United States Agency for International Development
Bureau of Democracy, Conflict and Humanitarian Assistance
Office of Food for Peace

Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA)

Endline Report

First Draft: August 31, 2021
Second Draft: September 29, 2021
Final Draft: December 21, 2021

Harold Alderman, Lucy Billings, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Alemayehu Seyoum Taffesse, and Heleene Tambet

Submitted by the
International Food Policy Research Institute (IFPRI)*

On behalf of
IFPRI, Hawassa University, Ambo University,
Teachers College at Columbia University, Johns Hopkins University,
World Vision, ORDA, CARE

* We 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. Michael Mulford, Chief of Party of the SPIR project, made many contributions to the impact evaluation design and shares intellectual credit for overall research design. Helen Verdeli and Paul Bolton contributed to the design of the maternal depression substudy and its integration into the overall study design. We thank Laterite, our survey data collection partner, for excellent work managing the conduct of the midline survey. For correspondence, contact Daniel Gilligan, International Food Policy Research Institute, 1201 Eye St., NW, Washington, DC 20005. email: d.gilligan@cgiar.org. phone: 202-862-8146.

Executive Summary

The Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA) in Ethiopia is a five-year project (2016-2021) supporting implementation of the fourth phase of the Productive Safety Net Programme (PSNP4) as well as providing complementary livelihood, nutrition, gender and climate resilience activities to strengthen the program and expand its impacts. The main objectives of SPIR are to enhance livelihoods, increase resilience to shocks, and improve food security and nutrition for rural households vulnerable to food insecurity. Activities under SPIR are organized into four Purposes: 1) livelihoods, 2) nutrition, 3) women's and youth empowerment, and 4) climate resilience. Across these Purposes, SPIR provides community-level programming, training of government staff involved in public service delivery at the woreda (district) and kebele (subdistrict) level, and targeted livelihood transfers.

IFPRI is conducting an experimental, quantitative impact evaluation of SPIR designed to measure the causal impact of multisectoral "graduation model" packages of livelihoods, nutrition, gender equity, and mental health interventions for improving outcomes in several domains, including livelihoods, food security, child nutrition, women's empowerment, mental health, and intimate partner violence (IPV). The impact evaluation uses a clustered randomized controlled trial (RCT) design with four intervention arms (three treatments and a control group) to test the relative effectiveness of these packages of interventions to improve outcomes for PSNP4 beneficiaries. This endline report of the impact evaluation presents evidence on the impact of three combinations of packages of core or enhanced gender-sensitive livelihood and nutrition activities on all primary and secondary outcomes for the evaluation after three years of implementation.[1] The endline survey for the impact evaluation was delayed due to the COVID-19 pandemic from mid-2020 and was conducted in February and April 2021, during which time a total of 3,812 households were interviewed out of the target of 3,996 households for the entire study sample.

SPIR Interventions

For learning purposes, the SPIR impact evaluation combined major core components and innovative new activities under Purpose 1 on livelihoods and Purpose 2 on nutrition, along with selected activities under Purpose 3 on gender and youth and Purpose 4 on climate resilience, into a study design of overlapping interventions to learn which combination of activities had the greatest impact and was most cost-effective at improving SPIR outcomes.[2] The randomized controlled trial (RCT) evaluates combinations of four interventions described below; L and N correspond to the primary SPIR interventions around livelihoods and nutrition, respectively, while L* and N* represent enhanced versions of these interventions.

Intervention L: SPIR livelihood activities: starting Village Economic and Social Associations (VESAs), financial literacy training, agriculture and livestock value chain development, home gardening and forage production

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

[2] A cost-effectiveness analysis related to the SPIR impact evaluation is underway, but is not included in this endline report.

Intervention L*:	SPIR livelihoods activities **plus** (i) social analysis and action (SAA) to improve women's access to markets, (ii) aspirations promotion activities in randomly selected kebeles (subdistricts), and (iii) targeted poultry or cash livelihood transfers
Intervention N:	SPIR nutrition activities: Nutrition Behavior Change Communication (BCC); WASH activities
Intervention N*:	SPIR nutrition activities **plus** (i) Timed and Targeted Counseling (TTC) (more intensive nutrition BCC), (ii) Community-based Participatory Nutrition Promotion (CPNP), (iii) male engagement in BCC, and (iv) Interpersonal Therapy in Groups (IPT-G) interventions for women screened for depression (provided after the midline survey), all supported by a Community Health Facilitator (CHF)

The main SPIR livelihood activities (L) under Purpose 1 were organized around VESA groups, which were used as a platform for trainings and other project activities around financial literacy, promotion of savings and credit use, agriculture and livestock value chain development (e.g., developing business skills and production skills), improving social capital, and catalyzing women's empowerment.

The SPIR health and nutrition package (N) included integrated nutrition social behavior change communication (SBCC) as well as water, sanitation and health (WASH) activities. Topics covered in SBCC included optimal infant and young child feeding (IYCF) practices, adolescent and maternal nutrition, diversified sources of nutritious foods, and utilization of health and nutrition services. The WASH component included providing support to village-level WASH management activities, limited support to improving sanitation infrastructure (water sources and latrines) and implementation of the community-led total sanitation and hygiene (CLTSH) approach.

The enhanced livelihoods (L*) interventions included all livelihoods activities as well as the following: ***Social Analysis and Action (SAA)*** enabled individuals and communities to explore and challenge social norms, beliefs, and practices around gender, including women's role in intrahousehold decision-making, mobility, choice of livelihood activities, and access to markets. The ***aspirations*** activities involved screening of short documentary films in the Amharic and Afaan Oromo languages designed to motivate individuals to undertake actions that will improve their well-being in the future. Livelihoods transfers were provided under L* in the form of a ***poultry or cash livelihoods packages*** for poor women (based on a baseline asset index), including either a poultry start-up package of US$200 worth of poultry start-up inputs and training or a one-time unconditional cash transfer of equivalent value.

The enhanced nutrition (N*) interventions included all nutrition activities as well as the following: In N* kebeles, BCC activities were delivered through a ***Timed and Targeted Counseling (TTC)*** model, including lessons on IYCF practices and adolescent and maternal nutrition. TTC was conducted at the household level, while the SBCC included in N activities was conducted at community events. TTC also encouraged men to support their wives in childcare and child feeding practices. ***Community-Based Participatory Nutrition Promotion (CPNP)*** sessions were two-week intensive feeding sessions for acutely malnourished children that include trainings for their mothers on complementary feeding and

caring practices. SPIR N* activities were coordinated by recruited and trained ***Community Health Facilitators (CHF)*** assigned to each of the N* kebeles to support the health extension workers (HEWs) in planning, coordinating, and facilitating health and nutrition activities. The CHF supported local Health Development Army (HDA) volunteers in conducting household-level counseling (involving both husband and wife) to promote ***male engagement in BCC*** related to IYCF and maternal nutrition using the TTC approach. In addition, male advocates conducted ***male engagement as men champions in men's groups*** to facilitate eight sessions designed to critically reflect on cultural gender norms and explore the positive and perceived negative effects of male involvement. Women screened for depression in the midline survey using the Patient Health Questionnaire (PHQ-9) tool were invited to enroll in 12-week ***Interpersonal Therapy in Group (IPT-G)*** sessions in each of the N* kebeles to address maternal depression.

Evaluation Design

The impact evaluation used a clustered RCT design to learn about the effect of different combinations of the SPIR interventions on the well-being of PSNP4 households: the livelihoods package (L), the nutrition package (N), and enhanced versions of each package (L* and N*, respectively). These packages were combined into multisectoral graduation model programs and randomized at the kebele level into four treatment arms: T1: L*+N*, T2: L*+N, T3: L+N*, T4: PSNP only. The evaluation sample comprises 192 kebeles in the Amhara and Oromia regions (Figure ES.1).

In addition, two supplemental interventions were cross-randomized across 94 kebeles in the experimental arms receiving the enhanced livelihoods intervention L*: a one-time poultry package and a one-time cash transfer. Both interventions were targeted to extremely poor households (or more specifically, the poorest 10 out of 18 sample households in each kebele). Half of the L* kebeles were randomly selected to receive cash transfers targeted to women in these extremely poor households, and half of the L* kebeles were randomly selected to receive poultry packages targeted to women in the extremely poor households.

Lastly, 50 percent of the L* kebeles (n=47) were randomly assigned to receive an aspirations treatment (also described in more detail below). Randomization of the poultry/cash intervention and randomization of the aspirations treatment were balanced such that approximately 25 percent of L* kebeles were assigned to either poultry only, poultry + aspirations, cash only, and cash + aspirations. The evaluation design and sample are summarized in Figure ES.1.

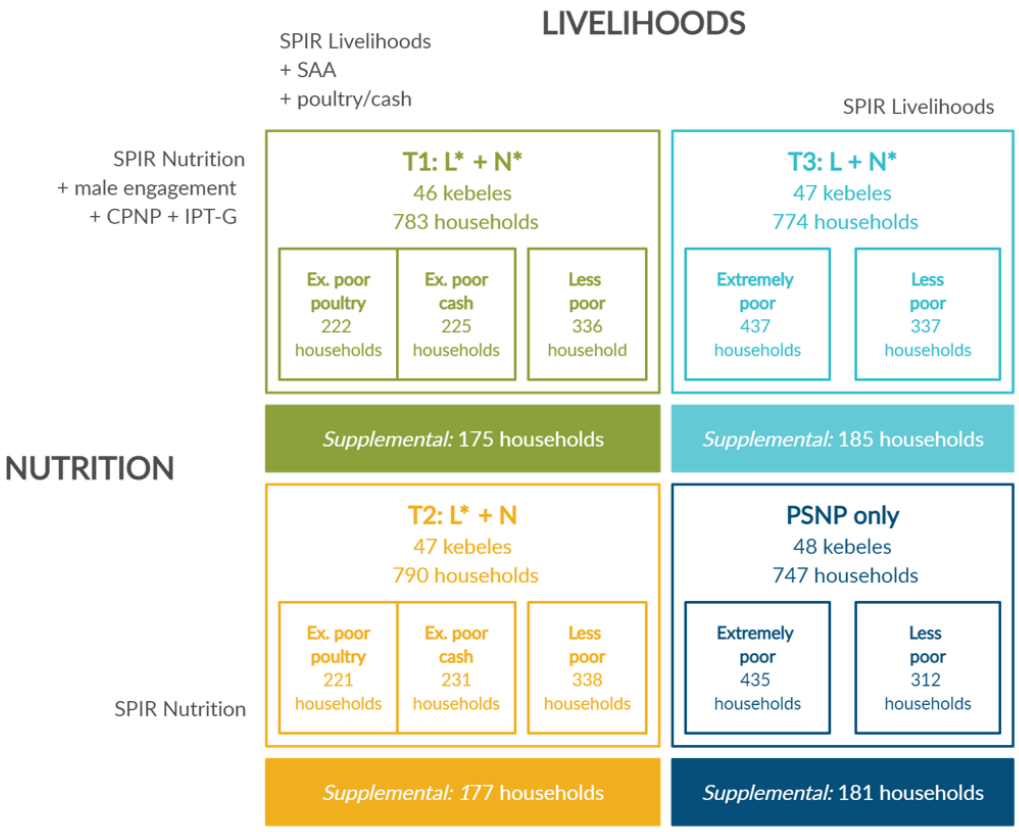
The Endline Survey and Household Panel Sample

As noted in the Baseline Report, the study takes place in 13 (original) woredas and 192 kebeles across the Amhara and Oromia regions of Ethiopia.[3] The baseline sampling process led to 3,314 households in the sample, or just over 17 households of PSNP4 beneficiaries with at least one child age 0–35 months in each kebele. The midline survey sample was designed to include all 3,314 baseline households to create a household level panel. Also, in order to assess the impact of SPIR on the diet and nutritional status of the

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

high-priority reference group of children under age 2 years, a supplemental sample of households was added in the midline survey. This made it possible to assess the impact of the SPIR treatment arms on children using a repeat cross-sectional analysis of children under age 2 years at midline. This sample was drawn from the original beneficiary lists used to draw the baseline household sample. The eligibility criteria for the supplemental sample were that the household had to have a member who is a PSNP4 beneficiary, the household had to have a child age 0–23 months, and the mother or primary female caregiver of that child had to be a household member. The midline survey target sample aimed to add 4 supplemental households in each kebele. The midline survey achieved an overall sample of 3,968 households.

Figure ES.1: SPIR experimental impact evaluation design[4]



The endline survey sought to re-interview all households in the midline sample. Of the 3,968 midline survey households, 3,812 were able to be located and interviewed at endline, leading to an attrition rate of 4.6 percent relative to the target sample. A large portion of the attrition (80 households) at endline was

[4] For this endline report, we have omitted the aspirations intervention, which was randomly assigned to half of the kebeles in L*, stratified by T1 and T2. Analysis of the impact of the aspirations intervention at midline showed no significant effects.

due to unrest in parts of Amhara that resulted in a decision not to visit four kebeles with reports of unrest. Among the remaining 104 attrited households, 55 had moved out of the study area, 26 were temporarily unavailable, and other households had either dissolved or were unavailable for the interview for other reasons; one household refused consent. In total, 3,706 primary female respondents and 2,465 primary male respondents were surveyed, and 1,064 new *endline index* children (under 24 months) were identified. In addition, the *midline* and *baseline index* children from previous survey rounds were measured for the anthropometrics module, if still part of the household.

The Changing Context and Exposure to Shocks

The woredas in Amhara and Oromia that are home to the SPIR project faced many significant shocks in the roughly 19 months that passed between the midline and endline surveys. In addition to COVID-19, the study area faced potentially significant pest infestations from fall armyworm and desert locusts. Many households also experienced weather, health and income shocks during this period. Finally, an armed conflict in Tigray region that began in November 2020 continued through the period of endline data collection and affected neighboring areas in Amhara region, including some in the SPIR operational areas. The endline survey collected information about exposure to these shocks in order to understand the extent to which they may have disrupted SPIR project activities or affected study outcomes.

Households reported numerous shocks related to the COVID-19 pandemic and related lockdowns, with most households reporting experiencing lengthy school closures, food shortages, and unemployment or income loss. Roughly half of respondents reported closures of markets and churches or mosques, and travel restrictions were also common. Pest infestations were a significant problem for some households in the survey. Desert locusts did not significantly affect households in Amhara, but roughly one in three households in Oromia lost cropland to desert locusts. Fall armyworm led to crop losses for 15 percent of households in Amhara but 60 percent of households in Oromia. Roughly half of households in the sample reported a significant drought event and similarly half reported a significant flood and associated erosion in the last 15 months. Despite the conflict in Tigray, its effects on study households, mostly in neighboring Amhara, were limited to a small share of households outside the four kebeles that study teams were unable to visit.

Balance tests for exposure to these shocks showed that the prevalence of each of these shocks was relatively balanced across study treatment arms, suggesting that these shocks are unlikely to lead to bias in estimated impacts of the SPIR program.

Experience with the SPIR program

The SPIR project expanded its reach since the midline study in 2019 for some program activities, consistent with the sequential nature of VESA discussions and other activities. In the main treatment arms (T1, T2, T3), reported participation in VESA groups is high but not universal, with 80–86 percent of households reporting that they have a member participating in a group.[5] The SPIR project also closed the

[5] Given that according to self-reports about 6 percent of households are likely receiving Permanent Direct Support and 4 percent are not part of either Public Works or Direct Support, a membership rate of in the range of 90–94 percent would be considered universal. See more detail in Section 6.1.

gender gap in participation in the period after the midline survey: VESA membership rates reported by men and women were roughly the same at endline. This was not the case at baseline and midline, when men participated at higher rates. Similarly, most respondents report high numbers of women in their VESA management committees, and approximately three-fourths of men and women say that their VESA has a female leader.

While participation in value chain trainings has risen considerably since the midline study and is now around a quarter of the sample, the share of those who are part of a producer marketing group has dropped to less than 8 percent. For general VESA discussions, three-fourths of members indicate attending meetings regularly (weekly or monthly) and the survey respondents show rather high recall rates for topics of interest to the program: IYCF practices, hygiene and sanitation, and gender dynamics within the household. Male engagement groups, which were part of the N* interventions in T1 and T3, became more active after the midline survey. At endline, roughly 40 percent of men in T1 or T3 kebeles had participated in male engagement groups in the last 12 months.

Impacts on Livelihood Outcomes

This report presents evidence about the medium-term effects of an integrated nutrition-sensitive social protection program on a range of livelihoods-related outcomes, measured approximately five years following the initiation of programming, and two years following one-time poultry and cash transfers targeted to the poorest households in the sample. The evidence suggests that SPIR had a range of positive effects, particularly on livestock-related production (notably for cash and poultry households) and for financial inclusion (for all households).

For extremely poor households (who were transfer recipients), we see evidence of persistent and large increases in livestock assets and engagement in livestock production: these effects are concentrated in poultry for poultry recipients, and are concentrated in non-poultry livestock for cash recipients. Extremely poor households also show evidence of substantial increases in membership in VESAs and the probability of reporting any savings. There is, however, no robust evidence of any increase in ownership of other durable goods (in a context in which ownership of these goods is rapidly increasing across the sample), or any increased consumption two years post-transfer.

For less poor households who did not receive transfers but were exposed to SPIR programming, we largely do not observe any substantial shifts in livestock assets or production. However, these households also show substantial increases in savings as well as some enhanced access to credit, and there is some weak evidence of improvement in housing characteristics.

Impacts on Nutrition and Child Welfare Outcomes

The SPIR program increased access to health services, including visits by HDA volunteers, BCC exposure, food demonstrations, and WASH. This despite the intervening COVID-19 strain on healthcare resources and the decreased mobility that the pandemic imposed. But this is almost literally a “glass half-filled” story; for no indicator of access to health services studied shows more than half the target population participating. Moreover, despite the TTC, key measures of IYCF, such as the age at which semi-solid or solid foods are introduced or child diet diversity, have not improved in the communities

where the intervention has been prioritized. Since proper complementary feeding is an essential element of nutritional care, this barrier likely contributes to the stagnating stunting rate. There are no indications of improvement in anthropometric outcomes. The SPIR project has, however, made modest advances in responding to underweight when it is identified. But, again, with child-weighing apparently infrequent, this improved service delivery does not fully cover the eligible population. Thus, identifying the gaps in coverage as well as improving the messaging on weaning appear to be ways that the intensified nutritional service delivery can achieve initial progress in improving nutritional outcomes.

Evidence on Mental Health, Relationship Dynamics, Agency, and Gender Equitable Attitudes and Roles

We investigate the impacts of the SPIR intervention on men's and women's mental health, marital dynamics, women's agency, and gender equitable attitudes and roles. We find that the SPIR intervention, and in particular T1 and T3, improved men's gender equitable attitudes and roles. These impacts are significantly different from T2, indicating that N* was needed for these transformative changes. However, we do not see any improvements in other dimensions of empowerment particularly related to women's decision-making or self-efficacy.

Impacts on mental health were mixed. We find no evidence that the SPIR intervention improved the mental health of the primary male or primary female at endline for the full sample or the less-poor sample. However, among the subsample of extremely poor households, the combination of T1 and poultry and the combination of T2 and cash transfers led to decreases in women's (but not men's) depressive symptoms, as measured by the probability of a woman reporting mild or moderate-to-severe depressive symptoms, and to decreases in reported unhappiness. The average impact of T2 across poultry and cash is large and significant as is the average impact of poultry across T1 and T2. Impacts of T2 (L*+N) are significantly different from T3 (L+N*), suggesting that L* is needed for improvements in women's mental health among the extremely poor subsample.

In terms of marital dynamics and intimate partner violence (IPV), we find no impacts of the SPIR intervention (T1, T2, T3) on these outcomes for either the full sample or subsample of less-poor households. For the subsample of extremely poor households, there is some indication that marital relations worsened, especially among poultry households where IPV has increased and the primary male is less likely to report that his spouse respects him.

