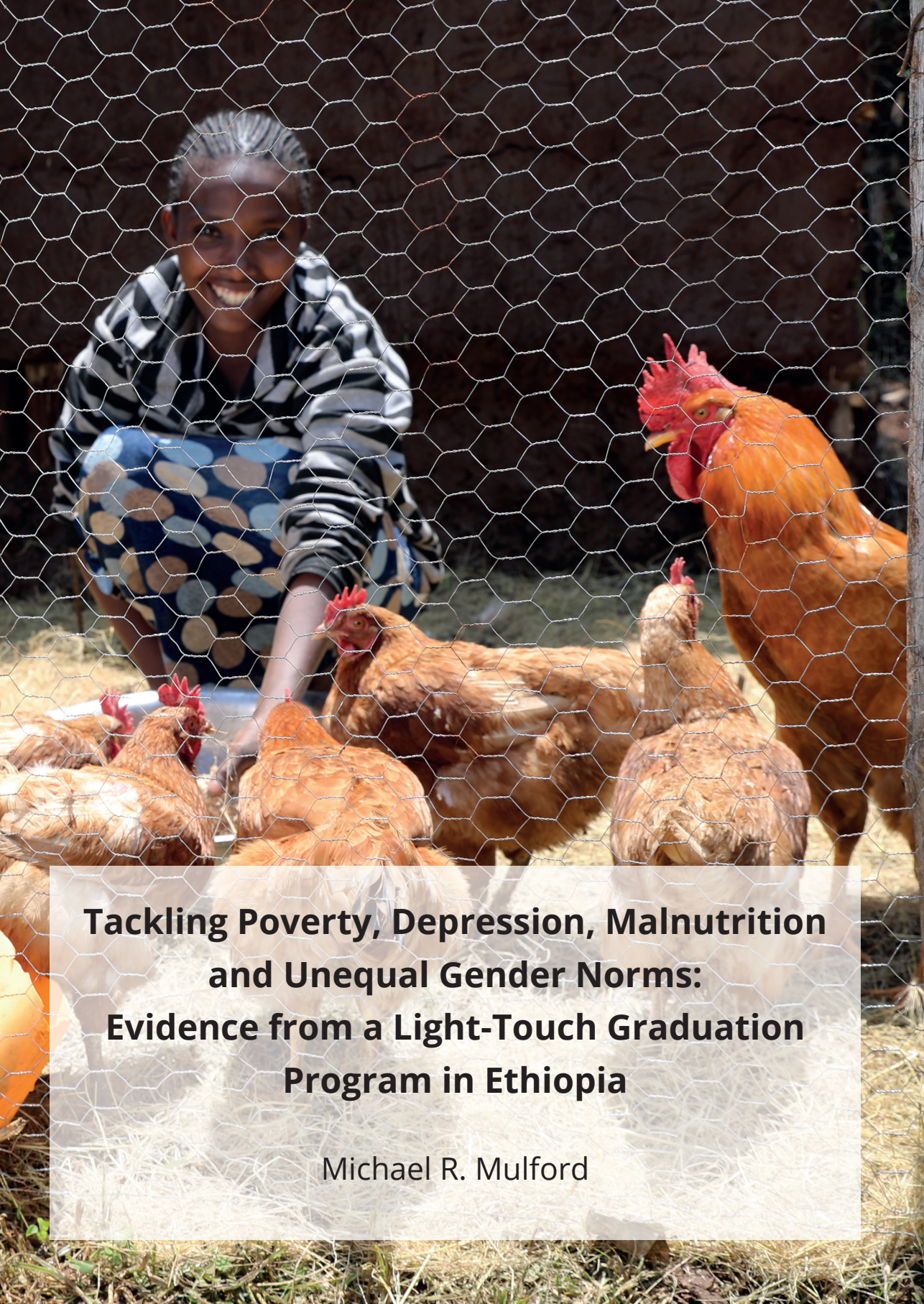


Tackling Poverty, Depression, Malnutrition and Unequal Gender Norms: Evidence from a Light-Touch Graduation Program in Ethiopia

Michael R. Mulford

Propositions

1. Graduation programs must address economic constraints to fully realize benefits of complementary mental health and nutrition interventions.
   (this thesis)

2. In graduation programs, cash transfers are as effective as the more common in-kind asset transfers.
   (this thesis)

3. Industrialized agriculture undermines the resilience of Earth's natural systems.

4. Adopting technologies such as hydroponics and vertical farming is critical in meeting growing food demands sustainably.

5. Urban climate vulnerability is predominantly determined by socio-economic factors.

6. Increasing urban density is critical to effective climate change mitigation and adaptation.

Propositions belonging to the thesis, entitled

How to Tackle Poverty, Depression, Malnutrition and Unequal Gender Norms: Evidence from a Light-Touch Graduation Program in Ethiopia

Michael Mulford
Wageningen, 25 June 2024

Tackling Poverty, Depression, Malnutrition and Unequal Gender Norms: Evidence from a Light-Touch Graduation Program in Ethiopia

Michael R. Mulford

Thesis committee

Promotor
Prof. Dr Erwin Bulte
Professor of Development Economics
Wageningen University & Research

Co-promotor
Dr Daniel Gilligan
Director of the Poverty, Gender, and Inclusion Unit
International Food Policy Research Institute (IFPRI), Washington, D.C., USA

Other members
Prof. Dr William Parienté, Université Catholique de Louvain, Belgium
Dr Patricio Dalton, Tilburg University
Dr Andrew Simons, Fordham University, USA
Dr Koos Gardebroek, Wageningen University & Research

This research was conducted under the auspices of Wageningen School of Social Sciences (WASS).

Tackling Poverty, Depression, Malnutrition and Unequal Gender Norms: Evidence from a Light-Touch Graduation Program in Ethiopia

Michael R. Mulford

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 25 June 2024
at 11 a.m. in the Omnia Auditorium.

Michael R. Mulford
Tackling Poverty, Depression, Malnutrition and Unequal Gender Norms:
Evidence from a Light-Touch Graduation Program in Ethiopia
206 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English

DOI: 10.18174/657866

Summary

Among the many ongoing efforts to alleviate poverty, one view is that multiple barriers trap households in poverty and a 'big push' is needed to facilitate their exit from these poverty traps. Specifically, a multifaceted approach that combines one-time livelihood transfers with targeted training and consumption support may be necessary to help poor households withstand shocks, and sustainably improve their lives and livelihoods. These are core elements and rationale for what have come to be known as graduation programs. The seminal study of graduation programming was a multi-country evaluation that found significant effects across several families of outcomes that persisted two years after the end of the interventions (Banerjee et al. 2015).

While many organizations and governments are now implementing similar programs, the underlying mechanisms through which these programs work are not yet well understood. Facing their own budget constraints, policy makers are especially interested to know the minimum investment necessary to achieve similar effects when implementing these programs at larger scale. For example, could some of the high-cost livelihoods components be reduced without jeopardizing the overall program goals? Conversely, rather than reducing components, perhaps there are missing elements such as those addressing mental health, that could enhance people's ability to benefit from these programs. While many of these programs have had a significant impact on livelihoods outcomes, the results for health and women's empowerment outcomes have been more mixed. It is possible that including additional complementary interventions could enable graduation programs to achieve a more comprehensive impact for these ultra poor households.

These are among the questions that motivated the study presented here. The principal aim of this research was to unpack and assess the comparative effectiveness of different program elements and their combinations as incorporated in a light-touch graduation program implemented at scale in Ethiopia. This program supported and built upon the government's Productive Safety Net Program (PSNP), targeting the poorest households in the most chronically food insecure areas in rural Ethiopia. Employing a cluster randomized controlled trial (RCT), we assessed the effectiveness of three distinct combinations of basic and enhanced livelihoods and nutrition program elements. These included the novel inclusion of group psychotherapy for treating depression, intensive nutrition behaviour change communication (BCC) targeting child malnutrition, and men's engagement groups designed to foster gender equitable attitudes and practices among men. Ours is also the first study that we are aware of to experimentally vary the modality of the livelihoods transfer, comparing the more common approach of an in-kind asset to an equivalent value cash transfer.

In Chapter 2, we address the question: can a lighter-touch version of graduation programming still impact household economic outcomes and facilitate pathways out of poverty? Notably, the one-time livelihood transfers are considerably smaller in our study compared with other graduation programs and are targeted only to the poorest PSNP households. Another characteristic of this program's light-touch approach is its use of group-based support, which contrasts with the more intensive, household-level visits typical of other graduation programs. We observed significant treatment effects on savings across all households, irrespective of whether they received livelihoods transfers or not. This increased savings appears to serve as a primary mechanism

through which these households mitigated the adverse effects of drought, smoothed their consumption and partially protected livestock holdings (Hirvonen et al. 2023). Modest effects on income and asset outcomes were observed among extremely poor households that received livelihood transfers. These effects were similar across both cash and in-kind transfer modalities. Neither the basic nor enhanced livelihood interventions yielded any effects on consumption at endline. Thus, while households receiving livelihood transfers were able to augment their livestock assets and income, the scale of these improvements was insufficient to bring about a transformative impact in their lives and livelihoods.

In Chapter 3, we find that neither enhanced livelihoods programming including transfers, nor the enhanced nutrition programming with group psychotherapy independently reduced depression for women with depressive symptoms. When these interventions were combined, they had a substantial effect on the prevalence of depression as well as on psychosocial, economic and wellbeing outcomes. Notably, these results were observed more than one year after the completion of the group psychotherapy, underscoring the importance of alleviating both financial and mental health constraints to both achieve and sustain these effects over this timeframe. In Chapter 4, we explore the question of whether a light-touch graduation program with either basic or enhanced nutrition programming impacts child nutrition and growth. Perhaps unsurprisingly, we observe a pattern of effects similar to those of Chapter 3, where only in extremely poor households that received cash transfers and exposure to enhanced nutrition programming do we find significant effects on child growth. In Chapter 5, we observed that programming exposure led to improved gender equitable attitudes and actions by men in all treatment arms at the midline. However, by the endline, sustained effects were only observed in treatment arms which included men's engagement groups.

When comparing these results to the findings from other studies, it is worth noting that households in this program experienced numerous shocks alongside disruptions caused by COVID-19 during the period preceding the endline. These included crop losses due to pest infestations, drought, and floods. The conflict that began in 2020 in the northern part of the country had not yet directly impacted the majority of the Amhara region program areas by the endline, but social unrest and sporadic violence occurred across the Oromia region during the entire program period. Considering the high vulnerability of most PSNP households, it is likely that some of the modest gains in livelihoods and productive assets were eroded as households sought to cope with these concurrent shocks.

In conclusion, the findings from this light-touch graduation study indicate that consistent and sustained effects on overall livelihood outcomes are unlikely to be attained without the one-time livelihood transfer. There could be administrative cost savings in shifting from in-kind to cash transfers (without affecting intended outcomes), but smaller value transfers may fall short in catalyzing productive livelihoods that support a pathway out of poverty. Including targeted mental health or intensive nutrition interventions can support improvement across a range of additional health and well-being outcomes, but only when combined with livelihoods programming that includes livelihood transfers. As policymakers consider how to scale these programs across different settings, this research emphasizes the importance of retaining core livelihoods

components while also suggesting other innovations that can contribute to a broader set of holistic outcomes.

Contents

Chapter 1: Introduction

1. Introduction Outline

In this introductory chapter, I begin by defining the problem (Section 2) that motivated this research and identifying the evidence gaps that the core chapters of this thesis will address. Following this, Section 3 provides a brief description of the context of this research, with a particular emphasis on the individual topics covered in each of the following chapters. In Section 4, a background of the government of Ethiopia's Productive Saftey Net Program (PSNP) is provided, along with detailed descriptions of key components of the Strengthen PSNP Institutions and Resilience (SPIR) program. The overall research objective and specific research questions are delineated in Section 5. Section 6 provides a short overview of the study design and Section 7 describes the program implementation and study rollout, highlighting challenges that arose due to COVID-19 restrictions and sporadic insecurity. Finally, a preview of the main findings are provided in Section 8. In each of the following sections, the content corresponds and maps to the topics of the four core chapters in this thesis, namely livelihoods (Chapter 2), mental health (Chapter 3), nutrition (Chapter 4), and women's empowerment (Chapter 5) outcomes from a light-touch graduation program implemented at scale in Ethiopia.

2. Problem Definition

Over the past few decades, the number of people living in extreme poverty has fallen dramatically, with the World Bank estimating a decline from 2 billion living on less than $2.15 per day in 1990, to 650 million in 2019.[1] Hidden within these global gains, the incidence of extreme poverty has persisted and is projected to grow in sub-Saharan Africa.[2] Among the many ongoing efforts to alleviate poverty, some advocate an incremental approach of testing and refining targeted interventions that address specific constraints faced by impoverished households (Duflo and Banerjee 2011). Another view is that multiple, often self-reinforcing barriers may trap households in poverty (Barrett and Carter 2013; Barrett, Garg, and McBride 2016), and a 'big push' is necessary to facilitate their exit from these poverty traps.

In its simplest version, this push could take the form of a large unconditional cash transfer to poor households. There is a large and growing number of studies using randomized control trials (RCTs) to test the effects of these one-time transfers on a wide range of outcomes. One example that is illustrative of the intertemporal pattern of effects is a study conducted in Kenya that found initial positive results across several outcomes for households that received $700 (Haushofer and Shapiro 2016). Three years later, only the effects on assets persisted when recipient households were compared to control households in other villages (Haushofer and Shapiro 2018). Other post-intervention studies of one-time cash, or a slight variant of cash-plus-training programs, also find initial positive effects dissipate over time (Blattman, Dercon, and Franklin 2019; Blattman, Fiala,

[1] Source: Poverty - Our World in Data

[2] 62% of those living in extreme poverty are estimated to live in SSA in 2022, and this is projected to rise to 90% by 2030, with absolute numbers increasing from 433 million in 2018 to a projected 443-477 million in 2030.

and Martinez 2018; Brudevold-Newman et al. 2017; Karlan, Knight, and Udry 2015). Blattman, Dercon, and Franklin (2019) speculate that targeting just one or two market failures may be insufficient to overcome the powerful forces that eventually drag poor households back into low equilibriums. A larger, multifaceted approach that combines one-time transfers with extended training and consumption support may be necessary to help poor households overcome multiple barriers, withstand shocks, and sustainably exit out of poverty. This is the rationale and core elements of what have come to be known as 'graduation' programs. The foundational study of graduation programming was a multi-country evaluation that found significant effects across several families of outcomes that persisted two years after the end of the interventions (Banerjee et al. 2015). Longer term follow-up evaluations found that effects persisted and even grew 7 years later in Bangladesh and 10 years later in India (Bandiera et al. 2017; Banerjee, Duflo, and Sharma 2021).

While many organizations and governments are now implementing similar programs, the underlying mechanisms through which these programs work are not yet well understood. Facing their own budget constraints, policy makers are especially interested to know the minimum investment necessary to achieve similar effects when implementing these programs at a larger scale.[3] The costs of graduation programs are often significant, with the one-time transfers alone valued between $437 and $1228 PPP per household (Banerjee et al. 2015). Could a smaller transfer, combined with the other program components still be effective in improving **livelihoods** outcomes? In Bangladesh, evidence suggested that households needed to move above a $500 asset[4] threshold to sustainably 'escape' poverty in that rural setting (Balboni et al. 2022). Consistent with the poverty-trap hypothesis, households that were just below this threshold fell back to a lower equilibrium over time. Perhaps even more difficult to scale, graduation programs typically include high frequency home visits for technical training and coaching over a two-year period. Beyond the challenge of managing a large team of qualified field agents, the cost alone to support these household visits was on average twice as large as the transfer value (Banerjee et al. 2015). What if this or other components were lessened or removed completely? In one of the countries in the multi-country study, Ghana, an additional experiment tested whether a less expensive asset-only, or a savings-only treatment could achieve results compared to the full graduation program (Banerjee et al. 2022). Neither of these lower-cost individual treatments produced lasting effects by themselves.

In the same paper, the authors speculated that rather than trying to simplify or reduce the number of program components, perhaps there are missing elements that could prevent people benefiting from these programs. For example, they suggest that a **mental health** intervention may support even larger take-up of other program components (Banerjee et al. 2022). Growing evidence points to a bi-directional relationship between common mental health illnesses and poverty that could lead to mental health-based poverty traps among individuals in extremely poor households (Ridley

[3] Apart from BRAC's Ultra-Poor Graduation Program in Bangladesh, the majority of graduation programs studied have been implemented at a relatively small scale (often less than 1,000 households).
[4] Based on small differences in initial asset holdings, the transfer in this program (91% of women selected to receive a cow) placed a majority of participants in a distribution right around this inflection point.

et al. 2020; Haushofer and Fehr 2014).[5] In addition, targeted support for maternal mental health, especially during the perinatal period, has the potential to improve both mother's wellbeing as well as the health and nutrition of their young children during the critical period for growth and development (Black et al. 2009; Surkan et al. 2011).

While graduation program evaluations report household-level food consumption effects, they typically have not included individual, especially child **nutrition** outcomes (Bandiera et al. 2017; Banerjee et al. 2015). Healthy diets, including consumption of animal sourced proteins and micro-nutrient rich foods, are critical for children's growth and cognitive development. This is especially important in avoiding growth faltering in the 6–23-month age window when stunting rates can rise dramatically (Victora et al. 2010). Graduation programs could be described as layering livelihood programming (asset transfer, training, high-frequency household visits and savings support) on top of elements common to social safety net programs (consumption support, provision of health information and linkages to related services). Studies of the latter have found that social assistance cash transfers can improve child nutrition, but effects are often small in magnitude (Manley, Alderman, and Gentilini 2022; Olney et al. 2022). Another channel for improving child nutrition could be through increased availability or access to sufficient and nutritious food resulting from the livelihood programming components. In a related literature, increased attention on making agricultural production interventions 'nutrition sensitive' has not yet uncovered how to effectively address these multiple objectives and increase children's dietary diversity (Ruel, Quisumbing, and Balagamwala 2018; Rosenberg et al. 2018; Sibhatu and Qaim 2018).

Lastly, the one set of outcomes without any persistent effects in the original multi-country graduation study was **women's empowerment** (Banerjee et al. 2015). Small effects observed at the first endline faded by the second endline, one year later. Sustainably addressing restrictive gender norms could both improve outcomes for women and unlock potential for improved household-level coordination in addressing various constraints targeted by these programs. Asset and skills transfers (Bandiera et al. 2017) or nutrition BCC messages are often targeted directly to women, but there is increasing recognition that engaging men is important to achieve gender equity outcomes (Glinski et al. 2018). This is not commonly incorporated in graduation program designs. One exception is a study in DRC that found overall impacts of a women-focused graduation program, but no differential effects from an additional men's engagement intervention (Angelucci, Heath, and Noble 2023).

3. Ethiopia context

The research presented here was conducted in Ethiopia, where 27 percent of the population live below the international **poverty** line ($2.15/day).[6] Recent economic growth translated to improved living standards for urban households but largely left rural populations behind[7] (World Bank 2019).

[5] In a recent paper, Haushofer and Salicath (2023) conclude that the product of the two elasticities of average income effects on depression and depression effects on economic outcomes is likely not large enough to cause a poverty trap in most settings.

[6] Poverty - Our World in Data

[7] Urban households experienced a 6 percent annual increase of consumption expenditures compared with 1% for rural households between 2011 and 2016.

The poorest households in rural Ethiopia have been excluded from the benefits of national economic growth, with most experiencing a stagnant or decreasing rate of consumption over the same period. In rural Ethiopia, a majority work in small-scale agriculture production,[8] with very few households engaged in non-farm activities (6 percent) or non-agricultural wage labor (2 percent)(Davis, Di Giuseppe, and Zezza 2017). The poorest are subsistence farmers who struggle to produce enough for their own consumption. Low productivity of farming systems is due to numerous factors, including degraded soil fertility, small plot sizes and susceptibility to increasingly erratic weather patterns and climate events such as droughts or floods.

Representative epidemiological data on **mental illness** does not exist in Ethiopia, but across several studies prevalence rates for depression, for example, are around 20 percent (Kassa and Abajobir 2018). As in other similar settings, there is a large gap between mental health needs and available treatment options. According to the Ethiopia Ministry of Health, in 2019 there were only 111 practicing psychiatrists serving a population of 112 million people. Given the low awareness of and access to these services, it is unlikely that individuals in poor rural households receive any support in diagnosing and treating common mental health illnesses.

Prevalence of stunting,[9] a common measure of chronic **malnutrition**, among young children in Ethiopia declined from 51 percent in 2000 to 37 percent in 2019 (EPHI 2019). While this overall downward trend is encouraging, children in rural areas are still nearly twice as likely to be stunted than those in urban centers (40 percent vs. 26 percent). These high rates of malnutrition are in part driven by very low dietary diversity and meal frequency, with only 11 percent of children consuming a minimum acceptable diet according to the WHO standard. These recommended diets are largely unaffordable for poor households (Hirvonen et al. 2020) especially with increasing prices of animal sourced foods and fruits and vegetables (Bachewe, Minten, and Yimer 2017).[10] Common religious observations are another factor. Among Orthodox Christians, there is a strong adherence to fasting practices that include qualified prohibitions on animal sourced food consumption for half of the days in a year. In Muslim communities, fasting practices include abstaining from food and drink between dawn and dusk during the month of Ramadan.

Deeply **entrenched gender norms** allocate the majority of household tasks to women, especially in rural Ethiopia. This includes childcare, food preparation, water and firewood collection and other household chores. Women also contribute significantly to agricultural production and other economic activities. Nonfarm activities tend to be small, informal enterprises with low profit margins and survival rates (Nagler and Naudé 2017), and women-managed enterprises in these rural areas have much lower labor productivity than male-owned enterprises (Essers, Megersa, and Sanfilippo 2019; Rijkers, Söderbom, and Loening 2010).[11] In terms of educational attainment, 40 percent of women in Ethiopia have never attended school, and only 6 percent completed a primary education. Only 48 percent of women are literate, compared to more than two thirds of men, and

[8] 83% of rural households work in own-farm agriculture.
[9] Measured by a child's height-for-age z-score as below two standard deviations from a reference population median.
[10] Increases of 32 to 36 percent in prices for milk, eggs, and meat; and 19 to 62 percent in prices of fruits and vegetables between 2007-2016.
[11] There is a total factor productivity difference of 12 percent between female and male-owned firms.

this drops to 39 percent for women in rural Ethiopia (Ethiopian Public Health Institute (EPHI) [Ethiopia] and ICF 2021). One third of married women have experienced spousal violence and one third report that their husbands insist on knowing where they are at all times (Central Statistical Agency (CSA) [Ethiopia] and ICF 2016).

4. Program Description

4.1 Productive Safety Net Program (PSNP)

In response to high levels of vulnerability to frequent droughts and other shocks experienced by rural populations in food insecure areas of the country, the Government of Ethiopia and other development partners launched the Productive Safety Net Program (PSNP) in 2005. It is one of the largest and longest running safety net programs on the continent with nearly eight million individuals benefitting from food or cash transfers as payment for public works labor or direct support for households unable to contribute labor.[12] The public work activities occur in the agricultural off-season when rural workers are generally underemployed (McCullough 2017). The goals of the PSNP are to provide reliable, predictable transfers to extreme poor households[13] to prevent asset depletion and negative coping strategies in the face of shocks, and to create community-level assets through public works. Ethiopia has a long history of droughts that have led to conditions of famine and loss of lives. In addition to humanitarian assistance, the PSNP is judged to have effectively supported vulnerable households to withstand the effects of the 2015-16 drought and prevented loss of life due to hunger or related causes (De Waal 2017; Sabates-Wheeler et al. 2022). The PSNP has also reduced the months of food insecurity in other years (Berhane et al. 2014) and supported landscape-level investments such as improved tree cover (Hirvonen et al. 2022).

4.2 Strengthen PSNP Institutions and Resilience (SPIR)

Strengthen PSNP Institutions and Resilience (SPIR) was a five-year, USAID Bureau for Humanitarian Assistance (BHA) funded, program (2016-2021) that supported PSNP implementation in 13 districts in Ethiopia's Amhara and Oromia regions. Implemented by World Vision, CARE and ORDA, SPIR benefitted 500,000 PSNP clients and layered integrated livelihood and nutrition activities onto core PSNP components. In terms of annual transfers, all PSNP public work clients received six months of transfers, while permanent direct support beneficiaries received twelve months of transfers. The monthly transfers were a mix of either cash or food depending on various factors including remoteness and access to local markets. Monthly household food transfers were calculated based on an individual ration composed of 15 kg of wheat, 1.5 kg of pulses (yellow split peas), and 0.45 kg of fortified vegetable oil. In terms of the nutrient value for a family of four, this ration provided approximately 91 percent of household energy requirements, 133 percent of protein requirements, and 62 percent of fat requirements.[14] As those eligible for the PSNP typically exceed the designated

[12] Approximately 20 percent of the PSNP caseload is comprised of these households who are unable to supply the intensive manual labor due to different factors such as disability or old age.
[13] Typically, the poorest 10-15 percent of households in these remote, food insecure districts.
[14] Using NutVal 4.1, https://www.nutval.net/

quota for a given district, government officials often respond by not registering the full number of family members in a household (enabling them to reach more households, but with smaller household-level rations or cash payments). As such, even though an average public work household size is six members, they may only receive transfers calculated for a household size of three. In this section, I will selectively describe SPIR program components that were combined into different treatments that will be explained in the research study design section. These are livelihoods, mental health, nutrition, and men's engagement.

4.2.1 Livelihoods

One of the core components of standard graduation programming is a relatively large one-time cash or asset transfer accompanied by targeted training and weekly follow-up support and coaching. In contrast, SPIR tested a less intensive, or 'light touch,' approach that aligned with the PSNP4 livelihoods design. Specifically, PSNP4 envisioned a 'credit' track where more capable PSNP clients (often with assets that could be used as collateral) would receive training and support in livelihoods groups and linked to local financial institutions to access credit to grow or diversify their livelihoods activities. The poorest PSNP clients who did not have assets or ability to access credit were designated to be included in a 'livelihood transfer' track and receive training and a one-time cash transfer of $200.

Broadening its scope beyond a sole focus on livelihoods, SPIR formed groups known as Village Economic and Social Associations (VESAs) as a foundational platform for PSNP clients to save, access small loans, receive support initiating income generating activities, and discuss different social issues together. SPIR established more than 5,000 VESAs formed by 25 members from 11-14 households (with both husband and wife participating). The bi-weekly VESA meetings were also forums for discussion of topics including health, hygiene, nutrition and gender norms, facilitated by SPIR community facilitators or volunteers. VESAs established their own by-laws and built social cohesion through collection of social funds to support members during times of crisis or loss, including the death of a family member. Local financial institutions reported they were more confident providing credit to PSNP clients who demonstrated their ability to consistently repay loans in VESAs.

Mirroring the PSNP4 'credit' pathway, SPIR provided training in collaboration with local government staff to a subset of VESA members who developed business plans and were linked to local microfinance institutions (MFIs) to access loans. The majority expressed interest and self-selected livestock-focused activities, namely improved poultry production and/or short-cycle animal fattening and sales. This focus on livestock was driven in part by high and increasing demand for animal sourced foods in Ethiopia.[15] One of the most popular livestock activities, poultry

[15] With increasing population, urbanization, and changing diets, demand for animal sourced foods is projected to increase by 80 percent in Africa between 2010 and 2030 (World Economic Forum, 2019). In Ethiopia, where cereal prices had remained relatively stable, real prices of animal sourced foods (milk, eggs, and meat) rose between 32 to 36 percent over the decade before the start of this program (F. Bachewe, Minten, & Yimer, 2017). With the largest livestock population in Africa, livestock is also one of the highest contributions to national income in Ethiopia, as compared to other countries in the continent (Enahoro, Njiru, Thornton, & Staal, 2019).

production, has a low barrier to entry and set-up costs scalable to starting with a relatively affordable number of improved-breed chickens and other inputs. SPIR provided business planning support, training on animal management, poultry coop construction and market linkages to inputs such as improved breed chickens, feed, and veterinary services. As individuals grew their small poultry enterprises, producer marketing groups were formed for the collection and sale of eggs to larger urban markets. The improved breed chickens promoted by the program could produce four times as many eggs as local-breed chickens (240 vs. 60 annually), while the cockerels could achieve market weight (2-3 times the size of local breeds) in a quarter of the time. While small ruminant animal rearing is common in rural Ethiopia, it relies on extensive, often unsustainable grazing practices and is often used as a savings and wealth accumulation strategy rather than an economically productive venture. The zero-grazing, short-cycle animal fattening strategies promoted by SPIR included training on construction of low-cost, hygienic animal enclosures, nutritious feed composition, animal health and management. As both poultry production and short-cycle animal fattening activities were managed at the household premises, they were also perceived as appropriate for PSNP households, with their limited agricultural landholdings and underutilized labor.

Lastly, the poorest households in our study sample were identified using an index of assets, constructed using data from the baseline survey, and received either a $200 cash transfer or a poultry start-up package of equivalent value. This was done according to a study cluster sub-randomization that will be described further in the study design section (Section 6).

4.2.2 Mental Health

In light of the significant gaps in access to mental health treatment and related services in these types of settings, mental health practitioners have developed and tested relatively simple forms of psychotherapy that can be delivered by non-specialists after receiving one-time practical training. One of these evidence-based interventions is Interpersonal Psychotherapy (IPT). Developed in the 1970s, it has been tested and recommended by the World Health Organization (WHO) for treating depression in both individual and group settings (Cuijpers et al. 2016; World Health Organization 2016). In partnership with the Director of Clinical Training at Teachers College, Columbia University, SPIR IPT-G Officers and Community Health Facilitators (CHFs) received an intensive 2-week training in IPT for groups (IPT-G) followed by supervised practice sessions.

Female caregivers with mild to moderate depressive symptoms were invited to participate in an IPT group with 5-6 other women in T1 and T3 kebeles. Facilitated by CHFs, the 12-week program supported participants to recognize links between their depressive symptoms and current life problems such as grief, disputes, devastating life changes, social isolation and loneliness. The sessions focused on enhancing interpersonal skills to better manage these problems. Throughout the intervention, IPT-G Officers provided supportive supervision to CHFs, with remote support from the IPT-G Coordinator and Columbia University Teachers College psychologists. Even after the 12 sessions, group participants continued to meet together autonomously, receiving monthly follow-up visits from the CHFs for the first 6 months after the end of the sessions.

4.2.3 Nutrition

Aligning with the Government of Ethiopia's National Nutrition Programme (NNP), explicit nutrition components were added in the fourth phase of the PSNP. These included an increased focus on linkages to health services and soft-conditionalities for pregnant and lactating women to follow recommended antenatal health facility visits while they were excluded from public works requirements and placed in a temporary direct support category. There were also nutrition behavioural change communication (BCC) sessions introduced that counted towards 4 days of public works requirements over a six-month public work period. SPIR augmented these by reinforcing key messages during VESA discussions including optimal infant and young child feeding (IYCF) practices, maternal nutrition, and utilization of health and nutrition services. Local health development army (HDA) volunteers were invited to become VESA members to support these nutrition BCC discussions. An 'enhanced' nutrition treatment added activities such as timed and targeted counselling (TTC) and community-based participatory nutrition promotion (CPNP). Evidence from nutrition-specific programming indicated that these activities could improve child nutritional outcomes (Kim et al. 2019; Kang et al. 2017) in similar settings. CHFs supported HDA volunteers to provide the timed and targeted (TTC) household-level counselling to pregnant and lactating women and their spouses. The plan was to provide eleven household visits (4 during pregnancy and 7 afterwards) and promote behaviors such as adequate diet and rest during pregnancy, optimal breastfeeding and complementary feeding, and warning signs for seeking professional care during a child's infancy.

Caregivers of underweight children in enhanced nutrition treatment kebeles were invited to participate in CPNP, a two-week, learning-by-doing nutritious meal demonstration and nutrition promotion model for rehabilitating malnourished children using locally available, participant-contributed foods. In addition to cooking demonstration and feeding, promoted BCC messages included diet diversification, importance of animal sourced food, hygiene, sanitation and child stimulation. Hygiene considerations were incorporated into the design of activities such as poultry production to prevent young children's exposure to animal faecal matter.

4.2.4 Men's engagement

In the past, many women's empowerment activities tended to focus predominantly on women for skill development, and access to savings and lending in female-only groups. In response to a growing appreciation that meaningfully engaging men is critical in changing social norms around gender and achieving women's empowerment goals, SPIR intentionally included both men and women in VESAs and in nutrition BCC at household-level TTC counselling. Within VESAs, gender-focused dialogue sessions included topics on the workloads of men and women, improved communication skills, shared household decision making, and men's engagement in childcare and household chores. In addition, in enhanced nutrition treatment kebeles (T1 and T3), men were invited to participate in men's engagement groups. Twenty men per group participated in eight discussion sessions where they reflected on gender norms in their community and the origins of their own beliefs of gender roles. Men's engagement facilitators followed a curriculum that challenged men to consider the unequal workload born by women, and to support their wives, especially in household tasks that were most commonly associated as women-only activities. After these sessions, some self-identified men's engagement champions described their own

experiences and demonstrated their commitment to challenging some of the social norms by cooking and preparing food during community-level events.

5. Research Objective and Questions

The overall objective of this research is to explore and unpack the relative contribution of different program elements and their combinations within a multi-sectoral graduation program. Specifically, we aim to bridge critical gaps in the graduation literature by testing the inclusion of mental health, nutrition and men's engagement activities within this program. In contrast to most previous studies, an important contribution of this research is that the program was implemented at scale, benefitting over 150,000 households, and imbedded in the Government of Ethiopia's national Productive Safety Net Program (PSNP).

The first research question, addressed in Chapter 2, is can a light-touch version of graduation programming still impact household economic outcomes and facilitate pathways out of poverty, similar to the achievements observed in many 'full package' graduation programs? In this study, one-time livelihood transfers are notably smaller in size[16] compared to other graduation programs and exclusively targeted to the poorest segment of PSNP households.[17] Another light-touch feature of this program is the absence of high-frequency, household-level mentoring and coaching visits. Furthermore, in contrast to typical graduation program evaluations, which involve comparisons to a pure control group, our study incorporates control clusters that also receive core PSNP programming and transfers.[18] Thus, this chapter explores whether a set of basic livelihood support activities involving participation in savings groups and livelihood transfers for extremely poor households are sufficient to impact household savings, income, assets and consumption in this light-touch graduation program implemented at scale.

In Chapter 3, we examine the question: Does the inclusion of group psychotherapy into a light-touch graduation program lead to improved outcomes for women experiencing depression? While a number of studies have looked at the impact of economic interventions on mental health outcomes (Romero et al. 2021; McGuire, Kaiser, and Bach-Mortensen 2022) or the impacts of mental health interventions independently, only a limited number of studies have compared the impact of economic and psychotherapy interventions.[19] One study cross randomized cash transfers with an eight-week group psychotherapy for criminally engaged men in Liberia (Blattman, Jamison, and Sheridan 2017), and another experimentally interacts cash transfers and individual five-week psychotherapy for a rural population in Kenya (Haushofer, Mudida, and Shapiro 2020) yielding mixed results. To the best of our knowledge, our study represents the first to incorporate a

[16] $200 ($374 in 2017 PPP)

[17] By comparison, other graduation programs more commonly provide transfers in the range of $600-1,600 to all households in the program.

[18] Either 6 months of transfers for public works PSNP clients, or 12 months of transfers for permanent direct support clients.

[19] In addition to combined cash transfer and psychotherapy studies, there is a recent study that includes a light touch livelihood component with pharmacotherapy (Angelucci and Bennett 2024) and one that cross-randomizes a large cash transfer with a workshop to raise aspirations (Orkin et al. 2023).

targeted psychotherapy intervention in a graduation program.[20] By experimentally varying livelihood transfers, we explore whether either intervention is independently effective in improving outcomes for women with depressive symptoms in the context of a light-touch graduation program, or if the combination is necessary to achieve effects across various economic and wellbeing outcomes.

The primary research question in Chapter 4 is: Does the addition of intensive nutrition BCC to a light-touch graduation program improve child nutrition and growth? Graduation programs aim to improve household economic and social wellbeing, yet little is known about their potential to influence nutritional outcomes for children in these households. It is unclear what the optimal type or level of intensity of nutrition BCC could effectively enhance nutritional outcomes within graduation programs, and whether there exists potential complementarity with other components such as livelihood programming. In a related literature, a meta-analysis found no additional effect on child anthropometric outcomes from incorporating nutrition BCC in cash transfer programs (Little et al. 2021). This contrasts with three separate studies in Myanmar, Bangladesh and Ethiopia. In Myanmar, cash alone failed to impact child malnutrition, but treatment effects were observed when cash was combined with BCC (Field and Maffioli 2021). In Bangladesh, neither transfers nor nutrition BCC individually improved child nutrition, but the combination of cash and nutrition BCC had a 0.25 SD effect on child stunting (Ahmed, Hoddinott, and Roy 2019). A similar pattern was found in Ethiopia, where BCC or food vouchers alone had limited effects, but the combination resulted in improvements in child feeding practices and growth (Han, Kim, and Park 2021).

In Chapter 5, we explore the question: Does incorporating a relatively low-intensity men's engagement activity into a light-touch graduation program lead to changes in men's gender-equitable attitudes and encourage their involvement in household chores that are considered taboo for men in highly patriarchal settings? Development programs seeking to empower women by providing them productive asset transfers and related livelihood activities risk adding to an already overburdened workload and creating tension with their partners who may resent being excluded from these activities. Graduation programs have generally not achieved sustained improvements in women's empowerment outcomes. Moreover, the sole study that incorporated an explicit men's engagement activity into a graduation program found no differential effects from the additional 16-session men's engagement program (Angelucci, Heath, and Noble 2023).

6. Research Study Design

With the goal of determining the effectiveness of different combinations of program elements on a range of outcomes, an impact evaluation employing a cluster randomized controlled trial (RCT) was conducted with four intervention arms (three treatments and a control). Treatment one (T1) included the full set of enhanced livelihoods (L*) and enhanced nutrition (N*) activities. Treatment two (T2) combined the enhanced livelihoods interventions (L*) with the basic nutrition activities (N). Treatment three (T3) included basic livelihood interventions (L) with a full package of enhanced nutrition activities (N*). Treatment four (T4), the control arm, received only core PSNP support and

[20] In a graduation program in Niger, Bossuroy et al. (2022) study the effects of adding a psychosocial intervention targeting aspirations and social norms, but this was not intended to treat mental illnesses.

transfers. The study sample was comprised of PSNP households in 192 subdistricts (kebeles) in 13 districts (woredas) in Amhara and Oromia regions in Ethiopia. Randomization was done with stratification at the district level, with a rerandomization to achieve balance in subdistrict covariates. A sub randomization within L* clusters (T1 and T2 kebeles) assigned a cluster-level designation of either 'cash' or 'poultry' to compare a $200 cash transfer with an equivalent value, in-kind poultry start-up package for extremely poor households. The figure below captures these different treatment arms and the number of kebeles and households assigned in each.

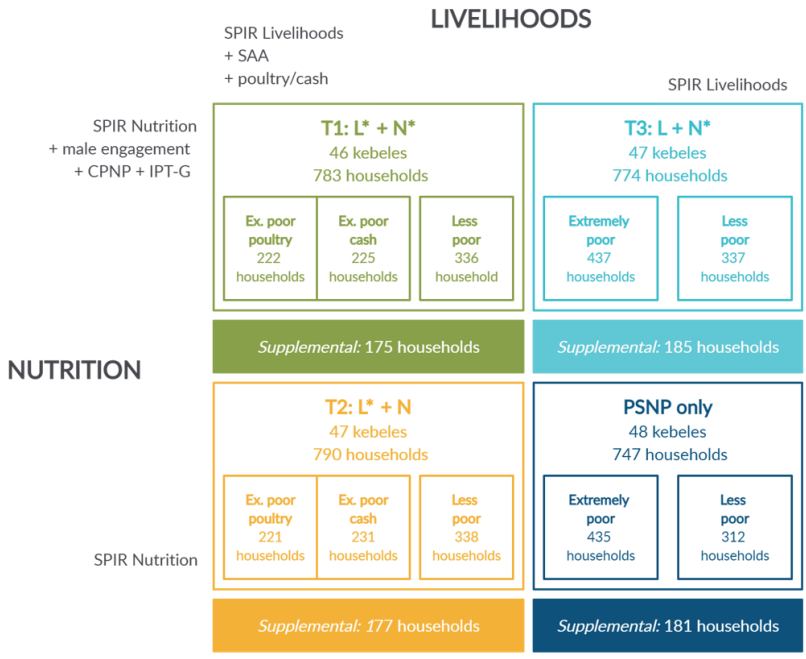


Figure 1: SPIR Impact Evaluation Study Design

In Chapter 2, the families of outcomes analysed were assets, financial inclusion, agriculture income (including livestock), non-agriculture income, and consumption. Average standard treatment effects (ASTE) for these outcomes and individual treatment effects for each variable were reported. For extremely poor households, we pooled all clusters assigned to receive poultry (in T1 and T2) and denote this as 'poultry.' Similarly, extremely poor households in clusters who received cash in T1 and T2 were pooled and analysed as 'cash.' These were compared to the extreme poor households in T3 that received only basic livelihoods support, and extreme poor households in the control arm (T4). Secondly, for the less poor sample of households in all treatment arms (T1-T3), a pooled treatment effect is estimated, as this set of households only received basic livelihoods support.

In Chapter 3, our primary outcome of interest is depression, as measured by the PHQ-9 instrument. We report effects on both the continuous score as well as on the prevalence of individuals with PHQ-9 ≥ 8. This threshold has precedence among other studies of depression in Ethiopia (Tiki, Taye, and Duko 2020; Woldetensay, Belachew, Biesalski, et al. 2018). We also report impacts on major depressive disorder (MDD), a more restrictive and clinically relevant measure of depression. We also report effects on psychosocial, economic, and anthropometric outcomes. After the first results table that presents findings on the entire sample of women at midline and endline (N=2,644), subsequent tables report treatment effects on the subsample of women with elevated depressive symptoms at midline (N=444).

In Chapter 4, we report on stunting for children 6-35 months of age as the primary outcome, together with other anthropometric outcomes. Given our interest in the interaction of the nutrition programming with the livelihood transfers, we restrict our analysis primarily to the extremely poor households (as less poor households were not eligible to receive the transfers). For each table of results, we present treatment effects for 5 different interactions: enhanced nutrition (EN) with poultry, EN with cash, EN without grants, core nutrition (CN) with poultry, and CN with cash. Intermediate outcomes that were analysed include mother's access to healthcare services, child's health history, mother's knowledge of infant and young child feeding (IYCF) practices, mother's dietary diversity, and child's dietary diversity. We also report average standard treatment effects for these key families of outcomes.

In Chapter 5, the primary outcomes of interest are men's involvement in domestic chores and men's gender equitable attitudes. As men's responses could be subject to social desirability bias, we include treatment effects based both on men's report of their own participation in select activities and women's report of their spouse's involvement in the same activities. In addition to reporting on the individual indicator, we also create summary indices for the two families of outcomes, men's involvement in household chores and men's gender equitable attitudes.

7. Description of Program Implementation and Study Rollout

The SPIR program broadly aligned with the fourth phase of the PSNP (PSNP4) and was implemented between January 2017 – September 2021. The study evaluation included three rounds of data collection at baseline (January-April 2018), midline (July-October 2019), and endline (February-April 2021). As SPIR programming began in certain kebeles in 2017, prior to the baseline, these kebeles were excluded from the study. Of the planned 3,474 households (18 eligible households selected using a simple random sample per kebele, in 193 kebeles), 3,314 were interviewed at baseline in 192 kebeles. One kebele was dropped due to insecurity during the baseline and on average 17 households met the selection criteria and were surveyed in the remaining kebeles. The inclusion criteria stipulated the household had to be a PSNP beneficiary household, having a child aged 0-35 months, and with the mother or primary female caregiver as a member of the household. Beginning in 2018, program implementation included VESA establishment, livelihoods training and linkage to financial institutions for targeted[21] PSNP clients, establishment of agro-dealers to improve access to quality inputs, health and nutrition activities

[21] less than 5% of PSNP households in this year

including CPNP, and supporting HDAs to conduct household-level nutritional counselling in enhanced nutrition kebeles. Water, sanitation and hygiene (WASH) activities were conducted in all treatment arms, including the control kebeles. This included construction of potable water points, latrines, and community-led total sanitation and hygiene promotion activities.

In April 2019, livelihood transfers were provided to the poorest 10 households in our sample in T1 and T2 arms. Depending on the random assignment to either a 'cash' or 'poultry' kebele within T1 and T2, recipients received either an unconditional cash transfer of $200 or training and a poultry start-up kit of equivalent value. The poultry kit comprised sixteen improved-breed, vaccinated chickens, an initial supply of feed (75 kgs), coop construction materials, feeding and watering troughs, handwashing station materials, and $36 in cash for veterinary services or other associated costs. The midline survey was conducted from July to October 2019, during which 3,220 of the initial 3,314 baseline study households were successfully interviewed. This reflects an attrition rate of 2.8%, with the most common reason being households relocating out of the study area.

During the midline, women were screened for depressive symptoms using the PHQ-9 depression assessment tool. This short depression screening instrument is used widely in nonpsychiatric settings and has been adapted and validated in Ethiopia (Woldetensay et al. 2018). In T1 and T3 treatment arms, 227 women exhibited elevated depressive symptoms and were invited to participate in IPT-G. Of these women, 188 attended at least one of the 12 weekly sessions facilitated by trained CHFs and participated in 10 sessions on average. Among the 39 women that did not attend, 17 were unable to participate in IPT-G as the CHF assigned to their kebele was on maternity leave or unavailable for other reasons. In kebeles where less than four women in the study sample were eligible to participate, CHFs screened in additional non-study participants in order to form groups with the recommended minimum size of 5-6 participants.

Another activity that started in T1 and T3 kebeles after the midline was men's engagement group discussions. In each of these kebeles, approximately 7-8 men's engagement groups were formed each having 20 men from both PSNP and non-PSNP households. This translated to about 40 percent of men in PSNP households participating in men's engagement groups in these kebeles. These groups met for eight sessions on a bi-weekly basis. After completion of the sessions, the plan was to conduct public events for men to share testimonials and promote their supportive household level practices. In most cases, these planned events were disrupted due to the restrictions on public gatherings in response to COVID-19. The endline survey, planned for mid-2020, was also delayed due to COVID-19, and was conducted instead between February and April 2021. At endline, 3,098 sample households were interviewed, signifying a 6.5% attrition rate from the original 3,314 households at baseline.[22] This attrition was due primarily to kebeles that became inaccessible due to conflict in the north of the country,[23] and is balanced across treatment and control arms.

[22] A supplemental sample of households was interviewed at both midline and endline surveys in order to achieve a parallel cross-section sample of households with young children for analysis of impacts on child anthropometric measures. This supplemental sample was not used in the analysis of the results with the exception of chapter 5.

[23] A conflict in the northern region of Tigray started in November 2020, and armed actors began raids in neighboring Amhara region, in SPIR operational areas, in February 2021. Militants later entered the SPIR implementation district of Tsagbji in March and temporarily occupied the district capital town, Tsata.

During the period preceding the endline, households experienced numerous shocks alongside disruptions caused by COVID-19 and conflict in the north. In the Oromia region, approximately one third of the study households lost crops due to desert locust infestations, while nearly two thirds reported crop losses due to fall armyworm. Crop losses were less prevalent among study households in the Amhara region, with only 15 percent reporting losses due to fall armyworm. Weather shocks also significantly impacted these predominantly subsistence, rainfed agriculture producing households. More than half of respondents in both regions reported experiencing drought, and a similar number of households reported experiencing floods and significant erosion events in the previous 15 months. Furthermore, 60 percent of households reported substantial increases in agricultural input prices, with a higher prevalence of such reports (85 percent of households) in the Oromia region. Social unrest and sporadic violence also occurred across Oromia region during this same period.[24] SPIR implementing partners employed various strategies to successful adapt programming in response to COVID-19 restrictions or temporary access issues due to insecurity. Given the high vulnerability of most PSNP households, it is likely that some of the modest gains in livelihoods and productive assets were eroded during this period of overlapping shocks.

8. Main Findings

As a preview of the findings in Chapter 2, the basic set of livelihoods activities involving participation in VESAs had significant effects solely in financial inclusion outcomes in the less poor households, and in financial inclusion and small effects on non-agricultural income in the extremely poor households. The enhanced livelihood treatment, including livelihood transfers, had slightly larger effects on financial inclusion as well as modest effects on income and asset outcomes among extreme poor households. These effects were similar across both cash and poultry modalities. Neither the basic nor enhanced set of livelihood interventions had any effects on consumption at endline.

In Chapter 3, we find that neither the enhanced livelihoods programming including transfers (T2), nor the enhanced nutrition programming including group psychotherapy (T3) independently reduced depression for women with depressive symptoms. When these interventions were combined (T1), they had a substantial effect on the prevalence of depression as well as on psychosocial, economic and wellbeing outcomes. Notably, these results were observed more than one year after the completion of the group psychotherapy, underscoring the importance of alleviating both financial and mental health constraints to both achieve and sustain these effects over this timeframe. A recent meta-analysis of the wellbeing effects of cash transfers (between $664 - $995 PPP) found an average effect size of 0.1 standard deviations, with larger transfers having larger effects on subjective wellbeing (McGuire, Kaiser, and Bach-Mortensen 2022). While the $200 livelihood transfer may have been too small, even when packaged with other SPIR

[24] A period of protests, often involving large numbers of Oromo youth, was sparked by the government's plan to displace farmers while expanding the Addis Ababa municipality into the Oromia region in 2014. These protests became enflamed by other grievances and often experienced violent suppression by security forces. One particularly widespread period of violence followed the murder of popular Oromo singer, Hachalu Hundessa, on 29 June 2020.

livelihood program components, to have treatment effects on wellbeing, its combination with group psychotherapy yielded notable effects on mental health as well as other outcomes for women with depression.

In Chapter 4, we find an increase in mothers' access to healthcare services across all treatment arms (T1, T2, T3). Specifically, the average standardized treatment effects for enhanced nutrition programming, with or without transfers, were approximately twice the size of those observed for basic nutrition programming with transfers. Although there were modest increases in maternal nutritional knowledge across both types of nutrition programming, they were often small in magnitude and weakly measured. No impacts on child nutrition were observed, except for the treatment that combined cash transfers with enhanced nutrition programming, reducing the prevalence of child stunting by approximately one third.

As a preview of the findings in Chapter 5, SPIR programming across all treatment arms T1-T3 yielded small effects on men's equitable attitudes (0.08-0.09 SD) and men's involvement in domestic tasks (0.12-0.18 SD) at the midline. For those in the treatment arm (T2) that were not exposed to the men's engagement intervention after the midline, these effects faded by endline. In contrast, in treatment arms that included the men's engagement activity (T1 and T3), the effects on men's involvement in domestic tasks persisted and increased to 0.26-0.29 SD. Given the plausibility of social desirability bias influencing men's self-reported behaviors, we include analysis of women's responses concerning their partner's involvement in household chores, food preparation, and collecting firewood or water. While the effect sizes are slightly less, the pattern of impact across all domestic tasks closely mirrors that reported by men.

References

Ahmed, Akhter, John Hoddinott, and Shalini Roy. 2019. "Food Transfers, Cash Transfers, Behavior Change Communication and Child Nutrition." *IFPRI Discussion Papers*, no. September: 40. http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/133420/filename/133631.pdf.

Angelucci, Manuela, and Daniel Bennett. 2024. "The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy." *American Economic Review* 114 (1): 169–98. https://doi.org/10.1257/AER.20210687.

Angelucci, Manuela, Rachel Heath, and Eva Noble. 2023. "Multifaceted Programs Targeting Women in Fragile Settings: Evidence from the Democratic Republic of Congo." *Journal of Development Economics* 164 (September): 103146. https://doi.org/10.1016/j.jdeveco.2023.103146.

Bachewe, Fantu, Bart Minten, and Feiruz Yimer. 2017. "The Rising Costs of Animal-Source Foods in Ethiopia: Evidence and Implications." *ESSP Working Papers*, no. August. https://ideas.repec.org/p/fpr/esspwp/108.html.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil. 2022. "Why Do People Stay Poor?" *Quarterly Journal of Economics* 137 (2): 785–844. https://doi.org/10.1093/qje/qjab045.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor Markets and Poverty in Village Economies*." *The Quarterly Journal of Economics* 132 (2): 811–70. https://doi.org/10.1093/qje/qjx003.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236). https://doi.org/10.1126/science.1260799.

Banerjee, Abhijit, Esther Duflo, and Garima Sharma. 2021. "Long-Term Effects of the Targeting the Ultra Poor Program." *American Economic Review: Insights* 3 (4): 471–86. https://doi.org/10.1257/aeri.20200667.

Banerjee, Abhijit, Dean Karlan, Robert Osei, Hannah Trachtman, and Christopher Udry. 2022. "Unpacking a Multi-Faceted Program to Build Sustainable Income for the Very Poor." *Journal of Development Economics* 155 (March): 102781. https://doi.org/10.1016/j.jdeveco.2021.102781.

Barrett, Christopher B., and Michael R. Carter. 2013. "The Economics of Poverty Traps and Persistent Poverty: Empirical and Policy Implications." *Journal of Development Studies* 49 (7): 976–90. https://doi.org/10.1080/00220388.2013.785527.

Barrett, Christopher B., Teevrat Garg, and Linden McBride. 2016. "Well-Being Dynamics and Poverty Traps." *Annual Review of Resource Economics* 8 (1): 303–27. https://doi.org/10.1146/annurev-resource-100815-095235.

Berhane, Guush, Daniel O. Gilligan, John Hoddinott, Neha Kumar, and Alemayehu Seyoum Taffesse. 2014. "Can Social Protection Work in Africa? The Impact of Ethiopia's Productive Safety Net Programme." *Economic Development and Cultural Change* 63 (3): 1–26. https://doi.org/10.1086/677753.

Black, Maureen M., Abdullah H. Baqui, K. Zaman, Shams El Arifeen, and Robert E. Black. 2009. "Maternal Depressive Symptoms and Infant Growth in Rural Bangladesh." *American Journal of Clinical Nutrition* 89 (3): 951–57. https://doi.org/10.3945/ajcn.2008.26692E.

Blattman, Christopher, Stefan Dercon, and Simon Franklin. 2019. "Impacts of Industrial and Entrepreneurial Jobs on Youth: 5-Year Experimental Evidence on Factory Job Offers and Cash Grants in Ethiopia." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3380592.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2018. "The Long Term Impacts of Grants on Poverty: 9-Year Evidence from Uganda's Youth Opportunities Program." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3223028.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (4): 1165–1206. https://doi.org/10.1257/aer.20150503.

Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand, et al. 2022. "Tackling Psychosocial and Capital Constraints to Alleviate Poverty." *Nature* 605 (7909): 291–97. https://doi.org/10.1038/s41586-022-04647-8.

Brudevold-Newman, Andrew, Maddalena Honorati, Pamela Jakiela, and Owen Ozier. 2017. "A Firm of One's Own Experimental Evidence on Credit Constraints and Occupational Choice." *World Bank Policy Research Working Paper Series*, no. 10583. https://doi.org/10.1596/1813-9450-7977.

Central Statistical Agency (CSA) [Ethiopia] and ICF. 2016. "Ethiopia Demographic and Health Survey, 2016." Addis Ababa, Ethiopia, and Rockville, Maryland, USA.

Cuijpers, Pim, Tara Donker, Myrna M. Weissman, Paula Ravitz, and Ioana A. Cristea. 2016. "Interpersonal Psychotherapy for Mental Health Problems: A Comprehensive Meta-Analysis." *American Journal of Psychiatry* 173 (7). https://doi.org/10.1176/appi.ajp.2015.15091141.

Davis, Benjamin, Stefania Di Giuseppe, and Alberto Zezza. 2017. "Are African Households (Not) Leaving Agriculture? Patterns of Households' Income Sources in Rural Sub-Saharan Africa." *Food Policy* 67: 153–74. https://doi.org/10.1016/j.foodpol.2016.09.018.

Duflo, Esther, and Abhijit V. Banerjee. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.* PublicAffairs.

Ethiopian Public Health Institute (EPHI) [Ethiopia] and ICF. 2021. "Ethiopia Mini Demographic and Health Survey 2019: Final Report." Addis Ababa, Ethiopia: EPHI/FMoH/ICF. https://www.dhsprogram.com/pubs/pdf/FR363/FR363.pdf.

Essers, Dennis, Kelbesa Megersa, and Marco Sanfilippo. 2019. "The Productivity Gaps of Female-Owned Firms: Evidence from Ethiopian Census Data." *Economic Development and Cultural Change*. https://doi.org/10.1086/703101.

Field, Erica M., and Elisa M. Maffioli. 2021. "Are Behavioral Change Interventions Needed to Make Cash Transfer Programs Work for Children? Experimental Evidence from Myanmar," February. https://doi.org/10.3386/W28443.

Glinski, A., C. Schwenke, L. O'Brien-Milne, and K. Farley. 2018. "Gender Equity and Male: It Only Works When Everyone Plays." Washington, D.C.

Han, Yaeeun, Hyuncheol Bryant Kim, and Seollee Park. 2021. "The Roles of Nutrition Education and Food Vouchers in Improving Child Nutrition: Evidence from a Field Experiment in Ethiopia." *Journal of Health Economics* 80 (December): 102545. https://doi.org/10.1016/J.JHEALECO.2021.102545.

Haushofer, Johannes, and Ernst Fehr. 2014. "On the Psychology of Poverty." *Science* 344 (6186): 862–67. https://doi.org/10.1126/science.1232491.

Haushofer, Johannes, Robert Mudida, and Jeremy Shapiro. 2020. "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-Being." Cambridge, MA. https://doi.org/10.3386/w28106.

Haushofer, Johannes, and Daniel Salicath. 2023. "The Psychology of Poverty: Where Do We Stand?," December. https://doi.org/10.3386/W31977.

Haushofer, Johannes, and Jeremy Shapiro. 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. https://doi.org/10.1093/qje/qjw025.

———. 2018. "The Long-Term Impact of Unconditional Cash Transfers: Experimental Evidence from Kenya." Nairobi, Kenya.

Hirvonen, Kalle, Yan Bai, Derek Headey, and William A Masters. 2020. "Affordability of the EAT–Lancet Reference Diet: A Global Analysis." *The Lancet Global Health* 8 (1): e59–66. https://doi.org/10.1016/S2214-109X(19)30447-4.

Hirvonen, Kalle, Elia A. Machado, Andrew M. Simons, and Vis Taraz. 2022. "More than a Safety Net: Ethiopia's Flagship Public Works Program Increases Tree Cover." *Global Environmental Change* 75 (July): 102549. https://doi.org/10.1016/j.gloenvcha.2022.102549.

Kang, Yunhee, Sungtae Kim, Sisay Sinamo, and Parul Christian. 2017. "Effectiveness of a Community-Based Nutrition Programme to Improve Child Growth in Rural Ethiopia: A Cluster Randomized Trial." *Maternal and Child Nutrition* 13 (1): 1–15. https://doi.org/10.1111/mcn.12349.

Karlan, Dean, Ryan Knight, and Christopher Udry. 2015. "Consulting and Capital Experiments with Microenterprise Tailors in Ghana." *Journal of Economic Behavior and Organization* 118: 281–302. https://doi.org/10.1016/j.jebo.2015.04.005.

Kassa, Getachew Mullu, and Amanuel Alemu Abajobir. 2018. "Prevalence of Common Mental Illnesses in Ethiopia: A Systematic Review and Meta-Analysis." *Neurology Psychiatry and Brain Research* 30 (December 2017): 74–85. https://doi.org/10.1016/j.npbr.2018.06.001.

Kim, Sunny S., Phuong Hong Nguyen, Yisehac Yohannes, Yewelsew Abebe, Manisha Tharaney, Elizabeth Drummond, Edward A. Frongillo, Marie T. Ruel, and Purnima Menon. 2019. "Behavior Change Interventions Delivered through Interpersonal Communication, Agricultural Activities, Community Mobilization, and Mass Media Increase Complementary Feeding Practices and Reduce Child Stunting in Ethiopia." *Journal of Nutrition* 149 (8): 1470–81. https://doi.org/10.1093/jn/nxz087.

Little, Madison T., Keetie Roelen, Brittany C.L. Lange, Janina I. Steinert, Alexa R. Yakubovich, Lucie Cluver, and David K. Humphreys. 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. https://doi.org/10.1371/JOURNAL.PMED.1003698.

Manley, James, Harold Alderman, and Ugo Gentilini. 2022. "More Evidence on Cash Transfers and Child Nutritional Outcomes: A Systematic Review and Meta-Analysis." *BMJ Global Health*. BMJ Publishing Group. https://doi.org/10.1136/bmjgh-2021-008233.

McCullough, Ellen B. 2017. "Labor Productivity and Employment Gaps in Sub-Saharan Africa." *Food Policy* 67: 133–52. https://doi.org/10.1016/j.foodpol.2016.09.013.

McGuire, Joel, Caspar Kaiser, and Anders M. Bach-Mortensen. 2022. "A Systematic Review and Meta-Analysis of the Impact of Cash Transfers on Subjective Well-Being and Mental Health in Low- and Middle-Income Countries." *Nature Human Behaviour* 6 (3): 359–70. https://doi.org/10.1038/s41562-021-01252-z.

Nagler, Paula, and Wim Naudé. 2017. "Non-Farm Entrepreneurship in Rural Sub-Saharan Africa: New Empirical Evidence." *Food Policy* 67: 175–91. https://doi.org/10.1016/j.foodpol.2016.09.019.

Olney, Deanna K., Aulo Gelli, Neha Kumar, Harold Alderman, Ara Go, and Ahmed Raza. 2022. "Social Assistance Programme Impacts on Women's and Children's Diets and Nutritional Status." *Maternal and Child Nutrition*. John Wiley and Sons Inc. https://doi.org/10.1111/mcn.13378.

Orkin, Kate, Robert Garlick, Mahreen Mahmud, Richard Sedlmayr, Johannes Haushofer, Stefan Dercon, Paul Brimble, et al. 2023. "Aspiring to a Better Future: Can a Simple Psychological Intervention Reduce Poverty? *." https://www.socialscienceregistry.org/trials/996.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms." *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.aay0214.

Rijkers, Bob, Måns Söderbom, and Josef L. Loening. 2010. "A Rural-Urban Comparison of Manufacturing Enterprise Performance in Ethiopia." *World Development* 38 (9): 1278–96. https://doi.org/10.1016/j.worlddev.2010.02.010.

Romero, Jimena, Kristina Esopo, Joel McGuire, and Johannes Haushofer. 2021. "The Effect of Economic Transfers on Psychological Well-Being." *Working Paper*.

