–No grant, T3–Grant). The PAP listed grant-eligibility heterogeneity as exploratory. This change was motivated by the observation that the simpler specification obscures economically meaningful heterogeneity by grant status, particularly for child HAZ.
ECD instrument	The PAP specified the Strengths and Difficulties Questionnaire (SDQ). The paper uses the CREDI short form (McCoy et al., 2018) instead, which is better suited to the age range of the sample (24–48 months at the medium-run follow-up).
Exploratory analyses	Several exploratory analyses were added post hoc: treatment effects on diarrhea prevalence, paternal IYCF knowledge, the child stimulation index, open defecation, household dietary diversity (HDDS), PSNP participation, formal and informal employment, formal credit access, and crop income. Exploratory analyses of maternal engagement with health and nutrition services were also added, as were detailed attrition analyses.
Cost-benefit analysis	Not pre-specified. This analysis was added to contextualize findings relative to the broader graduation model literature.

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).	One year, 2.5 years
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.	One year
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.	2.5 years
Secondary outcomes: Nutrition and child development		
Height-for-age (children 6-23 months)	See above.	One year
Stunting (children 24-48 months)	A binary indicator obtaining value 1 if HAZ <-2 and zero otherwise	2.5 years
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.	One year
Weight-for-height (children 24-48 months)	See above.	2.5 years
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.	One year
Meal frequency (children 6–23 months)	Number of times the child ate solid, semi-solid, or soft foods in the 24 hours preceding the interview, as reported by the caregiver.	One year
Breastfed yesterday (children 6–23 months)	Binary indicator equal to one if the child was breastfed during the day or night preceding the interview.	One year

Continued on next page

Outcome Variable	Definition	When assessed?
Food group consumption (children 6–23 months)	Seven binary indicators, each equal to one if the child consumed any food from the respective group in the 24 hours preceding the interview: (1) grains, roots, and tubers; (2) legumes and nuts; (3) dairy products; (4) flesh foods; (5) eggs; (6) vitamin A rich fruits and vegetables; (7) other fruits and vegetables.	One year
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 2.5 years. 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).	2.5 years
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. This measure also includes food away from home, captured through household expenditures on prepared foods and meals consumed outside the household during the previous 7 days. 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 2021 purchasing power parity (PPP) dollars, using PPP exchange rates for household final consumption expenditure provided by the World Bank.	One year and 2.5 years
Food consumption (per capita)	Daily per capita food consumption, derived from the food component of the consumption aggregate described above. Measured in 2021 PPP dollars.	One year and 2.5 years

Continued on next page

Outcome Variable	Definition	When assessed?
Non-food expenditure (per capita)	Daily per capita non-food expenditure, derived from the non-food component of the consumption aggregate described above. Measured in 2021 PPP dollars.	One year and 2.5 years
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 2.5 years rounds. The index is then normalized with respect to baseline so that at baseline it gets zero mean and SD of 1.	2.5 years
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 2021 PPP dollars) held in these formal institutions.	2.5 years
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).	One year and 2.5 years
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.	2.5 years
Gross income from livestock production	Total revenues from livestock production, including imputed value of live animal sales (valued at current unit values), slaughtered meat (sold and consumed), cow/goat milk, and eggs (milk and eggs valued using district-median prices). Measured in 2021 PPP dollars.	2.5 years
Total livestock costs	Total expenditure on livestock inputs in the past 12 months, including hired labor, purchased feed, medicines, veterinary services, and other inputs. Measured in 2021 PPP dollars.	2.5 years

Continued on next page

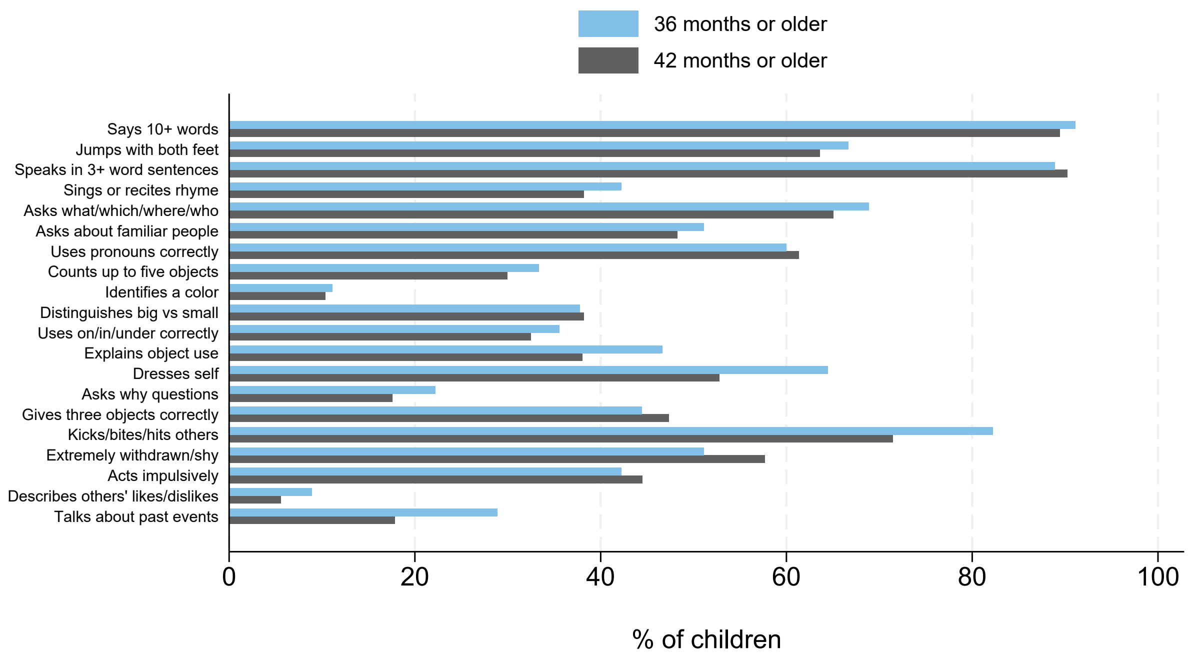
Outcome Variable	Definition	When assessed?
Livestock counts (by type)	Number of animals owned by the household, reported separately for cattle, sheep and goats, pack animals, and poultry.	2.5 years
Net income from any non-agricultural production	Calculated as the sum of household formal wage income and 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.	2.5 years
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.	2.5 years
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 2021 PPP dollars).	2.5 years
Additional outcomes (not prespecified): Nutrition and child development		
Early childhood development score	See above.	One year
Paternal IYCF knowledge	Same as caregiver IYCF knowledge described above, but the respondent is primary male.	2.5 years
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.	2.5 years
Diarrhea prevalence	Binary indicator equal to one if the caregiver reported that the index child had diarrhea in the two weeks preceding the interview.	One year and 2.5 years
Open defecation	A binary indicator equal to one if the household has no toilet facility and practices open defecation.	2.5 years

Continued on next page

Outcome Variable	Definition	When assessed?
Child has health card	Binary indicator equal to one if the index child has a health card or child health booklet.	One year
Contact with health extension worker	Binary indicator equal to one if the primary female had contact with a Health Extension Worker (HEW) in the three months preceding the interview.	One year and 2.5 years
Visited health post	Binary indicator equal to one if the primary female visited a health post for herself or her child in the three months preceding the interview.	One year and 2.5 years
Contact with Health Development Army	Binary indicator equal to one if the primary female had contact with a Health Development Army (HDA) member in the three months preceding the interview.	One year and 2.5 years
Attended food demonstration	Binary indicator equal to one if the primary female attended a food demonstration in the three months preceding the interview.	One year and 2.5 years
Attended community conversation	Binary indicator equal to one if the primary female attended a community conversation session in the three months preceding the interview.	One year and 2.5 years
Additional outcomes (not prespecified): Livelihoods		
Any formal credit	Binary indicator equal to one if the household obtained credit from a formal source (bank, microfinance institution, or other formal financial institution) in the past 12 months.	2.5 years
Formal employment	Binary indicator equal to one if any household member held formal wage employment in the past 12 months, defined as work with a written contract or a contract lasting at least two weeks.	2.5 years
Informal employment	Binary indicator equal to one if any household member held informal wage employment in the past 12 months, defined as work without a written contract and with a contract lasting less than two weeks.	2.5 years
Crop income	Total value of crops harvested, measured in 2021 PPP dollars.	2.5 years
PSNP beneficiary	Binary indicator equal to one if the household was a Productive Safety Net Programme beneficiary in 2025, through either public works or direct support.	2.5 years

B Additional results

Figure B1: 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 B1: Baseline balance: grant-eligible vs. non-eligible households

Variable	(1) Non-eligible Mean/(SE)	(2) Eligible Mean/(SE)	(1)-(2) Pairwise t-test Mean difference
Currently pregnant	0.529	0.539	-0.010
	(0.007)	(0.016)	
Child is 0-3 months old	0.199	0.189	0.010
	(0.008)	(0.013)	
Child is 4-6 months old	0.147	0.133	0.014
	(0.007)	(0.011)	
Child is 7-9 months old	0.123	0.136	-0.013
	(0.007)	(0.011)	
Household size	5.607	4.912	0.695***
	(0.069)	(0.091)	
Primary female has some formal education	0.436	0.424	0.011
	(0.013)	(0.018)	
Mother's IYCF knowledge score (max 11 points)	8.511	8.263	0.248***
	(0.097)	(0.100)	
Food Insecurity Experience Scale (FIES), raw score	5.987	6.471	-0.484***
	(0.067)	(0.076)	
Household daily per capita consumption (2021 PPP)	2.820	2.591	0.229***
	(0.048)	(0.052)	
Tropical Livestock Units	0.792	0.234	0.559***
	(0.027)	(0.015)	
Durable asset index (PCA)	0.133	-0.302	0.435***
	(0.033)	(0.040)	
Household reports formal savings	0.555	0.460	0.095***
	(0.018)	(0.022)	
F-test of joint significance (F-stat)			99.019***
Number of observations	2091	924	3015
Number of clusters	234	234	234