Table of Contents

List of Tables

List of Figures

Acronyms

ANCOVA	Analysis of covariance
ASTE	Average standardized treatment effect
BCC	Behavior change communication
CAPI	Computer-assisted personal interview
CHF	Community health facilitator
CPNP	Community-based Participatory Nutrition Promotion
DFSA	Development Food Security Activity
DHS	Demographic and Health Survey
DID	Difference-in-difference
DS	Direct support
EC	Ethiopian calendar
HAZ	Height-for-age Z-score
HDA	Health Development Army
HEW	Health extension worker
IFPRI	International Food Policy Research Institute
IPT-G	Interpersonal Therapy in Groups
IPV	Intimate partner violence
IYCF	Infant and young child feeding
MDD	Minimum dietary diversity
MUAC	mid-upper arm circumference
ORDA	Organization for Rehabilitation and Development in Amhara
PCA	Principal components analysis
PIM	Program on Policies, Institutions, and Markets
PSNP	Productive Safety Net Programme
PSNP4	Fourth phase of the Productive Safety Net Programme
RCT	Randomized controlled trial

RuSACCO	Rural savings and credit cooperative
SAA	Social analysis and action
SPIR	Strengthen PSNP4 Institutions and Resilience
TDS	Temporary direct support
TTC	Timed and Targeted Counselling
VESA	Village Economic and Social Association
VSLA	Village Savings and Loan Association
WASH	Water, sanitation and health
WHZ	Weight-for-height Z-score

1 Introduction

The Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA) in Ethiopia is a five-year project (2016–2021) supporting implementation of the fourth phase of the Productive Safety Net Programme (PSNP4) as well as providing complementary livelihood, nutrition, gender and climate resilience activities to strengthen the program and expand its impacts.[6] Under funding from USAID's Bureau of Humanitarian Assistance and in close collaboration with the Government of Ethiopia, World Vision leads implementation of the SPIR DFSA, in partnership with the Organization for Rehabilitation and Development in Amhara (ORDA) and CARE. In the locations being studied in this report, SPIR targeted nearly 500,000 PSNP4 beneficiaries in 13 of the most vulnerable *woredas* (districts) in the Amhara and Oromia regions of Ethiopia. The main objectives of SPIR are to enhance livelihoods, increase resilience to shocks, and improve food security and nutrition for rural households vulnerable to food insecurity.

As part of the SPIR learning agenda, IFPRI has conducted an experimental, quantitative impact evaluation of SPIR designed to measure the causal impact of multisectoral "graduation model" packages of livelihoods, nutrition, gender equity, and mental health interventions for improving outcomes in several domains, including livelihoods, food security, child nutrition, women's empowerment, mental health, and intimate partner violence (IPV). The impact evaluation uses a clustered randomized controlled trial (RCT) design with four intervention arms (three treatments and a control group) to test the relative effectiveness of these packages of interventions to improve outcomes for PSNP4 beneficiaries. The gender-sensitive livelihood interventions draw from a set of activities including savings promotion, financial literacy, and agriculture and livestock value chain development, as well as enhanced livelihood activities to change norms around women's access to markets, promote changes in aspirations, and test the effectiveness of one-time targeted poultry or cash grants. The nutrition interventions draw from combinations of health worker trainings on nutrition behavior change communication (BCC) activities; water, sanitation and hygiene (WASH) activities; intensive household-level nutrition BCC; recuperative community-based nutrition promotion; and male engagement in BCC.[7]

This endline report presents the results of the final round of evaluation of the SPIR project. The endline survey was conducted in February–April 2021, three years after the 2018 baseline survey and roughly 19 months after the 2019 midline survey. The endline survey was delayed from June 2020 because the COVID-19 pandemic made it impossible to safely conduct in-person interviews until early 2021. In addition, the conflict in Tigray region and surrounding areas temporarily disrupted delivery of SPIR services in some study sites and made it impossible to conduct the endline survey in four study *kebeles* (communities). Despite these significant challenges, most components of the SPIR interventions, including most seasonal PSNP4 transfers, continued to be delivered during the period leading up to the endline survey, despite some disruptions. Thus, while COVID-19 and the Tigray conflict interfered with program delivery in several study locations, this study is still able to evaluate the impact of the three SPIR graduation model intervention packages, which were introduced through experimental assignment in 192 kebeles in early 2018, on the project's primary and secondary outcomes. The SPIR interventions were

[6] With more than 7 million beneficiaries, PSNP4 is the Government of Ethiopia's flagship safety net program. It provides food or cash transfers targeted to poor households in the form of payments for seasonal labor on public works or as direct support to households whose primary income earners are elderly or disabled.

[7] In addition, after the midline survey, two study arms introduced group-based therapy for women screened for depression, in part as a strategy to increase their ability to improve the nutritional status of their children.

delivered in a phased rollout, with some components including poultry and cash livelihood transfers delivered only months before the midline survey and other components including male engagement and linkages to health services strengthened after the midline. As a result, the midline survey report presented preliminary estimates of SPIR at an intermediate stage. This endline report presents the impact of the full effect of three years of phased interventions and contributes, along with monitoring and implementation studies conducted by the implementation team, to the culminating lessons from the SPIR learning agenda.[8] This report briefly summarizes the SPIR interventions, the evaluation study design, and the process of data collection for the endline survey before presenting information on beneficiaries' experience with the program, exposure to shocks, and then evidence of program impact on primary and secondary outcomes in livelihoods, nutrition, and other dimensions of well-being (e.g., control over decision-making, mental health, and intimate partner violence). More details on the study design, SPIR interventions and the program context can be found in the SPIR Baseline Report (Alderman et al. 2019) and SPIR Midline Report (Alderman et al. 2020a).

The SPIR impact evaluation study is designed to contribute evidence to several knowledge gaps and active debates on a number of important topics around social protection and promising approaches to poverty alleviation, including the effectiveness of "graduation model" programs; the promise of poultry value chains as a female-friendly investment for the poor; poverty and mental health; and the effect of cash transfers on intimate partner violence (IPV). Graduation model programs complement targeted cash or food transfers with multisectoral investments in asset building, income generation, and access to markets to provide a "big push" to promote sustained poverty alleviation. Graduation model programs are being tested in many countries following the publication of results from a six-country study of programs styled after BRAC's Targeting the Ultra-Poor (TUP) program in Ethiopia, Ghana, Honduras, India, Pakistan, and Peru (Banerjee et al. 2015) and related papers on the same interventions (Banerjee et al. 2016; Bandiera et al. 2017; Banerjee et al. 2019). These programs led to substantial improvements in household economic outcomes, including consumption, food security, assets, financial inclusion, labor supply, and income as well as some measures of mental health related to stress. The Ethiopia SPIR graduation approach is distinct from these BRAC programs in two important ways. First, SPIR includes substantial, integrated programming designed to improve nutrition and women's empowerment. The BRAC programs showed no effects on women's empowerment, and nutrition outcomes were not assessed because they were not an objective of those programs. The impact evaluation of the SPIR project is the first study we are aware of that tests the impact of a nutrition-sensitive graduation program. Second, SPIR provides smaller resource transfers than many graduation model programs, the largest being a one-time poultry or cash grant of US$200 (Ethiopian birr equivalent value), which is roughly one-sixth of the value of livelihood transfers provided in the BRAC study. Instead of larger transfers, SPIR provides programming that prioritizes trainings and information offered through Village Economic and Social

[8] This endline report plays a critical role in the SPIR learning agenda's implementation of USAID's methodology for collaborating, learning and adapting (CLA) to improve project effectiveness. The CLA approach extends traditional M&E practices and learning-based impact evaluations to develop a more integrated approach to communication between the project implementation, M&E, and research teams, in order to provide feedback to improve project delivery and effectiveness during the implementation period. The lessons from this impact evaluation will also inform the next phase of USAID investments in Ethiopia through the upcoming Resilience Food Security Activities.

Associations (VESA) and other local institutions, as well as linking PSNP clients to improved public services, especially in health, and access to finance through local institutions.

Poultry has gained renewed attention as a promising value chain for women because it is an asset that is widely accessible to women, has low start-up costs, and provides a good source of nutritious animal-source foods for children in chicken meat and, especially, eggs. In 2106, Bill Gates promoted investment in chickens to help increase incomes for poor women (https://www.gatesnotes.com/Development/Why-I-Would-Raise-Chickens). In response, Chris Blattman suggested that large cash grants of the kind provided by Give Directly (Haushofer and Shapiro 2016) may be more effective at improving outcomes for more women, given heterogeneity in their needs and capacity to raise chickens (https://www.vox.com/the-big-idea/2017/3/14/14914996/bill-gates-chickens-cash-africa-poor-development). The SPIR study contributes evidence to this debate by testing a promising poultry start-up package (including improved breed chickens sourced from EthioChicken pullet growers) and benchmarking this package against cash transfers of similar value.[9]

A growing literature has addressed the relationship between poverty and stress, with the hypothesis that poverty-induced stress reduces the capacity for decision-making enough to contribute to a mental-health-induced poverty trap (Ong, Theseira and Ng 2019). Related evidence shows that the prevalence of various mental health problems is greater among the poor and that positive income shocks reduce depression (Christian, Hansel and Roth 2019). This has implications for investments in children, given evidence that maternal depression postpartum and in the first year of life, for example, is associated with low child nutrition outcomes (Nguyen et al. 2018; Black et al. 2009; Wachs et al. 2009). More recently, researchers have tested whether interventions to reduce depression in low-income countries with few mental health services can improve mental health and various welfare outcomes. Haushofer, Mudida and Shapiro (2019) show that large cash transfers are more effective than cognitive behavioral therapy at improving mental health and measures of life satisfaction. Baranov et al. (2020) show that psychotherapy for prenatally depressed mothers in Pakistan led to persistent improvements (seven years later) in women's mental health, financial empowerment, and parental investments in children. Angelucci (2020) finds positive impacts of eight months of psychiatric care in India on child human-capital investment and indebtedness, but no impacts on earnings, time use, consumption or hygiene. Here, we test whether the transfers and other SPIR interventions reduce the prevalence of depression in the endline survey.[10]

Also, recent research has shown that cash transfers and other modalities reduce IPV (Hidrobo, Peterman and Heiss 2016), including when combined with participation in nutrition BCC sessions in Bangladesh (Roy et al. 2019) and for second and later wives in polygamous households in Mali (Heath, Hidrobo and

[9] See McIntosh and Zeitlin (2018), for example, who conducted an experiment to benchmark a nutrition intervention in Rwanda against large cash grants.

[10] In addition, two treatment arms in the study provided a group therapy intervention (Bolton et al. 2007), Interpersonal Therapy in Groups (IPT-G) to women suffering from mild-to-severe depression using the PHQ-9 depression screening tool during the midline survey. We attempted to revisit these women for a follow-up survey in March 2020 after the 12-week group therapy session was completed, but this survey was canceled due to the COVID-19 pandemic. A phone survey of women in this sample only reached roughly one third of the target sample, or 166 women. We do not examine the impact of that intervention in this report, where the focus is on measuring the impact of the integrated packages of interventions provided to households across the experimental treatment arms.

Roy forthcoming). Here, we will test whether the SPIR treatment arms provide a comparable reduction in IPV.

This endline report is organized as follows. Chapter 2 provides an overview of the SPIR interventions implemented since the baseline survey. Chapter 3 describes the impact evaluation design. Chapter 4 describes the endline survey data collection. Chapter 5 reports on delivery of program components and participants' program exposure. Chapter 6 summarizes the prevalence of COVID-19 and pest-related shocks and discusses how these might affect impact estimates in this experiment. Chapter 7 presents impacts on livelihood outcomes. Chapter 8 presents impacts on nutrition and child welfare outcomes. Chapter 9 presents impacts on mental health, relationship dynamics, female agency, and gender equitable attitudes. Chapter 10 concludes.

2 The SPIR interventions

Activities under SPIR are organized into four Purposes: 1) livelihoods, 2) nutrition, 3) women's and youth empowerment, and 4) climate resilience. In each of these Purposes, SPIR provides community-level programming, training of government staff involved in public service delivery at the woreda (district) and kebele (subdistrict) level, and targeted livelihood transfers. Resource transfers received by SPIR participants come primarily from transfers received from the PSNP4. As an analogue to the targeted livelihood transfers received by the poorest PSNP4 beneficiaries, in 2019 SPIR delivered a poultry kit that included improved breed chickens from EthioChicken or an equivalent one-time cash grant of US$200 (birr equivalent value). Most other benefits of the SPIR project took the form of improved public service delivery and trainings to promote learning and support for community-level groups.

For learning purposes, the SPIR impact evaluation combines major core components and innovative new activities under Purpose 1 on livelihoods and Purpose 2 on nutrition, along with selected activities under Purpose 3 on gender and youth, into a study design of overlapping interventions to learn which combination of activities has the greatest impact and is most cost-effective at improving SPIR outcomes.[11] The RCT evaluates combinations of four interventions described below; L and N correspond to the primary SPIR interventions around livelihoods and nutrition, respectively, while L* and N* represent enhanced versions of these interventions.

Intervention L:	SPIR livelihood activities: establishing Village Economic and Social Associations (VESAs), financial literacy training, agriculture and livestock value chain development, home gardening and forage production
Intervention L*:	SPIR livelihoods activities **plus** (i) social analysis and action (SAA) to improve women's access to markets, (ii) aspirations promotion activities in randomly selected kebeles, and (iii) targeted poultry or cash livelihood transfers
Intervention N:	SPIR nutrition activities: Nutrition behavior change communication (BCC); WASH activities
Intervention N*:	SPIR nutrition activities **plus** (i) Timed and Targeted Counseling (TTC) (more intensive nutrition BCC), (ii) Community-based Participatory Nutrition Promotion (CPNP), (iii) male engagement in BCC, and (iv) Interpersonal Therapy in Groups (IPT-G) interventions for women screened for depression (provided after the midline survey), all supported by a community health facilitator (CHF)

The main SPIR livelihood activities (L) under Purpose 1 are organized around VESAs. The majority of VESAs were formed in the first two years of the program, soon after the baseline survey. These groups

[11] A cost-effectiveness analysis related to the SPIR impact evaluation is underway, but is not included in this endline report.

were then used as a platform for trainings and other project activities around financial literacy, promotion of savings and credit use, agriculture and livestock value chain development (e.g., developing business skills and production skills), improving social capital, and catalyzing women's empowerment. VESAs include men and women (often the husband and wife from the same household). See the Baseline Report for more details.

The SPIR health and nutrition package (N) includes integrated nutrition social behavior change communication (SBCC) as well as WASH activities. Topics covered in SBCC include optimal infant and young child feeding (IYCF) practices, adolescent and maternal nutrition, diversified sources of nutritious foods, and utilization of health and nutrition services. The WASH component includes providing support to village-level WASH management activities, limited support to improving sanitation infrastructure (water sources and latrines), and implementation of the community-led total sanitation and hygiene (CLTSH) approach in which health extension workers (HEWs) and volunteers in the Health Development Army (HDA) are trained to foster improvement in community sanitation and hygiene and reductions in the practice of open defecation. Lessons from these trainings are provided at VESA groups, public works sites, and school and religious gatherings. Nutritious food preparation demonstrations are also offered in all SPIR implementation kebeles, but the coverage and the frequency differ according to the responsibilities of the different HEWs and HDA agents.

In order to inform the presentation of results in this report, we briefly summarize each of the additional enhanced components of delivery in L* and N*:[12]

2.1 L* enhanced livelihood activities

Social analysis and action: In the SPIR program, social analysis and action (SAA) was used to enable individuals and communities to explore and challenge social norms, beliefs, and practices around gender and nutrition that shape their lives. SAA is a community-led social change strategy that addresses constraints on women's role in intrahousehold decision-making, mobility, and choice of livelihood activities, as well as restrictions on access to markets that derive from cultural and social norms.[13]

Aspirations: IFPRI researchers and others have conducted experiments in Ethiopia showing substantial and long-lived effects of an aspirations intervention based on short documentary films designed to motivate individuals to undertake actions that will improve their well-being in the future.[14] These documentaries, in the Amharic and Afaan Oromo languages, provide true, inspirational stories about the returns to hard work and the benefit of aiming high, and constitute the aspirations intervention within L*. The experimental design randomized access to the aspirations intervention to households in half of the kebeles within the L* design. Aspirations videos were screened in selected kebeles in December 2018.

Poultry or cash livelihoods packages for poor women: Livelihoods transfers were provided under L* in the form of poultry start-up packages or one-time unconditional cash transfers of equivalent value. These

[12] See the Baseline Report (Alderman et al. 2019) for more details on these interventions.
[13] In practice, implementation of the SAA interventions was not as intensive as originally planned.
[14] See Bernard et al. (2017, 2019) and Taffesse and Tadesse (2017) for the results of recent aspirations experiments conducted in Ethiopia.

transfers were targeted toward the poorest SPIR project participants. This project component was designed to mirror the PSNP4 targeted (rationed) livelihoods transfer, although targeting and programming of these transfers in the SPIR project differs from the PSNP4 approach. In the SPIR project, these livelihood transfers were given to women in the poorest households (based on a baseline asset index) either as a transfer of US$200 in cash (birr equivalent) or as US$200 worth of poultry start-up inputs and training. These livelihood packages were provided in April 2019.

2.2 N* enhanced nutrition activities

Timed and Targeted Counseling: In the N* treatment, BCC activities were organized under an intervention model referred to as Timed and Targeted Counseling (TTC). Under TTC, community HEWs and HDA leaders provided lessons in health posts at the community level and through household visits on IYCF practices and maternal nutrition. As with the primary SPIR Purpose 2 trainings, topics included diversifying diets into sources of nutritious foods (including cooking demonstrations) and promotion of use of health and nutrition services. TTC differed from the SBCC offered in all SPIR communities in that it was conducted at the household level rather than at larger group or community events. TTC also encouraged men to support their wives in childcare and child feeding practices. TTC endeavored to conduct 11 home-visits in the first 1,000 days, including 3 visits during pregnancy.

Community-based Participatory Nutrition Promotion: Community-based Participatory Nutrition Promotion (CPNP) sessions were two-week intensive feeding sessions for mildly to severely underweight children that included hands-on trainings for their mothers on complementary feeding and caring practices. The sessions, as well as many other N* community-level nutrition activities, were led by CHFs.

Male engagement in nutrition BCC: Household-level counseling (involving both husband and wife) related to IYCF and maternal nutrition were conducted using the TTC approach to support shared decision-making. Because TTC conducted nutrition trainings directly in the community, it was more intensive than the SBCC provided in the SPIR nutrition package (N). SPIR hired CHFs for **each of the N* kebeles** to provide supportive supervision and monitoring of HDA volunteers in their household-level counseling and other community health activities. These CHFs also supported training of CPNP for nutritious food preparation at Growth Monitoring and Promotion sessions.