Rosenberg, Adam M., John A. Maluccio, Jody Harris, Marjolein Mwanamwenge, Phuong H. Nguyen, Gelson Tembo, and Rahul Rawat. 2018. "Nutrition-Sensitive Agricultural Interventions, Agricultural Diversity, Food Access and Child Dietary Diversity: Evidence from Rural Zambia." *Food Policy* 80 (October 2017): 10–23. https://doi.org/10.1016/j.foodpol.2018.07.008.

Ruel, Marie T., Agnes R. Quisumbing, and Mysbah Balagamwala. 2018. "Nutrition-Sensitive Agriculture: What Have We Learned so Far?" *Global Food Security* 17 (January): 128–53. https://doi.org/10.1016/j.gfs.2018.01.002.

Sabates-Wheeler, Rachel, Kalle Hirvonen, Jeremy Lind, and John Hoddinott. 2022. "Expanding Social Protection Coverage with Humanitarian Aid: Lessons on Targeting and Transfer Values from Ethiopia." *The Journal of Development Studies* 58 (10): 1981–2000. https://doi.org/10.1080/00220388.2022.2096443.

Sibhatu, Kibrom T., and Matin Qaim. 2018. "Review: The Association between Production Diversity, Diets, and Nutrition in Smallholder Farm Households." *Food Policy* 77 (October 2017): 1–18. https://doi.org/10.1016/j.foodpol.2018.04.013.

Surkan, Pamela J., Caitlin E. Kennedy, Kristen M. Hurley, and Maureen M. Black. 2011. "Maternal Depression and Early Childhood Growth in Developing Countries: Systematic Review and Meta-Analysis." *Bulletin of the World Health Organization* 89 (8): 608-615E. https://doi.org/10.2471/BLT.11.088187.

Victora, Cesar Gomes, Mercedes De Onis, Pedro Curi Hallal, Monika Blössner, and Roger Shrimpton. 2010. "Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions." *Pediatrics* 125 (3). https://doi.org/10.1542/peds.2009-1519.

Waal, Alex De. 2017. *Mass Starvation : The History and Future of Famine*. Newark: Polity Press. http://rbdigital.rbdigital.com.

Woldetensay, Yitbarek Kidane, Tefera Belachew, Markos Tesfaye, Kathryn Spielman, Hans Konrad Biesalski, Eva Johanna Kantelhardt, and Veronika Scherbaum. 2018. "Validation of the Patient Health Questionnaire (PHQ-9) as a Screening Tool for Depression in Pregnant Women: Afaan Oromo Version." Edited by Steven LoBello. *PLOS ONE* 13 (2): e0191782. https://doi.org/10.1371/journal.pone.0191782.

World Bank. 2019. "Poverty & Household Welfare in Ethiopia 2011-2016." Washington, DC.

World Health Organization. 2016. "Group Interpersonal Therapy (IPT) for Depression." *World Health Organization* 1.

Chapter 2: Livelihoods

Can a Light-Touch Graduation Model Enhance Livelihood Outcomes? Evidence from Ethiopia

Jessica Leight, Harold Alderman, Daniel Gilligan, Melissa Hidrobo, and Michael Mulford

Abstract

In recent years, a growing literature has examined the potential of multifaceted, intensive "graduation model" interventions that simultaneously address multiple barriers constraining households' exit from poverty. In this paper, we present new evidence from a randomized trial of a lighter-touch graduation model implemented in rural Ethiopia. The primary experimental arms are a bundled intervention including a transfer valued at $374 (randomly assigned to be cash or an equivalent value in poultry), training, and savings groups; a simpler intervention including training and savings groups only; and a control arm. We find that three years post-baseline, the intervention inclusive of the transfer leads to some increases in assets, savings, and cash income from livestock, though there is no shift in consumption or household food security; these effects are consistent regardless of the modality of the transfer (cash versus poultry). The effects of training and savings groups alone are minimal.

1. Introduction

In recent years, a growing literature in development economics has examined the multifaceted constraints faced by households in extreme poverty. Given the salience of these constraints, layered "graduation model" interventions that simultaneously address several barriers to exit from poverty are broadly viewed as promising. However, the effectiveness of these interventions at a larger scale has not been widely explored to date, and given that they generally center around the provision of a large asset transfer (valued between $500 and $1000) in conjunction with intensive household-level support, there may be important trade-offs between intensity and feasible scale.

In this paper, we present new evidence from a randomized trial of a light-touch graduation model implemented at scale in rural Ethiopia. Strengthen PSNP4 Institutions and Resilience (SPIR) targeted households who were beneficiaries of the government of Ethiopia's Productive Safety Net Program (PSNP), a safety net that provides food and cash transfers to roughly the poorest 20% of rural Ethiopian households during the agricultural off-season in targeted food insecure districts. Building on the base PSNP support, SPIR delivered a multifaceted graduation model encompassing both livelihoods and nutrition interventions and centering around the formation of village-level savings groups. While lighter touch than other models, SPIR was also delivered at relatively large scale, serving more than 150,000 households; this is larger than other graduation and asset transfer projects analyzed in the literature to date, with the exception only of BRAC's Targeting the Ultrapoor program in Bangladesh, serving nearly 400,000 households.[1]

Our trial seeks to further unpack this multifaceted model, analyzing four experimental arms. The first two arms included the full set of livelihoods interventions: a one-time livelihoods transfer provided only to the poorest 60% of sample households within each experimental cluster (using a within-subdistrict ranking of sample households by wealth), training on livestock production and marketing, and the formation of village-level savings groups, village economic and social associations or VESAs.[2] The transfer was valued at $374 in 2017 purchasing power parity dollars,[3] and a further randomization at the cluster level assigned the transfer modality to be either cash, or a poultry package (16 chickens and complementary inputs) of comparable value.[4] This generates two experimental arms designated as cash and poultry, respectively. The third experimental arm included savings groups and training only; and the fourth arm serves as the control arm. In addition to the livelihoods interventions, all treatment arms except the control received nutrition behavioral

[1] This estimate draws from the Bandiera et al. paper and refers to the scale of implementation at the time of the trial. A trial conducted in Niger was implemented in the context of a national cash transfer program that reached 22,000 households in the most recent phase of programming (100,000 households overall). Table A1 summarizes the characteristics and scale of graduation programs evaluated in recent randomized controlled trials.

[2] More specifically, transfers were designed to target 10 out of 18 households in each cluster randomly assigned to a transfer, or 56%, and the achieved targeting was 58% of households in those clusters.

[3] All outcomes of interest will be valued using PPP-adjusted dollars with 2017 as the base year, as discussed in more detail in Section B5. 2017 is the most recent year for which a purchasing power parity conversion is available, and it is proximate to the timing of our baseline (conducted in early 2018).

[4] For the poultry package, the chickens themselves and the materials (pen construction, feeders, watering trough, handwashing station, etc.) each constituted about 30% of the estimated value, while the feed was 20% of the total value. The remaining 20% was provided in cash for any additional associated costs such as veterinary services.

change counseling (BCC).[5] The sample included 3,314 households in 192 clusters (subdistricts) in two regions, Amhara and Oromia.

Given that the full set of livelihoods interventions inclusive of the transfer was delivered only to extremely poor households (the poorest 60% within a sample of PSNP beneficiaries who are themselves among the poorest 20% in rural Ethiopia), our primary findings focus on this subsample. We also report additional results for the less poor sample (the upper 40%); the less poor sample received the same uniform set of livelihoods interventions (savings groups and livestock training) in conjunction with nutrition interventions in all three treatment arms, and thus we estimate a single pooled treatment effect.

The key outcomes of interest were measured in three large-scale surveys at baseline (2018), and one and three years following program initiation (2019 and 2021).[6] The primary outcomes of interest, pre-specified in a registered analysis plan, include assets, financial inclusion, agricultural income, non-agricultural income, and consumption and food security, and we report average standard treatment effects to facilitate interpretation of effects within each outcome family.

Our primary findings suggest that the full set of livelihoods interventions (inclusive of asset transfers) rolled out to extremely poor households generated some modest asset accumulation, increased financial inclusion, and increased cash income from livestock, and these effects do not vary with respect to the transfer modality of poultry versus cash. The effects on overall asset value (driven by livestock) are modest and not statistically significant, but the effects on financial inclusion are large: the probability of credit access increases by eight to ten percentage points, relative to a mean of 45% in the control arm, while the probability of reporting any savings increases by more than 30 percentage points (relative to a mean of 40%). Both poultry and cash recipients also experience a 25% increase in past-year income from livestock. However, given that this income source constitutes only around 11% of the value of total consumption on average, even a large relative increase here is minor vis-à-vis the value of overall consumption, and thus unsurprisingly, there is no significant effect on household-level consumption or food security. In fact, the coefficients estimated are small in magnitude and often opposite in sign relative to the increase in consumption that was hypothesized ex ante.

For extremely poor households who received the more limited set of interventions (savings' groups and training only), the only significant effect observed is an increase in savings. A similar pattern is observed for the less poor households who accessed savings' groups and training only in all treatment arms. Thus, the treatment effects for both subsamples for this set of interventions are identical – an increase in the probability of reporting any savings of around 30 percentage points.

We conclude that the implementation of a light-touch graduation model at scale generally did not lead to an exit from poverty for the targeted households, a finding that makes several contributions to the broader graduation model literature. First, we analyze a lighter-touch model implemented at larger scale that is arguably more policy-feasible. In Ethiopia, even sustaining the base PSNP at a large scale has proved extremely challenging in recent years, and implementing transfers of up to $500 or $1000 – as is seen in some of the most successful graduation model programs – is likely

[5] There is some variation in the intensiveness of nutrition programming in different arms, and we will subsequently demonstrate that this variation has little impact on the livelihood outcomes of interest.
[6] The endline survey was delayed approximately nine months from its scheduled date in 2020 due to the onset of the COVID-19 pandemic.

impossible.[7] The current phase of PSNP programming, PSNP5, targets only 6% of beneficiary households for livelihoods transfers of $300. Second, we provide novel evidence about the effects of varying the modality of the main transfer (cash versus poultry), and our findings suggest that the impact of transfer modality is minimal. Third, we analyze the effects of the graduation model intervention for both an extremely poor and less poor sample.

In the previous literature, there is evidence that more intensive graduation model programs analyzed in a six-country evaluation in Ethiopia, Ghana, Honduras, India, Pakistan, and Peru had large positive effects in a range of domains in both the medium and long-term (Banerjee et al. 2015; Banerjee, Duflo, and Sharma 2021), as well as similar evidence from a large-scale trial of BRAC Targeting the Ultrapoor conducted in Bangladesh (Bandiera et al. 2017; Balboni et al. 2022). There is also evidence of positive shorter- and medium-term effects of a graduation model emphasizing enhanced psychosocial support in Niger (Bossuroy et al. 2022), and positive effects of a base graduation model in Afghanistan (Bedoya et al. 2019). In Ghana, another recent paper shows that a graduation model implemented without a productive transfer did not have positive effects, and an intervention centering only around the formation of savings groups had positive effects that did not persist (Banerjee et al. 2022). In Yemen, a graduation model implemented in a high-conflict setting had positive effects only on savings and assets four years post-transfer (Brune et al. 2022).

Our findings also contribute to a literature examining medium- or long-term effects of one-time transfers that are offered in the absence of any overarching graduation model programming, but normally with much larger transfers. In Kenya, a randomized controlled trial of unconditional cash transfers offered by GiveDirectly (valued at $700 in nominal terms or $600 in 2017 PPP) found positive effects only on assets three years post-transfer when transfer recipient households are compared to households in other villages in order to abstract from intravillage spillovers (Haushofer and Shapiro 2016; 2018). Blattman et al. (2020) report the long-term effects of cash grants ($400, or $594 in 2017 PPP) for youth in Uganda after nine years, and again find significant effects only on durable assets and skilled work, despite large effects on earnings four years post-transfer (Blattman, Fiala, and Martinez 2014; 2020).

There is also a growing literature on transfers of animal assets that shows somewhat heterogeneous medium-term effects, though many papers here focus primarily on effects on nutrition (including child anthropometry) or food security (Rawlins et al. 2014; Miller et al. 2014; Jodlowski et al. 2016). Phadera et al. (2019) analyze the effects of an extremely large livestock transfer ($1600, or $934 in 2017 PPP) implemented in conjunction with skill trainings and supplementary services in Zambia. They find significant positive effects on consumption and assets as well as enhanced resilience approximately three years post-transfer (Phadera et al. 2019). Another more recent paper analyzes the effects of another substantial in-kind transfer of chickens and associated materials (valued at $500, or $717 in 2017 PPP) in Guatemala and find no significant effects on livelihood-related outcomes, on average, about a year post-transfer (Mullally, Rivas, and McArthur 2021).

[7] Recent publicly released papers by the World Bank and the UNDP have highlighted that in the fifth phase of the PSNP (the phase following PSNP4, during which SPIR was implemented), significant funding shortfalls have led to meaningful curtailment in multiple program dimensions. Thus far from expanding to encompass a more intensive graduation model intervention with a larger transfer, the PSNP is facing budgetary pressure to offer leaner transfers that have not kept pace with steadily escalating inflation (United Nations Development Program 2023; World Bank 2022).

2. Context and experimental design

2.1 Context

This trial was conducted in rural Ethiopia and focuses on a sample of largely subsistence agricultural producer households that are beneficiaries of the Productive Safety Net Program (PSNP). Launched in 2005, the PSNP is one of the largest safety net programs in sub-Saharan Africa, now serving eight million people annually. It provides cash and/or food transfers to rural households in the form of payment for labor on public works or direct transfers for households who do not have an eligible worker (Hoddinott and Mekasha 2020). The program is targeted both geographically (in districts that are often drought affected and chronically food-insecure) and at the household level, employing community-based targeting to select households that meet certain criteria, particularly food insecurity (Berhane et al. 2013).

A large existing literature analyzing the effects of the PSNP itself generally suggests it has some modest effects on enhancing food security and assets, but the effects are not large (Berhane et al. 2014; Gilligan, Hoddinott, and Taffesse 2009). The pattern of effects for child nutritional outcomes (Porter and Goyal 2016; Bahru et al. 2020) and agricultural productivity and yields (Hoddinott et al. 2012; Gazeaud and Stephane 2020) is similarly mixed. A recent paper suggests that during the period of pandemic-related restrictions and associated disruptions in 2020 (the year prior to the endline survey conducted for this trial), PSNP beneficiary households did not experience the deterioration in food security observed for non-beneficiary households, consistent with a protective effect of PSNP participation (Abay et al. 2021).

In general, the PSNP is structured around the provision of six months of payments (in food, cash, or a mix of the two) to rural households as payments for labor or direct transfers during the agricultural off-season (January to June). Across the PSNP, cash accounted for about 75% of this transfer value between 2010 and 2014, though there is considerable variation in the mix of cash and food provided across regions and over time (Hirvonen and Hoddinott 2021). Median annual transfers per household in our sample were valued at around $277 in 2017 PPP terms at baseline, with about 58% of the value of transfers received in cash.

2.2 Interventions

This randomized trial focuses on interventions conducted as part of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA), a five-year program (2016–2021) funded by USAID's Bureau of Humanitarian Assistance and led by World Vision, in partnership with ORDA Ethiopia and CARE. SPIR targeted nearly 150,000 households in 13 of the most vulnerable *woredas* (districts) in the Amhara and Oromia regions of Ethiopia, supporting PSNP implementation and providing complementary livelihood, nutrition, gender, and climate resilience activities.

Given its multiple objectives, SPIR encompasses a large number of program elements. This paper focuses particularly on the effects of livelihoods-related interventions, though we will also provide a brief overview of the nutrition and health-related programming, and more details are provided in project documents (Alderman et al. 2021). All interventions described here and evaluated as part of this randomized trial were delivered only to PSNP clients.

The primary livelihoods interventions include three components: a one-time transfer (valued at $374 PPP)[8] targeted at extremely poor households; livestock production and marketing training; and the formation of village economic and social associations (VESAs). The targeting of each of these interventions is distinct. VESAs were designed to be near-universal for PSNP households,[9] including 25—30 members per association (both men and women), and were used as a platform for facilitated discussions on financial literacy, income generation activities, and nutrition, health and gender-related topics. By contrast, livestock trainings were targeted only to households who demonstrated capacity and interest, and the topics (primarily improved poultry production and short-cycle goat and sheep fattening) reflected households' preference for diversification into productive livestock activities.[10]

Finally, the productive transfer was targeted to the poorest 60% of sampled PSNP beneficiary households.[11] These households (denoted the extremely poor) were identified using an asset index constructed from baseline data; the identification process was implemented in in every sample cluster, independent of treatment status, with extremely poor households then receiving transfers in select clusters based on cluster-level random assignment. The transfer was structured either as a lump-sum cash payment or an in-kind poultry package of comparable value (and the modality determined by further random assignment, as described further below). The cash transfer consisted of 5600 birr provided in cash (not via bank transfer), equivalent to approximately $200 at the prevailing market exchange rate or $374 in 2017 PPP, a sum equivalent to about 1.4 months of average household consumption expenditure. The poultry start-up package included 16 45-day old improved-breed chickens from EthioChicken (on average, eight male and eight female) and complementary inputs (75 kilograms of feed, chicken coop construction materials, a feeding trough, and cash to purchase veterinarian services). As previously noted, the chickens themselves and the materials (pen construction, feeders, watering trough, handwashing station, etc.) each constituted about 30% of the estimated value of the poultry package (60% of the total), while the feed and cash each constituted about 20% of the value. The EthioChicken Sasso breeds were expected to reach a higher market weight (for cockerels) and produce up to four times the number of eggs in comparison to local chicken breeds, and preliminary evidence from other evaluations has suggested that they are in fact meaningfully more productive (IDInsight 2018).

The experimental design also included a number of additional interventions in the nutrition and health sectors. The core nutrition intervention (denoted CN) included health and nutrition BCC messages provided in facilitated discussions at VESAs. In clusters assigned to enhanced nutrition and health programming (denoted EN), these interventions were supplemented by a timed and targeted counseling (TTC) strategy that trained local health workers and volunteers to provide BCC at health posts and in home visits around early childhood and maternal nutrition; households

[8] The value of the transfer was 5600 birr in 2019; based on a market exchange rate in that period of around 27 birr per dollar, this was slightly over $200. Converted to 2017 dollars in purchasing power terms, the value is estimated to be $374.

[9] Households that were included in the PSNP for permanent direct income support but who did not participate in public works due to the absence of an eligible household laborer (generally, elderly and disabled households) were not invited to join VESAs.

[10] While the training was offered at the household level, households would identify a single individual to participate in the training and then be the primary actor in managing the new productive activity. In general, a majority of participants were female, particularly for training around poultry.

[11] More specifically, 10 out of 18 households received the transfer. If there were fewer than 18 households sampled, the poorest 10 would still receive the transfer.

including a child who was underweight (based on weight for age and mid-upper arm circumference) were invited to participate in community-based participatory nutrition promotion that included growth monitoring and nutritious food preparation. Following the first year, the EN interventions also included targeted male engagement discussions designed to shift norms around gendered roles, particularly in terms of engagement in household tasks and child-rearing, and the provision of interpersonal therapy in groups to women (and later men) who were identified as eligible based on the prevalence of depressive symptoms.

2.3 Study Design

The study was a cluster randomized controlled trial designed to unpack the effectiveness of various elements of a multisectoral graduation model. The sample included 192 subdistricts (kebeles) in 13 districts (woredas) in Amhara and Oromia regions.[12] The subdistricts were selected based on programmatic criteria, restricting to those areas in which programming (in particular, the formation of VESAs) had not yet launched at baseline.

In the original study, subdistricts were randomized to four arms as summarized graphically in Figure A1. T1 included both the full (enhanced) set of livelihoods interventions and enhanced nutrition programming; T2 included full livelihoods interventions and core nutrition programming; T3 included core livelihoods interventions and enhanced nutrition programming; and the fourth arm constitutes the control arm of households who receive the PSNP only. An additional cross-randomization then assigned clusters within T1 and T2 to receive either cash or poultry transfers; as noted above, these transfers were rolled out only to the poorest 60% of households in each cluster.[13]

This paper focuses solely on livelihoods outcomes, and accordingly we analyze this design somewhat differently. First, we separate the extremely poor and less poor samples. For the extremely poor sample, we pool all clusters assigned to receive poultry and those assigned to receive cash (in both T1 and T2), to identify two treatment arms of analytical interest: in the first arm (denoted Poultry) a full set of livelihoods interventions was rolled out, including a poultry transfer; in the second arm (denoted Cash), a full set of livelihoods interventions was rolled out, including a cash transfer. In the third arm (denoted L), a core set of livelihoods interventions (savings groups and training) was rolled out. Second, less poor households (the other 40% of the sample) were offered only the core set of interventions (savings groups and trainings) in all three arms, and thus a pooled treatment effect across all three arms (L) is estimated for these households. The design of the trial as analyzed is summarized graphically in Figure 1.

[12] SPIR was operational in seven woredas in Amhara (these seven later became nine, due to some administrative divisions implemented within pre-existing woredas) and six woredas in Oromia, for a total of 13 woredas. The average subdistrict population is 903 households at baseline, of which 20% on average (around 180 households) are eligible for the PSNP.

[13] There was another cross-randomization assigning subdistricts in T1 and T2 to receive a targeted video-based intervention designed to enhance aspiration; sample households were invited to a screening of these videos with the objective of encouraging them to raise their aspirations and make investments that have positive long-term returns. Separate findings have shown that even at the one year follow-up, the effects of this intervention were not significantly different from zero (Leight et al. 2021) . Accordingly, we do not analyze this dimension of randomization any further in this paper.

As noted above, there is some variation in the random assignment of core or enhanced nutritional programming within the primary experimental arms analyzed here. Within the poultry and cash arms, 50% of subdistricts (those originally assigned to T1) received enhanced nutritional programming, while 50% (those originally assigned to T2) received core nutritional programming. The extremely poor sample in the L arm (drawn from T3) received enhanced nutritional programming. For the less poor households, we are estimating a pooled treatment effect of L, within which two thirds of clusters (those originally assigned to T1 and T3) received enhanced nutrition programming. In our analysis, we will explore in detail heterogeneity with respect to the assignment of subdistricts to varying nutrition interventions.

Randomization was conducted in Stata by the research team using stratification at the district level and a rerandomization procedure designed to achieve balance in kebele-level covariates. More details about randomization can be found in Appendix A.1.

2.4 Data Collection and Timeline

The evaluation included three major survey rounds: a baseline survey conducted between January and April 2018; a midline survey conducted following one year of program implementation between July and October 2019; and an endline survey following three years of program implementation conducted between February and March 2021. The primary intervention activities including the formation of VESAs in the study subdistricts commenced immediately followed the baseline, and implementation continued through summer 2021. The one-time transfers of poultry and cash were delivered in April 2019, approximately four months prior to the one year follow-up survey. Figure 2 summarizes the overall timeline.

The sampling frame was constituted by PSNP households reporting a child aged 0 to 36 months at baseline, and a simple random sample of households meeting these eligibility criteria was selected.[14] The target sample included 3,494 households, and the realized sample was 3,314 households or roughly 17 households per subdistrict.[15] In the one-year follow-up, 3,220 of the realized sample households were surveyed for an attrition rate of 3%, and in the three-year follow-up, 3,098 of the sample households were surveyed for an attrition rate of 7%. Figure A2 provides a flow chart summarizing the sampling. The rate of attrition is substantially driven by clusters that were entirely lost to follow-up due to conflict-related disruptions, and is balanced across the treatment and control arms.[16]

In each survey round, surveys were conducted with the primary female (defined as the primary caretaker, usually the mother, of the index child aged 0—3 at baseline) and the primary male, if present (defined as her husband or partner). Key outcomes of interest included demographic characteristics, participation in the PSNP and SPIR interventions, savings and assets, agriculture, livestock production, consumption and food security, infant and young child feeding practices,

[15] The target sample was 18 households per subdistrict in 193 subdistricts; however, one subdistrict was also dropped during the baseline survey due to insecurity.

[16] At both follow-up surveys, additional samples were also constructed with households including young children (0—24 months) in order to constitute a parallel repeat cross-section sample for analysis of intervention effects for anthropometrics. The supplemental samples were not included in any analysis reported in this paper because in the absence of baseline data, we cannot construct an asset index to identify them as extremely poor or less poor.

maternal nutrition, and depression and well-being. Appendix Table A2 provides an overview of the survey modules administered to each respondent in each round. Each survey respondent provided informed oral consent; ethical review and approval of the trial was provided by the Institutional Review Boards at IFPRI and at Hawassa University.

2.5 Outcome Variables

The analysis in this paper focuses on five pre-specified families of outcomes: assets, financial inclusion, income from agriculture and livestock, non-agricultural income, and consumption and food security. We will report the average standard treatment effects for these five outcome families as well as the treatment effects for each specific variable. While the majority of outcomes of interest are measured only at the three year follow-up, a subset of outcomes were also measured in the one-year follow-up survey. For the five average standard treatment effect coefficients, we also report q-values robust to the analysis of multiple hypothesis estimated across the set of ASTEs (Simes 1986).[17]

For assets, we analyze total asset value and the value of productive assets, consumer durables, and livestock assets as measured at the three year follow-up; at the one year follow-up, only livestock assets are reported.[18] For financial inclusion, we analyze any access to credit, access to formal credit, and total credit accessed over the past year, as well as a binary variable for any savings and the current savings balance. For income, we analyze income from livestock sales, income from sales of livestock products, crop income, and total livestock and crop income, all reported over the past year; for livestock products and crops, we separately analyze both the value sold for cash and the imputed value of the full amount produced. For non-agricultural income, we analyze binary variables for any income from any non-agricultural business, and any income from either formal and casual wage work (both reported for any household member over the past year). Agricultural income and income from a non-agricultural household business are not reported at the one year follow-up.

For consumption and food security, we analyze total, food and non-food consumption expenditure (reported per adult equivalent over the past month), and the food insecurity experience scale (FIES) as a continuous measure (also reported over the past month); consumption and food security are not reported at the one year follow-up. All continuous measures of value and income are converted to 2018 birr using Ethiopian CPI data, given that 2018 is the baseline year of this evaluation, and then to US dollars using a purchasing power-adjusted exchange rate from 2017, the most recent year for which the International Comparison Program has published PPP exchange rates.[19] (More details about the measurement of assets and consumption are provided in Appendix A.2.)

[17] We estimate these using the command qqvalue.
[18] Assets are valued at a market price based on price reported in complementary market surveys; for livestock, self-reported estimates of value are also reported by respondent households, and we can demonstrate that the estimated treatment effects are robust to using an alternate valuation strategy.
[19] Given that the baseline survey was conducted beginning in January 2018, the use of an exchange rate from 2017 is also logical in terms of timing.

3. Empirical results

3.1 Characterizing the Sample

Table 1 presents summary statistics, including simple demographic statistics as well as the baseline values of outcomes of interest, when available. The average household in the sample includes six members, and 20% of households are headed by a female. The average level of education is extremely low: more than 70% of household heads do not report any formal education, and nearly 70% report that their main activity is crop production. Baseline economic characteristics highlight that the sample is extremely poor in terms of consumption levels and characterized by limited access to financial services or formal employment. Average consumption expenditure per month per adult equivalent is $60. The estimated baseline value of assets owned is $1596, of which only 6% is constituted by non-livestock productive assets, 7% is constituted by durable consumption assets, and 86% is constituted by livestock assets.[20] Only 20% of households report any access to credit over the past year, and only 9% report access to formal credit; only 28% report any savings. Between 5% and 10% of households report any non-agricultural business or any engagement in wage labor over the past year (formal or informal).

To assess balance in baseline characteristics, we estimate balance tests separately for the extremely poor and less poor sample. For the extremely poor sample (Panel A of Table 1), the baseline variables of interest are regressed on binary variables for assignment to the poultry, cash and L arms (conditional on woreda fixed effects), and the joint p-value corresponding to the test $\beta_{poultry} = \beta_{cash} = \beta_L$ is reported in Column (6). For the less poor sample (Panel B of the same table), the same variables of interest are regressed on a single binary variable capturing assignment to a pooled treatment arm (again, less poor households were offered the same interventions in all three treatment arms). There is no evidence of any imbalance across treatment arms, and the joint test of balance yields a p-value of 0.108 for the extremely poor sample and 0.807 for the less-poor sample.

Table A3 reports baseline characteristics comparing across the extremely poor and less poor samples (pooled across treatment arms). As previously noted, the extremely poor households were identified using an asset proxy index at baseline that is presumptively correlated with a range of other socioeconomic characteristics, and the summary statistics here suggests that is indeed the case and that for the majority of indicators, the differences are highly statistically significant. Less poor households are larger in size and are much less likely to be female-headed (10% versus 27%): the household head in a less poor household is four percentage points more likely to report some education. Asset value is more than double for less poor households vis-à-vis extremely poor households ($2271 versus $1052), and the probability of reporting any savings is about 11 percentage points higher. Baseline consumption is not observed to be significantly different in

[20] Note that the estimated value of assets does not include cash savings, nor any imputed value of land or housing; land and housing are not privately owned or freely traded in rural Ethiopia, and thus valuing these assets is extremely challenging. The estimated value of all assets relies on prices reported in market prices; for livestock, self-reported valuations by households were also collected, and suggest a total estimated value that is around 25% lower vis-à-vis the market price valuation.

levels comparing across the two subsamples, though a Kolmogorov-Smirnov test does narrowly reject the hypothesis that the distributions are equal (p=0.094).[21]

3.2 Intervention Exposure

Table A4 reports summary statistics around intervention exposure to assess household engagement in the primary SPIR interventions; we generally draw on data reported at the three year follow-up survey, though data on receipt of transfers (poultry or cash) is from the one year follow-up, as this survey was more proximate to the date of the transfer. For extremely poor households as reported in Panel A, membership in VESA groups is uniformly high (80% or higher), though attendance over the past year is slightly lower (between 60% and 70%).[22] Between 30% and 40% of households in the kebeles assigned to any treatment arm report participating in financial education (implemented within VESA meetings) or livestock trainings, and this is relatively balanced across treatment arms with the exception of participation in livestock training; the latter is notably higher in the poultry arm (42%) compared to the other two treatment arms (28—30%).[23] This pattern may be unsurprising given that presumably there are strong complementarities between the receipt of a poultry package and training. Participation in livestock training is also notably balanced across men and women.

With respect to transfer receipt as reported in the one year follow-up survey, 94% of households assigned to the poultry arm report receipt of a poultry package and the average number of poultry reported received conditional on reporting receipt is very close to the target of 16. There are minimal reported spillovers to subdistricts in the control arm, and some reported spillovers (15-18%) to subdistricts in the other treatment arms, with these households reporting receipt of only six chickens; these responses presumably reflect a comprehension error in which households are reporting participation in an earlier SPIR nutrition-oriented activity where pregnant and lactating women where provided a subsidized provision of six improved breed hens. For the cash transfers, only around 63% of households report receipt of cash; there are minimal reported spillovers in the other arms. Administrative records suggest that 100% of targeted households received cash transfers, and so this may be suggestive of some recall error.[24]

Panel B reports parallel data for less poor households. VESA membership and participation in value chain and financial trainings is broadly similar for these households. No less poor households were eligible for either poultry or cash transfers, and this is broadly what is observed,

[21] Figure A3 shows the estimated kernel density of consumption at baseline (truncated at the 99th percentile) and some gap between the distribution for extremely poor and less poor households is observed, albeit noisy.

[22] Again, the reported level of non-participation in VESAs is consistent with the observation that around 15% of households served by SPIR are permanent direct support (PDS) households that receive PSNP benefits without the requirement of providing labor to public works due to the absence of an eligible adult laborer (generally, these households are headed by elderly and/or disabled individuals). PDS households would generally not participate in VESAs.

[23] Participation in financial training was reported only in the one year follow-up survey due to an error in survey design.

[24] Disbursement of cash was monitored by implementing partners, with each sampled household providing written signature confirmation of the receipt of the transfer and records cross-checked by supervisory staff. However, we cannot rule out the hypothesis that some cash was not delivered as planned, particularly in clusters characterized by a high concentration of reported non-receipt.

though again around 20% of households report receipt of six poultry, presumably with reference to the other intervention described previously.

3.3 Primary Results

Our primary specification uses the randomized design and estimates intent-to-treat effects separately for extremely poor and less poor households. For extremely poor households, we estimate the following specifications first excluding and then including the interactions with enhanced nutrition programming, where k denotes subdistrict or kebele (the level of randomization), and w denotes woreda.

$$Y_{ikwt} = \beta_1 Cash_{kw} + \beta_2 Poultry_{kw} + \beta_3 L_{kw} + Y_{ikw,t-1} + \eta_w + \varepsilon_{ikwt} \quad (1)$$

$$Y_{ikwt} = \beta_1 Cash_{kw} + \beta_2 Cash_{kw} \times EN_{kw} + \beta_3 Poultry_{kw} + \beta_4 Poultry_{kw} \times EN_{kw} + \beta_5 L_{kw} + Y_{ikw,t-1} + \eta_w + \varepsilon_{ikwt} \quad (2)$$

We also report tests of equality for coefficients β_1, β_2 and β_3 in the first model.

For extremely poor households, 50% of households in the cash and poultry arm were exposed to enhanced nutrition programming (denoted EN) and 50% were exposed to core nutrition programming (denoted CN); all households in the L arm were exposed to EN. In general, we hypothesize that the plausible interaction effects with enhanced nutrition interventions are small, though there may be some increased investment in livestock production in households exposed to nutrition BCC who seek to also increase the consumption of animal-source foods by pregnant and lactating women and infants and young children. Following recent guidance, we report both the fully interacted ("long") and non-interacted ("short") models to avoid incorrect inference based on the significance and sign of the interaction terms (Muralidharan, Romero, and Wüthrich 2023).

For the less poor households, we estimate the following specifications, again both excluding and including interactions with enhanced nutrition programming. L_{kw} here captures a pooled treatment effect for any livelihood intervention in the three treatment arms.

$$Y_{ilwt} = \beta L_{kw} + Y_{ikw,t-1} + \eta_w + \varepsilon_{ikwt} \quad (3)$$

$$Y_{ilwt} = \beta_1 L_{kw} + \beta_2 L_{kw} \times EN_{kw} + Y_{ikw,t-1} + \eta_w + \varepsilon_{ikwt} \quad (4)$$

For less poor households, 66% of all households pooling across treatment arms were exposed to enhanced nutrition programming, and 33% were exposed to core nutrition programming. All specifications include strata (woreda) fixed effects η_w, and standard errors are clustered at the subdistrict level (corresponding to the level of treatment assignment). For continuous variables (including the value of assets, income, savings, and consumption), we present inverse hyperbolic sine transformations to allow for non-linear effects. We then assess the magnitude by calculating the semi-elasticity to estimate the effect of assignment to this treatment arm in terms of a percentage increase relative to the mean of the outcome in the control arm (Bellemare and Wichman 2020).[25]

[25] For any treatment effect estimated using the inverse hyperbolic sine transformation, the elasticity of the outcome of interest with respect to treatment can be calculated using the formula exp (β-var(β))-1. The

The primary results are summarized in Table 2 and Table 3, for the average standard treatment effects as observed three years post-program for the short and long (fully interacted) model, respectively. Figure 3 then summarizes the findings from the short model in graphical form. The corresponding graphs for the long, fully interacted model are presented in Figure A4 in the Appendix.

For the extremely poor sample, the average standard treatment effects suggest that there are weakly positive effects (around .1 standard deviations) on assets in the poultry and cash arms, statistically significant primarily in the poultry arm. All three treatment arms exhibit a large and precisely estimated increase (around .3 standard deviations) in financial inclusion, comprised of variables reporting access to credit and savings. Finally, there is an increase in income from agriculture and livestock that is significant in the cash and poultry arms (.2 standard deviations), but generally no significant effects observed on non-agricultural income or consumption. (The only exception is a significant increase in engagement in non-agricultural income in the L arm.) More detailed results reported below will suggest that the divergent patterns for income and consumption reflect the fact that the increase in income from livestock observed for households in the poultry and cash arms is not large enough to manifest as a higher level of consumption. All of the statistically significant treatment effects are robust to correction for multiple hypothesis testing.

The estimated results of the long specification reported in Table 2 further suggest there is very little evidence of any meaningful heterogeneity in impacts on livelihoods outcomes generated by experimentally varying exposure to enhanced nutrition programming: the interaction effects are not only statistically insignificant, but are generally extremely small in magnitude, and thus the estimated coefficients on the treatment arms cash, poultry, and L remain highly consistent comparing across the long and short models. The only exception is that the interaction term between cash and enhanced nutrition is highly significant and negative for assets, a pattern that will be further discussed in Section 3c below.

For the less poor sample, there is generally no evidence of any significant treatment effects other than a positive and significant effect on financial inclusion, again of magnitude around .3 standard deviations. This effect is comparable in magnitude to the effect observed for the extremely poor households. Again, there is little evidence of any meaningful heterogeneity with respect to exposure to nutrition programming, and the primary estimated coefficients remain consistent in both the long and short models.

Table 4 through Table 8 then present the disaggregated results, focusing on the short model that does not include the interaction effects with enhanced nutrition programming. Here, the coefficient estimates of interest should be interpreted as the effect of the livelihoods interventions conditional on the distribution of nutrition programming; the full models including interaction effects are presented in Tables A11 through A15, and are discussed in Section 3c. For each set of outcomes, we analyze first the results for the extremely poor sample, followed by the results for the less poor sample. Note that the mean of all continuous outcomes in the control arm is reported in the tables for the untransformed variables (without the inverse hyperbolic sine transformations).

---

percentage effect is then calculated relative to the mean of the IHS-transformed variable in the control arm (Bellemare and Wichman 2020).

The findings on assets are presented in Table 4, for extremely poor households in Panel A and less poor households in Panel B. There is a large increase in the estimated value of livestock assets at the one year follow-up (10% in the cash arm, and 31% in the poultry arm), suggesting that even for households that receive cash transfers, livestock is a primary productive activity of interest. (The magnitudes of these coefficients are calculated as follows: we calculate exp (β-var(β))-1, equal to 1.912 for poultry; we then assess the mean relative to the mean of the IHS-transformed variable in the control arm, 6.191, yielding an effect of 31%.) Two years later, however, these effects have considerably attenuated and converged across poultry and cash recipients: there is a 9% increase in livestock value in the poultry arm and a 7% increase in the cash arm, and the difference across the two arms is no longer statistically significant. Consistent with this pattern, the difference in the estimated coefficients on livestock assets at the one and three year follow-ups is highly significant in the poultry arm but is not significant for the cash arm, where there is little variation over time. Moreover, the increase in total asset value at the three year follow-up is not statistically significant in either the cash or poultry arm, despite the significant increase in the value of livestock assets. For households in the L arm (who received no lump-sum transfer), there are no significant effects on assets at one year, and some evidence of a decline in assets at three years.

In the control arm, the mean value of total assets at the three year follow-up is $1105 for extremely poor households– nearly identical to the mean at baseline -- and this value is again overwhelmingly constituted (88%) by livestock. The treatment effect observed thus implies that the absolute gain in livestock assets is between $100 and $77 in the poultry and cash arms, respectively, and thus around 25% of the original transfer has persisted in livestock value. We can also assess persistence using a different set of data by analyzing effects on the counts of different livestock types in the short- and long-term follow-ups, as reported in Table A5. In the short-term, we see that poultry recipients report an increase of nine poultry owned relative to the control, and cash recipients report an increase of slightly under one poultry; both report an increase of around .6 sheep or goats, and no increase in oxen or cows. (The magnitude of the increase for poultry recipients is consistent with the recommendation provided during training that recipients sell the cockerels, roughly half of the flock, following a roughly three month period of fattening, while retaining the hens for eggs.) By the long-term follow-up, poultry recipients now have only two more chickens vis-à-vis control households, while their holdings of other livestock are unchanged; cash recipients have experienced some attenuation in both ownership of chickens and sheep and goats. The most parsimonious interpretation of this pattern is that livestock holdings have dwindled over time due to either mortality or buffer stock sales to fund consumption, and have not been replaced. Columns (4) and (8) in the same table also demonstrate that the treatment effects on estimated value of livestock are robust to valuing livestock at self-reported prices rather than market prices.[26]

Panel B of Table 4 presents parallel results for the less poor households (who, again, did not receive any lump-sum transfer). There is little evidence of any significant effect on assets at either

[26] The mean levels of livestock asset value in the same table suggest that respondents themselves estimate a value for their livestock that is about 25% lower than the implied mean in market prices: i.e., the mean estimated value in the control arm in the long-term follow up is $1103 when valued at own-reported prices, versus $1495 when valued at market-reported prices. The treatment effects remain similar in magnitude and precision using own-reported prices, suggesting a larger effect relative to the control mean.

one or three years, though there is an increase in the reported value of consumer durables owned at the three year follow-up, seemingly counterbalanced by a decline in livestock value.[27]

Moving on to financial inclusion, Table 5 presents the findings from the one year follow-up survey and Table 6 from the three year follow-up survey; in both tables, the findings from extremely poor households are presented in Panel A, and from less poor households in Panel B. For extremely poor households, there is no evidence of any meaningful effects on credit access at one year. However, a significant increase in credit access is observed at three years in the poultry and cash arms: this is an increase of seven to nine percentage points in the probability of accessing any credit, relative to a mean of 45%, and an eight percentage point increase in the probability of accessing formal credit, relative to a mean of 13%. The differences across waves in the estimated treatment effects are generally statistically significant for credit for the poultry and cash arms (as reported in Table 6), though more precisely estimated for poultry recipients.

For savings, we find positive effects for extremely poor households that are consistent across all three treatment arms and across both waves. In the one year follow-up, there is a 30 percentage point increase in the probability of reporting any savings relative to a mean of around 50%, and this pattern is persistent at three years. In continuous terms, the mean level of savings in the control arm in the three year follow-up survey is about $34, and the (unconditional) treatment effect implies a roughly threefold increase relative to this mean. (Interestingly, the level of savings in the control arm also dramatically declines by roughly half from the one to the three year follow-up, from $75 to $35, and thus the relative treatment effect commensurately increases over time.) For less poor households, a similar positive effect on savings is observed in both periods that is of parallel magnitude (roughly 30 percentage points) to the effect observed for extremely poor households, but no effect on credit access is observed for the less poor households in either wave.

Findings around income at the three year follow-up are presented in Table 7. (Supplementary results around income from wage work at one year are presented in Table A6 in the Appendix; income from cropping and livestock was not measured in that survey round.) In Panel A of Table 7, extremely poor households in the poultry and cash arms show evidence of a significant increase in past-year income from sales of livestock and livestock products, but no increase in income from cropping. There is no heterogeneity in the effect comparing across the poultry and cash arms, where the implied semi-elasticity suggests around a 25% increase relative to the control mean. However, the total estimated value of income from livestock, livestock products and crops (Column 7) shows a significant effect only in the poultry arm, a difference attributable primarily to the fact that the increase in the sales of livestock products (eggs) is somewhat larger in this arm. For households in the L arm who did not receive any transfers, there is no significant effect on any form of income, and a similar pattern is observed in Panel B for less poor households. Importantly, there is no evidence of any experimental effects in either sample or in either survey wave on the (extremely low) levels of participation in non-agricultural activities. Across both rounds, only 3% of households report any non-agricultural business (6% for less poor households), only 3—4% report any engagement in regular wage labor, and around 25% report any engagement in casual wage

[27] Interestingly, a weakly positive (but insignificant) effect on livestock asset value at the one year follow-up is subsequently observed to be weakly negative (but insignificant) at three years, and the difference across waves is statistically significant. While weak evidence, this could be suggestive of negative spillover effects on livestock asset holdings of the less poor households driven by growth in livestock for extremely poor households that received transfers.

labor. These low levels of participation remain unchanged.

Finally, Table 8 presents findings around consumption and food security, and there is no evidence of any meaningful experimental effect here; the estimated coefficients are small in magnitude and varying in sign. In fact, there is some evidence of an increase in food insecurity (a higher FIES score) among extremely poor households in the cash arm. While potentially somewhat surprising, these findings are consistent with the absence of any meaningful difference in consumption between households categorized as extremely poor and less poor based on their asset holdings – suggestive of a potentially low marginal utility for additional consumption -- and also consistent with the broader literature suggestive of a relatively low elasticity of consumption, inclusive of food consumption, for extremely poor households (Banerjee and Duflo 2007; Subramanian and Deaton 1996).

Given recent evidence that treatment effects estimated using the inverse hyperbolic sine may not be robust to alternate rescalings (Chen and Roth 2022), we also use two simpler specifications to assess the robustness of the treatment effects reported for continuous variables: a linear specification, and a linear specification for a binary variable equal to one if the variable of interest is strictly positive (non-zero). (The latter specification is not estimated for the total value of livestock product and crop production or for consumption, given that there are few or no zero values.) The findings around assets (Table A7) suggest that the observed increases in asset values remain positive and significant when using a binary specification, but are not significant in a linear specification. The increases in savings at one and three years (Table A8) and income at three years (Table A9) remain significant in both robustness specifications. The findings around consumption (Table A10) remain null.

To interpret the disparate findings on income vis-à-vis consumption, it is useful to note that an accounting of income sources for the sample households indicates that cash income from livestock constitutes a relatively small share of total income. Total estimated household consumption per year is around $2528 in this sample; given an average household size of 5.5, this corresponds to roughly $1.26 in consumption per person per day. Summing up the enumerated sources of income in-kind and in-cash reported in the household survey (including income from cropping over both seasons, income from sales of livestock and/or livestock products, income from cash and non-agricultural businesses, and income from PSNP transfers) accounts for around 60% of this total. Some possible sources of income --- e.g., remittances or informal transfers --- were not enumerated; in addition, underestimation of income vis-à-vis consumption is a common challenge in rural household surveys (Deaton 2019). Unsurprisingly, the sample households are substantially dependent on crop production, largely for subsistence (accounting for nearly 60% of enumerated income, or around a third of consumption); income from livestock and income from PSNP transfers each account for around 20% of enumerated income, or 11% of estimated consumption. Income from non-agricultural sources is negligible. Thus, even a meaningful (25%) relative increase in cash income from livestock such as that reported above would plausibly fail to be large enough to generate a detectable shift in consumption. In fact, this is exactly the pattern we observe.

3.4 Heterogeneity With Respect to Exposure to Enhanced Nutrition Programming

As described in some detail above, the experimental design also entails some variation across arms in exposure to enhanced nutrition programming (denoted EN) vis-à-vis core nutrition

programming (denoted CN). Among cash and poultry recipients, 50% of households were in subdistricts randomly assigned to receive EN, and 50% were in kebeles randomly designed to receive CN. All extremely poor households in subdistricts in the L arm are also exposed to EN. Among less poor households, two thirds were exposed to enhanced nutrition, and one third to core nutrition.

Findings around the effects of the nutritional interventions are reported in a separate paper and suggest there is little evidence of a significant effect on the majority of nutritional outcomes of interest for the less poor, other than some effects on knowledge and dietary diversity for women; for the extremely poor who received cash transfers and enhanced nutrition services, there is some evidence of reduced stunting (Alderman et al. 2023).[28] As previously noted, we report in Table 2 the estimated average standard treatment effects for the primary outcomes using a "long" specification that includes the interaction terms with the nutrition cross-randomization. In general, these interaction terms are not significant. They are also very small in magnitude relative to the main effects: for savings and income, for example, the interaction effects range in size between one tenth and one third the magnitude of the primary effect. Though achieving adequate statistical power to detect statistical significance in an interaction term can be challenging in randomized controlled trials (Muralidharan, Romero, and Wüthrich 2023), there is no evidence in this case that the interactions with nutrition programming are large and noisily estimated; rather, they seem to be consistently small.

The only exception to the above pattern is the coefficient on cash and the interaction between cash and EN in the specification employing assets as the dependent variable. In Column 1 of Table 2, we can observe that the coefficient on the cash binary variable is large (.2 standard deviations) and very precisely estimated for the sample that was not exposed to EN programming, while the interaction effect suggests an effect that is zero or slightly negative for the sample that is exposed to EN. This pattern could suggest that cash recipients also exposed to enhanced nutritional interventions are characterized by a higher level of expenditure on, for example, higher-cost or nutritious foods, and thus a lower level of investment in productive assets. This would be consistent with the evidence of an increase in child anthropometry in the subarm of cash recipients exposed to enhanced nutrition programming (Alderman et al. 2023).[29]

3.5 Attrition

As noted above, the base level of attrition in the sample is extremely low; only 3% of households attrite from the baseline to the one-year follow-up, and slightly over 5% attrite by the three year follow-up. The majority of the attrition at three years reflects entire kebeles rendered inaccessible to conflict, rather than household-level attrition. Among the 114 attrited households at one year, the top reasons for attrition included moving outside the survey area (n=44) and refusing consent (n=17). Among the 220 attrited households at three years, the top reason for attrition included

[28] There is some evidence in the one year follow-up that households that received the poultry transfers also consumed more eggs, but this pattern does not persist until endline (Alderman et al. 2022).

[29] As consumption data was not collected at midline, we are not able to identify if cash recipient households in the EN arms devoted relatively more resources to purchasing higher-value or nutritious foods for some period following the receipt of the cash transfer, at the expense of investing in asset accumulation. Such an increase in consumption would be plausible given the anthropometric effects, but the findings presented in Table 6 in this paper suggest that any such shift in consumption had dissipated by the point of the endline survey.

unrest in Amhara (n=72) and moving outside the survey area at any point during the evaluation period (n=92).

To assess any bias introduced by loss to follow-up, we regress binary variables for attrition at one and three years on baseline covariates, binary variables for the three primary treatment coefficients of interest, and the interaction between the two; we implement this test for a concise set of demographic characteristics and baseline values of outcomes of interest. The results are reported in Table A18 and show that there is relatively little evidence of any selective attrition comparing across the treatment and control arms. Female-headed households are significantly more likely to attrite, as are households in the L arm. The only estimated coefficients on an interaction term between a baseline covariate and a treatment arm that are significant are suggest that households without a formally educated head are somewhat less likely to attrite in the cash arm; and households with a higher level of baseline assets are somewhat less likely to attrite in both the poultry and L arms. All three coefficients are significant at only the ten percent level, and the latter two are also small in magnitude. Overall, we assess that there is little evidence of any meaningful bias introduced by selective attrition.

4. Discussion

The findings here suggest that a lighter-touch graduation model implemented at scale in rural Ethiopia generally did not catalyze a substantial shift out of poverty for the targeted households, a pattern at odds with the findings of a number of other recent trials. A summary of recent relevant trials and the transfer sizes analyzed is provided in Table A1, and it is evident that SPIR did effectively target an extremely poor population; monthly per capita consumption in this sample is around $47 at baseline, comparable to a number of the sites in the original multicountry graduation model trial as well as the sample analyzed in a recent trial in Niger (Bossuroy et al. 2022), and around 40% lower than consumption at baseline as measured in the Bangladeshi sample analyzed in Bandiera et al. (2017).

However, the intervention is lighter touch than other graduation models in a number of respects. First, the value of the lump-sum cash or asset transfer ($375 in purchasing power terms and about eight times monthly per capita consumption) is the second-smallest in this literature in absolute terms; the previous literature generally analyzes transfers that are between eight and 26 times monthly per capita consumption, with the exception only of Bossuroy et al. (2022) in Niger, where this ratio is around four. Second, there is no high-frequency household-level coaching included, in contrast to other models that entail weekly coaching visits over a period of two years. While a majority of the households participated in facilitated discussions around financial literacy, saving and lending, and business planning within VESAs, only a subset of households (about 30—40%, by design) participated in a one-time, intensive training on selected livelihood activities.[30] Third, all households receive consumption support transfers (including households in the control arm), a design feature also observed in Bossuroy et al. but generally not in the other papers in this literature.[31] This further narrows the gap between households in the control and treatment arms.

[30] Nor is any psychosocial support provided by the majority of the sample, though women who were identified as eligible for group therapy based on the prevalence of depressive symptoms were offered therapy.
[31] In the Ethiopian sample analyzed in Banerjee et al. (2015), all households were similarly PSNP beneficiaries and thus receiving regular PSNP transfers, inclusive of households in the control arm. A subset of households in the control arm in the trial sample in Peru also received regular cash transfers.

Importantly, the SPIR intervention did generate extremely large effects on financial inclusion, where there are increases of between .2 and .4 standard deviations, comparable to or slightly larger than the effects observed in Banerjee et al. (2015) and Bandiera et al. (2017). The treatment effects on cash income from livestock (around .2 standard deviations in this paper) also compare favorably to the treatment effect estimates observed in Banerjee et al. (2015) (around .35 standard deviations), notable given that the asset transfers in Banerjee et al. are generally much larger. However, in general cash income from livestock represents a small share of total consumption for households in this sample, who are heavily dependent on subsistence agricultural production as well as the transfers disbursed through the PSNP itself. Accordingly, even quite substantial effects on livestock income are insufficient to generate any shift in consumption or catalyze a movement out of poverty.

It may be informative to consider a more detailed comparison of the economic context faced by the households reached by SPIR and analyzed in this trial, vis-à-vis the economic context observed for ultra-poor Bangladeshi households, where a graduation model intervention analyzed in Bandiera et al. (2017) catalyzes much larger positive effects on consumption. In addition to being somewhat poorer in terms of monthly consumption compared to those households, this rural Ethiopian sample is notably different in its underlying economic characteristics. Ultra-poor Bangladeshi households are characterized by an almost total absence of assets: only 7% own land, and less than 10% report cow or goats, with an estimated total value of livestock of only around $40. By contrast, landlessness is rare in rural Ethiopia, where land is formally owned by the state with households having use rights that were locally allocated (Kebede 2008); 88% of households in our sample own land in this sense of allocated use rights, and livestock ownership is also relatively common, with 75% of even extremely poor households in our sample reporting ownership of some livestock at baseline. The women in extremely poor Bangladeshi households are predominantly engaged in casual labor in agriculture or domestic work (by contrast to women in richer households in the same villages, who are predominantly engaged in livestock rearing), and 40% of ultrapoor households report the woman as the sole income earner; whereas the households in our sample are predominantly engaged in subsistence agricultural production at baseline, with some additional income support from livestock rearing and the PSNP itself, and only about 20% are female-headed.

This comparison has two implications: first, relative to the ex ante asset endowment, the livestock transfer analyzed in Bandiera et al. (at $500 in PPP terms, or about ten times the ex ante livestock asset stock valued at around $40) is much larger than the transfer analyzed in this paper (around $400 in PPP terms, but less than half of households' ex ante livestock asset stock valued at around $900). Second, given the near absence of any meaningful non-agricultural or wage work in the sample subdistricts, the only feasible shift in household economic opportunities would be a very large substitution from cropping to livestock, or a substantial increase in income from cropping itself, and neither pattern is observed. The transfer rolled out as part of this light-touch model does have an appreciable effect on livestock income (a roughly 25% increase), but it is not large enough to catalyze any dramatic shift in household economic welfare.

Another informative comparison can be made to the sample of ultra-poor households in Niger included in the recent analysis of a graduation model enriched by targeted psychosocial interventions (Bossuroy et al. 2022). Consumption in the Nigerien sample is similar to consumption in our sample at baseline, and the lump-sum transfer implemented as part of the intervention (and provided to households in cash) is in fact slightly lower in magnitude than the

transfer provided here. The composition of household revenues is dramatically different in the Nigerien sample, however: even in the control arm, only around a quarter of household revenue is generated by agricultural production (harvest value), while nearly 60% is generated by non-agricultural businesses, with the remainder split between livestock revenue and wage revenue. Thus, while this sample of households is notably poor, it is not accurately described as dependent on subsistence agricultural production at baseline.

The graduation model intervention analyzed by Bossuroy et al., in turn, has particularly large positive effects on non-agricultural business revenue, with some (smaller) positive effects on crop value seemingly driven by more intensive use of cropping inputs, as well as positive effects on livestock revenue. The latter effects are roughly in the same range as, though somewhat larger than, the positive effects on revenue from livestock observed in this paper; the ex ante value of livestock assets (around $700 in the control arm) is also roughly similar to, though slightly lower than, the asset value of livestock observed in this sample, and the magnitude of the increase in livestock asset value observed in the Nigerien sample is similar (somewhat higher). Again, however, the positive effects on livestock assets and revenue observed in the Niger trial are relatively minor relative to the much larger positive effects on non-agricultural businesses and associated revenue reported there, and those effects are completely absent in this sample.

To sum up, the sample that we analyze here seems notable for two key characteristics: even relative to other samples of extremely poor households in other contexts, these rural PSNP beneficiary households are highly dependent on subsistence agricultural production at baseline (with some, but not substantial, income from livestock despite high livestock holdings); and they have almost no involvement in any form of non-agricultural business or wage labor, with these low rates of involvement remaining highly persistent over time. The light-touch graduation model implemented did not succeed in dramatically shifting households' portfolio of income-generating activities as in Bangladesh (where women shifted away from casual labor to livestock production), nor did it have a consistently positive effect across multiple income sectors as in Niger (where households already earning substantial non-agricultural income increased that income, while also increasing income from crops and livestock). Rather, the only channel through which the intervention was seemingly able to increase household resources was through expansion of livestock production and associated income from livestock and livestock products, a source that remains too small to substantially shift the trajectory of overall household income.

Our findings can also be usefully linked to a recent review paper and meta-analysis analyzing the effects of unconditional cash transfers and multifaceted graduation programs on consumption (Loeser and Kondylis 2021). In a pooled sample of studies, the authors find that the effects of transfers on consumption is around $0.35 per unit of transfer for UCTs, or $0.52 per unit of transfer of graduation programs, implying in this case an effect of between $130 and $195; relative to mean consumption in the control arm, this is a relative increase of 5—8%. The estimated increase in consumption corresponding to the 95% upper bound of our estimated treatment effects for poultry and cash recipients in the long-run follow-up are a roughly 2.5% and 1% in household consumption, respectively. Thus we can only narrowly rule out the hypothesis that the effect estimates here are consistent with the existing literature, though the effects on cash recipients seem more clearly to be below the level implied by previous estimates. The meta-analysis also usefully highlights that there is considerable dispersion in the estimated effects of any intervention of interest (particularly for multifaceted graduation models, a point the authors explicitly note), and thus focusing on the largest and most salient effect estimates --- such as those reported in the

original six-country evaluation or for BRAC's Targeting the Ultrapoor in Bangladesh --- may not be fully informative.

5. Conclusion

In this paper, we analyze the effects of a lighter-touch graduation model targeted to the poorest households in rural Ethiopia, and find that this model did not lead to a significant shift in consumption or an exit from poverty. There is an increase in livestock related income and access to credit that could point to a potential pathway for households to expand livestock production over time. However, households that had access to savings and training only showed no positive effects other than increased savings. Our findings also suggest that there is no difference in the effects of cash versus an alternate transfer modality (in this case, an improved breed poultry production package): both showed parallel effects on livelihood outcomes.[32]

More broadly, other contexts in which graduation models have been implemented successfully are often characterized by economic growth, even among poor rural households. By contrast, the setting of this study is characterized by stagnation or reversals, with consumption stationary among households in the control arm during the trial period (2018—2021); more dramatically, the already meager stocks of savings reported by households in control communities in the one year follow-up survey were halved by the three year follow-up, and the estimated value of livestock assets, the main asset owned by sample households, declined by 33%. This may reflect the fact that in the period preceding the endline, households in our sample were buffeted by numerous weather and man-made shocks. In Oromia, 35 percent of households surveyed in the three year follow-up reported losing crops to desert locusts and almost 60 percent reported losses due to army worm; in both Oromia and Amhara, more than half of respondents reported experiencing a flood event and associated erosion in the previous 15 months. In addition, more than 60 percent reported large increases in agricultural input prices, a shock that may itself reflect COVID-19 restrictions and associated supply chain disruptions. (Following the conclusion of this trial, the expansion of the Tigrayan conflict into northern Amhara generated a new and devastating round of conflict-related shocks affecting much of our Amhara study region.)

Even during periods of overall macroeconomic growth, the rural population in Ethiopia has often been left behind, with the poorest fifth of households experiencing stagnant or negative consumption growth compared to an average six percent annual increase in consumption for urban households between 2011-2016 (World Bank 2019). Paths to a sustainable income stream outside of agriculture also continue to be minimal, as rural non-farm enterprises in Ethiopia are often characterized by low productivity and only sporadic operation (Nagler and Naudé 2017; Rijkers, Söderbom, and Loening 2010). While there may be potential gains to income from domestic migration to urban areas (de Brauw, Mueller, and Woldehanna 2018; Mueller et al. 2018), remittance rates in Ethiopia are notably low (de Brauw, Mueller, and Woldehanna 2013). Given these broader economic trends limiting the ability of poor rural households to meaningfully diversify into any other productive economic activity, a lighter-touch graduation model may simply be insufficient to stimulate any meaningful exit from poverty.

[32] Separate findings suggest that differential effects on nutrition at the one year follow-up, including increased egg consumption for children and their mothers among the poultry recipient households, did not persist (Alderman et al. 2022; 2023).

References

Abay, Kibrom, Guush Berhane, John Hoddinott, and Kibrom Tafere. 2021. "COVID-19 and Food Security in Ethiopia: Do Social Protection Programs Protect?" *Economic Development and Cultural Change*, June. https://doi.org/10.1086/715831.

Alderman, Harold, Lucy Billings, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Alemayehu Seyoum Taffesse, and Heleene Tambet. 2021. "Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA): Endline Report."

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

Alderman, Harold, Daniel O Gilligan, Jessica Leight, Michael Mulford, and Heleene Tambet. 2022. "The Role of Poultry Transfers in Diet Diversity: A Cluster Randomized Intent to Treat Analysis." *Food Policy* 107: 102212.

Bahru, Bezawit Adugna, Mulusew G Jebena, Regina Birner, and Manfred Zeller. 2020. "Impact of Ethiopia's Productive Safety Net Program on Household Food Security and Child Nutrition: A Marginal Structural Modeling Approach." *SSM - Population Health* 12: 100660. https://doi.org/https://doi.org/10.1016/j.ssmph.2020.100660.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil. 2022. "Why Do People Stay Poor?" *The Quarterly Journal of Economics* 137 (2): 785–844.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor Markets and Poverty in Village Economies." *The Quarterly Journal of Economics* 132 (2): 811–70.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236): 1260799.

Banerjee, Abhijit, Esther Duflo, and Garima Sharma. 2021. "Long-Term Effects of the Targeting the Ultra Poor Program." *American Economic Review: Insights* 3 (4): 471–86.

Banerjee, Abhijit, Dean Karlan, Robert Osei, Hannah Trachtman, and Christopher Udry. 2022. "Unpacking a Multi-Faceted Program to Build Sustainable Income for the Very Poor." *Journal of Development Economics* 155: 102781.

Banerjee, Abhijit V, and Esther Duflo. 2007. "The Economic Lives of the Poor." *Journal of Economic Perspectives* 21 (1): 141–67.