Notes: Pairwise mean differences with strata fixed effects. Grant eligibility = bottom 4 households per kebele by baseline asset index. Standard errors clustered at the kebele level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table B2: Child nutrition variables in control arm: grant and non-grant eligible households

	(1) Non-eligible		(2) Eligible		(1)-(2)	
	N / Clusters	Mean/ (SE)	N / Clusters	Mean/ (SE)	N / Clusters	Mean difference
Panel A: Short-run follow-up						
	(1) Non-elig.		(2) Elig.		(1)-(2) Pairwise t-test	
IYCF knowledge score	539	9.469	232	9.595	771	-0.125
	80	(0.100)	79	(0.122)	80	
Dietary diversity score (8 groups)	539	3.308	232	3.108	771	0.200**
	80	(0.067)	79	(0.083)	80	
Minimum acceptable diet	536	0.101	228	0.061	764	0.039**
	80	(0.015)	79	(0.017)	80	
Height-for-age z-score	537	-1.592	232	-1.653	769	0.061
	80	(0.060)	79	(0.074)	80	
Stunted (HAZ < -2)	537	0.378	232	0.362	769	0.016
	80	(0.021)	79	(0.034)	80	
Weight-for-height z-score	537	-0.526	232	-0.553	769	0.027
	80	(0.050)	79	(0.069)	80	
Panel B: Medium-run follow-up						
	(1) Non-elig.		(2) Elig.		(1)-(2) Pairwise t-test	
IYCF knowledge score	517	9.250	222	8.950	739	0.299*
	79	(0.093)	77	(0.140)	79	
Height-for-age z-score	526	-2.158	225	-2.317	751	0.159**
	79	(0.066)	77	(0.086)	79	
Stunted (HAZ < -2)	526	0.557	225	0.613	751	-0.056*
	79	(0.025)	77	(0.037)	79	
Weight-for-height z-score	526	-0.414	225	-0.477	751	0.063
	79	(0.038)	77	(0.068)	79	
CREDI z-score	526	0.084	225	-0.151	751	0.234***
	79	(0.042)	77	(0.077)	79	

Notes: Sample restricted to control arm (T1). Column (1) shows the mean (standard error) for grant-ineligible households. Column (2) shows the mean (standard error) for grant-eligible households. Column (3) shows the pairwise mean difference. Children are 6–23 months at one-year follow-up and 24–48 months at the 2.5-year follow-up. All tests include strata fixed effects. Standard errors clustered at the kebele level. *** p<0.01, ** p<0.05, * p<0.10.

Table B3: Program exposure

	T1		T2		T3	
	%	(N)	%	(N)	%	(N)
Ever participated in neighborhood group (1-yr)	22.7	(961)	73.6	(927)	87.3	(948)
Ever participated in neighborhood group (3-yr)	24.1	(999)	59.6	(935)	78.5	(968)
Ever participated in neighborhood group (1-yr or 3-yr)	36.2	(1028)	84.4	(989)	92.8	(989)
Ever received maternal cash grant (1-yr)	1.2	(961)	1.8	(924)	88.7	(947)
Ever received maternal cash grant (3-yr)	2.6	(999)	3.4	(935)	80.6	(968)
Ever received maternal cash grant (1-yr or 3-yr)	3.5	(1028)	5.0	(989)	92.6	(989)
Ever participated in VESA (1-yr)	15.8	(1004)	63.2	(970)	71.4	(980)
Ever participated in VESA (3-yr)	32.0	(999)	78.3	(935)	82.6	(968)
Ever participated in VESA (1-yr or 3-yr)	38.8	(1028)	87.2	(989)	91.1	(990)
Receipt of livelihoods cash transfer (1-yr)	4.0	(1004)	38.4	(970)	35.2	(980)

Note: 1-yr = one-year follow-up; 2.5-yr = 2.5-year follow-up. VESA = Village Economic and Social Association.

Table B4: Nutrition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: One-year follow-up (Children are 6-23 months)								
	IYCF know. †	Health card[e]	Diet diversity†	MAD	HAZ	Stunted	WHZ	CREDI[e]
T2	0.28***	-0.01	0.11*	0.01	-0.01	0.01	-0.09	0.03
	(0.09)	(0.03)	(0.06)	(0.02)	(0.06)	(0.02)	(0.06)	(0.04)
q-value	[0.001]		[0.017]	[0.217]	[0.288]	[0.235]	[0.060]	
T3	0.43***	0.03	0.58***	0.11***	0.09	-0.03	0.05	0.19***
	(0.09)	(0.03)	(0.06)	(0.02)	(0.06)	(0.02)	(0.06)	(0.05)
q-value	[0.001]		[0.001]	[0.001]	[0.063]	[0.075]	[0.163]	
ANCOVA?	Yes	No	No	No	No	No	No	No
T2 vs T3 (p-value)	0.052*	0.146	0.000***	0.000***	0.093*	0.055*	0.011**	0.000***
Control mean	9.51	0.29	3.25	0.09	-1.61	0.37	-0.53	-0.00
Observations	2231	2233	2233	2221	2226	2226	2226	2212
Panel B: 2.5-year follow-up (Children are 24-48 months)								
	IYCF knowledge†	Paternal IYCF know.[e]	HAZ†	Stunted	WHZ	Stimulation[e]	CREDI[e]	
T2	0.29***	0.24**	-0.03	0.01	-0.07	0.03	0.03	
	(0.08)	(0.11)	(0.06)	(0.02)	(0.05)	(0.06)	(0.04)	
q-value	[0.001]		[0.128]	[0.235]	[0.088]			
T3	0.27***	0.22**	0.07	-0.03	0.01	0.22***	0.12***	
	(0.07)	(0.11)	(0.06)	(0.02)	(0.05)	(0.06)	(0.04)	
q-value	[0.001]		[0.068]	[0.100]	[0.258]			
ANCOVA?	Yes	Yes	No	No	No	No	No	
T2 vs T3 (p-value)	0.824	0.888	0.106	0.097*	0.126	0.005***	0.041**	
Control mean	9.16	8.63	-2.21	0.57	-0.43	1.82	0.01	
Observations	2163	1316	2203	2203	2203	2210	2195	

Notes: † = Primary outcome. [e] = Exploratory outcome (not pre-specified). IYCF knowledge = maternal knowledge of infant and young child feeding practices; Health card = child has a health card; 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; Paternal IYCF knowledge = paternal knowledge of IYCF practices; Stimulation = child stimulation activities index; CREDI = Caregiver Reported Early Development Instruments 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 2.5-year 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); not reported for exploratory outcomes.

Table B5: Consumption and food insecurity

	(1)	(2)	(3)	(4)	(5)
Panel A: Consumption and food insecurity, one year follow-up					
	Per capita cons.	Food cons.[e]	Nonfood cons.[e]	Food insecurity binary	Food insecurity cont.
Pooled T	0.10**	0.09**	0.01*	-0.01	-0.10
	(0.04)	(0.04)	(0.00)	(0.01)	(0.07)
q-value	[0.011]			[0.103]	[0.060]
ANCOVA?	Yes	Yes	Yes	Yes	Yes
Control mean	2.27	2.05	0.22	0.91	5.91
Observations	2954	2954	2954	2954	2954
Panel B: Consumption and food insecurity, three year follow-up					
	Per capita cons.	Food cons.[e]	Nonfood cons.[e]	Food insecurity binary	Food insecurity cont.
Pooled T	0.14***	0.12**	0.03***	0.01	-0.03
	(0.05)	(0.05)	(0.01)	(0.01)	(0.07)
q-value	[0.006]			[0.198]	[0.217]
ANCOVA?	Yes	Yes	Yes	Yes	Yes
Control mean	2.25	1.90	0.35	0.88	5.96
Observations	2902	2902	2902	2902	2902

Notes: [e] = Exploratory outcome (not pre-specified). 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 2.5-year 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); not reported for exploratory outcomes.

Table B6: Assets, financial inclusion, income, and employment

	(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***	0.12***	0.22***	23.32***	0.07***	74.34***
	(0.04)	(0.03)	(0.02)	(2.55)	(0.02)	(14.53)
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	31.23	0.68	173.30
Observations	2902	2902	2902	2902	2902	2902
Panel B: Income and employment, three year follow-up						
	Any net income: livestock	Net income: livestock	Any net income: non-ag.prod.	Net income: non-ag.prod.	Formal empl.[e]	Informal empl.[e]
Pooled T	0.10***	50.04***	0.02	-19.91	0.00	-0.07***
	(0.01)	(8.67)	(0.02)	(12.52)	(0.01)	(0.02)
q-value	[0.001]	[0.001]	[0.144]	[0.060]		
ANCOVA?	No	No	No	No	No	No
Control mean	0.77	89.93	0.28	116.95	0.05	0.39
Observations	2902	2902	2902	2902	2902	2902

Notes: [e] = Exploratory outcome (not pre-specified). 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. Formal empl. = binary indicator for formal wage employment; Informal empl. = binary indicator for informal wage employment. Standard errors clustered at the kebele level (unit of treatment) are reported in parentheses, and the 2.5-year 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); not reported for exploratory outcomes.