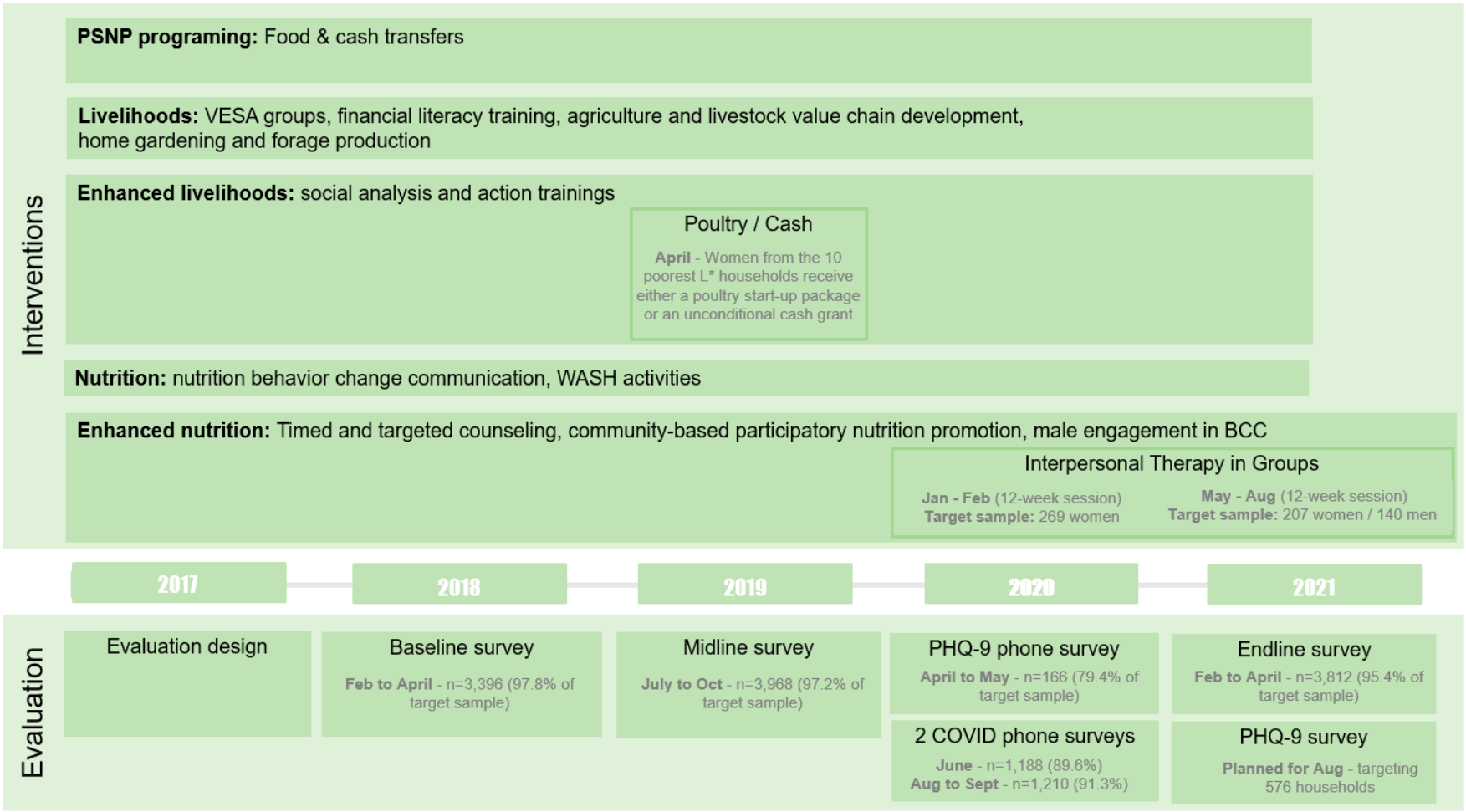
Male engagement through men champions, men's groups, and public awareness campaigns: Trained male advocates facilitated eight sessions for newly established men's groups in **each of the N* kebeles**. The sessions provided an opportunity for men to critically reflect on cultural gender norms and explore the positive and perceived negative effects of male involvement, seeking to better understand how gender inequity affects the lives of women, children and men. This intervention was scaled up over time. Prior to the midline, SPIR formed one pilot men's engagement group per N* kebele, which may not have had any sample households invited. After the midline, this was scaled up, targeting 7 to 10 men's engagement groups in each N* kebele (supported by newly hired men's engagement facilitators based at the kebele level). For capacity reasons, this scaling took place in half of N* kebeles (randomly selected) from December 2019 to March 2020 and began in the other half of the N* kebeles after April 2020. Men in the sample households were intentionally invited to participate in these groups, along with both PSNP and non-PSNP community members.

Interpersonal Therapy in Groups (IPT-G) to address maternal depression: Depression of female and male respondents was assessed in all three survey rounds (baseline, midline, endline) using the Patient Health Questionnaire (PHQ-9) tool to measure depressive symptoms. Women whose screening indicated mild-to-severe depression (PHQ-9 score ≥8) during the midline survey (July–October 2019) were invited to enroll in 12-week IPT-G sessions (approximately 5 to 8 women in each group) starting in December 2019 in **each of the N* kebeles.** The sessions were followed up by another round of data collection targeting only those who had scored ≥8. After the endline survey, both female and male respondents who had a PHQ-9 score at or above 8 were invited to participate in therapy groups, formed for women and men separately and lasting for 12 weeks again. This was then followed up by another survey round to measure the depression scores post-treatment, although it was only possible to conduct the survey in Oromia due to unrest in Amhara at the time of the survey.

Community health facilitators: SPIR N* activities were coordinated by project hired and trained community health facilitators (CHFs). SPIR assigned a CHF to each of the N* kebeles in order to support the HEWs in planning, coordinating, and facilitating health and nutrition activities, and to help the HDA leaders in their role as facilitators in VESA groups, CPNP sessions, and TTC visits.

For the timeline of the main SPIR activities as well as surveys conducted under SPIR learning agenda, see Figure 2.1.

Figure 2.1: Timeline of SPIR interventions and evaluation activities



3 Evaluation design

The impact evaluation used a clustered randomized controlled trial (RCT) design to learn about 1) the impact of three combinations of SPIR livelihoods (L) and nutrition (N) activities and enhanced livelihood (L*) and nutrition (N*) activities; 2) the effect of adding an aspirations intervention to the enhanced livelihoods activities; and 3) the differential effect on the poorest 10 out of 18 households in each community of providing an adult female household member with a one-time transfer of a poultry start-up package or a cash transfer of equivalent value. 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.[15] The Pre-analysis Plan with a complete list of hypotheses to be tested is available in the AEA registry (Alderman et al. 2020b).

We do not evaluate the impact of the aspirations intervention (topic 2, above) in this endline report. We examined the impact of the aspirations intervention, which was introduced early in the project, in the midline report and found no impacts on any of the project primary or secondary outcomes. We expect that the intervention was too "light touch" to have any sustained effects after having no short-term effects, so we dropped plans for analysis of the aspirations intervention in this endline report. We have kept the description of the aspirations intervention in this chapter because it was part of the original study design.

3.1 Experimental design

The impact evaluation design compares combinations of these activity packages by randomly assigning kebeles to one of the following four intervention arms (see Figure 3.1):

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

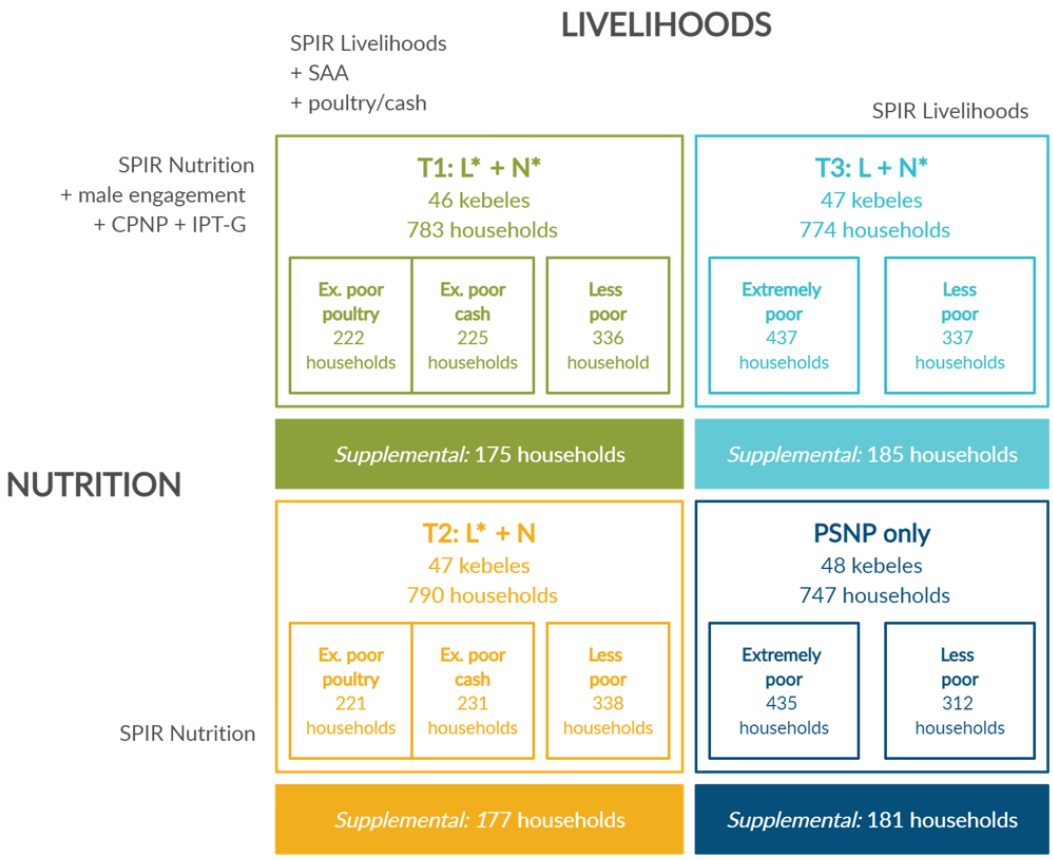
Consistent with the graduation model design, the treatment arms in the experiment are integrated combinations of L, L*, N, and N*. The evaluation tests the relative effectiveness of those combinations. At endline, as at midline, we examine the impact of the fully enhanced nutrition and livelihood models that include SAA, aspiration activities, livelihood transfers, TTC, CPNP, and male engagement in BCC against the control (T1 v C).[16] We also measure the impact of adding TTC, CPNP, and male engagement (T2 v C) or adding only SAA, aspirations, and poultry start-up kits or cash grants (T3 v C) to the main SPIR model. These last two comparisons tell us the effect of the combined L+N interventions when SAA,

[15] https://www.socialscienceregistry.org/trials/8281.

[16] The IPT-G group therapy intervention was introduced after the midline survey to women in N* kebeles (T1 or T3) who were screened for mild-to-severe depression at midline. As noted earlier, we do not examine the impact of that intervention on this subgroup in this report, where the focus is on measuring the impact of the integrated packages of interventions provided to households across the experimental treatment arms.

aspirations, and livelihood transfers are added or when male engagement and IPT-G are added on top of the main integrated SPIR intervention, respectively.

Figure 3.1: SPIR experimental impact evaluation design



3.1.1 Cash benchmarking of the poultry livelihood intervention

The PSNP4 includes a Livelihood Transfer component, in which a fraction of the poorest PSNP beneficiaries, identified through community targeting, also received an asset transfer designed to promote business development. The SPIR project implemented enhanced livelihood transfers in the form of a poultry start-up package or unconditional cash grant to mirror this feature of the PSNP4. These livelihood

packages were provided to the 10 poorest out of 18 households in each L* kebele in the SPIR study, with targeted households selected according to an asset index developed from the baseline data.[17]

The poultry start-up package was selected in part because of the availability of the promising EthioChicken breeds (Sasso breed), which are known to be highly productive in the adverse, poor conditions prevalent in rural Ethiopia. These chickens perform well under a semi-scavenging feeding system and are not disease-prone. In addition, this approach to providing the livelihood transfer allows the SPIR impact evaluation to provide evidence in the current debate on the potential of poultry value chains to improve outcomes for poor women, as described in Chapters 1 and 2.

The SPIR endline data provide an opportunity to contribute evidence to this debate, comparing valuable EthioChicken poultry start-up packages to an equivalent cash grant within the context of an integrated graduation model social protection program. Although several other interventions are taking place in the SPIR study, the randomized assignment of poultry or cash grant packages to the poorest households in the L* treatment arms make it feasible to identify the impact of either livelihood package, when combined with related complementary interventions that are also experimentally assigned to poultry and cash grant recipients.

Figure 3.1 shows the structure of the experimental impact evaluation design, including the allocation of the endline sample across the extremely poor and less poor subgroups in each arm, the poultry and cash interventions in T1 and T2, and the supplemental midline sample in each arm. The aspirations intervention is omitted from the figure for simplicity. Adding it would split the boxes for extremely poor–poultry, extremely poor–cash and less–poor in half again, with one of each sub-box randomly assigned to the aspirations intervention as well, in T1 and T2.

3.1.2 Substudy on depression

Recent evidence has identified maternal depression, particularly in the postpartum period, as a potentially important determinant of child growth and development outcomes in low-income settings. In one study, infants of mothers with depressive symptoms had 2.17 higher odds of being stunted (95% CI: 1.24, 3.81) than did infants of mothers with few symptoms (Wachs et al. 2009). A prior study showed that reducing depression can reduce child stunting by 27 percent (Black et al. 2009). It has also been shown that an interpersonal therapy in groups (IPT-G) intervention was highly effective at reducing depression in Uganda (Bass et al. 2006; Bolton et al. 2003), and that an IPT-G intervention significantly reduced depression for adolescent girls (but not adolescent boys) living in internally displaced persons camps in war-affected northern Uganda (Bolton et al. 2007).

Addressing maternal depression is consistent with the SPIR project objectives under Purpose 3 to strengthen the capacity of women to improve outcomes for themselves and their families. Based on this evidence, we added a substudy on the impact of IPT-G on maternal depression and child nutrition

[17] The asset index was constructed using ownership data on more than 30 asset categories, including consumer durables, productive assets, livestock and land. The asset index was constructed using principal components analysis, which reduces the influence of ownership of assets in the index that are shown to be highly correlated with ownership of other assets (Filmer and Pritchett 2001).

outcomes, collaborating with psychologists Lena Verdeli (Teachers College, Columbia University) and Paul Bolton (Johns Hopkins University) for the substudy related to maternal depression. As part of this study, we also collected data on depression for the primary male respondent, to learn about the correlation of adult female and male depression in the household and how the mental health of both females and males affects household welfare 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) and the primary male respondent (usually, her partner) were 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, with severity classified by intervals of: none (0), minimal (1–4), mild (5–9), moderate (10–14), moderately severe (15–19), and severe (20–27). The PHQ-9 tool questions were also asked of the primary female respondent and primary male respondent in the baseline survey. The baseline scores, available in the Baseline Report, were substantially lower than reported for this tool elsewhere in Ethiopia (Adewuya et al. 2006; Gelayea et al. 2013). One reason for lower average scores is that the recall time period was erroneously shortened from two weeks to one week in the baseline survey. This would have systematically lowered scores, although it is not known whether this was the main reason for the unexpectedly low scores. As a result, in the midline and endline surveys, we corrected the recall period to two weeks. We also undertook a more careful training of the survey team at midline to improve the approach to enumeration for these sensitive questions. The endline survey team comprised many of the enumerators from the midline survey, and the training was repeated, though without the participation of the research team due to COVID-19-related travel restrictions. Mean midline depression scores were substantially higher than at baseline, as reported in the midline report (Alderman et al. 2020a). Mean endline PHQ-9 scores are reported in Chapter 9, along with estimates of the impact of SPIR on mean severity of depression symptoms at endline.

3.2 Implementing the SPIR evaluation study design

The experimental study design involves randomly assigning 192 kebeles into the four treatment arms (see the Baseline Report for details). All kebeles assigned to L* (T1 and T2) (n=96) received the SAA intervention. In addition, study households in half of the L* kebeles (n=48) received the aspirations intervention, stratified across T1 and T2. The livelihood transfer of the poultry start-up package or unconditional cash grant was provided to the 10 poorest households in the L* study communities. Randomization of poultry or cash grants was done at the kebele level, with women in one half of the L* kebeles randomly assigned to receive the poultry-start up package, and women in the remaining half of the L* kebeles (48) to receive the cash grant. Randomization of the poultry/cash livelihood intervention was done using rerandomization, in which more than 100 randomized assignments were conducted and one such randomized assignment was randomly selected from among those that demonstrated balance across the aspirations and non-aspirations kebeles in the L* intervention arms.

All households in kebeles assigned to N* (T1 and T3) received both targeted male engagement through TTC household visits that specifically include men, together with their spouses, the CPNP promotion, and

the pilot men's engagement groups (one group per kebele prior to midline), and public awareness campaigns. In addition, women who are mothers of children age 0–35 months and their male partners were screened for depressive symptoms during the midline survey and women with mild depression or worse (PHQ-9≥8) were invited to enroll in IPT-G in December 2019. The same target women and men were screened for depression again in the endline survey. A second round of IPT-G was implemented after the endline survey, starting in June 2021.[18] The intervention to promote male engagement was further scaled up after the midline, with primary male respondents in half of the N* kebeles invited to participate in new men's groups from December 2019 to March 2020, and the other half scaled up men's engagement groups starting in April 2020.

3.3 Household eligibility and sampling

3.3.1 Kebele and household eligibility criteria

As noted in the Baseline Report, the study takes place in 13 (original) woredas across the Amhara and Oromia regions of Ethiopia.[19] In designing the study sample, we began with a list of all kebeles in which the PSNP operated in these woredas and dropped kebeles where VESAs had already been formed, the first step in implementation of the SPIR project. A total of 196 kebeles (115 in Amhara and 81 in Oromia) remained as a part of the study. Because the implementation team had started VESA groups in some of the more accessible kebeles that were ultimately dropped from the sample, the kebeles that were retained for the sample were often very remote. 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. The baseline sampling process led to 3,314 households in the sample, or just over 17 households of PSNP4 beneficiaries with at least one child age 0–35 months in each kebele.

3.3.2 Balance of baseline covariates

The baseline report presented evidence that key household covariates were balanced at baseline when comparing across the four main treatment arms. The midline report presents additional evidence of balance in baseline covariates in the achieved midline sample for the various intervention sub-arms and for the two subsamples of interest (poor households who were eligible for asset transfers in the L* arms, and nonpoor households who were not eligible for asset transfers). The midline balance tables showed no robust evidence of any meaningful difference in baseline household characteristics. There was also considerable balance for livelihood and nutrition outcomes. There was some imbalance in baseline IPV and in baseline PHQ-9 scores for men in the midline sample. Based on this evidence of substantial balance in the achieved sample at midline and relatively low attrition in the sample between the midline

[18] An in-person follow-up survey was conducted to gather the PHQ-9 depression scale again after completion of the 12-week IPT-G sessions in September 2021. As of the time of this report, the Tigray conflict had spilled into Amhara, making it unsafe to conduct this PHQ-9 follow-up survey there, so the follow-up survey is only being conducted in Oromia.

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

and endline surveys (as described below), we did not examine balance on baseline covariates in the achieved endline sample in this report, with one exception. The study kebeles are quite remote and there was some concern among implementation partners that SPIR treatment kebeles were more remote than control kebeles. We investigated the balance in the sample with respect to remoteness, using three measures: distance from the household to the nearest town, distance from the household to the kebele office, and travel time to get to the main market. See the balance table below and the table of summary statistics. The data confirmed that the kebeles are remote, being roughly 75 minutes travel time from the nearest moderately sized town. However, we found no statistically significant difference in either distance or travel time between any of the treatment and control arms, suggesting no potential bias due to differences in remoteness.

In general, we conclude that the randomization was effective and that the sample was balanced across key dimensions of demographics, livelihoods, and nutrition outcomes at baseline. However, there is some evidence of imbalance in baseline depression scores for some intervention sub-arms.

3.3.2 Supplemental midline sample

In order to assess the impact of SPIR on the diet and nutritional status of the high priority reference group of children under age 2 years, a supplemental sample of households was added in the midline survey. This makes it possible to assess the impact of the SPIR treatment arms on children using a repeat cross-sectional analysis of children under age 2. This sample was drawn from the original beneficiary lists used to draw the baseline household sample. The eligibility criteria for the supplemental sample were that the household had to have a member who is a PSNP4 beneficiary, the household had to have a child age 0–23 months, and the mother or primary female caregiver of that child had to be a household member. The target sample aimed to add 4 such households in each kebele to add another 768 households to the sample, for 4,082 households overall in the midline sample. Balance tests conducted at midline showed that the supplemental sample households are very similar to the panel sample households: supplemental sample households were somewhat less likely to be married, had older primary females, and were less likely to have improved roof materials on their homes.

3.4 Empirical strategy

In this endline report, impacts of the SPIR project are estimated on the baseline and endline data using analysis of covariance (ANCOVA) models controlling for the baseline outcome variable in the regression when baseline data on the outcome are available. When the outcome was not measured at baseline, we use simple difference models estimated on the midline data. The ANCOVA model provides better model fit and more statistical power than a difference-in-difference (DID) model, particularly when the autocorrelation in the outcome is low, or when the current level of the outcome is not highly determined by the past-period level of the outcome (McKenzie 2012). This is likely to be true for most of the outcomes we study. Using the ANCOVA model, we will estimate intent to treat (ITT) effects in accordance with assigned treatment.