Bedoya, Guadalupe, Aidan Coville, Johannes Haushofer, Mohammad Isaqzadeh, and Jeremy Shapiro. 2019. "No Household Left Behind Afghanistan: Targeting the Ultra Poor Impact Evaluation."

Bellemare, Marc F, and Casey J Wichman. 2020. “Elasticities and the Inverse Hyperbolic Sine Transformation.” *Oxford Bulletin of Economics and Statistics* 82 (1): 50–61.

Berhane, Guush, Daniel O Gilligan, John Hoddinott, Neha Kumar, and Alemayehu Seyoum Taffesse. 2014. “Can Social Protection Work in Africa? The Impact of Ethiopia’s Productive Safety Net Programme.” *Economic Development and Cultural Change* 63 (1): 1–26. https://doi.org/10.1086/677753.

Berhane, Guush, K Hirvonen, John Hoddinott, Neha Kumar, A Seyoum Taffesse, Yisehac Yohannes, M Tefera, B Nishan, J Lind, and R Sabates-Wheeler. 2013. “The Implementation of the Productive Safety Nets Programme and the Household Asset Building Programme in the Ethiopian Highlands, 2012: Program Performance Report.” *Unpublished, International Food Policy Research Institute, Washington, DC*.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. “Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda.” *The Quarterly Journal of Economics* 129 (2): 697–752.

———. 2020. “The Long-Term Impacts of Grants on Poverty: Nine-Year Evidence from Uganda’s Youth Opportunities Program.” *American Economic Review: Insights* 2 (3): 287–304.

Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand, Catherine C Thomas, Christopher Udry, and Julia Vaillant. 2022. “Tackling Psychosocial and Capital Constraints to Alleviate Poverty.” *Nature* 605 (7909): 291–97.

Brauw, Alan de, Valerie Mueller, and Tassew Woldehanna. 2013. “Motives to Remit: Evidence from Tracked Internal Migrants in Ethiopia.” *World Development* 50: 13–23. https://doi.org/https://doi.org/10.1016/j.worlddev.2013.04.008.

———. 2018. “Does Internal Migration Improve Overall Well-Being in Ethiopia?†.” *Journal of African Economies* 27 (3): 347–65. https://doi.org/10.1093/jae/ejx026.

Brune, Lasse, Dean Karlan, Sikandra Kurdi, and Christopher Udry. 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.

Chen, Jiafeng, and Jonathan Roth. 2022. “Logs with Zeros? Some Problems and Solutions.”

Deaton, Angus. 2019. “The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.” Washington, D.C.: World Bank.

Gazeaud, Jules, and Victor Stephane. 2020. “Productive Workfare? Evidence from Ethiopia’s Productive Safety Net Program.” *American Journal of Agricultural Economics*.

Gilligan, Daniel O, John Hoddinott, and Alemayehu Seyoum Taffesse. 2009. “The Impact of Ethiopia’s Productive Safety Net Programme and Its Linkages.” *The Journal of Development Studies* 45 (10): 1684–1706. https://doi.org/10.1080/00220380902935907.

Haushofer, Johannes, and Jeremy Shapiro. 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.

———. 2018. "The Long-Term Impact of Unconditional Cash Transfers: Experimental Evidence from Kenya."

Hirvonen, Kalle, and John Hoddinott. 2021. "Beneficiary Views on Cash and In-Kind Payments: Evidence from Ethiopia's Productive Safety Net Programme." *The World Bank Economic Review* 35 (2): 398–413. https://doi.org/10.1093/wber/lhaa002.

Hoddinott, John, Guush Berhane, Daniel O Gilligan, Neha Kumar, and Alemayehu Seyoum Taffesse. 2012. "The Impact of Ethiopia's Productive Safety Net Programme and Related Transfers on Agricultural Productivity." *Journal of African Economies* 21 (5): 761–86. https://doi.org/10.1093/jae/ejs023.

Hoddinott, John, and Tseday J Mekasha. 2020. "Social Protection, Household Size, and Its Determinants: Evidence from Ethiopia." *The Journal of Development Studies* 56 (10): 1818–37. https://doi.org/10.1080/00220388.2020.1736283.

IDInsight. 2018. "EthioChicken Impact Evaluation: Understanding Income and Nutritional Effects of Expanding Access to High Productivity Chickens for Ethiopian Smallholder Farmers in the Early Days of Ownership."

Jodlowski, Margaret, Alex Winter-Nelson, Kathy Baylis, and Peter D Goldsmith. 2016. "Milk in the Data: Food Security Impacts from a Livestock Field Experiment in Zambia." *World Development* 77: 99–114. https://doi.org/https://doi.org/10.1016/j.worlddev.2015.08.009.

Kebede, Bereket. 2008. "Land Reform, Distribution of Land and Institutions in Rural Ethiopia: Analysis of Inequality with Dirty Data." *Journal of African Economies* 17 (4): 550–77.

Leight, Jessica, Daniel O Gilligan, Michael Mulford, Alemayehu Seyoum Taffesse, and Heleene Tambet. 2021. *Aspiring to More? New Evidence on the Effect of a Light-Touch Aspirations Intervention in Rural Ethiopia*. Vol. 2070. Intl Food Policy Res Inst.

Loeser, John, and Florence Kondylis. 2021. "Intervention Size and Persistence." No. 9769. Policy Research Working Paper.

Miller, Laurie C, Neena Joshi, Mahendra Lohani, Beatrice Rogers, Meghan Loraditch, Robert Houser, Padma Singh, and Shubh Mahato. 2014. "Community Development and Livestock Promotion in Rural Nepal: Effects on Child Growth and Health." *Food and Nutrition Bulletin* 35 (3): 312–26.

Mueller, Valerie, Cheryl Doss, Agnes Quisumbing, and null null. 2018. "Youth Migration and Labour Constraints in African Agrarian Households." *The Journal of Development Studies* 54 (5): 875–94. https://doi.org/10.1080/00220388.2018.1430770.

Mullally, Conner, Mayra Rivas, and Travis McArthur. 2021. "Using Machine Learning to Estimate the Heterogeneous Effects of Livestock Transfers." *American Journal of Agricultural Economics* n/a (n/a).

Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich. 2023. "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments." *The Review of Economics and Statistics*, March, 1–44. https://doi.org/10.1162/rest_a_01317.

Nagler, Paula, and Wim Naudé. 2017. "Non-Farm Entrepreneurship in Rural Sub-Saharan Africa: New Empirical Evidence." *Food Policy* 67: 175–91. https://doi.org/https://doi.org/10.1016/j.foodpol.2016.09.019.

Phadera, Lokendra, Hope Michelson, Alex Winter-Nelson, and Peter Goldsmith. 2019. "Do Asset Transfers Build Household Resilience?" *Journal of Development Economics* 138: 205–27. https://doi.org/https://doi.org/10.1016/j.jdeveco.2019.01.003.

Porter, Catherine, and Radhika Goyal. 2016. "Social Protection for All Ages? Impacts of Ethiopia's Productive Safety Net Program on Child Nutrition." *Social Science & Medicine* 159: 92–99. https://doi.org/https://doi.org/10.1016/j.socscimed.2016.05.001.

Rawlins, Rosemary, Svetlana Pimkina, Christopher B Barrett, Sarah Pedersen, and Bruce Wydick. 2014. "Got Milk? The Impact of Heifer International's Livestock Donation Programs in Rwanda on Nutritional Outcomes." *Food Policy* 44: 202–13. https://doi.org/https://doi.org/10.1016/j.foodpol.2013.12.003.

Rijkers, Bob, Måns Söderbom, and Josef L Loening. 2010. "A Rural–Urban Comparison of Manufacturing Enterprise Performance in Ethiopia." *World Development* 38 (9): 1278–96. https://doi.org/https://doi.org/10.1016/j.worlddev.2010.02.010.

Simes, R J. 1986. "An Improved Bonferroni Procedure for Multiple Tests of Significance." *Biometrika* 73 (3): 751–54. https://doi.org/10.1093/biomet/73.3.751.

Subramanian, Shankar, and Angus Deaton. 1996. "The Demand for Food and Calories." *Journal of Political Economy* 104 (1): 133–62.

United Nations Development Program. 2023. "Quarterly Economic Profile: Ethiopia."

World Bank. 2019. "Poverty & Household Welfare in Ethiopia 2011-2016." Washington, DC.

———. 2022. "Project Paper on a Proposed Second Additional Grant; Report No: PAD5072." Https://Documents1.Worldbank.Org/Curated/En/099745008162221711/Pdf/BOSIB0c263dc410b90b4680ff886901cf67.Pdf. 2022.

Figures and Tables

Figure 1: Experimental design

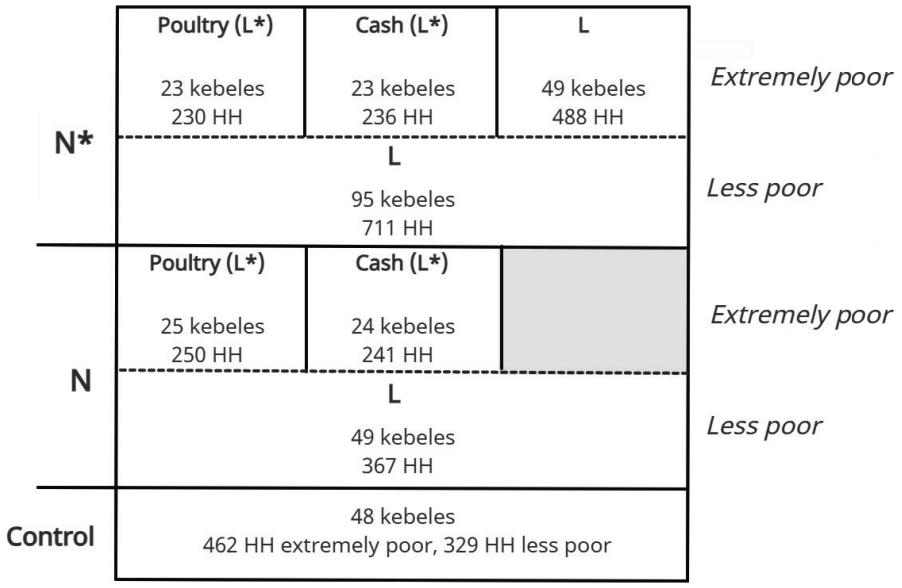


Figure 2: Timeline

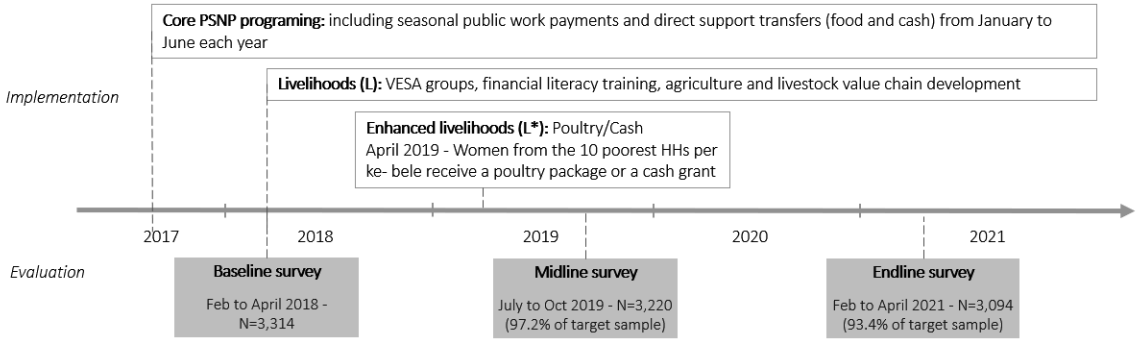


Figure 3: Primary treatment effects at endline

Figure 3a: Extremely poor, short model

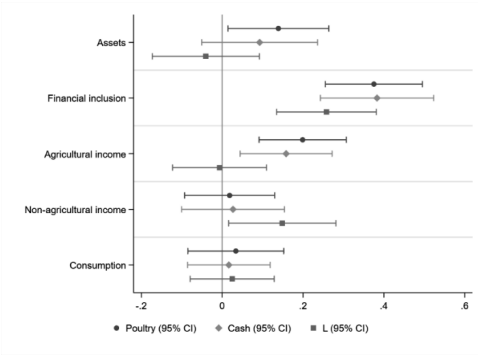
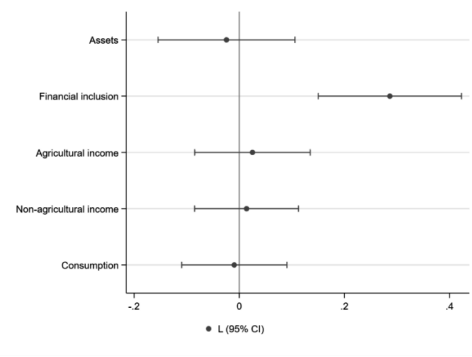


Figure 3b: Less poor, short



Notes: This figure reports the average standard treatment effects for each primary outcome family in conjunction with 95% confidence intervals. Coefficient estimates for the extremely poor are reported in Figure 3a, and for the less poor are reported in Figure 3b.

Table 1: Balance in baseline characteristics

	(1) N	(2) Control mean	(3) Poultry mean	(4) Cash mean	(5) L mean	(6) P-value on joint test
Panel A: Extremely poor households						
Household size	1,907	5.40	5.66	5.47	5.27	0.084*
Female-headed household	1,907	0.29	0.22	0.27	0.31	0.138
Household head: married, monogamous	1,905	0.75	0.79	0.75	0.73	0.305
Household head has no formal education	1,907	0.72	0.72	0.76	0.73	0.203
Value of livestock, productive assets, and consumer durables (market prices)	1,907	1,030.52	1,156.24	1,061.29	962.24	0.188
Total value of all productive assets	1,907	54.65	60.91	59.41	53.02	0.453
Total value of all consumer durable assets	1,907	88.21	94.68	94.67	94.42	0.995
Estimated value of all livestock owned by the household (market prices)	1,907	887.67	995.96	907.22	810.31	0.189
Household took out any type of loan (past year)	1,416	0.22	0.22	0.21	0.18	0.753
Household has any savings	1,465	0.22	0.27	0.26	0.26	0.993
Household reports any non-agricultural business	1,465	0.05	0.03	0.03	0.06	0.262
Household reports any regular wage work (past year)	1,465	0.06	0.03	0.05	0.03	0.508
Consumption expenditure per month per adult equivalent	1,880	59.44	60.63	64.72	62.46	0.894
	(1) N	(2) Control mean	(3) Pooled treatment mean			
Panel B: Less poor households						
Household size	1,407	6.23	6.18			0.808
Female-headed household	1,406	0.06	0.08			0.309
Household head: married, monogamous	1,405	0.95	0.94			0.336
Household head has no formal education	1,407	0.71	0.69			0.817
Value of livestock, productive assets and consumer durables (market prices)	1,407	2,410.23	2,215.95			0.436
Total value of all productive assets	1,407	141.36	145.11			0.683
Total value of all consumer durable assets	1,407	133.49	148.49			0.810
Estimated value of all livestock owned by the household (market prices)	1,407	2,096.09	1,901.01			0.409
Household took out any type of loan (past year)	1,313	0.18	0.21			0.865
Household has any savings	1,342	0.35	0.36			0.807
Household reports any non-agricultural business	1,342	0.05	0.05			0.642
Household reports any regular wage work (past year)	1,342	0.04	0.04			0.848
Consumption expenditure per month per adult equivalent	1,395	61.13	64.11			0.713

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The final column reports a p-value derived from a regression in which the covariate of interest is regressed on the binary variables for treatment assignment, conditional on strata dummies; the p-value corresponds to the joint test of the hypothesis $\beta_1 = \beta_2 = \beta_3 = 0$, in Panel A, or the hypothesis that $\beta=0$ in Panel B. All continuous variables are reported in 2017 dollars adjusted for purchasing power parity.

Table 2: Average standard treatment effects, no interactions

	(1) Assets	(2) Financial inclusion	(3) Agricultural and livestock income	(4) Non-agricultural income	(5) Consumption and food security
Panel A: Extremely poor households					
Poultry	0.139**	0.375***	0.272***	0.018	0.034
p-value	(0.029)	(0.000)	(0.000)	(0.746)	(0.577)
q-value	[0.062]	[0.000]	[0.000]	[0.751]	[0.751]
Cash	0.093	0.383***	0.242***	0.027	0.016
p-value	(0.205)	(0.000)	(0.000)	(0.680)	(0.751)
q-value	[0.385]	[0.000]	[0.000]	[0.751]	[0.751]
L	-0.041	0.258***	0.063	0.149**	0.025
p-value	(0.549)	(0.000)	(0.295)	(0.028)	(0.641)
q-value	[0.751]	[0.000]	[0.492]	[0.062]	[0.751]
Test: Poultry = Cash	0.481***	0.900***	0.653***	0.892***	0.751***
Test: Cash = L	0.050***	0.046***	0.005***	0.096***	0.856***
Test: Poultry = L	0.003	0.023	0.002	0.047	0.868
N	1,765	1,765	1,765	1,030	1,771
Panel B: Less poor households					
L	-0.024	0.286***	0.066	0.014	-0.009
p-value	(0.714)	(0.000)	(0.285)	(0.783)	(0.853)
q-value	[0.853]	[0.000]	[0.713]	[0.853]	[0.853]
N	1,322	1,322	1,322	951	1,323

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. All average standard treatment effect estimates are calculated following the method of Katz, Kling and Liebman (2007) and present the effect size relative to the standard deviation of the control arm. All treatment effects are estimated conditional on strata fixed effects; both conventional p-values clustered at the subdistrict level and sharpened q-values adjusted for multiple hypothesis testing are reported. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the conventional standard errors and p-values.

Table 3: Average standard treatment effects, interactions with respect to enhanced nutrition

	(1) Assets	(2) Financial inclusion	(3) Agricultural and livestock income	(4) Non-agricultural income	(5) Consumption and food security
Panel A: Extremely poor households					
Poultry	0.151**	0.321***	0.264***	0.029	0.051
p-value	(0.024)	(0.000)	(0.001)	(0.690)	(0.498)
q-value	[0.075]	[0.000]	[0.005]	[0.877]	[0.877]
Cash	0.217***	0.434***	0.264***	-0.004	0.023
p-value	(0.001)	(0.000)	(0.001)	(0.958)	(0.701)
q-value	[0.005]	[0.000]	[0.005]	[0.958]	[0.877]
Poultry x EN	-0.023	0.108	0.016	-0.022	-0.034
p-value	(0.774)	(0.120)	(0.872)	(0.779)	(0.700)
q-value	[0.886]	[0.300]	[0.909]	[0.886]	[0.877]
Cash x EN	-0.251***	-0.101	-0.044	0.068	-0.014
p-value	(0.007)	(0.323)	(0.611)	(0.494)	(0.822)
q-value	[0.026]	[0.673]	[0.877]	[0.877]	[0.893]
L	-0.039	0.259***	0.064	0.148**	0.025
p-value	(0.569)	(0.000)	(0.292)	(0.029)	(0.640)
q-value	[0.877]	[0.000]	[0.665]	[0.081]	[0.877]
Test: Poultry = L	0.291	0.187	0.995	0.699	0.719
Test: Cash = L	0.000	0.026	0.015	0.063	0.978
Test: Poultry = L	0.004	0.305	0.018	0.133	0.711
N	1,765	1,765	1,765	1,030	1,771
Panel B: Less poor households					
L	-0.025	0.306***	0.063	-0.035	-0.000
p-value	(0.749)	(0.000)	(0.370)	(0.523)	(0.997)
q-value	[0.997]	[0.001]	[0.997]	[0.997]	[0.997]
L x EN	0.001	-0.030	0.005	0.073	-0.014
p-value	(0.991)	(0.576)	(0.930)	(0.101)	(0.786)
q-value	[0.997]	[0.997]	[0.997]	[0.504]	[0.997]
N	1,322	1,322	1,322	951	1,323

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. All average standard treatment effect estimates are calculated following the method of Katz, Kling and Liebman (2007) and present the effect size relative to the standard deviation of the control arm. Both conventional p-values clustered at the subdistrict level and sharpened q-values adjusted for multiple hypothesis testing are reported. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the conventional standard errors and p-values.

Table 4: Assets

	One year follow-up	*Three year follow-up*			
	(1) Value of livestock	(2) Value of total assets	(3) Value of productive assets	(4) Value of consumer durables	(5) Value of livestock
	(IHS transformation)				
Panel A: Extremely poor households					
Poultry	1.084***	0.111	0.150	0.038	0.428**
	(0.174)	(0.103)	(0.103)	(0.096)	(0.173)
Cash	0.508**	0.059	0.071	0.008	0.359*
	(0.226)	(0.115)	(0.122)	(0.084)	(0.204)
L	-0.190	-0.244**	0.015	-0.020	-0.314
	(0.206)	(0.110)	(0.101)	(0.090)	(0.198)
Test: Poultry = Cash	0.004***	0.622	0.515	0.756	0.714
Test: Poultry = L	0.000***	0.001***	0.180	0.554	0.000***
Test: Cash = L	0.002***	0.008***	0.630	0.736	0.001***
Test: β_1 1 year = 3 years [poultry]					0.000
Test: β_2 1 year = 3 years [cash]					0.432
Test: β_3 1 year = 3 years [L]					0.519
Mean of control	1,453.86	1,105.66	35.21	96.54	971.87
N	1,847	1,765	1,765	1,765	1,765
Panel B: Less poor households					
L	0.148	-0.074	0.044	0.167**	-0.233
	(0.193)	(0.091)	(0.078)	(0.080)	(0.154)
Test: β 1 year = 3 years					0.018
Mean of control	2,168.13	1,653.85	45.41	100.84	1,495.62
N	1,373	1,322	1,322	1,322	1,322

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The dependent variables are calculated as the value of the specified asset class, valued using locally reported market prices; the means in the control arm are reported in 2017 dollars in purchasing power parity terms. The estimated regressions employ an inverse hyperbolic sine transformation. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table 5: Financial inclusion; one year follow-up

	(1) Any credit (past year)	(2) Any formal credit (past year)	(3) Total credit (past year, IHS transformation)	(4) Any savings	(5) Total savings (IHS transformation)
Panel A: Extremely poor households					
Poultry	0.024	0.014	0.177	0.318***	1.406***
	(0.043)	(0.036)	(0.261)	(0.041)	(0.186)
Cash	-0.006	0.008	0.011	0.349***	1.362***
	(0.043)	(0.036)	(0.276)	(0.038)	(0.197)
L	0.064	0.046	0.450	0.260***	1.144***
	(0.044)	(0.036)	(0.277)	(0.042)	(0.210)
Test: Poultry = Cash	0.462	0.839	0.500	0.368	0.821
Test: Poultry = L	0.321	0.326	0.257	0.134	0.204
Test: Cash = L	0.088*	0.227	0.087*	0.013**	0.302
Mean of control	0.42	0.17	156.35	0.44	40.98
N	1,291	1,285	1,290	1,289	1,296
Panel B: Less poor households					
L	0.046	0.014	0.211	0.278***	1.018***
	(0.037)	(0.032)	(0.247)	(0.034)	(0.175)
Mean of control	0.436	0.194	184.986	0.504	74.990
N	1,182	1,176	1,182	1,182	1,185

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ an inverse hyperbolic sine transformation. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table 6: Financial inclusion; three year follow-up

	(1) Any credit (past year)	(2) Any formal credit (past year)	(3) Total credit (past year, IHS transformation)	(4) Any savings	(5) Total savings (IHS transformation)
Panel A: Extremely poor households					
Poultry	0.075**	0.077**	0.516**	0.335***	1.404***
	(0.036)	(0.031)	(0.231)	(0.046)	(0.215)
Cash	0.098**	0.083***	0.620**	0.333***	1.317***
	(0.040)	(0.032)	(0.250)	(0.046)	(0.231)
L	-0.006	0.032	0.016	0.327***	1.275***
	(0.034)	(0.030)	(0.216)	(0.043)	(0.213)
Test: Poultry = Cash	0.576	0.839	0.692	0.955	0.647
Test: Poultry = L	0.019**	0.112	0.027**	0.799	0.450
Test: Cash = L	0.007***	0.083*	0.014**	0.845	0.819
Test: β_1 1 year = 3 years [poultry]	0.297	0.112	0.252	0.718	0.993
Test: β_2 1 year = 3 years [cash]	0.016	0.027	0.019	0.737	0.848
Test: β_3 1 year = 3 years [L]	0.129	0.690	0.118	0.131	0.555
Mean of control	0.45	0.13	147.03	0.40	34.07
N	1,765	1,760	1,765	1,765	1,765
Panel B: Less poor households					
L	0.019	0.036	0.196	0.302***	1.396***
	(0.042)	(0.036)	(0.268)	(0.042)	(0.204)
Test: β 1 year = 3 years	0.567	0.561	0.959	0.602	0.085
Mean of control	0.48	0.19	170.37	0.47	34.73
N	1,322	1,320	1,321	1,322	1,322

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ an inverse hyperbolic sine transformation. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table 7: Income, three-year follow-up

	(1) Income from sales of livestock (past year)	(2) Income from sales of livestock products (past year)	(3) Income from crops (past year)	(4) Total estimated value of livestock products produced (past year)	(5) Total estimated value of crops harvested (past year)	(6) Total income from livestock and crops	(7) Total estimated value of livestock sales and livestock and crop products	(8) Any income from non-agricultural business	(9) Any income from formal wage work (past year)	(10) Any income from casual wage work (past year)
	(IHS transformation)				(IHS transformation)					
Panel A: Extremely poor households										
Poultry	0.787***	0.508***	0.226	0.458**	0.421	0.678***	0.557**	0.014	-0.007	0.025
	(0.243)	(0.141)	(0.198)	(0.188)	(0.267)	(0.256)	(0.218)	(0.017)	(0.018)	(0.047)
Cash	0.791***	0.408***	0.146	0.310*	0.294	0.668**	0.341	0.018	-0.008	0.020
	(0.243)	(0.128)	(0.201)	(0.174)	(0.309)	(0.258)	(0.262)	(0.021)	(0.019)	(0.043)
L	0.075	0.172	-0.289	0.166	-0.226	-0.114	-0.114	0.044**	0.016	0.045
	(0.242)	(0.119)	(0.224)	(0.169)	(0.280)	(0.277)	(0.250)	(0.020)	(0.018)	(0.045)
Test: Poultry = Cash	0.987	0.514	0.665	0.425	0.664	0.967	0.348	0.835	0.962	0.903
Test: Poultry = L	0.006***	0.023**	0.015**	0.108	0.016**	0.004***	0.003***	0.168	0.141	0.670
Test: Cash = L	0.005***	0.078*	0.047**	0.388	0.090*	0.004***	0.076*	0.325	0.172	0.563
Test: β_1 1 year = 3 years [poultry]									0.650	0.358
Test: β_2 1 year = 3 years [cash]									0.505	0.441
Test: β_3 1 year = 3 years [L]									0.623	0.843
Mean of control	266.87	6.48	82.53	27.76	475.02	360.17	773.88	0.03	0.04	0.26
N	1,765	1,761	1,765	1,762	1,765	1,765	1,765	1,030	1,029	1,030
Panel B: Less poor households										
L	0.224	0.147	-0.424*	0.053	0.166	0.032	0.246	-0.019	0.008	0.033
	(0.255)	(0.171)	(0.215)	(0.176)	(0.253)	(0.275)	(0.217)	(0.017)	(0.016)	(0.033)
Test: β 1 year = 3 years									0.914	0.956
Mean of control	373.76	-39.87	126.95	49.41	639.40	465.83	1,067.56	0.06	0.03	0.20
N	1,322	1,319	1,322	1,322	1,322	1,322	1,322	951	951	951

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ an inverse hyperbolic sine transformation. The estimate of total income in Column (6) corresponds to the sum of Columns (1) through (3), and the estimate of total crop and product value in Column (7) corresponds to the sum of Columns (1), (4), and (5). All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table 8: Consumption and food security, three year follow-up

	(1) Monthly food consumption per adult equivalent (IHS transformation)	(2) Monthly non-food consumption per adult equivalent (IHS transformation)	(3) Monthly consumption per adult equivalent (IHS transformation)	(4) Food Insecurity Experience Scale (0-8)
Panel A: Extremely poor households				
Poultry	0.032	0.020	0.025	0.102
	(0.068)	(0.084)	(0.062)	(0.199)
Cash	-0.085	0.052	-0.059	0.517**
	(0.069)	(0.073)	(0.061)	(0.208)
L	-0.071	0.140*	-0.038	0.226
	(0.067)	(0.077)	(0.060)	(0.175)
Test: Poultry = Cash	0.091*	0.646	0.181	0.057*
Test: Poultry = L	0.120	0.096*	0.309	0.507
Test: Cash = L	0.842	0.140	0.722	0.139
Mean of control	44.87	7.68	52.86	3.36
N	1,706	1,764	1,701	1,748
Panel B: Less poor households				
L	-0.069	0.074	-0.036	0.006
	(0.075)	(0.060)	(0.064)	(0.192)
Mean of control	47.878	7.840	55.872	3.318
N	1,291	1,322	1,290	1,314

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The dependent variables in Columns (1) through (3) are household-level consumption variables; the dependent variable in Column (4) is the continuous FIES score. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ an inverse hyperbolic sine transformation. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Appendix

A.1 Randomization

This description of the randomization process draws substantially on the baseline report. Within the 196 study kebeles that were randomized, 49 were assigned to each of the four treatment arms: T1: L*+EN; T2: L*+N; T3: L+EN; and C: PSNP only. Randomization was stratified at the district (woreda) level to provide balance of treatment assignment geographically, and the first version of the randomization was conducted based on subdistrict (kebele)-level implementation data provided in September 2017. Kebeles in which VESA groups had already formed were removed from the sample, leading to a total of 158 clusters. In November 2017, a second dataset was provided by the implementing partners including an additional woreda in Oromia, Daro Lebu, where implementation would start a few months later, which added 37 additional clusters to the study. In addition, original kebele level data on VESA group formation was incorrect in 4 of the Oromia kebeles. As a result, it was agreed to redo the randomization for Oromia region only, since VESA program formation had already begun in Amhara based on the initial randomization. In the second dataset, VESA programs in Oromia that were marked as "very new" were now kept in the eligible list of kebeles. Thus, the original randomization for the 115 kebeles in Amhara was retained and the new 81 kebeles in Oromia were re-randomized, ending up with a total of 196 clusters.

Using the initial data, 1,000 potential treatment allocations were generated, stratified by woreda. A uniform random number between 0 and 1 was drawn for each of the clusters in each stratum. These were ordered and then allocated 1/4th of the sample to each treatment arm. Since some strata did not consist of clusters that were evenly divisible by four, we randomly allocated the leftover clusters within each stratum to one of the treatment arms ensuring that balance across arms within strata would be preserved (i.e., no treatment arms gets more than one leftover cluster within the strata) and that the allocation would be random. When there was one leftover cluster, a random number between 0 and 1 was drawn, and if it was less than 0.25 it was allocated to treatment group 1, if it was between 0.25 and 0.5 it was allocated to treatment group 2, if between 0.5 and 0.75 it was allocated to treatment group 3, and if between 0.75 and 1 it was allocated to treatment group 4. With two leftover clusters, there were 6 possible allocations across the 4 treatment groups (4 choose 2). Again, we drew a random number between 0 and 1 at the strata level, and if this was less than 0.1667 then the clusters go in treatment groups 1 and 2, if between 0.1667 and 0.3333 then the clusters go in treatment groups 1 and 3, and so on. A second random number is selected and ranked to decide the order of the allocation to each of the treatment arms. The same procedure was followed for strata with 3 leftover clusters.

Using the share of PSNP beneficiaries in each kebele and the distance from the kebele to the district capital to balance the treatment arms, the relative efficiency of each of the 1,000 potential allocations was calculated. For any treatment allocation, the relative efficiency provides a measure of the balance in observable characteristics between potential treatment groups. The maximum t-statistic from the regression of the observed characteristic on the treatment allocations (with strata dummies) is calculated at the sample level. Allocations with the most equal allocations across regions were kept from these 1,000 allocations – that is, allocations with more than 29 kebeles per treatment arm in Amhara were dropped, and allocations with more than

12 kebeles per treatment arm in Oromia were dropped. At the sample level, allocations that resulted in less than 39 clusters in each treatment arm were also dropped. From the remaining allocations, the one with the highest relative efficiency – the minimum maximum t-statistic - was retained (Bruhn and McKenzie 2009). This allocation is used as the final randomization allocation for Amhara.

In the second set of data received on November 1, the procedure was modified to take as given the previous assignment of kebeles to treatment groups in Amhara. For each stratum in Oromia, 1,000 potential treatment allocations were generated using the same procedure that was used in the initial randomization; leftover clusters within each stratum were also managed similarly. For each of the 1,000 potential treatment allocations generated for Oromia in the second set of randomizations, the relative efficiency is calculated using the potential treatment allocation for previously unassigned kebeles in Oromia and the actual treatment assignment for kebeles in Amhara. Allocations with the most equal balance across Oromia were kept – that is allocations that resulted in a treatment group with 19 or fewer clusters were dropped. At the sample level, allocations that resulted in 49 clusters per treatment arm were kept. From the remaining allocations, the one with the highest relative efficiency – the minimum maximum t-statistic - was retained. This allocation is used as the final randomization allocation, maintaining the original Amhara randomization and combining it with the new randomization for Oromia (Table A22).

Following the initial randomization of kebeles across the four treatment arms, two kebeles were dropped because they had no PSNP beneficiaries and one was dropped for security reasons. The randomization across the 193 remaining kebeles is shown in Table A23.

Within the 95 L* kebeles assigned to the treatment arms T1: L* + EN and T2: L*+N, half were randomized, by woreda, to (1) receive aspirations treatment; and (2) not receive aspirations treatment. Again, 1000 potential treatment allocations were calculated, by woreda, of which only allocations with balance between the number of kebeles in the aspirations treatment in L*+EN and L*+N treatment groups, were kept. Of the allocations that remained, a random number between 0 and 1 was assigned to each randomization and the one with the lowest random number assigned was kept as the realized allocation.

The L* kebeles were also randomized into poultry and cash treatment arms. Again, 1000 potential allocations were drawn, stratified by woreda. Only the allocations with balance between the number of poultry treatment kebeles in L*+EN and L*+N treatment groups, as well as balance between the number of poultry treatment kebeles in L*+EN aspirations vs non-aspirations, and L*+N aspirations vs non-aspirations groups were retained. Of the allocations that remained, a random number between 0 and 1 was assigned to each randomization and the one with the lowest random number assigned was retained as the realized allocation.

The resulting final kebele-level randomization is presented in Tables A15 and A16.

A.2 Measuring assets and income

We measure assets and income as follows.

To construct estimates of asset value, we use count values of assets owned provided by households, in conjunction with data on asset prices obtained from local markets. For livestock,

the price data was obtained in a market survey administered at the same time as the three year follow-up survey (in 2021) and is adjusted for inflation to correspond to 2018 prices at baseline. For non-livestock assets, the price data was obtained from the midline survey corresponding to the main IFPRI evaluation of the PSNP (PSNP4 Highland Outcomes Report); this survey was conducted in 2018, and thus no inflation adjustment is required.

To construct income data, we use data that is directly reported by households about their income from sales of livestock products, livestock, and crops in the most recent season; as well as data about their overall output of livestock products and crops. Output and sales of different livestock products are reported for shorter recall periods varying from one week to 3 months, and these reports are imputed to production over the past year assuming consistent production over the year; sales of livestock are directly reported for the past year.

In the two follow-up surveys, output and sales for crops are reported for the mehr season, generally the primary season for production in this region. A secondary growing season, the belg, is characterized by relatively low levels of production; at baseline, belg production constituted less than 10% of total annual production on average, though with considerable variation across woredas. (Woredas in Oromia in which chat production is common are generally characterized by more output in the belg season.) We adjust our total estimates of production upward to encompass the belg season using a woreda-level adjustment factor constructed from baseline data.

Estimates of the total value of livestock products and the total crop value are calculated by valuing each reported crop produced at the local market price. In all cases, we employ price data from the market most proximate to each subdistrict. If a particular market did not report price data for a particular commodity, we replace that missing value with the average price reported in that woreda, or if not reported in that woreda, in that region.

A.3 Comparing consumption and transfers in different graduation models

Table A2 presents data on consumption per capita and transfer size comparing across various graduation model interventions evaluated in the existing literature. This table is constructed as follows. First, data on transfer size and consumption was compiled from the papers published. Importantly, for interventions that include both a lump-sum asset or cash transfer and ongoing cash support in the treatment arm that are not provided to the control arm, we calculate the total sum of both components. (Thus a graduation model that includes a $500 asset transfer and two years of monthly consumption support payments of $25 a month, neither of which are provided to the control arm, would be described as a $1100 transfer here.)

Second, all estimates are then converted to 2017 dollars in purchasing power terms. We do this by identifying the value of the transfer in the original currency in the year recorded, converting this currency value to a 2017 currency value (using exchange rates reported by the World Bank), and then converting to purchasing power parity terms using data from the International Comparison Program.

Figure A1: Study design

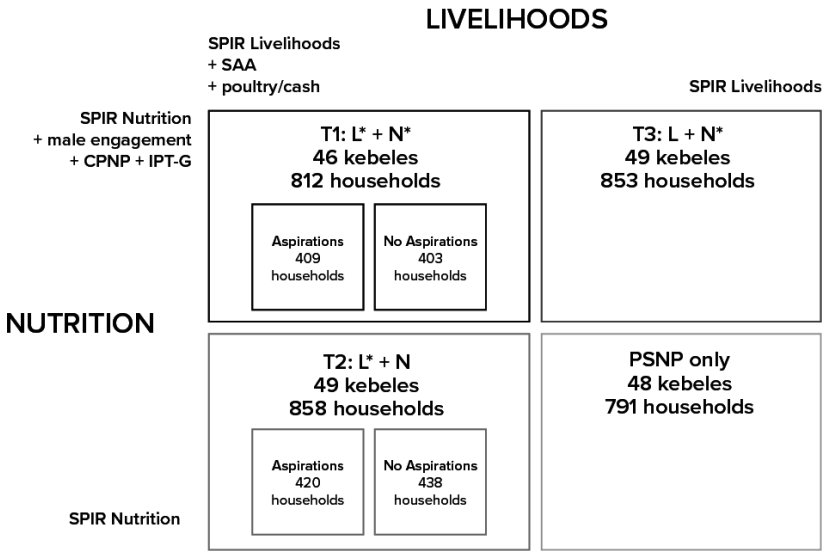


Figure A2: Flow diagram

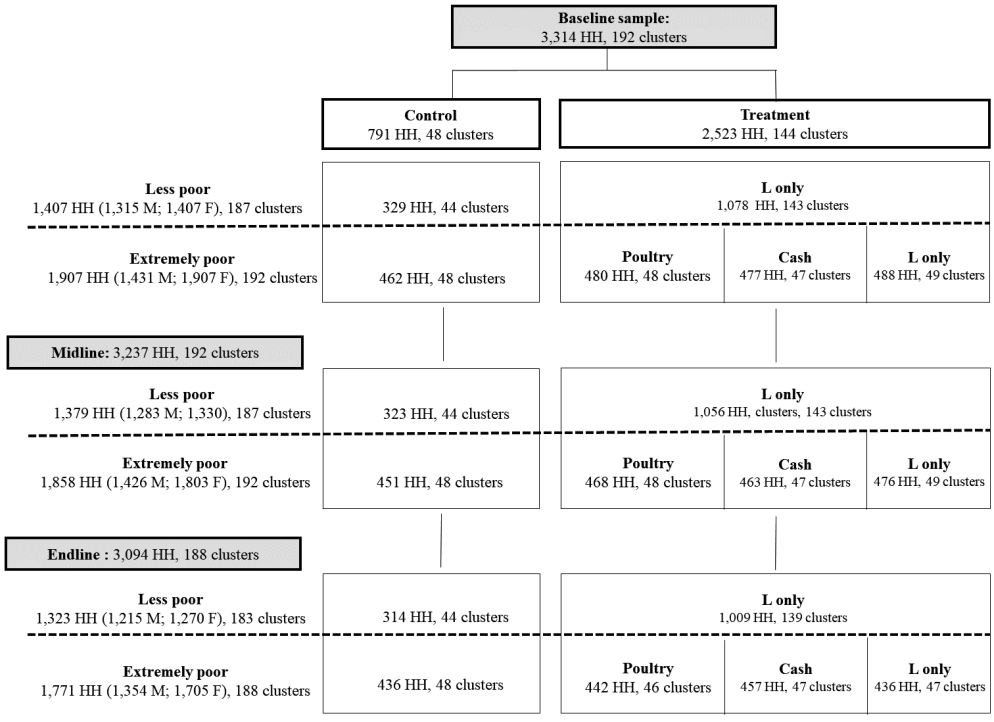
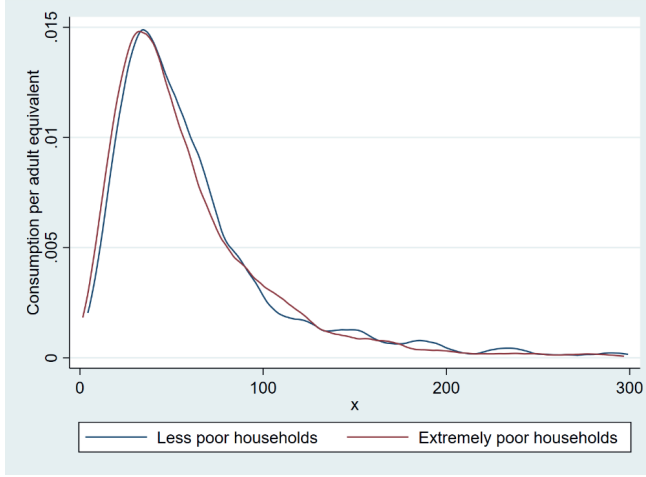


Figure A3: Baseline consumption for extremely poor versus non-poor households



Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The graph truncates consumption per adult equivalent at approximately the 99th percentile observed in the pooled sample.

Figure A4: Primary treatment effects, three year follow-up

Figure A4a: Extremely poor, long model

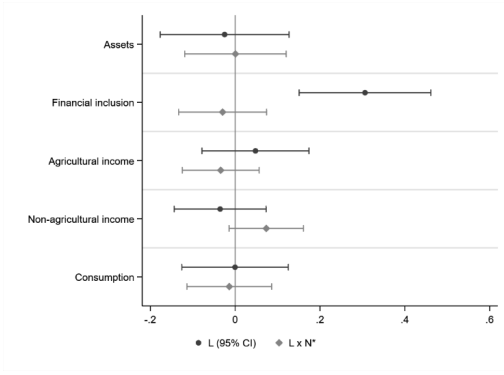
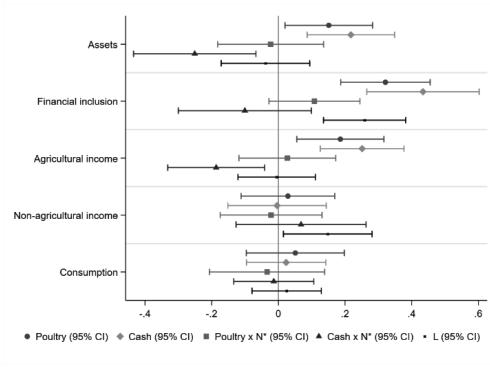


Figure A4b: Less poor, long model



Notes: This figure reports the average standard treatment effects for each primary outcome family in conjunction with 95% confidence intervals. Coefficient estimates for the extremely poor are reported in Figure 3a, and for the less poor are reported in Figure 3b.

Table A1: Transfer size in the graduation model literature

Author	Context	Year project launched	Monthly per capita consumption (PPP $2017)	Monthly household consumption (PPP $2017)	Transfer value (PPP $2017)	Transfer / consumption	Intervention population (number of households)
Bossuroy et al. 2022	Niger	2016	71.30	NR	296.85	4.16	22,507
Bandiera et al. 2017	Bangladesh	2007	78.43	NR	629.66	8.03	360,000
Alderman et al. 2022	Ethiopia	2019	47.30	272.94	374.20	7.91	150,000
Brune et al. 2022	Yemen	2010	160.83	1254.50	1600.02	9.95	505
Banerjee et al. 2015	Peru	2011	99.51	517.46	1095.57	11.01	785
Banerjee et al. 2015	Ghana	2011	38.18	318.42	481.06	12.60	666
Banerjee et al. 2015	Pakistan	2008	99.53	626.07	1537.88	15.45	660
Banerjee et al. 2015	Honduras	2009	49.94	293.67	781.47	15.65	800
Banerjee et al. 2015	India	2007	40.56	160.63	712.01	17.55	541
Bedoya et al. 2019	Afghanistan	2016	87.70	NR	1960.00	22.35	1,219
Banerjee et al. 2015	Ethiopia	2010	50.10	280.07	1295.55	25.86	458

Notes: This table summarizes characteristics of the sample and the transfer included in other recent papers in the graduation model literature. More details about how this information was compiled is provided in the above Appendix Section A.3. In cases where average household size was not reported, monthly household consumption cannot be reported.

Table A2: Main survey modules administered at each round

FEMALE	Baseline	Midline	Endline
Housing, water & sanitation	X	—	X
Livestock ownership and management	X	X	X
Cost of livestock production	X	X	X
Income from livestock & livestock products	X	X	X
Agricultural extension related to livestock	X	X	—
Own business activities	X	—	—
Wage employment	—	X	X
Credit for productive purposes	X	—	X
Credit for consumption purposes	X	—	—
Access to savings	X	X	X
Household consumables (monthly)	X	—	—
Women's dietary diversity (24-hour recall)	X	X	X
Household food consumption & expenditure	X	—	X
Household food security	X	X	X
Female agency & decision-making	X	X	X
IYCF practices, knowledge & child health history	X	X	X
Childcare activities & exposure to health services	X	X	X
Anthropometry	X	X	X
Participation in VESA groups and SPIR activities	X	X	X
Exposure to SPIR programming	—	X	X
MALE	**Baseline**	**Midline**	**Endline**
Productive assets	X	—	X
Consumer durables	X	—	X
Livestock ownership & management	X	X	X
Income from livestock & livestock products	X	—	X
Cost of livestock production	X	—	X
Agricultural extension related to livestock	X	X	—
Land characteristics and tenure	X	X	X
Crop choice, inputs & production – Belg Season	X	—	—
Crop choice, inputs & production – Mehr Season	X	X	X
Wage employment	X	X	X
Own business activities	X	—	X
Credit for productive purposes	X	X	X
Credit for consumption purposes	X	X	X
Access to savings	X	X	X
Durables and services (annual)	X	X	X
Household consumables (monthly)	X	—	X
Intrahousehold dynamics & gender norms	X	X	X
IYCF knowledge & childcare activities	X	X	X
Exposure to health and nutrition services	X	X	X
Participation in PSNP	X	X	X
Participation in VESA groups and SPIR activities	X	X	X
Experience with shocks	X	X	X

Table A3: Baseline characteristics for extremely poor versus less poor households

	(1) N	(2) Extremely poor mean	(3) Less poor mean	(4) P-value on test of significance
Household size	3,314	5.45	6.18	0.000***
Female-headed household	3,313	0.27	0.10	0.000***
Household head: married, monogamous	3,310	0.75	0.93	0.000***
Household head has no formal education	3,314	0.73	0.69	0.007***
Value of livestock, productive assets and consumer durables (market prices)	3,314	1,052.39	2,271.56	0.000***
Total value of all productive assets	3,314	57.00	143.02	0.000***
Total value of all consumer durable assets	3,314	93.04	144.28	0.000***
Estimated value of all livestock owned by the household (market prices)	3,314	900.02	1,957.90	0.000***
Household took out any type of loan (past year)	2,729	0.21	0.21	0.490
Household has any savings	2,807	0.25	0.36	0.000***
Household reports any non-agricultural business	2,807	0.04	0.04	0.945
Household reports any regular wage work (past year)	2,807	0.04	0.04	0.825
Food consumption expenditure per month per AE	3,275	49.72	51.31	0.399
Non-food consumption expenditure per month per AE	3,275	12.10	13.33	0.021**
Consumption expenditure per month per AE	3,275	61.81	64.63	0.167

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The final column reports a p-value derived from a simple t-test of equality across subsamples.

Table A4: Intervention exposure

	N	Control mean	Poultry mean	Cash mean	L mean
Panel A: Extremely poor households					
At least one household member is a member of a VESA group	1,765	0.14	0.86	0.81	0.79
Member attended >=1 VESA meeting in last 12 months	1,806	0.08	0.72	0.67	0.65
Female participated in livestock trainings	1,720	0.05	0.35	0.20	0.23
Male participated in livestock trainings	1,030	0.06	0.28	0.29	0.24
Household participated in livestock trainings	1,769	0.07	0.42	0.28	0.30
Household participated in financial education trainings	1,825	0.10	0.41	0.37	0.34
Household received a poultry start-up package	1,846	0.05	0.94	0.18	0.15
Number of chickens household received from start-up package	614	8.38	15.77	5.35	6.11
Household received an unconditional cash grant	1,847	0.02	0.12	0.63	0.04
Amount of money household reported receiving	376	221.07	139.48	395.84	178.76

Panel B: Less poor households	N	Control mean	Pooled treatment mean
At least one household member is a member of a VESA group	1,315	0.13	0.86
Member attended >=1 VESA meeting in last 12 months	1,354	0.09	0.69
Female participated in livestock trainings	1,287	0.05	0.28
Male participated in livestock trainings	950	0.05	0.30
Household participated in livestock trainings	1,366	0.09	0.38
Household participated in financial education trainings	1,322	0.07	0.37
Household received a poultry start-up package	1,372	0.05	0.21
Number of chickens household received from start-up package	206	5.73	6.55
Household received an unconditional cash grant	1,373	0.03	0.03
Amount of money household reported receiving	39	148.39	110.69

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. All variables reported are measured in the three year follow-up survey, with the exception of variables capturing the cash and poultry transfers and participation in financial education, all reported in the one year follow-up. (Participation in financial education was not measured in the three year follow-up due to a survey error.)

Table A5: Livestock ownership: sub-sample of extremely poor households

	One year follow-up				*Three year follow-up*			
	(1) Poultry owned	(2) Goats and sheep owned	(3) Cows and oxen owned	(4) Value of livestock (self-reports, IHS transformation)	(5) Poultry owned	(6) Goats and sheep owned	(7) Cows and oxen owned	(8) Value of livestock (self-reports, IHS transformation)
<u>Extremely poor households</u>								
Poultry	9.133***	0.586***	-0.023	1.193***	1.811***	0.460**	0.017	0.634***
	(0.577)	(0.202)	(0.113)	(0.206)	(0.291)	(0.216)	(0.121)	(0.228)
Cash	0.871**	0.513***	-0.008	0.583**	0.552**	0.112	0.020	0.429
	(0.395)	(0.193)	(0.117)	(0.267)	(0.234)	(0.182)	(0.113)	(0.262)
L	0.675*	0.029	-0.220**	-0.176	0.186	-0.069	-0.173	-0.223
	(0.357)	(0.176)	(0.110)	(0.246)	(0.222)	(0.196)	(0.113)	(0.260)
Test: Poultry = Cash	0.000***	0.739	0.886	0.010**	0.000***	0.110	0.975	0.401
Test: Poultry = L	0.000***	0.008***	0.035**	0.000***	0.000***	0.019**	0.093*	0.000***
Test: Cash = L	0.583	0.014**	0.030**	0.004***	0.126	0.356	0.058*	0.015**
Mean of control	1.67	1.50	1.02	1,079.61	1.76	1.49	1.08	705.53
N	1,825	1,825	1,825	1,825	1,765	1,765	1,765	1,765
<u>Less poor households</u>								
L	0.925***	-0.031	-0.169	-0.009	0.569**	0.005	-0.291**	-0.310*
	(0.238)	(0.215)	(0.109)	(0.197)	(0.236)	(0.203)	(0.122)	(0.169)
Mean of control	2.091	2.170	1.562	1,591.816	2.166	2.016	1.777	1,103.525
N	1,364	1,364	1,364	1,364	1,322	1,322	1,322	1,322

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The dependent variables are calculated as the value of the specified asset class, using households' own estimated valuations; the means in the control arm are reported in 2017 dollars in purchasing power parity terms. The estimated regressions employ a linear specification in Columns (1) to (3) and (5) through (7), and an inverse hyperbolic sine specification in Columns (4) and (8). All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A6: Non-agricultural income at one year follow-up

	(1) Any income from formal wage work (past year)	(2) Any income from casual wage work (past year)
Panel A: Extremely poor households		
Poultry	0.002	-0.019
	(0.016)	(0.047)
Cash	0.007	-0.011
	(0.018)	(0.043)
L	0.007	0.036
	(0.018)	(0.047)
Test: Poultry = L	0.771	0.224
Test: Cash = L	0.984	0.248
Test: Poultry = Cash	0.764	0.842
Mean of control	0.04	0.30
N	1,261	1,260
Panel B: Less poor households		
L	0.006	0.030
	(0.013)	(0.038)
Mean of control	0.03	0.24
N	1,178	1,177

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A7: Assets, without IHS transformation

	One year follow-up		*Three year follow-up*							
	Linear outcome	*Binary for positive*	*Linear outcome variable*			*Binary variable for positive value*				
	(1) Value of livestock	(2) Value of livestock	(3) Value of total assets	(4) Value of productive assets	(5) Value of consumer durables	(6) Value of livestock	(7) Value of total assets	(8) Value of productive assets	(9) Value of consumer durables	(10) Value of livestock
Panel A: Extremely poor households										
Poultry	218.189*	0.159***	3.571	2.431	-1.112	4.800	-0.000	0.029	-0.006	0.088***
	(115.320)	(0.021)	(90.286)	(2.044)	(9.032)	(85.321)	(0.001)	(0.024)	(0.015)	(0.023)
Cash	71.876	0.080***	84.526	2.034	9.698	74.549	-0.005*	0.019	0.002	0.059**
	(102.333)	(0.028)	(90.638)	(1.881)	(9.930)	(86.884)	(0.003)	(0.028)	(0.013)	(0.029)
L	-179.966	0.009	-127.681	-0.895	3.489	-125.844	-0.005	0.022	0.003	-0.023
	(109.645)	(0.028)	(87.568)	(1.901)	(10.164)	(83.667)	(0.003)	(0.022)	(0.014)	(0.029)
Test: Poultry = Cash	0.149	0.002***	0.366	0.860	0.218	0.415	0.115	0.685	0.565	0.273
Test: Poultry = L	0.000***	0.000***	0.133	0.147	0.613	0.117	0.104	0.705	0.547	0.000***
Test: Cash = L	0.007***	0.017**	0.015**	0.159	0.522	0.017**	0.997	0.889	0.959	0.006***
Test: β_1 1 year = 3 years [poultry]						0.014				0.001
Test: β_2 1 year = 3 years [cash]						0.974				0.417
Test: β_3 1 year = 3 years [L]						0.509				0.240
Mean of control	1,453.86	0.82	1,105.66	35.21	96.54	971.87	1.00	0.91	0.97	0.83
N	1,847	1,847	1,765	1,765	1,765	1,765	1,765	1,765	1,765	1,765
Panel B: Less poor households										
L	-88.429	0.027	-120.729	0.305	19.670*	-135.548	0.003	0.009	0.001	-0.010
	(130.076)	(0.021)	(98.944)	(2.021)	(10.792)	(94.048)	(0.003)	(0.011)	(0.009)	(0.017)
Test: β 1 year = 3 years						0.569				0.056
Mean of control	2,168.13	0.93	1,653.85	45.41	100.84	1,495.62	1.00	0.97	0.98	0.94
N	1,373	1,373	1,322	1,322	1,322	1,322	1,322	1,322	1,322	1,322

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The dependent variables are calculated as the value of the specified asset class, valued using locally reported market prices; the means in the control arm are reported in 2017 dollars in purchasing power parity terms. The estimated regressions employ a linear specification or a binary variable for a strictly positive value, as specified. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A8: Financial inclusion, without IHS transformation

	One year follow-up				*Three year follow-up*			
	Linear outcome variable		*Binary variable for positive value*		*Linear outcome variable*		*Binary variable for positive value*	
	(1) Total credit (past year)	(2) Total savings	(3) Total credit (past year)	(4) Total savings	(5) Total credit (past year)	(6) Total savings	(7) Total credit (past year)	(8) Total savings
Panel A: Extremely poor households								
Poultry	20.994	25.087**	0.025	0.321***	29.845	20.553***	0.077**	0.340***
	(27.446)	(10.816)	(0.044)	(0.041)	(24.800)	(6.430)	(0.036)	(0.046)
Cash	-18.605	29.336*	-0.002	0.351***	19.535	15.994**	0.098**	0.336***
	(29.573)	(16.136)	(0.044)	(0.039)	(23.365)	(7.019)	(0.040)	(0.046)
L	26.189	38.642	0.059	0.266***	-9.543	19.105**	-0.009	0.328***
	(28.603)	(25.025)	(0.044)	(0.042)	(22.105)	(7.484)	(0.034)	(0.044)
Test: Poultry = Cash	0.129	0.777	0.503	0.380	0.678	0.494	0.597	0.901
Test: Poultry = L	0.829	0.575	0.405	0.158	0.092*	0.839	0.012**	0.730
Test: Cash = L	0.100*	0.724	0.128	0.020**	0.183	0.677	0.005***	0.833
Test: β_1 1 year = 3 years [poultry]					0.746	0.677	0.296	0.683
Test: β_2 1 year = 3 years [cash]					0.128	0.359	0.021	0.740
Test: β_3 1 year = 3 years [L]					0.163	0.378	0.136	0.162
Mean of control	156.35	40.98	0.42	0.44	147.03	34.07	0.45	0.40
N	1,290	1,296	1,295	1,296	1,765	1,765	1,765	1,765
Panel B: Less poor households								
L	8.517	-0.374	0.051	0.278***	20.727	30.448***	0.022	0.301***
	(31.591)	(16.057)	(0.037)	(0.034)	(26.092)	(7.433)	(0.042)	(0.042)
Test: β 1 year = 3 years					0.732	0.024	0.542	0.625
Mean of control	184.986	74.990	0.436	0.504	170.37	34.73	0.48	0.47
N	1,182	1,185	1,185	1,184	1,321	1,322	1,321	1,322

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ a linear specification or a binary variable for a strictly positive value, as specified. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A9: Income at three year follow-up, without IHS transformation

	Linear outcome variable				*Binary variable for positive value*			
	(1) Income from livestock sales	(2) Income from sales of livestock products	(3) Income from crops cultivated in last year	(4) Total agricultural and livestock income	(5) Income from livestock sales	(6) Income from sales of livestock products	(7) Income from crops cultivated in last year	(8) Total agricultural and livestock income
Panel A: Extremely poor households								
Poultry	45.096	30.665***	7.242	135.117**	0.141***	0.113***	0.048	0.133***
	(32.992)	(11.763)	(21.866)	(66.411)	(0.037)	(0.025)	(0.034)	(0.036)
Cash	85.488**	29.538**	49.605	162.786***	0.112***	0.068***	0.027	0.084**
	(35.569)	(12.564)	(34.317)	(58.472)	(0.037)	(0.021)	(0.033)	(0.036)
L	-19.332	19.342	9.872	13.481	0.023	0.031	-0.043	-0.006
	(33.759)	(11.793)	(48.836)	(69.174)	(0.038)	(0.021)	(0.036)	(0.039)
Test: Poultry = Cash	0.285	0.913	0.224	0.708	0.451	0.091*	0.508	0.147
Test: Poultry = L	0.071*	0.282	0.958	0.149	0.003***	0.002***	0.011**	0.000***
Test: Cash = L	0.006***	0.335	0.540	0.075*	0.023**	0.108	0.045**	0.018**
Mean of control	266.87	6.48	83.99	361.63	0.46	0.08	0.28	0.58
N	1,765	1,761	1,765	1,765	1,765	1,740	1,765	1,752
Panel B: Less poor households								
L	19.759	74.762	-1.233	115.542	0.038	0.021	-0.079**	0.011
	(39.094)	(72.152)	(37.034)	(101.721)	(0.038)	(0.028)	(0.036)	(0.037)
Mean of control	373.77	-39.87	126.64	465.52	0.58	0.15	0.36	0.68
N	1,322	1,319	1,322	1,322	1,322	1,294	1,322	1,306

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ a linear specification or a binary variable for a strictly positive value, as specified. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A10: Consumption and food security, without IHS transformation

	Linear outcome variable		
	(1) Monthly food consumption per adult equivalent	(2) Monthly non-food consumption per adult equivalent	(3) Monthly consumption per adult equivalent
Panel A: Extremely poor households			
Poultry	3.879	-0.117	3.705
	(4.598)	(0.694)	(4.866)
Cash	-1.268	-0.058	-1.511
	(4.040)	(0.643)	(4.235)
L	-2.265	0.882	-1.414
	(3.368)	(0.666)	(3.619)
Test: Poultry = Cash	0.289	0.927	0.306
Test: Poultry = L	0.169	0.128	0.280
Test: Cash = L	0.793	0.122	0.981
Mean of control	44.87	7.68	52.86
N	1,706	1,764	1,701
Panel B: Less poor households			
L	-2.321	0.635	-1.629
	(4.466)	(0.561)	(4.601)
Mean of control	47.878	7.840	55.872
N	1,291	1,322	1,290

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ a linear specification. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A11: Heterogeneity with respect to nutrition exposure: assets

	One year follow-up	*Three year follow-up*			
	(1) Value of livestock	(2) Value of total assets	(3) Value of productive assets (IHS transformation)	(4) Value of consumer durables	(5) Value of livestock
Panel A: Extremely poor households					
Poultry	0.932***	0.057	0.145	0.167*	0.285
	(0.197)	(0.120)	(0.111)	(0.086)	(0.204)
Cash	0.564**	0.208*	0.165	0.062	0.682***
	(0.251)	(0.108)	(0.110)	(0.106)	(0.183)
Poultry x EN	0.315*	0.112	0.011	-0.257*	0.289
	(0.188)	(0.137)	(0.144)	(0.142)	(0.219)
Cash x EN	-0.109	-0.296*	-0.188	-0.110	-0.643**
	(0.343)	(0.157)	(0.180)	(0.109)	(0.284)
L	-0.189	-0.241**	0.016	-0.019	-0.309
	(0.206)	(0.110)	(0.101)	(0.090)	(0.199)
Mean of control	1,453.86	1,105.66	35.21	96.54	971.87
N	1,847	1,765	1,765	1,765	1,765
Panel B: Less poor households					
L	0.074	-0.032	0.002	0.163	-0.200
	(0.214)	(0.095)	(0.088)	(0.100)	(0.173)
L x EN	0.113	-0.063	0.063	0.005	-0.050
	(0.145)	(0.075)	(0.063)	(0.075)	(0.158)
Mean of control	2,168.13	1,653.85	45.41	100.84	1,495.62
N	1,373	1,322	1,322	1,322	1,322

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The dependent variables are calculated as the value of the specified asset class, valued using locally reported market prices; the means in the control arm are reported in 2017 dollars in purchasing power parity terms. The estimated regressions employ an inverse hyperbolic sine transformation. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A12: Heterogeneity with respect to nutrition exposure: financial inclusion, one year follow-up

	(1) Any credit (past year)	(2) Any formal credit (past year)	(3) Total credit (past year, IHS transformation)	(4) Any savings	(5) Total savings (IHS transformation)
Panel A: Extremely poor households					
Poultry	-0.041	-0.030	-0.186	0.305***	1.339***
	(0.048)	(0.040)	(0.302)	(0.051)	(0.237)
Cash	-0.013	0.015	-0.058	0.367***	1.549***
	(0.050)	(0.039)	(0.315)	(0.043)	(0.218)
Poultry x EN	0.133**	0.091**	0.746**	0.027	0.140
	(0.053)	(0.042)	(0.308)	(0.051)	(0.253)
Cash x EN	0.017	-0.016	0.164	-0.042	-0.423
	(0.059)	(0.049)	(0.380)	(0.044)	(0.282)
L	0.064	0.046	0.451	0.260***	1.148***
	(0.044)	(0.036)	(0.277)	(0.042)	(0.210)
Mean of control	0.42	0.17	156.35	0.44	40.98
N	1,291	1,285	1,290	1,289	1,296
Panel B: Less poor households					
L	0.072*	0.026	0.362	0.276***	1.072***
	(0.042)	(0.039)	(0.288)	(0.037)	(0.202)
L x EN	-0.039	-0.019	-0.232	0.003	-0.084
	(0.038)	(0.034)	(0.252)	(0.030)	(0.168)
Mean of control	0.44	0.19	184.99	0.50	74.99
N	1,182	1,176	1,182	1,182	1,185

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ an inverse hyperbolic sine transformation. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A13: Heterogeneity with respect to nutrition exposure: financial inclusion, three year follow-up

	(1) Any credit (past year)	(2) Any formal credit (past year)	(3) Total credit (past year, IHS transformation)	(4) Any savings	(5) Total savings (IHS transformation)
Panel A: Extremely poor households					
Poultry	0.043	0.033	0.282	0.343***	1.398***
	(0.045)	(0.036)	(0.291)	(0.053)	(0.257)
Cash	0.132***	0.089**	0.827***	0.354***	1.388***
	(0.048)	(0.037)	(0.312)	(0.053)	(0.251)
Poultry x EN	0.064	0.088**	0.470	-0.014	0.012
	(0.053)	(0.041)	(0.348)	(0.054)	(0.251)
Cash x EN	-0.069	-0.010	-0.411	-0.042	-0.142
	(0.062)	(0.045)	(0.397)	(0.054)	(0.288)
L	-0.006	0.032	0.018	0.327***	1.276***
	(0.034)	(0.030)	(0.216)	(0.043)	(0.213)
Mean of control	0.45	0.13	147.03	0.40	34.07
N	1,765	1,760	1,765	1,765	1,765
Panel B: Less poor households					
L	0.017	0.035	0.134	0.321***	1.538***
	(0.050)	(0.041)	(0.312)	(0.049)	(0.239)
L x EN	0.003	0.002	0.093	-0.029	-0.213
	(0.037)	(0.028)	(0.230)	(0.034)	(0.178)
Mean of control	0.48	0.19	170.37	0.47	34.73
N	1,322	1,320	1,321	1,322	1,322

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ an inverse hyperbolic sine transformation. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A14: Heterogeneity with respect to nutrition exposure: income, three year follow-up

	(1) Income from sales of livestock	(2) Income from sales of livestock products	(3) Income from crops cultivated in last year (Mehr and Belg seasons)	(4) Total estimated value of livestock products produced in last 12 months	(5) Total estimated value of crops harvested in last year (Mehr and Belg seasons)	(6) Total income from livestock and crops	(7) Total value of livestock and crops	(8) Household reports any non-agricultural business	(9) Any income from formal wage work in past 12 months	(10) Any income from casual wage work in past 12 months
	IHS transformation									
Panel A: Extremely poor households										
Poultry	0.729**	0.511**	0.161	0.528**	0.287	0.629**	0.476*	0.027	-0.008	-0.001
	(0.317)	(0.202)	(0.251)	(0.259)	(0.339)	(0.310)	(0.244)	(0.023)	(0.022)	(0.058)
Cash	0.785**	0.484***	0.462*	0.522**	0.890***	0.838***	0.724***	0.014	-0.025	0.030
	(0.305)	(0.181)	(0.241)	(0.221)	(0.299)	(0.317)	(0.260)	(0.021)	(0.019)	(0.053)
Poultry x N*	0.116	-0.004	0.132	-0.137	0.272	0.100	0.164	-0.029	0.002	0.055
	(0.359)	(0.237)	(0.261)	(0.281)	(0.365)	(0.361)	(0.258)	(0.027)	(0.021)	(0.066)
Cash x N*	0.013	-0.153	-0.632**	-0.429*	-1.194***	-0.341	-0.767**	0.010	0.038	-0.022
	(0.348)	(0.199)	(0.250)	(0.229)	(0.427)	(0.334)	(0.352)	(0.038)	(0.029)	(0.058)
L	0.075	0.173	-0.284	0.169	-0.217	-0.112	-0.109	0.044**	0.016	0.045
	(0.242)	(0.119)	(0.224)	(0.169)	(0.280)	(0.277)	(0.250)	(0.020)	(0.018)	(0.045)
Mean of control	266.87	6.48	82.53	27.76	475.02	360.17	773.88	0.03	0.04	0.26
N	1,765	1,761	1,765	1,762	1,765	1,765	1,765	1,030	1,029	1,030
Panel B: Less poor households										
L	0.345	0.054	-0.347	0.009	0.293	0.324	0.314	-0.034*	0.002	0.023
	(0.300)	(0.196)	(0.255)	(0.217)	(0.277)	(0.326)	(0.244)	(0.019)	(0.018)	(0.038)
L x N*	-0.181	0.139	-0.116	0.065	-0.190	-0.436*	-0.101	0.022	0.008	0.014
	(0.229)	(0.158)	(0.193)	(0.181)	(0.200)	(0.237)	(0.171)	(0.016)	(0.014)	(0.034)
Mean of control	373.77	-39.87	126.95	49.41	639.40	465.83	1,067.56	0.06	0.03	0.20
N	1,322	1,319	1,322	1,322	1,322	1,322	1,322	951	951	951

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ an inverse hyperbolic sine transformation. The estimate of total income in Column (6) corresponds to the sum of Columns (1) through (3), and the estimate of total crop and product value in Column (7) corresponds to the sum of Column (1), (4), and (5). All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A15: Heterogeneity with respect to nutrition exposure: consumption and food security

	(1) Monthly food consumption per adult equivalent (IHS transformation)	(2) Monthly non-food consumption per adult equivalent (IHS transformation)	(3) Monthly consumption per adult equivalent (IHS transformation)	(4) Food Insecurity Experience Scale (0-8)
Panel A: Extremely poor households				
Poultry	0.030	0.062	0.039	0.108
	(0.078)	(0.104)	(0.071)	(0.242)
Cash	-0.049	0.061	-0.032	0.343
	(0.075)	(0.085)	(0.069)	(0.231)
Poultry x EN	0.003	-0.085	-0.029	-0.014
	(0.098)	(0.111)	(0.091)	(0.296)
Cash x EN	-0.071	-0.021	-0.055	0.347
	(0.096)	(0.079)	(0.086)	(0.319)
L	-0.071	0.140*	-0.038	0.224
	(0.067)	(0.077)	(0.061)	(0.176)
Mean of control	44.87	7.68	52.86	3.36
N	1,706	1,764	1,701	1,748
Panel B: Less poor households				
L	-0.059	0.075	-0.027	0.033
	(0.087)	(0.080)	(0.079)	(0.245)
L x EN	-0.016	-0.002	-0.015	-0.040
	(0.069)	(0.075)	(0.063)	(0.213)
Mean of control	47.88	7.84	55.87	3.32
N	1,291	1,322	1,290	1,314

Notes: The sample of extremely poor households is defined by ranking households within each subdistrict based on an asset index constructed at baseline; the poorest 10 in each sample of 18 are denoted extremely poor, with the remaining households denoted less poor. The dependent variables in Columns (1) through (3) are household-level consumption variables; the dependent variable in Column (4) is the continuous FIES score. The continuous dependent variables are reported in 2017 dollars in purchasing power parity terms, and the estimated regressions employ an inverse hyperbolic sine transformation. All specifications are estimated conditional on strata fixed effects and employing standard errors clustered at the subdistrict level; asterisks indicate significance at the 10, 5 and 1 percent level.