Table B7: Consumption by food group: one-year follow-up (children 6–23 months)

	(1) Breastmilk	(2) Staples	(3) Legumes or nuts	(4) Dairy	(5) Meat	(6) Eggs	(7) Vit. A F&V	(8) Other F&V
T2 – No grant	0.008	0.001	-0.009	-0.012	-0.001	0.011	0.026	0.019
	(0.015)	(0.009)	(0.031)	(0.029)	(0.008)	(0.021)	(0.022)	(0.028)
T2 – Grant	-0.012	-0.005	0.046	0.028	0.001	0.065***	0.056	0.052
	(0.019)	(0.018)	(0.038)	(0.039)	(0.009)	(0.025)	(0.040)	(0.038)
T3 – No grant	0.013	0.020**	0.020	0.169***	-0.001	0.215***	0.115***	0.028
	(0.014)	(0.010)	(0.028)	(0.029)	(0.008)	(0.025)	(0.023)	(0.026)
T3 – Grant	0.000	-0.005	-0.015	0.157***	0.026*	0.239***	0.066*	0.106***
	(0.018)	(0.016)	(0.038)	(0.043)	(0.014)	(0.033)	(0.039)	(0.037)
ANCOVA	No	No	No	No	No	No	No	No
p-value: T2 = T3	0.581	0.108	0.936	0.000	0.416	0.000	0.003	0.358
p-value: T2-G = T2-NG	0.414	0.784	0.238	0.407	0.824	0.051	0.515	0.468
p-value: T3-G = T3-NG	0.567	0.171	0.450	0.801	0.022	0.507	0.276	0.066
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

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). Each outcome is a binary indicator for consumption from the indicated food group in the previous day. Sample is children aged 6–23 months at the one-year follow-up. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, education indicators, child sex, and child age in months. Standard errors clustered at the kebele level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table B8: Child health and health care utilization

	(1) Diarrhea (one year)	(2) Open def. (2.5 years)	(3) Diarrhea (2.5 years)	(4) Health card	(5) Contact with HEW	(6) Health post visit	(7) Contact with HDA	(8) Food demo.	(9) Community conv.
T2 – No grant	-0.024	-0.021	-0.017	-0.032	0.134***	0.077**	0.138***	0.250***	0.193***
	(0.025)	(0.033)	(0.029)	(0.029)	(0.029)	(0.032)	(0.022)	(0.028)	(0.025)
T2 – Grant	-0.052	-0.026	0.066	0.050	0.027	0.066	0.195***	0.186***	0.179***
	(0.042)	(0.048)	(0.043)	(0.042)	(0.044)	(0.044)	(0.034)	(0.042)	(0.041)
T3 – No grant	-0.041*	-0.080**	-0.128***	0.033	0.158***	0.067**	0.168***	0.400***	0.312***
	(0.025)	(0.031)	(0.028)	(0.031)	(0.030)	(0.033)	(0.021)	(0.028)	(0.026)
T3 – Grant	-0.077*	-0.069	-0.018	0.022	0.093**	0.058	0.180***	0.335***	0.285***
	(0.045)	(0.043)	(0.041)	(0.040)	(0.042)	(0.042)	(0.033)	(0.039)	(0.035)
ANCOVA	No	Yes	No	No	No	No	No	No	No
p-value: T2 = T3	0.410	0.054	0.000	0.145	0.141	0.741	0.412	0.000	0.000
p-value: T2-G = T2-NG	0.544	0.925	0.104	0.083	0.027	0.825	0.117	0.177	0.754
p-value: T3-G = T3-NG	0.468	0.826	0.019	0.803	0.151	0.867	0.741	0.126	0.486
Control mean	0.37	0.43	0.32	0.29	0.55	0.60	0.12	0.08	0.08
Observations	2233	2210	2209	2233	2227	2231	2213	2231	2231

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). Column 1 and columns 4–9 are measured at the one-year follow-up (children 6–23 months); columns 2–3 are measured at the 2.5-year follow-up (children 24–59 months). Health card = child has a health card. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, education indicators, child sex, and child age in months, and their interactions with grant eligibility. Standard errors clustered at the kebele level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table B9: Food insecurity and dietary diversity: One-year follow-up

	(1) Worried about food	(2) Ate unhealthy food	(3) Ate few kinds of food	(4) Skipped a meal	(5) Ate less than should	(6) Ran out of food	(7) Went hungry	(8) Whole day without eating	(9) HDDS	(10) Food gap
T2 – No grant	-0.003	0.009	0.007	0.018	0.046***	0.023	0.012	-0.021	0.040	0.008
	(0.009)	(0.009)	(0.010)	(0.021)	(0.016)	(0.024)	(0.023)	(0.017)	(0.098)	(0.079)
T2 – Grant	-0.020	-0.021	-0.037**	-0.029	-0.057***	-0.017	-0.006	-0.028	0.265**	-0.309***
	(0.014)	(0.013)	(0.016)	(0.025)	(0.022)	(0.037)	(0.035)	(0.028)	(0.107)	(0.107)
T3 – No grant	0.012	0.003	0.005	-0.018	0.009	-0.052**	-0.040*	-0.039**	0.862***	-0.226***
	(0.008)	(0.010)	(0.011)	(0.020)	(0.017)	(0.024)	(0.024)	(0.018)	(0.098)	(0.077)
T3 – Grant	-0.001	-0.003	0.000	-0.061**	-0.081***	-0.089***	-0.092**	-0.042*	0.998***	-0.523***
	(0.011)	(0.011)	(0.012)	(0.025)	(0.021)	(0.032)	(0.037)	(0.025)	(0.110)	(0.097)
ANCOVA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value: T2 = T3	0.052	0.898	0.308	0.023	0.008	0.000	0.005	0.217	0.000	0.001
p-value: T2-G = T2-NG	0.208	0.047	0.009	0.147	0.000	0.325	0.623	0.812	0.083	0.015
p-value: T3-G = T3-NG	0.249	0.644	0.740	0.173	0.001	0.327	0.196	0.899	0.270	0.005
Control mean	0.98	0.97	0.96	0.83	0.88	0.56	0.61	0.12	6.03	3.28
Observations	2954	2954	2954	2954	2954	2954	2954	2954	2953	2953

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). Columns 1–8 are binary indicators for specific food insecurity experiences in the past 30 days (FIES items). HDDS = Household Dietary Diversity Score (number of food groups consumed in past 24 hours, range 0–12). Food gap = number of months with inadequate food provisioning in past 12 months (range 0–12). All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, and education indicators, and their interactions with grant eligibility. ANCOVA denotes whether the lagged dependent variable was controlled for. Standard errors clustered at the kebele level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table B10: Food insecurity and dietary diversity: 2.5-year follow-up

	(1) Worried about food	(2) Ate unhealthy food	(3) Ate few kinds of food	(4) Skipped a meal	(5) Ate less than should	(6) Ran out of food	(7) Went hungry	(8) Whole day without eating	(9) HDDS	(10) Food gap
T2 – No grant	0.004	-0.002	0.006	0.033*	0.025	0.030	0.034	0.004	0.265***	0.050
	(0.012)	(0.011)	(0.011)	(0.020)	(0.016)	(0.021)	(0.025)	(0.020)	(0.086)	(0.097)
T2 – Grant	-0.023**	-0.000	0.007	0.003	0.009	-0.048	-0.019	-0.026	0.481***	-0.161
	(0.011)	(0.011)	(0.012)	(0.027)	(0.022)	(0.032)	(0.040)	(0.035)	(0.129)	(0.149)
T3 – No grant	-0.010	-0.009	-0.009	-0.004	-0.006	-0.018	-0.001	-0.004	0.704***	-0.087
	(0.013)	(0.011)	(0.011)	(0.022)	(0.019)	(0.023)	(0.024)	(0.018)	(0.088)	(0.098)
T3 – Grant	-0.007	-0.012	-0.008	-0.040	-0.002	-0.076**	-0.092**	-0.048	0.894***	-0.235*
	(0.012)	(0.011)	(0.013)	(0.029)	(0.024)	(0.032)	(0.037)	(0.030)	(0.128)	(0.141)
ANCOVA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value: T2 = T3	0.608	0.365	0.093	0.040	0.119	0.024	0.040	0.491	0.000	0.203
p-value: T2-G = T2-NG	0.124	0.914	0.962	0.367	0.567	0.046	0.248	0.432	0.081	0.210
p-value: T3-G = T3-NG	0.854	0.851	0.918	0.300	0.895	0.138	0.029	0.179	0.121	0.395
Control mean	0.95	0.95	0.95	0.79	0.87	0.62	0.60	0.22	5.98	3.26
Observations	2902	2902	2902	2902	2902	2902	2902	2902	2901	2902

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). Columns 1–8 are binary indicators for specific food insecurity experiences in the past 30 days (FIES items). HDDS = Household Dietary Diversity Score (number of food groups consumed in past 24 hours, range 0–12). Food gap = number of months with inadequate food provisioning in past 12 months (range 0–12). All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, and education indicators, and their interactions with grant eligibility. ANCOVA denotes whether the lagged dependent variable was controlled for. Standard errors clustered at the kebele level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table B11: Supplemental livelihoods outcomes: 2.5-year follow-up