For most outcomes reported (e.g., livelihoods, food security, women's diets, depression and IPV), we estimate impacts of SPIR on the full sample, including the supplemental midline sample, to be able to measure the average treatment effect for the full sample. For households from the midline supplemental

sample and those with missing baseline outcome data, the regression models include a dummy variable indicating missing baseline data, so that all observations can be included in the ANCOVA model.

For these outcomes, we estimate three models:

1) the average treatment effect of each treatment (T1, T2 and T3) over all its cross-randomized subtreatments (the 'pooled' treatment effect of that treatment) on the full endline sample, including midline supplemental sample households;
2) the impact of each treatment and the cross-randomized subtreatments (poultry/cash in T1 and T2) on the sample of households identified as "extremely poor" at baseline; and
3) the average treatment effect of each treatment (the pooled treatment effect) on the sample of households identified as "less poor" at baseline.[20]

Note that we do not account for the subrandomization of the aspirations intervention in T1 and T2 in these estimates. Results in the midline report show null effects of the aspirations treatment across virtually all outcomes. This suggests that separately accounting for exposure to the aspirations intervention in the T1 and T2 samples would not change the average treatment effect of T1 and T2 or of the cross-randomized poultry and cash interventions. Thus, we ignore the presence of the randomized aspirations intervention in these models, but note that the pooled effects reported for T1 and T2 include the presence of the aspirations treatment.

The specification for the pooled model (model 1) estimates the impact of each treatment arm separately against the control (T4) using the ANCOVA model with the following specification:

$$Y_{1hvd} = \beta_0 + \beta_1 T1_{vd} + \beta_2 T2_{vd} + \beta_3 T3_{vd} + \beta_4 Y_{0hvd} + \beta_5 X_{0hvd} + \mu_d + \varepsilon_{hvd,} \quad (1)$$

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. β_1, β_2, and β_3 provide ITT estimates of the impact of T1, T2, and T3 respectively. To test whether the ITT estimators are statistically different across treatment arms T1 and T2, for example, we conduct a Wald test of equality of the estimates β_1 and β_2.

In this report, we do not present results from a model estimating the impact of a combined treatment group T∈{T1,T2,T3} against the control group (T4) for several reasons. First, the three treatment arms include distinct packages of interventions that are not clear substitutes. Although all three are graduation model interventions, it is not clear what we learn by estimating the combined effect. We also do not expect a substantial gain in power, which is largely determined by the smaller number of clusters in the

[20] The "extremely poor" sample includes the poorest 10 out of 18 households based on a baseline PCA asset index and the "less poor" sample includes the 8 out of 18 least poor households based on the same index.

control group in this pooled model. On a practical level, including estimates on the combined treatment effect for all outcomes would further expand the number of tests presented, with many of the results being difficult to interpret.

For model 2, we estimate the fully interacted model on the extremely poor household sample as follows:

$$Y_{1hv} = \begin{matrix}\beta_0 + \beta_1 T1_v * P_v + \beta_2 T1_v * C_v + \beta_3 T2_v * P_v + \beta_4 T2_v * C_v \\ +\beta_5 T3_v + \beta_{10} Y_{0hv} + \beta_{11} X_{0hv} + \mu_d + \varepsilon_{hv}\end{matrix} \quad (2)$$

where P_v is an indicator for the poultry treatment and where C_v is an indicator for the cash treatment. Here, $T1_v * P_v$ is an indicator for a household in a T1 treatment kebele assigned to the poultry transfer. β_1 is an estimate of the average impact of T1 and the poultry transfer relative to the control group. β_2 is an estimate of the average impact of T1 and the cash transfer. β_3 and β_4 estimate the analogous impacts for T2, and β_5 estimates the average impact of T3. We also report, for the extremely poor sample, the average impact of T1, T2, poultry and cash, using linear combinations of the estimates in equation (2).

For model 3, we estimate the pooled effect of T1, T2 and T3 using equation (1) on only the less poor sample.

For child nutrition and dietary outcomes measured for children age 0–24 months or 6–23 months, we estimate impacts using a repeated cross-section model from the baseline and endline data, estimated using DID. In these models we control for baseline kebele mean outcomes since baseline data are not available for children under age 24 months in the endline sample.

The DID approach utilizes the following regression framework.

$$\begin{matrix}Y_{cvd} = \beta_0 + \beta_1 T1_{vd} + \beta_2 T2_{vd} + \beta_3 T3_{vd} + \beta_4 Time \quad + \beta_5 T1_{vd} xTime + \beta_6 T2_{vd} xTime + \\ \beta_7 T3_{vd}\, xTime + \beta_8 XAge_{cvd} + \beta_9 Gender_{cvd} \; + \; \mu_d + \varepsilon_{cvd,}\end{matrix} \quad (3)$$

where $T1_{vd}$ is an indicator for whether a child in kebele *v* in woreda *d* was randomly assigned to treatment T1.[21] Similarly, $T2_{vd}$ indicates randomized assignment to T2, and $T3_{vd}$ indicates randomized assignment to T3. Time is a dummy variable defined as 0 if the observation is from the baseline and 1 if the observation is from the endline. β_1, β_2, and β_3 provide estimates of the difference between T1, T2, or T3 and the control group at baseline, respectively. β_4 provides a measure of secular trends in the outcome between rounds and β_5, β_6, and β_7 indicate the impact of treatment arms T1, T2, and T3 at endline, respectively. Furthermore, both age and gender are included in regressions covering breastfeeding, complementary feeding, and nutritional status since both characteristics improve the regression precision and are uncorrelated with treatment. Finally, 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

[21] Child subscripts are omitted to simplify notation.

– T4. As T1 and T3 both had N* programs, the result tables also report the joint significance of β_5 and β_7.

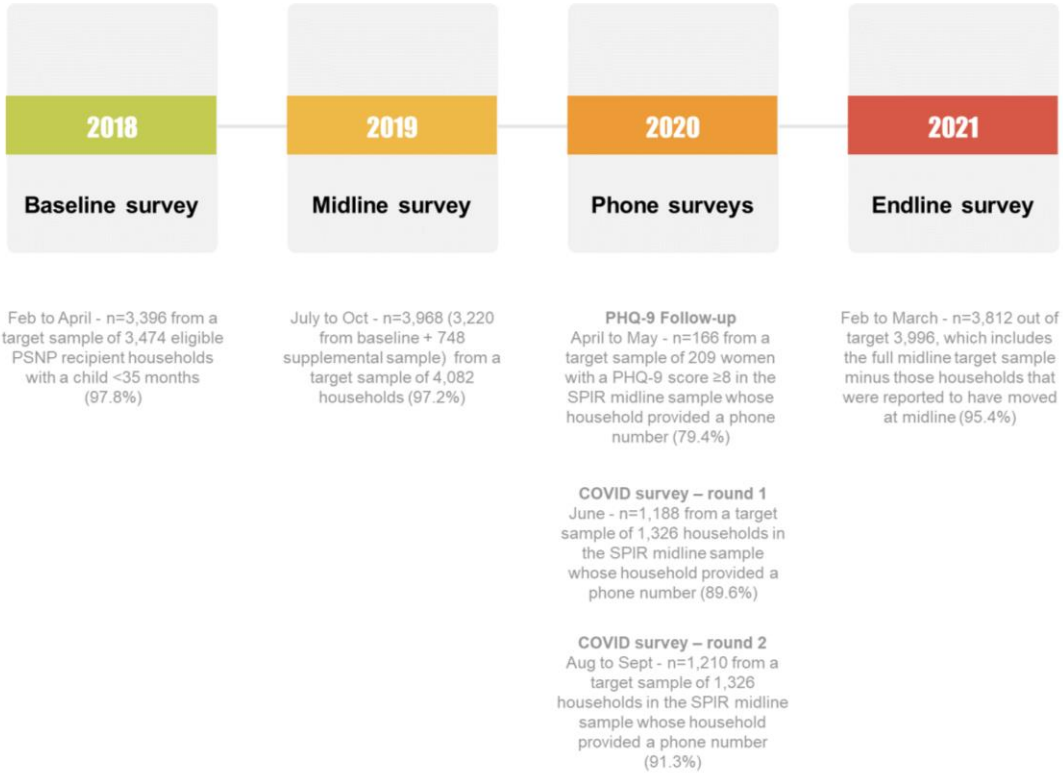
This report presents hundreds of estimates of treatment effects of the SPIR project. With such a large number of estimates, it is common to consider adjusting for multiple hypothesis testing. When conducting many hypothesis tests, the probability of rejecting the null hypothesis may be increasing in the number of tests. We use an approach to correcting for multiple hypothesis testing 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. For nearly all treatment effect estimates, we report both the conventional p-value and the sharpened q-value. We also present average standardized treatment effects that reduce the number of tests by creating indices of families of outcomes (see Kling et al. 2007).

4 Data collection

The SPIR impact evaluation included three household surveys of the full study sample in addition to three shorter surveys conducted by phone with a subsample of households. Details on data collection for the baseline and midline surveys are covered in the baseline and midline reports, respectively. In this chapter, we provide a summary of the data collection for the three phone surveys conducted in 2020 and the endline survey conducted in February–April 2021. These surveys were adapted to respond to the COVID-19 crisis. Laterite served as the in-country partner, leading the fieldwork for these surveys in close cooperation with the quantitative evaluation team from IFPRI and the SPIR implementation team.

A short follow-up survey was initially planned for March 2020 aiming to reach all study households in a subsample of 143 kebeles assigned to treatment arms 1, 3, and 4. The aim of the follow-up visit was to measure maternal depression in order to assess the impact of a 12-week IPT-G on the prevalence of depression. However, the follow-up visits were canceled due to the onset of the COVID-19 crisis and related travel restrictions. In place of the field visits, a short phone survey was conducted from late April to early May 2020 targeting a subsample of roughly 209 IPT-G participants. This subsample included all women who were screened to have at least mild depression at midline (PHQ-9≥8) and whose household provided a phone number in a past survey round.

Figure 4.1: Overview of surveys conducted for SPIR project



Two additional phone surveys were administered targeting all households that reported a phone number in a previous survey round (n=1,326) to assess changes in household well-being during the COVID-19 pandemic. The first COVID-19 phone survey was conducted in June 2020 and the second survey was conducted from August to September 2020. The endline household survey was originally planned for June to August 2020, but since it was not possible to conduct fieldwork during this period due to COVID-19 restrictions, the survey was postponed until February–April 2021 targeting the full midline sample including the original baseline sample and the supplemental midline sample (n=3,996).

4.1 Survey instruments

4.1.1 Phone surveys

PHQ-9 follow-up – The primary female respondent from each sample household was invited to participate in a brief phone interview, which included administration of the PHQ-9 and other questions about their well-being. Respondents were offered an incentive of airtime credits worth 100 Ethiopian birr at the completion of the interview.

COVID-19 surveys – The target respondent for the COVID-19 phone surveys was the primary male respondent from the midline survey. If the primary male respondent from the midline survey was not

available, the midline primary female respondent was interviewed. The questionnaire included questions on economic activity, food security, agricultural activities, and knowledge, perceptions, and behaviors related to containment of the COVID-19 virus. The survey also included questions about access to PNSP4 transfers as well as past exposure to desert locusts and expected future risks. Respondents were offered an airtime credit of 100 birr upon completion of each survey wave and were provided public health information on recommended social distancing and hygiene practices at the end of the call.

4.1.2 Endline survey

Similar to the baseline and midline surveys, 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 identified primary male respondent; and a set of questions for the identified primary female respondent. Primary female and male respondents were the same as those identified at baseline or midline for the supplemental sample. The primary female was the caregiver of the index child and the primary male was her spouse. If this individual was no longer a member of the household a new primary male or female respondent was identified as the caregiver of the endline index child.

The index child identified at baseline (*baseline index child*) was a randomly selected child from the sampled household between the ages of 0 and 35.9 months. The baseline index child was 34–70 months of age at the time of the endline survey. The *midline index child* was a sibling of the baseline index child between the ages of 0 and 23.9 months at midline (or if there was no sibling in this age range, a different child from the household in this age range, randomly selected at the time of the midline survey). The *endline index child* was identified as a randomly selected sibling of the *baseline* or *midline index child* between the ages of 0 and 23.9 months at endline. If there was no sibling in this age range, a different child from the household was randomly selected, and if there was no child in the age range of 0–23.9 months, a household had no *endline index child;* that applied to almost three-quarters of the households at endline. A separate, specifically trained team conducted anthropometry measurements on the *baseline, midline,* and *endline index children* and the primary female respondent.

Most of the questions included in the endline household questionnaire appeared in either or both the baseline and midline household questionnaires. Modules on PSNP participation, crop choice, crop production and sales, access to financial services, childcare activities, experience with depression, and household dynamics and agency were covered in all three rounds of the household survey. Other select modules were included in either the baseline or midline, with the second measurement taken at endline. New modules only included in the endline questionnaire covered experience of fall armyworm and desert locusts, experience of unrest and social cohesion, land tenure and investments, stress and happiness, and time and risk preferences. The list of modules from the endline household questionnaire is presented in Figure 4.2. The endline household survey questionnaire was administered by enumerators using tablets with a computer-assisted personal interview (CAPI) programmed in SurveyCTO. The CAPI enabled enumerators to easily access pre-loaded data, follow interview skip patterns according to interviewee responses, and back-up survey data after each day of interviews.

Figure 4.2: Lists of modules in the endline household, female, and male surveys

Strengthen PSNP4 Institutions and Resilience (SPIR)
Development Food Security Activity (DFSA)
ENDLINE SURVEY: Household Questionnaire – January 20, 2020

DRAFT: For Research Purpose only

Outline:

Module A: Household identification and consent
Part 1: Household identification, location, and consent
Part 2: Sample verification

Module B: Household composition and characteristics
Part 1: Household roster

Module C: Sample structure and result of randomization
Part 1: Sample structure and result of randomization

Module Z: Household location (GPS recording)

Universal Codes (Include with all CAPI options):

-97=Refuse to respond	-98=Don't know	-99= Not applicable

Strengthen PSNP4 Institutions and Resilience (SPIR)
Development Food Security Activity (DFSA)
ENDLINE SURVEY: Male Questionnaire – February 5, 2021

DRAFT: For Research Purpose only

Sample Variables *(from household questionnaire)*

Module A: Household assets
- Part 1: Productive assets
- Part 2: Consumer durables

Module B: Livestock production
- Part 1: Livestock ownership and management
- Part 2: Income from livestock and specified agricultural products
- Part 3: Cost of livestock production

Module C: Agriculture
- Part 1: Land characteristics and tenure
- Part 3a: Crop choice – Mehr Season
- Part 3b: Crop inputs and labor – Mehr Season
- Part 3c: Crop production, sales, and use – Mehr Season

Module D: Income apart from own-agricultural activities
- Part 1: Wage employment

Module E: Business and youth employment
- Part 1: Own business activity

Module G: Access to credit and financial services
- Part 1: Credit for production purposes
- Part 2: Credit for consumption purposes
- Part 3: Access to savings
- Part 4: Access to insurance

Module H: Expenditure and markets
- Part 1: Durables and services (annual)
- Part 2: Household consumables (monthly)
- Part 3: Food markets

Module L: Wellbeing
- Part 4: Experience with depression and emotional wellbeing
- Part 5: Safety protocol
- Part 6: Stress and happiness

Module I: Household dynamics and empowerment
- Part 2: Agency, risk and time preferences
- Part 3: Intrahousehold Dynamics and Attitudes
- Part 4: Gender norms

Module J: Nutrition, health, and care of child
- Part 2: Childcare activities

Module K: Access to the PSNP and SPIR activities
- Part 1: Public Works
- Part 2: Direct Support
- Part 4: Other Public Transfers
- Part 6: Participation in VESA groups and SPIR activities

Module M: Program exposure
- Part 1: Financial education and livelihoods
- Part 2: Health and nutrition
- Part 3: Social analysis and action

Module L: Wellbeing
- Part 2: Experience with shocks
- Part 7: Social cohesion and experience with social unrest

Universal Codes (Include with all CAPI options):

97=Refuse to respond	98=Don't know	99= Not applicable

Strengthen PSNP4 Institutions and Resilience (SPIR)
Development Food Security Activity (DFSA)
ENDLINE SURVEY: Female Questionnaire – February 23, 2020

DRAFT: For Research Purpose on

Sample Variables *(from household questionnaire)*

Module A: Housing, water, sanitation and hygiene
- Part 1: Housing and water
- Part 2: Sanitation and hygiene

Module C: Livestock owned by the woman
- Part 0: Livestock ownership and management
- Part 1: Income from specified agricultural products

Module D: Income apart from own-agricultural activities
- Part 2: Wage employment

Module F: Access to credit and financial services
- Part 1: Productive and consumption credit
- Part 3: Access to savings

Module L: Mental wellbeing
- Part 3: Experience with depression and emotional wellbeing
- Part 4: Safety protocol
- Part 5: Stress & happiness

Module H: Dietary diversity and food security
- Part 1: Women's dietary diversity (24-hour recall)
- Part 2: Household food consumption and expenditure
- Part 3: Household food security & FIES

Module I: Household activities, decision-making and empowerment
- Part 1: Decision-making on value chains and market access
- Part 2: Agency, time and risk preferences
- Part 5: Intrahousehold dynamics and attitudes

Module J: Nutrition, health, and care of child
- Part 1: PSNP during pregnancy and lactation
- Part 3: Infant and young child feeding (IYCF) practices
- Part 4: Child health history
- Part 5: Maternal IYCF knowledge and perceptions
- Part 6: Childcare activities
- Part 7: Exposure to health and nutrition services
- Part 8: Anthropometry

Module K: Participation in VESA groups

Module M: Program exposure
- Part 1: Financial education and livelihoods
- Part 2: Health and nutrition
- Part 3: Social analysis and action
- Part 5: Participation costs

Universal Codes (Include with all CAPI options):

-97=Refuse to respond	-98=Don't know	-99= Not applicable

4.2 Ethical approval

The IFPRI Institutional Review Board (IRB) reviewed the study protocol at baseline for the SPIR DFSA quantitative evaluation and granted ethical approval for the study. The protocol was updated for the midline survey, PHQ-9 follow-up, COVID-19 phone surveys, and the endline survey and resubmitted to the IFPRI IRB for review. The IFPRI IRB approval for the endline survey required additional documentation of clearance from World Vision, the Ethiopian Ministry of Agriculture, and the IFPRI Senior Management Team due to the higher level of risks involved in data collection during the pandemic. IFPRI also received ethics approval from the Institutional Review Board (IRB) at Hawassa University.

Informed oral consent was collected from all participants prior to the start of any interview by phone or in-person. The entire field team was trained on ethical data collection prior to the start of any interviews. Before beginning a survey, enumerators read each respondent a brief description of the study that was being conducted, informed them that their participation in the study was voluntary and that they could discontinue participating at any time, and asked whether they agreed to respond to the interview questions. The enumerator only completed a survey if they received verbal consent from the target respondent to participate in the study.

Confidentiality of the data is protected by recording survey interview responses using CAPI, so no hard copy versions of survey questionnaires are available. All files containing raw and analyzed data are securely stored in password-protected databases. Access to the complete data is restricted to two members of the IFPRI evaluation team. A unique household ID is assigned to each household. The name and geographic location of the respondent is kept in a separate data file. Anonymized versions of the datasets that exclude personal identifiers will be made available for public access.

4.3 Enumeration teams and trainings

4.3.1 Phone surveys

PHQ-9 follow-up – Five enumerators were trained to administer the PHQ-9 follow-up survey by phone. All enumerators had been part of the midline survey enumeration team and participated in an in-person training for the PHQ-9 survey in March 2020 prior to the COVID-19 shutdown. Enumerators conducted mock interviews with one another to practice the interview by phone and test the questionnaire.