Table A16: Number of kebeles in each treatment arm, by woreda

Region	Woreda	T1: L*+EN	T2: L*+N	T3: L+EN	C: PSNP4	Total
Amhara	Bugna	2	1	1	1	5
	Dahena	3	4	4	4	15
	Gaz Gibla	2	1	2	2	7
	Lasta	3	4	4	4	15
	Meket	10	9	9	10	38
	Sekota	5	5	5	4	19
	Wadla	4	4	4	4	16
Oromia	Chiro	3	3	3	3	12
	Daro Lebu	9	10	9	9	37
	Gemechis	2	3	3	3	11
	Grawa	4	3	3	4	14
	Kurfachelle	0	1	1	0	2
	Siraro	2	1	1	1	5
Total		49	49	49	49	196

Note: After the sampling was completed, two new woredas – Gazo and Tsagabji – were created from the existing woredas. This led to a reshuffling of some kebeles in Meket, Wadla and Lasta. For the purpose of this study, we refer to these kebeles by the woreda to which they belonged at the time of sampling. Kebele treatment assignments and associated implementation remain unchanged despite these administrative changes.

Table A17: Number of kebeles in each treatment arm, by region

	Amhara	Oromia	Total
T1: L*+EN	27	19	46
T2: L*+N	28	21	49
T3: L+EN	29	20	49
C: L+N	29	20	49
Total	113	80	193

Table A18: Attrition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Binary variable: Attrited at midline					Binary variable: Attrited at endline				
Poultry	-0.001	-0.009	-0.009	0.000	-0.000	0.005	-0.010	0.092	0.009	-0.085
	(0.009)	(0.016)	(0.022)	(0.010)	(0.100)	(0.014)	(0.022)	(0.056)	(0.016)	(0.143)
Cash	0.012	0.013	-0.026	0.009	0.083	-0.006	0.024	-0.022	-0.011	0.049
	(0.017)	(0.017)	(0.018)	(0.016)	(0.075)	(0.020)	(0.023)	(0.034)	(0.019)	(0.110)
L	0.008	-0.003	-0.001	0.008	0.052	0.027	0.019	0.084**	0.018	0.082
	(0.008)	(0.011)	(0.021)	(0.008)	(0.065)	(0.021)	(0.018)	(0.041)	(0.017)	(0.104)
Female-headed household	0.044***					0.031				
	(0.015)					(0.026)				
Head has no formal education		0.001					0.012			
		(0.008)					(0.013)			
Baseline value of prod. assets			-0.003					0.002		
			(0.002)					(0.003)		
Household reports any savings				-0.002					-0.005	
				(0.009)					(0.014)	
Household consumption					0.004					-0.002
					(0.007)					(0.010)
Baseline variable x Poultry	-0.010	0.012	0.001	-0.013	0.000	0.063	0.044	-0.014*	-0.032	0.016
	(0.034)	(0.019)	(0.003)	(0.014)	(0.015)	(0.068)	(0.033)	(0.007)	(0.023)	(0.022)
Baseline variable x Cash	-0.044	-0.010	0.005	0.015	-0.013	-0.044	-0.052*	0.002	0.016	-0.011
	(0.028)	(0.019)	(0.004)	(0.020)	(0.011)	(0.041)	(0.028)	(0.006)	(0.024)	(0.017)
Baseline variable x L	-0.025	0.012	0.001	-0.000	-0.007	0.043	0.026	-0.008*	0.032	-0.007
	(0.022)	(0.014)	(0.003)	(0.017)	(0.009)	(0.046)	(0.030)	(0.004)	(0.028)	(0.015)
Mean of control (T4)	0.030	0.030	0.030	0.030	0.030	0.052	0.052	0.052	0.052	0.052
N	3,313	3,314	3,314	2,807	3,275	3,313	3,314	3,314	2,807	3,275

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. Models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Chapter 3: Mental Health

Treating depression among the extreme poor: Evidence from a graduation program with group therapy in Ethiopia

Michael Mulford[c], Daniel O. Gilligan[a], Harold Alderman[a], Paul Bolton[b], Melissa Hidrobo[a], Jessica Leight[a], Heleene Tambet[a], Lena Verdeli[d], and Abaydar Workie[c]

[a]International Food Policy Research Institute
[b]Johns Hopkins University
[c]World Vision International
[d]Teachers College, Columbia University

Abstract

We report the results of an experiment testing the impact of a light-touch graduation program of livelihood and nutrition programming, that includes group psychotherapy for extremely poor women in Ethiopia. Among women who initially exhibited depressive symptoms, a treatment arm combining enhanced livelihood activities that included targeted cash or asset transfers and enhanced nutrition activities that included group psychotherapy reduced the prevalence of depressive symptoms by 46 percent and of major depressive disorder by 75 percent more than one year later. This bundled intervention also improved women's psychosocial outcomes, income from livestock sales, and physical health. Neither the enhanced livelihood programming alone nor the enhanced nutrition programming alone reduced depression among women with depressive symptoms. These results document the effectiveness of a scalable model of group therapy delivered within a multifaceted graduation program and point to the importance of addressing multiple constraints for those living in poverty in order to effectively target depression.

1. Introduction

Depression and related mental health disorders are a significant global public health challenge. It is estimated that one out of seven individuals experience depression over their lifetime, and that depression is among the leading causes of disease burden and disability worldwide (Vos et al. 2012). The prevalence of depressive disorders grew from 170.8 million in 1990 to 279.6 million in 2019 (GBD 2019 Mental Disorders Collaborators 2022).

Growing evidence documents that poverty and common mental health disorders are mutually reinforcing, suggesting the potential for a psychological pathway for poverty traps (Haushofer and Fehr 2014; Patel and Kleinman 2003; Lund et al. 2011; Barrett et al. 2018). Poverty induced stress contributes to higher prevalence of depression and other mental health challenges in low- and middle-income countries (LMICs), especially among the poor in those countries (Lund et al. 2010). At the same time, depression has dire economic consequences. Common somatic symptoms associated with depression such as fatigue and disrupted sleeping and eating can reduce labor supply and impair cognitive function, which in turn affect decision making abilities and overall productivity. Other symptoms such as anhedonia, hopelessness, and pessimism reduce individuals' capacity to act on economic opportunities or make longer-term investments (de Quidt and Haushofer 2016).

There is some evidence of effective interventions to address mental health in LIMCs, but most of the existing solutions are limited, and few operate at a sufficiently large scale to meaningfully address the extent of untreated mental health problems. Psychosocial interventions have been effective in addressing mental health outcomes in LIMCs (Barbui et al. 2020), including psychotherapies treating depression (Cuijpers et al. 2018). Availability of skilled mental healthcare providers and related services is however extremely limited in these countries, particularly in rural areas. While not primarily motivated by addressing mental health, a growing literature has documented impacts of cash transfer programs on depression and anxiety among poor households targeted by these programs (McGuire, Kaiser, and Bach-Mortensen 2022; Ridley et al. 2020; Haushofer, Mudida, and Shapiro 2020; Angeles et al. 2019; Haushofer and Shapiro 2016; Baird, de Hoop, and Özler 2013; Fernald and Gunnar 2009; Ozer et al. 2011). The impact of these transfers on depression are often small or not sustained. More intensive graduation model programming is promising, but there is limited evidence of their impact on depression, and no graduation model has specifically targeted depression.

In this paper, we use a cluster randomized controlled trial (cRCT) to study the impact of a novel graduation program in Ethiopia that includes group psychotherapy for women who exhibited elevated depressive symptoms. The target population for this study comprised households participating in Ethiopia's Productive Safety Net Program (PSNP), the national safety net that provides food and cash transfers either as payment for public works labor or as direct support for the elderly or disabled. The graduation program, Strengthen PSNP Institutions and Resilience (SPIR), layered additional livelihoods, health, nutrition, and other programming on top of the PSNP consumption-support transfers. As a graduation program, SPIR had several unique features: (i)

cash or asset 'livelihood transfers' were only provided to the poorest half of PSNP households[1] and their value, $200, was less[2] than in many other graduation programs; (ii) the program included targeted group-based interpersonal psychotherapy (IPT-G)[3] facilitated by trained non-specialists to address depression in mothers[4] of young children, who were screened in based on their reported prevalence of depressive symptoms; and (iii) the overall program operated at scale, supporting more than 500,000 beneficiaries.

The study sample included 3,314 households in 192 *kebeles* (sub-districts). All households in the study, including the control, received PSNP consumption-support transfers. Three treatment arms each provided graduation model programming that differed in the combination of livelihood and nutrition components; two of these treatment arms included group psychotherapy. Basic livelihood programming (L) was delivered through Village Economic and Social Associations (VESAs), including access to savings and credit and limited livelihood training. An enhanced livelihood component (L*) added a $200 livelihood transfer targeted to women in extremely poor households. Core nutrition programming (N) included nutrition information and sanitation and hygiene services, while the enhanced nutrition component (N*) added IPT-G for women screened for depression, and additional nutrition support. These components were combined and randomized into four treatment arms: T1 combined the enhanced livelihoods and nutrition (L* and N*), T2 combined the enhanced livelihoods and basic nutrition (L* and N), T3 combined the basic livelihoods and enhanced nutrition (L and N*), and T4 was the control. Data collection was conducted before the program began (2018), just before the group therapy intervention (2019), and one year after the completion of the 12-week group therapy sessions (2021).

We report results for two samples of women: the full sample and the subsample of women screened for depression during the second survey round. For women in the full sample, we find no impacts on average of SPIR programming on depression at midline – just before the group therapy treatment for treatment arms T1 and T3 – or at endline, either from the pooled treatment or from the three distinct treatment arms. For women exhibiting elevated depressive symptoms at midline, we find large effects on depression and economic outcomes for the treatment arm that combined enhanced livelihood and nutrition components (T1). This contrasts with mostly null effects on depression and other outcomes in the other treatment arms.

In terms of mental health outcomes, for the subsample of women with elevated depressive symptoms at midline, the prevalence of these symptoms fell by 46 percent in T1. This effect was even larger, a 75 percent prevalence reduction in T1, using the more restrictive measure of Major

[1] The PSNP program targets the poorest 10-15% of households in chronically food insecure areas in Ethiopia. Thus, the poorest 55% of PSNP households that are targeted for the cash or asset transfer in this study represent the poorest 5-8% of households in these rural, food insecure communities.
[2] This is approximately $374 in 2017 purchasing power parity (PPP) terms. The cost of the asset transfer in the six graduation programs studied in (Banerjee et al. 2015) ranged from US$481 to US$1537 per household (all in 2017 PPP terms).
[3] Interpersonal Psychotherapy for Groups (IPT-G) is a scalable intervention included in the WHO's mental health Gap Action Programme (mhGAP) to address depression in low-resource settings.
[4] As 98 percent of children under 36 months were cared for by their mother, we use the term mother rather than female caregiver in this paper.

depressive disorder[5] (MDD) (Otte et al. 2016). Regarding psychosocial outcomes, women in T1 reported higher levels of happiness but no effect on a measure of self-efficacy. For economic outcomes, no effects were measured for any treatment arm on measures of consumption or assets, but a positive and similar effect (72 percent increase) was observed in both T1 and T3 on income from livestock sales. It is plausible that women in these treatment arms were able to improve their own livestock-focused economic activities, but in relative terms, this increased income was not enough to translate into overall household-level impacts. Lastly, in terms of health and nutrition outcomes, there was no effect on improving women's dietary diversity in any treatment arm, but in T1, women's Body Mass Index[6] increased by 0.41 SD – a large effect in a sample where more than 30 percent of women are underweight. This pattern of results suggests that amongst the subset of women who exhibited elevated depressive symptoms, the combination of enhanced nutrition and enhanced livelihoods support in T1 was necessary to address the multiple constraints they faced to improve their wellbeing as measured across multiple dimensions.

This study makes several important contributions to the literature. Banerjee et al. (2018) speculate that mental health illnesses could be one of the unaddressed constraints that prevent the extreme poor from successfully taking up economic opportunities in graduation programs. Our study is the first[7] that we are aware of to add a targeted mental health intervention to graduation programming which also experimentally varies the provision of livelihoods transfers. As people in extreme poverty experience a higher burden of stress and depression (Sipsma et al. 2013), targeted interventions to address both mental health illnesses and socioeconomic constraints may be necessary to unlock potentially self-reinforcing mental illness-induced poverty traps (Haushofer and Fehr 2014; Genicot and Ray 2017).

Availability of treatment for depression in LMICs is severely limited, due in part to the dearth of trained psychologists or psychiatrists.[8] This has motivated the development of low-cost, scalable forms of psychotherapy that can be delivered by trained non-specialists to effectively treat depression, anxiety, and other mental disorders (Singla et al. 2017). Growing evidence points toward the short-term[9] effectiveness of these psychological interventions in treating common mental disorders (Singla et al., 2017), including treatment of maternal depression through IPT-G (Bolton et al. 2003). A much smaller number of studies have evaluated longer-term effects of these interventions, with no significant effects measured after 3 years in Pakistan (Maselko et al. 2020), but persistent effects observed after 5 and 7 years from psychotherapy interventions to address depression in India (Bhat et al. 2022) and Pakistan (Baranov et al. 2020) respectively. Measured

[5] MDD is characterized by expression of 5 or more depressive symptoms that are present nearly every day for a two week period and cause clinically significant distress and loss of normal functioning in social or occupational settings.
[6] BMI is a ratio of weight-to-height that is used to identify conditions of underweight and overweight in adults.
[7] Bossuroy et al. (2022) add a light-touch psychosocial intervention on aspirations and social norms to a graduation program in the Sahel, but this is not intended to treat mental illnesses.
[8] In Ethiopia, for example, there were only 111 practicing psychiatrists in 2019 for a population of 112 million people, according to the Ministry of Health.
[9] Based on one to six months post treatment evaluations.

approximately one year after completion of the IPT-G sessions, our results suggest that addressing income constraints may be critical to sustaining initial gains from mental health interventions.

Studies that are most closely related to our graduation program that includes a focus on mental health are two that looked at the individual and combined effects of psychotherapy treatment and cash transfers for high-risk young men in Liberia (Blattman, Jamison, and Sheridan 2017), and a general-population sample from rural Kenya (Haushofer, Mudida, and Shapiro 2020). In Liberia, therapy or a $200 cash transfer alone had no persistent effect on outcomes of crime and violence, but therapy plus cash led to large treatment effects after one year. The study in Kenya found significant effects on psychological well-being from a $1,076 cash transfer, but no effects from the therapy treatment alone. Nor were differential effects observed in the combined therapy plus cash treatment after one year.

Many questions remain about the targeting and design of these interventions. Most studies evaluate the impact on individuals targeted due to their vulnerability (Bryant et al. 2017) or self-reported mental illness symptoms. There is also a question of intensity with interventions varying from a few short, individual counselling sessions (Vlassopoulos et al. 2021) to three months of weekly group therapy sessions as was implemented in this study. Haushofer et al. (2020) hypothesize that null effects measured one year after treatment could be due to their lower intensity intervention that was provided for a non-targeted rural population in Kenya. This partially contrasts with the positive short-term effects from a more intensive CBT-based intervention that was provided for a general population in rural Ghana (Barker et al. 2022). [10]

In the few studies that combine livelihoods and mental health interventions, there are also questions about what level of livelihoods programming intensity is required to enhance or sustain effects. A minimal livelihoods assistance on its own did not improve mental health among adults with depression in India, but when combined with pharmacotherapy treatment, reduced depression more than pharmacotherapy treatment on its own (Angelucci and Bennett 2024). A $200 cash transfer in the Liberia study mentioned above did not have sustained economic effects, but its combination with therapy sustained a 0.25 SD decrease in antisocial behaviors after ten years for high-risk young men targeted by the intervention (Blattman et al. 2022). One study that did not include a livelihoods component found sustained mental health benefits, but no effects on economic outcomes (Bhat et al. 2022). We contribute to this literature by providing evidence that a targeted, relatively intense psychotherapy intervention combined with livelihood transfers in a graduation program has sustained impacts across several categories of outcomes for women in poor households in rural Ethiopia.

2. The graduation program with group therapy

2.1 The SPIR program

[10] This is not a direct comparison, as the study in Kenya may have also had positive short-term effects that faded by the time of the measurement one year later.

In recognition of the high levels of vulnerability among rural households largely engaged in subsistence, rain-fed agriculture, the Government of Ethiopia and other development partners launched the Productive Safety Net Program (PSNP) in 2005 to prevent household asset depletion and negative coping strategies in the face of recurrent shocks such as drought. In its fourth phase, the program benefitted nearly eight million individuals through food or cash transfers associated with payment for public works labor in the agricultural off-season or as direct support for households unable to contribute labor.[11] The PSNP was targeted both geographically and at a household level to the most food insecure, representing the poorest 15% of households in chronically food insecure areas of the country. In spite of administrative challenges, such as providing transfers on a timely and predictable schedule, the program has reduced the months of food insecurity during the lean season (Gilligan, Hoddinott, and Taffesse 2009; Berhane et al. 2014) and the negative impact of shocks (Knippenberg and Hoddinott 2019). Through the public work requirement, the program also supports community level investment in soil protection and watershed restoration. This has resulted in landscape level effects such as increased tree cover in PSNP areas (Hirvonen et al. 2022).

Supporting the PSNP in 13 districts in Amhara and Oromia regions, the SPIR program was funded by USAID's Bureau for Humanitarian Assistance and implemented by World Vision, CARE and ORDA. SPIR layered several livelihood and nutrition activities on top of the core PSNP components.[12] Additional livelihood activities (L) included the formation of Village Economic and Social Associations (VESAs), the primary platform supporting financial literacy training for women and men, income generation activity promotion and business plan development. The core nutrition activities (N) included nutritional behavior change communication (BCC) and limited water, sanitation, and hygiene (WASH) activities to promote health. For learning purposes, SPIR introduced enhanced models of the livelihoods and nutrition activities.

The enhanced livelihood activity (L*) included all livelihood activities in L and livelihood transfers provided to extremely poor households according to an asset index measured at baseline.[13] The livelihood transfers were provided either as a one-time unconditional cash transfer of $200 or an improved poultry production package of start-up inputs of equivalent value[14]. The enhanced nutrition activities (N*) included a more targeted program of nutrition BCC, a 2-week community-based participatory nutrition promotion (CPNP) activity for caregivers with underweight children to learn and practice improved child feeding practices; promotion of male engagement in household chores and childcare; and an invitation to participate in Interpersonal Psychotherapy in Groups

[11] Approximately 20% of the PSNP caseload is comprised of these households who are unable to supply the intensive manual labor due to different factors such as disability or old age.
[12] Additional activities related to sustainable land management and integrated watershed management were included in all treatment arms, including the control, as a part of the public works-supported community assets construction.
[13] The L* activities also included a one-time aspirations promotion event in the form of inspirational documentary films. Analysis of the impact of this aspirations intervention found null effects, so we exclude them from this analysis (Leight et al. 2021).
[14] The poultry package included 16 improved breed pullets (45-54 day-old chickens), poultry feed, chicken coop construction materials and support for veterinary services and training.

(IPT-G) for women who were screened for elevated depressive symptoms during the midline survey. A timeline of the introduction of SPIR program components is provided in Figure 1.

2.2 The IPT-G intervention

Interpersonal psychotherapy (IPT) was originally developed in the 1970s for individual treatment of depression and has since been adapted for group settings and recommended by WHO as an evidence-based psychological intervention for treating mild to moderate-severe depression (Cuijpers et al. 2016; World Health Organization 2016). IPT focuses on the connection between the onset of depressive symptoms with interpersonal problems, broadly categorized into grief over the death of a loved one, disputes, devastating life changes, or social isolation. The adaptation to delivery of IPT in groups provides opportunity for participants to recognize the contribution of these stressful interpersonal experiences to their depressive symptoms, identify and share personal goals, and role play practical skills to help address these problems.

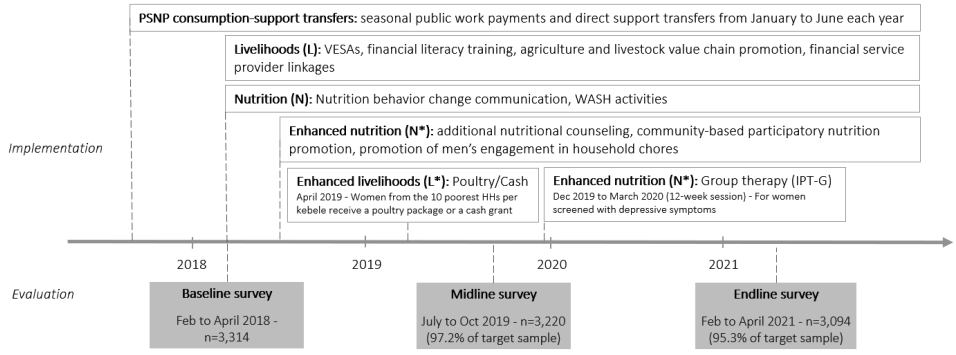
The IPT-G intervention started after the midline survey (Figure 1) and the sessions were conducted by non-specialist Community Health Facilitators (CHFs) who received two weeks of intensive training to support this mental health intervention. CHFs supported participants to focus on strategies to address their most pressing interpersonal problem, seeking ideas and support from other group members and sharing any progress or relational challenges from the past week. The CHFs were closely supervised and supported by IPT-G officers,[15] who provided detailed feedback on the quality of their facilitation. IPT-G participants also received monthly follow-up support from the CHFs for the 6 months after the last IPT-G session either through a monthly group meeting, a home visit, or a phone call.

The IPT-G groups included 6 women on average; in kebeles where less than 4 women were screened into treatment, CHFs conducted additional screening among non-SPIR households in order to form an optimally sized group within a reasonable distance from screened women's homes. As a result, 239 additional women with elevated depression symptoms participated in the sessions but were not part of the evaluation. Of the 227 women invited to the group therapy in T1 and T3 treatment arms, 188 women (83 percent) attended at least one of the 12 weekly sessions. The average number of sessions attended was 10.3, indicating very low levels of dropout from the intervention among those who had started it. Due to some unrest in the study area, almost half of the groups did not hold the last session. Among the 39 women in T1 and T3 with elevated depression symptoms who did not participate in group therapy, 17 were unable to participate because the assigned CHF was unavailable to facilitate the group therapy intervention, 6 dropped because of pregnancy and 9 refused. Other reasons include security concerns or that the woman had moved. One CHF went on maternity leave prior to the IPT-G training. Of the 94 trained CHFs, one resigned, one was transferred, one was sick for an extended period and two were on maternity leave. From our study sample, there was a small imbalance in this delivery failure across the two treatment arms, as this affected 4 women in two T1 kebeles and 13 women in four T3 kebeles.

[15] Prior to the start of this intervention, the IPT-G Officers facilitated pilot IPT-G sessions for 8 weeks with non-study participants as part of their own training.

It should be noted that in the study period, households experienced numerous shocks including pest infestations, extreme weather events, and sporadic incidents of violent conflict. For example, in the Oromia region, 35 percent of households reported losing crops to desert locusts and almost 60 percent reported losing crops to fall army worm. The reported crop losses due to pests was less in the Amhara region, but in both regions more than half of respondents reported experiencing either a significant drought or flood event in the past 15 months. Social unrest, protests, and violence occurred sporadically across the Oromia region during this period and a major civil conflict with the northern region of Tigray started in November 2020.

Figure 1: Timeline of SPIR interventions and study activities



3. Empirical Strategy

3.1 Evaluation design

The impact evaluation design compares combinations of SPIR livelihoods and nutrition activities by randomly assigning 192 kebeles to one of the following four intervention arms (see Figure 2)[16] :

Treatment 1 (T1):	L* + N* + PSNP
Treatment 2 (T2):	L* + N + PSNP
Treatment 3 (T3):	L + N* + PSNP
Control (T4):	PSNP only

[16] This experimental trial is registered in the AEA RCT registry under number AEARCTR-0008281 and digital object identifier (DOI): 10.1257/rct.8281-1.0. The Pre-analysis Plan with a complete list of hypotheses to be tested is available in the AEA registry (Alderman et al. 2020).

Consistent with the graduation model design, the treatment arms in the experiment are integrated combinations of L, L*, N, and N*. The evaluation tests the relative effectiveness of those combinations.

As mentioned above, L* included a livelihood transfer for the poorest households which were identified based on an asset index constructed for all treatment arms from the baseline data; the poorest 10 out of 18 households in each kebele were classified as 'extremely poor', while the other 8 out of 18 are referred to as 'less poor'. Within the L* treatment arm, kebeles were further randomized, with extremely poor households in one half of the L* kebeles randomly assigned to receive the poultry-start up package, and extremely poor households in the remaining half of the L* kebeles receiving the cash grant.

Figure 2: SPIR experimental impact evaluation design

Intervention description	T1: L*+N* +PSNP	T2: L*+N +PSNP	T3: L+N* +PSNP	T4: PSNP only
PSNP consumption-support transfers for public work clients and direct support beneficiaries	✓	✓	✓	✓
L: VESA, targeted value chain training & financial service linkages	✓	✓	✓	
N: Nutrition behavior change communication (BCC), WASH	✓	✓	✓	
L*: L + poultry/cash livelihood transfers to extremely poor women	✓	✓		
N*: N + enhanced BCC + CPNP + male engagement + IPT-G for women with depression	✓		✓	

3.2 Sample

The study took place in 13 woredas across the Amhara and Oromia regions of Ethiopia.[17] In designing the study sample, we began with a list of all kebeles in which the PSNP operated within these woredas and dropped kebeles where VESAs had already been formed, the first step in implementation of the SPIR program. A total of 196 kebeles (115 in Amhara and 81 in Oromia) remained as a part of the study. Two of the 196 kebeles were subsequently dropped for having no PSNP clients, and 1 kebele (Ejartii in Daro Lebu) was later dropped for security reasons. A fourth kebele was dropped at the time of the baseline survey, leaving 192 kebeles in the study.

[17] After the baseline survey, 2 new woredas were created from the existing 13 woredas included in the study design at baseline, leading to 15 woredas in the study sample. We retain the original 13 woreda strata when controlling for study design in the treatment effect models during analysis.

Within the 192 kebeles, households were randomly selected based on the following inclusion criteria (1) be a PSNP client household, (2) have at least one child aged 0-35 months (index child), and (3) have the mother or primary female caregiver of the 0-35-month-old child be a member of the household. The last two criteria enabled measurement of maternal and child nutrition, a primary outcome in the overall evaluation. The baseline sampling process led to 3,314 households in the sample, or just over 17 households of PSNP clients in each kebele.

3.3 Primary outcomes

During the midline and endline surveys, the primary female respondent (mother or caregiver of the baseline index child or midline index child, see below) was screened for depression symptoms and functional effects of depression using the PHQ-9 symptom assessment tool. The PHQ-9 asks subjects to report the frequency with which they experienced each symptom of depression (e.g., feeling bad about yourself; feeling that you would be better off dead) over the previous two weeks, with coded responses ranging from 0 (not at all) to 3 (nearly every day). The tool yields a depression severity score from 0 to 27. Using this score, severity is sometimes classified in intervals as: none (0), minimal (1–4), mild (5–9), moderate (10–14), moderately severe (15–19), and severe (20–27) (Kroenke, Spitzer, and Williams 2001). We use the continuous PHQ-9 score as a primary outcome for depression. In addition, we include as a primary outcome the prevalence of individuals with PHQ-9 ≥ 8, and refer to individuals in this group as experiencing elevated depressive symptoms. This threshold level was also used by other recent studies of depression in rural Ethiopia (Tiki, Taye, and Duko 2020; Woldetensay, Belachew, Biesalski, et al. 2018). This decision was supported by the validation of the PHQ-9 in Afaan Oromo, where the optimal cut-off value for meeting desired sensitivity and specificity measures of the tool was recommended to be eight (Woldetensay, Belachew, Tesfaye, et al. 2018). Analysis of the sensitivity of the use of different cut-off scores of the PHQ-9 to screen for Major depressive disorder (MDD) found that there was no loss of sensitivity for studies that used a cut-off of 8 versus a commonly used cut-off score of 10 (Moriarty et al. 2015).[18]

In addition, we report impacts on MDD, a leading measure of depression indicating persistently depressed mood or loss of interest in activities causing significant impairment in daily life. MDD can be assessed using the responses on the PHQ-9, following an algorithm based on the Diagnostic and Statistical Manual of Mental Disorders fourth edition (DSM-IV) criteria. The algorithm requires that at least 5 symptoms are scored at least two or more, with the exception of the suicidal ideation question which qualifies with a score of one or more.[19] Another requirement is that one of the symptoms scored as either two or more must be either the question on no interest or pleasure in doing things or on feeling down or depressed.

3.4 Baseline balance

[18] Choosing this slightly lower cut-off also allowed the study to screen in the desired number of participants more efficiently in a given community – meeting the trade-offs of having a minimum recommended group size, while not requiring participants to travel far to get to the weekly group sessions.

[19] If survey respondents indicated any suicide ideation, enumerators would immediately follow a safety protocol by asking follow-up questions to determine the severity and in high-risk cases would connect these individuals to health services with the support of local SPIR staff and logistics.

Randomization of the four treatment arms provided balance in baseline individual and household characteristics. As shown in Table 1, imbalance occurs for only 3 out of 210 comparisons for 33 characteristics across 4 treatment arms, far less than the 5 percent of tests that would be expected to have p-values below 0.05 by chance. The imbalance occurs for variables on the number of household members aged 18-60 years old (between T2 and T3), the number of bedrooms (between T2 and control), and non-food expenditure per adult equivalent (between T3 and control). Balance at baseline is also achieved for the sample of women who were later screened with elevated depressive symptoms at midline (Table A1). The small degree of imbalance is only modestly higher than in the sample as a whole, with only 7 out of 210 comparisons leading to rejection of equality of means across arms. Again, this is less than would be predicted by chance. Variables demonstrating some imbalance include household size, share of female members of age 0-17 years, the value of nonfood consumption, whether the household experienced any illness in the last 2 years, and whether a female member experienced emotional violence or sexual violence in the past year. The prevalence of shocks was also balanced across treatment arms.

3.5 Attrition

From the baseline sample of 3,314 households, 3,220 households were interviewed at midline and 3,094 households at endline, three years post-baseline, representing a modest attrition rate of 6.6 percent. A sizeable portion of this attrition was due to insecurity in Amhara that prevented the survey team from conducting the endline survey in four of the kebeles. We test for the potential for attrition bias, comparing the attrition rates by treatment arm from baseline to endline. In the full sample, there is no statistically significant difference in attrition rates between T2 or T3 and the control group, but attrition in T1 is 2.04 percent lower than in the control group (p=0.010). In the sample of women screened for depression at midline, attrition is 2.6 percent lower in T1 than in the control group and 2.3 percent higher in T3 than in the control group, but these differences are not significant. These patterns suggest that selective attrition may lead to small differences in the composition of the sample between T1 and control in the full sample.

3.6 Estimation strategy

The specification for the treatment effects model estimates the impact of each treatment arm separately against the control (T4) using a simple difference model with the following specification:

$$Y_{1hvd} = \beta_0 + \beta_1 T1_{vd} + \beta_2 T2_{vd} + \beta_3 T3_{vd} + \beta_4 X_{0hvd} + \mu_d + \varepsilon_{hvd,} \qquad (1)$$

where $T1_{vd}$ is an indicator for whether household *h* in kebele *v* in woreda *d* was randomly assigned to treatment T1, $T2_{vd}$ indicates randomized assignment to T2, $T3_{vd}$ indicates randomized assignment to T3, and μ_d is a vector of dummy variables controlling for woreda fixed effects, which was the level of stratification used in the randomization of T1–T4. Standard errors are clustered at the kebele level. In addition, we correct for multiple hypothesis testing using a method introduced by Simes (1986), which adjusts inference for the multiplicity of tests estimated and results in a modified measure of statistical significance, the sharpened q-value. We also estimate equation (1) for multiple subsamples. This includes the subsample of women exhibiting elevated depressive symptoms at midline; the subsample of extremely poor women; and the intersection of the two (extremely poor women exhibiting elevated depressive symptoms).

4. Results

4.1 Depression outcomes

We measure the intent-to-treat effect of the SPIR graduation model on depression by estimating equation (1) using OLS. For the full sample of women, we find no significant effects of any of the three treatment arms, T1, T2, or T3, on the continuous PHQ-9 depression scale or on the probability of elevated depressive symptoms (PHQ-9 ≥ 8) either at midline or endline (Table 2). A similar pattern is observed when we separate the sample into the extremely poor and less poor sample, as shown in Columns (3) through (6) of the same table: for women in households in the extremely poor sample, there is no impact of the graduation program in any of the three treatment arms on the mean PHQ-9 score. Estimates show a significant impact in reducing the prevalence of elevated depressive symptoms (PHQ-9 ≥ 8) from T2 at both midline and endline, but neither effect survives adjustment for multiple hypothesis testing. Similarly, in the sample of women in less poor households, there is no significant impact of any of the graduation model treatments after controlling for multiple hypothesis testing.

Next, we examine the impact of SPIR on women who exhibited elevated depressive symptoms in the midline survey (Table 3). We find that T1, the arm that included enhanced livelihood programming (inclusive of transfers) and enhanced nutrition programming (inclusive of IPT-G), is the only treatment arm to lead to a significant reduction in depressive symptoms at endline. These combined components in T1 reduced women's depression severity score by 1.15 relative to a control group mean of 5.15 and reduced the prevalence of elevated depressive symptoms by 12.1 percentage points relative to a control group mean of 26.0 percent at endline. The average ITT effect of T1 on women who exhibited depression more than one year earlier are driven by effects on women in extremely poor households. For women in extremely poor households, T1 reduced depression severity by 1.73 (control mean of 6.1) and reduced the prevalence of depression by 17.3 percentage points (control mean of 32 percent). This is a large effect measured one year after the completion of IPT-G sessions and nearly two years after the livelihood transfers. Moreover, the impact of T1 on the prevalence of elevated depressive symptoms is meaningfully greater than the impact of T2 or of T3. Given the small samples and associated larger standard errors, these differences are in some cases noisily estimated, but they are clearly large.

During the intervention, we measured a steady and consistent decrease in average PHQ-9 scores during the 12 weekly sessions for participants in the IPT-G intervention, It is important to note that in this sample of women with elevated depressive symptoms at midline, there is a substantial secular improvement in mental health in the control group, where two thirds of these women no longer exhibited elevated depressive symptoms at endline, 15 months later. This likely reflects a common pattern in which, for many individuals, depressive episodes can recur periodically rather than persisting in an unremitting fashion over time (Otte et al. 2016). The results in Table 3 show that the combined treatment in T1 further reduced the prevalence of depression among this group at endline by 46.5 percent, including 54.1 percent among women in extremely poor households.

Among the same group of women with elevated depressive symptoms at midline, we find that the treatment arm (T3) that included group psychotherapy for these women but did not include livelihood transfers had no statistically significant effect on women's depression measured one-

year post-treatment. There were also no significant effects on depression measured for women in the T2 treatment, either among those that received livelihood transfers in the extremely poor sample of households or the full sample.

In Table 4, we report the effect of SPIR on Major depressive disorder, a more selective measure of depression, at endline for women that were screened for elevated depressive symptoms at midline. A pattern of results similar to Table 3 can be observed even when using this more restrictive measure of depression. T1 leads to a reduction in the prevalence of MDD at endline by 5.6 percentage points (from a control mean of 7.3 percent) and 8.9 percentage points for women in the extremely poor households (compared to a control mean of 10.7 percent), meaning that the combined enhanced livelihood and enhanced nutrition treatment in T1 addressed MDD for at least three out of four women. There is also a similar reduction of 8.7 percentage points among women in the extremely poor households in T2 that received the livelihood transfers, however, this effect is not significant after controlling for multiple hypothesis testing.

4.2 Other psychosocial outcomes

Next, we examine the impact of the SPIR program on stress, happiness, and self-efficacy at endline for women who had elevated depressive symptoms at midline (Table 5). Consistent with the pattern highlighted in previous tables, T1 is the only treatment arm with significant positive effects on these psychosocial outcomes. T1 increases women's reported happiness on a 1-4 scale by 0.364 in the full sample (with a control mean of 2.33), by 0.34 (weakly significant) in the extremely poor sample, and by 0.48 in the less poor sample of households. While not statistically significantly measured, the sign on the point estimate for reported stress levels is negative in T1, as compared to positive in T2 and T3 in the full sample, and this effect is statistically different from both the T2 and T3 effects. A statistically significant increase of 1.28 in stress is measured in the T2 arm for these women, with a control mean of 5.69 on 1-10 scale. No effects were measured on the self-efficacy scale for women in any of the treatment arms.

4.3 Economic outcomes

We also report the impact of SPIR on household's economic outcomes at endline for women who exhibited depressive symptoms at midline (Table 6). There are no effects on consumption expenditure or several asset indices (consumer durable, productive, and livestock assets). As several of the key livelihood activities supported by SPIR focused on livestock, including promotion of improved poultry production and short-cycle animal fattening and sales, we also present results for the impact on income from livestock sales or livestock product sales in the past year. Unlike traditional animal rearing, these livestock production activities can be managed directly at the household premises which allow women to play a more active role and take ownership of these activities. No statistically significant effect is measured in T2, but T1 and T3 both have a significant positive effect on the income reported for livestock sales (1.96 and 1.94 respectively on a control mean of 2.69[20]). This represents a 72 percent increase in income from these sales for women in the treatment arms that received the group therapy intervention. This effect size is nearly three times

[20] Treatment effects for income from livestock sales and from the sale of animal products are estimated using the inverse hyperbolic sine transformation to allow for non-linear effects.

the size of the effect measured in T1 and T2 in the full sample of women at endline.[21] No effects were measured at endline on the net income from sales of livestock products for these women.

4.4 Nutrition and health outcomes

Finally, we examine the impact of SPIR on women's dietary diversity and BMI, an important indicator of women's health in rural Ethiopia.[22] While T1 and T3 had an effect on dietary diversity[23] for the overall sample of women at midline (Alderman et al. 2021), none of the treatment arms had any effect on either the dietary diversity or minimum dietary diversity scores for women who exhibited elevated depressive symptoms at midline. The results in Table 7 show that the combined livelihood transfer and enhanced nutrition activities including the group therapy intervention in T1 significantly increases women's BMI for women that were screened for elevated depressive symptoms at midline, while the treatment arms that do not include group therapy (T2) or the livelihood transfer (T3) are not effective at improving this aspect of women's health. In the sample of women exhibiting depression at midline, T1 increases the z-score for women's BMI (BMIZ) by 0.41 standard deviations, a large effect. This increase in BMI is a positive result in this sample of extremely poor women, who exhibit substantial undernutrition: nearly one third of women in the control group were underweight (BMI<18.5) at endline, while only 2.4 percent of women were overweight or obese. These results show that relaxing the budget constraint alone is not enough to improve nutritional status for women who exhibited depression at midline, and enhanced nutrition activities without a livelihood transfer is similarly ineffective. It is only when depressed women participate in enhanced nutrition activities that includes group therapy and receive livelihoods transfers that they likely experience combined improvements in appetite and consumption that increases their BMI.

4.5 Robustness checks

We implement three robustness checks to determine the sensitivity of the findings to model specification, household exposure to a male engagement intervention, and the choice of the depression threshold score. Results are presented in the appendix. First, we estimate the models on the depression scale and depression prevalence including additional covariates to improve precision in the estimates and test the robustness of these findings to confounding factors. Added covariates include the household size, respondent's age, marital status, and education; the age of the primary respondent of the opposite sex (usually their spouse); an indicator of livestock assets; and exposure to shocks in the last 15 months (drought, flood, death of a family member, illness of a family member, or divorce). In the full sample, including these covariates produces a modest improvement in precision (Table A2). In addition, in the sample of women exhibiting depression at

[21] A weakly significant effect of 23-25 percent increase in livestock sales in T2 and T1 was measured in the full sample of women at endline.

[22] Body Mass Index is a widely used measure of adult nutrition. Using data from multiple countries and contexts, a WHO expert committee noted that a BMI below 16 was associated with extremely poor health, while a BMI below 17 was associated with a moderate risk and below 18.5 with a mild risk of ill health (WHO Expert Committee on Physical Status : the Use and Interpretation of Anthropometry (1993 : Geneva and Organization 1995). The cut-off of 18.5 became the accepted measure of underweight among adult populations.

[23] Effect of 4.6 percentage points over a control mean of 5.2 percent.

midline, estimated treatment effects of T1 increase somewhat in the model with covariates, leading to an estimated decline in prevalence of depression in the full sample of 15.5 percentage points and an estimated decline in prevalence of depression in the extremely poor sample of 17.9 percentage points (Table A3).

Second, we present results for models including a control variable for whether the female respondent's male partner or another adult male household member participated in a men's engagement intervention that was introduced as part of the enhanced nutrition program in T1 and T3 around the same time that women with depressive symptoms in T1 and T3 participated in the group therapy intervention (Table A4). The men's engagement intervention consisted of the formation of men's groups in which participants reflected on gender norms in their community and were encouraged to participate in household chores and other activities that were commonly assigned to women or girls. At endline, 40 percent of men in these treatment arms reported participating in at least one men's engagement group session. Results change only modestly after controlling for selected participation in the men's engagement intervention. The estimated impact on the mean depression score drops modestly in the full sample and the impact on the probability of a PHQ-9 score of 8 or higher increases from 17.3 to 20.2 percent in the extremely poor sample. The q-value of the depression score increases to 0.055 for the full sample and extremely poor sample, but the q-value for the threshold depression measures remain significant after controlling for multiple hypothesis testing.

Third, we examine how the impacts on the prevalence of depression change if a PHQ-9 cutoff of 10 is used instead of 8 (Table A5). With this change in threshold, the prevalence of depression changes from 26.0 percent to 14.6 percent in the control group of the full sample, and from 32.0 percent to 18.7 percent in the control group of the extremely poor sample. At the higher depression threshold score, the impact of T1 is smaller, falling to 6.6 percent in the full sample and to 9.5 percent in the extremely poor sample; neither estimate is significant after controlling for multiple hypothesis testing. Several factors could be responsible for this pattern of results. These results are consistent with the interpretation that the impacts of the T1 treatment on the prevalence of depression were driven primarily by impacts at the mild-to-moderate level of depression, notwithstanding the impacts estimated on Major depressive disorder. Alternatively, our model may be underpowered in this sample to detect the smaller impacts on the prevalence of depression at the threshold of 10 or higher. Given the substantial impacts on Major depressive disorder, a measure of more severe depression, we think it is unlikely that the impacts on prevalence of depression are driven primarily by those with a PHQ-9 score of 8 or 9.

4.6 Further mechanisms: Improvement in depression symptoms

To provide further evidence of the pathway of psychological improvement that produced these effects, we examined the impact of the program on the individual depression symptoms reported in the PHQ-9 scale. Table A6a reports the impacts for all 9 depression symptoms, or sub-components of the PHQ-9. These results suggest the reduction in depression caused by T1 is driven primarily by a large reduction in the frequency of feeling tired or having little energy, having a poor appetite or overeating, and feeling bad about yourself or that you are a failure. Impacts of T1 on each of these symptoms is statistically significant, although only impacts on feeling tired or having little energy remain significant after controlling for multiple hypothesis testing. We see these

results as suggestive of pathways of psychological improvement that may have contributed to the improvements in other outcomes. In particular, the large estimated reduction in frequency of having a poor appetite (which would be more common than overeating in this sample) is consistent with the impact of T1 on women's BMI. The reduction in frequency of feeling tired or having little energy is also consistent with improvements in other psychosocial outcomes and plausibly with more engagement in economic activities as measured by increased income from livestock sales.

We also test the impact of the treatment arms on the 9 depression symptoms in a model including covariates (Table 6b). In this model, the impact of T1 on having a poor appetite or overeating is somewhat larger and is statistically significant even when controlling for multiple hypothesis testing. This further supports that the impacts of T1 on women's BMI operate through women recovering their appetite. The pattern of impacts on feeling tired or having little energy and on feeling bad about yourself or that you are a failure are largely the same after adding covariates to the model. In this model, the impact of T1 on feeling down, depressed, or hopeless is larger than in the model without covariates, but is not significant when controlling for multiple hypothesis testing (q=0.056).

5. Conclusion

Our findings suggest that SPIR, a large, multifaceted graduation program, had no impact on women's depression on average. This result contrasts with results from a multi-country graduation program evaluation that found small, but persistent effects on mental health, an increase of 0.1 standard deviation at endline and declining to 0.07 standard deviations a year later (Banerjee et al. 2015). Two meta-analyses analyzing the effects of economic interventions, including graduation programs[24] and cash transfers, on mental health, similarly estimated average effect sizes of 0.07 to 0.1 standard deviations (Romero et al. 2021; McGuire, Kaiser, and Bach-Mortensen 2022). While we generally cannot reject the hypothesis that similar effects are estimated here, our study is also characterized by relatively small transfers ($200 provided to extremely poor households in two treatment arms), while both studies suggest that larger transfers were associated with larger effect sizes on mental health and psychological wellbeing outcomes, and the average value of the asset transfer in the multi-country study was $758 (Banerjee et al. 2015). The SPIR endline also followed a period of multiple, overlapping shocks, including social tension and conflict, desert locust and fall army worm infestations, and COVID-19 disruptions. These factors may have eroded any mental health gains and contributed to this null result, in terms of SPIR's average impact on depression.

Our study also provides estimates of impacts of SPIR treatments on depression for women who exhibited elevated depressive symptoms at midline, measured on average 18 months later. As our endline evaluation came more than one year after the midline, when these symptoms where initially identified, many of the depressive symptoms experienced by women in the control arm diminished over this period. For women who exhibited elevated depressive symptoms at midline but did not receive the enhanced health and nutrition programming, we also do not observe any statistically significant effects on depression at endline (T2).

[24] This meta-analysis included only 3 graduation program studies out of the 57 analyzed, including Banerjee et al. (2015).

Lastly, we turn to the effects on women identified with depressive symptoms at midline in treatment arm (T1) that received both enhanced livelihood and enhanced health and nutrition programming, including the invitation to participate in IPT-G. Here we observe large and statistically significant effects both on the continuous depression severity score and depression, as measured by a PHQ-9 score of 8 or higher. The prevalence of depression in the full sample in this group decreased 13 percentage points from a control mean of 25.8 percent (an average decrease of 50 percent). For those in the extremely poor households that also received livelihoods transfers, the decrease in depression was 17 percentage points, from a control mean of 32 percent (an average decrease of 53 percent). In the less poor households, the magnitude of the effects was not as large, and not statistically significantly measured. This may be in part due to the size of this sub-sample and the fact that many of the depressive symptoms appear to have resolved on their own in this group of women in less poor households.[25]

Evaluations of similar scalable group psychotherapy interventions have measured significant effects on mental health outcomes, including depression, immediately after treatment and 3-6 months afterwards (Singla et al. 2017). The results here contribute to a smaller set of studies that have assessed effects at one year or more after the intervention (Baranov et al. 2020; Bhat et al. 2022; Haushofer, Mudida, and Shapiro 2020; Blattman, Jamison, and Sheridan 2017). The sustained effects in the combined treatment arm (T1) for women screened with elevated depressive symptoms were significant in both the PHQ-9 cumulative score with a cut-off and the more restrictive algorithm to measure MDD. Reducing depressive symptoms, including somatic ones such as low energy and appetite, combined with improved economic activities measured by increased income from livestock sales among these women in T1 is a plausible pathway for the significantly improved wellbeing outcome, body mass index: we observe a 0.41 standard deviation increase in BMI z-score for women who expressed elevated depressive symptoms at midline. These results point to the importance of employing scalable group psychotherapy interventions to address depression within larger graduation model programs which combine a cash or asset transfer to achieve improved mental health and wellbeing outcomes for individuals in extremely poor and food insecure households.

For women suffering from depression in these areas, participation in a social safety net program that provided livelihoods, nutrition, and consumption support was not enough to achieve the objectives of the program. Even adding either a livelihood transfer or targeted group therapy to these programs did not achieve sustained effects. The combination of relaxing financial constraints and addressing depression through group psychotherapy was critical in achieving other psychosocial, economic, and physical well-being outcomes for these women in rural Ethiopia.

[25] The depression control mean in the less poor group of women at endline was 16.7 percent, as compared to the control mean of 32 percent in the extreme poor at endline.

References

Alderman, Harold, Lucy Billings, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Alemayehu Seyoum Taffesse, and Heleene Tambet. 2021. "Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA): Endline Report." https://ebrary.ifpri.org/digital/collection/p15738coll5/id/8084.

Alderman, Harold, Daniel O Gilligan, Melissa Hidrobo, Jessica Leight, Michael Mulford, and Alemayehu Seyoum Taffesse. 2020. "A Randomized Impact Evaluation of the Graduation Program Strengthen Productive Safety Net Program Phase 4 (PSNP4) Institutions and Resilience (SPIR) in Ethiopia." https://doi.org/https://doi.org/10.1257/rct.8281-2.1.

Angeles, Gustavo, Jacobus de Hoop, Sudhanshu Handa, Kelly Kilburn, Annamaria Milazzo, Amber Peterman, Sara Abdoulayi, et al. 2019. "Government of Malawi's Unconditional Cash Transfer Improves Youth Mental Health." *Social Science and Medicine* 225 (August 2018): 108–19. https://doi.org/10.1016/j.socscimed.2019.01.037.

Angelucci, Manuela, and Daniel Bennett. 2024. "The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy." *American Economic Review* 114(1):169-98.

Baird, Sarah, Jacobus de Hoop, and Berk Özler. 2013. "Income Shocks and Adolescent Mental Health." *Journal of Human Resources* 48 (2): 370–403. https://doi.org/10.3368/jhr.48.2.370.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236). https://doi.org/10.1126/science.1260799.

Banerjee, Abhijit V., Dean Karlan, Robert Darko Osei, Hannah Trachtman, and Christopher Udry. 2018. "Unpacking a Multi-Faceted Program to Build Sustainable Income for the Very Poor." *SSRN Electronic Journal*, 1–37. https://doi.org/10.2139/ssrn.3127053.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial." *American Economic Review* 110 (3): 824–59. https://doi.org/10.1257/aer.20180511.

Barbui, Corrado, Marianna Purgato, Jibril Abdulmalik, Ceren Acarturk, Julian Eaton, Chiara Gastaldon, Oye Gureje, et al. 2020. "Efficacy of Psychosocial Interventions for Mental Health Outcomes in Low-Income and Middle-Income Countries: An Umbrella Review." *The Lancet Psychiatry* 7 (2): 162–72. https://doi.org/10.1016/S2215-0366(19)30511-5.

Barker, Nathan, Gharad Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher Udry. 2022. "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress." *American Economic Review: Insights* 4 (4): 527–45. https://doi.org/10.1257/aeri.20210612.

Barrett, Christopher B., Michael Carter, Jean-Paul Chavas, and Michael R. Carter. 2018. *The Economics of Poverty Traps*. Chicago: University of Chicago Press: University of Chicago Press. https://doi.org/10.7208/chicago/9780226574448.001.0001.

Bedoya, Guadalupe, Aidan Coville, Johannes Haushofer, Mohammad Isaqzadeh, and Jeremy Shapiro. 2019. "No Household Left Behind: Afghanistan Targeting the Ultra Poor Impact Evaluation." Cambridge, MA. https://doi.org/10.3386/w25981.

Berhane, Guush, Daniel O. Gilligan, John Hoddinott, Neha Kumar, and Alemayehu Seyoum Taffesse. 2014. "Can Social Protection Work in Africa? The Impact of Ethiopia's Productive Safety Net Programme." *Economic Development and Cultural Change* 63 (3): 1–26. https://doi.org/10.1086/677753.

Bhat, Bhargav, Jonathan de Quidt, Johannes Haushofer, Vikram Patel, Gautam Rao, Frank Schilbach, and Pierre-Luc Vautrey. 2022. "The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes." *NBER Working Paper Series*. Cambridge, MA. https://doi.org/10.3386/w30011.

Blattman, Christopher, Sebastian Chaskel, Julian C Jamison, and Margaret Sheridan. 2022. "Cognitive Behavior Therapy Reduces Crime and Violence over 10 Years: Experimental Evidence *."

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (4): 1165–1206. https://doi.org/10.1257/aer.20150503.

Bolton, Paul, Judith Bass, Richard Neugebauer, Helen Verdeli, Kathleen F. Clougherty, Priya Wickramaratne, Liesbeth Speelman, Lincoln Ndogoni, and Myrna Weissman. 2003. "Group Interpersonal Psychotherapy for Depression in Rural Uganda." *JAMA* 289 (23): 3117. https://doi.org/10.1001/jama.289.23.3117.

Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand, et al. 2022. "Tackling Psychosocial and Capital Constraints to Alleviate Poverty." *Nature* 605 (7909): 291–97. https://doi.org/10.1038/s41586-022-04647-8.

Bryant, Richard A., Alison Schafer, Katie S. Dawson, Dorothy Anjuri, Caroline Mulili, Lincoln Ndogoni, Phiona Koyiet, et al. 2017. "Effectiveness of a Brief Behavioural Intervention on Psychological Distress among Women with a History of Gender-Based Violence in Urban Kenya: A Randomised Clinical Trial." *PLoS Medicine* 14 (8): 1–20. https://doi.org/10.1371/journal.pmed.1002371.

Cuijpers, Pim, Tara Donker, Myrna M. Weissman, Paula Ravitz, and Ioana A. Cristea. 2016. "Interpersonal Psychotherapy for Mental Health Problems: A Comprehensive Meta-Analysis." *American Journal of Psychiatry* 173 (7). https://doi.org/10.1176/appi.ajp.2015.15091141.

Cuijpers, Pim, Eirini Karyotaki, Mirjam Reijnders, Marianna Purgato, and Corrado Barbui. 2018. "Psychotherapies for Depression in Low- and Middle-Income Countries: A Meta-Analysis." *World Psychiatry* 17 (1): 90–101. https://doi.org/10.1002/wps.20493.

Fernald, Lia C.H., and Megan R. Gunnar. 2009. "Poverty-Alleviation Program Participation and Salivary Cortisol in Very Low-Income Children." *Social Science and Medicine* 68 (12): 2180–89. https://doi.org/10.1016/j.socscimed.2009.03.032.

GBD 2019 Mental Disorders Collaborators. 2022. "Global, Regional, and National Burden of 12 Mental Disorders in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019." *The Lancet Psychiatry* 9 (2): 137–50. https://doi.org/10.1016/S2215-0366(21)00395-3.