	(1) Any formal credit	(2) Crop harvest PPP	(3) Gross LS income	(4) PSNP
T2 – No grant	0.235***	-9.073	18.354	0.002
	(0.030)	(28.060)	(14.713)	(0.012)
T2 – Grant	0.141***	37.645	77.855***	0.005
	(0.031)	(31.435)	(15.794)	(0.012)
T3 – No grant	0.173***	31.015	68.232***	0.010
	(0.026)	(26.178)	(13.259)	(0.011)
T3 – Grant	0.132***	26.723	89.693***	0.014
	(0.031)	(32.910)	(17.519)	(0.010)
ANCOVA	No	No	No	No
p-value: T2 = T3	0.100	0.283	0.003	0.333
p-value: T2-G = T2-NG	0.008	0.233	0.002	0.824
p-value: T3-G = T3-NG	0.221	0.901	0.300	0.697
Control mean	0.14	500.71	116.48	0.97
Observations	2902	2902	2902	2902

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). 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 2.5-year control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table B12: Disaggregated livestock ownership: 2.5-year follow-up

	(1) Num. poultry	(2) Num. sheep	(3) Num. goats	(4) Num. calves/ heifers	(5) Num. cows	(6) Num. oxen
T2 – No grant	0.358**	-0.085**	0.157*	0.002	0.007	-0.010
	(0.164)	(0.035)	(0.083)	(0.029)	(0.024)	(0.016)
T2 – Grant	0.540**	-0.037	0.401***	0.099***	0.079***	0.008
	(0.236)	(0.044)	(0.091)	(0.038)	(0.028)	(0.019)
T3 – No grant	0.922***	-0.058	0.274***	0.085***	0.038	0.018
	(0.180)	(0.039)	(0.081)	(0.030)	(0.026)	(0.016)
T3 – Grant	1.091***	0.042	0.501***	0.106***	0.060**	0.014
	(0.279)	(0.056)	(0.113)	(0.037)	(0.028)	(0.017)
ANCOVA	Yes	Yes	Yes	Yes	Yes	Yes
p-value: T2 = T3	0.004	0.198	0.113	0.014	0.459	0.073
p-value: T2-G = T2-NG	0.408	0.300	0.042	0.042	0.029	0.453
p-value: T3-G = T3-NG	0.511	0.167	0.092	0.666	0.563	0.881
Control mean	1.87	0.19	0.55	0.26	0.21	0.09
Observations	2902	2902	2902	2902	2902	2902

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). Calves/heifers includes local and crossbred heifers and calves. Cows includes local and crossbred cows. All regressions control for strata fixed effects, the baseline value of the outcome variable interacted with grant eligibility, 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 2.5-year control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table B13: Livestock production outcomes: 2.5-year follow-up

	(1) Any egg production	(2) Total eggs	(3) Any milk production	(4) Any cow milk	(5) Cow milk (liters/day)	(6) Any goat milk	(7) Goat milk (liters/day)
T2 – No grant	0.102***	63.212***	-0.017	-0.006	-0.025	-0.006	-0.005
	(0.028)	(15.405)	(0.026)	(0.024)	(0.041)	(0.012)	(0.010)
T2 – Grant	0.102***	41.604***	0.047*	0.036*	0.041	0.020	-0.004
	(0.038)	(13.023)	(0.027)	(0.021)	(0.032)	(0.021)	(0.027)
T3 – No grant	0.184***	81.370***	0.047*	0.027	0.053	0.032**	0.033***
	(0.025)	(14.228)	(0.025)	(0.021)	(0.040)	(0.015)	(0.013)
T3 – Grant	0.138***	61.612***	0.087***	0.058**	0.111**	0.033	0.012
	(0.036)	(14.584)	(0.028)	(0.023)	(0.055)	(0.020)	(0.023)
ANCOVA	No	No	No	No	No	No	No
p-value: T2 = T3	0.006	0.274	0.007	0.116	0.027	0.009	0.002
p-value: T2-G = T2-NG	1.000	0.246	0.075	0.166	0.172	0.301	0.991
p-value: T3-G = T3-NG	0.261	0.283	0.260	0.301	0.401	0.960	0.377
Control mean	0.45	64.06	0.20	0.16	0.23	0.05	0.03
Observations	2902	2890	2902	2902	2902	2902	2902

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). 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 control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

Table B14: Primary use of livestock production: 2.5-year follow-up

	Cow milk		Goat milk		Eggs	
	(1)	(2)	(3)	(4)	(5)	(6)
	Own consumption	Sales	Own consumption	Sales	Own consumption	Sales
T2 – No grant	0.120*	-0.106*	0.046	-0.046	0.008	0.006
	(0.064)	(0.059)	(0.043)	(0.043)	(0.035)	(0.034)
T2 – Grant	0.039	-0.039	-0.071	0.071	0.013	-0.015
	(0.126)	(0.126)	(0.064)	(0.064)	(0.059)	(0.058)
T3 – No grant	0.087	-0.085	-0.051	0.051	0.126***	-0.099***
	(0.059)	(0.057)	(0.037)	(0.037)	(0.034)	(0.033)
T3 – Grant	-0.071	0.071	0.032	-0.032	0.178***	-0.182***
	(0.123)	(0.123)	(0.048)	(0.048)	(0.056)	(0.055)
ANCOVA	No	No	No	No	No	No
p-value: T2 = T3	0.405	0.499	0.532	0.532	0.000	0.000
p-value: T2-G = T2-NG	0.556	0.620	0.146	0.146	0.935	0.748
p-value: T3-G = T3-NG	0.216	0.218	0.176	0.176	0.386	0.167
Control mean	0.75	0.23	1.00	0.00	0.41	0.55
Observations	495	495	175	175	1579	1579

Notes: Each column reports estimates from a single regression of a binary indicator for the primary use of production on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). Samples are conditional on producing: columns (1)–(2) include households that milked cows; columns (3)–(4) include households that milked goats; columns (5)–(6) include households with egg-laying hens. All regressions control for strata fixed effects and baseline covariates: household size, primary female's age, and education indicators. Standard errors clustered at the kebele level in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table B15: Household egg and milk consumption

	One-year follow-up				2.5-year follow-up			
	(1) Any eggs	(2) Eggs (number)	(3) Any milk	(4) Milk (liters)	(5) Any eggs	(6) Eggs (number)	(7) Any milk	(8) Milk (liters)
T2 – No grant	0.029	0.191	0.003	0.055	0.091***	0.722***	0.003	0.074
	(0.022)	(0.225)	(0.028)	(0.064)	(0.021)	(0.201)	(0.022)	(0.082)
T2 – Grant	0.093***	0.465	0.032	0.036	0.085***	0.832**	0.096***	0.202***
	(0.026)	(0.284)	(0.031)	(0.078)	(0.030)	(0.353)	(0.031)	(0.074)
T3 – No grant	0.290***	2.502***	0.132***	0.265***	0.227***	2.289***	0.089***	0.305***
	(0.024)	(0.253)	(0.029)	(0.070)	(0.024)	(0.244)	(0.022)	(0.086)
T3 – Grant	0.283***	2.196***	0.117***	0.247***	0.224***	2.563***	0.129***	0.216***
	(0.030)	(0.329)	(0.034)	(0.087)	(0.033)	(0.460)	(0.031)	(0.072)
ANCOVA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value: T2 = T3	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.041
p-value: T2-G = T2-NG	0.037	0.375	0.430	0.835	0.863	0.773	0.010	0.274
p-value: T3-G = T3-NG	0.819	0.397	0.697	0.851	0.931	0.578	0.257	0.403
Control mean	0.11	0.86	0.26	0.38	0.12	0.98	0.20	0.37
Observations	2953	2954	2953	2954	2901	2902	2901	2902

Notes: Each column reports estimates from a single regression of the outcome on four treatment-by-grant-eligibility interaction terms (T2 – No grant, T2 – Grant, T3 – No grant, T3 – Grant) and a main effect for grant eligibility. Grant eligibility refers to household eligibility for the livelihood transfer (bottom 4 households per kebele by poverty ranking). Outcomes are from the household consumption module (7-day recall). Columns 1 and 5 are binary indicators for whether the household consumed eggs; columns 2 and 6 report the number of eggs consumed; columns 3 and 7 are binary indicators for whether the household consumed fresh milk; columns 4 and 8 report liters of milk consumed. All regressions control for the baseline value of the outcome variable interacted with grant eligibility, 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 control mean is shown. Statistical significance denoted with *** < 0.01, ** < 0.05, and * < 0.10.