COVID-19 surveys – The enumeration team for this survey included 13 enumerators (4 female and 9 male), 2 senior field supervisors, a data manager, a research analyst, and a research associate. The enumerator training was conducted remotely with a PowerPoint presentation and audio recordings covering the content for the survey and expected conduct of the enumerators. Conference calls were held to review the material. Enumerators' comprehension of the material was tested with quizzes. Finally, mock interviews were conducted by phone for enumerators to practice administering the questionnaire. The same team was re-trained in the same remote fashion to administer the updated questionnaire for the second round of the COVID-19 phone survey.

4.3.2 Endline survey

The enumeration team for the endline household survey included 14 field coordinators and 62 enumerators (28 females and 34 males), as well as 16 anthropometric enumerators. Eight enumeration teams covered the Amhara region, and 6 teams covered the Oromia region. Each team included a field coordinator, 4 survey enumerators, and an anthropometric enumerator. Three senior field coordinators oversaw all teams.

Laterite prepared training material and conducted the enumerator training in Addis Ababa from January 14 to 28, 2021. To keep group sizes small and mitigate the risk of spreading COVID-19, the training was delivered in three separate sessions. Training material developed by Laterite and reviewed by IFPRI included a training manual, a training PowerPoint, a fieldnote template, and the SurveyCTO instruments. After training, the team was deployed to pilot test the survey, which took place over two days in two separate locations not far from Addis Ababa. A total of 100 households were included in the pilot test, allowing enumerators to practice administering the survey and to update estimates on survey duration and assess logistical needs. Small updates were made to the questionnaire to adjust response options and improve the phrasing of the translation.

Enumerators and trainers were tested for COVID-19 before the start of training. Temperatures were taken at the start of the training session each day. During training, rooms were well ventilated, individuals were spaced for recommended social distancing and face masks and hand sanitizer was used. Data collection materials were sanitized before and after every use.

4.4 Fieldwork experience

4.4.1 Phone surveys

PHQ-9 follow-up – Enumerators reached 173 households, but only completed 166 interviews because the primary female was unavailable for interview in 6 of the households, and 1 respondent did not consent to participate in the interview. This resulted in a response rate 79.4 percent of the 209 target respondent households.

COVID-19 surveys – In the first round, the survey team reached 1,190 households or 89.7 percent of the target sample, deemed a very high response rate for a phone survey. Of these households, 2 did not consent to participate in the interview. The response rate was even better for the second survey round, with 1,211 households reached or 91.3 percent of the target sample, with only 1 household refusing consent.

4.4.2 Endline survey

Enumeration teams worked Monday through Saturday completing two household surveys per day on average. The endline survey enumeration team interviewed a total of 3,812 households out of the target of 3,996 households for the entire sample (see Table 4.1). The target sample included the full midline sample except for the households reported to have permanently moved out of the study area at midline. The attrition rate was 4.6 percent, and a large part of it (80 households) was due to unrest in parts of Amhara that resulted in a decision not to visit four kebeles with reports of unrest. Among the remaining 104 attrited households, 55 had moved out of the study area, 26 were temporarily unavailable, and other

households had either dissolved or were unavailable for the interview for other reasons; 1 household refused consent. In total, 3,706 primary female respondents and 2,465 primary male respondents were surveyed, and 1,064 new *endline index* children were identified, in addition to the *midline* and *baseline index* children from previous surveys who also were measured for the anthropometrics module, if still part of the household. Through the course of the data collection, 39 index children were found to have a low mid-upper-arm circumference, indicating possible severe acute malnutrition, and were referred to the local health post in the kebele. In addition, 52 primary females were reported for IPV. Seven primary females and 3 primary males were referred for severe depression or risk of suicide.

Several measures were taken to monitor and mitigate the risk of COVID-19 transmission throughout the fieldwork. Weekly calls were held between Laterite, World Vision, and IFPRI to share updates from local and national monitoring systems. During data collection, all members of the field team had their temperatures taken daily. Midway through the fieldwork, a COVID-19 test was administered to all members of the field team. As part of community entry, the enumeration team would contact local health officials and the SPIR project field offices to inquire if there were any COVID-19 outbreaks. When conducting household visits, enumerators screened each respondent for COVID-19 symptoms before starting the interview. However, no infected cases were detected through screening. Enumerators and respondents were required to wear masks throughout the interview and maintain recommended social distance. Enumerators used hand sanitizer before and after every interview and sanitized all data collection materials.

Table 4.1: Summary of endline household sample

	Amhara	Oromia	Total
Number of EAs targeted	112	80	192
Intended households per kebele	--	--	--
Completed households per kebele	--	--	--
Intended household interviews	2,320	1,676	3,996
Household/right members not found	69	34	103
No interview due to civil unrest	80	--	80
Consent not given	1	--	1
Completed household interviews	2,170	1,642	3,812
T1: L*+ N*	469	321	790
T2: L*+ N	484	353	837
T3: L + N*	485	341	826
C: PSNP only	424	343	767
Number of primary female respondents*	2,123	1,583	3,706
Number of primary male respondents*	1,256	1,209	2,465
Number of children 0–23 months	646	730	1,376
Number of children 24–60 months	1,412	1,321	2,733

* Where primary female or male wasn't available, certain modules (such as household's asset ownership and housing characteristics), were administered to their spouse. This table only reflects the presence of the designated primary female or male.

4.5 Data quality and cleaning

For the phone surveys and endline survey, data were recorded during the interviews on tablets using SurveyCTO. All data were synced by enumerators (unless there were internet connectivity issues) to a remote server in Dropbox. Senior field supervisors reviewed the survey with enumerators to address any concerns that may have been raised during the interviews. A research analyst from the survey firm ran data checks daily on the uploaded data to identify data errors, check inconsistencies with the field team, and communicate any patterns to avoid future errors. A different analyst from IFPRI checked incoming data regularly to make sure that all modules were completed and that the answers were within logical boundaries.

Once final datasets were received from Laterite, the IFPRI team carefully cleaned the data and constructed any necessary indicators for the analysis.

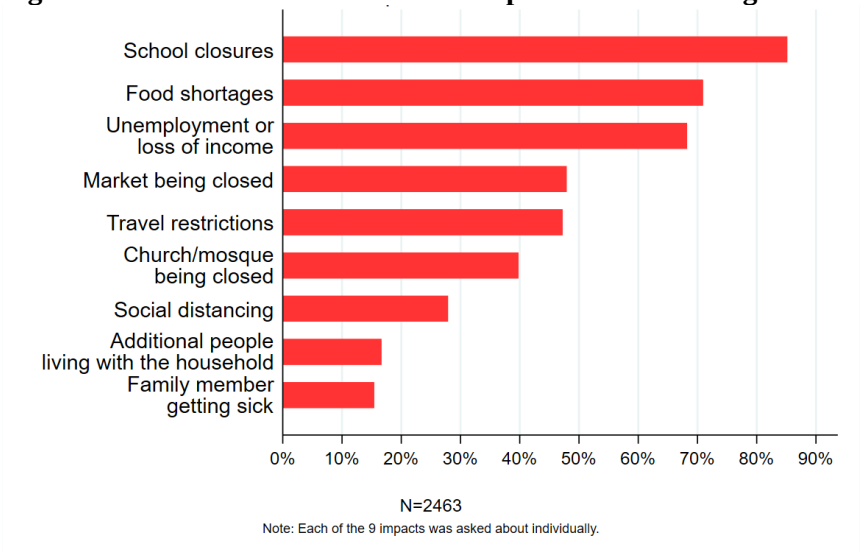
5 Changing context: COVID-19, pests, social unrest, and other shocks

The woredas in Amhara and Oromia that are home to the SPIR project faced many significant shocks in the roughly 19 months that passed between the midline and endline surveys. In addition to COVID-19, the study area faced potentially significant pest infestations from fall armyworm and desert locusts. Many households also experienced weather, health and income shocks during this period. Finally, an armed conflict in Tigray region that began in November 2020 continued through the period of endline data collection and affected neighboring areas in Amhara region, including some in the SPIR operational areas. In this chapter, we report the extent of these shocks as reported by respondents in the endline survey in order to inform the context for the impact analysis. In addition, we examine whether the prevalence of these shocks differed by treatment arm in the study. Although these shocks may have affected many of the outcomes in this report, they would not bias the impact estimates for the SPIR project if their prevalence is balanced across treatment arms.

5.1 COVID-19

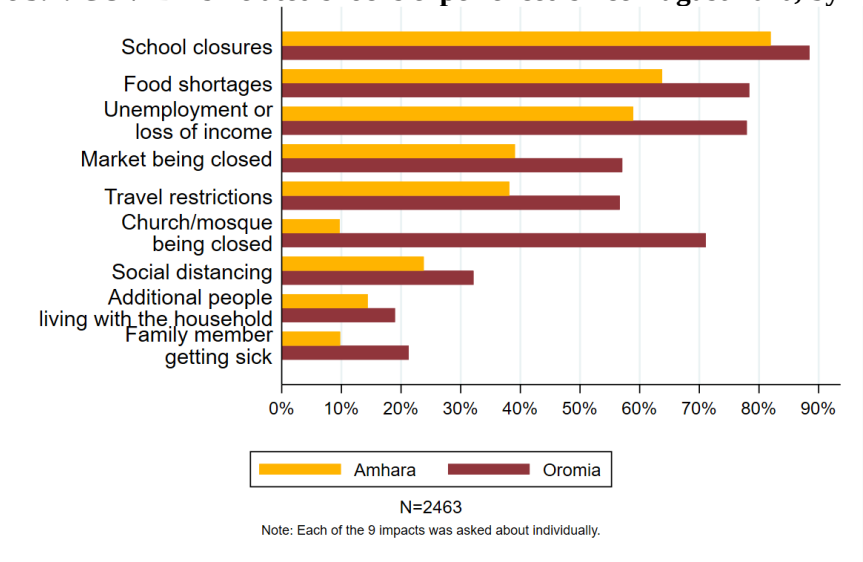
Households reported a number of shocks related to the COVID-19 pandemic and related lockdowns. Two phone surveys of a subsample of the study respondents conducted in the first six months of the pandemic in 2020 found that households faced reductions in food security, foods shortages, and income loss as a result of the pandemic. The endline survey asked respondents to indicate the shocks they experienced since August 2020, the time of the last phone survey. Figure 5.1 shows that more than 85 percent of households reported experiencing school closures, which is expected because schools were closed nationally until October 2020. In addition, more than 70 percent of households reported food shortages and 68 percent reported unemployment or income loss. Closures of markets and churches or mosques affected almost half of the respondents and travel restrictions were also common.

Figure 5.1: COVID-19-related shocks experienced since August 2020



These patterns in relative shock prevalence were similar across the two regions, but the exposure to most COVID-19-related shocks was higher in Oromia than in Amhara (Figure 5.2). Church and mosque closures were much more common in Oromia than Amhara, for example.

Figure 5.2: COVID-19-related shocks experienced since August 2020, by region



5.2 Desert locusts and fall armyworm

Pest infestations from desert locusts and fall armyworm threatened large swaths of Ethiopia in 2020/21, with the challenges of desert locusts making international news. Figure 5.3 shows that desert locusts were not a significant problem in the study locations in Amhara, but in Oromia, 35 percent of households reported losing crops and 22 percent reported losing grazing land to the locusts. Fall armyworm turned out to be a bigger threat in the SPIR locations, with 15 percent of Amhara households and almost 60 percent of Oromia households reporting losing crops to fall armyworm. However, Figures 5.4a and 5.4b show that the extent of crop loss among those losing crops to desert locusts was far greater on average than the crop loss from fall armyworm for those who were exposed.

Figure 5.3: Exposure to pests in the Mehr season in 2020

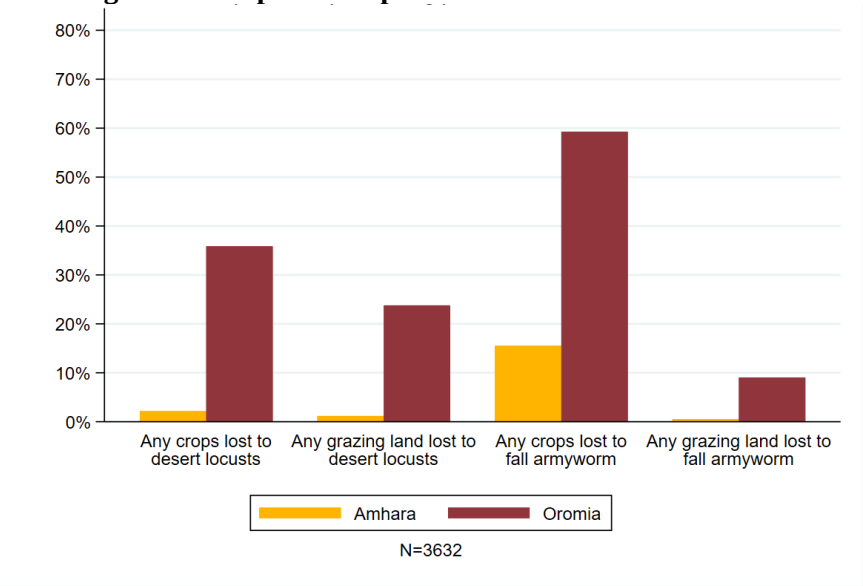


Figure 5.4a: Crop loss due to desert locusts

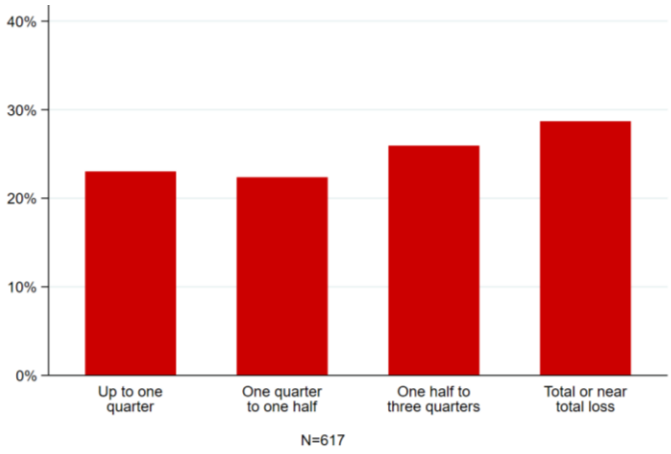
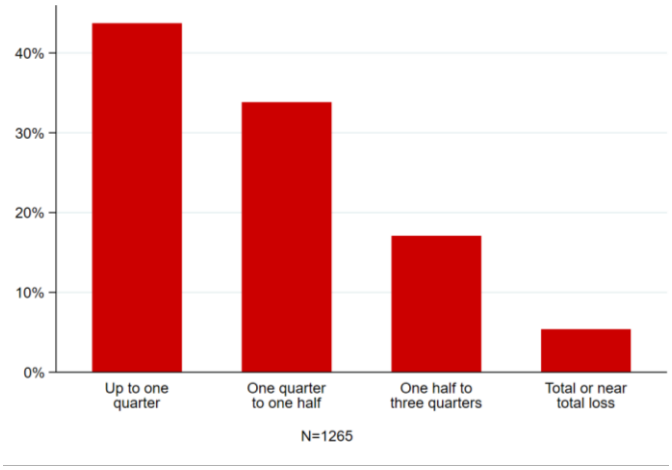


Figure 5.4b: Crop loss due to fall armyworm



5.3 Weather, health, and income shocks

Weather shocks were a common challenge in the sample (Figure 5.5), with more than 50 percent of households independently reporting experiencing drought. Fifty percent of households also reported floods and erosion as significant shocks in the last 15 months. These weather shocks were common in both Amhara and Oromia (Figure 5.6). Large increases in inputs prices were reported by more than 60 percent of the sample, but this problem was mostly concentrated in Oromia, where more than 85 percent of households reported the problem. These price increases may have been precipitated by supply chain disruptions due to COVID-19.

Figure 5.5: Shocks experienced in the previous 15 months

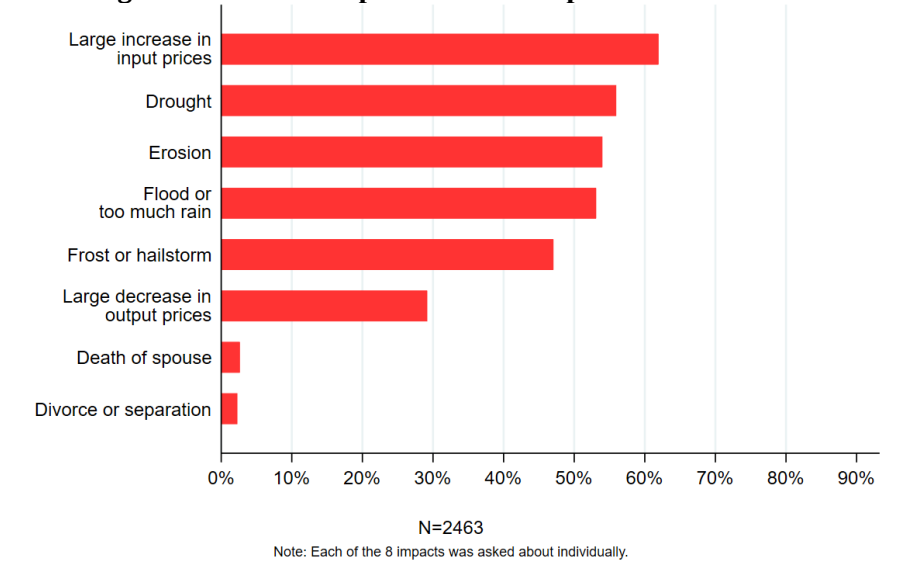
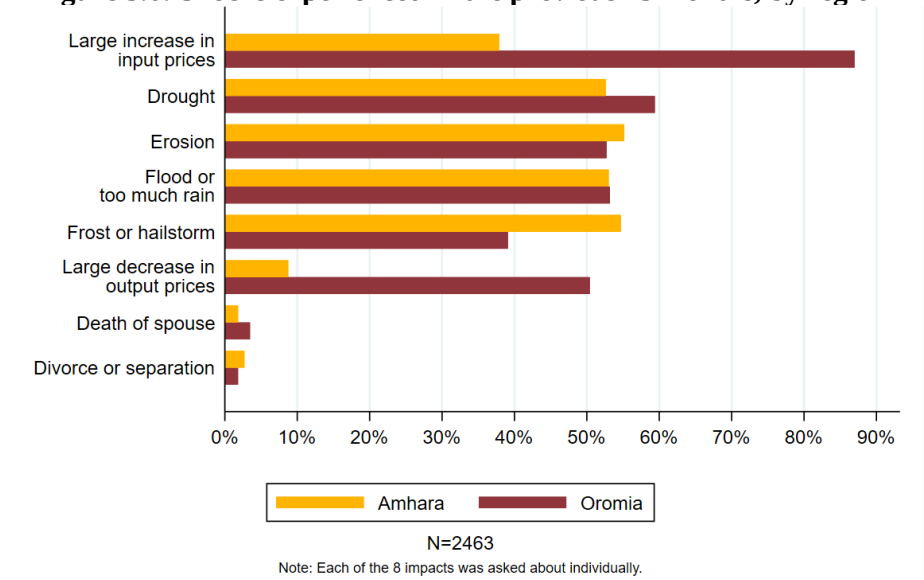


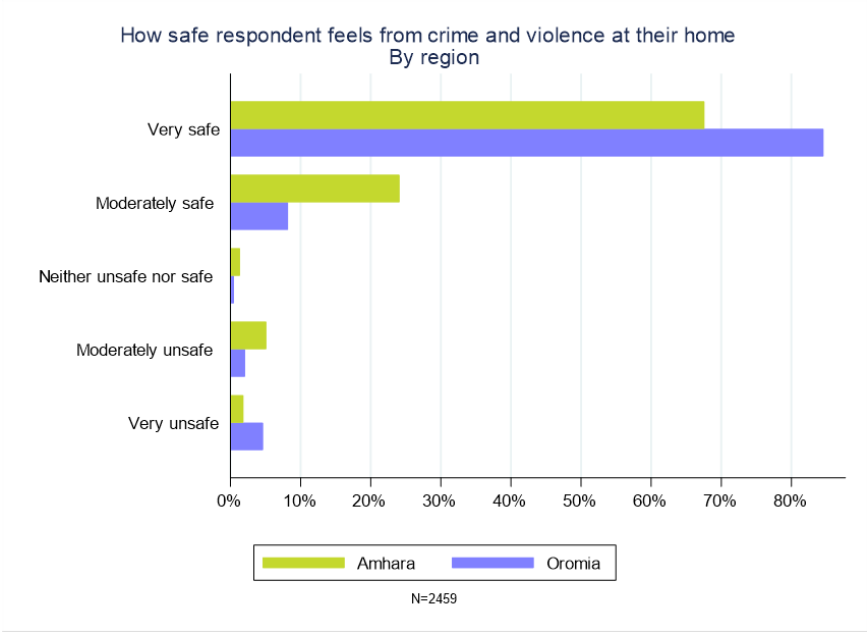
Figure 5.6: Shocks experienced in the previous 15 months, by region



5.4 Conflict and social unrest

Despite the growing conflict in Tigray, its effects on study households, mostly in neighboring Amhara were localized, with only 2.8 percent of households reporting experiencing violence since January 2020. Concerns about safety were generally not high in the sample (Figure 5.7) but were higher in Amhara than Oromia.