Genicot, Garance, and Debraj Ray. 2017. "Aspirations and Inequality." *Econometrica* 85 (2): 489–519. https://doi.org/10.3982/ecta13865.

Ghosal, Sayantan, Smarajit Jana, Anandi Mani, Sandip Mitra, and Sanchari Roy. 2022. "Sex Workers, Stigma, and Self-Image: Evidence from Kolkata Brothels." *The Review of Economics and Statistics* 104 (3): 431–48. https://doi.org/10.1162/rest_a_01013.

Gilligan, Daniel O., John Hoddinott, and Alemayehu Seyoum Taffesse. 2009. "The Impact of Ethiopia's Productive Safety Net Programme and Its Linkages." *Journal of Development Studies* 45 (10): 1684–1706. https://doi.org/10.1080/00220380902935907.

Haushofer, Johannes, and Ernst Fehr. 2014. "On the Psychology of Poverty." *Science* 344 (6186): 862–67. https://doi.org/10.1126/science.1232491.

Haushofer, Johannes, Robert Mudida, and Jeremy Shapiro. 2020. "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-Being." *SSRN Electronic Journal*. Cambridge, MA. https://doi.org/10.3386/w28106.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *Quarterly Journal of Economics* 131 (4): 1973–2042. https://doi.org/10.1093/qje/qjw025.Advance.

Hirvonen, Kalle, Elia A. Machado, Andrew M. Simons, and Vis Taraz. 2022. "More than a Safety Net: Ethiopia's Flagship Public Works Program Increases Tree Cover." *Global Environmental Change* 75 (July): 102549. https://doi.org/10.1016/j.gloenvcha.2022.102549.

Knippenberg, Erwin, and John Hoddinott. 2019. "Shocks, Social Protection, and Resilience: Evidence from Ethiopia." https://doi.org/10.21955/gatesopenres.1115281.1.

Kroenke, Kurt, Robert L Spitzer, and Janet B W Williams. 2001. "The PHQ-9." *Journal of General Internal Medicine* 16 (9): 606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Leight, Jessica, Daniel O. and Gilligan, Michael Mulford, Alemayehu Seyoum Taffesse, and Heleene Tambet. 2021. "Aspiring to More? New Evidence on the Effect of a Light-Touch Aspirations Intervention in Rural Ethiopia." https://ssrn.com/abstract=3981827.

Lund, Crick, Alison Breen, Alan J. Flisher, Ritsuko Kakuma, Joanne Corrigall, John A. Joska, Leslie Swartz, and Vikram Patel. 2010. "Poverty and Common Mental Disorders in Low and Middle Income Countries: A Systematic Review." *Social Science and Medicine* 71 (3): 517–28. https://doi.org/10.1016/j.socscimed.2010.04.027.

Lund, Crick, Mary De Silva, Sophie Plagerson, Sara Cooper, Dan Chisholm, Jishnu Das, Martin Knapp, and Vikram Patel. 2011. "Poverty and Mental Disorders: Breaking the Cycle in Low-Income and Middle-Income Countries." *The Lancet* 378 (9801): 1502–14. https://doi.org/10.1016/S0140-6736(11)60754-X.

Maselko, Joanna, Siham Sikander, Elizabeth L. Turner, Lisa M. Bates, Ikhlaq Ahmad, Najia Atif, Victoria Baranov, et al. 2020. "Effectiveness of a Peer-Delivered, Psychosocial Intervention on Maternal Depression and Child Development at 3 Years Postnatal: A Cluster Randomised Trial in Pakistan." *The Lancet Psychiatry* 7 (9): 775–87. https://doi.org/10.1016/S2215-0366(20)30258-3.

McGuire, Joel, Caspar Kaiser, and Anders M. Bach-Mortensen. 2022. "A Systematic Review and Meta-Analysis of the Impact of Cash Transfers on Subjective Well-Being and Mental Health in Low- and Middle-Income Countries." *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01252-z.

Moriarty, Andrew Stephen, Simon Gilbody, Dean McMillan, and Laura Manea. 2015. "Screening and Case Finding for Major Depressive Disorder Using the Patient Health Questionnaire (PHQ-9): A Meta-Analysis." *General Hospital Psychiatry* 37 (6): 567–76. https://doi.org/10.1016/j.genhosppsych.2015.06.012.

Ohrnberger, Julius, Eleonora Fichera, Matt Sutton, and Laura Anselmi. 2020. "The Worse the Better? Quantile Treatment Effects of a Conditional Cash Transfer Programme on Mental Health." *Health Policy and Planning* 35 (9): 1137–49. https://doi.org/10.1093/heapol/czaa079.

Otte, Christian, Stefan M. Gold, Brenda W. Penninx, Carmine M. Pariante, Amit Etkin, Maurizio Fava, David C. Mohr, and Alan F. Schatzberg. 2016. "Major Depressive Disorder." *Nature Reviews Disease Primers* 2 (1): 16065. https://doi.org/10.1038/nrdp.2016.65.

Ozer, Emily J., Lia C.H. Fernald, Ann Weber, Emily P. Flynn, and Tyler J. VanderWeele. 2011. "Does Alleviating Poverty Affect Mothers' Depressive Symptoms? A Quasi-Experimental Investigation of Mexico's Oportunidades Programme." *International Journal of Epidemiology* 40 (6): 1565–76. https://doi.org/10.1093/ije/dyr103.

Patel, Vikram, and Arthur Kleinman. 2003. "Poverty and Common Mental Disorders in Developing Countries." *Bulletin of the World Health Organization* 81 (8): 609–15. https://doi.org/10.1590/S0042-96862003000800011.

Quidt, Jonathan de, and Johannes Haushofer. 2016. "Depression for Economists." Cambridge, MA. https://doi.org/10.3386/w22973.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms." *Science* 370 (6522). https://doi.org/10.1126/science.aay0214.

Romero, Jimena, Kristina Esopo, Joel McGuire, and Johannes Haushofer. 2021. "The Effect of Economic Transfers on Psychological Well-Being." *Working Paper*.

Simes, R. John. 1986. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3): 751-754.

Singla, Daisy R., Brandon A. Kohrt, Laura K. Murray, Arpita Anand, Bruce F. Chorpita, and Vikram Patel. 2017. "Psychological Treatments for the World: Lessons from Low- and Middle-Income Countries." *Annual Review of Clinical Psychology* 13: 149–81. https://doi.org/10.1146/annurev-clinpsy-032816-045217.

Sipsma, Heather, Angela Ofori-Atta, Maureen Canavan, Isaac Osei-Akoto, Christopher Udry, and Elizabeth H. Bradley. 2013. "Poor Mental Health in Ghana: Who Is at Risk?" *BMC Public Health* 13 (1). https://doi.org/10.1186/1471-2458-13-288.

Tiki, Takele, Kefyalew Taye, and Bereket Duko. 2020. "Prevalence and Factors Associated with Depression among Pregnant Mothers in the West Shoa Zone, Ethiopia: A Community-Based Cross-Sectional Study." *Annals of General Psychiatry* 19 (1): 1–7. https://doi.org/10.1186/s12991-020-00275-6.

Vlassopoulos, Michael, Abu Siddique, Tabassum Rahman, Debayan Pakrashi, Asad Islam, and Firoz Ahmed. 2021. "Improving Women's Mental Health during a Pandemic." www.iza.org.

Vos, Theo, Abraham D Flaxman, Mohsen Naghavi, Rafael Lozano, Catherine Michaud, Majid Ezzati, Kenji Shibuya, et al. 2012. "Years Lived with Disability (YLDs) for 1160 Sequelae of 289 Diseases and Injuries 1990–2010: A Systematic Analysis for the Global Burden of Disease Study 2010." *The Lancet* 380 (9859): 2163–96. https://doi.org/10.1016/S0140-6736(12)61729-2.

WHO Expert Committee on Physical Status: the Use and Interpretation of Anthropometry (1993 : Geneva, Switzerland), and World Health Organization. 1995. "Physical Status : The Use of and Interpretation of Anthropometry , Report of a WHO Expert Committee." WHO Technical Report Series ; 854. Geneva: World Health Organization. https://iris.who.int/handle/10665/37003.

Woldetensay, Yitbarek Kidane, Tefera Belachew, Hans Konrad Biesalski, Shibani Ghosh, Maria Elena Lacruz, Veronika Scherbaum, and Eva Johanna Kantelhardt. 2018. "The Role of Nutrition, Intimate Partner Violence and Social Support in Prenatal Depressive Symptoms in Rural Ethiopia: Community Based Birth Cohort Study." *BMC Pregnancy and Childbirth* 18 (1): 374. https://doi.org/10.1186/s12884-018-2009-5.

Woldetensay, Yitbarek Kidane, Tefera Belachew, Markos Tesfaye, Kathryn Spielman, Hans Konrad Biesalski, Eva Johanna Kantelhardt, and Veronika Scherbaum. 2018. "Validation of the Patient Health Questionnaire (PHQ-9) as a Screening Tool for Depression in Pregnant Women: Afaan Oromo Version." Edited by Steven LoBello. *PLOS ONE* 13 (2): e0191782. https://doi.org/10.1371/journal.pone.0191782.

World Health Organization. 2016. "Group Interpersonal Therapy (IPT) for Depression." *World Health Organization* 1.

Table 1: Balance in baseline characteristics

	Mean and standard deviation				Tests of equality of means (p-value)					
	T1: L*+N*	T2: L*+N	T3: L+N*	T4: Control	T1 vs Control	T2 vs Control	T3 vs Control	T1 vs T2	T1 vs T3	T2 vs T3
Household size	5.824 (2.019)	5.877 (1.959)	5.718 (1.928)	5.799 (1.927)	0.909	0.699	0.690	0.822	0.650	0.459
Number of household members aged 0-5 years	1.640 (0.689)	1.671 (0.703)	1.664 (0.693)	1.695 (0.689)	0.468	0.756	0.677	0.677	0.751	0.918
Number of household members aged 6-17 years	2.059 (1.608)	2.042 (1.585)	1.979 (1.524)	2.023 (1.531)	0.819	0.890	0.746	0.920	0.637	0.676
Share of females among household members	0.513 (0.192)	0.503 (0.182)	0.519 (0.188)	0.521 (0.183)	0.477	0.068	0.814	0.338	0.619	0.109
Age of primary female	30.892 (7.752)	30.787 (6.848)	30.802 (7.326)	30.690 (7.033)	0.659	0.824	0.800	0.830	0.855	0.975
Age of primary male	38.418 (9.361)	38.158 (8.834)	38.148 (8.768)	38.034 (8.634)	0.524	0.839	0.829	0.695	0.643	0.985
Primary female is single	0.149 (0.357)	0.122 (0.328)	0.159 (0.366)	0.143 (0.350)	0.863	0.527	0.666	0.373	0.781	0.238
Primary male is single	0.005 (0.068)	0.010 (0.101)	0.009 (0.097)	0.006 (0.080)	0.664	0.422	0.526	0.212	0.286	0.877
Share of female adults with primary education (4 or more grades)	0.100 (0.287)	0.124 (0.320)	0.112 (0.305)	0.095 (0.282)	0.813	0.180	0.386	0.296	0.565	0.597
Share of male adults with primary education (4 or more grades)	0.182 (0.368)	0.210 (0.388)	0.171 (0.364)	0.179 (0.369)	0.903	0.250	0.706	0.305	0.625	0.140
Primary female has any education	0.187 (0.390)	0.231 (0.422)	0.207 (0.405)	0.197 (0.398)	0.746	0.280	0.730	0.172	0.514	0.466
Primary male has any education	0.334 (0.472)	0.343 (0.475)	0.352 (0.478)	0.329 (0.470)	0.890	0.720	0.511	0.825	0.624	0.818
Distance to household's nearest town (km)	13.952 (7.996)	11.391 (7.039)	13.361 (8.456)	14.230 (6.510)	0.854	0.041	0.566	0.102	0.725	0.207
Land size per capita (hectares)	0.207 (0.918)	0.220 (2.232)	0.360 (2.751)	0.160 (0.601)	0.356	0.469	0.211	0.889	0.355	0.431
Standardized Livestock Units	0.913 (0.929)	0.947 (0.854)	0.907 (0.848)	0.971 (0.960)	0.507	0.764	0.427	0.682	0.945	0.597
Household has no electricity	0.503 (0.500)	0.409 (0.492)	0.424 (0.495)	0.382 (0.486)	0.081	0.688	0.545	0.153	0.258	0.813
Household has improved roof material	0.460 (0.499)	0.459 (0.499)	0.418 (0.494)	0.371 (0.483)	0.177	0.177	0.444	0.980	0.504	0.512
Household has improved source of water - dry season	0.650 (0.477)	0.648 (0.478)	0.705 (0.456)	0.591 (0.492)	0.430	0.423	0.083	0.976	0.424	0.376

Number of bedrooms	1.414 (0.677)	1.501 (0.779)	1.407 (0.620)	1.346 (0.620)	0.311	0.045	0.384	0.246	0.913	0.222
Consumption expenditure per month per adult equivalent (Birr)	607.640 (521.091)	602.816 (638.539)	604.405 (501.966)	567.968 (493.788)	0.421	0.489	0.443	0.922	0.945	0.973
Household faced drought in the last two years	0.447 (0.498)	0.455 (0.498)	0.453 (0.498)	0.427 (0.495)	0.771	0.697	0.698	0.914	0.932	0.977
Household faced flood in the last two years	0.056 (0.231)	0.044 (0.205)	0.048 (0.214)	0.044 (0.205)	0.575	0.995	0.833	0.569	0.697	0.826
Household experienced a death or divorce/separation in the last two years	0.059 (0.236)	0.036 (0.187)	0.045 (0.208)	0.056 (0.230)	0.799	0.077	0.343	0.036	0.210	0.370
Household experienced illness in the last two years	0.101 (0.302)	0.076 (0.265)	0.095 (0.293)	0.105 (0.307)	0.865	0.190	0.646	0.283	0.784	0.421

Notes: Estimates from the DFSA SPIR baseline survey sample, N=3,314. Standard deviations are in parentheses. P-value is from the test of difference of means between the treatment arms. Standard errors are clustered at the kebele level.

Table 2: Treatment effects on female depression outcomes

	(1) Full sample: Depression severity score (PHQ-9, 0-27)	(2) Full sample: Depression severity score is 8 or higher	(3) Extremely poor: Depression severity score (PHQ-9, 0-27)	(4) Extremely poor: Depression severity score is 8 or higher	(5) Less poor: Depression severity score (PHQ-9, 0-27)	(6) Less poor: Depression severity score is 8 or higher
Endline						
T1	0.015	-0.017	-0.090	-0.040	0.154	0.013
	(0.256)	(0.020)	(0.346)	(0.028)	(0.269)	(0.022)
	[0.954]	[0.853]	[0.953]	[0.853]	[0.853]	[0.853]
T2	0.018	-0.026	-0.136	-0.054**	0.222	0.013
	(0.252)	(0.019)	(0.302)	(0.024)	(0.290)	(0.025)
	[0.942]	[0.525]	[0.784]	[0.148]	[0.784]	[0.784]
T3	0.181	0.006	0.118	-0.003	0.261	0.018
	(0.244)	(0.021)	(0.312)	(0.027)	(0.269)	(0.025)
	[0.919]	[0.919]	[0.919]	[0.919]	[0.919]	[0.919]
Test: T1 = T2	(0.989)	(0.672)	(0.891)	(0.592)	(0.820)	(0.974)
	[0.989]	[0.989]	[0.989]	[0.989]	[0.989]	[0.989]
Test: T2 = T3	(0.521)	(0.140)	(0.382)	(0.039)	(0.895)	(0.847)
	[0.782]	[0.420]	[0.765]	[0.232]	[0.895]	[0.895]
Test: T1 = T3	(0.519)	(0.293)	(0.538)	(0.190)	(0.697)	(0.853)
	[0.807]	[0.807]	[0.807]	[0.807]	[0.836]	[0.853]
Mean of control (T4)	2.920	0.126	3.169	0.147	2.574	0.096
N	3,011	3,011	1,723	1,723	1,288	1,288
Midline						
T1	-0.279	-0.027	-0.317	-0.044	-0.239	-0.003
	(0.280)	(0.025)	(0.333)	(0.030)	(0.388)	(0.037)
	[0.513]	[0.513]	[0.513]	[0.513]	[0.647]	[0.928]
T2	-0.200	-0.042*	-0.431	-0.065**	0.074	-0.015
	(0.281)	(0.024)	(0.324)	(0.028)	(0.411)	(0.038)
	[0.718]	[0.222]	[0.369]	[0.137]	[0.857]	[0.829]
T3	-0.420	-0.049*	-0.194	-0.027	-0.728*	-0.078**
	(0.292)	(0.026)	(0.342)	(0.032)	(0.398)	(0.037)
	[0.228]	[0.138]	[0.570]	[0.488]	[0.138]	[0.138]
Test: T1 = T2	(0.771)	(0.529)	(0.729)	(0.456)	(0.395)	(0.740)
	[0.771]	[0.771]	[0.771]	[0.771]	[0.771]	[0.771]
Test: T2 = T3	(0.431)	(0.808)	(0.470)	(0.189)	(0.036)	(0.082)
	[0.564]	[0.808]	[0.564]	[0.377]	[0.214]	[0.246]
Test: T1 = T3	(0.618)	(0.412)	(0.718)	(0.559)	(0.171)	(0.028)
	[0.718]	[0.718]	[0.718]	[0.718]	[0.514]	[0.170]
Mean of control (T4)	4.227	0.217	4.240	0.218	4.207	0.215
N	2,644	2,644	1,507	1,507	1,137	1,137

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table 3: Treatment effects on female depression outcomes at endline, women with elevated depression symptoms at midline

	(1) Full sample: Depression severity score (PHQ-9, 0-27)	(2) Full sample: Depression severity score is 8 or higher	(3) Extremely poor: Depression severity score (PHQ-9, 0-27)	(4) Extremely poor: Depression severity score is 8 or higher	(5) Less poor: Depression severity score (PHQ-9, 0-27)	(6) Less poor: Depression severity score is 8 or higher
T1	-1.149**	-0.121**	-1.728**	-0.173**	-0.637	-0.062
	(0.513)	(0.051)	(0.783)	(0.077)	(0.804)	(0.080)
	[0.044]	[0.044]	[0.044]	[0.044]	[0.444]	[0.444]
T2	0.114	0.026	-1.178	-0.062	1.168	0.103
	(0.706)	(0.069)	(0.959)	(0.102)	(0.822)	(0.076)
	[0.872]	[0.850]	[0.444]	[0.821]	[0.444]	[0.444]
T3	-0.124	-0.014	-0.517	-0.047	0.302	0.028
	(0.615)	(0.068)	(0.816)	(0.099)	(0.774)	(0.085)
	[0.840]	[0.840]	[0.840]	[0.840]	[0.840]	[0.840]
Test: T1 = T2	(0.039)	(0.016)	(0.545)	(0.248)	(0.026)	(0.024)
	[0.058]	[0.052]	[0.545]	[0.297]	[0.052]	[0.052]
Test: T2 = T3	(0.726)	(0.583)	(0.480)	(0.898)	(0.256)	(0.331)
	[0.871]	[0.871]	[0.871]	[0.898]	[0.871]	[0.871]
Test: T1 = T3	(0.039)	(0.068)	(0.100)	(0.163)	(0.206)	(0.256)
	[0.200]	[0.200]	[0.200]	[0.244]	[0.247]	[0.256]
Mean of control (T4)	5.146	0.260	6.053	0.320	3.729	0.167
N	444	444	249	249	195	195

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table 4: Treatment effects on female Major depressive disorder at endline, women with elevated depression symptoms at midline

	(1) Full sample: Meets criteria for Major depressive disorder	(2) Extremely poor: Meets criteria for Major depressive disorder	(3) Less poor: Meets criteria for Major depressive disorder
T1	-0.056**	-0.089**	-0.017
	(0.025)	(0.041)	(0.022)
	[0.048]	[0.048]	[0.434]
T2	-0.027	-0.087**	0.033
	(0.027)	(0.042)	(0.036)
	[0.356]	[0.121]	[0.356]
T3	-0.024	-0.035	-0.014
	(0.032)	(0.048)	(0.022)
	[0.542]	[0.542]	[0.542]
Test: T1 = T2	(0.117)	(0.942)	(0.062)
	[0.176]	[0.942]	[0.176]
Test: T2 = T3	(0.905)	(0.173)	(0.065)
	[0.905]	[0.260]	[0.195]
Test: T1 = T3	(0.138)	(0.104)	(0.583)
	[0.207]	[0.207]	[0.583]
Mean of control (T4)	0.073	0.107	0.021
N	444	249	195

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table 5: Stress, happiness, and locus of control at endline, women with elevated depression symptoms at midline

	(1) Full sample: Female's current stress level (1-10)	(2) Full sample: Female's current happiness level (1-4)	(3) Full sample: New General Self-Efficacy Scale for female (1-10)	(4) Extremely poor: Female's current stress level (1-10)	(5) Extremely poor: Female's current happiness level (1-4)	(6) Extremely poor: New General Self-Efficacy Scale for female (1-10)	(7) Less poor: Female's current stress level (1-10)	(8) Less poor: Female's current happiness level (1-4)	(9) Less poor: New General Self-Efficacy Scale for female (1-10)
T1	-0.593	0.364***	-0.168	-0.571	0.336*	0.046	-0.552	0.482***	-0.431
	(0.426)	(0.138)	(0.317)	(0.543)	(0.172)	(0.443)	(0.548)	(0.168)	(0.505)
	[0.372]	[0.043]	[0.672]	[0.473]	[0.158]	[0.918]	[0.473]	[0.043]	[0.507]
T2	0.439	0.047	-0.027	-0.389	0.215	0.353	1.283**	-0.073	-0.478
	(0.449)	(0.152)	(0.336)	(0.556)	(0.187)	(0.436)	(0.549)	(0.183)	(0.568)
	[0.729]	[0.855]	[0.937]	[0.729]	[0.729]	[0.729]	[0.194]	[0.855]	[0.729]
T3	0.312	0.021	0.264	-0.053	-0.060	0.310	0.928	0.208	0.227
	(0.448)	(0.144)	(0.388)	(0.509)	(0.176)	(0.468)	(0.656)	(0.208)	(0.612)
	[0.917]	[0.918]	[0.917]	[0.918]	[0.918]	[0.917]	[0.917]	[0.917]	[0.918]
Test: T1 = T2	(0.021)	(0.025)	(0.670)	(0.760)	(0.520)	(0.483)	(0.001)	(0.001)	(0.928)
	[0.056]	[0.056]	[0.855]	[0.855]	[0.780]	[0.780]	[0.003]	[0.003]	[0.928]
Test: T2 = T3	(0.784)	(0.859)	(0.477)	(0.564)	(0.153)	(0.927)	(0.542)	(0.118)	(0.253)
	[0.927]	[0.927]	[0.846]	[0.846]	[0.687]	[0.927]	[0.846]	[0.687]	[0.760]
Test: T1 = T3	(0.041)	(0.009)	(0.241)	(0.345)	(0.021)	(0.567)	(0.015)	(0.130)	(0.221)
	[0.092]	[0.064]	[0.310]	[0.389]	[0.064]	[0.567]	[0.064]	[0.234]	[0.310]
Mean of control (T4)	6.276	2.325	4.098	6.653	2.253	3.787	5.688	2.438	4.583
N	444	444	444	249	249	249	195	195	195

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table 6: Treatment effects on economic outcomes at endline, households with a woman with depression symptoms at midline

	(1) Total consumption expenditure per month per adult equivalent (Birr)	(2) Consumer Durable Asset Index	(3) Household Productive Asset Index	(4) Household Livestock Asset Index	(5) Net income from sales of livestock in past 12 months[1]	(6) Net income from sales of livestock products (eggs, milk, dairy, honey) in past 12 months[1]
T1	-0.049	0.130	0.178	0.086	1.963**	-0.121
	(0.096)	(0.206)	(0.230)	(0.208)	(0.901)	(0.471)
	[0.797]	[0.797]	[0.797]	[0.797]	[0.186]	[0.797]
T2	-0.022	0.217	0.309	-0.095	1.257	0.044
	(0.113)	(0.205)	(0.217)	(0.197)	(0.901)	(0.432)
	[0.918]	[0.583]	[0.496]	[0.918]	[0.496]	[0.918]
T3	0.009	0.063	0.236	-0.074	1.936**	0.432
	(0.096)	(0.161)	(0.194)	(0.218)	(0.836)	(0.471)
	[0.927]	[0.882]	[0.672]	[0.882]	[0.132]	[0.722]
Test: T1 = T2	(0.796)	(0.719)	(0.592)	(0.375)	(0.447)	(0.701)
	[0.796]	[0.796]	[0.796]	[0.796]	[0.796]	[0.796]
Test: T2 = T3	(0.766)	(0.424)	(0.726)	(0.921)	(0.451)	(0.386)
	[0.919]	[0.902]	[0.919]	[0.921]	[0.902]	[0.902]
Test: T1 = T3	(0.507)	(0.724)	(0.791)	(0.479)	(0.977)	(0.240)
	[0.949]	[0.949]	[0.949]	[0.949]	[0.977]	[0.949]
Mean of control (T4)	7.134	-0.172	-0.688	-0.101	2.693	1.662
N	429	441	442	442	442	441

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986). Inverse hyperbolic sine transformation has been applied to outcome variables presented in columns (5) and (6).

Table 7: Treatment effects on anthropometry and dietary diversity at endline, women with elevated depression symptoms at midline

	(1) Women's Dietary Diversity Score (1-10)[1]	(2) Met Minimum Dietary Diversity for Women (MDD-W)[1]	(3) Body Mass Index (BMI) z-score	(4) Woman is underweight (BMI<18.5)	(5) Woman is overweight or obese (BMI>=25)
T1	0.031	-0.003	0.414***	-0.127*	0.043
	(0.161)	(0.037)	(0.151)	(0.067)	(0.033)
	[0.927]	[0.927]	[0.034]	[0.147]	[0.315]
T2	-0.201	-0.019	0.087	0.027	0.043*
	(0.168)	(0.035)	(0.135)	(0.078)	(0.025)
	[0.582]	[0.731]	[0.731]	[0.731]	[0.408]
T3	-0.025	-0.030	0.114	-0.022	0.042
	(0.196)	(0.038)	(0.146)	(0.070)	(0.033)
	[0.900]	[0.724]	[0.724]	[0.900]	[0.724]
Test: T1 = T2	(0.145)	(0.659)	(0.034)	(0.014)	(0.998)
	[0.242]	[0.824]	[0.084]	[0.070]	[0.998]
Test: T2 = T3	(0.376)	(0.768)	(0.849)	(0.444)	(0.969)
	[0.969]	[0.969]	[0.969]	[0.969]	[0.969]
Test: T1 = T3	(0.755)	(0.481)	(0.058)	(0.036)	(0.974)
	[0.944]	[0.802]	[0.145]	[0.145]	[0.974]
Mean of control (T4)	2.659	0.089	-0.154	0.317	0.024
N	444	444	442	442	442

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Appendix

Table A1: Balance in baseline characteristics, women with elevated depression symptoms at midline

	Mean and standard deviation				Tests of equality of means (p-value)					
	T1: L*+N*	T2: L*+N	T3: L+N*	T4: Control	T1 vs Control	T2 vs Control	T3 vs Control	T1 vs T2	T1 vs T3	T2 vs T3
Household size	6.596 (2.345)	6.893 (2.166)	6.229 (2.220)	6.142 (1.855)	0.295	0.043	0.785	0.524	0.392	0.069
Number of household members aged 0-5 years	1.833 (0.786)	1.875 (0.761)	1.762 (0.741)	1.811 (0.639)	0.859	0.597	0.671	0.761	0.589	0.379
Number of household members aged 6-17 years	2.649 (1.735)	2.741 (1.759)	2.448 (1.641)	2.236 (1.477)	0.187	0.047	0.330	0.785	0.519	0.247
Share of females among household members	0.510 (0.192)	0.473 (0.179)	0.523 (0.202)	0.544 (0.177)	0.120	0.001	0.304	0.125	0.581	0.031
Age of primary female	32.407 (7.828)	32.220 (6.172)	32.171 (7.677)	32.294 (7.309)	0.920	0.943	0.906	0.857	0.821	0.959
Age of primary male	39.883 (8.853)	39.788 (8.501)	38.734 (9.177)	39.623 (9.329)	0.853	0.913	0.527	0.942	0.339	0.422
Primary female is single	0.155 (0.363)	0.110 (0.314)	0.229 (0.422)	0.145 (0.354)	0.879	0.572	0.244	0.424	0.261	0.076
Primary male is single	0.011 (0.103)	0.020 (0.141)	0.000 (0.000)	0.000 (0.000)	0.300	0.122		0.562	0.301	0.122
Share of female adults with primary education (4 or more grades)	0.110 (0.310)	0.062 (0.234)	0.060 (0.224)	0.058 (0.222)	0.192	0.886	0.933	0.252	0.208	0.947
Share of male adults with primary education (4 or more grades)	0.218 (0.404)	0.268 (0.413)	0.162 (0.357)	0.152 (0.347)	0.264	0.076	0.815	0.511	0.348	0.107
Primary female has any education	0.182 (0.387)	0.165 (0.373)	0.162 (0.370)	0.172 (0.379)	0.848	0.888	0.829	0.754	0.698	0.949
Primary male has any education	0.372 (0.486)	0.364 (0.483)	0.278 (0.451)	0.330 (0.473)	0.606	0.689	0.485	0.921	0.231	0.289
Distance to household's nearest town (km)	14.458 (8.100)	10.687 (6.041)	12.683 (10.347)	13.297 (6.205)	0.560	0.106	0.784	0.045	0.465	0.347

Land size per capita (hectares)	0.315 (2.042)	0.107 (0.088)	0.185 (0.605)	0.115 (0.097)	0.296	0.639	0.262	0.278	0.515	0.209
Standardized Livestock Units	0.810 (0.786)	0.989 (0.918)	0.921 (1.025)	0.884 (0.900)	0.552	0.506	0.803	0.249	0.453	0.704
Household has no electricity	0.321 (0.469)	0.315 (0.467)	0.357 (0.482)	0.241 (0.430)	0.329	0.349	0.153	0.941	0.679	0.615
Household has improved roof material	0.500 (0.502)	0.402 (0.492)	0.476 (0.502)	0.433 (0.497)	0.456	0.728	0.609	0.281	0.779	0.385
Household has improved source of water - dry season	0.658 (0.477)	0.661 (0.476)	0.581 (0.496)	0.622 (0.487)	0.713	0.679	0.641	0.978	0.424	0.387
Number of bedrooms	1.465 (0.598)	1.420 (0.666)	1.514 (0.590)	1.457 (0.732)	0.932	0.708	0.555	0.632	0.596	0.322
Consumption expenditure per month per adult equivalent (Birr)	686.883 (752.055)	538.786 (561.182)	606.333 (588.668)	514.610 (485.947)	0.115	0.783	0.291	0.180	0.461	0.445
Household faced drought in the last two years	0.614 (0.489)	0.571 (0.497)	0.638 (0.483)	0.630 (0.485)	0.874	0.562	0.928	0.684	0.800	0.486
Household faced flood in the last two years	0.088 (0.284)	0.062 (0.243)	0.057 (0.233)	0.031 (0.175)	0.156	0.351	0.323	0.589	0.465	0.882
Household experienced a death or divorce/separation in the last two years	0.061 (0.241)	0.062 (0.243)	0.057 (0.233)	0.055 (0.229)	0.836	0.820	0.951	0.971	0.889	0.869
Household experienced illness in the last two years	0.123 (0.330)	0.036 (0.186)	0.114 (0.320)	0.150 (0.358)	0.541	0.008	0.491	0.036	0.865	0.110

Notes: Estimates from the DFSA SPIR baseline survey sample, N=475. Standard deviations are in parentheses. P-value is from the test of difference of means between the treatment arms. Standard errors are clustered at the kebele level.

Table A2: Treatment effects on female depression outcomes with covariates

	(1) Full sample: Depression severity score (PHQ-9, 0-27)	(2) Full sample: Depression severity score is 8 or higher	(3) Extremely poor: Depression severity score (PHQ-9, 0-27)	(4) Extremely poor: Depression severity score is 8 or higher	(5) Less poor: Depression severity score (PHQ-9, 0-27)	(6) Less poor: Depression severity score is 8 or higher
T1: L* + N*	-0.033	-0.020	-0.117	-0.038	0.124	0.012
	(0.246)	(0.020)	(0.337)	(0.027)	(0.266)	(0.023)
	[0.894]	[0.874]	[0.874]	[0.874]	[0.874]	[0.874]
T2: L* + N	0.016	-0.025	-0.041	-0.046*	0.121	0.008
	(0.237)	(0.019)	(0.292)	(0.024)	(0.277)	(0.025)
	[0.945]	[0.545]	[0.945]	[0.306]	[0.945]	[0.945]
T3: L + N*	0.124	0.003	0.111	-0.000	0.165	0.014
	(0.233)	(0.020)	(0.299)	(0.026)	(0.265)	(0.025)
	[0.994]	[0.994]	[0.994]	[0.994]	[0.994]	[0.994]
Household size	0.036	0.000	0.046	0.004	0.033	-0.005
	(0.044)	(0.004)	(0.055)	(0.006)	(0.066)	(0.006)
	[0.748]	[0.996]	[0.748]	[0.748]	[0.748]	[0.748]
Age of primary female	0.022*	0.001	0.028*	0.001	0.015	0.002
	(0.013)	(0.001)	(0.017)	(0.001)	(0.022)	(0.002)
	[0.294]	[0.428]	[0.294]	[0.428]	[0.487]	[0.428]
Age of primary male	0.017	0.001	0.015	0.001	0.018	0.001
	(0.011)	(0.001)	(0.016)	(0.001)	(0.017)	(0.002)
	[0.573]	[0.573]	[0.573]	[0.736]	[0.573]	[0.736]
Primary female is single	0.067	-0.025	-0.100	-0.054	0.514	0.074
	(0.428)	(0.043)	(0.548)	(0.057)	(0.582)	(0.048)
	[0.875]	[0.840]	[0.875]	[0.758]	[0.758]	[0.758]
Primary female has some education	0.075	0.001	0.194	0.009	-0.067	-0.012
	(0.176)	(0.015)	(0.245)	(0.020)	(0.238)	(0.021)
	[0.932]	[0.932]	[0.932]	[0.932]	[0.932]	[0.932]
Standardized Livestock Units	-0.235***	-0.011*	-0.176	-0.011	-0.234***	-0.010
	(0.064)	(0.006)	(0.137)	(0.012)	(0.087)	(0.008)
	[0.002]	[0.121]	[0.272]	[0.331]	[0.023]	[0.272]
Household was affected by drought(s) in the last 15 months	0.839***	0.034**	0.960***	0.024	0.696***	0.043**
	(0.184)	(0.015)	(0.253)	(0.021)	(0.233)	(0.021)
	[0.000]	[0.042]	[0.001]	[0.247]	[0.006]	[0.052]
Household was affected by flood(s) in the last 15 months	0.200	-0.002	0.095	0.002	0.322	-0.004
	(0.180)	(0.015)	(0.245)	(0.020)	(0.229)	(0.020)
	[0.806]	[0.914]	[0.914]	[0.914]	[0.806]	[0.914]
Household was affected by death in the last 15 months	-0.715	-0.050	-0.178	0.057	-1.366**	-0.164***
	(0.522)	(0.054)	(0.921)	(0.099)	(0.530)	(0.024)
	[0.345]	[0.528]	[0.847]	[0.684]	[0.032]	[0.000]
Household was affected by illness in the last 15 months	0.747***	0.044**	0.771***	0.042*	0.760**	0.050*
	(0.209)	(0.018)	(0.293)	(0.023)	(0.298)	(0.029)
	[0.003]	[0.021]	[0.021]	[0.076]	[0.021]	[0.089]
Household was affected by divorce in the last 15 months	0.167	-0.017	0.479	-0.021	-0.476	-0.003
	(0.813)	(0.057)	(1.082)	(0.072)	(0.837)	(0.096)
	[0.978]	[0.978]	[0.978]	[0.978]	[0.978]	[0.978]
Mean of control (T4)	2.920	0.126	3.169	0.147	2.574	0.096
N	3,011	3,011	1,723	1,723	1,288	1,288

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are

calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table A3: Treatment effects on female depression outcomes with covariates, women with depression symptoms at midline

	(1) Full sample: Depression severity score (PHQ-9, 0-27)	(2) Full sample: Depression severity score is 8 or higher	(3) Extremely poor: Depression severity score (PHQ-9, 0-27)	(4) Extremely poor: Depression severity score is 8 or higher	(5) Less poor: Depression severity score (PHQ-9, 0-27)	(6) Less poor: Depression severity score is 8 or higher
T1: L* + N*	-1.335***	-0.139***	-2.090***	-0.205**	-0.219	-0.023
	(0.441)	(0.046)	(0.773)	(0.078)	(0.757)	(0.080)
	[0.009]	[0.009]	[0.015]	[0.015]	[0.774]	[0.774]
T2: L* + N	0.112	0.023	-0.899	-0.044	1.201	0.110
	(0.616)	(0.061)	(0.916)	(0.099)	(0.748)	(0.071)
	[0.855]	[0.849]	[0.657]	[0.849]	[0.368]	[0.368]
T3: L + N*	-0.421	-0.034	-0.928	-0.069	0.489	0.054
	(0.571)	(0.066)	(0.827)	(0.104)	(0.744)	(0.079)
	[0.608]	[0.608]	[0.608]	[0.608]	[0.608]	[0.608]
Household size	-0.121	-0.011	-0.020	0.003	-0.269	-0.037*
	(0.113)	(0.014)	(0.162)	(0.018)	(0.169)	(0.020)
	[0.574]	[0.655]	[0.902]	[0.902]	[0.347]	[0.347]
Age of primary female	0.043	0.003	0.046	0.001	-0.015	0.005
	(0.040)	(0.004)	(0.043)	(0.004)	(0.105)	(0.009)
	[0.886]	[0.886]	[0.886]	[0.886]	[0.886]	[0.886]
Age of primary male	-0.022	-0.001	-0.031	-0.000	0.015	-0.001
	(0.037)	(0.003)	(0.046)	(0.004)	(0.087)	(0.008)
	[0.973]	[0.973]	[0.973]	[0.973]	[0.973]	[0.973]
Primary female is single	-1.935	-0.161	-2.366	-0.168	-1.942	-0.093
	(1.267)	(0.138)	(1.654)	(0.177)	(1.247)	(0.137)
	[0.310]	[0.371]	[0.310]	[0.413]	[0.310]	[0.497]
Primary female has some education	0.214	0.035	0.898	0.087	-0.927	-0.059
	(0.670)	(0.069)	(0.866)	(0.086)	(1.001)	(0.099)
	[0.750]	[0.740]	[0.713]	[0.713]	[0.713]	[0.740]
Standardized Livestock Units	-0.744***	-0.061***	-0.785**	-0.082**	-0.734**	-0.057**
	(0.210)	(0.020)	(0.374)	(0.035)	(0.284)	(0.028)
	[0.003]	[0.008]	[0.041]	[0.033]	[0.022]	[0.041]
Household was affected by drought(s) in the last 15 months	0.914	0.088	1.615	0.099	0.789	0.109*
	(0.583)	(0.054)	(1.038)	(0.091)	(0.567)	(0.056)
	[0.183]	[0.183]	[0.183]	[0.279]	[0.201]	[0.183]
Household was affected by flood(s) in the last 15 months	1.448***	0.131***	0.630	0.100	1.964***	0.146**
	(0.495)	(0.050)	(0.755)	(0.075)	(0.669)	(0.069)
	[0.012]	[0.020]	[0.406]	[0.223]	[0.012]	[0.055]
Household was affected by death in the last 15 months	-1.868	0.003	-1.678	0.110	-2.408	-0.133
	(1.446)	(0.215)	(2.254)	(0.340)	(1.546)	(0.099)
	[0.397]	[0.990]	[0.687]	[0.895]	[0.397]	[0.397]
Household was affected by illness in the last 15 months	0.760	0.070	1.034	0.081	0.125	0.037
	(0.592)	(0.060)	(0.847)	(0.087)	(0.792)	(0.074)
	[0.493]	[0.493]	[0.493]	[0.534]	[0.875]	[0.743]
Household was affected by divorce in the last 15 months	-0.222	0.043	0.372	0.059	-1.263	-0.071
	(1.468)	(0.201)	(1.748)	(0.281)	(1.797)	(0.111)
	[0.880]	[0.880]	[0.880]	[0.880]	[0.880]	[0.880]
Mean of control (T4)	5.146	0.260	6.053	0.320	3.729	0.167
N	444	444	249	249	195	195

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level.

All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table A4: Treatment effects on female depression outcomes, women with elevated depression symptoms at midline: Controlling for spouse's participation in male engagement

	(1) Full sample: Depression severity score (PHQ-9, 0-27)	(2) Full sample: Depression severity score is 8 or higher	(3) Extremely poor: Depression severity score (PHQ-9, 0-27)	(4) Extremely poor: Depression severity score is 8 or higher	(5) Less poor: Depression severity score (PHQ-9, 0-27)	(6) Less poor: Depression severity score is 8 or higher
T1	-1.107**	-0.136**	-1.697**	-0.202**	-0.459	-0.056
	(0.525)	(0.052)	(0.802)	(0.080)	(0.849)	(0.081)
	[0.055]	[0.037]	[0.055]	[0.037]	[0.590]	[0.590]
T2	0.137	0.016	-1.145	-0.083	1.308	0.110
	(0.730)	(0.069)	(0.980)	(0.104)	(0.841)	(0.074)
	[0.851]	[0.851]	[0.490]	[0.640]	[0.429]	[0.429]
T3	-0.102	-0.023	-0.474	-0.070	0.488	0.041
	(0.622)	(0.067)	(0.823)	(0.093)	(0.804)	(0.085)
	[0.869]	[0.869]	[0.869]	[0.869]	[0.869]	[0.869]
Spouse participated in male engagement	-0.097	0.051	-0.092	0.079	-0.215	0.012
	(0.499)	(0.051)	(0.669)	(0.073)	(0.656)	(0.072)
	[0.891]	[0.891]	[0.891]	[0.891]	[0.891]	[0.891]
Missing value for spouse's participation	0.106	0.046	-0.268	0.048	0.852	0.077
	(0.542)	(0.059)	(0.743)	(0.086)	(0.843)	(0.098)
	[0.845]	[0.845]	[0.845]	[0.845]	[0.845]	[0.845]
Test: T1 = T2	(0.040)	(0.014)	(0.543)	(0.216)	(0.031)	(0.027)
	[0.060]	[0.060]	[0.543]	[0.259]	[0.060]	[0.060]
Test: T2 = T3	(0.727)	(0.585)	(0.475)	(0.906)	(0.296)	(0.379)
	[0.872]	[0.872]	[0.872]	[0.906]	[0.872]	[0.872]
Test: T1 = T3	(0.043)	(0.045)	(0.103)	(0.128)	(0.224)	(0.243)
	[0.136]	[0.136]	[0.192]	[0.192]	[0.243]	[0.243]
Mean of control (T4)	5.146	0.260	6.053	0.320	3.729	0.167
N	444	444	249	249	195	195

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table A5: Treatment effects on female depression outcomes, women with elevated depression symptoms at midline: Switching depression score threshold from 8 to 10

	(1) Full sample: Depression severity score (PHQ-9, 0-27)	(2) Full sample: Depression severity score is 10 or higher	(3) Extremely poor: Depression severity score (PHQ-9, 0-27)	(4) Extremely poor: Depression severity score is 10 or higher	(5) Less poor: Depression severity score (PHQ-9, 0-27)	(6) Less poor: Depression severity score is 10 or higher
T1	-1.149**	-0.066*	-1.728**	-0.095	-0.637	-0.043
	(0.513)	(0.037)	(0.783)	(0.063)	(0.804)	(0.051)
	[0.087]	[0.156]	[0.087]	[0.203]	[0.430]	[0.430]
T2	0.114	0.003	-1.178	-0.058	1.168	0.051
	(0.706)	(0.045)	(0.959)	(0.073)	(0.822)	(0.056)
	[0.951]	[0.951]	[0.640]	[0.640]	[0.640]	[0.640]
T3	-0.124	-0.063	-0.517	-0.079	0.302	-0.050
	(0.615)	(0.041)	(0.816)	(0.061)	(0.774)	(0.055)
	[0.840]	[0.587]	[0.792]	[0.587]	[0.836]	[0.727]
Test: T1 = T2	(0.039)	(0.095)	(0.545)	(0.582)	(0.026)	(0.109)
	[0.117]	[0.164]	[0.582]	[0.582]	[0.117]	[0.164]
Test: T2 = T3	(0.726)	(0.134)	(0.480)	(0.750)	(0.256)	(0.117)
	[0.750]	[0.402]	[0.720]	[0.750]	[0.512]	[0.402]
Test: T1 = T3	(0.039)	(0.943)	(0.100)	(0.753)	(0.206)	(0.888)
	[0.233]	[0.943]	[0.300]	[0.943]	[0.412]	[0.943]
Mean of control (T4)	5.146	0.146	6.053	0.187	3.729	0.083
N	444	444	249	249	195	195

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table A6a: Treatment effects on female depression outcomes, women with elevated depression symptoms at midline: Individual questions from the PHQ-9 scale

	(1) Having little interest or pleasure in doing things	(2) Feeling down, depressed or hopeless	(3) Having trouble falling asleep, staying asleep or sleeping too much	(4) Feeling tired or having little energy	(5) Having a poor appetite or overeating	(6) Feeling bad about yourself or that are a failure	(7) Having trouble concentrating on things, such as cooking food, etc.	(8) Moving or speaking so slowly that other people could have noticed, or the opposite	(9) Having thoughts that you would be better off dead or of hurting yourself
T1	-0.132	-0.165*	-0.064	-0.245***	-0.163**	-0.206**	-0.096	-0.086	0.018
	(0.092)	(0.089)	(0.103)	(0.080)	(0.072)	(0.104)	(0.091)	(0.093)	(0.073)
	[0.273]	[0.151]	[0.598]	[0.025]	[0.109]	[0.149]	[0.442]	[0.459]	[0.805]
T2	0.082	-0.092	-0.073	0.098	0.030	-0.012	-0.083	0.018	0.146
	(0.125)	(0.088)	(0.086)	(0.109)	(0.095)	(0.128)	(0.092)	(0.104)	(0.116)
	[0.773]	[0.717]	[0.717]	[0.717]	[0.928]	[0.928]	[0.717]	[0.928]	[0.717]
T3	0.026	-0.018	0.009	0.016	0.039	-0.091	-0.068	-0.079	0.044
	(0.103)	(0.085)	(0.115)	(0.098)	(0.099)	(0.110)	(0.094)	(0.087)	(0.081)
	[0.940]	[0.940]	[0.940]	[0.940]	[0.940]	[0.940]	[0.940]	[0.940]	[0.940]
Test: T1 = T2	(0.059)	(0.437)	(0.933)	(0.000)	(0.024)	(0.064)	(0.887)	(0.232)	(0.211)
	[0.144]	[0.562]	[0.933]	[0.001]	[0.109]	[0.144]	[0.933]	[0.348]	[0.348]
Test: T2 = T3	(0.651)	(0.408)	(0.459)	(0.428)	(0.938)	(0.474)	(0.862)	(0.228)	(0.340)
	[0.837]	[0.711]	[0.711]	[0.711]	[0.938]	[0.711]	[0.938]	[0.711]	[0.711]
Test: T1 = T3	(0.095)	(0.096)	(0.549)	(0.001)	(0.032)	(0.189)	(0.758)	(0.917)	(0.666)
	[0.216]	[0.216]	[0.823]	[0.011]	[0.145]	[0.340]	[0.853]	[0.917]	[0.853]
Mean of control (T4)	0.553	0.748	0.667	0.764	0.496	0.659	0.561	0.423	0.276
N	444	443	444	444	444	444	444	444	444

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Table A6b: Treatment effects on female depression outcomes, women with elevated depression symptoms at midline: Individual questions from the PHQ-9 scale with covariates

	(1) Having little interest or pleasure in doing things	(2) Feeling down, depressed or hopeless	(3) Having trouble falling asleep, staying asleep or sleeping too much	(4) Feeling tired or having little energy	(5) Having a poor appetite or overeating	(6) Feeling bad about yourself or that are a failure	(7) Having trouble concentrating on things, such as cooking food, etc.	(8) Moving or speaking so slowly that other people could have noticed, or the opposite	(9) Having thoughts that you would be better off dead or of hurting yourself
T1: L* + N*	-0.157*	-0.194**	-0.091	-0.251***	-0.192***	-0.231**	-0.092	-0.117	-0.003
	(0.088)	(0.085)	(0.099)	(0.076)	(0.072)	(0.098)	(0.091)	(0.086)	(0.072)
	[0.136]	[0.056]	[0.408]	[0.012]	[0.038]	[0.056]	[0.401]	[0.261]	[0.966]
T2: L* + N	0.105	-0.081	-0.045	0.095	0.027	-0.012	-0.080	-0.005	0.108
	(0.124)	(0.088)	(0.086)	(0.100)	(0.092)	(0.116)	(0.092)	(0.100)	(0.109)
	[0.718]	[0.718]	[0.902]	[0.718]	[0.962]	[0.962]	[0.718]	[0.962]	[0.718]
T3: L + N*	0.020	-0.059	-0.024	-0.037	0.033	-0.149	-0.105	-0.112	0.015
	(0.104)	(0.082)	(0.115)	(0.095)	(0.102)	(0.112)	(0.098)	(0.084)	(0.081)
	[0.851]	[0.851]	[0.851]	[0.851]	[0.851]	[0.848]	[0.850]	[0.848]	[0.851]
Household size	-0.029	-0.016	-0.019	-0.007	0.000	-0.027	-0.032	0.005	0.005
	(0.020)	(0.018)	(0.020)	(0.019)	(0.019)	(0.023)	(0.020)	(0.017)	(0.018)
	[0.669]	[0.669]	[0.669]	[0.887]	[0.993]	[0.669]	[0.669]	[0.887]	[0.887]
Age of primary female	0.008	-0.002	0.008	0.001	0.009	0.002	0.008	0.006	0.003
	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)
	[0.485]	[0.850]	[0.485]	[0.938]	[0.485]	[0.850]	[0.485]	[0.485]	[0.850]
Age of primary male	-0.003	-0.005	-0.004	-0.005	-0.006	-0.002	0.001	-0.000	0.002
	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)
	[0.974]	[0.974]	[0.974]	[0.974]	[0.974]	[0.974]	[0.974]	[0.974]	[0.974]
Primary female is single	-0.496*	-0.239	-0.334	-0.060	-0.016	-0.055	-0.471**	-0.302	0.039
	(0.284)	(0.241)	(0.242)	(0.258)	(0.169)	(0.237)	(0.214)	(0.206)	(0.137)
	[0.373]	[0.581]	[0.385]	[0.919]	[0.923]	[0.919]	[0.262]	[0.385]	[0.919]
Primary female has some education	0.100	-0.069	0.051	-0.167*	0.116	0.060	-0.050	0.045	0.130
	(0.102)	(0.101)	(0.108)	(0.093)	(0.116)	(0.118)	(0.099)	(0.100)	(0.117)
	[0.656]	[0.656]	[0.656]	[0.656]	[0.656]	[0.656]	[0.656]	[0.656]	[0.656]
Standardized Livestock Units	-0.080**	-0.055	-0.120***	-0.108***	-0.061*	-0.121***	-0.069*	-0.083***	-0.049*
	(0.037)	(0.041)	(0.031)	(0.039)	(0.034)	(0.034)	(0.038)	(0.030)	(0.028)
	[0.057]	[0.184]	[0.002]	[0.016]	[0.089]	[0.002]	[0.089]	[0.016]	[0.089]

Household was affected by drought(s) in the last 15 months	0.190**	0.140	0.111	0.041	0.119	0.125	0.105	0.055	0.030
	(0.086)	(0.089)	(0.097)	(0.089)	(0.092)	(0.091)	(0.082)	(0.078)	(0.088)
	[0.253]	[0.363]	[0.385]	[0.725]	[0.363]	[0.363]	[0.363]	[0.617]	[0.731]
Household was affected by flood(s) in the last 15 months	0.078	0.156*	0.105	0.252***	0.038	0.227***	0.242***	0.184**	0.162**
	(0.090)	(0.084)	(0.095)	(0.089)	(0.089)	(0.080)	(0.071)	(0.073)	(0.073)
	[0.436]	[0.096]	[0.350]	[0.016]	[0.672]	[0.016]	[0.008]	[0.029]	[0.048]
Household was affected by death in the last 15 months	-0.500***	-0.109	-0.307**	-0.572***	0.139	-0.150	-0.298	-0.208	0.136
	(0.162)	(0.217)	(0.149)	(0.204)	(0.214)	(0.275)	(0.202)	(0.228)	(0.360)
	[0.023]	[0.693]	[0.124]	[0.026]	[0.693]	[0.693]	[0.319]	[0.652]	[0.707]
Household was affected by illness in the last 15 months	0.128	0.235**	0.125	0.022	0.036	0.107	0.127	0.013	-0.038
	(0.089)	(0.098)	(0.099)	(0.087)	(0.112)	(0.098)	(0.091)	(0.086)	(0.079)
	[0.466]	[0.155]	[0.466]	[0.883]	[0.883]	[0.499]	[0.466]	[0.883]	[0.883]
Household was affected by divorce in the last 15 months	0.257	-0.152	-0.178	-0.067	-0.195	-0.066	0.056	0.045	0.075
	(0.269)	(0.198)	(0.267)	(0.194)	(0.210)	(0.195)	(0.193)	(0.147)	(0.160)
	[0.771]	[0.771]	[0.771]	[0.771]	[0.771]	[0.771]	[0.771]	[0.771]	[0.771]
Mean of control (T4)	0.553	0.748	0.667	0.764	0.496	0.659	0.561	0.423	0.276
N	444	443	444	444	444	444	444	444	444

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. False Discovery Rate corrected q-values, reported in brackets, control for multiple hypothesis testing (Simes 1986).

Chapter 4: Nutrition

Including Scalable Nutrition Interventions in a Graduation Model Program: Experimental Evidence from Ethiopia

Harold Alderman[1], Daniel O. Gilligan[1], Jessica Leight[1], Tigist Mamo[2], Michael Mulford[2], Heleene Tambet[1]

1 International Food Policy Research Institute

2 World Vision

Abstract

We explore the impact of different models of scalable nutrition services embedded within a light-touch graduation program, implemented at scale in Ethiopia. The graduation program provided poor households enrolled in Ethiopia's national safety net, the Protective Safety Net Program (PSNP), with additional livelihood programming including savings groups, business skills training and linkages to financial services. In addition, extremely poor households received a one-time livelihood grant on an experimental basis, as cash transfers or in-kind poultry grants, at a value much smaller than lump sum transfers in other graduation model programs in recent literature. The experiment compared a core nutrition model of nutrition information and sanitation and hygiene activities to an enhanced model that added more intensive nutrition messaging, supplementary feeding of malnourished children, mental health services, and a male engagement activity . Results show that interaction with health care workers and participation in community health activities increased significantly under the enhanced nutrition model, as did maternal nutritional knowledge. Nevertheless, neither nutrition model led to significant improvements in child dietary diversity or anthropometric outcomes on average. However, cash livelihood grants combined with the enhanced nutrition model reduced childhood stunting.

List of Acronyms

ASTE: Average Standard Treatment Effects

BHA: Bureau for Humanitarian Assistance

CHF: Community Health Facilitators

CLTSH: Community-Led Total Sanitation and Hygiene

CN: Core Nutrition

CPNP: Community-Based Participatory Nutrition Promotion

cRCT: Cluster Randomized Control Trial

EN: Enhanced Nutrition

HDA: Health Development Army

HEWs: Health Extension Workers

IYCF: Infant and Young Child Feeding

ODF: Open Defecation Free

ORDA: Organization for Rehabilitation and Development in Amhara

LMICs: Low- and Middle-Income Countries

PSNP: Productive Safety Net Program

TTC: Timed and Targeted Counseling

TUP: Targeting the Extremely poor program

VESAs: Village Economic and Social Associations

WASH: Water, Sanitation, and Hygiene

1. Introduction

Developing countries spend, on average, 1.5 percent of GDP on social safety net programs and reach over 1 billion people with at least one program (World Bank, 2018). Such programs have multiple goals. They seek to both address current consumption and poverty and to facilitate investments that will reduce future poverty (Das, Do, Ozler 2005). Their success in the first goal is readily apparent; transfers reduce the incidence of absolute poverty (US$1.90 PPP per day) by 36 percent (World Bank, 2018). Impacts on a diverse set of productive investments are also apparent, albeit often less pronounced, partially reflecting their multiple objectives. These range from increased schooling (Baird et al. 2014), enhanced agricultural earnings (Gertler, Martinez, and Rubio-Codina 2012), and additional migration for employment (Ardington, Case, and Hosegood 2009). Another explicit objective for many transfer programs is improved nutrition. Favorable impacts on anthropometric outcomes are commonly noted although, on average, these are small. For example, Manley et al. (2022) found a significant, but modest 1.35 percent average reduction in child stunting from cash transfer programs in a recent meta-analysis of 129 studies.

Graduation model programs expand on the concept of targeted cash or food transfers supporting consumption, by providing an additional lump sum asset or cash transfer designed to promote and sustain poverty alleviation. These graduation programs often include intensive livelihood training or mentoring to support higher earnings as well as measures to promote financial inclusion (Bandiera et al. 2017; Banerjee et al. 2015; Bossuroy et al. 2022). Occasionally, they also include explicit measures to increase women's empowerment or reduce stress. Graduation model programs have been tested in many countries following the publication of results from a six-country study of programs styled after BRAC's Targeting the Extremely poor (TUP) program in Ethiopia, Ghana, Honduras, India, Pakistan, and Peru (Banerjee et al. 2015). Evaluations of these programs have demonstrated their potential to improve economic outcomes, but most of these programs were not designed to include nutrition interventions and generally do not report impacts on nutrition outcomes.

Graduation model programs may be strengthened by including nutrition objectives and explicit nutrition programming for several reasons. Graduation programs are frequently targeted to the extreme poor in low- and middle-income countries (LMICs). Malnutrition is often, but not always, co-located with this poverty. Thus, reducing malnutrition is consistent with the objective of graduation programs to achieve sustainable long-term improvements in wellbeing for participants and their communities and to address outcomes beyond monetary poverty. There is some evidence that transfers must be accompanied with other design features such as nutrition training and access to health services to be effective (Manley et al. 2022; Ahmed et al. 2019). Still, currently little is known about how to design graduation programs in terms of the size of the asset transfer and the approach to nutrition services to successfully achieve improvements in child nutrition.

The current study aims to address this lacuna. It reports the results of a cluster randomized control trial (cRCT) to assess the impact on child nutrition of a 'light-touch' graduation program that includes nutrition interventions of varying intensity, provided to beneficiaries of Ethiopia's Productive Safety Net Program (PSNP), one of the largest social protection programs on the

continent.[1] Implemented at scale, the graduation program provided PSNP beneficiaries with additional livelihood programming including savings groups, business skills training and linkages to financial services. In addition, extremely poor households received a one-time livelihood transfer on an experimental basis, either as cash transfers or in-kind poultry grants. These were at a value much smaller than those studied in Banerjee et al. (2015), which were up to 4 times larger (in 2017 PPP dollars) and had accompanying intensive livelihoods training and technical support.

Child stunting, and infant and young child feeding (IYCF) practices and knowledge were the primary and secondary nutrition outcomes[2] evaluated by this cRCT, which was adequately powered to assess a significant difference in these outcomes. Using this data this paper makes two main contributions to the literature on graduation. First, it documents the contribution of both cash and in-kind livelihood support on nutrition. Second, it indicates whether such grants can be enhanced with the provision of nutritional services. Results show that interaction with health care workers and participation in community health activities increased significantly under the enhanced nutrition model, as did maternal nutritional knowledge. Neither nutrition model led to significant improvements in child stunting on average. However, cash livelihood grants combined with the enhanced nutrition model reduced childhood stunting.

2. Study Context in Relation to Current Literature

Behavioral change communication (BCC) is a regular component of nutrition interventions. Keats et al. (2021) deem the evidence on the contribution of BCC for effective implementation of breastfeeding promotion and for education on complementary feeding in conjunction with food provision to be strong in food insecure settings but only moderate without food provision in more food secure populations. This implies potential program complementarity with transfer programs (Olney et al. 2022). A few programs that include regular grants during the critical years for child growth as well as packages of services along with nutrition focused BCC have reported significant impacts on child nutrition (Cahyadi et al. 2020; Carneiro et al. 2021). However, only a few studies have directly assessed the impact of BCC in conjunction with transfers compared to transfers alone (Little et al. 2021; Field and Maffioli 2021; Ahmed, Hoddinott, and Roy 2019). Although ultimately the goal of behavioral change is generally to improve nutritional outcomes, studies often focus on knowledge and behaviors around IYCF which appear more responsive to interventions than do anthropometric indicators. For example, Han, Kim, and Park (2021) look at BCC and food vouchers in Ethiopia and conclude that neither intervention alone influences child feeding practices or chronic child undernutrition, but the combination does.

This is the first study we are aware of to provide systematic evidence on the effectiveness of nutrition BCC delivered at scale as an integral part of a graduation program for extremely poor households on child growth. While the literature on graduation programs regularly examines

[1] The PSNP provides monthly consumption support for six months per year as payment for participation in labor-intensive public works (80 percent of beneficiaries) or for twelve months as direct support to labor scarce households (20 percent of beneficiaries).

[2] The pre-analysis plan for the evaluation is linked at the AEA RCT registry for this trial (registry number AEARCTR-0008281): https://www.socialscienceregistry.org/trials/8281.

impacts on food security, relatively few studies look at child nutritional status or diet diversity among transfer recipients. For example, none of the six country studies in the seminal paper by Banerjee et al. (2015) nor those by Bandiera et al. (2017), Blattman, Fiala, and Martinez (2020), or Bosseroy et al. (2022)[3] include evidence on child stunting or underweight[4]. Haushofer and Shapiro (2016) report results on an overall health index in which nutrition is a component but find no impact from either the monthly or the larger lump sum transfer on this composite measure. McIntosh and Zeitlin (2021) show the impact of lump sum transfers on nutritional outcomes and compare the impact to an in-kind nutritional and maternal health intervention in Rwanda with cash transfers and find significant effects on anthropometric outcomes in one sub-arm of the cash component in the study. However, they do not explore the complementarity of cash in health interventions.

As well as comparing cash transfers to in-kind services, McIntosh and Zeitlin (2021) include a range of transfer allocations to benchmark the size of transfers that might result in improved nutrition. They find that only the largest transfer, which was more than twice as large as the one in the light-touch graduation program studied here, had an impact of 0.1 standard deviations on anthropometry, significant at the 10 percent level.[5] The current study builds on those findings by investigating whether a considerably smaller transfer than provided in several published graduation programs can enhance child nutritional status by exploiting complementarities in scalable nutrition-focused services.