C Disentangling grant effects from baseline asset poverty

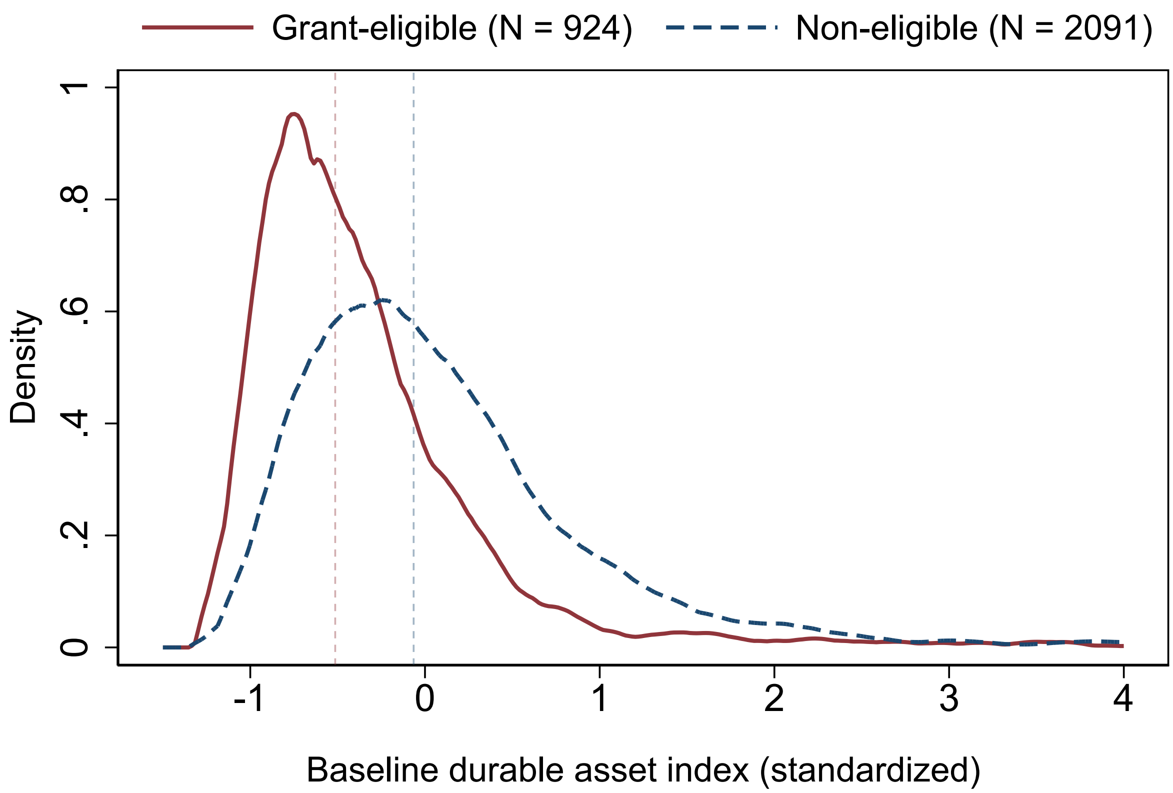
One challenge in interpreting the larger treatment effects on child growth among grant-eligible households is that these households are poorer at baseline: an alternative explanation for the differential effects is that the elasticity of child growth with respect to household investment is simply larger for poorer households. This does not affect the validity of the experimental comparison, since eligible households are identified symmetrically in treatment and control clusters. However, it would change the interpretation: the larger effects would reflect baseline poverty rather than the grant itself.

The entire sample consists of extremely poor households —- at baseline, 72 percent of the sampled individuals lived below the international extreme poverty line -— rendering it it difficult to motivate a mechanism through which relatively small differences in baseline assets would generate meaningfully different physiological responsiveness to nutrition interventions. Beyond this point, we also provide empirical evidence against this interpretation, exploiting the fact that grant eligibility was determined by ranking households within each cluster by their baseline asset index and classifying the poorest 33 percent as eligible. Because this cutoff is cluster-specific, households with the same absolute asset level can fall on different sides of the eligibility threshold depending on the wealth distribution of their cluster. Figure C1 shows that this generates substantial overlap in the baseline asset distributions of eligible and non-eligible households: 22 percent of eligible households have baseline assets above the non-eligible group's median, and 23 percent of non-eligible households fall below the eligible group's median.

Table C1 then uses this overlap to test whether the differential treatment effects by eligibility status can be explained by the baseline asset gradient. Column (1) replicates the main specification without any baseline asset controls. Column (2) adds a linear control for the baseline asset index, allowing the level of HAZ to vary with baseline wealth. Column (3) further interacts the treatment indicators with the baseline asset index, allowing treatment effects themselves to vary with baseline assets within each eligibility group. Moreover, all regressions include strata fixed effects, defined by the interaction of woreda, PSNP targeting intensity, and remoteness, so the baseline asset controls capture residual within-strata variation in household wealth rather than differences across woredas or agroecological zones.

The treatment effect estimates remain remarkably stable across these specifications, challenging the notion that the larger effects observed among eligible households reflected a simple mechanical relationship between baseline asset poverty and growth responsiveness.

Figure C1: Distribution of baseline asset index, by livelihood grant eligibility



Notes: Kernel density estimates of the baseline asset index (standardized PCA score) for grant-eligible and non-eligible households. Grant eligibility was determined by ranking households within each cluster (kebele) based on their asset index value and classifying the poorest 33 percent as eligible. Dashed vertical lines indicate group medians. Because the eligibility cutoff is cluster-specific, households with the same absolute asset level can have different eligibility status across clusters, generating the overlap visible in the figure. Of grant-eligible households, 22 percent have baseline assets above the non-eligible group's median; 23 percent of non-eligible households fall below the eligible group's median.

Table C1: Grant effect or poverty responsiveness? HAZ at 2.5 years

	Original	Asset control	Asset interactions
	(1)	(2)	(3)
T2 – No grant	-0.111	-0.108	-0.100
	(0.077)	(0.076)	(0.077)
T2 – Grant	0.168	0.181	0.196*
	(0.114)	(0.114)	(0.115)
T3 – No grant	0.012	0.009	0.018
	(0.070)	(0.069)	(0.071)
T3 – Grant	0.226**	0.236**	0.205**
	(0.102)	(0.102)	(0.104)
Treatment × asset index interactions			
T2 – No grant × Assets			-0.070
			(0.077)
T2 – Grant × Assets			0.023
			(0.123)
T3 – No grant × Assets			-0.066
			(0.057)
T3 – Grant × Assets			-0.106
			(0.091)
p-value: T2-G = T2-NG	[0.035]	[0.028]	[0.026]
p-value: T3-G = T3-NG	[0.064]	[0.050]	[0.111]
p-value: Tx × asset = 0			[0.470]
Observations	2,203	2,203	2,203
Baseline asset control	No	Yes	Yes
Tx × asset slopes	No	No	Yes

Notes: Dependent variable is height-for-age z-score (HAZ) in the 2.5-year follow-up. All columns include strata × eligibility fixed effects, household and child controls interacted with eligibility. Column (1) replicates the main specification. Columns (2)–(3) add a linear control for the baseline durable asset index (standardized PCA score), interacted with eligibility. Column (3) further interacts each treatment-eligibility cell with the asset index, allowing treatment-response gradients to differ across groups; the equality test evaluates the eligible vs. non-eligible gap at the sample mean of the asset index. Standard errors clustered at kebele level in parentheses. *** < 0.01, ** < 0.05, * < 0.10.

D Attrition and bounds

As noted in the main text, attrition in the conventional sense (at the household level) is extremely low in this trial: 2% at one year and 4% at two and a half years. 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% and 26% in two follow-up surveys). This is primarily due to high rates of miscarriage or feigned pregnancy in the sample enrolled as 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 in both follow-up surveys and regress these variables on the four treatment-by-grant-eligibility cell indicators, a main effect for grant eligibility, selected baseline covariates, and the interaction between covariates and the four treatment cells; we report the findings in Tables D1 and D2.[42] The first and fourth columns of each table report attrition rates across the four treatment-by-eligibility cells: the constant represents attrition in the control group, while the coefficients on T2–No grant, T2–Grant, T3–No grant, and T3–Grant show differences relative to control. At both follow-ups, these differences are generally small and not statistically significant in most cases. At the household level, the joint test of the four treatment cells is significant at the one year follow-up ($p = 0.006$), driven by modestly lower attrition among T3 no-grant households; the joint test is not significant at the two and a half year follow-up ($p = 0.121$). One notable exception is that grant-eligible T2 households exhibit six percentage points higher attrition in the medium-run survey ($p < 0.05$).

For child-level attrition (Table D2), 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 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.[43] Again, the joint test across all covariate-by-treatment interaction terms fails to reject the null ($p = 0.260$ at one year, $p = 0.436$ at two and a half years).

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

[42] This analysis was not prespecified in the registered report.

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

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, both pooled across grant eligibility status and within each eligibility stratum. This differs from the fully interacted specification used in the main analysis, so the naïve (unbounded) estimates underlying the bounds may not align exactly with the previously reported estimates.[44] The findings in Table D3 suggest that the previously reported positive effects on IYCF knowledge and diet diversity can generally be bounded away from zero in both T2 and T3 in the pooled comparisons, though the T2 effect on dietary diversity at one year among grant-ineligible households cannot be bounded away from zero.[45]

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

[45] Given the limited differential attrition across arms (see columns 1 and 4 in Table D2), the trimming fractions used in constructing the Lee bounds are small, and the resulting bounds are correspondingly narrow.