Figure 5.7: Concerns over safety, by region



5.5 Testing balance of shock exposure across treatment arms

We tested whether the prevalence of the shocks reported above differed by treatment arms in the study. Finding an imbalance should be unlikely because the allocation of treatment arms was random, but an imbalance could still occur by chance. This would potentially lead to bias in estimating the impact of SPIR treatments. Results of these tests are reported in Tables 5.1 and 5.2.

Table 5.1 shows that shocks due to weather (e.g., drought, flood, erosion) were balanced across treatment arms, but frost was significantly more prevalent in T3 than in the other treatment arms. There were no significant imbalances for price shocks or pests across treatment arms, except that input price increases were weakly significantly higher in T3 than in the control and the effect of fall armyworm on grazing lands was slightly higher in T1 (2 percent more households) and weakly higher in T2 (1.5 percent more households). Similarly, in Table 5.2, the prevalence of COVID-19, health and conflict-related shocks was balanced across arms T1, T2 and T3. The only exception is that the prevalence of divorce is weakly significantly lower in T1, but the difference is small.

5.6 Conclusion

Households in SPIR woredas faced many significant shocks in the period between the midline and endline surveys, including the COVID-19 pandemic and associated lockdowns, desert locusts, fall armyworm, droughts, floods, and civil unrest. Disruptions due to COVID-19 included closures to schools, markets, mosques and churches and also led to food shortages and periods of income loss. Many of these effects were worse in Oromia than Amhara. Desert locusts did not significantly affect households in

Amhara, but roughly one in three households in Oromia lost crops to desert locusts. Fall armyworm led to crop losses for 15 percent of households in Amhara but 60 percent of households in Oromia. Roughly half of households in the sample reported a significant drought event, and similarly, half reported a significant flood and associated erosion in the last 15 months. Despite the conflict in Tigray, its effects on study households, mostly in neighboring Amhara, were limited to a small share of households outside the four kebeles that study teams were unable to visit.

Balance tests for exposure to these shocks showed that the prevalence of each of these shocks was relatively balanced across study treatment arms, suggesting that these shocks are unlikely to lead to bias in estimated impacts of the SPIR program.

Table 5.1: Experience of weather shocks and pest infestations

	(1) Household was affected by drought(s) in the last 15 months	(2) Household was affected by excessive rain in the last 15 months	(3) Household was affected by erosion in the last 15 months	(4) Household was affected by frost in the last 15 months	(5) Household was affected by inputs' price increase in the last 15 months	(6) Household was affected by outputs' price decrease in the last 15 months	(7) Any crops were lost to desert locusts in the last Mehr season	(8) Any grazing land was lost to desert locusts in the last Mehr season	(9) Any crops were lost to fall armyworm in the last Mehr season	(10) Any grazing land was lost to fall armyworm in the last Mehr season
T1	0.031	-0.044	-0.051	0.036	-0.005	-0.009	0.004	0.025	0.027	0.020**
	(0.036)	(0.056)	(0.053)	(0.046)	(0.031)	(0.029)	(0.024)	(0.019)	(0.026)	(0.009)
T2	0.012	0.024	0.010	0.050	0.002	-0.032	-0.008	0.024	0.025	0.015*
	(0.040)	(0.055)	(0.053)	(0.045)	(0.029)	(0.027)	(0.026)	(0.019)	(0.027)	(0.008)
T3	-0.007	0.054	0.032	0.093**	0.051*	-0.027	0.023	0.012	0.023	0.000
	(0.037)	(0.056)	(0.051)	(0.044)	(0.028)	(0.029)	(0.032)	(0.020)	(0.027)	(0.008)
Mean of control (T4)	0.560	0.535	0.552	0.409	0.614	0.329	0.163	0.091	0.336	0.037
N	2,463	2,463	2,462	2,463	2,462	2,461	3,632	3,339	3,635	3,296

* p<0.1 ** p<0.05 *** p<0.01

Table 5.2: Experience of COVID-19 related shocks and conflict

	(1) Household was affected by death in the last 15 months	(2) Household was affected by illness in the last 15 months	(3) Household was affected by divorce in the last 15 months	(4) Household was affected by unemploy-ment in the last 6 months	(5) Household was affected by food shortages in the last 6 months	(6) Household was affected by market closures in the last 6 months	(7) Household was affected by travel restrictions in the last 6 months	(8) Household was affected by church or mosque closures in the last 6 months	(9) Household was affected by school closures decrease in the last 6 months	(10) Household experien-ced violence (crime, unrest) in the last 12 months
T1	-0.018	0.036	-0.013*	0.046	0.050	0.017	-0.003	0.007	0.015	0.006
	(0.013)	(0.027)	(0.008)	(0.040)	(0.033)	(0.035)	(0.038)	(0.030)	(0.031)	(0.011)
T2	-0.015	0.026	0.000	0.056	0.051	-0.033	-0.060	-0.001	0.021	0.010
	(0.012)	(0.025)	(0.009)	(0.039)	(0.035)	(0.035)	(0.038)	(0.030)	(0.029)	(0.010)
T3	-0.006	0.039	0.003	0.041	0.028	-0.041	-0.051	-0.051	-0.029	-0.000
	(0.012)	(0.024)	(0.010)	(0.036)	(0.029)	(0.043)	(0.046)	(0.038)	(0.033)	(0.009)
Mean of control (T4)	0.039	0.229	0.025	0.647	0.684	0.494	0.502	0.435	0.850	0.025
N	2,463	2,463	2,461	2,463	2,463	2,463	2,460	2,463	2,459	2,461

* p<0.1 ** p<0.05 *** p<0.01

6 Experience with the SPIR Program

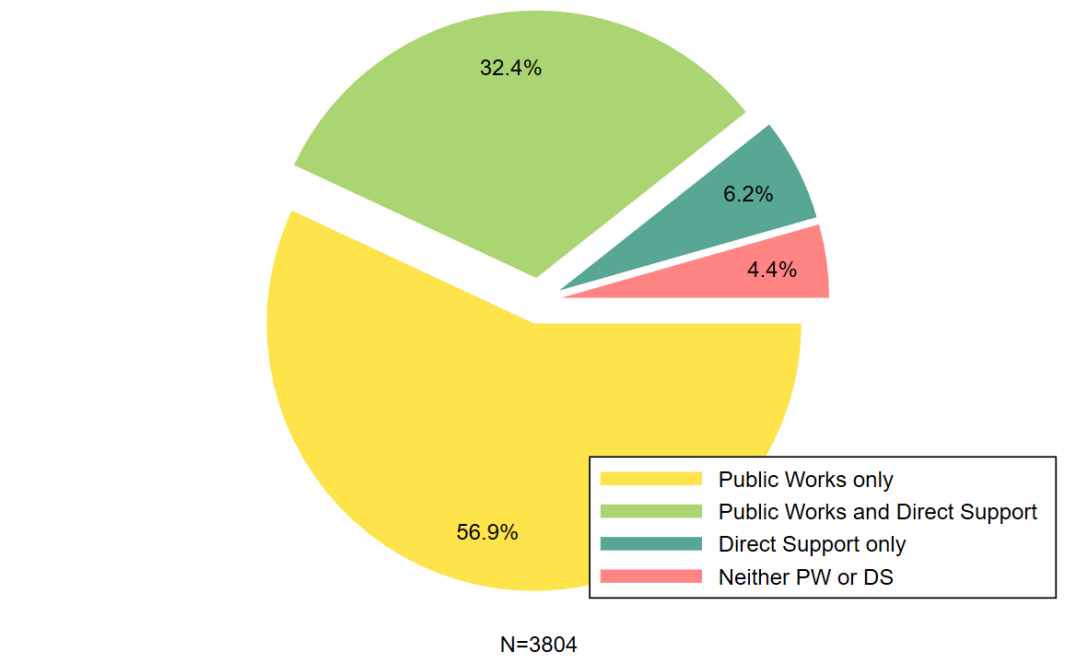
This chapter first briefly covers the reach of core PSNP operations, after which we delve into exposure to different SPIR interventions as reported by respondents of the endline survey. We broadly categorize the programming into livelihood activities, implemented in the enhanced livelihood (L*) kebeles, and nutrition activities, implemented in the enhanced nutrition (N*) kebeles.

6.1 Participation in PSNP

By design, the core components of PSNP programming, including public works (PW) and/or direct support (DS) transfers, are available to all households in the SPIR study sample. Following the PSNP eligibility criteria, most households are eligible for PW, which entail food or cash payments for seasonal labor, while permanent direct support is available to households with only elderly/disabled members, and temporary direct support to those with a pregnant or lactating woman.

As shown in Figure 6.1, 89 percent of households indicated having participated in PW during the previous year. A little under a third of these households, or 32 percent, report that they also received direct support payments during the same year, likely meaning that they were enrolled in TDS due to a lactating/pregnant household member. Six percent said that they only received direct support—most likely households that are on PDS—and 4 percent indicated not being part of either program. Since in the baseline study virtually all households reported being part of one or the other, these households have likely graduated from the program in the past four years.

Figure 6.1: Household was part of PSNP public works or direct support during previous year



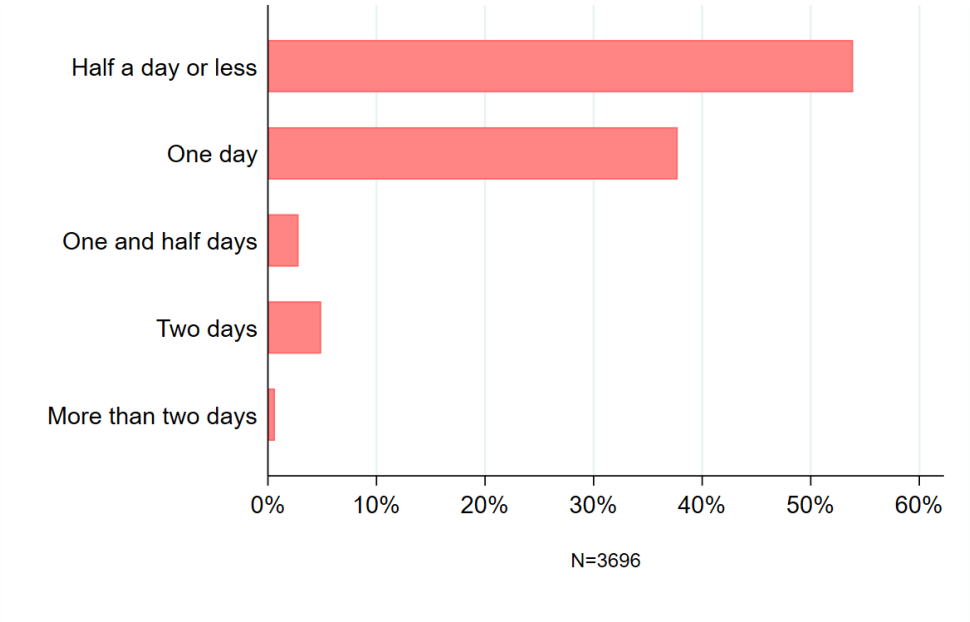
In terms of regional differences, households in Oromia are more likely than households in Amhara to have participated in a PW activity over the previous 12 months (94 percent vs. 86 percent) and less likely

to have received direct support payments (36 percent vs. 40 percent). The lower share of direct support recipients in Oromia can likely be explained by both the low share of female-headed households in Oromia and the much higher proportion of households in Amhara already receiving payments before the start of PSNP4.

The reported receipt rates of any direct support have been rising since the baseline study. In Oromia, the rate has gone from 4 percent at baseline (2018) to 30 percent at midline (2019) to 36 percent at endline. For Amhara, these numbers are 20 percent, 43 percent, and 40 percent, respectively. However, while we do not have explicit answers about whether the transfers that the respondents are receiving are TDS or PDS, looking at the conjunction of answers to questions about participation in PW and in direct support, we see a clear upward trend in those who indicate being part of both programs—meaning that they have likely been receiving temporary support due to pregnant or lactating household members. At baseline, the share of such households was 6 percent, at midline 28 percent, and as indicated above, at endline 32 percent. The impact of efforts to get eligible women enrolled in receiving additional support thus are evident in the data.

At endline, only 55 out of 3,793 respondents describe their household as having graduated from the PSNP program, the most prevalent reason for this being graduation based on livestock ownership. Only 13 men say their household had self-graduated from the program. Similar to the midline, roughly 36 percent of respondents report that they have made a complaint about the PSNP program or the food transfer. Of these people, only 12 percent say that the complaint was successfully resolved; in over one-third of the cases, the case was reportedly resolved by a kebele grievance/appeal committee.

Figure 6.2: How long it took to collect the food ration (public works payment or direct support transfer)

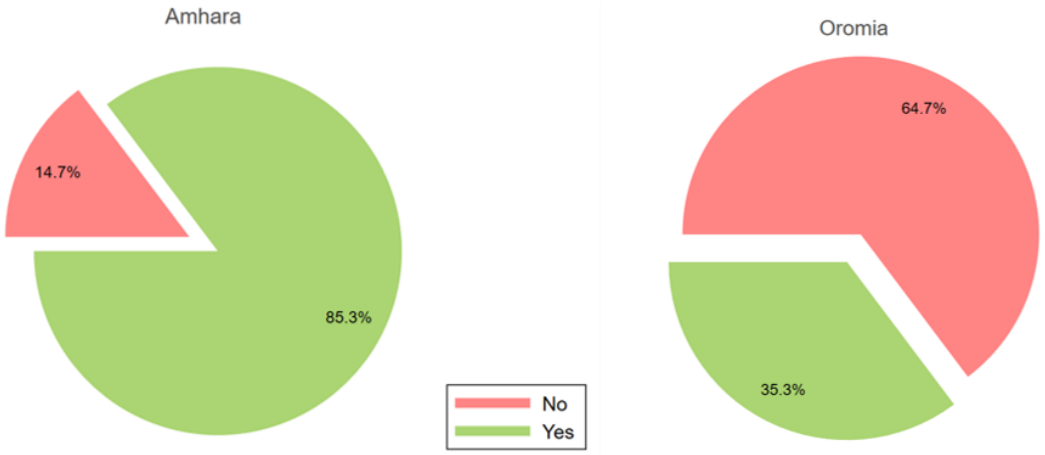


Respondents were also asked how long it takes them to collect the food ration they receive as either their PW payment or direct support transfer. As shown in Figure 6.2, for 54 percent of households it takes half

a day or less. Roughly 8 percent of households—most of them from Oromia—say that it takes them more than a day. These figures present an improvement from the midline study in 2019, when 49 percent of households indicated that the travel takes them half a day or less, while 10 percent said it takes two days or even more. By the endline study, this last figure had dropped to 5 percent.

During the coronavirus pandemic, SPIR PSNP households were also targeted to receive a hygiene kit; Figure 6.3 shows that reported receipt rates were very high in Amhara at over 85 percent, while in Oromia, only 35 percent reported receiving the kit. This aligns with the specific regional programming as the response in Amhara was more focused on distributing soaps and jerry-cans to households, while in Oromia the focus was on strengthening health facilities through personal protective equipment and handwashing stations.

Figure 6.3: Household received a hygiene kit (soap bars or a jerry-can) during the pandemic



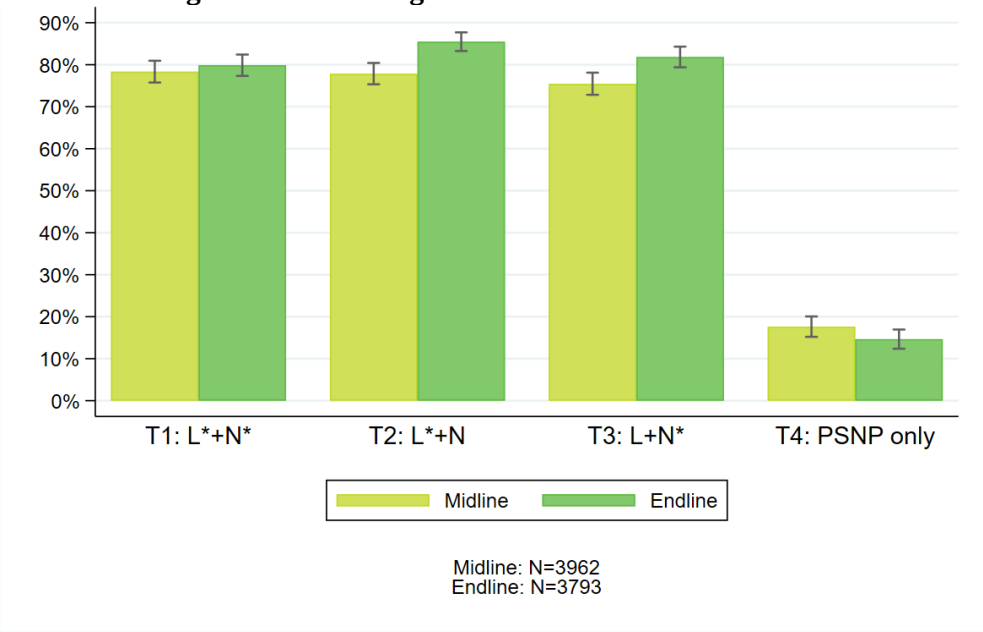
6.2 Core of SPIR programming: VESAs

Village Economic and Social Associations (VESAs) are one of the main platforms through which SPIR programming is delivered. Each group has 25 to 30 members and includes both men and women (often the husband and wife from a single household). In addition to facilitating savings and lending, the SPIR project works with VESAs to foster financial literacy, develop business skills, enhance production and marketing skills, improve social capital, and catalyze women's empowerment. VESAs also serve as a platform for other trainings and services provided by development agents and private sector actors, and create an enabling environment for trainings on social and cultural norms.

Error! Reference source not found. reports both women's and men's responses to questions about their membership in VESAs, as well as a combined indicator of either a male or a female in the same household being a member. At midline, 79 percent, 78 percent, and 76 percent of households in treatment groups T1, T2 and T3, respectively, report having a VESA member in the household. By endline, these numbers were 80 percent, 86 percent, and 82 percent (see Figure 6.4). Contrary to midline, when women in all treatment groups were 6–7 percentage points less likely than men to report being a VESA member, by endline women's membership rates are as high as men's. Similar to midline, around 15 percent of

households in the control group also report having a member in a VESA at endline, but these reports are highly concentrated in a few specific kebeles in Meket, Grawa, and Sekota woredas.

Figure 6.4: How long it took to collect the food ration



Respondents who indicated not being part of a VESA were asked for the reason. Among the control group, T4, by far the most prevalent answer was not having a VESA established in one's vicinity, with 69 percent of respondents choosing this option. For treatment groups T1, T2 and T3, the most common answer, at 60 percent, was "other reason for not wanting to join." Among these groups, non-VESA membership is also highly correlated with reports of receiving direct support payments but not being part of PW—that is, likely being a household on PDS—which aligns with the program design; SPIR specifically targeted PW households for enrolment in VESAs and excluded PDS households. Among those who report being part of PW, the VESA membership rate is 83 percent, while the rate for direct-support-only households is 50 percent.