In addition to exploring complementarity within a graduation program, the study's experimental design compares the impact of two models of nutrition investment, an enhanced model of more intensive BCC along with components designed to strengthen male support to investments in child nutrition and support to women's wellbeing, including mental health as well as a less intensive community nutrition program that did not directly support household visits. The goal of the intensive mode of delivery is in keeping with Kim et al. (2020) who report results of trials in Bangladesh, Ethiopia, and Vietnam designed to directly compare intensive interpersonal contact relative to alternate standard nutrition BCC programs. All three of these studies confirmed that the frequency of contact matters for diet diversity and breastfeeding practices. Kim et al. (2019) also modeled dose response for improved height for age with a combination of platforms in a companion study in Ethiopia. The current study – investigating a program that reached over half a million beneficiaries - adds to the literature on nutrition program intensity within the context of an integrated graduation model program.

3. Evaluation Design

While the PSNP was initiated in 2005, it did not include a specific nutrition component until the fourth phase (PSNP4) launched in 2015 (Berhane et al. 2020). To date, neither the earlier stages nor the fourth phase of PSNP have shown any acceleration of the national trend in reducing stunting

[3] The working paper version, Bosseroy et al. (2021), however, did report on anthropometrics. That paper found no significant effects on height or weight for age of children.
[4] Raza, Van de Poel, and Van Ourti (2018) report reduced wasting after 4 years in the communities Bandiera et al. (2017) studied. However, wasting is a short-term measure generally concentrated among children under 2.
[5] Plausibly, the 13-month study might be a lower bound since some anthropometric impacts are cumulative (Alderman and Headey 2018).

(Berhane et al. 2020; Abay et al. 2023). USAID's Bureau of Humanitarian Assistance (BHA) supported the PSNP4 in selected woredas (districts) through the Strengthen PSNP4 Institutions and Resilience (SPIR) graduation program. In addition to supporting implementation of PSNP4, SPIR also provided additional complementary livelihood, nutrition, gender, and natural resource management activities led by World Vision in partnership with the Organization for Rehabilitation and Development (ORDA) and CARE. SPIR targeted more than 500,000 PNSP clients in 15 of the most vulnerable woredas in Amhara and Oromia regions of Ethiopia and provided community-level programming and capacity building of government staff involved in public service delivery at the woreda (district) and kebele (subdistrict) level.

The trial included three treatment arms and a control group. One treatment arm provided a package of enhanced nutrition (EN) services including support for household-level BCC, recuperative feeding for acutely malnourished children, promotion of male engagement in household tasks, and group therapy for mothers screened for depressive symptoms. A subset of the poorest beneficiaries within this arm also received an additional one-time grant either in cash or as poultry assets as a part of the enhanced livelihood program component. A second treatment arm provided the enhanced livelihood program treatment with a less intensive and less costly core nutrition service (CN). A third treatment arm provided the same enhanced package of nutrition services as the first treatment arm but without the enhanced livelihood program including grants. All three treatment arms also provided a set of core services including promotion of savings and access to credit through Village Economic and Social Associations (VESAs).

Across all treatment arms, as well as the control group, PSNP clients participating in public works were eligible to attend six monthly 2-hour BCC sessions in place of 4 person-days of work over the six-month period. Also in all treatment arms, a water, sanitation, and hygiene (WASH) component included implementation of a community-led total sanitation and hygiene (CLTSH) activity and Open Defecation Free (ODF) public education events, as well as the provision of support to village-level WASH management activities and limited support to improving potable water and sanitation infrastructure (water sources and latrines). In both the CN and EN treatments these services were augmented by IYCF and nutrition BCC discussions in VESAs. Topics covered included optimal IYCF practices, adolescent and maternal nutrition, diversified sources of nutritious foods, and utilization of health and nutrition services.

The EN interventions built upon these core nutrition activities by supporting government health extension staff and volunteers to additionally deliver BCC lessons on IYCF practices and adolescent and maternal nutrition through a Timed and Targeted Counseling (TTC) model at the household level. The TTC model envisions 11 home visits in the first 1,000 days after conception, 4 of which should be during pregnancy, 2 in the first month post-partum, 4 more when the child is between 1 and 12 months and another visit when the child is between 14 and 18 months. Additional services in the EN treatment kebeles were targeted to caregivers of children screened for acute malnutrition based on mid-upper arm circumference or underweight based on weight for age. These Community-Based Participatory Nutrition Promotion sessions (CPNP) included training on complementary feeding and caring practices within two-week intensive feeding sessions.

These activities were coordinated by trained Community Health Facilitators (CHF) who were recruited and assigned to each of the kebeles to support the health extension workers (HEWs) in planning, coordinating, and facilitating health and nutrition activities. The CHF also supported local

Health Development Army (HDA) volunteers to conduct household-level counseling related to IYCF and maternal nutrition using the TTC approach. Both husband and wife were encouraged to be present for these counseling sessions to promote male engagement in targeted IYCF and maternal nutrition actions. In addition, local facilitators supported the formation of men's engagement groups and facilitated eight sessions designed to critically reflect on cultural gender norms and explore the positive and perceived negative effects of male involvement in household chores and childcare activities that were typically assigned as women's only tasks.[6] Figure 1 summarizes the key programs by treatment arm. These services were not provided in the control.

Comparatively poor households among the already food insecure PSNP client households in the two enhanced livelihood program treatment arms received a livelihood transfer ($200 equivalent in local currency). The kebeles in these arms of the study were randomly selected to provide either a one-time cash transfer to the subset of relatively poor households or a poultry package of equivalent value. In each of these kebeles the 10 poorest households of the baseline survey sample of 18 households were chosen to receive these transfers through a ranking assessment based on an asset index. The asset index included ownership data on more than 30 asset categories, including consumer durables, productive assets, livestock, and land. It was constructed using principal components analysis, in which the first component maximizes the variation of all variables explained by that component, and therefore gives high weights to variables that are highly correlated with each other (Filmer and Pritchett 2001). These households are the main focus of this study and are deemed extremely poor in subsequent discussion.

SPIR's theory of change posits that improving household income through livelihood interventions is necessary for households to access nutritious foods and complementary to nutrition BCC interventions. While the subsistence farming households in this program produce staple crops for their own consumption, they often depend on local markets for a majority of non-staple foods (Sibhatu and Qaim 2017).[7] Relative to the price per calorie of the cheapest starchy staple, one study in Ethiopia found that leafy green vegetables were 11-18 percent more costly per calorie, and eggs nearly 30 percent higher in cost per calorie (Headey et al. 2019). Beginning with regular saving and access to small loans through VESAs, PSNP households gain experience initiating and managing small enterprises. Aligning with the livelihood component of the PSNP4 design, livelihood transfers targeted extremely poor households who would otherwise be unable to engage in additional livelihood activities through access to formal credit due to lack of collateral or productive assets. Associated technical training accompanied these transfers with the aim of supporting household's engagement in profitable economic activities. The expected increase in household income is deemed as a critical complement to improved knowledge on provision of diversified and nutritious diets, enabling extremely poor PSNP households to access nutrient rich foods in a sustainable manner.

[6] Furthermore, women were screened for depression in the midline survey and those with symptoms of depression – roughly 20 percent - were invited to participate in group therapy sessions.

[7] Given the small target number of households relative to the overall population in a kebele (over 1,000 households on average), and the small size of the transfer, we do not expect or observe any local market price effects (such as lowering the price of eggs in a kebele where extremely poor households received the poultry package transfer).

4. Data

The data were collected over three rounds, with a baseline collected between February and April 2018, a midline survey between July and October 2019, and an endline survey originally planned for 2020, 24 months after the baseline, but delayed an additional 12 months due to COVID-19. Initially, 196 kebeles in Amhara (115 kebeles) and Oromia (81 kebeles) regions were selected for the trial and subsequently randomized into treatment groups. However, two kebeles had no PSNP clients and thus were not eligible for the program. Two other kebeles experienced ongoing civil unrest and were necessarily dropped from the project. Thus, the evaluation sample comprises 192 kebeles. In each kebele, 18 households were randomly sampled, leading to a planned baseline sample of 3,456 households in the study. The inclusion criteria for the sample were that households had to (1) be a PSNP client household, (2) have at least one child aged 0-35 months (the *index* child), and (3) have the mother or primary female caregiver[8] of the 0-35-month-old child as a member of the household. The last criterion enabled measurement of maternal and child diets as well as child-care for nutritionally vulnerable ages in all sample households. The actual baseline sample consisted of 3,314 households.

The endline survey sought to re-interview any baseline household unless it had permanently moved or dissolved by the time of the midline survey. Of the 3,248 households in the endline target sample, 3,094 were located and interviewed, implying an attrition rate of 4.7 percent relative to the target sample, or 6.6 percent relative to the original baseline sample. Nearly half of the attrition (70 households or 2 percent of the target sample) at endline was due to insecurity in one of the operational woredas in Amhara that resulted in a decision not to visit four kebeles. Among the remaining 84 attrited households, 48 had moved out of the study area, 24 were temporarily unavailable, and other households had either dissolved or were unavailable for the interview for other reasons; one household refused consent. In total, 3,062 primary female respondents and 2,482 primary male respondents were surveyed. The delay in the endline data collection necessitated by COVID-19 meant that most of the additional midline sample children were older than 24 months by the endline contrary to the initial intent. However, 905 of the baseline index children had younger siblings less than 24 months who were included in the endline sample and contributed to outcome measures for children less than 2 years old at endline. Even so, the sample for the analysis of current feeding practices and contact with healthcare workers is smaller than at the baseline[9].

The endline household interview was conducted in three parts: household-level questions covering household and respondent identification and household demographics; a set of questions for the

[8]As 99.1 percent and 97.9 percent of children under 24 months and under 36 months respectively were cared for by their mother, the word mother will used instead of caregiver in the rest of this study, including tables. Similarly, father is used in lieu of primary male caregiver irrespective of biological roles.

[9] A supplemental sample of households was added at midline to refresh the age-appropriate sample for aspects of analysis of nutrition such as dietary diversity of children in keeping with IYCF norms. This sample was drawn from the original beneficiary lists that were used for the baseline household sample with the same eligibility criteria as in the baseline, with the exception being that children must be under 2 years old at the time of the survey. The midline survey aimed to add 4 supplemental households in each kebele (another 768 households). As the supplemental sample did not receive any livelihood grants that are the main focus of this paper, that sample is not included in the main body of this study. Results including the supplemental households, however, are presented in Annex Table 3.

identified primary male respondent; and a set of questions for the identified primary female respondent. The primary female was the mother of the index child in most cases and the primary male was her partner. A separate, specifically trained team conducted anthropometry measurements.

5. Analytical Approach

The first step in the analysis was to verify that outcomes and contextual variables were balanced across treatment arms at baseline. Next, we report on the primary outcome of interest, stunting, defined, as is generally done, as a child's height for age being two or more standard deviations below the age and gender specific median. We also include other anthropometric outcomes in this step. The regressions use the following model:

$$Y_{i1} = \beta_0 + \beta_1 EN + \beta_2 CN + \beta_3 X_{i0} + \sigma_w + \varepsilon_i, \quad (1)$$

where Y_{i1} is the outcome variable at endline, EN indicates randomized assignment to the enhanced nutrition intervention, and CN indicates randomized assignment to the core nutrition intervention delivered at the community level. X_{i0} is a vector of controls, including child age, child gender, maternal education, and woreda level fixed effects. The anthropometric outcomes are reported for the entire sample. However, in as much as the focus of this analysis is the contribution of livelihood grants to nutrition, we repeat the overall assessment of the EN and CN programs using only the subsample of extremely poor. As any outcome, or absence of impact, reflects the additional services supported through SPIR, we then investigated the impact of the EN and CN models on access to the services provided and on any changes in diets. These secondary results reflect the major inputs into the production of health that the project aspired to influence. We focus on the extremely poor[10], testing the effects of the interventions on recipients of cash and of poultry relative to each other. In the communities that received enhanced nutrition we also compare the impacts of the EN intervention for grant recipients with the results for other extremely poor households. As there are no extremely poor households who did not receive grants in the CN arm of the study, this comparison is not available.

Child age is handled in slightly different ways depending on features of the outcome studied. In the case of anthropometry, age is included to accommodate differences in growth velocity and biological needs of children as well as the cumulative impact of programs (Alderman and Headey 2018). Thus, the anthropometry regression which cover children up to 36 months includes a variable for age in months. In contrast, the diet data covers a smaller sample of children up to 24 months since the module was not administered neither to parents of children above 24 months nor to parents of children less than 6 months for whom exclusive breastfeeding is strongly recommended. In addition, since the timing of TTC implies that contact with service providers should be greatest when a child is under a year old, although some contact is expected after the

[10] The overall impacts of the EN and CN programs on the respective communities are of interest to the SPIR program planners and have been reported in the endline report (Alderman et al. 2021). We also report summary results for the less poor and the extremely poor at the end of this paper.

first birthday as well, we test this by interacting a dummy variable for older children in the regressions that assess access to health services.

All regressions include woreda level fixed effects (σ_w) and adjust standard errors by clustering at the kebele level. The duration of the study ruled out child fixed effects or ANCOVA approaches for anthropometric outcomes since children who were in the baseline were older than 24 months by the endline[11] All models are intent to treat models estimated by OLS, using the randomized assignment for the identification of causal impact. The study design included twice as many kebeles in the EN treatment as in the CN treatment or the control group to allow for interaction of the EN treatment with the enhanced livelihood treatment.

While we present individual outcomes, where appropriate we also report average standard treatment effects (ASTE) estimated for key families of outcomes, following Kling, Liebman, and Katz (2007). These are aggregated summaries of key outcomes of interest in a couple of tables, equivalent to outcomes presented in tables where all variables could be added up into a single measure. The ASTEs serve as summary indices that aggregate information over multiple outcomes and present effect sizes relative to the standard deviation of the control arm. This aggregation improves statistical power to detect effects that go in the same direction within a domain. As this calculation requires that the signs of the coefficients all have the same welfare interpretation it is occasionally necessary to redefine the outcome for the purpose of computing an ASTE. We report ASTE for both nutrition programs as well as the two transfers, poultry, or cash. Where the dependent variable is a count item – either knowledge scores or the number of food categories consumed - the total score serves in lieu of the ASTE.

6. Results

Before presenting the results from the midline and endline surveys, Table 1 shows that the sample is well balanced across treatment arms in baseline characteristics. There are a few differences across treatment arms significant at p <0.10, but these do not appear to be systematic in any direction. Moreover, there are no significant differences in the key anthropometry measures. These summary statistics provide an overview of the context at baseline. Households in the sample are poor; the low average of expenditures is consistent with the mean prevalence of poverty in the sample (below $1.90 per capita per day) of 45 percent. The kebeles are fairly remote at 11-14 km from the nearest town. Households are quite large, averaging 5.7 members. Education levels are low, with only 1 in 5 women and 1 in 3 men having any education. Women's diets are limited, with only 3.4 percent of women meeting the standard for minimum dietary diversity (FAO and USAID 2015). Child diets are also severely limited, with index children aged 6-35 months consuming fewer than 2 food groups per day. Only 1.0 percent of these index children consumed a minimally acceptable diet (WHO and UNICEF 2017). Child malnutrition is a substantial public health problem in these communities: across the different treatment arms 36.2-39.7 percent of index children were stunted and 11.6-16.6 percent were wasted.

[11] As indicated, all the baseline households that were available at the endline were included in the final survey regardless of the ages of their children. In principle, it would be possible to have a child level panel of anthropometric measures using data collected for children up to 60 months. Similarly, we could not use a panel on child health care since this was collected only for children < 24 months and thus did not cover the initial child sample by the endline.

As indicated in Table 2a there is no indication that the SPIR program had a measurable impact on anthropometric status on average. Indeed, all the point estimates for the coefficients of EN in the table are close to zero. Although the sample of children 6-35 months is smaller than anticipated in the original design due to the delay in the endline survey caused by COVID precautions, alternative estimates with a larger sample of children 6-59 months, consistent in size with the original power calculations, also indicate no significant impacts on anthropometry overall.

However, Table 2b, which focuses on a sample restricted to the extremely poor households in each kebele, provides experimental evidence on the impact of program subcomponents. While there was no observed improvement in anthropometry relative to the control among the extremely poor households in the kebeles that received enhanced nutrition but did not receive the livelihood grants, those that received cash transfers and the EN program had significantly higher height for age and reduced stunting. This combination of interventions reduced stunting by 18 percentage points, compared to the 54 percent of the extremely poor in the control group that was stunted. In contrast, cash transfers provided to households in the CN nutrition program did not lead to improvement in these outcomes.[12] We focus on this sample of extremely poor households in the results that follow to inform the mechanisms behind this finding. However, we return to a comparison of ASTE results for the extremely poor and less poor households which had no access to cash or poultry livelihoods support at the conclusion of the presentation of results on livelihood transfers.

Does the limited impact reflect a deviation from the planned implementation of the program components in some treatment arms? The ASTEs reported in Table 3 indicate increased access to nutrition services for the extremely poor households. This improved access is across all the components of the ASTE for the kebeles receiving the EN nutrition program while the ASTE results for the households in CN arm are largely driven by increased participation in nutrition discussions within the VESAs. Unlike the results for stunting, the increase in service access observed in the EN program does not differ between the households receiving livelihood grants and those in the EN arms that were not provided grants. That is, there is no evidence that cash transfers drove increased health service utilization.

Table 3 also indicates increased probabilities of attending community food demonstrations or attending a community BCC session as well as participating in water, sanitation, and hygiene meetings. The WASH activities indicated in columns 6 and 7 were offered in all treatment arms as well as in the control. The EN program, nevertheless, facilitated increases in participation in these activities, while the CN did not. Grants apparently had no complementary impact on participation. Column 8 indicates a substantial increase in the probability of discussing health and nutrition messages for both the EN and CN programs through the VESAs that were facilitated in all treatment arms of the intervention. However, unlike WASH activities, these were not organized in the control arm of the study.

As shown by the ASTE in Table 4, the EN program did not contribute to an increase in a suite of child specific health and nutrition activities. For example, there was no increase in the probability of a child being weighed or having their mid-upper arm circumference (MUAC) measured for all children

[12] All extremely poor households in the CN arm of the study received either poultry or cash grants. Thus, we cannot assess the impact of CN on extremely poor household in the absence of livelihood support.

less than 24 months. The age for which this information was obtained was more restricted than for family access to services in Table 3. Children less than 6 months either were not expected to receive the services, as with vitamin A, or the 6-month recall did not apply fully. Thus, the table covers children 6-23 months. There is suggestive evidence of an increase in participation in recuperative child feeding involving cooking demonstrations in the EN arm mediated by cash transfers. The two-week cooking demonstrations and child feeding were provided to mothers whose children screened as underweight or at risk of malnutrition (based on MUAC) through weight screenings. Thus, they reflect the probability of being weighed – which the program hopes to increase – as well as the probability of being undernourished – which the program desired to decrease – as well as the take up of the CPNP offer. Participation in these recuperative child feeding sessions required mothers to supply the ingredients for the meals being prepared each day, suggesting a plausible pathway for an interaction effect with cash transfers.

Table 5 indicates that there was an insignificant increase in maternal knowledge in the EN treatment areas. The average total increase in the knowledge score in EN kebeles offering grants of 0.154 reported in column 1 measures the sum of the 7 questions on nutritional knowledge. This effect is less than a 4 percent of the mean score for the control population despite the EN arms receiving grants and having increased contact with HEW and HDA workers as well as participation in facilitated discussions on nutrition in the VESA meetings. This compares to the similarly modest effect of 0.226 (6.6 percent) on having attended any school.[13]

The overall EN program also contributed to an increase in the maternal diet diversity score (column 1 of Table 6), which is based on the sum of results in columns 3-12, as well as in the share or women meeting the guideline for minimum diet diversity. The increase in MDD among women receiving poultry of 0.044 is small, yet it is a 77 percent increase over the value in the control communities, reflecting the very poor quality of diets among mothers in the study. However, as in the results on the nutrition knowledge score, there is no significant difference if the EN community was randomized to receive grants or only the enhanced nutrition intervention. Nor was there an increase in the CN communities.

Nevertheless, a similar increase in the diversity of diets for children 6-23 months was not observed on average in the EN arms although this was a key focus of the IYCF training (Table 7). There was also no overall change in the share of children consuming at least five of eight food groups daily, including breast milk, which is the recommended minimum diet diversity (MDD) for a child of 6–24 months (WHO and UNICEF, 2017). Virtually no child in the sample met this guideline. In addition to MDD, the guideline for a minimal acceptable diet includes minimum meal frequency, defined as proportion of children aged 6–23 months who receive solid, semi-solid, or soft foods at least two (three) times for children aged 6–8 (9–23) months. While over 44 percent of the children in the control group were fed in accord with this guideline, the SPIR nutrition programs by themselves did not lead to any increase in the share who were provided meals at a frequency that is keeping with the recommendation.

However, children in households within the EN arms of the study that were randomized to receive poultry grants did see an increase in the diet diversity score. This was driven, in part, by an increase

[13] The interaction of EN and education was neither significantly positive nor negative. Thus, there was no evidence that the program complements education nor substitutes for it.

in the frequency of egg consumption but also milk intake. This was not the case in other households within EN program arms. Despite the higher probability of both egg and dairy consumption among children in kebeles randomized for poultry grants, there is no significant difference between these kebeles and those selected for cash in the nutritional outcomes report in Table 2b. The diet diversity scores, however, do not report quantities consumed nor is the sample designed to assess heterogeneity by child age which maps with the duration of program coverage.

Previous work has shown that such in-kind assistance in combination with BCC has a different impact on egg consumption than equivalent cash in these villages (Alderman et al. 2022). This is consistent with market imperfections that make consumption decisions not separable from production decisions, a special case of agricultural household models, albeit not a rare one. As the short run response to an in-kind grant may differ from longer term behavior, the key results on the dietary impact in the endline are compared to the previous midline results using the published specification. These are reported in Annex 1.

The diet diversity tables for both mothers and for children include a dummy variable for religious fasting days in Orthodox households. These indicate a significant reduction in the probability of consuming dairy, meat, and eggs with little corresponding increase in other food groups among adult women. Fasting among Orthodox households did not affect dairy consumption by children in keeping with previous evidence (D'Haene et al., 2020). This likely reflects the fact that cows provide milk according to their own biology and the milk is not easily stored. In contrast, the timing of animal purchases or slaughtering is at the household's discretion.

In Table 8, we provide a summary of ASTEs or aggregate impact for each family of outcomes by receipt of EN and CN for the extremely poor and less poor samples. These do not focus on the nature of livelihood grants but are relevant for understanding the overall impact of SPIR on nutrition. Results show that the only significant difference in program response between the extremely poor and the less poor is in the results on stunting. While this may be partially due to the significant complementarity of cash transfers and the EN intervention, it also reflects the unexpected – but statistically insignificant - increase in stunting among the less poor. The marginally significant differences between overall EN and CN ASTE results among the extremely poor in columns 5 and 6 were not observed in tables 6 and 7. This is because the tests in Table 8 include extremely poor households who were not randomized for grants. This adds a subsample that was no less responsive than the EN response among grant recipients and in so doing increases the sample size. In each case where the EN results differed from CN results for the extremely poor, including the substantial ASTE results on access to health care, they also differed for the less poor.

7. Concluding Discussion

The livelihood cash grants were designed to encourage asset formation and had no intrinsic features specifically tied to access to or utilization of enhanced health and nutrition services. There was, however, complementarity of cash and enhanced nutrition for the anthropometric outcomes, even though this was not seen in diet diversity. This leaves a puzzle as to the actual mechanism of the complementarity of cash and nutritional services. Plausibly the recipients of cash were able to purchase more food. The diet diversity index does not measure changes in quantities only the probability that a food group was included in the daily diet. Also, cash recipients were more likely to participate in the CPNP recuperative child feeding sessions, perhaps drawing upon the cash to

provide the expected food inputs in this two-week program. Thus, the overall results provide a basis for tempered nutritional outcome expectations for this additional support to extremely poor households, valued at well over one year of transfers in the PSNP, when combined with enhanced nutrition programming.

The grant of poultry did enhance the EN program in regard to the inclusion of eggs in a child's diet, but this was not sufficient to have a significant impact on height for age. The challenge to the advocacy of household poultry production as a means to foster better child nutrition may hinge on the ability to sustain flocks. The susceptibility of the birds to illness was noted by Passarelli et al. (2022) in a similar program in Ethiopia. However, the final SPIR survey did not report on the number of birds that died. The survey did ascertain the number of birds sold in the previous 12 months; poultry grant recipients sold, on average, two more birds than their neighbors. Due to the delay in data collection necessitated by COVID the time between surveys was greater than the 12 months recall so there is a possibility that sales in this period partially accounts for the convergence of flock size between in-kind grant recipients and their counterparts. Nevertheless, the potential to enhance consumption of animal source foods with these transfers did not prove to be sustainable (Annex 1).

The clearest evidence of the impact of the EN treatment distinct from the complementarity with grants was shown in the increase of contact between mothers of young children and service providers despite the challenges of COVID as well as the improvement in women's IYCF knowledge. Within the EN kebeles, the overall probability of any mother meeting with a HEW or member of the HDA in the previous 3 months increased from the pre-COVID midline (Figure 2). Still, even though the contact in EN communities was 28 percent above the 43 percent rate of contact for mothers in the control it was not sufficient to improve anthropometric indicators of child nutrition status. This likely reflects the fact that the frequency of contact with health service providers (either HEWs or HDA volunteers) still fell well short of intent. Although the TTC model envisions 11 home visits in the first 1,000 days after conception, the endline survey revealed that approximately only a quarter of mothers of children under 12 months in the EN communities reported a home visit in the last 3 months and this probability declined for older children.

This contrasts with an average of 2 visits in the last 3 months in Kim et al. (2020) and likely accounts for the larger impact measured in that study. However, there is a positive association of HAZ and program participation for children whose household reported having received a home visit recently. This hints at the possibility that more frequent service provision might have more favorable impacts, but it does not give any indication as to how this intensification may be achieved. As the current study cannot assess the determinants of home visits by HEW or HDA volunteers, this can be considered a limitation of the study and an area for future research.

Restricting the sample in Table 2a to the 598 young children in the households that reported they had contact with a HEW or HDA in the preceding three months – a selected subsample that is potentially biased towards larger impacts – finds an increase in HAZ of 0.286 (SE=0.163) and a reduction of stunting of -0.082 (SE=0.045). Both these increases are significant at 10 percent level, while lower statistical power than the results in Table 2a is expected as the regressions are filtered to have a smaller sample. A further restriction of the sample to the 253 households where the HEW or a HDA visited at the household indicated much larger point estimates of improvements in HAZ and reduction in stunting, 0.616 (SE=0.275) and -0.165 (SE=0.072) respectively, with both of these

estimates significant at 5 percent level. While the unknown nature of the selectivity into this subsample, possibly due to better trained staff and supervisors or more motivated mothers or both, makes these results only suggestive at best, they may provide guidance for future inclusion of nutrition within the PSNP.

Drilling down further, results indicate that contact with HEW or HDA workers in the EN program was concentrated among households with children less than a year old. The increase for all extremely poor households with these young infants is 25 percentage points higher than the 47 percent share of contact for corresponding households with children in that age group in the control group (results not shown). While the net increase in contact with a HEW or HDA when the youngest child was less than twelve months was significant at $p < 0.01$, the overall increase for the extremely poor households within the EN communities for children 12-23 months is only 0.030 [SE=0.0820.189-0.116] and not significant.

While the less intensive CN intervention led to increased participation in community events relative to the control, it did not lead to increased contact with government health staff or volunteers, nor increased women's IYCF knowledge or dietary diversity. Even though women in the CN arm reported the inclusion of nutrition topics in VESAs similar to those in EN, all the increases in contact with health service providers in the EN program were larger than those in CN at $P < 0.1$. This difference likely explains the absence of an impact on mother's knowledge score or her diet diversity in the CN program, outcomes that were influenced by the EN arm of the study.

Despite the increased nutritional knowledge amongst women as well as changes in their diets, there was not a clear impact on dietary diversity for young children in the EN program. This remains extremely low. Hirvonen et al. (2021) document late introduction of solid and semi-solid foods in Ethiopia which is linked to lower length for age of young children. In the current study, 26 percent of the children 6-8 months consumed no other food than breastmilk in the previous day indicated in the list of 25 foods in the appropriate IYCF module for diet diversity.

Thus, in conclusion, the success in enhancing coverage of nutrition services to low-income mothers in remote areas can be considered a partial achievement of the SPIR program. It remains, however, a challenge to maximize the nutritional impact of this service delivery. Livelihood grants make a modest contribution to the challenge, but in the longer run, it is a larger task to build upon this contribution.

References

Abay, Kibrom, Guush Berhane, John Hoddinott, and Kibrom Tafere. COVID-19 and food security in Ethiopia: do social protection programs protect?. 2023. *Economic Development and Cultural Change* 71(2): 373-402.

Ahmed, Akhter; Hoddinott, John F.; and Roy, Shalini. 2019. Food transfers, cash transfers, behavior change communication and child nutrition: Evidence from Bangladesh. IFPRI Discussion Paper 1868. Washington, DC: International Food Policy Research Institute.

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

Alderman, Harold, Lucy Billings, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Alemayehu Seyoum Taffesse, and Heleene Tambet. 2021. "Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA): Endline Report."

Alderman, Harold and Derek Headey. The timing of growth faltering has important implications for observational analyses of the underlying determinants of nutrition outcomes. 2018. *PLOS ONE*. 13(4): e0195904 Apr 25.

Ardington, Cally, Anne Case, and Victoria Hosegood. 2009. Labor supply responses to large social transfers: Longitudinal evidence from South Africa. *American economic journal: Applied economics* 1 (1): 22-48.

Baird, Sarah, Francisco HG Ferreira, Berk Özler, and Michael Woolcock. 2014. Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness* 6 (1): 1-43.

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

Banerjee, Abhijit, Duflo E, Goldberg N, Karlan D, Osei R, Pariente W, Shapiro J, Thuysbaert B, and C. Udry. 2015. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science* 348(6236): 1260799-1-15.

Berhane, Guush; Golan, Jenna; Hirvonen, Kalle; Hoddinott, John; Kim, Sunny; Taffesse, Alemayehu Seyoum; et al. 2020. Evaluation of the nutrition-sensitive features of the fourth phase of Ethiopia's Productive Safety Net Programme. ESSP Working Paper 140. Washington, DC: International Food Policy Research Institute.

Blattman, C., Fiala, N. and Martinez, S., 2020. The Long-Term Impacts of Grants on Poverty: Nine-Year Evidence from Uganda's Youth Opportunities Program. *American Economic Review: Insights*, 2(3), pp.287-304.

Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand et al. 2022. Tackling psychosocial and capital constraints to alleviate poverty. *Nature*: 1-7.

Bossuroy, Thomas, Markus Goldstein, Dean Karlan, Harounan Kazianga, William Pariente, Patrick Premand, Catherine Thomas, Christopher Udry, Julia Vaillant, and Kelsey Wright. 2021. Pathways out of Extreme Poverty: Tackling Psychosocial and Capital Constraints with a Multi-faceted Social Protection Program in Niger. Policy Research Working Paper 9562. The World Bank.

Cahyadi, Nur, Rema Hanna, Benjamin A. Olken, Rizal Adi Prima, Elan Satriawan, and Ekki Syamsulhakim. 2020. Cumulative impacts of conditional cash transfer programs: Experimental evidence from Indonesia. *American Economic Journal: Economic Policy* 12 (4): 88-110.

Carneiro, Pedro Manuel, Lucy Kraftman, Giacomo Mason, Lucie Moore, Imran Rasul, and Molly Scott. 2021. The impacts of a multifaceted pre-natal intervention on human capital accumulation in early life. *American Economic Review*. 111(8): 2506-2549.

Das, Jishnu, Quy-Toan Do and Berk Özler. 2005. Reassessing Conditional Cash Transfer Programs, *World Bank Research Observer*, 20(1): 57-80.

D'Haene, E., Vandevelde, S. and B. Minten. 2020. Fasting, food, and farming: Evidence from Ethiopian producers on the link of food taboos with dairy development. Strategy Support Program Working Paper #141. International Food Policy Research Institute, 2020.

FAO and USAID. 2015. Minimum Diet Diversity. A Guideline for Measurement. Rome.

Field, Erica M., and Elisa M. Maffioli. 2021. Are behavioral change interventions needed to make cash transfer programs work for children? Experimental evidence from Myanmar. No. w28443. National Bureau of Economic Research.

Filmer, Deon and Lant Pritchett. 2001. Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography* 38(1): 115-132.

Gertler, Paul J., Sebastian W. Martinez, and Marta Rubio-Codina. 2012. Investing cash transfers to raise long-term living standards. *American Economic Journal: Applied Economics* 4 (1): 164-92.

Han, Yaeeun, Hyuncheol Bryant Kim, and Seollee Park. 2021. The Roles of Nutrition Education and Food Vouchers in Improving Child Nutrition: Evidence from a Field Experiment in Ethiopia. *Journal of Health Economics* 80: 102545.

Haushofer, Johannes, and Jeremy Shapiro. 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, Derek, Kalle Hirvonen, John Hoddinott, and David Stifel. 2019. "Rural Food Markets and Child Nutrition." *American Journal of Agricultural Economics* 101 (5): 1311–27.

Hirvonen, Kalle, Abdulazize Wolle, Arnaud Laillou, Vincenzo Vinci, Stanley Chitekwe, and Kaleab Baye. 2021. Understanding delays in the introduction of complementary foods in rural Ethiopia. *Maternal & Child Nutrition*: e13247.

Iannotti, L.L., Chapnick, M., Nicholas, J., Gallegos-Riofrio, C.A., Moreno, P., Douglas, K., Habif, D., Cui, Y., Stewart, C., Lutter, C.K., Waters, W.F., 2020. Egg intervention effect on linear growth no longer present after two years. *Maternal & Child Nutrition* 16 (2). https://doi.org/10.1111/mcn.v16.210.1111/mcn.12925

Keats, Emily C., Jai K. Das, Rehana A. Salam, Zohra S. Lassi, Aamer Imdad, Robert E. Black, and Zulfiqar A. Bhutta. 2021. Effective interventions to address maternal and child malnutrition: an update of the evidence. *The Lancet Child & Adolescent Health* 5.

Kim, Sunny S., Phuong Hong Nguyen, Yisehac Yohannes, Yewelsew Abebe, Manisha Tharaney, Elizabeth Drummond, Edward A. Frongillo, Marie T. Ruel, and Purnima Menon. 2019. Behavior change interventions delivered through interpersonal communication, agricultural activities, community mobilization, and mass media increase complementary feeding practices and reduce child stunting in Ethiopia. *The Journal of nutrition* 149 (8): 1470-1481.

Kim, Sunny S., Phuong Hong Nguyen, Lan Mai Tran, Silvia Alayon, Purnima Menon, and Edward A. Frongillo. 2020. Different combinations of behavior change interventions and frequencies of interpersonal contacts are associated with infant and young child feeding practices in Bangladesh, Ethiopia, and Vietnam. *Current developments in nutrition* 4 (2): nzz140.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. Experimental analysis of neighborhood effects. *Econometrica* 75 (1): 83-119.

Leight, Jessica, Harold Alderman, Daniel Gilligan, Melissa Hidrobo, and Michael Mulford. 2023. The Effects of a Light-Touch Graduation Model on Livelihoods Outcomes: Evidence from Ethiopia. IFPRI.

Little, Madison T., Keetie Roelen, Brittany CL Lange, Janina I. Steinert, Alexa R. Yakubovich, Lucie Cluver, and David K. Humphreys. 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.

Manley, James, Harold Alderman, Ugo Gentilini. 2022. More Evidence on Cash Transfers and Child Nutritional Outcomes: A Systematic Review and Meta-Analysis. *BMJ Global Health*. 7 (4): e008233.

McIntosh, Craig, and Andrew Zeitlin. 2021.Cash versus Kind: Benchmarking a Child Nutrition Program against Unconditional Cash Transfers in Rwanda. *arXiv preprint arXiv:2106.00213*.

Passarelli, Simone, Semira Abdelmenan, Dagmawit Tewahido, Yohannes Mulugeta Demmu, Helen Abreham, Ramya Ambikapathi, Nilupa S. Gunaratna, Yemane Berhane, and Wafaie Fawzi. 2022. Nutrition-sensitive chicken production in Ethiopia: a qualitative evaluation. *Agriculture & Food Security* 11 (1): 1-12.

Olney, Deanna, Aulo Gelli, Neha Kumar, Harold Alderman, Ara Go, Ahmed Raza, Jessica Owens, Alejandro Grinspun, Garima Bhalla and Omar Benammour. 2022. Impacts of social protection programs on women's and children's nutrition: A review. *Maternal and Child Nutrition*. e13378.

Raza, W. A., Van de Poel, E., & Van Ourti, T. (2018). Impact and spill-over effects of an asset transfer program on child undernutrition: Evidence from a randomized control trial in Bangladesh. *Journal of Health Economics*, *62*, 105-120.

Ruel, Marie and Harold Alderman. Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition?. 2013. *The Lancet* 382, no. 9891: 536-551.

Sibhatu, Kibrom T., and Matin Qaim. 2017. "Rural Food Security, Subsistence Agriculture, and Seasonality." *PLoS ONE* 12 (10): 1–15.

WHO and UNICEF. 2017. Global Nutrition Monitoring Framework. Operational Guidelines for meeting Targets for 2025. WHO Geneva.

World Bank. 2018. *The State of Social Safety Nets 2018*. Washington, DC: World Bank. doi:10.1596/978-1-4648-1254-5.

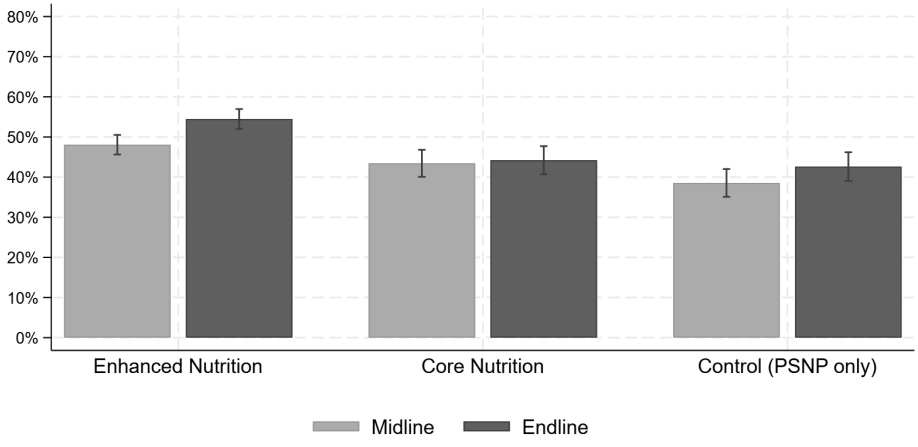
Figures and Tables

Figure 4: Programming by treatment arm

	Public works	WASH	Credit/savings	Grants	Nutritional	Men's engagement
Treatment 1	✓	✓	✓	✓	Enhanced	✓
Treatment 2	✓	✓	✓	✓	Core	✗
Treatment 3	✓	✓	✓	✗	Enhanced	✓
Control	✓	✓	✗	✗	✗	✗

Source: authors

Figure 5: Mother had contact with a HEW in the last 3 months, by survey round



Midline: N=3201
Endline: N=3061

Table 1: Balance in baseline characteristics

	N	Mean and standard deviation			p-value		
		Enhanced Nutrition	Core Nutrition	Control	EN vs Control	CN vs Control	EN vs CN
Log of total monthly consumption expenditure per adult equivalent	3,275	6.136 (0.728)	6.120 (0.731)	6.092 (0.705)	0.589	0.975	0.803
Distance to nearest town (x0.1km)	3,299	0.136 (0.083)	0.114 (0.069)	0.142 (0.067)	0.422	0.037	0.080
Household size	3,314	5.742 (1.979)	5.808 (1.956)	5.747 (1.924)	0.808	0.710	0.725
Mother's age	3,286	30.767 (7.548)	30.717 (6.774)	30.522 (7.129)	0.647	0.935	0.898
Mother has some education	3,282	0.195 (0.396)	0.223 (0.417)	0.197 (0.398)	0.472	0.307	0.321
Father's age	2,813	38.284 (9.475)	38.160 (9.072)	37.874 (9.234)	0.548	0.987	0.825
Father has some education	2,812	0.344 (0.475)	0.345 (0.476)	0.335 (0.472)	0.864	0.922	0.997
Number of food groups (out of 10) mother consumed during the previous day	3,296	2.094 (1.171)	1.958 (1.118)	2.040 (1.286)	0.156	0.070	0.059
Mother met the Minimum Dietary Diversity for Women (MDD-W)	3,296	0.031 (0.174)	0.033 (0.180)	0.040 (0.197)	0.479	0.919	0.819
Child's age in months	3,314	18.561 (10.543)	19.073 (10.260)	18.501 (10.096)	0.602	0.258	0.315
Child is male	3,301	0.505 (0.500)	0.525 (0.500)	0.507 (0.500)	0.500	0.332	0.350
Child is stunted (HAZ<-2SD)	3,179	0.370 (0.483)	0.398 (0.490)	0.362 (0.481)	0.685	0.317	0.378
Child is wasted (WHZ<-2SD)	3,158	0.134 (0.341)	0.166 (0.372)	0.114 (0.319)	0.756	0.209	0.312
Child's weight was measured in past 3 months	3,296	0.271 (0.445)	0.271 (0.445)	0.273 (0.446)	0.980	0.987	0.999
Child's height was measured in past 3 months	3,296	0.240 (0.427)	0.238 (0.426)	0.262 (0.440)	0.753	0.786	0.952
Number of food groups (out of 7) child consumed during the previous day	3,296	1.634 (0.817)	1.681 (0.814)	1.718 (0.880)	0.099	0.646	0.303
Child met the minimum meal frequency for 6-23 months old children	1,430	0.429 (0.495)	0.449 (0.498)	0.445 (0.498)	0.568	0.657	0.578

Notes: Estimates from the DFSA SPIR baseline survey sample. Standard deviations are in parentheses. P-value is from the test of difference of means between the treatment arms. Standard errors are clustered at the kebele level.

Table 2a: Anthropometrics, children 6-35 months of age, full sample

	(1) Height-for-age z-score (HAZ)	(2) Proportion stunted (HAZ<-2SD)	(3) Weight-for-height z-score (WHZ)	(4) Proportion wasted (WHZ<-2SD)	(5) Weight-for-age z-score (WAZ)	(6) Proportion underweight (WAZ<-2SD)	(7) Mid-upper arm circumference (MUAC)	(8) Proportion of severe acute malnutrition (MUAC<11.5 cm)
Enhanced Nutrition (EN)	0.075	-0.023	0.073	-0.005	0.096	-0.025	0.044	-0.008
	(0.126)	(0.036)	(0.091)	(0.023)	(0.099)	(0.029)	(0.094)	(0.012)
Core Nutrition (CN)	0.080	-0.003	0.159	-0.027	0.152	-0.021	0.022	-0.006
	(0.160)	(0.043)	(0.101)	(0.023)	(0.112)	(0.034)	(0.093)	(0.013)
Test: EN = CN	0.967	0.562	0.353	0.262	0.567	0.895	0.786	0.849
Mean of control	-1.853	0.486	-0.499	0.114	-1.357	0.314	13.681	0.031
N	1,140	1,140	1,140	1,140	1,140	1,140	1,140	1,140

Notes: Estimates from the DFSA SPIR endline survey sample for all children of age 6 to 35 months. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, child's sex and age in months, and whether child's mother has any education. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 2b: Anthropometrics, children 6-35 months of age, extremely poor households

	(1) Height-for-age z-score (HAZ)	(2) Proportion stunted (HAZ<-2SD)	(3) Weight-for-height z-score (WHZ)	(4) Proportion wasted (WHZ<-2SD)	(5) Weight-for-age z-score (WAZ)	(6) Proportion underweight (WAZ<-2SD)	(7) Mid-upper arm circumference (MUAC)	(8) Proportion of severe acute malnutrition (MUAC<11.5 cm)
EN x Poultry	0.256	-0.070	0.119	0.007	0.224	-0.028	0.323**	-0.012
	(0.263)	(0.071)	(0.172)	(0.033)	(0.158)	(0.055)	(0.143)	(0.019)
EN x Cash	0.431*	-0.179***	0.017	0.005	0.257	-0.060	0.170	-0.011
	(0.235)	(0.055)	(0.140)	(0.042)	(0.167)	(0.052)	(0.127)	(0.020)
EN without grants	-0.244	0.007	0.046	0.013	-0.082	-0.041	-0.085	-0.015
	(0.151)	(0.049)	(0.152)	(0.034)	(0.140)	(0.046)	(0.132)	(0.017)
CN x Poultry	0.182	-0.095	0.073	-0.002	0.154	-0.089	0.140	-0.000
	(0.261)	(0.059)	(0.177)	(0.032)	(0.187)	(0.064)	(0.146)	(0.021)
CN x Cash	-0.043	-0.011	0.113	-0.021	0.068	-0.038	-0.078	-0.002
	(0.196)	(0.055)	(0.133)	(0.029)	(0.148)	(0.048)	(0.130)	(0.023)
Average effect of EN with grants	0.345*	-0.125**	0.067	0.006	0.241*	-0.044	0.245**	-0.011
	(0.197)	(0.052)	(0.129)	(0.031)	(0.135)	(0.044)	(0.111)	(0.016)
Average effect of CN with grants	0.072	-0.054	0.093	-0.011	0.112	-0.064	0.033	-0.001
	(0.182)	(0.049)	(0.128)	(0.026)	(0.140)	(0.047)	(0.116)	(0.018)
Test: EN with grants = CN with grants	0.207	0.152	0.842	0.534	0.360	0.664	0.050**	0.551
Test: EN with grants = EN without grants	0.002***	0.006***	0.887	0.834	0.020**	0.942	0.005***	0.801
Test: EN x Poultry = EN x Cash	0.567	0.134	0.567	0.952	0.858	0.586	0.320	0.943
Test: CN x Poultry = CN x Cash	0.431	0.162	0.829	0.541	0.650	0.416	0.150	0.927
Mean of control	-1.990	0.543	-0.438	0.087	-1.395	0.353	13.638	0.029
N	637	637	637	637	637	637	637	637

Notes: Estimates from the DFSA SPIR endline survey for children of age 6 to 35 months from the subsample of extremely poor households. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, child's sex and age in months, and whether child's mother has any education. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. Sample of poor households is determined by ranking households within kebeles based on land and asset index constructed at baseline where 10 poorest households out of 18 in each kebele are classified as 'poor'.

Table 3: Mother's access to healthcare, extremely poor households

	(1) Average Standardized Treatment Effect: Access to healthcare	(2) Has had contact with a HEW or a HDA in past 3 months	(3) Has been visited by a HEW or a HDA at home in past 3 months	(4) Has attended a food demonstration in her community in last 3 months	(5) Has attended a BCC session in past 3 months	(6) Has attended a community led total sanitation and hygiene (CLTSH) event	(7) Has participated in a Open Defecation Free (ODF) event	(8) Has discussed nutrition topics at VESA meetings in past 12 months
EN x Poultry	0.570***	0.151***	0.109**	0.093***	0.192***	0.154***	0.098*	0.386***
	(0.094)	(0.053)	(0.045)	(0.036)	(0.038)	(0.051)	(0.054)	(0.048)
EN x Cash	0.535***	0.141***	0.094**	0.116***	0.149***	0.163***	0.144**	0.333***
	(0.107)	(0.042)	(0.037)	(0.037)	(0.046)	(0.056)	(0.060)	(0.056)
EN without grants	0.500***	0.132***	0.101***	0.132***	0.155***	0.112**	0.137***	0.289***
	(0.070)	(0.036)	(0.032)	(0.028)	(0.032)	(0.045)	(0.049)	(0.032)
CN x Poultry	0.225***	-0.016	0.025	-0.005	0.009	0.053	0.107*	0.249***
	(0.078)	(0.050)	(0.038)	(0.021)	(0.036)	(0.053)	(0.054)	(0.035)
CN x Cash	0.301***	0.081*	0.063	0.023	0.073*	0.055	0.042	0.258***
	(0.073)	(0.046)	(0.039)	(0.028)	(0.040)	(0.049)	(0.055)	(0.042)
Average effect of EN with grants	0.552***	0.146***	0.102***	0.105***	0.170***	0.158***	0.121**	0.359***
	(0.076)	(0.038)	(0.032)	(0.028)	(0.033)	(0.044)	(0.047)	(0.040)
Average effect of CN with grants	0.262***	0.031	0.044	0.009	0.041	0.054	0.075*	0.253***
	(0.060)	(0.038)	(0.031)	(0.020)	(0.030)	(0.042)	(0.045)	(0.030)
Test: EN with grants = CN with grants	0.000***	0.005***	0.086*	0.000***	0.000***	0.017**	0.302	0.011**
Test: EN with grants = EN without grants	0.549	0.726	0.989	0.408	0.691	0.312	0.751	0.102
Test: EN x Poultry = EN x Cash	0.790	0.868	0.768	0.624	0.416	0.889	0.476	0.439
Test: CN x Poultry = CN x Cash	0.404	0.107	0.413	0.280	0.175	0.982	0.290	0.838
Mean of control		0.399	0.159	0.063	0.109	0.307	0.336	0.046
N	1,662	1,662	1,662	1,662	1,662	1,662	1,662	1,662

Notes: Estimates from the DFSA SPIR endline survey for the subsample of extremely poor households. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects and whether mother has any education. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. Sample of poor households is determined by ranking households within kebeles based on land and asset index constructed at baseline where 10 poorest households out of 18 in each kebele are classified as 'poor'.

Table 4: Child's health history, extremely poor households

	(1) Average Standardized Treatment Effect: Access to health and nutrition services	(2) Received dose of Vitamin A in past 6 months	(3) Child's weight was measured in past 3 months	(4) Child's height was measured in past 3 months	(5) Child's MUAC was measured in past 3 months	(6) Child identified as severely or moderately underweight in past 6 months	(7) Child participated in a 2-week cooking demonstration and feeding session (CPNP)
EN x Poultry	0.059	0.101	-0.003	0.025	-0.010	0.014	0.016
	(0.119)	(0.089)	(0.084)	(0.063)	(0.085)	(0.057)	(0.047)
EN x Cash	0.106	-0.033	-0.036	0.033	0.002	0.029	0.138**
	(0.118)	(0.095)	(0.101)	(0.066)	(0.076)	(0.072)	(0.061)
EN without grants	-0.012	-0.069	-0.020	0.003	-0.079	0.028	0.043
	(0.104)	(0.087)	(0.071)	(0.056)	(0.064)	(0.061)	(0.048)
CN x Poultry	-0.130	-0.166	-0.064	-0.018	-0.104	-0.005	-0.007
	(0.111)	(0.103)	(0.089)	(0.064)	(0.084)	(0.054)	(0.035)
CN x Cash	-0.074	-0.097	-0.119	-0.081	-0.123	0.118	0.042
	(0.122)	(0.102)	(0.082)	(0.059)	(0.077)	(0.078)	(0.052)
Average effect of EN with grants	0.083	0.033	-0.019	0.029	-0.004	0.022	0.078*
	(0.097)	(0.073)	(0.074)	(0.053)	(0.065)	(0.053)	(0.044)
Average effect of CN with grants	-0.102	-0.132	-0.091	-0.049	-0.113*	0.056	0.017
	(0.097)	(0.082)	(0.070)	(0.051)	(0.065)	(0.056)	(0.038)
Test: EN with grants = CN with grants	0.070*	0.044**	0.377	0.132	0.135	0.572	0.128
Test: EN with grants = EN without grants	0.364	0.233	0.990	0.637	0.270	0.923	0.461
Test: EN x Poultry = EN x Cash	0.732	0.234	0.768	0.917	0.901	0.848	0.062*
Test: CN x Poultry = CN x Cash	0.670	0.574	0.587	0.354	0.843	0.102	0.278
Mean of control		0.505	0.308	0.143	0.363	0.176	0.055
N	341	341	341	341	341	341	341

Notes: Estimates from the DFSA SPIR endline survey for children aged 6-13 months from the extremely poor subsample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, child's gender and age in months, and whether mother has any education. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. Sample of poor households is determined by ranking households within kebeles based on land and asset index constructed at baseline where 10 poorest households out of 18 in each kebele are classified as 'poor'.

Table 5: Mother's IYCF knowledge, extremely poor households

	(1) Female IYCF knowledge score (0-7)	(2) Knows how long after birth should a baby start breastfeeding	(3) Knows until what age a baby should be exclusively breastfed	(4) Knows what to do if a mother thinks her baby is not getting enough breast milk	(5) Knows which foods are rich in vitamin A	(6) Knows the common problem with gruels given as first foods to babies	(7) Knows how often a baby 6-23 months old should eat animal source foods	(8) Knows how often a child should be fed when sick
EN x Poultry	0.245*	0.012	0.036	0.009	0.054	-0.003	0.069**	0.067*
	(0.128)	(0.034)	(0.023)	(0.048)	(0.041)	(0.044)	(0.033)	(0.035)
EN x Cash	0.066	0.027	-0.003	-0.078*	0.086*	0.018	0.006	0.010
	(0.117)	(0.028)	(0.027)	(0.039)	(0.045)	(0.030)	(0.037)	(0.045)
EN without grants	0.161*	0.046*	0.026	-0.061	0.091***	-0.002	0.006	0.055
	(0.097)	(0.026)	(0.022)	(0.038)	(0.035)	(0.034)	(0.029)	(0.036)
CN x Poultry	0.001	0.002	-0.014	-0.023	0.036	0.024	-0.001	-0.023
	(0.121)	(0.033)	(0.029)	(0.049)	(0.050)	(0.042)	(0.034)	(0.037)
CN x Cash	0.079	0.026	0.012	-0.039	0.124***	-0.072**	-0.026	0.053
	(0.107)	(0.031)	(0.028)	(0.041)	(0.043)	(0.035)	(0.047)	(0.041)
Average effect of EN with grants	0.154	0.019	0.016	-0.035	0.070**	0.008	0.037	0.038
	(0.102)	(0.026)	(0.021)	(0.037)	(0.035)	(0.032)	(0.030)	(0.033)
Average effect of CN with grants	0.039	0.014	-0.001	-0.031	0.079**	-0.023	-0.013	0.015
	(0.097)	(0.026)	(0.023)	(0.037)	(0.038)	(0.032)	(0.034)	(0.032)
Test: EN with grants = CN with grants	0.207	0.812	0.430	0.916	0.810	0.279	0.085*	0.465
Test: EN with grants = EN without grants	0.941	0.274	0.624	0.468	0.540	0.744	0.196	0.645
Test: EN x Poultry = EN x Cash	0.186	0.651	0.154	0.072*	0.528	0.600	0.071*	0.211
Test: CN x Poultry = CN x Cash	0.521	0.490	0.427	0.755	0.106	0.023**	0.588	0.096*
Mean of control	3.967	0.865	0.841	0.278	0.596	0.292	0.715	0.380
N	1,693	1,693	1,693	1,693	1,693	1,693	1,693	1,693

Notes: Estimates from the DFSA SPIR endline survey for the subsample of extremely poor households. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects and whether mother has any education. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. Sample of poor households is determined by ranking households within kebeles based on land and asset index constructed at baseline where 10 poorest households out of 18 in each kebele are classified as 'poor'.

Table 6: Mother's dietary diversity, extremely poor households

	(1) Women's Dietary Diversity Score, WDDS (1-10)	(2) Meets Minimum Dietary Diversity for Women, MDD-W	(3) Mother consumed tubers and grains	(4) Mother consumed pulses	(5) Mother consumed nuts and seeds	(6) Mother consumed dairy	(7) Mother consumed meat, fish, poultry	(8) Mother consumed eggs	(9) Mother consumed green leafy vegetables	(10) Mother consumed vitamin A rich fruits and vegetables	(11) Mother consumed other vegetables	(12) Mother consumed other fruits
EN x Poultry	0.257**	0.044**	-0.001	0.010	0.102**	-0.006	0.012	0.023	0.021	0.014	0.075	0.007
	(0.121)	(0.022)	(0.014)	(0.027)	(0.048)	(0.031)	(0.020)	(0.022)	(0.019)	(0.023)	(0.058)	(0.016)
EN x Cash	0.151	0.020	-0.028	-0.029	0.027	0.018	0.025	-0.012	0.033	0.024	0.096*	-0.003
	(0.131)	(0.023)	(0.020)	(0.047)	(0.040)	(0.028)	(0.015)	(0.016)	(0.027)	(0.029)	(0.054)	(0.014)
EN without grants	0.171	0.037	-0.011	-0.025	0.027	0.025	0.016	0.018	0.021	-0.002	0.080*	0.021
	(0.111)	(0.023)	(0.013)	(0.025)	(0.032)	(0.027)	(0.015)	(0.016)	(0.018)	(0.022)	(0.042)	(0.014)
CN x Poultry	0.099	0.031	-0.029	-0.034	-0.022	0.005	0.023	0.008	0.026	0.029	0.075	0.017
	(0.138)	(0.027)	(0.019)	(0.036)	(0.039)	(0.028)	(0.016)	(0.016)	(0.020)	(0.030)	(0.055)	(0.026)
CN x Cash	-0.033	0.016	-0.017	-0.066**	0.027	0.048	0.004	-0.010	0.001	0.011	-0.039	0.008
	(0.128)	(0.026)	(0.016)	(0.031)	(0.042)	(0.033)	(0.014)	(0.014)	(0.022)	(0.030)	(0.051)	(0.018)
Average effect of EN with grants	0.203**	0.032	-0.015	-0.009	0.064*	0.006	0.019	0.005	0.027	0.019	0.086*	0.002
	(0.103)	(0.019)	(0.013)	(0.030)	(0.036)	(0.025)	(0.015)	(0.015)	(0.019)	(0.021)	(0.046)	(0.012)
Average effect of CN with grants	0.034	0.024	-0.023	-0.050*	0.002	0.026	0.014	-0.001	0.013	0.020	0.019	0.013
	(0.107)	(0.022)	(0.014)	(0.026)	(0.033)	(0.026)	(0.013)	(0.013)	(0.018)	(0.024)	(0.044)	(0.017)
Test: EN with grants = CN with grants	0.120	0.674	0.618	0.219	0.084*	0.379	0.704	0.678	0.458	0.953	0.132	0.509
Test: EN with grants = EN without grants	0.773	0.807	0.786	0.633	0.284	0.445	0.839	0.452	0.744	0.349	0.900	0.136
Test: EN x Poultry = EN x Cash	0.467	0.328	0.215	0.432	0.154	0.433	0.516	0.140	0.667	0.751	0.745	0.521
Test: CN x Poultry = CN x Cash	0.411	0.615	0.581	0.439	0.300	0.191	0.234	0.280	0.251	0.623	0.062*	0.734
Mean of control	2.606	0.057	0.979	0.568	0.207	0.109	0.036	0.045	0.052	0.086	0.494	0.031
N	1,693	1,693	1,693	1,693	1,693	1,693	1,693	1,693	1,693	1,693	1,693	1,693

Notes: Estimates from the DFSA SPIR endline survey for the subsample of extremely poor households. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, whether mother has any education, and whether the day before the survey was a fast day. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. Sample of poor households is determined by ranking households within kebeles based on land and asset index constructed at baseline where 10 poorest households out of 18 in each kebele are classified as 'poor'.

Table 7: Child's dietary diversity, children 6-23 months of age, extremely poor households

	(1) Number of food groups (of 8) consumed	(2) Child meets meet minimum dietary diversity	(3) Child meets minimum meal frequency	(4) Child consumed breast milk	(5) Child consumed grains, roots or tubers	(6) Child consumed legumes or nuts	(7) Child consumed dairy	(8) Child consumed fish or meat	(9) Child consumed eggs	(10) Child consumed vitamin A rich fruits or vegetables	(11) Child consumed other fruits or vegetables
EN x Poultry	0.375**	0.039	0.131**	-0.003	0.016	0.065	0.159***	0.011	0.140**	-0.014	0.000
	(0.181)	(0.030)	(0.061)	(0.028)	(0.064)	(0.061)	(0.057)	(0.023)	(0.059)	(0.019)	(0.002)
EN x Cash	-0.066	-0.008	0.020	0.015	-0.024	-0.097**	0.067	-0.011	-0.041*	0.005	0.020
	(0.117)	(0.012)	(0.068)	(0.017)	(0.068)	(0.044)	(0.063)	(0.012)	(0.025)	(0.026)	(0.019)
EN without grants	-0.043	-0.012	0.044	0.002	-0.051	-0.073	0.098	-0.001	-0.009	-0.008	-0.001
	(0.119)	(0.015)	(0.054)	(0.019)	(0.047)	(0.050)	(0.078)	(0.020)	(0.029)	(0.022)	(0.002)
CN x Poultry	0.153	-0.008	0.113	0.023	0.068	-0.083*	0.150*	0.016	-0.033	0.015	-0.003
	(0.120)	(0.011)	(0.077)	(0.015)	(0.060)	(0.048)	(0.077)	(0.022)	(0.046)	(0.026)	(0.004)
CN x Cash	0.136	0.037	0.044	-0.053	0.015	0.006	0.075	0.013	0.026	0.055	-0.002
	(0.131)	(0.030)	(0.075)	(0.039)	(0.051)	(0.063)	(0.081)	(0.025)	(0.040)	(0.037)	(0.003)
Average effect of EN with grants	0.152	0.015	0.075	0.006	-0.004	-0.017	0.112**	0.000	0.049	-0.004	0.010
	(0.123)	(0.018)	(0.051)	(0.020)	(0.054)	(0.043)	(0.053)	(0.015)	(0.035)	(0.018)	(0.010)
Average effect of CN with grants	0.145	0.014	0.079	-0.014	0.042	-0.039	0.113*	0.014	-0.004	0.035	-0.002
	(0.106)	(0.018)	(0.062)	(0.023)	(0.046)	(0.045)	(0.068)	(0.018)	(0.035)	(0.026)	(0.003)
Test: EN with grants = CN with grants	0.953	0.977	0.948	0.406	0.425	0.635	0.979	0.422	0.158	0.179	0.304
Test: EN with grants = EN without grants	0.113	0.071*	0.569	0.817	0.399	0.262	0.833	0.925	0.063*	0.881	0.305
Test: EN x Poultry = EN x Cash	0.014**	0.092*	0.165	0.428	0.593	0.011**	0.097*	0.282	0.002***	0.462	0.307
Test: CN x Poultry = CN x Cash	0.897	0.100*	0.443	0.030**	0.387	0.178	0.341	0.931	0.242	0.301	0.693
Mean of control	2.283	0.011	0.674	0.978	0.859	0.174	0.185	0.011	0.054	0.022	0.000
N	353	353	353	353	353	353	353	353	353	353	353

Notes: Estimates from the DFSA SPIR endline survey for children aged 6-23 months from the subsample of extremely poor households. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, child's gender and age in months, whether mother has any education, and whether the day before the survey was a fast day. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. Sample of poor households is determined by ranking households within kebeles based on land and asset index constructed at baseline where 10 poorest households out of 18 in each kebele are classified as 'poor'.