Table D1: Household level attrition analysis

	Short-run			Medium-run		
	(1)	(2)	(3)	(4)	(5)	(6)
T2 – No grant	-0.004	-0.005	-0.023	0.030	0.029	-0.062
	(0.008)	(0.008)	(0.072)	(0.024)	(0.024)	(0.066)
T2 – Grant	-0.022	-0.022	0.084	0.062**	0.060**	0.305**
	(0.015)	(0.015)	(0.133)	(0.026)	(0.026)	(0.149)
T3 – No grant	-0.019**	-0.020***	-0.084	-0.001	-0.002	-0.039
	(0.007)	(0.007)	(0.058)	(0.020)	(0.019)	(0.059)
T3 – Grant	-0.020	-0.021	-0.031	0.007	0.005	-0.016
	(0.014)	(0.014)	(0.084)	(0.021)	(0.021)	(0.101)
Grant eligible	0.019	0.018	0.020	-0.002	0.006	-0.003
	(0.015)	(0.015)	(0.016)	(0.008)	(0.011)	(0.015)
Cons. per capita		-0.000	0.004		0.003	0.000
		(0.002)	(0.004)		(0.003)	(0.002)
× T2-NG			-0.007			0.014**
			(0.006)			(0.007)
× T2-G			-0.013			0.000
			(0.008)			(0.015)
× T3-NG			-0.004			-0.001
			(0.004)			(0.005)
× T3-G			-0.007			-0.012
			(0.005)			(0.007)
FIES score		0.000	0.002		0.004***	0.002
		(0.001)	(0.003)		(0.002)	(0.003)
× T2-NG			-0.006			0.004
			(0.005)			(0.004)
× T2-G			-0.000			0.001
			(0.005)			(0.009)
× T3-NG			-0.001			0.004
			(0.004)			(0.006)
× T3-G			-0.001			-0.004
			(0.004)			(0.005)
Asset index		0.002	0.004		0.009	-0.001
		(0.004)	(0.009)		(0.006)	(0.011)
× T2-NG			-0.001			0.029
			(0.011)			(0.023)
× T2-G			0.046			0.105**
			(0.029)			(0.047)
× T3-NG			-0.007			-0.003
			(0.009)			(0.015)

Table D1 continued

× T3-G		-0.030		-0.001
		(0.019)		(0.026)
IYCF knowledge	-0.003**	-0.004	0.001	-0.000
	(0.001)	(0.003)	(0.002)	(0.003)
× T2-NG		-0.001		0.005
		(0.004)		(0.004)
× T2-G		0.002		-0.006
		(0.005)		(0.007)
× T3-NG		0.003		0.001
		(0.003)		(0.004)
× T3-G		0.003		0.005
		(0.004)		(0.006)
Household size	-0.001	0.002	-0.002	0.003
	(0.003)	(0.005)	(0.003)	(0.005)
× T2-NG		-0.008		-0.012
		(0.006)		(0.009)
× T2-G		-0.007		-0.024**
		(0.010)		(0.012)
× T3-NG		-0.001		0.001
		(0.006)		(0.006)
× T3-G		-0.008		-0.005
		(0.008)		(0.007)
Age of the primary female	-0.002*	-0.003*	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)
× T2-NG		0.004*		0.002
		(0.003)		(0.002)
× T2-G		-0.000		0.001
		(0.003)		(0.003)
× T3-NG		0.002		-0.000
		(0.002)		(0.002)
× T3-G		0.001		0.001
		(0.004)		(0.002)
Primary female has 1 to 3 years of education	-0.018***	-0.020	-0.011	-0.004
	(0.007)	(0.013)	(0.010)	(0.016)
× T2-NG		0.009		-0.001
		(0.018)		(0.025)
× T2-G		-0.029		-0.049
		(0.026)		(0.041)
× T3-NG		0.021		0.012
		(0.015)		(0.023)
× T3-G		-0.013		-0.029

Table D1 continued

			(0.021)			(0.029)
Primary female has 4 to 7 years of education		-0.007	-0.019		-0.007	-0.002
		(0.008)	(0.017)		(0.009)	(0.013)
× T2-NG			0.029			-0.016
			(0.024)			(0.026)
× T2-G			0.007			-0.037
			(0.031)			(0.039)
× T3-NG			0.024			0.015
			(0.019)			(0.019)
× T3-G			-0.001			0.017
			(0.026)			(0.033)
Primary female has 8 or more years of education		0.005	0.016		-0.013	-0.004
		(0.012)	(0.025)		(0.013)	(0.019)
× T2-NG			-0.010			-0.033
			(0.038)			(0.039)
× T2-G			-0.023			-0.091**
			(0.031)			(0.038)
× T3-NG			-0.015			0.024
			(0.026)			(0.033)
× T3-G			-0.034			-0.010
			(0.037)			(0.030)
Enrolled on the basis of a pregnant woman		-0.007	-0.006		-0.004	-0.011
		(0.006)	(0.010)		(0.005)	(0.009)
× T2-NG			-0.013			-0.000
			(0.017)			(0.016)
× T2-G			0.004			-0.006
			(0.020)			(0.028)
× T3-NG			0.008			0.005
			(0.011)			(0.016)
× T3-G			-0.009			0.060***
			(0.030)			(0.022)
Constant	0.024***	0.107***	0.117**	0.024**	0.018	0.040
	(0.006)	(0.028)	(0.056)	(0.011)	(0.035)	(0.032)
Observations	3015	3015	3015	3015	3015	3015
Joint p-value (4 cells)	0.006	0.005	0.303	0.121	0.145	0.139
p-value: T2 = T3	0.355	0.321	.	0.082	0.087	.
Joint F-test interactions p	.	.	0.401	.	.	0.242

Notes: Dependent variable is a binary indicator for attrition. Columns 1–3: one year follow-up. Columns 4–6: 2.5 year follow-up. All regressions include strata fixed effects and use the 4-way treatment-by-grant-eligibility specification. All independent variables are measured at baseline. Standard errors clustered at the kebele level in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table D2: Child level attrition analysis

	Short-run			Medium-run		
	(1)	(2)	(3)	(4)	(5)	(6)
T2 – No grant	-0.024	-0.017	0.077	0.018	0.022	0.104
	(0.019)	(0.019)	(0.149)	(0.027)	(0.026)	(0.168)
T2 – Grant	0.006	-0.018	-0.403	0.032	0.007	-0.110
	(0.033)	(0.030)	(0.253)	(0.038)	(0.035)	(0.266)
T3 – No grant	-0.036*	-0.039**	-0.044	-0.038	-0.042*	-0.050
	(0.019)	(0.018)	(0.140)	(0.025)	(0.025)	(0.159)
T3 – Grant	-0.016	-0.009	0.188	-0.031	-0.026	0.106
	(0.031)	(0.029)	(0.197)	(0.034)	(0.032)	(0.205)
Grant eligible	0.010	0.008	-0.001	0.017	0.026	0.017
	(0.028)	(0.029)	(0.033)	(0.027)	(0.028)	(0.032)
Cons. per capita		0.003	0.010		0.003	0.009
		(0.005)	(0.008)		(0.005)	(0.009)
× T2-NG			-0.020			-0.013
			(0.013)			(0.015)
× T2-G			0.016			-0.003
			(0.019)			(0.021)
× T3-NG			-0.004			0.000
			(0.011)			(0.013)
× T3-G			-0.023			-0.023
			(0.015)			(0.015)
FIES score		0.007**	0.011**		0.011***	0.011*
		(0.003)	(0.005)		(0.004)	(0.006)
× T2-NG			-0.000			0.004
			(0.009)			(0.011)
× T2-G			0.011			0.010
			(0.012)			(0.014)
× T3-NG			-0.011			-0.002
			(0.007)			(0.010)
× T3-G			-0.015			-0.013
			(0.012)			(0.014)
Asset index		-0.008	-0.017		0.005	-0.008
		(0.009)	(0.014)		(0.011)	(0.018)
× T2-NG			0.018			0.040
			(0.022)			(0.031)
× T2-G			0.031			0.064
			(0.062)			(0.069)
× T3-NG			0.013			0.001
			(0.021)			(0.026)

Table D2 continued

× T3-G		-0.013		-0.031
		(0.046)		(0.047)
IYCF knowledge	0.003	0.001	0.003	0.002
	(0.003)	(0.006)	(0.004)	(0.006)
× T2-NG		-0.007		-0.005
		(0.008)		(0.009)
× T2-G		0.024**		0.015
		(0.010)		(0.012)
× T3-NG		0.007		0.003
		(0.008)		(0.009)
× T3-G		-0.009		-0.002
		(0.012)		(0.013)
Household size	-0.003	0.002	-0.004	0.005
	(0.005)	(0.008)	(0.005)	(0.009)
× T2-NG		-0.022*		-0.038***
		(0.012)		(0.014)
× T2-G		-0.007		-0.013
		(0.017)		(0.018)
× T3-NG		0.008		0.009
		(0.013)		(0.014)
× T3-G		-0.012		-0.020
		(0.014)		(0.015)
Age of the primary female	0.006***	0.005*	0.005***	0.004
	(0.002)	(0.003)	(0.002)	(0.003)
× T2-NG		0.005		0.006
		(0.004)		(0.004)
× T2-G		0.004		0.001
		(0.006)		(0.006)
× T3-NG		-0.001		-0.003
		(0.004)		(0.005)
× T3-G		0.003		0.003
		(0.006)		(0.006)
Primary female has 1 to 3 years of education	-0.018	-0.027	-0.034*	-0.049
	(0.017)	(0.029)	(0.020)	(0.034)
× T2-NG		0.011		0.012
		(0.048)		(0.053)
× T2-G		-0.022		-0.020
		(0.072)		(0.074)
× T3-NG		0.029		0.054
		(0.045)		(0.052)
× T3-G		0.020		0.035