For the first time, the endline survey also asked about presence of women in the management of VESAs. Encouragingly, 75 percent of female and 74 percent of male respondents indicated that their VESA has a female chairperson or a female vice chairperson/secretary. Furthermore, the median number of women reported to be in VESA management committees was 3.8. Finally, 66 percent of participating women and 69 percent of men say their VESA has conducted a share-out of the group's savings/shares.

6.2.1 Value chain development

The primary VESA activities, including trainings on financial literacy, agricultural and livestock value chain development discussions, and training on home gardening and forage production, were designed as L interventions and were therefore targeted toward all three treatment arms (T1, T2 and T3). While the goal of VESAs is for every member to be exposed to the basics of the value chain approach, a smaller

percentage of households are also invited to come to specific multiday trainings on targeted value chains—primarily poultry production and goat and sheep fattening—depending on cost implications, recipients' interest, and perceived capacity to participate in respective value chains. These were assessed through participative value chain assessments that were conducted in each woreda at the beginning of SPIR, with the goal of identifying the top three targeted value chains for each woreda; in the majority of cases, by far the most popular and consistently selected were poultry and sheep/goat fattening.

The rates of both women and men who had participated have significantly increased since midline. As Table 6.2 and Table 6.3 show, in treatment groups T1, T2 and T3, 22–29 percent of women and 25–31 percent of men had participated in the previous 12 months. At midline, these figures were 12–16 percent for women and 11–16 percent for men on average. Participation rates are considerably higher in Oromia, and among T2 beneficiaries. Among males, by far the most frequently attended types of trainings were on sheep/goat fattening and on poultry production, with 60 percent and 53 percent of those who had been to value chain trainings attending these types, respectively. Ox fattening, vegetable production, beekeeping, and haricot bean production were significantly less attended by both men and women. Among females, poultry production training was the most popular type by a wide margin, attended by 75 percent of those who took part in any value chain trainings, while sheep/goat fattening was attended by 40 percent of participating women.

In addition to various trainings, SPIR also supported the formation of producer marketing groups (PMGs) associated with the targeted value chains. Such groups relevant to specific value chains are, however, not available in all locations. Membership rates in these groups observed at endline are lower than at midline—roughly 6–7 percent of females in T1 through T3 report being a member of a marketing group, compared to roughly 10 percent at midline. Similarly, for males, these rates are 7–8 percent. While at midline participation in T1 and T2 was almost twice as high as in T3, by endline the gap has disappeared. Membership in PMGs is generally expected from those who also participate in value chain trainings, which is also apparent in the data: 20 percent of value chain trainings' participants report being a PMG member, compared to only 2 percent of nonparticipants. Expectedly, the most common producer marketing group is associated with poultry production (64 percent of participating women and 50 percent of men) and sheep/goat fattening (29 percent of participating women and 40 percent of men).

6.2.2 Discussions on health and nutrition

Another important component of VESA meetings are discussions on hygiene and health, nutrition, child feeding, and childcare. Table 6.4 and Table 6.5 show the reported exposure to nutrition- and hygiene-related topics as reported by females and males, respectively.

Roughly 61 percent of all women and men in T1–T3 have attended a VESA meeting or discussion in the previous 12 months. When we look at only women who report being VESA members, the respective rates are 81 percent in T1, and 74 percent in T2 and T3. For men who are VESA members, these numbers are 75, 74 and 73 percent. Among those who have attended, nearly 40 percent say the meetings have been happening weekly, and only 1 percent say they have occurred less than monthly. Females who did attend at least one meeting report the number of times that health and nutrition topics were discussed being roughly 2.5, while for men this same figure is only 0.8 times.

The share of women reporting that nutrition topics were covered at VESA meetings has increased sharply since midline, from just over 40 percent to near 80 percent (see Figure 6.5). There was a similar increase

in the share of women reporting that health or hygiene topics were covered at VESA meetings (see Figure 6.6). Both genders report very similar coverage of topics, while recall of specific themes has gone up since midline: 41 percent of males and 45 percent of females report discussing IYCF practices (compared to 30 percent at midline), 32 percent of men and 39 percent of women recall talking about maternal nutrition, and 19 percent of both men and women report covering COVID-19 and prevention of its spread. While most topics were recalled at a higher rate by respondents from T3 as compared to T1 and T2 at midline, by endline these differences are gone; one explanation for this is the timing of the rollout of discussion modules, which largely happened after the midline survey.

Figure 6.5: Women reporting nutrition topics were covered at VESA meetings

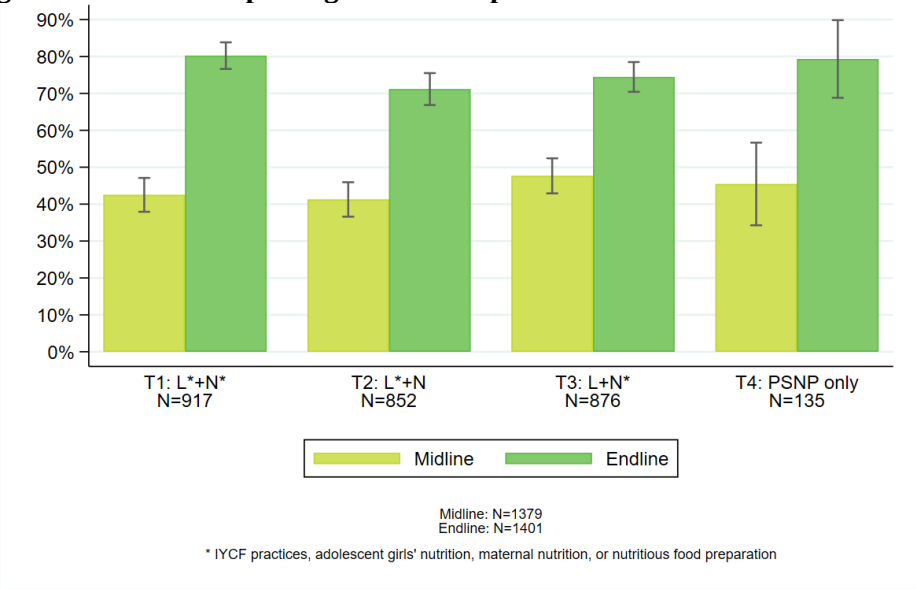
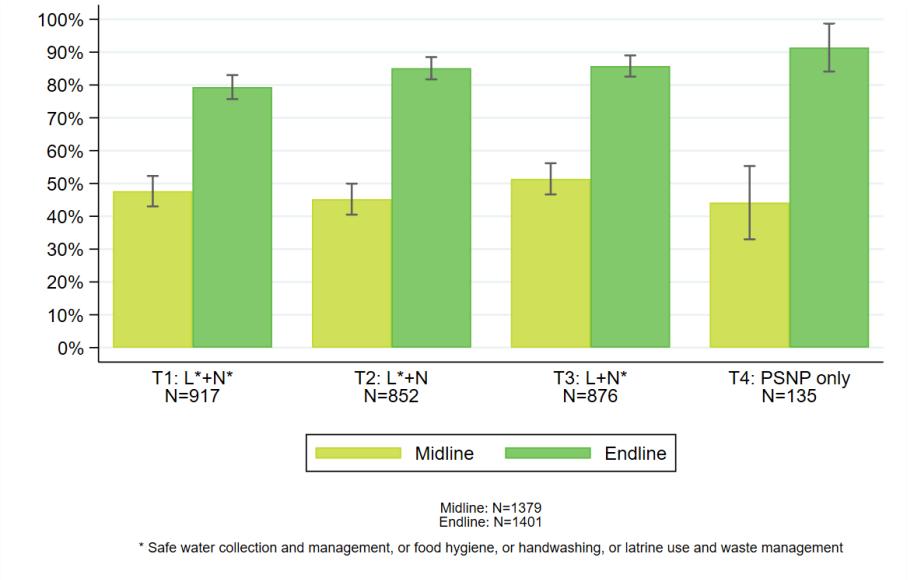


Figure 6.6: Women reporting health or hygiene topics were covered at VESA meetings



6.2.3 Discussions on gender topics

Social analysis and action (SAA) is an L* livelihood activity with the goal of addressing constraints on women's role in intrahousehold decision-making, mobility, and choice of livelihood. While the SAA program was designed to be facilitated by the Food Security Task Force, due to implementation difficulties it was eventually provided by SPIR community facilitators as the VESA discussion module on gender equity.

Table 6.6 and Table 6.7 report answers to questions on SAA-related activities given by females and males, respectively. While at midline less than half of both females and males said that gender topics or couples' relationship questions were discussed at least once at VESA meetings, at endline all women who had attended VESA meetings universally claim discussing such topics, the average number of times discussed being 2.7—higher than for health and nutrition topics. For men, 77 percent say that these topics were discussed at least once. Those who indicate discussing gender or couples' issues at the meetings are asked about coverage of specific subtopics. Just as at midline, the most broadly covered topic appears to be workload sharing, with well over 70 percent discussing it, after which male engagement in childcare was the second-most widely recalled topic.

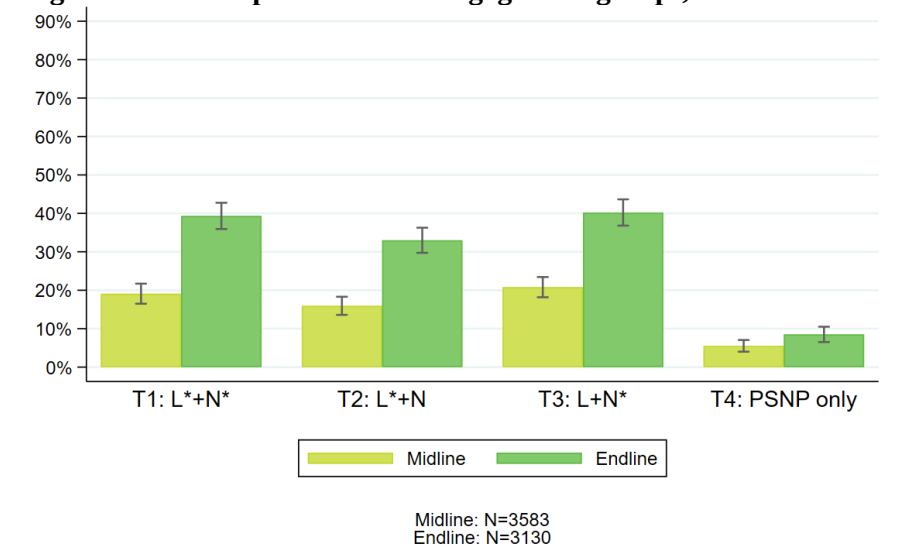
Lastly, respondents were asked if they have observed any changes in their spouse's behavior or actions as a result of these discussions. Among respondents whose partners also report attending, 58 percent of women and 67 percent of men indicate observing changes in their spouse's behavior, respectively.

6.2.4 Male engagement groups

Male engagement groups are SPIR-facilitated men's groups where topics such as gender norms and men's role in caregiving and household chores are discussed in depth. Male engagement groups were formed as part of the N* interventions in T1 and T3. Table 6.8 shows that at endline, roughly 40 percent of men in

T1 or T3 kebeles participated in male engagement groups in the last 12 months. This participation rate was only somewhat higher than men reported in T2, where 33 percent of men claimed to participate in men's engagement groups. These rates of participation reflect the scaling-up of male engagement groups after the midline survey (see Figure 6.7). Even in the control group, more than 8 percent of men indicated being involved in a men's engagement group. Rather than representing spillover of this intervention, we expect that men in T2 and the control group were more likely to be reporting participation in VESA group discussions that touched on similar topics, rather than participating in the more intensive male engagement groups. Among those participating in a male engagement group, the average number of times these men attended in the last 12 months is only 2.7. The recalled topics that were asked about largely overlap with the topics covered at the main VESA discussions, so again respondents' answers might be confounding the two. Nearly 74 percent of men who responded indicate covering gender roles and division of labor, roughly 51 percent recall caregiving and related division of labor, and 47 percent indicate discussing decision-making within the household. This roughly aligns with what was observed regarding VESA gender discussions (see Table 6.7).

Figure 6.7: Participation in male engagement groups, last 12 months



6.3 Costs of participation to participants

For the first time, the endline survey asked questions about any costs that respondents incurred due to their participation in SPIR activities. The respective module was administered to women in treatment groups T1, T2, and T3, although in the interest of saving time, only a random subsample of all households got the module. The summary statistics are reported in Table 6.9.

During the previous months, the SPIR activity with the highest number of times attended were the VESA group discussions, with the average being two times across all household members. For all other activities (SPIR value chain trainings, male engagement groups, food demonstrations, BCC sessions, CLTSH events), the average number of times attended was less than one. In terms of how long it took the respondents to travel to SPIR activities and events, across the six activities the average travel time one

way was 25 minutes. SPIR value chain trainings had the longest reported travel time of 34 minutes, and VESA group discussions the shortest time of 21 minutes.

Finally, the respondents were asked about any out-of-pocket costs—childcare, transportation, meals and such—that they incurred due to participating in SPIR events. CLTSH events and VESA group discussions have the lowest associated costs at 2.5 and 4.3 birr per event, respectively. Similar to the travel time, SPIR value chain trainings also come with the highest reported out-of-pocket cost at 20.9 birr per event.

6.4 Conclusion

This chapter summarized results from the endline survey on participation in the PSNP4 and SPIR activities, focusing on effectiveness of delivery and fidelity of the interventions to the experimental design. The evidence shows that the SPIR project has expanded its reach since the midline study in 2019 for some program activities, consistent with the sequential nature of VESA discussions and other activities. In the main treatment arms (T1, T2, T3), reported participation in VESA groups is high, but not universal, with 80–86 percent of households reporting that they have a member participating in a group.[22] The SPIR project also closed the gender gap in participation in the period after the midline survey: VESA membership rates reported by men and women were roughly the same at endline. This was not the case at baseline and midline, when men participated at higher rates. Similarly, most respondents report high numbers of women in their VESA management committees, and approximately three-fourths of men and women say that their VESA has a female leader.

While participation in value chain trainings has risen considerably since the midline study and is now around a quarter of the sample, the share of those who are part of a producer marketing group has dropped to less than 8 percent. For general VESA discussions, three-fourths of members indicate attending meetings regularly (weekly or monthly) and the survey respondents show rather high recall rates of topics of interest to the program: IYCF practices, hygiene and sanitation, and gender dynamics within the household. Unfortunately, male engagement groups seem not to have taken off to the extent expected and reported attendance rates are rather low.

[22] Given that according to self-reports about 6 percent of households are likely receiving PDS and 4 percent are not part of either PW or direct support, a membership rate in the range of 90–94 percent would be considered universal. See more detail in Section 6.1.

Table 6.1: VESA membership and group characteristics

	N	All	Amhara	Oromia	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Either primary male or primary female is a member of a VESA group	3,793	0.660 (0.474)	0.662 (0.473)	0.657 (0.475)	0.799 (0.401)	0.855 (0.352)	0.818 (0.386)	0.147 (0.354)
Primary female is a member of a VESA group	3,674	0.642 (0.480)	0.638 (0.481)	0.646 (0.478)	0.778 (0.416)	0.828 (0.378)	0.799 (0.401)	0.137 (0.344)
Number of females that hold positions on primary female's VESA group's Management Committee	2,285	3.889 (5.769)	3.713 (6.995)	4.122 (3.546)	4.175 (4.613)	3.628 (7.779)	3.777 (3.817)	4.527 (6.717)
Primary female's VESA group has a female chairperson or a female vice chairperson	2,313	0.745 (0.436)	0.650 (0.477)	0.869 (0.338)	0.766 (0.424)	0.731 (0.444)	0.762 (0.426)	0.602 (0.492)
Primary female's VESA group conducted a share-out of the group savings/shares	2,230	0.659 (0.474)	0.502 (0.500)	0.847 (0.360)	0.684 (0.465)	0.678 (0.468)	0.652 (0.477)	0.408 (0.494)
Primary male is a member of a VESA group	3,776	0.635 (0.481)	0.646 (0.478)	0.621 (0.485)	0.765 (0.424)	0.823 (0.382)	0.799 (0.401)	0.137 (0.344)
Number of females that hold positions on primary male's VESA group's Management Committee	2,332	3.714 (4.004)	3.422 (4.166)	4.123 (3.727)	3.951 (4.278)	3.345 (3.375)	3.772 (3.863)	4.333 (6.181)
Primary male's VESA group has a female chairperson or a female vice chairperson	2,371	0.739 (0.440)	0.635 (0.482)	0.879 (0.326)	0.773 (0.419)	0.724 (0.447)	0.745 (0.436)	0.587 (0.494)
Primary male's VESA group conducted a share-out of the group savings/shares	2,395	0.688 (0.463)	0.578 (0.494)	0.839 (0.367)	0.733 (0.443)	0.692 (0.462)	0.675 (0.469)	0.484 (0.502)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses.

Table 6.2: Value chain development activities, females

	N	All	Amhara	Oromia	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Participated in SPIR value chain trainings	3,699	0.203 (0.403)	0.138 (0.345)	0.291 (0.454)	0.217 (0.412)	0.288 (0.453)	0.252 (0.434)	0.049 (0.216)
Participated in sheep/goat fattening or rearing training	752	0.419 (0.494)	0.249 (0.433)	0.527 (0.500)	0.356 (0.480)	0.417 (0.494)	0.478 (0.501)	0.409 (0.497)
Participated in poultry production training	752	0.750 (0.433)	0.717 (0.451)	0.771 (0.420)	0.756 (0.430)	0.760 (0.428)	0.763 (0.426)	0.591 (0.497)
Participated in ox fattening training	752	0.016	0.017	0.015	0.015	0.026	0.000	0.045

		(0.125)	(0.130)	(0.123)	(0.120)	(0.159)	(0.000)	(0.211)
Participated in vegetable production training	752	0.073	0.082	0.068	0.054	0.085	0.091	0.000
		(0.261)	(0.275)	(0.251)	(0.226)	(0.279)	(0.288)	(0.000)
Member of a producer marketing group associated with a SPIR promoted value chain	3,701	0.052	0.019	0.097	0.056	0.073	0.067	0.010
		(0.222)	(0.136)	(0.296)	(0.230)	(0.261)	(0.251)	(0.100)
Producer marketing group is associated with sheep/goat fattening	192	0.286	0.200	0.309	0.250	0.377	0.194	0.444
		(0.453)	(0.405)	(0.464)	(0.437)	(0.488)	(0.398)	(0.527)
Producer marketing group is associated with poultry production	192	0.635	0.650	0.632	0.712	0.522	0.726	0.444
		(0.483)	(0.483)	(0.484)	(0.457)	(0.503)	(0.450)	(0.527)
Producer marketing group is associated with ox fattening	192	0.005	0.000	0.007	0.000	0.000	0.016	0.000
		(0.072)	(0.000)	(0.081)	(0.000)	(0.000)	(0.127)	(0.000)
Producer marketing group is associated with vegetable production	192	0.068	0.125	0.053	0.038	0.087	0.065	0.111
		(0.252)	(0.335)	(0.224)	(0.194)	(0.284)	(0.248)	(0.333)
Bought inputs or sold products collectively as a group	193	0.280	0.475	0.229	0.283	0.261	0.274	0.444
		(0.450)	(0.506)	(0.421)	(0.455)	(0.442)	(0.450)	(0.527)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses.