Table 8: Summary by family of outcome

	(1) Height-for-age z-score (HAZ)	(2) Proportion stunted (HAZ<-2SD)	(3) ASTE: Mother's access to healthcare	(4) ASTE: Child's health and services	(5) Mother's IYCF knowledge score (0-7)	(6) Mother's Dietary Diversity Score (1-10)	(7) Child's total food groups consumed (0-8)
Extremely poor							
Enhanced Nutrition (EN)	0.087	-0.069	0.526***	0.039	0.157**	0.188**	0.027
	(0.166)	(0.045)	(0.048)	(0.083)	(0.069)	(0.073)	(0.103)
Core Nutrition (CN)	0.078	-0.055	0.265***	-0.101	0.041	0.031	0.155
	(0.186)	(0.051)	(0.055)	(0.095)	(0.080)	(0.084)	(0.118)
Mean of control	-1.990	0.543	-0.000	-0.000	3.967	2.606	2.290
N	637	637	1,662	341	1,693	1,693	355
Less poor							
Enhanced Nutrition (EN)	0.009	0.054	0.597***	0.145	0.156	0.150	0.168
	(0.176)	(0.051)	(0.059)	(0.107)	(0.083)	(0.079)	(0.132)
Core Nutrition (CN)	0.084	0.078	0.294***	-0.007	-0.090	-0.079	0.157
	(0.205)	(0.059)	(0.068)	(0.119)	(0.095)	(0.091)	(0.149)
Mean of control	-1.650	0.402	0.000	-0.000	3.960	2.739	2.200
N	503	503	1,238	251	1,256	1,256	260
Extremely poor: EN = CN	0.957	0.764	0.000***	0.105	0.089*	0.030**	0.229
Less poor: EN = CN	0.662	0.622	0.000***	0.126	0.002***	0.003***	0.930
EN: Extremely poor = less poor	0.746	0.074*	0.183	0.421	0.992	0.689	0.394
CN: Extremely poor = less poor	0.981	0.054*	0.617	0.516	0.235	0.287	0.992

Notes: Estimates from the DFSA SPIR endline survey sample. Columns (1) and (2) are restricted to children aged 6-35 months only, columns (4) and (7) are restricted to children aged 6-23 months only. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects and whether child's mother has any education; columns (1), (2), (4) and (7) additionally control for child's sex and age in months; columns (6) and (7) control for whether the day preceding the survey was a fast day. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors

Annex 1: Testing the persistence of results of poultry ownership: midline and endline

Annex table 1 compares the midline results from Alderman et al. (2022) with the endline impact a year and a half later. The specification replicates the earlier results using the full sample including a supplementary sample added at midline to refresh the sample of children less than 24 months. Thus, the specification differs slightly from the results presented in the main text. The opportunity to return to the initial results indicates that the significant increase in egg consumption for all households eligible for poultry grants that was observed at midline is no longer apparent. This is likely driven by the fact that at midline – less than 6 months after the grants were distributed – poultry recipients had 7.5 more birds than household that did not receive the transfer including those who had a cash grant of equivalent value. Two years after the grants were disbursed this difference had declined to only 1.5 birds, although the increase in poultry owned remained statistically significant (column 8). However, the coefficient of the interaction of enhanced nutrition and poultry in column 6 reproduces the result in Table 7 despite the slight differences in sample and specification.

Annex Table 1: Testing the persistence of results of poultry ownership: midline and endline

	Midline (Food Policy 2022)					Endline				
	(1) Child consumed eggs	(2) Primary female consumed eggs	(3) Total number of poultry owned by household, midline	(4) Child consumed eggs	(5) Primary female consumed eggs	(6) Child consumed eggs	(7) Primary female consumed eggs	(8) Total number of poultry owned by household, midline	(9) Child consumed eggs	(10) Primary female consumed eggs
Assigned to poultry transfer	0.122***	0.071***	7.560***			-0.009	0.010	1.479***		
	(0.046)	(0.022)	(0.496)			(0.044)	(0.014)	(0.275)		
Assigned to cash grant	0.033	-0.008	0.358			0.032	-0.005	0.226		
	(0.034)	(0.012)	(0.269)			(0.036)	(0.012)	(0.200)		
Enhanced Livelihood (excluding poultry and cash)	0.012	-0.005	0.714***			0.020	-0.001	0.480***		
	(0.021)	(0.010)	(0.223)			(0.022)	(0.009)	(0.172)		
Enhanced Nutrition	0.049***	0.012	0.289	0.039**	0.007	0.002	0.017*	0.084	0.003	0.012
	(0.019)	(0.010)	(0.221)	(0.019)	(0.009)	(0.021)	(0.009)	(0.162)	(0.018)	(0.008)
Interaction of Enhanced Nutrition and poultry	-0.011	-0.029				0.131*	0.001			
	(0.069)	(0.027)				(0.076)	(0.024)			
Interaction of Enhanced Nutrition and cash	-0.007	0.025				-0.081*	-0.021			
	(0.047)	(0.022)				(0.042)	(0.018)			
Predicted baseline expenditure	0.081	0.042	0.263	0.099	0.053					
	(0.067)	(0.034)	(0.548)	(0.069)	(0.034)					
Interview was conducted on a day after fasting	-0.024	-0.069***		-0.016	-0.068***	-0.033	-0.056***		-0.034*	-0.054***
	(0.027)	(0.014)		(0.029)	(0.014)	(0.022)	(0.010)		(0.020)	(0.010)
Mother has some education	0.031	0.010	-0.011	0.013	0.001	0.030	0.007	-0.022	0.020	0.007
	(0.028)	(0.013)	(0.163)	(0.027)	(0.012)	(0.026)	(0.010)	(0.149)	(0.025)	(0.010)
Father has some education	0.002	0.003	0.086	0.005	0.001	0.036	0.020**	0.200	0.038*	0.018**
	(0.021)	(0.011)	(0.160)	(0.020)	(0.010)	(0.022)	(0.009)	(0.163)	(0.022)	(0.008)
Household size	0.009	0.005	0.154**	0.008	0.005	0.004	-0.003	0.093**	0.005	-0.004**
	(0.009)	(0.004)	(0.064)	(0.009)	(0.004)	(0.004)	(0.002)	(0.037)	(0.004)	(0.002)
Child age in months	0.003**			0.003**						
	(0.002)			(0.002)						
Male child	0.012			0.008						
	(0.018)			(0.019)						

Baseline value of outcome	0.064	0.005		0.059	0.007					
	(0.084)	(0.028)		(0.090)	(0.030)					
Total number of poultry owned by household, baseline			0.180***					0.121***		
			(0.033)					(0.028)		
Total number of poultry owned by household, midline				0.007***	0.005***				0.016***	0.011***
				(0.003)	(0.001)				(0.004)	(0.002)
Endline Index Child's age (months)						0.006**			0.006**	
						(0.003)			(0.003)	
Endline Index Child is male						0.002			0.006	
						(0.019)			(0.018)	
Constant	-0.619	-0.208	0.497	-0.768	-0.252	0.006	0.111***	2.072***	-0.113***	0.075***
	(0.418)	(0.216)	(3.751)	(0.475)	(0.217)	(0.058)	(0.041)	(0.408)	(0.042)	(0.021)
R^2	0.06	0.04	0.40	0.06	0.04	0.07	0.03	0.08	0.09	0.05
N	1,009	3,114	3,114	974	3,013	722	3,704	3,080	721	3,695
Mean of control	0.041	0.039	1.814	0.041	0.039	0.044	0.04	1.971	0.044	0.04

Notes: Estimates from the DFSA SPIR endline and midline survey samples; the midline results are from Alderman et al 2022, Food Policy. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors. Where primary female's/male's education is missing, it is replaced with 0 and respectively controlled for.

Annex Table 2: Balance in baseline characteristics
Extremely poor sample

	N	Mean and standard deviation			p-value		
		Enhanced Nutrition	Core Nutrition	Control	EN vs Control	CN vs Control	EN vs CN
Log of total monthly consumption expenditure per adult equivalent	1,880	6.095 (0.754)	6.107 (0.725)	6.077 (0.719)	0.960	0.795	0.873
Distance to nearest town (x0.1km)	1,897	0.136 (0.085)	0.115 (0.071)	0.140 (0.069)	0.414	0.050	0.094
Household size	1,907	5.434 (1.970)	5.534 (1.942)	5.405 (1.865)	0.825	0.558	0.625
Mother's age	1,901	30.576 (7.814)	30.380 (7.204)	30.128 (7.128)	0.417	0.916	0.696
Mother has some education	1,899	0.194 (0.395)	0.227 (0.419)	0.222 (0.416)	0.244	0.464	0.338
Father's age	1,474	37.519 (9.436)	37.265 (8.703)	36.856 (9.074)	0.401	0.957	0.707
Father has some education	1,473	0.335 (0.472)	0.329 (0.471)	0.358 (0.480)	0.799	0.737	0.894
Number of food groups (out of 10) mother consumed during the previous day	1,897	1.947 (1.082)	1.800 (1.028)	1.955 (1.229)	0.367	0.075	0.103
Mother met the Minimum Dietary Diversity for Women (MDD-W)	1,897	0.018 (0.132)	0.017 (0.128)	0.035 (0.183)	0.319	0.388	0.884
Child's age in months	1,907	18.829 (10.633)	18.973 (10.370)	18.009 (10.026)	0.553	0.438	0.804
Child is male	1,901	0.522 (0.500)	0.512 (0.500)	0.522 (0.500)	0.836	0.725	0.745
Child is stunted (HAZ<-2SD)	1,826	0.376 (0.485)	0.399 (0.490)	0.371 (0.484)	0.754	0.480	0.521
Child is wasted (WHZ<-2SD)	1,813	0.124 (0.330)	0.173 (0.379)	0.118 (0.323)	0.370	0.114	0.143
Child's weight was measured in past 3 months	1,897	0.247 (0.432)	0.291 (0.455)	0.271 (0.445)	0.286	0.345	0.282
Child's height was measured in past 3 months	1,897	0.224 (0.417)	0.254 (0.436)	0.264 (0.441)	0.271	0.672	0.471

Number of food groups (out of 7) child consumed during the previous day	1,897	1.595 (0.792)	1.628 (0.719)	1.673 (0.876)	0.256	0.886	0.541
Child met the minimum meal frequency for 6-23 months old children	793	0.433 (0.496)	0.402 (0.492)	0.392 (0.489)	0.367	0.702	0.518

Notes: Estimates from the DFSA SPIR baseline survey sample. Standard deviations are in parentheses. P-value is from the test of difference of means between the treatment arms. Standard errors are clustered at the kebele level.

Annex Table 3: Summary by family of outcome
Including the supplemental sample

	(1) Height-for-age z-score (HAZ)	(2) Proportion stunted (HAZ<-2SD)	(3) ASTE: Mother's access to healthcare	(4) ASTE: Child's health and services	(5) Mother's IYCF knowledge score (0-7)	(6) Mother's Dietary Diversity Score (1-10)	(7) Child's total food groups consumed (0-8)
Extremely poor							
Enhanced Nutrition (EN)	-0.041	-0.027	0.471***	0.084	0.162**	0.185***	-0.060
	(0.138)	(0.039)	(0.041)	(0.077)	(0.063)	(0.065)	(0.098)
Core Nutrition (CN)	0.088	-0.019	0.197***	-0.117	0.010	-0.004	-0.006
	(0.152)	(0.043)	(0.047)	(0.087)	(0.072)	(0.074)	(0.110)
Mean of control	-1.911	0.514	-0.000	-0.000	3.928	2.633	2.364
N	916	916	2,069	393	2,107	2,107	407
Less poor							
Enhanced Nutrition (EN)	-0.107	0.064	0.638***	0.182	0.174	0.261	0.138
	(0.153)	(0.043)	(0.054)	(0.101)	(0.074)	(0.072)	(0.125)
Core Nutrition (CN)	0.069	0.063	0.304***	-0.017	-0.055	0.021	0.100
	(0.182)	(0.051)	(0.063)	(0.114)	(0.086)	(0.085)	(0.143)
Mean of control	-1.602	0.395	-0.000	0.000	3.935	2.648	2.219
N	698	698	1,510	287	1,531	1,531	300
Extremely poor: EN = CN	0.341	0.838	0.000***	0.011**	0.013**	0.003***	0.591
Less poor: EN = CN	0.249	0.987	0.000***	0.034**	0.002***	0.001***	0.750
EN: Extremely poor = less poor	0.748	0.132	0.001***	0.401	0.886	0.415	0.175
CN: Extremely poor = less poor	0.936	0.178	0.074*	0.419	0.494	0.799	0.554

Notes: Estimates from the DFSA SPIR endline survey sample, including an additional sample of households with a child aged 0-24 months that was added at the time of the midline survey. Since these households' poverty status was not classified based on asset data from the baseline survey, we predicted the variable that determines whether they belong to the extremely poor or to the less poor subsample using their characteristics reported in the midline survey. Columns (1) and (2) are restricted to children aged 6-35 months only, columns (4) and (7) are restricted to children aged 6-23 months only. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample indicator, and whether child's mother has any education at baseline (or midline for the supplemental sample); columns (1), (2), (4) and (7) additionally control for child's sex and age in months; columns (6) and (7) control for whether the day preceding the survey was a fast day. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard error

Chapter 5: Men's Engagement

Men can cook: Effectiveness of a light-touch men's engagement intervention to change attitudes and behaviors in rural Ethiopia

Harold Alderman, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Michael Mulford, Heleene Tambet

Abstract

Graduation model interventions seek to address multiple barriers constraining households' exit from poverty, however, few explicitly target unequal gender norms. Using a randomized control trial design, combined with three rounds of data, we investigate the impacts on gender equitable attitudes and behaviors of a graduation program that seeks to simultaneously "push" households out of poverty and improve unequal gender norms in Ethiopia. We find that at midline all treatment arms lead to improvements in men's gender equitable attitudes and their engagement in household domestic tasks as reported by both men and women; but at endline, impacts are only sustained in the treatment arms that introduced men's engagement groups after the midline survey to further promote improvements in equitable gender norms.

1. Introduction

Many development programs seek to improve gender equality by directly empowering women through investments in human capital, access to resources, or productive livelihood options. While several indicators point to growing improvements in gender equity in recent years, with increased girls' school enrollment and women's labor force participation, an area where gender inequality broadly persists is in the home. Even when both husband and wife work outside of the home, women often perform a larger share of home-based domestic tasks. In many settings, social norms around gender roles play a big part (Jayachandran, 2021). These restrictive gender norms can lead to inefficient allocation of household labor that is neither pareto-efficient for maximizing the individual utility of household members, nor for maximizing household joint-production functions such as in agricultural production or in raising healthy, well-educated children (Couprie et al., 2020; Duflo & Udry, 2004; Udry, 1996). Moreover, these restrictive gender norms persist over long time horizons and intergenerationally even outside of original country settings (Alesina et al., 2013). However, until recently, development programs, tended to work around restrictive social norms instead of directly targeting them.

In this paper we investigate the impact of a graduation model program that seeks to simultaneously "push" households out of poverty and improve gender equitable norms at scale. The graduation program, Strengthen PSNP Institutions and Resilience (SPIR), provided multi-sectoral livelihoods and nutrition support to nearly 150,000 food insecure households who were clients of Ethiopia's Productive Safety Net Program (PSNP). Approaches to promote gender equity were mainstreamed throughout SPIR activities, including bringing men and women together in Village Economic and Social Associations (VESAs). These served as a platform for facilitated discussions around financial literacy, savings and credit, income generating activities, and gender equitable norms. Moreover, as a part of its men's engagement strategy, SPIR formed men's engagement groups to provide an opportunity for men to critically reflect on cultural gender norms and explore the positive and perceived negative effects of male involvement in tasks traditionally assigned to women.

Using a randomized control trial (RCT) design, we investigate whether SPIR improves men's gender equitable attitudes and behavior. We take advantage of three rounds of data (baseline, midline, endline) and the timing of the men's engagement groups to identify the added impact of the men's engagement groups. We find that, at midline, SPIR leads to improvements in men's gender equitable attitudes and their involvement in household tasks. By endline, only treatment arms with the men's engagement groups continue to have significant impacts. Impacts at endline for the treatment arms with the men's engagement groups are large and range from 0.26-0.29 standard deviations (SD) for men's reported involvement in household tasks. Similar increases are found from women's reports of their partner's behavior, alleviating worries of self-reporting bias. Impacts on gender equitable norms are smaller in magnitude, ranging from 0.07-0.14 SD. For both set of outcomes, impacts at endline from the treatment arms receiving the men's engagement intervention are significantly different from the arm that did not receive the intervention. Thus, we attribute the sustained or larger impacts at endline to the men's engagement groups, which were rolled out after the midline survey.

Our findings contribute to the recent literature on the importance of men's engagement strategies in achieving improved gender equity and related outcomes (Glinski et al., 2018) including RCTs evaluating the impact of these interventions on gender attitudes (Dhar et al., 2022; Nguyen & Tarp, 2022), intimate partner violence (Abramsky et al., 2014; Christofides et al., 2020; Doyle et al., 2018; Hossain et al., 2014; Sharma et al., 2020; Vaillant et al., 2020), HIV-prevention (Jewkes et al., 2008;

Sharma et al., 2020) and family planning (Raj et al., 2016; Shattuck et al., 2011). We contribute to this literature in three notable ways. First, in contrast to the studies mentioned above, we study a men's engagement intervention in the context of a multicomponent anti-poverty program. Second, while gender equitable approaches were mainstreamed into different program components, the men's engagement activity is lighter touch than most of those mentioned above and implemented at a larger scale.[1] Lastly, there are only a few experimental studies that include outcomes related to men's involvement in household tasks and even fewer that include both women's and men's reports.[2]

We also contribute to the literature on graduation model programs which generally find mixed results on women's empowerment (Banerjee et al., 2015; Bossuroy et al., 2022). One potential reason is that these graduation model programs do not explicitly seek to improve gender equitable norms, which may create an enabling environment for women. The exception is a recent study of a women-focused graduation program in the Democratic Republic of Congo (DRC) which also consisted of a 16-week men's engagement program (Angelucci et al., 2023). They find large and significant impacts of the graduation program on consumption, women's employment and finances, empowerment, and health, but no additional effects from the men's engagement program. However, they did not directly report on men's attitudes or behaviors, and thus we do not know whether the male engagement component shifted these more proximal outcomes.

2. Experimental design

2.1 Context

The study takes place in rural Ethiopia where gender norms are patriarchal and traditional, with men considered heads of family and the main decisionmakers, while women are expected to be submissive to men (Alemu 2007; Mabsout et al 2010). Qualitative work on the study sample reveals that men and women find physical violence acceptable as a form of disciplinary action when women do not complete their household responsibilities or obey their husband (Ranganathan et al 2022). Moreover a third of men and more than one half of women in the study sample believe a husband is justified beating his wife under some circumstances, including perceived negligence of domestic duties such as burning food or neglecting children (Alderman et al 2021).

In terms of gender roles, women are responsible for domestic chores and men responsible for farm activities, to which women and children also contribute (Tefera, 2020). According to the Ethiopian Central Statistical Agency (2014), 94 percent of rural women were involved in domestic activities, spending an average of about 5 to 6 hours per day. While most men are accepting of men

[1] Most of the interventions cited above were time intensive with 14-16 sessions lasting 2-3 hours each (38 – 48 hours total), while the lighter touch men's engagement intervention studied here consisted of 8 sessions lasting about 2 hours each (16 hours total).

[2] A men's engagement intervention to improve health outcomes in Rwanda found a greater sharing of childcare and household tasks reported by both men and women at endline, but no statistically significant reduction in women's time spent on these tasks (Doyle et al., 2018). Two studies to reduce IPV through male engagement interventions found that although the interventions had no significant impact on IPV, the interventions did improve men's engagement in household tasks (Hossain et al., 2014; Vaillant et al., 2020). In Ethiopia, a gender transformative intervention to reduce IPV was tested with men's groups, women's groups, and couples' groups, and, while finding mixed results on IPV, women reported an increase in male involvement in childcare and household tasks in the couples' treatment arm, but these effects were not statistically significant in other treatment groups (Sharma et al., 2020).

contributing to domestic chores, men significantly underestimate the extent to which their peers hold this view (Mccullough et al., 2022). Still, only one third of men in rural Ethiopia are reported by their spouses to help with household chores, and when they do, this is a rare occurrence (Central Statistical Agency and ICF 2016). In our sample, only one fifth of women report that their spouses help with household chores (Alderman et al 2021).

2.2 Interventions

PSNP and SPIR

Initiated in 2005, the PSNP is one of the largest social protection programs in sub-Saharan Africa and is a key element of the Ethiopian government's strategy for poverty alleviation and rural development. The PSNP provides cash and/or food transfers to the poorest 10-15 percent of households in food insecure areas as payments for seasonal labor on public work sites or as direct support to vulnerable households who are unable to provide labor contributions. SPIR was a multisectoral graduation program that supported implementation of the fourth phase of the PSNP as well as provided complementary livelihood, nutrition, and gender equity activities to strengthen the program and expand its impacts. SPIR provided support to over 150,000 PNSP households in 15 food insecure districts in Amhara and Oromia regions.

The SPIR program was organized around a core set of livelihood and nutrition activities designed with gender equity considerations in mind, as well as specific interventions designed to shape gender norms. The core livelihood activities (L) were centered around VESAs, supporting group saving and lending activities as well as providing a platform for gender dialogues and trainings on income generation activities and financial literacy.[3] The core nutrition activities (N) include nutrition behavior change communication (BCC) and water, sanitation, and hygiene (WASH) activities. For learning purposes, SPIR introduced enhanced models of both the livelihood and nutrition activities (see Table 1). The enhanced livelihood activity (L*) added livelihood transfers to the core L activities. The livelihood transfers were $200 in Ethiopian birr or an equivalent value of improved poultry start-up kit provided to the poorest 60 percent of households in the study sample. The enhanced nutrition activities (N*) added to the core N activities more targeted nutrition counseling for pregnant and lactating women called timed and targeted counseling (TTC), a 2-week community-based participatory nutrition promotion (CPNP) activity for caregivers with underweight children, men's engagement groups, and Interpersonal Psychotherapy in Groups (IPT-G) for women who were screened for elevated depressive symptoms at midline.

VESA Gender dialogue

The gender dialogue sessions that were part of VESAs included 6 discussion sessions on 1) workloads of men and women, 2) cooperation and sharing household work; 3) household decision making; 4) improved listening, communication and understanding skills; 5) engaging men in childcare work; and 6) identifying restrictive social norms related to women's mobility. These one-hour discussions were incorporated into the bimonthly VESA meetings facilitated by SPIR Community Facilitators or VESA volunteer facilitators. Sessions were designed to engage women and men through activities that lead to reflection about their own situations. At midline, self-reported male participation in VESAs was about 50 percent and of these about 44 percent reported discussing topics related to gender (Alderman et al., 2020).

[3] Over 5,000 VESAs were formed with more than 117,000 members, including both husband and wife from PSNP households. Depending on the size of the sub-district administrative unit (kebele), between 10-25 VESAs were formed per kebele.

Men's engagement groups

The men's engagement activity consisted of eight sessions that facilitated men's reflections on gender norms in their community and sharing their own beliefs or opinions on gender-related expectations and roles. The sessions covered topics on gender roles, power and early marriage, father's legacy, caregiving, and nutrition. The session on father's legacy challenged men to identify things they would wish to emulate from their fathers and areas they would want to change in their own role as fathers. They also role-played doing household chores that are predominantly performed by girls or women and were challenged to try these out in their own home. Men later shared their experiences together in the group and encouraged and challenged each other to increase their efforts and continue practicing these counter-cultural actions in their home. While the sessions begin by creating awareness of the unequal workload and unfair burden of household tasks that rest on women, the motivation for men to engage in supporting these chores comes both from a sense that the status quo is unfair and a sense that they will themselves benefit from becoming a supportive partner and better father.

Men's engagement groups were facilitated by men's group facilitators (MGF) and male advocators. The MGFs began by identifying 'male advocators,' men who had the potential to become role models of attitude and behavior change in their communities. If they were willing, these men were selected into this role through consultation with VESA members and kebele administration leaders. One of the key responsibilities of the male advocators was to support the MGF with the co-facilitation of the men's engagement group, with two male advocators per group. With the goal of establishing 7-8 men's engagement groups in a kebele, between 14-16 male advocators were selected per kebele, and provided with a 3-day orientation and training on the men's engagement methodology and lessons. Afterwards, men's engagement groups were formed, with 20 men in each group, including the male advocators. As PSNP client households only represent the poorest 10-15 percent of households in a community, SPIR intentionally broadened the scope of the men's engagement intervention to include both PSNP and non-PSNP men in an effort to engage the broader community, and included local elders, religious leaders or kebele officials. These groups typically met twice per month. According to the endline survey, about 40 percent of men in N* kebeles reported attending at least one men's engagement session in the last 12 months.[4] Following the completion of the group sessions, public events were planned to promote these behaviors through testimonials and practical demonstration of men cooking the most common local staple food, *injera*. The initiation of COVID-19 related restrictions on public gatherings coincided with the intended timing of these events, and thus prevented most from taking place as originally planned.

2.3 Experimental design

The study was designed as a cluster RCT to test the impact of multisectoral graduation models that differ in the intensity of livelihood and nutrition interventions. The RCT included 192 kebeles, which were stratified by districts (woredas), and randomized to one of the following four intervention arms[5]:

[4] As 7-8 groups of 20 men were formed in each kebele, and on average 50-60% were from PSNP households, there would be an expected 70-100 men from PSNP HHs invited to participate. With an average number of 200-300 PSNP households in a study kebele, 40 percent is generally in the range of the average participation rate we expected.

[5] The study originally included 196 kebeles, but 2 were dropped for not having PSNP clients and 2 for security reasons.

Treatment 1 (T1):	L* + N*
Treatment 2 (T2):	L* + N
Treatment 3 (T3):	L + N*
Control (T4):	PSNP only

Table 1 shows how the livelihood and nutrition activities map to the intervention arms. For this paper, we are interested in the impact of the three treatment arms (T1, T2, T3) compared to the control group on men's gender equitable attitudes and behavior. All three treatment arms received PSNP and the core nutrition and livelihood interventions which included the VESA gender dialogue. In addition, we are interested in the added impact of N* activities in T1 and T3, which include the men's engagement groups. While most livelihood and nutrition activities started before the midline survey, a few enhanced nutrition (N*) activities (mainly the men's engagement groups and IPT-G) started after the midline survey.

In addition, due to budget considerations, men's engagement groups within N* were randomly rolled out, with MGFs forming men's engagement groups in half of the N* kebeles in December 2019 and with a plan to shift to the second half of N* kebeles in May 2020. COVID-19 restrictions on social gatherings disrupted the start of these sessions, with about one third of the groups starting in June 2020 and the rest starting closer to the end of the year. As a result, many of the men's engagement groups formed in the second half of N* kebeles were still in the process of completing their sessions when the endline survey was conducted between February and April 2021.

3. Data and empirical strategy

3.1 Data collection

The SPIR impact evaluation included three rounds of data collection; the baseline survey was conducted from February to April 2018, the midline survey from July to October 2019, and the endline survey from February to April 2021. The sampling frame included 192 study kebeles, from which households were randomly selected based on the following inclusion criteria (1) be a PSNP client household, (2) have at least one child aged 0-35 months (index child), and (3) have the mother or primary female caregiver of the 0-35-month-old child be a member of the household. The last two criteria enabled measurement of maternal and child nutrition, a primary outcome in the overall evaluation. In total 3,314 households were interviewed at baseline, of which 3,220 households were re-interviewed at midline. In addition, 748 supplemental households were added to the midline sample, for a total of 3,968 households at midline. These supplemental households were sampled based on the same criterion as above, with the exception that the index child must be under the age of 2 years. The endline survey sought to re-interview all households in the midline sample as well as any baseline household that was not available at midline but deemed likely to be available at endline.[6] Of the 3,996 households in the endline target sample, 3,812 were located and interviewed.

In all rounds, the questionnaire was structured in three parts: a brief household-level questionnaire for identification and household demographics, a male respondent questionnaire and a female

[6] Those deemed unlikely were households that permanently moved away or experienced the death of the primary respondent(s).

respondent questionnaire. The female questionnaire was administered to the mother or primary female caregiver of the index child ("primary female") and the male questionnaire was administered to her husband or partner ("primary male"). In households with a single adult female and no adult male, some of the modules for the primary male respondent were skipped. Each survey respondent was surveyed separately and provided informed oral consent.

Outcomes of interest

Our main outcomes of interest are men's involvement in domestic chores and men's gender equitable attitudes. Questions related to men's involvement in household domestic activities were asked of both the primary male and female. Men were asked about whether they helped with the following three activities in the last three days: household chores, meal preparation and cooking, and collecting firewood and water. Women were asked to report their spouse's involvement for the same three activities. We create three binary indicators that equal one if the primary male reports being involved in the specific activity. Given that men's responses may be subject to social desirability bias, we also create three binary indicators that equal one if the primary female reports her spouse was engaged in the specific activity.

We also construct three different indicators of equitable gender attitudes. The first is from a list of four questions asked to the primary male on whether a husband is justified in beating his wife under different circumstances: going out without telling him, burning the food, neglecting the children, and arguing with him. We create a binary indicator that equals one if the respondent says a husband is not justified in beating his wife under any of the four circumstances. The second is from a list of three questions on whether it is acceptable for a woman to travel alone to different places: market, friends/family, health center. We create a binary indicator that equals one if the primary male respondent says it is acceptable for a woman to travel alone to the three specified places. Lastly, we ask the primary male whether they agree with five gender inequitable statements (such as whether "Changing diapers, giving a bath, and feeding kids is only the mother's responsibility" or "A woman should obey her husband in all things") where the responses range from strongly disagree (0) to strongly agree (4).[7] Scores are reversely coded and summed up so that a higher score implies more equitable gender attitudes, ranging from 0-20.

In addition to the individual indicators on men's self-reported engagement in household chores and men's gender equitable attitudes, we create summary indices for the two families of outcomes following Kling et al. (2007). Analyzing summary indices has several advantages including reducing the probability of a false discovery by reducing the number of distinct hypothesis tests conducted, and providing more power to detect impacts as it can reduce the random error from each individual indicator (Anderson, 2008).

Sample, attrition and balance

The sample we use in the analysis are households with a primary male as the main outcomes are specific to men. For the main analysis we use the unbalanced panel sample that includes all primary males available at each time period, as this gives us the most power to detect impacts. However, we also conduct the analysis on the balanced sample of primary males that were available at midline and endline as a robustness check. Figure 1 illustrates the sample of primary males surveyed across all three rounds of data collection. At baseline there are 2,813 primary males, of which 2,398 are re-interviewed at midline in addition to 684 primary males that were added to the sample (600 from the supplemental sample and 84 from households without a

[7] The midline questionnaire includes 10 statements, but for consistency, we use only the 5 statements that were also used at endline.

primary male at baseline) for a midline sample of 3,082 primary males. Of the 3,082 primary males at midline, 2,235 were resurveyed at endline, plus 127 primary males from baseline that were not available at midline, for a total sample of 2,362.

While attrition of households in our sample is low (only 2.8 percent of baseline households attrited by midline and only 4 percent of households attrited from midline to endline), attrition of the primary male in the household is higher; 14.8 percent of primary males attrited from baseline to midline, and 27.5 percent attrited from midline to endline. According to field reports the main reason men were not surveyed at endline was that they were absent due to work. Appendix Table 1 reports attrition of the primary male at different time periods and whether it is correlated with the treatment status of households. We find that those in the treatment arms are less likely than the control arm to attrit at endline, however, across treatment arms attrition does not differ. We test whether differential attrition in baseline characteristics across the control arm and treatment arms threatens the internal validity of our study. We find little evidence of selective attrition from baseline to midline (Appendix Table 2) or baseline to endline (Appendix Table 3). Moreover, the first three columns of Appendix Tables 2 and 3 reveal that baseline characteristics are balanced for those that remain in the sample. The main exception is that primary males in T3 are less likely than the control arm to find violence not justified in any of the stated situations.

Estimation

To investigate the impact of SPIR and the added value of N* on male's gender equitable attitudes and behaviors, we use the midline and endline data respectively and estimate the following intent-to-treat (ITT) specification:

$$Y_{ikw}^{t} = \beta_0^t + \beta_1^t T1_{kw} + \beta_2^t T2_{kw} + \beta_3^t T3_{kw} + \beta_4^t X_{ikw}^0 + \mu_w + \varepsilon_{ikw,}$$

where Y_{ikw}^{t} is the outcome of interest for individual *i*, from kebele *k*, and woreda *w* at time *t* (which is either midline or endline). $T1_{kw}$ is an indicator for whether kebele *k* was randomly assigned to treatment T1 which includes L*and N* activities, $T2_{kw}$ indicates the randomized assignment to T2 which includes L* and N activities, and $T3_{kw}$ indicates the randomized assignment to T3 which includes L and N* activities. We control for design features of the study including a vector of woreda fixed effects, μ_w, which is the level of stratification used in the randomization, and a vector of indicators, X_{ikw}^0, for whether the household is part of the supplemental sample, whether it is the poorest 60 percent eligible for livelihood transfers as part of L*, and whether it was included in the first or second phase of the men's engagement group rollout (endline only specifications). β_1^t, β_2^t and β_3^t provide ITT estimates of the impact of T1, T2, and T3 respectively in each time period. We test for the equality of coefficients to see if impacts vary by treatment arm. At endline we are particularly interested in whether treatment arms with N*(β_1^t and β_3^t) are significantly different from treatment arms with only N (β_2^t). We also test for equality of coefficients across rounds, to see if impacts vary across rounds.

All estimates cluster standard errors at the kebele level. In addition, for specifications on individual indicators, we correct for multiple hypothesis testing using an approach introduced by (Simes, 1986), which adjusts inference for the multiplicity of tests estimated, resulting in a modified measure of statistical significance, the sharpened q-value. We report significance from both the conventional p-value (as stars in the tables) and the sharpened q-value (in brackets), while clustered standard errors are reported in parentheses.

4. Results

4.1 Main impacts

We first present the impacts at midline and endline on the summary index for men's engagement in domestic tasks. Figure 2 reveals that at midline all three treatment arms improve men's engagement in domestic tasks by 0.12-0.18 standard deviations (SD). Impacts do not vary across treatment arms at midline, suggesting that the core SPIR activities which include the VESA gender discussions improved men's self-reported behaviors. By endline, only T1 (which includes L* and N*) and T3 (which includes L and N*) still have positive and significant impacts on men's engagement in domestic tasks that range from 0.26 to 0.29 SD. The magnitudes of impact of T1 and T3 are similar, while the magnitudes of impact of T1 compared to T2 (which includes L* and N), and T3 compared to T2 are significantly different at endline, suggesting that N* is leading to larger impacts compared to the treatment arm without N*. For T1, impacts are significantly larger at endline compared to midline.

We next look at the individual indicators that make up the summary index for men's engagement in domestic tasks. Table 2 presents the impact estimates on men's self-reported engagement in household chores, cooking/meal preparation, and collecting water/firewood at midline (columns 1-3) and endline (columns 4-6), and reports tests of equality across rounds (columns 7-9). At midline all three treatment arms lead to significant improvements in men reporting that they helped cook and collect firewood/water. Point estimates range from 4.6 percentage points to 9.7 percentage points. At endline impacts for T1 and T3 remain significant for all three indicators, while for T2 impacts remain significant only for the indicator on collecting firewood and water. Impacts of T1 and T3 are significantly different from T2 at endline. For T1, impacts at endline are significantly larger than midline across all three indicators, consistent with the pattern on the summary index. Increased point estimates for the T1 and T3 arms at endline range from 12.7 to 20.8 percentage point increase over control means of 22.7 to 55.1 percent.

Figure 3 reveals a similar pattern for the summary index on gender equitable attitudes; at midline all three treatment arms significantly improve men's gender equitable attitudes by 0.08-0.09 SD, again suggesting that the core SPIR activities improved men's gender equitable attitudes. By endline, only T1 continues to have a significant impact. Impacts of T3 are positive but not significant at endline. Importantly, impacts of T1 compared to T2 and T3 compared to T2 are significantly different at endline, again suggesting that N* component is leading to larger impacts compared to N. For T2, impacts are significantly smaller at endline compared to midline, suggesting that without N* impacts found at midline are not sustained.

Table 3 presents impact estimates on the individual indicators that make the summary index for men's gender equitable attitudes. At midline, T1 and T2 lead to significant improvements on the total score for men's gender equitable attitudes (column 3). Impacts range from 0.56 to 0.91 points, which represents an increase of 6-11 percent compared to the mean of the control arm. Impacts on the other two indicators (whether a husband is not justified in beating his wife and whether it is acceptable for women to travel alone) are positive but not significant. Similar to the results on domestic tasks, impacts do not vary by treatment arm at midline. By endline the impact on the total score of gender equitable attitudes is sustained only for T1, which is significantly different from that of T2 (column 6). For T2, the impact on the total score of gender equitable attitudes is significantly smaller at endline compared to midline (column 9), consistent with the pattern on the summary index.

The experimental variation in the timing of the rollout of the men's engagement groups did generally not have any differential effects on these outcomes (Appendix Table 4). While not precisely measured, the interaction of the pooled T1 and T3 treatment (N*) with the assignment to the later rollout group has negative-signed point estimates for most outcomes, suggesting that effects from the men's engagement groups that were completed a year prior could be similar or even slightly larger than the effects measured by groups that were either just finishing or still meeting during the endline survey.

4.2 Robustness

For both sets of outcomes, we find similar patterns; at midline all treatment arms improve men's gender equitable attitudes and participation in domestic tasks, but impacts are only sustained and/or improved at endline for the treatment arms with N*. The main difference in the N* component between midline and endline is the men's engagement groups which were rolled out after the midline survey. Below we show that results are robust to alternative hypotheses.

One potential alternative reason for the difference in impact estimates between midline and endline is that the sample of men changes across rounds. To ensure our results are not driven by the different sample of men, we run the same analysis for the sample of men who were included in both the midline and endline rounds. Appendix Table 5 and 6, reveal that our results are robust to this panel sample.

Another potential explanation is that men's behavior did not actually change, and instead their responses were subject to social desirability bias. If this was more likely to occur in the treatment groups with N*, then this could bias our estimates. To investigate whether social desirability bias is likely to be an issue, we estimate impacts on women's reports of their partners' behavior. Appendix Table 7 reveals that at endline, impacts from women's reports of their husbands' behavior are similar to impacts from men's self-reports, and in particular the pattern of large impacts across T1 and T3, but not T2, holds.

Lastly, from the experimental design, the other component of N* that was implemented between midline and endline was the IPT-G intervention. Women screened at midline who were found to have mild-to-severe depression were invited to join IPT-G sessions. Although IPT-G was targeted to women and not men, women's improved mental health could affect men's involvement in household tasks. Given that only women with depression were screened into these groups, the percent of women invited to participate in IPT-G in N* communities is low, at 18 percent, so this is unlikely to be driving our results. Regardless, we run the same analysis on the sample of men whose partners were not screened into IPT-G. Appendix tables 8 and 9 reveal that results are robust to excluding households with women who were eligible to participate in IPT-G.

4.3 Other outcomes

While involving men in childcare activities was an integral part of the curriculum in both the VESA gender dialogues and the men's engagement groups, we do not include men's childcare activities in the main analysis as questions regarding childcare were only asked on the subsample of men in each time period who had a child less than 24 months, which is only about a quarter of households by endline. However, appendix table 10 reveals that the magnitude of impact on men's childcare activities is generally small and close to zero suggesting that regardless of the smaller sample, there are no impacts on these indicators. Thus, while SPIR was able to change men's involvement in chores and their attitudes about gender equitable norms, it does not appear to have led to

greater participation in childcare activities. One potential reason for why SPIR was not able to change men's behavior with respect to childcare, is that most men were already engaging in childcare activities (80 percent of men in the control group at midline and 98 percent at endline took part in at least one childcare activity in the last 3 days), which suggests there was just less room for impact.

5. Conclusion

In this paper we investigate the impact on men's gender equitable attitudes and behaviors of a graduation model program that seeks to simultaneously "push" households out of poverty and improve gender equitable norms. We find that at midline all treatment arms lead to improvements in men's gender equitable attitudes and their involvement in domestic tasks. Unfortunately, we cannot disentangle what component or activities lead to the improvements but given that impacts at midline do not vary by arm, we conclude that it is likely not the L* and N* components but instead, the core N and L components that were similar across all arms. By endline, only the treatment arms with the men's engagement groups continue to have significant impacts which are large and range from 0.26-0.29 SD for men's reported involvement in household tasks. Impacts at endline on gender equitable attitudes are smaller in magnitude and range from 0.07-0.14 SD. Across both sets of outcomes, the magnitudes of impact in the treatment arms receiving the men's engagement intervention (T1 and T3) are significantly different than that of the treatment arm that did not receive this intervention (T2). Thus, we attribute the sustained or larger impacts at endline in T1 and T3 to the men's engagement groups which were rolled out after the midline survey. Results are robust to using women's report of their husband's behaviors, a balanced sample, and a smaller sample that excludes households where women were also treated for depression.

The pattern of results in the T2 treatment arm mirrors that of Banerjee et al. (2015), where a 0.046 SD treatment effect on an index of women's empowerment in their first endline had dissipated by the second endline one year later. Due to the experimental variation in the rollout of the men's engagement groups after the midline in our study, and the disruption in the start of the second round of men's engagement groups due to COVID-19 restrictions, we can compare treatment effects in clusters where men's engagement groups where completed a year before the endline survey with those that had either just completed or were in the process of completing their sessions in the second half of N* kebeles. We find no differential effects, indicating that the overall large effects on men's behaviors were not driven by the group that just finished the sessions, and that effects persisted for at least one year after the completion of the sessions in the first half of N* kebeles.

Changing attitudes and behaviors around gender norms faces many challenges, especially in rural settings where traditions are deeply rooted in customs that have persisted for many generations. It is not only men facing ridicule from other men for engaging in these perceived 'women-only' household tasks, but even other women, especially the mother-in-law, that are strongly against men's participation in domestic activities such as cooking. This aberration is often intuited as an insult or direct rejection of his wife and her capabilities. We find that challenging traditional gender norms can be motivated and sustained through a peer group where men encouraged and supported each other to put into practice these counter-cultural behaviors. Anecdotal evidence reveals that the novelty of observing men baking the traditional staple food of *injera* was one of the most noteworthy signs that traditional gender norms could be challenged in rural settings. Beyond the timeline of this study, the impact could be even greater if these more equitable gender attitudes and norms are modeled and transmitted to children in these households. Longer term

follow-up studies of these men's engagement strategies should be prioritized to determine if these initially promising effects persist over time and intergenerationally.

References

Abramsky, T., Devries, K., Kiss, L., Nakuti, J., Kyegombe, N., Starmann, E., Cundill, B., Francisco, L., Kaye, D., Musuya, T., Michau, L., & Watts, C. (2014). Findings from the SASA! Study: A cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda. *BMC Medicine*, *12*(1), 1–17. https://doi.org/10.1186/S12916-014-0122-5/TABLES/5

Alderman, H., Billings, L., Gilligan, D. O., Hidrobo, M., Leight, J., Ramani, G., Taffesse, A. S., & Tambet, H. (2020). *Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA)*: Midline report.

Alderman, H., Billings, L., Gilligan, D. O., Hidrobo, M., Leight, J., Taffesse, A. S., & Tambet, H. (2021). Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA): Endline Report.

Alemu, A., 2007. Oral narrative as ideological weapon for subordinating women: the case of Jimma Oromo*. J. African Cult. Stud. 19 (1), 55–80.

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics*, *128*(2), 469–530. https://doi.org/10.1093/qje/qjt005

Angelucci, M., Heath, R., & Noble, E. (2023). Multifaceted programs targeting women in fragile settings: Evidence from the Democratic Republic of Congo. *Journal of Development Economics*, *164*, 103146. https://doi.org/10.1016/j.jdeveco.2023.103146

Anderson, Michael 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: 1481–95. https://doi.org/10.1198/016214508000000841.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., & Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, *348*(6236). https://doi.org/10.1126/science.1260799

Bedoya, G., Coville, A., Haushofer, J., Isaqzadeh, M., & Shapiro, J. (2019). No Household Left Behind : Afghanistan Targeting the Ultra Poor Impact Evaluation. In *Policy Research Working Paper* (Issue No. 8877). http://hdl.handle.net/10986/31867

Bossuroy, T., Goldstein, M., Karimou, B., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C. C., Udry, C., Vaillant, J., & Wright, K. A. (2022). Tackling psychosocial and capital constraints to alleviate poverty. *Nature*, *605*(7909), 291–297. https://doi.org/10.1038/s41586-022-04647-8

Central Statistical Agency. (2014). *Ethiopia Time Use Survey 2013*. www.csa.gov.et

Central Statistical Agency and ICF. (2016). *Ethiopia Demographic and Health Survey, 2016*. Addis Ababa, Ethiopia, and Rockville, Maryland, USA.

Christofides, N. J., Hatcher, A. M., Rebombo, D., McBride, R. S., Munshi, S., Pino, A., Abdelatif, N., Peacock, D., Levin, J., & Jewkes, R. K. (2020). Effectiveness of a multi-level intervention to

reduce men's perpetration of intimate partner violence: A cluster randomised controlled trial. *Trials*, *21*(1). https://doi.org/10.1186/s13063-020-4185-7

Couprie, H., Cudeville, E., & Sofer, C. (2020). Efficiency versus gender roles and stereotypes: an experiment in domestic production. *Experimental Economics*, *23*(1), 181–211. https://doi.org/10.1007/s10683-019-09612-3

Dhar, D., Jain, T., & Jayachandran, S. (2022). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review*, *112*(3), 899–927. https://doi.org/10.1257/AER.20201112

Doyle, K., Levtov, R. G., Barker, G., Bastian, G. G., Bingenheimer, J. B., Kazimbaya, S., Nzabonimpa, A., Pulerwitz, J., Sayinzoga, F., Sharma, V., & Shattuck, D. (2018). Gender-transformative bandebereho couples' intervention to promote male engagement in reproductive and maternal health and violence prevention in Rwanda: Findings from a randomized controlled trial. *PLoS ONE*, *13*(4). https://doi.org/10.1371/journal.pone.0192756

Duflo, E., & Udry, C. (2004). *Intrahousehold Resource Allocation in Côte d'Ivoire: Social Norms, Separate Accounts and Consumption Choices*. http://www.nber.org/papers/w10498

Glinski, A., Schwenke, C., O'Brien-Milne, L., & Farley, K. (2018). *Gender Equity and Male Engagement: It only works when everyone plays*.

Hossain, M., Zimmerman, C., Kiss, L., Abramsky, T., Kone, D., Bakayoko-Topolska, M., Annan, J., Lehmann, H., & Watts, C. (2014). Working with men to prevent intimate partner violence in a conflict-affected setting: a pilot cluster randomized controlled trial in rural Côte d'Ivoire. *BMC Public Health*, *14*(1), 339. https://doi.org/10.1186/1471-2458-14-339

Jayachandran, S. (2021). Social Norms as a Barrier to Women' s Employment in Developing Countries. *IMF Economic Review*, *69*, 576–595.

Jewkes, R., Nduna, M., Levin, J., Jama, N., Dunkle, K., Puren, A., & Duvvury, N. (2008). Impact of Stepping Stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: cluster randomised controlled trial. *BMJ*, *337*(aug07 1), a506–a506. https://doi.org/10.1136/bmj.a506

Kling, J., Liebman, J., & Katz, L. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119.

Mabsout R, and van Staveren I. 2010. "Disentangling Bargaining Power from Individual and Household Level to Institutions: Evidence on Women's Position in Ethiopia." World Development 38 (5): 783–96.

Mccullough, E., Mcgavock, T., Magnan, N., Assefa, T., Mcgavock, T. J., & Assefa, T. W. (2022). *Experimental Effects of Updating Husbands' Beliefs on Participation in Women's Household Work*.

Nguyen, C. V., & Tarp, F. (2022). Changing male perceptions of gender equality: Evidence from a randomised controlled trial study. *World Development*, *158*. https://doi.org/10.1016/j.worlddev.2022.106019

Raj, A., Ghule, M., Ritter, J., Battala, M., Gajanan, V., Nair, S., Dasgupta, A., Silverman, J. G., Balaiah, D., & Saggurti, N. (2016). Cluster Randomized Controlled Trial Evaluation of a Gender Equity and Family Planning Intervention for Married Men and Couples in Rural India. *PLOS ONE*, *11*(5), e0153190. https://doi.org/10.1371/journal.pone.0153190

Ranganathan, M., Pichon, M., Hidrobo, M., Tambet, H., Sintayehu, W., Tadesse, S., & Buller, A. M. (2022). Government of Ethiopia's public works and complementary programmes: a mixed-methods study on pathways to reduce intimate partner violence. Social Science & Medicine (1982), 294, 114708–114708.

Sharma, V., Leight, J., Verani, F., Tewolde, S., & Deyessa, N. (2020). Effectiveness of a culturally appropriate intervention to prevent intimate partner violence and HIV transmission among men, women, and couples in rural Ethiopia: Findings from a cluster-randomized controlled trial. *PLoS Medicine*, *17*(8). https://doi.org/10.1371/JOURNAL.PMED.1003274

Shattuck, D., Kerner, B., Gilles, K., Hartmann, M., Ng'ombe, T., & Guest, G. (2011). Encouraging contraceptive uptake by motivating men to communicate about family planning: The Malawi Male Motivator project. *American Journal of Public Health*, *101*(6), 1089–1095. https://doi.org/10.2105/AJPH.2010.300091

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, *73*(3), 751–754. https://doi.org/10.1093/biomet/73.3.751

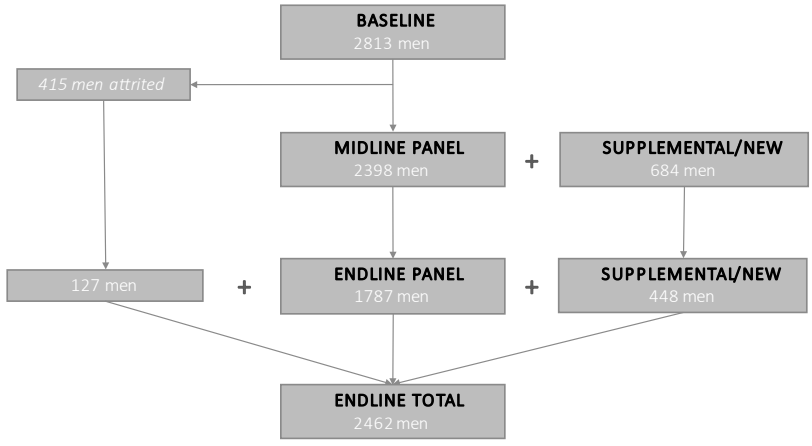
Tefera, N. (2020). *Gender-Differentiated Time Allocations on Daily Activities in Rural Ethiopia*.

Udry, C. (1996). Gender, Agricultural Production, and the Theory of the Household. In *Source: Journal of Political Economy* (Vol. 104, Issue 5). https://about.jstor.org/terms

Vaillant, J., Koussoubé, E., Roth, D., Pierotti, R., Hossain, M., & Falb, K. L. (2020). Engaging men to transform inequitable gender attitudes and prevent intimate partner violence: a cluster randomised controlled trial in North and South Kivu, Democratic Republic of Congo. *BMJ Global Health*, *5*(5). https://doi.org/10.1136/bmjgh-2019-002223

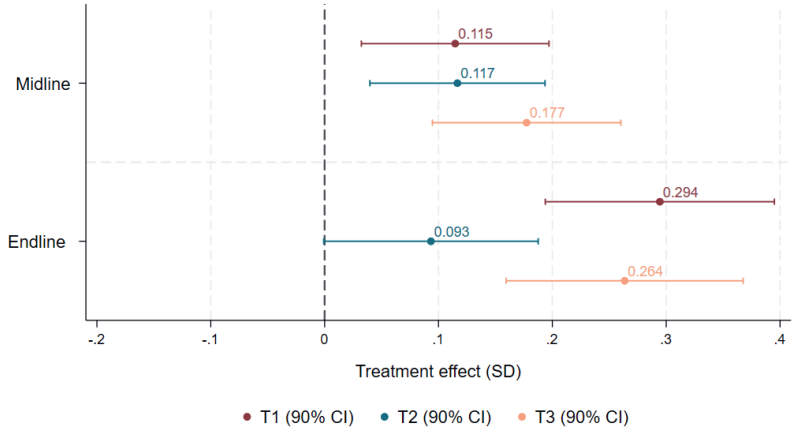
Figures

Figure 1: Flow chart of primary males in study



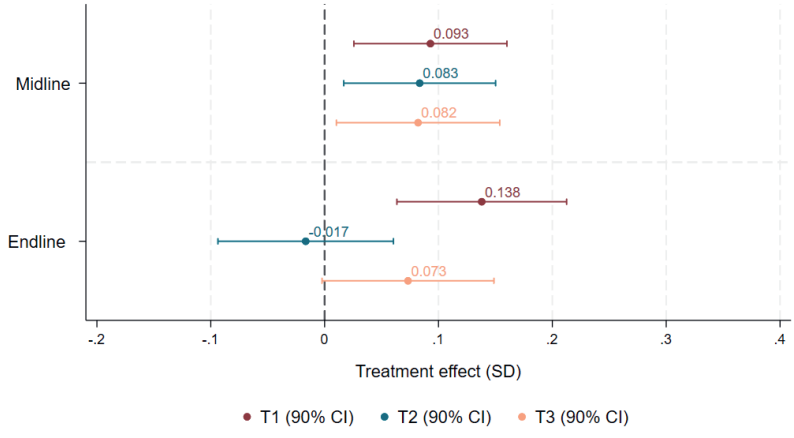
*Note: Supplemental and new men refers to primary males from the supplemental sample and primary males who were not surveyed at baseline but whose household was surveyed at baseline.

Figure 2: Impacts on summary index for men's involvement in domestic tasks



Notes: The model specifications control for the woreda-level fixed effects and the binary indicators of being eligible for the livelihood transfers and being in the supplemental sample. Endline specification also controls for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. Standard errors are clustered at the kebele level.

Figure 3: Impacts on summary index for men's gender equitable attitudes



Notes: The model specifications control for the woreda-level fixed effects and the binary indicators of being eligible for the livelihood transfers and being in the supplemental sample. Endline specification also controls for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. Standard errors are clustered at the kebele level.

Tables

Table 1: Study intervention arms and livelihood and nutrition components

Intervention description	T1: L*+N*	T2: L*+N	T3: L+N*	T4: PSNP only
PSNP consumption-support transfers	✓	✓	✓	✓
L: VESA+ gender dialogue +trainings	✓	✓	✓	
N: Nutrition BCC + WASH	✓	✓	✓	
L*: + livelihood transfers	✓	✓		
N*: + TTC + CPNP + men's engagement + IPT-G	✓		✓	

* VESA is Village Economic and Social Associations. BCC is behavior change and communication. WASH is water, sanitation, and hygiene. TTC is time and targeted nutrition counseling. CPNP is community-based participatory nutrition promotion, IPT-G is interpersonal psychotherapy in groups.

Table 2: Primary male's involvement in domestic tasks in the past 3 days

	Midline			Endline			Midline vs endline		
	(1) Helped with household chores	(2) Helped with cooking or meal preparation	(3) Helped with collecting firewood and water	(4) Helped with household chores	(5) Helped with cooking or meal preparation	(6) Helped with collecting firewood and water	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.046	0.049*	0.066**	0.151***	0.127***	0.208***	0.019	0.020	0.001
	(0.028)	(0.026)	(0.033)	(0.046)	(0.035)	(0.035)			
	[0.103]	[0.079]	[0.065]	[0.003]	[0.001]	[0.000]			
T2	0.046	0.054**	0.062**	0.030	0.003	0.104***	0.643	0.113	0.306
	(0.028)	(0.024)	(0.031)	(0.034)	(0.027)	(0.033)			
	[0.163]	[0.076]	[0.084]	[0.464]	[0.923]	[0.011]			
T3	0.069**	0.097***	0.079**	0.146***	0.138***	0.144***	0.048	0.236	0.108
	(0.028)	(0.024)	(0.036)	(0.038)	(0.029)	(0.035)			
	[0.018]	[0.000]	[0.027]	[0.000]	[0.000]	[0.000]			
Test: T1 = T2	(0.992)	(0.845)	(0.911)	(0.007)	(0.000)	(0.001)			
	[0.992]	[0.992]	[0.992]	[0.014]	[0.002]	[0.002]			
Test: T2 = T3	(0.439)	(0.091)	(0.637)	(0.001)	(0.000)	(0.188)			
	[0.527]	[0.182]	[0.637]	[0.004]	[0.000]	[0.282]			
Test: T1 = T3	(0.414)	(0.078)	(0.723)	(0.902)	(0.706)	(0.039)			
	[0.828]	[0.234]	[0.867]	[0.902]	[0.867]	[0.234]			
Mean of control	0.342	0.233	0.506	0.294	0.227	0.551			
N	3,067	3,067	3,068	2,452	2,450	2,451			

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 3: Primary Male's gender equitable attitudes

	Midline			Endline			Midline vs endline		
	(1) A husband is not justified in beating his wife in any of these situations [1]	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(3) Index of support for equitable gender norms	(4) A husband is not justified in beating his wife in any of these situations [1]	(5) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(6) Index of support for equitable gender norms	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.034	0.034	0.914***	0.057*	0.051	0.800**	0.565	0.689	0.794
	(0.033)	(0.028)	(0.319)	(0.033)	(0.043)	(0.380)			
	[0.306]	[0.284]	[0.028]	[0.181]	[0.284]	[0.109]			
T2	0.043	0.043	0.557*	-0.005	0.008	-0.252	0.156	0.343	0.042
	(0.030)	(0.027)	(0.336)	(0.032)	(0.035)	(0.347)			
	[0.312]	[0.312]	[0.312]	[0.876]	[0.876]	[0.703]			
T3	0.044	0.028	0.520	0.051	0.017	0.380	0.847	0.782	0.731
	(0.031)	(0.029)	(0.336)	(0.033)	(0.041)	(0.339)			
	[0.309]	[0.390]	[0.309]	[0.309]	[0.680]	[0.390]			
Test: T1 = T2	(0.778)	(0.730)	(0.245)	(0.047)	(0.258)	(0.002)			
	[0.778]	[0.778]	[0.386]	[0.142]	[0.386]	[0.009]			
Test: T2 = T3	(0.960)	(0.579)	(0.909)	(0.057)	(0.802)	(0.026)			
	[0.960]	[0.960]	[0.960]	[0.171]	[0.960]	[0.157]			
Test: T1 = T3	(0.749)	(0.845)	(0.201)	(0.852)	(0.327)	(0.098)			
	[0.852]	[0.852]	[0.604]	[0.852]	[0.654]	[0.590]			
Mean of control	0.602	0.356	8.696	0.666	0.426	10.333			
N	3,070	3,076	3,079	2,450	2,458	2,460			

[1] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband.