Table D2 continued

			(0.059)			(0.064)
Primary female has 4 to 7 years of education		0.019	-0.024		0.025	-0.025
		(0.018)	(0.034)		(0.021)	(0.035)
× T2-NG			0.085*			0.073
			(0.049)			(0.056)
× T2-G			0.086			0.098
			(0.071)			(0.079)
× T3-NG			0.059			0.078
			(0.048)			(0.054)
× T3-G			0.021			0.065
			(0.062)			(0.069)
Primary female has 8 or more years of education		-0.011	-0.037		-0.026	-0.039
		(0.023)	(0.042)		(0.025)	(0.044)
× T2-NG			0.021			-0.039
			(0.063)			(0.073)
× T2-G			0.042			0.001
			(0.079)			(0.084)
× T3-NG			0.069			0.083
			(0.061)			(0.068)
× T3-G			0.006			-0.014
			(0.071)			(0.074)
Enrolled on the basis of a pregnant woman		0.308***	0.312***		0.269***	0.277***
		(0.016)	(0.028)		(0.017)	(0.030)
× T2-NG			-0.014			-0.039
			(0.044)			(0.049)
× T2-G			0.005			0.006
			(0.049)			(0.055)
× T3-NG			-0.013			-0.019
			(0.042)			(0.047)
× T3-G			-0.008			0.032
			(0.054)			(0.055)
Constant	0.202***	-0.199***	-0.213**	0.237***	-0.138**	-0.165
	(0.014)	(0.061)	(0.096)	(0.015)	(0.068)	(0.108)
Observations	3015	3015	3015	3015	3015	3015
Joint p-value (4 cells)	0.313	0.248	0.321	0.192	0.200	0.864
p-value: T2 = T3	0.319	0.696	.	0.020	0.047	.
Joint F-test interactions p	.	.	0.260	.	.	0.436

Notes: Dependent variable is a binary indicator for child-level attrition. Columns 1–3: one year follow-up. Columns 4–6: 2.5 year follow-up. All regressions include strata fixed effects and use the 4-way treatment-by-grant-eligibility specification. All independent variables are measured at baseline. Standard errors clustered at the kebele level in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table D3: Lee bounds on treatment effects

		Pooled		No grant		Grant	
		T2 vs T1	T3 vs T1	T2 vs T1	T3 vs T1	T2 vs T1	T3 vs T1
IYCF knowledge (one year)	Lower bound	0.294***	0.416***	0.304***	0.452***	0.183	0.345**
		(0.084)	(0.088)	(0.095)	(0.102)	(0.140)	(0.139)
	Upper bound	0.268***	0.393***	0.304***	0.430***	0.127	0.345**
		(0.080)	(0.086)	(0.095)	(0.101)	(0.134)	(0.139)
Diet diversity (one year)	Lower bound	0.125**	0.593***	0.057	0.595***	0.240**	0.582***
		(0.048)	(0.062)	(0.064)	(0.073)	(0.094)	(0.102)
	Upper bound	0.125**	0.593***	0.057	0.595***	0.240**	0.582***
		(0.048)	(0.062)	(0.064)	(0.073)	(0.094)	(0.102)
IYCF knowledge (2.5 years)	Lower bound	0.305***	0.301***	0.246**	0.234***	0.498***	0.460***
		(0.071)	(0.064)	(0.099)	(0.081)	(0.156)	(0.133)
	Upper bound	0.421***	0.195***	0.378***	0.117	0.417***	0.385***
		(0.067)	(0.063)	(0.088)	(0.084)	(0.142)	(0.122)
HAZ (2.5 years)	Lower bound	-0.064	0.043	-0.154**	-0.010	0.099	0.173*
		(0.062)	(0.057)	(0.075)	(0.065)	(0.103)	(0.099)
	Upper bound	-0.020	0.084	-0.118	0.030	0.179	0.243**
		(0.064)	(0.057)	(0.076)	(0.066)	(0.113)	(0.102)

Notes: Each column reports Lee (2009) bounds on the treatment effect for the indicated comparison. The bounds use the reweighting approach of McKenzie (2026) to handle ties. compares all T2 (or T3) households to all T1 households; and restrict the comparison to within each eligibility stratum. Standard errors clustered at the kebele level in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

E Cash+BCC literature

To unpack the magnitude of the reported effects on diet diversity and stunting, we summarize the existing literature on cash-plus-BCC interventions that report effects on these outcomes.[46] Table E1 summarizes the key relevant papers,[47] and highlights there is wide variation in these interventions. A typical BCC intervention involves monthly group sessions delivered over a two-year period, though other models vary from weekly (Ahmed et al., 2019) to only two sessions total (Bouguen and Dillon, 2026). Total cash transfer amounts also vary substantially, from roughly 163 $PPP in Levere et al. (2024) to about 2,088 $PPP in Ahmed et al. (2025), with the maternal cash transfers provided in T3 in this trial (1,248 $PPP) near the middle of the distribution and the T3 and livelihoods grant package (1,982 $PPP) among the largest.

Figures E1a and E1b summarize the estimated treatment effects on dietary diversity and child height-for-age Z-scores (HAZ) from this literature.[48] For dietary diversity, nearly all studies report positive effects with effect sizes between 10 and 20 percent (Figure E1a) 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).[49] 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 child diets, it is clear in Figures E1b that this set of interventions has generally struggled to shift child HAZ — a pattern consistent with the disconnect between dietary diversity and linear growth documented in the main text. The studies that do report significant HAZ effects share a common feature: either substantially larger total transfer values (Ahmed et al., 2025), or households accumulating productive livestock assets through direct transfers (Bouguen and Dillon, 2026), or essentially, both: households using large transfers to accumulate livestock assets and thus generate predictable cash flows. The T3-Grant estimate in this trial falls in the same range, and notably, it is the only subgroup in

[46] This analysis then necessarily largely excludes the previous literature on graduation models, since these papers do not include BCC and typically do not report children's nutritional outcomes.

[47] Little 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.

[48] One study summarized in Table E1, 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 null effects on height-for-age.

[49] This indicator is not reported in Bouguen and Dillon (2026).

our design that combines cash, BCC, and a livelihoods grant —- mirroring the intensity of the programs that produce the largest effects in the cross-study comparison.

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. For child anthropometry, however, intensity alone does not appear sufficient: the composition of the intervention, and specifically the inclusion of a productive asset component, also seems important.

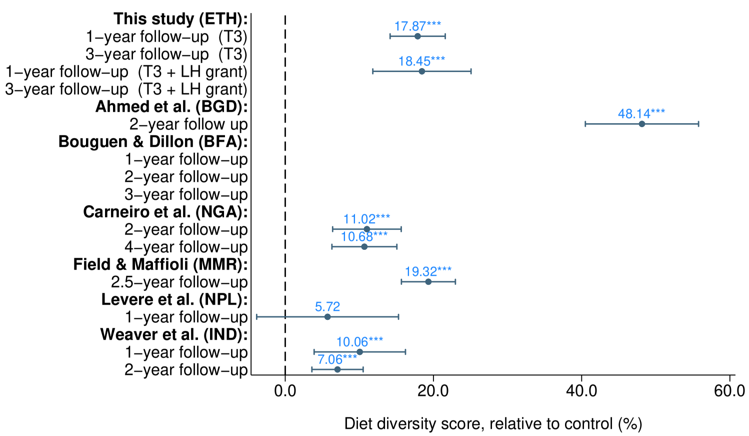
Table E1: Overview of the cash+BCC literature

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 nutrition 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 multisession training on IYCF, hygiene, health-seeking, and spending for mothers, 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)	1,248	2,088	743	788	912[a]	163	643	717
Livelihoods	VESAs; light-touch livelihoods and financial training; one-third received a $300 livelihoods grant (734 $PPP).	No	Livestock voucher redeemable for poultry or small ruminants; animals monitored for two weeks post-distribution.	No	No	No	No	No

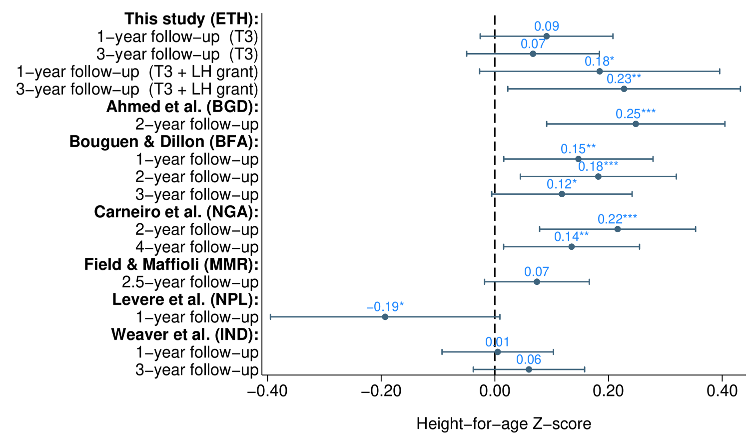
Notes: VESA = Village Economic and Social Association; BCC = Behaviour change communication; LH grant = Livelihoods grant. Several studies used multi-arm designs; we report estimates from the most expansive treatment arm relative to the control group. For Ahmed et al., this corresponds to the cash + BCC arm in North Bangladesh. For Bouguen & Dillon, the information 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 2021 purchasing power parity dollars ($PPP). For SPIR, the total cash and livelihoods grant values are drawn from the program's marginal cost accounting and are consistent with the figures reported in the cost-effectiveness and cost-benefit tables. [a] Myanmar's consumer price index is not reported by the World Bank for 2020 or 2021; the 2019 CPI is used as a proxy, which modestly understates inflation between 2019 and 2021.

Figure E1: Treatment effects relative to prior cash-plus-BCC evaluations

(a) Child dietary diversity



(b) Child height-for-age (HAZ)



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. In Panel (a), 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.

F Cost-effectiveness and cost-benefit analyses: Additional details

F.1 Cost-effectiveness analysis

This section provides additional details for the cost-effectiveness analysis reported in the main text. We restrict the cost-effectiveness analysis to the T3-Grant arm because it is the only SPIR treatment arm with a statistically significant effect on child HAZ. Given that the cost component of the cost-effectiveness ratio is known while the treatment effect on HAZ is estimated with sampling error, uncertainty in the ratio is driven entirely by the denominator. We construct 95 percent confidence intervals by CI inversion: the lower and upper bounds of the 95 percent CI on the HAZ treatment effect are plugged directly into the cost-effectiveness formula, with the bounds flipping because cost per 0.1 SD is inversely proportional to the treatment effect.

Cost is measured per household but HAZ effects are estimated at the child level. To express cost-effectiveness per beneficiary child, we divide cost per household by the average number of children in each study's estimation sample per enrolled household. SPIR and Carneiro et al. (2021) both use cohort designs that follow a single index per household, so the ratio equals one by construction. Ahmed et al. (2025)'s estimation sample is restricted to children exposed during their 1,000-day window, which in practice means the baseline index child aged 0–24 months plus a modest number of children born during the 24-month intervention; we take the ratio as approximately 1.10 based on the ratio of their North estimation sample size (2,218 children, from their Table 2) to the baseline sample (2,010 households). Bouguen and Dillon (2026)'s estimation sample includes all eligible children 0–5 years of age; their text reports approximately 4,000 eligible children in 3,500 eligible households, yielding a ratio of approximately 1.14. These corrections modestly reduce cost per 0.1 SD HAZ for Ahmed et al. and Bouguen & Dillon relative to a naive per-household calculation.

All costs are expressed in 2021 PPP dollars to match the monetary values reported elsewhere in the paper.

Only SPIR and Ahmed et al. (2025) have real program accounting data (reported in Ahmed et al. (2016)). Carneiro et al. (2021)'s cost figures are taken from their Table 9 Panel A, in which they explicitly assume that cash delivery and BCC delivery each cost 10 percent of the cash transfer value: a cash transfer of $536 per beneficiary in 2014 PPP dollars, $54 in assumed administrative costs, and $54 in assumed information/BCC costs. For Bouguen and Dillon (2026), no program accounting data are available. We use the component values reported in the program description section of the paper, but exclude the nutrition bundle (enriched flour, garden kit, and two training sessions) because no dollar values are reported. We then apply a 10 percent and 10 percent overhead markup on the cash component for BCC, delivery and administration, following Carneiro et al.

Because Carneiro et al. (2021)'s 10 percent + 10 percent overhead assumption is not anchored in measured costs, we present a second cost scenario that replaces their assumption with the overhead shares we observe in Ahmed et al. (2016)'s program accounting–29.0 percent of cash transfer value for cash delivery, and 27.9 percent for BCC delivery. These are nearly three times the Carneiro assumption. Columns (3a) and (3b) of Table 8 report Carneiro's cost-effectiveness ratio under the two assumptions. The difference between the two scenarios represents the sensitivity of published cost-effectiveness comparisons to unmeasured administrative and BCC delivery costs; comparing Carneiro and SPIR under the Ahmed-adjusted scenario provides a less optimistic but arguably more realistic picture of Carneiro's true cost-effectiveness. SPIR, Ahmed, and Bouguen and Dillon are each shown in a single column. We do not apply the same two-scenario treatment to Bouguen and Dillon because their BCC component is minimal (two training sessions rather than the weekly group sessions and home visits of Ahmed's design), and thus applying Ahmed et al.'s 27.9 percent BCC share to their cash transfer would not be a meaningful upper bound. The daggers on the BCC, delivery and administration cells in columns (3a), (3b), and (4) of Table 8 mark the cells where overhead costs are based on assumptions rather than measured program accounting.

Cost-benefit analysis

This section provides more details around the cost-benefit analysis described in the main body above. First, the analysis of benefits considers two pathways. The first operates through an immediate increase in household consumption: we treat consumption as the primary measure of program benefits at the household level.[50] As emphasized by Deaton (1997), both conceptual and practical considerations support the use of consumption as a preferred welfare indicator in low-income settings. Here, consumption captures not only the direct effects of transfers but also the returns to accumulated assets and savings (Bossuroy et al., 2022). In agrarian economies, where self-employment is widespread and income is often seasonal and derived from multiple sources, consumption is generally measured more reliably than income (Deaton, 1997; Carletto et al., 2022).

In addition, the assumptions regarding the time horizon and possible decay in treatment effects vary considerably in the earlier literature. Some assume that the observed effects ex-

[50] Earlier graduation evaluations also considered benefits associated with asset holdings and savings levels (Bandiera et al., 2017; Banerjee et al., 2015).

tend to perpetuity (Banerjee et al., 2015), while others assume a fixed time-horizon (Bandiera et al., 2017; Bedoya et al., 2019). Evidence from longer-term studies provides some guidance, but also suggests that persistence is context-dependent: intensive graduation programs yielded sustained effects in Bangladesh and India but fading effects in Ethiopia (Banerjee et al., 2021; Balboni et al., 2022; Barker et al., 2024). We therefore consider four scenarios: a 10-year and a 20-year time horizon, each without and with 10 percent annual decay. In the absence of decay, we assume benefits beyond the time frame observed in the trial persist at the year three level.[51]

Second, we consider the benefits arising from lifetime consumption gains due to improved child nutritional status. Chronic undernutrition in early childhood may affect adult economic outcomes through several distinct but related channels (Victora et al., 2008; Currie and Vogl, 2013; Hoddinott et al., 2013a). One channel operates via physical stature: growth faltering in childhood leads to shorter adult height, which is consistently associated with lower earnings across both low- and high-income (Persico et al., 2004; Lundborg et al., 2014; Vogl, 2014; King et al., 2026). A second channel reflects the neurological consequences of early nutritional deprivation. Undernutrition during critical developmental periods can impair cognitive development, contributing to reduced educational attainment and, consequently, weaker labor market outcomes (Cordero et al., 1993; Kar et al., 2008; Hoddinott et al., 2013b; Alderman et al., 2006). A third mechanism arises from the long-term health effects of early stunting. Early-life undernutrition is linked to a higher risk of chronic diseases in adulthood, which may diminish labor productivity and work capacity (see Hoddinott et al. (2013a)).

As noted above, to capture the future benefits associated with improved nutritional status, we focus on the HAZ ITT estimated in the medium-run follow-up when the children are 24–48 months of age. Building on Carneiro et al. (2021), we apply the long-run HAZ–consumption estimates provided by Hoddinott et al. (2013b).[52] We apply this semi-elasticity to the mean annual per capita consumption in the control group at baseline (998.5 $PPP) to obtain the annual consumption gain per child. [53] We then sum these gains over the adult life cycle (ages 18–65), applying age-specific survival probabilities (UN, 2024) and discounting back

[51] Detailed household level benefit calculations are presented in Table F1 in the appendix.

[52] We use the IV estimate of the effect of HAZ on log per capita consumption from Hoddinott et al. (2013b), an estimate that is statistically significant. Carneiro et al. (2021) use the earnings semi-elasticities from the same source (4% for men, 9% for women per SD), which is appropriate given that their study collects individual earnings data. In our setting, individual earnings are not available in the SPIR surveys nor in large-scale Ethiopian household surveys, as such data are difficult to measure reliably in agrarian economies where self-employment predominates (Deaton, 1997). The consumption semi-elasticity is therefore a natural choice, consistent with our use of household consumption as the primary welfare measure throughout the analysis.

[53] The SPIR evaluation surveys and measurements focused on following a single index child per household; we do not estimate within-household spillovers.

to year zero at the same five percent discount rate used for the first pathway.

Table F1: SPIR cost-benefit analysis: detailed calculations

	T2-NG†	T2-G†	T3-NG	T3-G
Panel A. Consumption ITTs and annualized benefits				
Per capita daily ITT, Year 1	-0.012	0.056	0.140**	0.276***
Per capita daily ITT, Year 3	0.043	0.030	0.206***	0.304***
Annual HH benefit, Year 1	-23.7	110.4	275.9	544.0
Annual HH benefit, Year 2 (interp.)	30.6	84.8	341.0	571.6
Annual HH benefit, Year 3	84.8	59.1	406.0	599.2
Panel B. Present value of consumption benefits (at 5%)				
PV of observed years (1–3)	78.4	233.1	922.8	1554.1
PV of projected years 4–10, no decay	423.6	295.6	2029.5	2995.0
PV of projected years 4–10, 10% decay	290.0	202.3	1389.1	2049.9
PV of projected years 4–20, no decay	825.4	575.9	3954.3	5835.4
PV of projected years 4–20, 10% decay	407.3	284.2	1951.3	2879.6
A: Total (10yr, no decay)	502.0	528.6	2952.3	4549.2
B: Total (10yr, 10% decay)	368.4	435.4	2311.9	3604.1
C: Total (20yr, no decay)	903.8	808.9	4877.1	7389.6
D: Total (20yr, 10% decay)	485.7	517.2	2874.1	4433.8
Panel C. Break-even annual decay rate (%)				
10-year horizon	24.1	Never	33.8	40.0
20-year horizon	26.4	Never	34.8	40.5

Notes: Discount rate: 5%. Annual HH benefit = daily per capita ITT × 5.4 × 365. Year 2 linearly interpolated between one-year and 2.5-year. Break-even decay is the annual rate at which consumption-only BCR = 1. indicates BCR < 1 even with 0% decay. † Consumption ITTs not statistically significant. *** p<0.01, ** p<0.05, * p<0.10.