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Tables

Appendix Table 1: Attrition rates by treatment arm of primary male respondents

	(1) Primary male respondent attrited baseline to midline	(2) Primary male respondent attrited baseline to endline	(3) Primary male respondent attrited midline to endline
T1	-0.002	-0.060**	-0.048*
	(0.021)	(0.030)	(0.027)
T2	-0.021	-0.068**	-0.058**
	(0.021)	(0.032)	(0.028)
T3	-0.003	-0.056*	-0.034
	(0.020)	(0.032)	(0.029)
Test: T1 = T2	0.358	0.777	0.692
Test: T2 = T3	0.336	0.685	0.379
Test: T1 = T3	0.978	0.885	0.609
Mean of control	0.132	0.313	0.264
N	2,813	2,813	3,082

Notes: Standard errors (in parentheses) are clustered at the kebele level. Models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 2: Selective attrition by baseline characteristics
Attrition from baseline to midline

	(1) N	(2) T1	(3) T2	(4) T3	(5) T1 x Attrition	(6) T2 x Attrition	(7) T3 x Attrition
Household size	2,916	0.211*	0.064	0.023	0.754	0.050	0.240
		(0.127)	(0.116)	(0.126)	(0.486)	(0.587)	(0.530)
Number of children under the age of 5	2,916	0.010	-0.010	0.004	0.074	0.240	0.127
		(0.032)	(0.032)	(0.034)	(0.139)	(0.210)	(0.151)
Log of total monthly consumption expenditure per adult equivalent	2,888	-0.003	-0.006	0.031	-0.242	-0.028	-0.127
		(0.057)	(0.060)	(0.062)	(0.198)	(0.331)	(0.204)
Age of primary male	2,813	0.582	0.229	0.198	7.144**	1.252	5.020
		(0.571)	(0.533)	(0.491)	(3.085)	(3.309)	(3.978)
Primary male is the household head	2,813	-0.021	0.006	0.010	0.225**	0.098	0.138
		(0.025)	(0.020)	(0.018)	(0.112)	(0.130)	(0.116)
Primary male has some education	2,812	0.006	-0.001	0.010	-0.152	0.248	-0.027
		(0.030)	(0.034)	(0.031)	(0.154)	(0.188)	(0.156)
Primary male's main activity is crop production	2,811	-0.000	0.012	0.030	0.019	0.050	-0.115
		(0.037)	(0.034)	(0.036)	(0.141)	(0.134)	(0.143)
Primary male is married and lives with their spouse	2,800	0.002	-0.012	0.014	0.050	-0.072	0.049
		(0.020)	(0.021)	(0.018)	(0.104)	(0.133)	(0.101)
Primary male finds it acceptable for a woman to travel alone to market, health center, or visit a friend	2,646	0.031	0.030	0.036	0.006	0.180	0.205
		(0.040)	(0.041)	(0.042)	(0.160)	(0.174)	(0.150)
Primary male doesn't find violence justified in any situation asked about	2,646	-0.040	-0.034	-0.064**	0.037	0.028	0.101
		(0.030)	(0.029)	(0.028)	(0.131)	(0.128)	(0.131)
Total number of childcare activities primary male participated in (0-10)	2,800	-0.060	0.025	0.060	1.273	1.041	0.278
		(0.157)	(0.173)	(0.161)	(1.070)	(0.961)	(0.740)

Notes: Standard errors (in parentheses) are clustered at the kebele level. Models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 3: Selective attrition by baseline characteristics
Attrition from baseline to endline

	(1) N	(2) T1	(3) T2	(4) T3	(5) T1 x Attrition	(6) T2 x Attrition	(7) T3 x Attrition
Household size	2,916	0.191	0.082	0.036	0.870**	-0.231	0.041
		(0.127)	(0.116)	(0.128)	(0.416)	(0.344)	(0.365)
Number of children under the age of 5	2,916	0.008	-0.000	0.011	0.071	-0.074	-0.038
		(0.032)	(0.032)	(0.035)	(0.132)	(0.125)	(0.107)
Log of total monthly consumption expenditure per adult equivalent	2,888	-0.005	-0.011	0.032	-0.046	0.073	-0.062
		(0.057)	(0.061)	(0.062)	(0.174)	(0.148)	(0.137)
Age of primary male	2,813	0.654	0.279	0.396	2.111	-0.318	-0.730
		(0.567)	(0.538)	(0.504)	(2.973)	(2.070)	(2.230)
Primary male is the household head	2,813	-0.016	0.006	0.011	-0.003	0.047	0.021
		(0.025)	(0.020)	(0.018)	(0.082)	(0.051)	(0.052)
Primary male has some education	2,812	0.001	-0.001	0.008	0.047	0.108	0.045
		(0.030)	(0.034)	(0.031)	(0.156)	(0.127)	(0.102)
Primary male's main activity is crop production	2,811	-0.003	0.012	0.027	0.080	0.042	0.013
		(0.038)	(0.035)	(0.036)	(0.111)	(0.107)	(0.099)
Primary male is married and lives with their spouse	2,800	0.006	-0.012	0.015	-0.069	-0.013	-0.004
		(0.020)	(0.022)	(0.018)	(0.062)	(0.051)	(0.043)
Primary male finds it acceptable for a woman to travel alone to market, health center, to visit a friend	2,646	0.036	0.022	0.036	-0.195	0.239*	0.105
		(0.040)	(0.042)	(0.043)	(0.125)	(0.140)	(0.132)
Primary male doesn't find violence justified in any situation asked about	2,646	-0.040	-0.036	-0.063**	0.018	0.062	0.030
		(0.030)	(0.029)	(0.029)	(0.105)	(0.088)	(0.107)
Total number of childcare activities primary male participated in (0-10)	2,800	-0.043	0.054	0.115	0.475	-0.213	-0.817
		(0.160)	(0.176)	(0.164)	(0.772)	(0.617)	(0.589)

Notes: Standard errors (in parentheses) are clustered at the kebele level. Models control for woreda level fixed effects. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 4: Male's involvement in tasks and gender equitable attitudes

	(1) Male reports that he helped with household chores (e.g. cleaning)	(2) Male reports that he helped with cooking or meal preparation	(3) Male reports that he helped with collecting firewood and water	(4) A husband is not justified in beating his wife in any of these situations [1]	(5) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(6) Gender equitable attitudes score (0-20)
Pooled T1 & T3	0.149***	0.133***	0.175***	0.054*	0.033	0.581*
	(0.038)	(0.028)	(0.032)	(0.030)	(0.039)	(0.338)
	[0.000]	[0.000]	[0.000]	[0.105]	[0.391]	[0.105]
Pooled T1 & T3 x Late treatment	-0.044	-0.040	-0.060**	-0.044	0.034	0.039
	(0.037)	(0.029)	(0.030)	(0.028)	(0.035)	(0.254)
	[0.358]	[0.334]	[0.289]	[0.334]	[0.406]	[0.878]
T2	0.030	0.003	0.105***	-0.005	0.008	-0.249
	(0.034)	(0.027)	(0.033)	(0.032)	(0.035)	(0.347)
	[0.925]	[0.925]	[0.011]	[0.925]	[0.925]	[0.925]
Mean of control (T4)	0.294	0.227	0.551	0.666	0.426	10.333
N	2,452	2,450	2,451	2,450	2,458	2,460

[1] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food.

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects and the baseline value of the outcome if the respective data was collected. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 5: Primary male's involvement in domestic tasks in the past 3 days
Midline to endline panel sample

	Midline			Endline			Midline vs endline		
	(1) Helped with household chores (e.g. cleaning)	(2) Helped with cooking or meal preparation	(3) Helped with collecting firewood and water	(4) Helped with household chores (e.g. cleaning)	(5) Helped with cooking or meal preparation	(6) Helped with collecting firewood and water	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.035	0.056*	0.083**	0.151***	0.123***	0.193***	0.010	0.053	0.017
	(0.034)	(0.031)	(0.036)	(0.046)	(0.035)	(0.034)			
	[0.299]	[0.083]	[0.033]	[0.003]	[0.001]	[0.000]			
T2	0.048	0.057**	0.077**	0.029	0.001	0.102***	0.606	0.095	0.554
	(0.035)	(0.027)	(0.034)	(0.034)	(0.028)	(0.033)			
	[0.250]	[0.076]	[0.070]	[0.467]	[0.982]	[0.014]			
T3	0.082**	0.095***	0.095**	0.154***	0.130***	0.125***	0.077	0.376	0.478
	(0.032)	(0.029)	(0.038)	(0.040)	(0.031)	(0.035)			
	[0.013]	[0.002]	[0.013]	[0.000]	[0.000]	[0.001]			
Test: T1 = T2	(0.715)	(0.977)	(0.870)	(0.006)	(0.000)	(0.003)			
	[0.977]	[0.977]	[0.977]	[0.011]	[0.002]	[0.010]			
Test: T2 = T3	(0.322)	(0.184)	(0.640)	(0.001)	(0.000)	(0.444)			
	[0.483]	[0.368]	[0.640]	[0.002]	[0.000]	[0.533]			
Test: T1 = T3	(0.152)	(0.223)	(0.761)	(0.941)	(0.809)	(0.024)			
	[0.447]	[0.447]	[0.941]	[0.941]	[0.941]	[0.144]			
Mean of control	0.342	0.227	0.487	0.292	0.227	0.558			
N	2,260	2,260	2,261	2,260	2,260	2,261			

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 6: Primary male's gender equitable attitudes
Midline to endline panel sample

	Midline			Endline			Midline vs endline		
	(1) A husband is not justified in beating his wife in any of these situations [1]	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(3) Index of support for equitable gender norms	(4) A husband is not justified in beating his wife in any of these situations [1]	(5) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(6) Index of support for equitable gender norms	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.041	0.048	0.713**	0.071**	0.055	0.921**	0.465	0.863	0.650
	(0.036)	(0.031)	(0.345)	(0.033)	(0.041)	(0.404)			
	[0.252]	[0.183]	[0.081]	[0.081]	[0.218]	[0.081]			
T2	0.039	0.035	0.426	-0.009	0.008	-0.289	0.169	0.468	0.073
	(0.036)	(0.032)	(0.352)	(0.032)	(0.035)	(0.363)			
	[0.554]	[0.554]	[0.554]	[0.817]	[0.817]	[0.639]			
T3	0.031	0.060*	0.302	0.055	0.014	0.440	0.545	0.239	0.733
	(0.037)	(0.033)	(0.360)	(0.033)	(0.042)	(0.361)			
	[0.484]	[0.301]	[0.484]	[0.301]	[0.742]	[0.450]			
Test: T1 = T2	(0.940)	(0.682)	(0.356)	(0.011)	(0.203)	(0.001)			
	[0.940]	[0.819]	[0.534]	[0.034]	[0.406]	[0.004]			
Test: T2 = T3	(0.820)	(0.432)	(0.707)	(0.037)	(0.881)	(0.017)			
	[0.881]	[0.865]	[0.881]	[0.110]	[0.881]	[0.102]			
Test: T1 = T3	(0.768)	(0.694)	(0.205)	(0.585)	(0.224)	(0.079)			
	[0.768]	[0.768]	[0.447]	[0.768]	[0.447]	[0.447]			
Mean of control	0.607	0.340	8.800	0.668	0.429	10.354			
N	2,260	2,270	2,275	2,260	2,270	2,275			

[1] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband.

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 7: Primary male's involvement in domestic tasks in the past 3 days as reported by primary female and primary male at endline

	Female's report			Male's report			Female's vs male's reports		
	(1) Spouse helped with household chores (e.g. cleaning)	(2) Spouse helped with cooking or meal preparation	(3) Spouse helped with collecting firewood and water	(4) Helped with household chores (e.g. cleaning)	(5) Helped with cooking or meal preparation	(6) Helped with collecting firewood and water	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.119***	0.104***	0.215***	0.151***	0.127***	0.208***	0.364	0.497	0.828
	(0.040)	(0.036)	(0.039)	(0.046)	(0.035)	(0.035)			
	[0.004]	[0.005]	[0.000]	[0.002]	[0.001]	[0.000]			
T2	-0.002	0.012	0.056*	0.030	0.003	0.104***	0.315	0.716	0.111
	(0.032)	(0.025)	(0.029)	(0.034)	(0.027)	(0.033)			
	[0.947]	[0.920]	[0.155]	[0.773]	[0.947]	[0.011]			
T3	0.090**	0.101***	0.166***	0.146***	0.138***	0.144***	0.073	0.233	0.500
	(0.036)	(0.030)	(0.037)	(0.038)	(0.029)	(0.035)			
	[0.013]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]			
Test: T1 = T2	(0.002)	(0.007)	(0.000)	(0.007)	(0.000)	(0.001)			
	[0.002]	[0.007]	[0.000]	[0.007]	[0.001]	[0.001]			
Test: T2 = T3	(0.006)	(0.001)	(0.002)	(0.001)	(0.000)	(0.188)			
	[0.008]	[0.002]	[0.002]	[0.002]	[0.000]	[0.188]			
Test: T1 = T3	(0.411)	(0.907)	(0.173)	(0.902)	(0.706)	(0.039)			
	[0.822]	[0.907]	[0.519]	[0.907]	[0.907]	[0.237]			
Mean of control	0.198	0.153	0.366	0.294	0.227	0.551			
N	3,070	3,070	3,069	2,452	2,450	2,451			

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 8: Primary Male's involvement in domestic tasks in the past 3 days
Excluding households where a female was eligible for IPT-G

	Midline			Endline			Midline vs endline		
	(1) Male reports that he helped with household chores (e.g. cleaning)	(2) Male reports that he helped with cooking or meal preparation	(3) Male reports that he helped with collecting firewood and water	(4) Male reports that he helped with household chores (e.g. cleaning)	(5) Male reports that he helped with cooking or meal preparation	(6) Male reports that he helped with collecting firewood and water	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.035	0.048	0.047	0.146***	0.137***	0.185***	0.009	0.014	0.003
	(0.030)	(0.030)	(0.035)	(0.044)	(0.035)	(0.035)			
	[0.246]	[0.166]	[0.215]	[0.002]	[0.000]	[0.000]			
T2	0.047	0.060**	0.059*	0.027	0.023	0.097***	0.567	0.290	0.385
	(0.031)	(0.026)	(0.033)	(0.036)	(0.029)	(0.034)			
	[0.197]	[0.062]	[0.163]	[0.460]	[0.460]	[0.027]			
T3	0.076**	0.103***	0.074*	0.131***	0.140***	0.125***	0.144	0.344	0.248
	(0.030)	(0.026)	(0.038)	(0.037)	(0.031)	(0.036)			
	[0.015]	[0.000]	[0.052]	[0.001]	[0.000]	[0.001]			
Test: T1 = T2	(0.691)	(0.677)	(0.744)	(0.006)	(0.001)	(0.005)			
	[0.744]	[0.744]	[0.744]	[0.011]	[0.006]	[0.011]			
Test: T2 = T3	(0.348)	(0.096)	(0.685)	(0.003)	(0.000)	(0.373)			
	[0.448]	[0.193]	[0.685]	[0.010]	[0.001]	[0.448]			
Test: T1 = T3	(0.169)	(0.062)	(0.493)	(0.686)	(0.917)	(0.052)			
	[0.338]	[0.186]	[0.740]	[0.823]	[0.917]	[0.186]			
Mean of control	0.339	0.238	0.521	0.312	0.228	0.579			
N	2,629	2,628	2,629	2,083	2,082	2,083			

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 9: Primary Male's gender equitable attitudes
Excluding households where a female was eligible for IPT-G

	Midline			Endline			Midline vs endline		
	(1) A husband is not justified in beating his wife in any of these situations [1]	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(3) Index of support for equitable gender norms	(4) A husband is not justified in beating his wife in any of these situations [1]	(5) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(6) Index of support for equitable gender norms	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.001	0.053*	0.805**	0.052	0.027	0.823**	0.248	0.556	0.967
	(0.033)	(0.030)	(0.347)	(0.036)	(0.044)	(0.380)			
	[0.965]	[0.163]	[0.094]	[0.226]	[0.659]	[0.094]			
T2	0.037	0.049*	0.454	0.006	0.001	-0.248	0.377	0.237	0.070
	(0.030)	(0.028)	(0.357)	(0.034)	(0.039)	(0.349)			
	[0.440]	[0.440]	[0.440]	[0.970]	[0.970]	[0.715]			
T3	0.032	0.028	0.286	0.070**	0.004	0.486	0.335	0.589	0.617
	(0.030)	(0.029)	(0.347)	(0.034)	(0.043)	(0.339)			
	[0.493]	[0.493]	[0.493]	[0.267]	[0.920]	[0.459]			
Test: T1 = T2	(0.289)	(0.894)	(0.309)	(0.168)	(0.520)	(0.001)			
	[0.464]	[0.894]	[0.464]	[0.464]	[0.624]	[0.006]			
Test: T2 = T3	(0.856)	(0.448)	(0.626)	(0.040)	(0.939)	(0.007)			
	[0.939]	[0.895]	[0.939]	[0.120]	[0.939]	[0.039]			
Test: T1 = T3	(0.372)	(0.413)	(0.123)	(0.579)	(0.536)	(0.180)			
	[0.579]	[0.579]	[0.539]	[0.579]	[0.579]	[0.539]			
Mean of control	0.613	0.347	8.970	0.657	0.430	10.412			
N	2,630	2,636	2,639	2,084	2,088	2,090			

[1] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband.

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 10: Primary Male's childcare activities

	Midline			Endline			Midline vs endline		
	(1) Number of childcare activities (out of 7) in the last 3 days	(2) Took part in any childcare activities in the last 3 days	(3) Took part in any child feeding activities in the last 3 days[1]	(4) Number of childcare activities (out of 7) in the last 3 days	(5) Took part in any childcare activities in the last 3 days	(6) Took part in any child feeding activities in the last 3 days[1]	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.138	0.011	0.034	0.092	0.018	-0.015	0.787	0.805	0.509
	(0.099)	(0.024)	(0.026)	(0.157)	(0.013)	(0.074)			
	[0.397]	[0.773]	[0.397]	[0.773]	[0.397]	[0.836]			
T2	-0.021	0.007	0.021	0.053	-0.004	0.075	0.645	0.703	0.345
	(0.106)	(0.026)	(0.026)	(0.113)	(0.014)	(0.058)			
	[0.840]	[0.840]	[0.840]	[0.840]	[0.840]	[0.840]			
T3	-0.037	0.019	-0.011	-0.002	-0.011	-0.009	0.829	0.345	0.974
	(0.097)	(0.026)	(0.026)	(0.140)	(0.018)	(0.064)			
	[0.990]	[0.990]	[0.990]	[0.990]	[0.990]	[0.990]			
Test: T1 = T2	(0.172)	(0.876)	(0.676)	(0.775)	(0.132)	(0.201)			
	[0.401]	[0.876]	[0.876]	[0.876]	[0.401]	[0.401]			
Test: T2 = T3	(0.887)	(0.656)	(0.262)	(0.639)	(0.709)	(0.170)			
	[0.887]	[0.851]	[0.786]	[0.851]	[0.851]	[0.786]			
Test: T1 = T3	(0.108)	(0.761)	(0.137)	(0.406)	(0.070)	(0.914)			
	[0.273]	[0.914]	[0.273]	[0.608]	[0.273]	[0.914]			
Mean of control	2.218	0.794	0.678	1.728	0.983	0.469			
N	2,076	2,076	1,842	750	750	457			

[1] Excludes children who are exclusively breastfed.

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Chapter 6: Conclusion

1. Conclusion Outline

In Section 2 of this concluding chapter, I briefly address the overall objective of this research and summarize the main findings as they relate to the principal research questions posed in Chapter's 2-5. Before considering the overall interpretation and delving into the implications of these results, Section 3 reviews the study setting, and Section 4 underscores unique features that differentiate this program from other graduation programs. In Section 5, results from different chapters are synthesized and explored in further depth. Finally, in Section 6, I present an analysis of policy implications and opportunities for future research on questions raised by this study.

2. Revisiting the Research Objective and Questions

Presented in the introduction, the primary objective of this research was to unpack and assess the relative efficacy of different program elements and their combinations as incorporated in a light-touch graduation program. Using the cross randomization of a set of basic and enhanced livelihoods and nutrition programming, we investigated the treatment effects of different programming combinations on various outcomes of interest. In most cases, the program components were closely aligned with the Productive Safety Net Program (PSNP) priorities and implementation strategies. Due in part to this programmatic alignment to a national-wide program, one potential limitation of the study was the light-touch nature of fundamental graduation program components. For example, even within the most intensive version of the livelihoods programming, livelihood transfers were notably smaller in value and targeted solely to a subset of extremely poor households. Consequently, there was a risk that even in the 'enhanced' versions of livelihoods and nutrition programming, the combined package could still have fallen short of a minimum level of investment necessary to achieve graduation program objectives.

In response to the research question addressed in Chapter 2, we partially attribute the null effects on household consumption, the overall livelihoods outcome, to this light-touch characterization. Other external factors described in the section below likely contributed as well to this result. We conclude that in this setting, the lighter-touch graduation program design was insufficient to catalyze a meaningful exit from poverty for a majority of households. In Chapter 3, we ask the question of whether the novel inclusion of targeted group psychotherapy intervention into this program could improve outcomes for women who experience depression. Large and sustained mental health and wellbeing effects were observed only in the treatment arm that combined both enhanced livelihoods, including transfers, and enhanced nutrition, including psychotherapy. This reinforces the notion that poverty and mental health are closely interlinked in this setting and only by relaxing constraints in both domains could we expect sustained impact for those already experiencing depressive symptoms.

In Chapter 4, we explore the question of whether a light-touch graduation program with either basic or enhanced nutrition programming impacts child nutrition and growth. Perhaps unsurprisingly, we observe a pattern of effects similar to those of Chapter 3, where only in extremely poor households

that received cash transfers and exposure to enhanced nutrition programming did we find significant effects on the primary outcome of child growth. Similar to other programs, SPIR embedded gender-related elements into its targeting, design, and implementation across all activities. In Chapter 5, we observed that exposure to SPIR programming led to improved gender equitable attitudes and actions by men in all treatment arms at the midline. However, by the endline, sustained effects were only observed in enhanced nutrition treatment arms which included men's engagement groups. Alleviating concerns of social desirability bias, the large treatment effects on men's self-reported actions were corroborated by their spouses' responses.

3. Study Context

The government's PSNP is targeted at a macro level to areas of the country that are chronically food insecure due to recurrent droughts and low productivity farming systems characterized by poor soil fertility and geographical remoteness from markets. Within these rural areas, the poorest 10-15 percent of households are targeted. The estimated $1.26 per person daily consumption among PSNP households in the study is significantly below the international poverty line of $2.15 per person per day. In contrast to many contexts of other graduation programs, there are very few opportunities for productive non-agricultural enterprises in these remote, rural areas in Ethiopia. These are mostly petty trading or other temporary, informal activities with low profit margins and revenue potential, born out of necessity to smooth consumption rather than responding to profitable business opportunities (Nagler and Naudé 2017). This is also reflected in our surveys where less than 5 percent of households report either engaging non-agricultural businesses or regular wage labor.[1]

Low rates of education, one measure of human capital, likely prevent engagement in the very few skilled labor opportunities. In our sample, 72 percent of household heads report they have no formal education. Households may also consider sending an adult member to migrate for work in a low-skill sector. In an attempt to follow the manufacturing, export-led growth example of several East Asian economies, the government of Ethiopia built industrial parks across the country as a part of its Growth and Transformation Plans (GTP I and II). The idea was that prioritized sectors (especially apparel and textiles) would benefit from a large supply of low-cost labor. One study that evaluated outcomes of workers that were offered employment in these industrial jobs by lottery reported that over three quarters of workers quit within the first year (Blattman and Dercon 2018). The main reason was that they found the jobs unpleasant and hazardous to their health. There was no effect on income for those that received job offers, and large negative health effects observed after one year.

Perhaps the most revealing evidence to the lack of labor market options for these households is their willingness to contribute labor for six months out of the year in strenuous manual work on public work sites for less than $300 in payment.[2] As described in the introduction chapter, these

[1] In comparison, Bossuroy et al. (2022) report that 60 percent of household's income is derived from non-agricultural businesses in their study.

[2] Median annual transfers were valued at around $277 per household in 2017 PPP terms at baseline. Eroding the value of this already limited payment, the government has not revised the person-day wage rate to keep up with inflation due to budget constraints.

already vulnerable households were beset by numerous shocks, especially in the period prior between the midline and endline surveys. This included market disruptions due to COVID-19, pest infestations, drought, flooding, and conflict. Average reported impacts also conceal the fact that some households lost their entire annual harvest. For example, nearly 30 percent of households that reported any crop loss due to desert locusts, indicated the loss was total or near total in magnitude. I visited one of these communities that described how a swarm of locusts descended on their fields like a large cloud right before the intended harvest period, and afterwards nothing remained. As predominantly subsistence farming households, these type of covariate pest or weather shocks push them beyond their limited coping strategies such as borrowing from family or neighbors.[3]

As mentioned in previous chapters, some effects from these multiple shocks are likely revealed in deteriorating conditions among households in the control kebeles (who themselves are still benefiting from core PSNP transfers). For example, the average level of savings among households in the control group decreased from $75 at midline to $35 at endline. The value of livestock assets declined by 33 percent among this group over the same period. The already extremely low rate of women achieving a minimal accepted diet dropped from 11.5 percent at midline to 5.2 percent at endline. It is against this stark backdrop that results from this study should be considered, especially when drawing implications for policy while comparing to results from other graduation program studies.

4. Unique Characteristics of the SPIR Program and Study

As mentioned before, one of the important differences between this and most other graduation program studies is that this program was implemented at scale. Additionally, it directly supported and layered on the government of Ethiopia's national productive safety net program (PSNP). Several design decisions were made with this this alignment in mind. Instead of uniform livelihoods support to all participants[4], SPIR followed the PSNP segmented approach, targeting less poor clients for one-time business skills training[5] and linking them to financial service providers, and identifying the poorest clients for livelihoods transfers (with the transfer size matching the amount chosen in the PSNP). Compared to the individual household coaching and mentoring visits of the standard graduation model,[6] SPIR facilitated the formation of savings groups and producer marketing groups where PSNP clients supported each other in self-initiated income generating activities and collectively purchased inputs or sold their products together. Lastly, the consumption support in standard graduation programs is unconditional and given weekly or monthly during the first year of the program. In our case, the majority of PSNP clients receive transfers for six months *each* year,

[3] In the example of the community devastated by desert locusts, they described how all able-bodied adult men left in desperate search of any form of day labor they could find in nearby towns.
[4] This is a common feature in standard graduation programs, but infeasible due to budget constraints in a program such as the PSNP. For example, in the PSNP5 Livelihoods Support Program Operations Manual, it states that "550,000 households will be selected (for livelihood support), including 167,000 in the livelihoods transfer track." This represents 6 percent of the approximately 2,667,000 households in the program.
[5] This was mostly focused on productive livestock activities including improved poultry production and short-cycle animal fattening and sales.
[6] 104 household visits over two years

and these transfers were made as payment for public works labor contributed during the same period. Importantly, both treatment and control arms in SPIR receive these core transfers.

In the remainder of this section, I outline several key distinctions between this study and others concerning livelihoods, mental health, nutrition, and women's empowerment. When comparing the results of this study to other graduation program studies, it is important to consider the heterogeneity of effects observed in other programs. For example, the original multi-country study found that treatment effects on income, assets, and consumption were driven by those in the top wealth quantiles (Banerjee et al. 2015). No treatment effects were measured on financial inclusion for those in the lowest two wealth quantiles, or on income for those in the lowest percentile. The effect on income was nearly 20 times larger for those in the 90th percentile than those in the 25th percentile. Similarly, the effect on an asset index was 14 times larger for those in the 90th percentile than those in the 10th percentile, and 4 times larger for the same comparison on consumption effects. In that study, it was clearly the less poor households that benefitted the most from the program, translating the large one-time asset transfer and other components into substantial gains in income, assets and consumption. By comparison, less poor households in SPIR were not targeted for livelihood transfers.[7] Perhaps it is not surprising that we find no overall effects on consumption, when SPIR only provided (smaller) transfers to extremely poor households in our sample.[8]

As described in Chapter 3, one of the novel contributions of this research was the introduction of a targeted group psychotherapy intervention to address interpersonal sources of depression. In the mental health literature, it is rare that these interventions are targeted uniquely to individuals in ultra poor households or evaluated in combination with other programming.[9] In the same literature, studies often target individuals experiencing severe symptoms, with evaluations typically conducted within a few months post-treatment (Singla et al. 2017; Huang et al. 2020; Cuijpers et al. 2018).[10] These study features are partially due to the nature of depression symptomology which can be episodic, lasting a few months, followed by periods of remission. By contrast, our study targets individuals experiencing moderate symptoms in poor households and evaluates the combined psychotherapy and graduation programming one year after the end of treatment. Two studies that combine cash with psychotherapy are more closely related to our research (Haushofer, Mudida, and Shapiro 2020; Blattman, Jamison, and Sheridan 2017). However, these also have design features (indicated by italics) that differ from our study, with *much larger* cash transfers and *individual* psychotherapy delivered to a *general* population in Kenya (Haushofer, Mudida, and

[7] At SPIR's baseline, households in the 'less poor' designation owned more than twice the level of assets as the 'extreme poor' households and had a higher-level savings.

[8] As an additional comparison, the ratio of the transfer value to consumption is about twice the size on average for the multi-country study compared to the ratio of transfer value to consumption in SPIR. So, while the multi-country study reported statistically significant, albeit small in magnitude, effects on consumption for the poorest households in their sample, the transfer value itself was meaningfully larger vis-à-vis their starting level of consumption.

[9] Also, median sample sizes for these studies tend to be around 200 or smaller (Singla et al. 2017), as compared to 444 in our study (women screened at midline with depressive symptoms).

[10] Notable exceptions are longer term follow-up evaluations of psychotherapy in India (4-5 years) and Pakistan (7 years) (Bhat et al. 2022; Baranov et al. 2020).

Shapiro 2020), and group psychotherapy targeted to *criminally engaged young men* in Liberia (Blattman, Jamison, and Sheridan 2017).

Livelihood creation (asset transfer, training, consumption support, savings account, coaching) is the main thrust of the original graduation model. However, the last of the six program components in the model was health education and linkage to basic health services (Banerjee et al. 2015). They also assessed effects on an index of physical health that combined measures of perceived health, days of work missed due to poor health, and an activities-of-daily-living score. There were no physical health treatment effects other than a small, weakly measured effect on self-perceived health.[11] As described in Chapter 4, our study is the first we are aware of that evaluates the addition of enhanced nutrition programming to a graduation program, and examines its effects on child growth and nutrition. The basic nutrition package in our study is similar to the standard graduation program of health education and linkage to basic health services. SPIR's additional enhanced nutrition programming draws elements from recent nutrition-specific programs that were evaluated in Ethiopia. This includes a study of Community-based Participatory Nutrition Promotion (CPNP) that found an 8 percent differential decline in child stunting (Kang et al. 2017) and an intensive set of nutrition BCC activities that led to a differential decline of 5.6 percentage points in child stunting after two years (Kim et al. 2019). Unlike the CPNP study that used the underweight eligibility qualification for CPNP as their study enrollment criteria, less than 10 percent of children in SPIR's enhanced nutrition treatment arms qualified and were enrolled in CPNP.[12] The intensive nutrition BCC in the other study combined interpersonal communication, community mobilization and mass media messaging, and was targeted towards a general population in a more productive area of Ethiopia. This contrasts with our study that targeted the poorest households in food insecure areas with a limited selection of the intensive nutrition BCC activities evaluated in this nutrition-specific program.[13]

Women's empowerment was an outcome whose initial treatment effects (0.046 SD) faded by the second endline in the multi-country graduation study (Banerjee et al. 2015). A similar pattern was observed in a graduation program in Afghanistan, but when a broader set of measures were included, stronger treatment effects were reported (Bedoya et al. 2019). A more recent graduation program that included both a cash transfer and psychosocial intervention found that women's control over earnings improved most in the cash arm, while social support improved most in the psychosocial arm; but there were no effects from any treatment on women's control over household resources (Bossuroy et al. 2022). At midline, we found effects on women's control over earnings in households that received either the cash or poultry transfer, but these effects faded by endline (Alderman et al. 2021). Targeting men's gender equitable attitudes and practices through men's engagement activities is not a common women's empowerment component of graduation programs. Apart from our study, the other program that included this strategy was a women's focused graduation program in DRC (Angelucci, Heath, and Noble 2023).[14] While they found large

[11] However, in one of the countries (India), a 10-year follow-up found a 0.2 SD effect on the physical health index.

[12] The pooled treatment effect of N* was increase of 6.5 percentage points in terms of reported CPNP participation at endline (Alderman et al. 2021).

[13] In fact, the macro targeting for this study intentionally selected agriculturally productive districts that were not participating in the Productive Safety Net Program (Kim et al. 2019).

[14] Their men's engagement program was twice as long as the one included in our study (16 vs. 8 sessions).

effects on women's empowerment (0.2 SD), there were no differential effects in the treatment arm that included the men's engagement activity.

5. Synthesizing Results from the Core Chapters

An interesting set of results linking findings from Chapter 2 and 4 sheds light on the tradeoffs faced by resource-constrained households as they navigate the multiple objectives promoted in this program. Considering the program's impact on livelihood outcomes in Chapter 2, we find very little heterogeneity with respect to variation in nutrition programming. A key exception is the interaction among extremely poor households between the cash and enhanced nutrition treatments. The effect of the cash treatment on assets without enhanced nutrition programming is large and precisely estimated (0.22 SD).[15] This effect is completely nullified when cash is combined with enhanced nutrition programming. This prompts the question: where did these resources go if not toward asset accumulation? One plausible explanation is that extremely poor households in this treatment arm allocated more resources to purchasing food for their young children. Chapter 4 corroborates this, as it was the only treatment where any effects on child growth were observed. Specifically, there was a large,18 percentage point reduction in stunting from a control mean of 54 percent.

An activity that may have contributed to this outcome was participation in CPNP, where statistically significant effects on participation were only observed in the same cash and enhanced nutrition treatment. In enhanced nutrition kebeles, households with underweight children were specifically targeted for participation in CPNP, the two-week recuperative feeding and nutrition promotion session. Since the meals provided during these sessions were sourced from local ingredients contributed by participating caregivers, there is a concern that extremely poor households may have been deterred from attending due to an inability to contribute to these daily feeding sessions. CHFs and health workers attempted to prevent this by raising awareness among community members to only contribute according to their means and allowing caregivers to provide varying types of food ingredients or other in-kind contributions such as water or firewood. The increased income generated from investments made with the cash transfer may have had a dual effect: first, it could have alleviated the stigma associated with being unable to contribute food during CPNP sessions, thereby boosting participation rates. Second, it may have eased the financial constraints associated with adopting the recommended feeding practices for their young children afterward.

Considering the substantial impact on child growth, it seems that extremely poor households in the cash and enhanced nutrition arm weighed their options and prioritized investing their supplemental income in food for their children rather than augmenting their assets. As our experimental design also included a treatment that combined the poultry transfer with enhanced nutrition, this raises the question of why we did not observe a similar pattern of effects[16] for a livelihood activity that is presumed to be particularly 'nutrition sensitive.' Poultry production is widely perceived as an intervention with potential to contribute to both income and nutrition objectives, with the latter met through increased consumption of eggs and use of poultry or egg sales income for regular

[15] Chapter 2, Table 3

[16] The treatment effects on assets remained consistent among poultry recipients, regardless of their exposure to enhanced nutrition programming.

household nutritional needs.[17] Commonly practiced in these settings, small-scale poultry production is characterized by indigenous breeds,[18] minimal inputs, and high susceptibility to diseases, all of which contribute to low productivity and production (Dessie and Ogle 2001). The poultry start-up package provided by SPIR included improved breed chickens[19] that maintain a much higher level of egg production even under semi-scavenging feeding systems. They were also fully vaccinated against all diseases except Newcastle disease, which requires periodic vaccination to maintain resistance over the life of the chicken. At the midline, a few months after the poultry start-up package was received, we observed significant effects on child egg consumption in the previous day. This effect, an 11 percentage point increase over a control mean of 4 percent, was observed only in the poultry treatment arms (Alderman et al. 2022). The child egg consumption effect persisted in the poultry arm that received enhanced nutrition at endline,[20] but faded in the poultry arm without enhanced nutrition programming (Chapter 4, Table 7, Column 9).[21]

There were also large and statistically significant effects on women's reporting income on poultry and egg sales in the poultry treatment arms at midline (Alderman et al. 2020),[22] with poultry sales (but not income from eggs) effects persisting at endline.[23] The poultry sales' effect is consistent with a decreasing flock size observed at both survey rounds. At midline, of the 16 chickens provided to the poultry transfer recipients, on average, 8 remain. As half of the chickens were male, this is in keeping with the program recommendation to sell the cockerels when they attained an optimal market weight. When the 8 hens reached the end of their productive period (approximately 18-24 months), it seems most poultry recipients sold these hens without replacing them. This is reflected by poultry transfer recipients having 3.5 chickens on average compared to 2 chickens in extremely poor household in the control kebeles. Given the sharp decline in poultry ownership, the absence of a persistent effect on egg sales is not surprising.

The puzzle remains as to why there were no statistically significant anthropometric treatment effects for young children in poultry recipient households in the enhanced nutrition programming arm. Even with a declining flock size, persistent effects on child egg consumption and interestingly, also dairy consumption are observed in this treatment group.[24] As many factors contribute to child

[17] Egg sales can provide regular (weekly) income. As poultry production is most commonly practiced by women, there is an expectation that regular income derived from this activity is more likely to support routine household nutrition needs than other livelihood activities with larger, lumpier income patterns that may be more often controlled by men.
[18] Indigenous breed poultry lay 40-60 eggs per year, taking time to rear chicks between laying periods.
[19] Primarily Ethiochicken's Sasso dual purpose breed, a French hybrid that produces four times as many eggs per year than indigenous breeds.
[20] A 14 pp increase over a control mean of 5.4 percent at the endline in the poultry and enhanced nutrition treatment arm.
[21] Female caregiver egg consumption effects (a 6 pp increase over a control mean of 4 percent) in the poultry treatment arm at midline faded by endline.
[22] A 15.6 pp increase over a control mean of 33 percent for women's reporting income from poultry sales, and a 18.5 pp increase over a control mean of 19 percent for women's reporting income from egg sales (Alderman et al. 2020).
[23] A 14 pp increase (control mean 30 percent) in women's reporting any income from poultry sales among women in extreme poor households in the poultry treatment group (Alderman et al. 2021).
[24] There are strong associations between consumption of animal sourced foods and child growth (Headey, Hirvonen, and Hoddinott 2018), and this was one of the pathways for treatment effects in the Alive and Thrive

nutrition, it is difficult to determine the underlying set of conditions that led to improved child growth in the enhanced nutrition treatment that received cash, but not in the poultry transfer kebeles. One explanation could be that the overall quantity of complementary feeding was more salient than an improved quality of potentially smaller amounts of nutritious food. While the poultry and enhanced nutrition treatment effects on egg and dairy consumption was large in magnitude, these were in comparison to very low initial averages. Even with nearly doubling the prevalence of milk consumption[25] and quadrupling that of egg consumption, this only resulted in one third of children consuming dairy products and one fifth of children consuming eggs in this treatment group.

While perhaps less likely, another explanation to this puzzle could be that nutritional gains from consuming poultry products were undermined by increased exposure to fecal matter leading to diarrhea or environmental enteric dysfunction[26] (Gelli et al. 2017; George et al. 2015). In traditional poultry systems, chickens are often kept inside the household dwelling at night and left to scavenge nearby the house during the day. Using a large agricultural survey of rural households in Ethiopia, one study found a positive association of poultry ownership with child growth, but a negative association with the practice of corralling poultry in the house overnight (Headey and Hirvonen 2016). Careful hygiene management was an integral component in the SPIR poultry start-up kit and training (including materials for constructing a chicken coop and handwashing station). While we observed a high level of compliance with coop construction and initial management of the chickens (providing chicken feed, keeping them contained in the coop and maintaining a clean environment), it is possible that households reverted to traditional practices of allowing chickens to scavenge for food as it became less profitable to purchase chicken feed and actively manage reduced flock sizes. One study deliberately explored this issue through an experiment involving poultry production training, combined poultry and livestock WASH training, and a control group. They observed slight improvements in livestock WASH knowledge and practices, but no impact on child stunting (Gelli et al. 2023).[27]

As described in Chapter 2, the large treatment effects (0.38 SD) on savings among extremely poor households that received transfers compare favorably to Banerjee et al. (2015)[28], while effects on income from livestock (0.2 SD) and livestock assets (0.1 SD) are notably smaller.[29] In the multi-country study, mental health treatment effects had declined from around 0.1 SD to 0.06 SD but were still statistically significant at the second endline. The prevalence of depression among women in extremely poor households that received transfers was reduced by about one third

study in Ethiopia: increased egg consumption associated with increased rates of minimum dietary diversity and increased HAZ scores (Kim et al. 2019).

[25] A 16 pp increase over an 18.5 percent control mean for index child dairy consumption in the poultry and enhanced nutrition treatment.

[26] Environmental enteric dysfunction is an inflammation of the small intestine leading to impaired gut immune function, malabsorption, and associated with child stunting, even among seemingly asymptomatic children without diarrhea (Korpe and Petri 2012).

[27] Unrelated to poultry production, another study compared different randomly assigned health, nutrition education, or hygiene interventions in Ethiopia found the only intervention to have a significant impact on child stunting was the WASH intervention (Fenn et al. 2012).

[28] Savings treatment effects of 0.38 SD

[29] As compared to around 0.35 SD on livestock income and 0.25 SD on productive assets (Banerjee et al. 2015).

(Chapter 3, Table 2, Colum 4).[30] Our program did not have overall consumption or food security impacts, however it appears that the moderate income and wealth (assets) effects for those that received transfers were enough to lead to reductions in female depression. When considering the more restrictive measure of major depressive disorder (Chapter 3, Table 4), the effect of receiving the transfers for women with depression at midline is similar with or without psychotherapy on this clinical measure of depression.[31] The livelihood transfers may have been particularly salient for women in these extremely poor households who faced different idiosyncratic or covariate shocks. At midline, having experienced an illness in the household in the last two years was associated with a 5 percentage point higher prevalence of depression, while facing a drought in the past two years was associated with a 7.6 percentage point higher prevalence of depression for women.[32]

6. Policy Implications and Opportunities for Future Research

Despite falling short of achieving the graduation programming goal of facilitating a sustainable exit from poverty, SPIR contributed to large treatment effects on savings for all households. This effect was similar for households that received transfers and those that did not. This increased savings, supported by participation in VESAs, appears to be the primary mechanism by which SPIR households mitigated negative effects of drought, smoothed consumption and partially protected their livestock holdings (Hirvonen et al. 2023). Using a measure of drought that captured relative dryness in comparison to a long-term mean[33] during the principal cropping season, a relatively small 0.25 SD deviation in dryness led to a 36 percent increase in food insecurity and a 35 percent decrease in livestock holdings for households in the control arm. These effects are partially or fully muted for households in the SPIR treatment arms (Hirvonen et al. 2023). In the absence of formal insurance products, these local savings groups appear to be instrumental in protecting households from significant negative effects of even moderate covariate shocks.

Moderate treatment effects on income and assets are observed for households that receive livelihood transfers, while no such effects are observed among either the comparable extremely poor households or the less poor households that did not receive livelihood transfers. This underscores the importance of livelihood transfers in graduation programs, suggesting that graduation programs lacking this component are unlikely to achieve core livelihood outcomes (Banerjee et al. 2022). Regarding the size of the one-time transfer, the conclusion in Chapter 2 is that even if households receiving livelihood transfers were able to increase their livestock assets

[30] This was a 4-5 pp reduction from a control mean of 15 percent, with effects only precisely measured in the T2 treatment arm.

[31] For the women who displayed depressive symptoms at midline, it was the combination of receiving livelihood transfers and participation in the psychotherapy (as a part of the enhanced nutrition programming) that led to a broader array of improvements in the prevalence of depression (PHQ-9 score of 8 or above), happiness levels, livestock income, and physical health.

[32] At endline, among the covariates added in Table A2 (Chapter 2) reporting the treatment effects on female depression, these two indicators were also statistically significant (4.4 pp increase for household sickness and 3.4 pp increase for household affected by drought).

[33] Using the Standardized Precipitation-Evapotranspiration Index (SPEI) this measure is able to capture effects of moderate drought events (Vicente-Serrano, Beguería, and López-Moreno 2010).

and income, the magnitude of these results was not large enough to catalyze a transformative impact in their lives and livelihoods.

One potential explanation for these results may lie in the relatively modest size of the livelihood transfer implemented in our program compared to the one-time transfers in other graduation programs.[34] If a poverty trap mechanism operates in this or similar contexts, an intervention that fails to push households above a certain asset threshold may prove ineffective in facilitating their escape from poverty. Instead, households may regress to a lower equilibrium over time (Balboni et al. 2022). Consequently, in the presence of a poverty trap, reducing the size of the livelihood transfer – for instance, from $600 to $300 – could jeopardize the entire program investment, particularly if a transfer of $500 or more is necessary to increase household assets above the critical threshold.[35] Given variations in country contexts and other graduation program design options, comparing the distinct impact of different transfer sizes in graduation programs poses a challenge. Nevertheless, the findings from this study imply that a larger transfer size, when combined with other components of graduation programs, is likely essential for achieving overarching poverty alleviation objectives.

Weekly household-level coaching visits over a two-year period represent one of the most expensive elements of the standard graduation model.[36] Although our study did not directly compare group-based livelihoods support with the individual coaching model, the treatment effects on financial inclusion, primarily achieved through participation in savings groups (VESAs), were similar to those observed in the standard graduation model. In a study that experimentally varied either group or individual coaching in a graduation program in Uganda, researchers found no statistically significant differences between the two treatments across a range of outcomes, including productive assets, income, consumption, and food security (TOPS Uganda Graduation Randomized Control Trial Associate Award 2022). Combined with the results of our study, these findings suggest that a less intensive, group-based livelihood support approach could be one way to reduce program costs without lessening the impact on livelihood outcomes. These groups may also provide social support to their members, including a supplemental 'social fund' savings mechanism to mitigate the negative effects of shocks for members.

Another contribution of this research is addressing the question of whether there are benefits or drawbacks to providing one-time livelihoods transfers in the form of in-kind assets versus cash. Most graduation programs studied to date have provided these as in-kind transfers, often in the form of livestock (Banerjee et al. 2015; Balboni et al. 2022; Bandiera et al. 2017; Bedoya et al. 2019; Brune et al. 2022).[37] To our knowledge, our study is the first to experimentally vary the type of livelihood transfer within a graduation program. In 2017, Bill Gates advocated for providing poultry to households in extreme poverty, citing its potential to support multiple outcomes including

[34] Refer to Chapter 2, Table A1 for a transfer size comparison with graduation programs in 10 other countries.
[35] In the PSNP5 livelihoods technical design group, the recommendation was to increase the transfer from $200 to $500-600, but due to budget constraints this was reduced to $300 for PSNP5 (the subsequent phase of PSNP programming).
[36] Representing more than one third of the total program budget in the canonical study (Banerjee et al. 2015).
[37] More recent exceptions include studies in Niger (Bossuroy et al. 2022), DRC (Angelucci, Heath, and Noble 2023)

livelihoods, nutrition and women's empowerment.[38] Chris Blattman responded by making the case instead for cash transfers and recommending a direct comparison of the two approaches.[39] This is precisely what we examine through the random assignment of enhanced livelihood treatment kebeles to either poultry or cash transfers of equivalent value. When assessing livelihoods outcomes, we find no difference in the treatment effects of cash or poultry. This may be attributed to participants perceived risks of managing a large number of poultry, compared to the less risky and potentially more profitable livestock fattening and sales activity.[40] This is evident in the large divestiture of chickens and transition to other livestock production activities among most households in the poultry treatment group.

Somewhat surprisingly, we found that when paired with enhanced nutrition, the cash transfer treatment had large effects on child anthropometric outcomes whereas the presumed 'nutrition sensitive' poultry transfers did not. The absence of treatment effects in the poultry and enhanced nutrition arm is somewhat puzzling, as we did observe effects on intermediate outcomes such as increased consumption of animal sourced foods (eggs and dairy) among young children in these households. There was no difference in the largely null effects on women's empowerment between cash and poultry.[41] Taking into account the additional procurement and logistics challenges associated with supplying the in-kind poultry package, cash emerges as the clear winner in this study setting. This finding holds particular relevance for policy formulation and program implementation. Given the similarity of results between the two transfer modalities, if political constraints render cash unfeasible in certain settings, in-kind asset transfers could serve as a viable alternative. We do not find evidence for multidimensional additive effects that favor poultry over cash in settings where cash could readily replace this asset transfer in graduation programs.

Given the potential bidirectional relationship between poverty and common mental health disorders, could a targeted mental health intervention play a pivotal role in unlocking a pathway to improved outcomes as a part of a larger graduation program? Without addressing financial poverty, one of the underlying causes of poor mental and physical health among these ultra poor households, our study finds that a group psychotherapy without livelihood transfers does not have sustained treatment effects on mental health outcomes. In contrast, the combination of group psychotherapy and livelihood transfers (in conjunction with the larger set of program components) has large effects on women's mental and physical health. This novel study highlights the potential impact of integrating mental health interventions into future graduation programs, although numerous design questions remain and warrant further research. Should the intervention be delivered in a group or through individual counseling sessions? Should it be provided to all program participants or targeted to individuals exhibiting symptoms of depression, stress, or anxiety? What should be the level of intensity? Twelve weeks of psychotherapy? Or perhaps 5-7 weeks would still

[38] Why I would raise chickens | Bill Gates (gatesnotes.com)
[39] Bill Gates wants to give the poor chickens. What they need is cash. - Vox
[40] In one of the country sites in the multi-country graduation study, the majority of participants chose to receive poultry for the asset transfer. A large number of the chickens died due to disease outbreak(s) and authors found no treatment effects on consumption and financial inclusion, and a negative treatment effect on assets in this country (Honduras) (Banerjee et al. 2015).
[41] While there are no average treatment effects on Intimate Partner Violence (IPV) of either poultry or cash, we do find a marginally significant 4 percentage point increase in emotional and physical violence experienced by women in the poultry treatment groups. These effects do not hold up under multiple hypothesis testing.

be effective in many cases. Should it focus on a specific condition such as depression, or perhaps a broader, transdiagnostic intervention may be more appropriate to address a range of interrelated mental health issues. Could a lighter touch psychosocial intervention achieve similar results? Which of these interventions are scalable or practical to embed in national government health services structures?

Based on our observations, one aspect that appeared to be instrumental in sustained mental health improvements was the supportive group environment. These intentionally small groups (6-8 members) consisted of women that were all dealing with depressive symptoms. Within these settings, there existed a profound level of transparent and open sharing facilitated by mutual trust and a collective commitment to maintaining confidentiality. The support and encouragement participants derived from these groups was evident in participants' regular meeting together even months after the facilitated IPT-G sessions concluded. Our observations suggest that beyond cost-saving benefits of the group model, there may be additional benefits to delivering these interventions in group settings.

While psychotherapy interventions are typically directed towards individuals experiencing mental distress or depression, one recent study found positive treatment effects on economic and psychological outcomes when psychotherapy was administered to a general population in rural Ghana (Barker et al. 2022). This contrasts with findings from a graduation program that offered psychotherapy in groups comprising both depressed and non-depressed individuals. They found null or even deteriorating mental health outcomes from exposure to these groups, with the hypothesis that the mixed nature of these groups exacerbated stigma for those experiencing depressive symptoms.[42] Considering additional factors such as cost and scalability, it seems policy makers and graduation program implementers should prioritize targeted psychotherapy interventions over those provided to a general population, at least until further evidence becomes available.

Regarding the question of model intensity, from a 3-month psychotherapy intervention to lighter touch psychosocial interventions, our study did not test or compare different models. However, we found no effects from a very light-touch aspirations treatment[43] in our enhanced livelihoods programming (Leight et al. 2021) that was similar in nature to other psychosocial treatments (John and Orkin 2022; Orkin et al. 2023; Bossuroy et al. 2022; Bernard et al. 2014). In a recent review of evidence on these lighter-touch psychosocial interventions, Haushofer and Salicath (2023) conclude that while effects may not be large and could fade quickly, the low-cost nature of these interventions may justify their application in terms of cost-effectiveness. Further research is necessary to address these and other relevant questions, such as scalability, when considering the integration of mental health interventions into graduation programs.

Our study also contributes novel evidence on the potential of intensive nutrition BCC inclusion in graduation programs to impact child nutrition outcomes in ultra-poor households. Consistent with

[42] Authors also cite deviations from the WHO implementation guidelines and anecdotal evidence of breaches of confidentiality as other potential factors behind these unexpected results.
[43] This was a one-time screening of videos featuring positive messages from farming households similar to the participants, and showcasing how they used their own resources to make investments that had positive long-term results. These videos were selected from those developed in an earlier study (Bernard et al. 2014).

the pattern of effects on other outcomes, anticipated results may only be achieved when nutrition programming is coupled with livelihoods support (notably including a one-time cash transfer in addition to other program components). Our findings suggest that while intensive nutrition BCC alone can improve intermediate outcomes such as knowledge of infant and young child feeding practices and access to health services, the combination of intensive nutrition and livelihood interventions, including livelihood transfers, is necessary to achieve the primary objective of improved child growth among extremely poor households targeted by these programs. Additionally, our results underscore existing evidence indicating that conventional, non-intensive nutrition programming is insufficient to achieve nutritional objectives within the framework of graduation programs.

While increased women's empowerment is often targeted as a standalone objective, it could also be instrumental in attaining many of the household level outcomes targeted by graduation programs. However, findings from our study indicate that conventional gender programming may not be adequate to sustainably change men's gender-equitable attitudes and behaviors. Adding a targeted men's engagement group intervention shows promising results in this regard. One possible mechanism is that through open discussions within a men's-only group, men may revise misconceptions about their peers' beliefs regarding gender-restrictive social norms (Bursztyn, González, and Yanagizawa-Drott 2020). Even substantial changes in men's involvement in household chores may not be sufficient to reduce women's overall workload (Doyle et al. 2018). Explicitly targeting women with livelihood transfers and associated training, while excluding their spouses, also risks increasing women's stress.[44] More research is necessary to understand how to better design existing graduation program components with a gender lens and identify what additional components are needed to improve women's empowerment outcomes.

[44] While frequently raised as a concern, the evidence does not substantiate the hypothesis that increasing women's economic empowerment through targeted livelihoods programming increases the incidence of intimate partner violence (Eggers del Campo and Steinert 2020; Buller et al. 2018).

References

Alderman, Harold, Lucy Billings, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Alemayehu Seyoum Taffesse, and Heleene Tambet. 2021. "Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA): Endline Report." https://ebrary.ifpri.org/digital/collection/p15738coll5/id/8084.

Alderman, Harold, Daniel O Gilligan, Melissa Hidrobo, Jessica Leight, Michael Mulford, and Alemayehu Seyoum Taffesse. 2020. "A Randomized Impact Evaluation of the Graduation Program Strengthen Productive Safety Net Program Phase 4 (PSNP4) Institutions and Resilience (SPIR) in Ethiopia." https://doi.org/https://doi.org/10.1257/rct.8281-2.1.

Alderman, Harold, Daniel O. Gilligan, Jessica Leight, Michael Mulford, and Heleene Tambet. 2022. "The Role of Poultry Transfers in Diet Diversity: A Cluster Randomized Intent to Treat Analysis." *Food Policy* 107 (January): 102212. https://doi.org/10.1016/j.foodpol.2021.102212.

Angelucci, Manuela, Rachel Heath, and Eva Noble. 2023. "Multifaceted Programs Targeting Women in Fragile Settings: Evidence from the Democratic Republic of Congo." *Journal of Development Economics* 164 (September): 103146. https://doi.org/10.1016/j.jdeveco.2023.103146.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil. 2022. "Why Do People Stay Poor?" *Quarterly Journal of Economics* 137 (2): 785–844. https://doi.org/10.1093/qje/qjab045.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor Markets and Poverty in Village Economies*." *The Quarterly Journal of Economics* 132 (2): 811–70. https://doi.org/10.1093/qje/qjx003.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236). https://doi.org/10.1126/science.1260799.

Barker, Nathan, Gharad Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher Udry. 2022. "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress." *American Economic Review: Insights* 4 (4): 527–45. https://doi.org/10.1257/aeri.20210612.

Bedoya, Guadalupe, Aidan Coville, Johannes Haushofer, Mohammad Isaqzadeh, and Jeremy Shapiro. 2019. "No Household Left Behind: Afghanistan Targeting the Ultra Poor Impact Evaluation." Cambridge, MA. https://doi.org/10.3386/w25981.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Seyoum Taffesse. 2014. "The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia." *SSRN Electronic Journal*, no. April. https://doi.org/10.2139/ssrn.2514590.

Blattman, Christopher, and Stefan Dercon. 2018. "The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia." *American Economic Journal: Applied Economics* 10 (3): 1–38. https://doi.org/10.1257/app.20170173.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (4): 1165–1206. https://doi.org/10.1257/aer.20150503.

Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand, et al. 2022. "Tackling Psychosocial and Capital Constraints to Alleviate Poverty." *Nature* 605 (7909): 291–97. https://doi.org/10.1038/s41586-022-04647-8.

Brune, Lasse, Dean Karlan, Sikandra Kurdi, and Christopher Udry. 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 (March): 102780. https://doi.org/10.1016/J.JDEVECO.2021.102780.

Buller, Ana Maria, Amber Peterman, Meghna Ranganathan, Alexandra Bleile, Melissa Hidrobo, and Lori Heise. 2018. "A Mixed-Method Review of Cash Transfers and Intimate Partner Violence in Low- and Middle-Income Countries." *The World Bank Research Observer* 33 (2): 218–58. https://doi.org/10.1093/WBRO/LKY002.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott. 2020. "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia." *American Economic Review* 110 (10): 2997–3029. https://doi.org/10.1257/AER.20180975.

Cuijpers, Pim, Eirini Karyotaki, Mirjam Reijnders, Marianna Purgato, and Corrado Barbui. 2018. "Psychotherapies for Depression in Low- and Middle-Income Countries: A Meta-Analysis." *World Psychiatry* 17 (1): 90–101. https://doi.org/10.1002/wps.20493.

Dessie, Tadelle, and B Ogle. 2001. "Village Poultry Production Systems in the Central Highlands of Ethiopia." *Trop Anim Health Prod* 33 (6): 521–37.

Doyle, Kate, Ruti G. Levtov, Gary Barker, Gautam G. Bastian, Jeffrey B. Bingenheimer, Shamsi Kazimbaya, Anicet Nzabonimpa, et al. 2018. "Gender-Transformative Bandebereho Couples' Intervention to Promote Male Engagement in Reproductive and Maternal Health and Violence Prevention in Rwanda: Findings from a Randomized Controlled Trial." *PLoS ONE* 13 (4). https://doi.org/10.1371/journal.pone.0192756.

Eggers del Campo, Isabel, and Janina Isabel Steinert. 2020. "The Effect of Female Economic Empowerment Interventions on the Risk of Intimate Partner Violence: A Systematic Review and Meta-Analysis." *Https://Doi.Org/10.1177/1524838020976088* 23 (3): 810–26. https://doi.org/10.1177/1524838020976088.

Fenn, Bridget, Assaye T. Bulti, Themba Nduna, Arabella Duffield, and Fiona Watson. 2012. "An Evaluation of an Operations Research Project to Reduce Childhood Stunting in a Food-Insecure Area in Ethiopia." *Public Health Nutrition* 15 (9): 1746–54. https://doi.org/10.1017/S1368980012001115.

Gelli, Aulo, Anissa Collishaw, Josue Awonon, Elodie Becquey, Ampa Diatta, Loty Diop, Rasmane Ganaba, et al. 2023. "Effects of an Integrated Poultry Value Chain, Nutrition, Gender and WASH Intervention (SELEVER) on Hygiene and Child Morbidity and Anthropometry in Burkina

Faso: A Secondary Outcome Analysis of a Cluster Randomised Trial." *Maternal & Child Nutrition* 19 (4): e13528. https://doi.org/10.1111/MCN.13528.

Gelli, Aulo, Derek Headey, Francis Ngure, Elodie Becquey, Rasmané Ganaba, Lieven Huybregts, Abdoulaye Pedehombga, et al. 2017. "Assessing the Health and Nutrition Risks of Smallholder Poultry Production in Burkina Faso: Insights from Formative Research." *IFPRI - Discussion Papers*, no. 1665: vi-pp.

George, Christine Marie, Lauren Oldja, Shwapon K. Biswas, Jamie Perin, Gwenyth O. Lee, Shahnawaz Ahmed, Rashidul Haque, et al. 2015. "Fecal Markers of Environmental Enteropathy Are Associated with Animal Exposure and Caregiver Hygiene in Bangladesh." *The American Journal of Tropical Medicine and Hygiene* 93 (2): 269–75. https://doi.org/10.4269/AJTMH.14-0694.

Haushofer, Johannes, Robert Mudida, and Jeremy Shapiro. 2020. "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-Being." Cambridge, MA. https://doi.org/10.3386/w28106.

Haushofer, Johannes, and Daniel Salicath. 2023. "The Psychology of Poverty: Where Do We Stand?," December. https://doi.org/10.3386/W31977.

Headey, Derek, and Kalle Hirvonen. 2016. "Is Exposure to Poultry Harmful to Child Nutrition? An Observational Analysis for Rural Ethiopia." *PLoS ONE*. https://doi.org/10.1371/journal.pone.0160590.

Headey, Derek, Kalle Hirvonen, and John Hoddinott. 2018. "Animal Sourced Foods and Child Stunting." *American Journal of Agricultural Economics* 100 (5): 1302–19. https://doi.org/10.1093/ajae/aay053.

Hirvonen, Kalle, Daniel O. Gilligan, Jessica Leight, Heleene Tambet, and Victor Villa. 2023. "Do Ultra-Poor Graduation Programs Build Resilience against Droughts? Evidence from Rural Ethiopia." International Food Policy Research Institute (IFPRI). https://doi.org/10.2499/P15738COLL2.137000.

Huang, Ruirui, Chunli Yan, Yumei Tian, Beimei Lei, Dongqi Yang, Dan Liu, and Jun Lei. 2020. "Effectiveness of Peer Support Intervention on Perinatal Depression: A Systematic Review and Meta-Analysis." *Journal of Affective Disorders* 276 (November): 788–96. https://doi.org/10.1016/J.JAD.2020.06.048.

John, Anett, and Kate Orkin. 2022. "Can Simple Psychological Interventions Increase Preventive Health Investment?" *Journal of the European Economic Association* 20 (3): 1001–47. https://doi.org/10.1093/jeea/jvab052.

Kang, Yunhee, Sungtae Kim, Sisay Sinamo, and Parul Christian. 2017. "Effectiveness of a Community-Based Nutrition Programme to Improve Child Growth in Rural Ethiopia: A Cluster Randomized Trial." *Maternal and Child Nutrition* 13 (1): 1–15. https://doi.org/10.1111/mcn.12349.

Kim, Sunny S., Phuong Hong Nguyen, Yisehac Yohannes, Yewelsew Abebe, Manisha Tharaney, Elizabeth Drummond, Edward A. Frongillo, Marie T. Ruel, and Purnima Menon. 2019. "Behavior

Change Interventions Delivered through Interpersonal Communication, Agricultural Activities, Community Mobilization, and Mass Media Increase Complementary Feeding Practices and Reduce Child Stunting in Ethiopia." *Journal of Nutrition* 149 (8): 1470–81. https://doi.org/10.1093/jn/nxz087.

Korpe, Poonum S., and William A. Petri. 2012. "Environmental Enteropathy: Critical Implications of a Poorly Understood Condition." *Trends in Molecular Medicine* 18 (6): 328–36. https://doi.org/10.1016/j.molmed.2012.04.007.

Leight, Jessica, Daniel O. and Gilligan, Michael Mulford, Alemayehu Seyoum Taffesse, and Heleene Tambet. 2021. "Aspiring to More? New Evidence on the Effect of a Light-Touch Aspirations Intervention in Rural Ethiopia." https://ssrn.com/abstract=3981827.

Nagler, Paula, and Wim Naudé. 2017. "Non-Farm Entrepreneurship in Rural Sub-Saharan Africa: New Empirical Evidence." *Food Policy* 67: 175–91. https://doi.org/10.1016/j.foodpol.2016.09.019.

Orkin, Kate, Robert Garlick, Mahreen Mahmud, Richard Sedlmayr, Johannes Haushofer, Stefan Dercon, Paul Brimble, et al. 2023. "Aspiring to a Better Future: Can a Simple Psychological Intervention Reduce Poverty? *." https://www.socialscienceregistry.org/trials/996.

Singla, Daisy R., Brandon A. Kohrt, Laura K. Murray, Arpita Anand, Bruce F. Chorpita, and Vikram Patel. 2017. "Psychological Treatments for the World: Lessons from Low- and Middle-Income Countries." *Annual Review of Clinical Psychology* 13: 149–81. https://doi.org/10.1146/annurev-clinpsy-032816-045217.

Technical and Operational Performance Support (TOPS) Uganda Graduation Randomized Control Trial Associate Award. 2022. "Endline Report of the Resilience Food Security Activity Graduating to Resilience in Uganda, Cohort 1." Washington, DC: The TOPS Program. www.fsnnetwork.org.

Vicente-Serrano, Sergio M., Santiago Beguería, and Juan I. López-Moreno. 2010. "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index." *Journal of Climate* 23 (7): 1696–1718. https://doi.org/10.1175/2009JCLI2909.1.

Acknowledgements

I and my coauthors gratefully acknowledge funding for the impact evaluation by USAID under Cooperative Agreement No AID-FFP-A-16-00008. This work was undertaken as part of the CGIAR Research Program on Policies, Institutions, and Markets (PIM) led by the International Food Policy Research Institute (IFPRI). Additional funding support for this study was also provided by PIM. We also gratefully acknowledge project management support by Lucy Billings and research support by Gayathri Ramani.

I would also like to thank my promoter and advisor, Erwin Bulte, for his insightful comments and support; and also my co-promoter, co-author and colleague, Daniel Gilligan for encouraging, advising and working together with me on this exciting body of research.