Table 6.3: Value chain development activities, males

	N	All	Amhara	Oromia	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Participated in SPIR value chain trainings	2,462	0.231	0.205	0.258	0.247	0.314	0.278	0.062
		(0.421)	(0.404)	(0.438)	(0.431)	(0.465)	(0.448)	(0.242)
Participated in sheep/goat fattening or rearing training	568	0.599	0.482	0.695	0.569	0.602	0.598	0.714
		(0.491)	(0.501)	(0.461)	(0.497)	(0.491)	(0.492)	(0.458)
Participated in poultry production training	568	0.525	0.514	0.534	0.575	0.521	0.515	0.371
		(0.500)	(0.501)	(0.500)	(0.496)	(0.501)	(0.501)	(0.490)
Participated in ox fattening training	568	0.048	0.019	0.071	0.046	0.057	0.036	0.057
		(0.213)	(0.138)	(0.257)	(0.210)	(0.232)	(0.186)	(0.236)
Participated in vegetable production training	568	0.100	0.082	0.116	0.072	0.090	0.154	0.029
		(0.301)	(0.274)	(0.320)	(0.259)	(0.287)	(0.362)	(0.169)
Member of a producer marketing group associated with a SPIR promoted value chain	2,462	0.061	0.018	0.104	0.068	0.082	0.074	0.012
		(0.238)	(0.134)	(0.306)	(0.251)	(0.275)	(0.262)	(0.111)
Producer marketing group is associated with sheep/goat fattening	147	0.395	0.318	0.408	0.366	0.527	0.227	0.571
		(0.490)	(0.477)	(0.493)	(0.488)	(0.504)	(0.424)	(0.535)

Producer marketing group is associated with poultry production	147	0.497 (0.502)	0.409 (0.503)	0.512 (0.502)	0.561 (0.502)	0.382 (0.490)	0.614 (0.493)	0.286 (0.488)
Producer marketing group is associated with ox fattening	147	0.007 (0.082)	0.000 (0.000)	0.008 (0.089)	0.000 (0.000)	0.000 (0.000)	0.023 (0.151)	0.000 (0.000)
Producer marketing group is associated with vegetable production	147	0.075 (0.264)	0.273 (0.456)	0.040 (0.197)	0.073 (0.264)	0.073 (0.262)	0.091 (0.291)	0.000 (0.000)
Bought inputs or sold products collectively as a group	149	0.523 (0.501)	0.522 (0.511)	0.524 (0.501)	0.643 (0.485)	0.509 (0.505)	0.444 (0.503)	0.429 (0.535)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses.

Table 6.4: VESA discussions on health and nutrition, females

	N	All	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Has attended a VESA group meeting/discussion in past 12 months	3,690	0.483 (0.500)	0.631 (0.483)	0.608 (0.488)	0.586 (0.493)	0.089 (0.285)
VESA group meetings have been happening weekly	1,781	0.380 (0.485)	0.397 (0.490)	0.373 (0.484)	0.396 (0.490)	0.188 (0.393)
Number of times health and nutrition topics were discussed in VESA group	1,746	2.514 (5.731)	2.637 (6.057)	2.343 (6.200)	2.663 (5.170)	1.736 (1.823)
IYCF practices were covered at VESA meetings	1,401	0.449 (0.498)	0.462 (0.499)	0.414 (0.493)	0.438 (0.497)	0.690 (0.467)
Maternal nutrition was covered at VESA meetings	1,401	0.390 (0.488)	0.434 (0.496)	0.345 (0.476)	0.389 (0.488)	0.362 (0.485)
Nutritious food preparation was covered at VESA meetings	1,401	0.370 (0.483)	0.383 (0.487)	0.381 (0.486)	0.349 (0.477)	0.362 (0.485)
COVID-19 and prevention of its spread were covered at VESA meetings	1,401	0.191 (0.393)	0.198 (0.399)	0.191 (0.394)	0.184 (0.388)	0.172 (0.381)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses.

Table 6.5: VESA discussions on health and nutrition, males

	N	All	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Has attended a VESA group meeting/discussion in past 12 months	3,128	0.480 (0.500)	0.583 (0.493)	0.621 (0.485)	0.594 (0.491)	0.101 (0.302)
VESA group meetings have been happening weekly	1,501	0.333 (0.471)	0.386 (0.487)	0.325 (0.469)	0.311 (0.464)	0.197 (0.401)
IYCF practices were covered at VESA meetings	1,136	0.409 (0.492)	0.372 (0.484)	0.368 (0.483)	0.456 (0.499)	0.589 (0.496)
Maternal nutrition was covered at VESA meetings	1,136	0.321 (0.467)	0.383 (0.487)	0.275 (0.447)	0.321 (0.467)	0.232 (0.426)
Nutritious food preparation was covered at VESA meetings	1,136	0.342 (0.474)	0.369 (0.483)	0.320 (0.467)	0.334 (0.472)	0.357 (0.483)
COVID-19 and prevention of its spread were covered at VESA meetings	1,136	0.194 (0.395)	0.184 (0.388)	0.239 (0.427)	0.164 (0.371)	0.161 (0.371)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses.

Table 6.6: VESA discussions on gender dynamics, females

	N	All	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Gender topics/couple's relationships were discussed at least once at VESA meetings	1,781	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
Number of times gender topics/couple's relationships were discussed at VESA meetings	1,781	2.734 (2.835)	2.818 (3.152)	2.736 (3.236)	2.639 (2.059)	2.725 (1.518)
Workload sharing was covered at VESA meetings	1,781	0.695 (0.461)	0.704 (0.457)	0.685 (0.465)	0.697 (0.460)	0.675 (0.471)
Male engagement in childcare was covered at VESA meetings	1,781	0.462 (0.499)	0.492 (0.500)	0.420 (0.494)	0.463 (0.499)	0.525 (0.503)
Respect or mutual understanding was covered at VESA meetings	1,781	0.490 (0.500)	0.478 (0.500)	0.466 (0.499)	0.541 (0.499)	0.412 (0.495)
Shared decision-making on household finance was covered at VESA meetings	1,781	0.280 (0.449)	0.301 (0.459)	0.285 (0.452)	0.262 (0.440)	0.200 (0.403)
Violence against women was covered at VESA meetings	1,781	0.241	0.231	0.241	0.243	0.300

		(0.428)	(0.422)	(0.428)	(0.430)	(0.461)
Has observed changes in spouse's behavior or actions as a result	1,610	0.576	0.622	0.533	0.593	0.433
		(0.494)	(0.485)	(0.499)	(0.492)	(0.499)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses.

Table 6.7: VESA discussions on gender dynamics, males

	N	All	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Gender topics/couple's relationships were discussed at least once at VESA meetings	1,181	0.754	0.785	0.688	0.813	0.667
		(0.431)	(0.411)	(0.464)	(0.390)	(0.477)
Number of times gender topics/couple's relationships were discussed at VESA meetings	1,181	2.199	2.202	2.002	2.391	2.511
		(3.574)	(2.215)	(5.169)	(2.205)	(3.174)
Workload sharing was covered at VESA meetings	891	0.799	0.771	0.797	0.830	0.800
		(0.401)	(0.421)	(0.403)	(0.376)	(0.407)
Male engagement in childcare was covered at VESA meetings	891	0.613	0.582	0.608	0.654	0.567
		(0.487)	(0.494)	(0.489)	(0.477)	(0.504)
Respect or mutual understanding was covered at VESA meetings	891	0.631	0.575	0.626	0.682	0.733
		(0.483)	(0.495)	(0.485)	(0.467)	(0.450)
Shared decision-making on household finance was covered at VESA meetings	891	0.340	0.322	0.332	0.375	0.267
		(0.474)	(0.468)	(0.472)	(0.485)	(0.450)
Violence against women was covered at VESA meetings	891	0.323	0.305	0.322	0.332	0.433
		(0.468)	(0.461)	(0.468)	(0.472)	(0.504)
Has changed their behavior or actions as a result	862	0.666	0.646	0.676	0.708	0.333
		(0.472)	(0.479)	(0.469)	(0.456)	(0.480)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses.

Table 6.8: Male engagement groups

	N	All	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Attended a men's engagement group in past 12 months	3,130	0.305	0.393	0.330	0.402	0.085
		(0.461)	(0.489)	(0.471)	(0.491)	(0.279)
Number of men's engagement group meetings attended in past 12 months	955	2.693	2.777	2.490	2.699	3.095
		(2.332)	(2.598)	(2.117)	(1.911)	(3.463)
Gender roles and division of labor were covered at engagement groups	956	0.736	0.742	0.700	0.777	0.656

		(0.441)	(0.438)	(0.459)	(0.417)	(0.479)
Caregiving and division of caregiving were covered at engagement groups	956	0.506 (0.500)	0.523 (0.500)	0.468 (0.500)	0.524 (0.500)	0.500 (0.504)
Understanding nutrition was covered at engagement groups	956	0.348 (0.477)	0.345 (0.476)	0.338 (0.474)	0.373 (0.484)	0.281 (0.453)
Household decision-making was covered at engagement groups	956	0.472 (0.499)	0.442 (0.497)	0.475 (0.500)	0.489 (0.501)	0.516 (0.504)
Control of income, assets and resources were covered at engagement groups	956	0.381 (0.486)	0.326 (0.469)	0.433 (0.496)	0.395 (0.490)	0.359 (0.484)
Gender norms, beliefs and expectations were covered at engagement groups	956	0.290 (0.454)	0.319 (0.467)	0.289 (0.454)	0.260 (0.439)	0.297 (0.460)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses.

Table 6.9: Program participation costs in the last 3 months

	N	Number of times attended across all household members	Average time to travel to one event (minutes)	Value of out-of-pocket cost spent on one event (Birr)
SPIR value chain training	774	0.523 (1.576)	33.993 (38.706)	20.929 (77.600)
VESA group discussion	773	2.008 (3.757)	20.708 (25.573)	4.283 (23.040)
Male engagement group	770	0.410 (1.304)	23.353 (25.502)	7.716 (56.751)
Food demonstration	774	0.714 (3.881)	23.043 (25.304)	6.759 (34.967)
Behavior change communication (BCC) session	772	0.289 (1.070)	26.376 (30.110)	6.667 (20.869)
Community-led total sanitation and hygiene (CLTSH) event	774	0.492 (1.392)	23.916 (28.680)	2.479 (14.137)

Note: Estimates from the DFSA SPIR endline survey sample. Standard deviations are in parentheses. The questionnaire module was administered to a randomly selected subsample of all households in treatment groups T1, T2, and T3.

7 Impacts on livelihood outcomes

7.1 Introduction

This chapter evaluates the effects of the SPIR interventions on livelihoods outcomes, focusing particularly on the effects of the base L* and L interventions and the effects of the interventions cross-randomized in the first and second treatment arms: the poultry and cash transfers targeted to extremely poor households.

We focus on outcomes of interest that are plausibly linked to these interventions. In particular, we report effects on a range of variables linked to livestock ownership and sales, given the salience of the poultry transfer and livestock production value chain prioritization in SPIR programming. We also analyze income from livestock raising and cropping activities and household non-agricultural businesses and outside labor. We report experimental impacts on savings (particularly via VESAs) and credit access as reported by both men and women, and household assets and consumption.

7.1.1 Interpreting tables

For each set of outcomes, we present three tables: pooled effects by treatment arm and effects reported for the extremely poor and less poor. Here, we briefly summarize how to interpret each table.

The tables on pooled effects report coefficients for T1, T2, and T3. These coefficients capture the average effect of all interventions implemented in each arm (including, in T1 and T2, the cross-randomized transfers of poultry and cash to poor households). The tables also report tests of equality (T1=T2, T2=T3, T1=T3) that allow us to identify whether the experimental effect on the outcome of interest is significantly different comparing across the different treatment arms. The sample for these tables reporting pooled effects include all households with available data at endline, including the supplemental households added to the study sample at midline.

The tables reporting effects for the extremely poor restrict the sample across all four treatment arms to the households that were identified as eligible for cash or poultry transfers (the poorest 10 out of 18 baseline sample households in each kebele).[23] The tables then report coefficients for the effects of poultry and cash transfers implemented in conjunction with SPIR programming; each coefficient is reported separately for T1 and T2, the two arms in which there were cash and poultry interventions. The table also reports the coefficient for T3. Using these coefficients, we can calculate linear combinations (means) to estimate the average effect of T1 and T2 for poor households, and the average effect of poultry and cash. We again report the same tests of equality across treatment arms.

The tables reporting effects for the less poor restrict the sample across all four treatment arms to the households that were **not** eligible for cash or poultry transfers (the richest 8 out of 18 households in each

[23] There were, however, a small number of kebeles in which fewer than 18 households were sampled at baseline, and in which fewer than 10 households received a targeted transfer. A count of 192 households are drawn from kebeles in which 16 or fewer households (minimum 2) were sampled, and there is accordingly some variation in the number of households that received transfers.

kebele at baseline). The tables report coefficients for the effects of T1, T2, and T3. No households in the less poor group were assigned to receive poultry or cash transfers.

Each table reports standard errors estimated using conventional methods, and denotes statistical significance at standard levels with asterisks: one asterisk denotes statistical significance at the 10 percent level, two asterisks statistical significance at the 5 percent level, and three asterisks statistical significance at the 1 percent level. In addition, we report sharpened q-values corrected for multiple hypothesis testing following the method of Simes (1986) to adjust inference for the multiplicity of tests estimated; this adjustment is conducted for all outcomes reported in a given table. The discussion of results here will generally focus on results that are also statistically significant when corrected for multiple hypothesis testing, but will highlight some results that are significant using standard methods but not when the correction is employed.

In the next subsection of the report, we also report average standard treatment effects estimated for each family of outcomes, following Katz et al. (2007). This enables us to assess broader patterns of significance and magnitude for a related set of variables, and provides a useful overview of the findings.

7.2 Impacts on summary indices

Tables 7.2a and 7.3a present average standard treatment effects for each variable family that enable us to summarize the average magnitude and significance of effects for each outcome area. The outcome variables are defined in Table 7.1 below.

Table 7.1: Summary indices

Poultry production	Primary female reports owning any poultry Primary male reports household owns any poultry Total number of poultry reported owned by primary female Total number of poultry reported owned by household Primary female reports any income from poultry sales Primary male reports household has any income from poultry sales Primary female reports any income from egg sales
Other livestock production	Primary female reports owning any sheep or goats Primary male reports household owns any sheep or goats Primary male reports household owns any oxen Primary male reports household owns any other livestock Primary female reports any income from sales of sheep and goats Primary male reports any household income from sales of sheep and goats Primary male reports any household income from sales of oxen Primary male reports any household income from sales of other livestock
Income from livestock production and crop cultivation	Net income from sales of livestock products Income from all livestock sales Costs associated with livestock production Net income from livestock sales Household earned any income from crops cultivated (Mehr season) Income earned from crops cultivated (Mehr season)
Business and wage work	Household has non-agricultural business Household member reports regular wage work Primary female reports regular wage work

	Household member reports irregular/casual wage work Primary female reports irregular/casual wage work Primary male reports irregular/casual wage work
Household savings	Household has any savings Amount of savings Household has any savings at home or with a relative Household has any savings in the bank Household has any savings with an MFI Household has any savings with a VESA/VSLA Household has any savings with a RUSACCO Primary male reports household membership of VESA/VSLA Primary male reports household membership of RUSACCO Primary male reports household has a bank account
Female savings	Primary female has personal savings of her own Primary female amount of savings Primary female deposits her savings with a VESA/VSLA Primary female deposits her savings with RUSACCO Primary female has any savings with an MFI
Household credit access	Household reports any loan for productive purposes Household obtained a productive loan from VESA/VSLA Amount of productive loan Household reports any loan for consumption purposes
Female credit access	Primary female reports any loan for productive purposes Primary female obtained a productive loan from VESA/VSLA Amount of productive loan Primary female reports any loan for consumption purposes
Household aggregate assets	Estimated value of livestock owned Estimated value of livestock owned by the female Household livestock asset index Household productive asset index Consumer durable asset index Household total asset index
Most owned productive assets and consumer durables	Household owns at least one sickle Household owns at least one axe Household owns at least one spade or shovel Household owns at least one solar panel Household owns at least one blanket Household owns at least one flashlight/torch
Housing characteristics	Household has an improved source of water (rainy season) Household has improved roof material Household's number of bedrooms Household has access to electricity
Consumption	Total value of monthly food consumption per adult equivalent (birr) Monthly expenditure on non-food items per adult equivalent (birr) Total consumption expenditure per month per adult equivalent (birr) Calories of daily food consumption per adult equivalent (kcal)

We can observe in Table 7.2a that there are substantial effects of all three treatment arms on poultry production: the effect is around .2 standard deviations in T1 and T2, and around .1 standard deviations in T3. For other (nonpoultry) livestock production and income from livestock production, we observe

increases of around .06 standard deviations in T1 and T2 only. There is no evidence of any significant shift in income from cropping or labor and wage work in any arm.

Tables 7.2b and 7.2c show the same average standard treatment effects (ASTEs) for the subsamples of extremely poor and less poor households. In Table 7.2b, we observe that the effects for livestock variables are larger in magnitude and more precisely estimated for extremely poor households: an increase of between .2 and .4 standard deviations for poultry production, and around .1 standard deviations for other livestock production and livestock income. Interestingly, we observe that for poultry-related outcomes, the estimated coefficients are significantly larger for poultry households vis-à-vis cash recipient households, as evident in the p-value reported at the base of the table; while for secondary outcomes linked to nonpoultry livestock, the estimated coefficients are generally larger for cash recipients vis-à-vis poultry recipients, and the difference is significant at the 10 percent level (p=0.095). For less poor households as reported in Table 7.2c, the effects on all livestock variables are smaller in magnitude and generally not statistically significant, with the exception of a significant increase in poultry production in T3 and a significant increase in outside labor and wage work in T1.

In Table 7.3a, we report effects for a series of additional variables for the pooled sample. We see significant shifts in both household savings and female savings, as well as household credit and female credit. There is, however, no evidence of any substantial effect on assets or housing characteristics. In Tables 7.3b and 7.3c, the effects on savings are relatively consistent across subsamples, though for the extremely poor sample (reported in Table 7.3b), the positive effect on female savings is significantly larger in T1 and T2 vis-à-vis T3. For credit access, the effects are larger (.2 to .4 standard deviations) and often statistically significant only for the less poor subsample.[24]

For aggregate assets, there is an increase of around .1 standard deviation for the extremely poor sample only (primarily in T1 and T3), but not in the less poor sample. The less poor sample shows some enhancement in housing characteristics in arms T1 and T2 (magnitude around .15 standard deviations, but significant only at the 10 percent level). There are no significant effects on consumption for any subsample.

7.3 Livestock ownership and sales

Next, we analyze a series of variables capturing the household's ownership and sales of livestock. Note that for all variables, women report information about livestock that they own (solely or jointly), while men report information about poultry owned at the household level.

In analyzing these effects, it is important to note two key contextual points around the SPIR program. First, the cash transfer provided was unconditional; there was no requirement that households use the cash to engage in livestock fattening or rearing. Second, households could self-select into livestock-specific value chain trainings, and as noted previously in the evidence around program participation in Chapter 6,

[24] There is, however, a large and statistically significant increase in the index of female access to credit for cash recipients in T1.

only around 25–30 percent of households participated. The observed magnitudes of effects should be interpreted with these points in mind.

7.3.1 Pooled sample

Table 7.4a reports pooled effects for the three treatment arms for variables linked to poultry production. It is evident that in the first two treatment arms (T1 and T2), there are generally positive effects on a range of variables linked to livestock production, and the observed effects are parallel across arms. We observe an increase in the probability the woman or household reports any poultry owned of between 8 and 11 percentage points in columns (1) and (2), relative to a base probability of 66 percent for the household owning any poultry in the control arm. The total number of poultry reported owned by the woman herself and the household increases by between 0.5 and 0.8, as reported in columns (3) and (4), relative to a mean in the control arm of around two, suggesting the average size of poultry flocks increased by about 25 percent.

When we examine variables linked to reported income from sales of poultry, we also observe a substantial increase in the probability that women and men report income from poultry sales in T1, between 7 and 10 percentage points relative to a mean of 28 percent (at the household level) in the control arm. There is also some weaker evidence of an increase in the probability of household sales of poultry that is around 5 percentage points in T2 and T3, significant at the 10 percent level in T2, and insignificant when corrected for multiple hypothesis testing in T3. There is no evidence of any statistically significant shift in the probability of reporting income from sales of eggs. These effects are captured graphically in Figure 7.1.

Figure 7.1: Household livestock production: Full sample

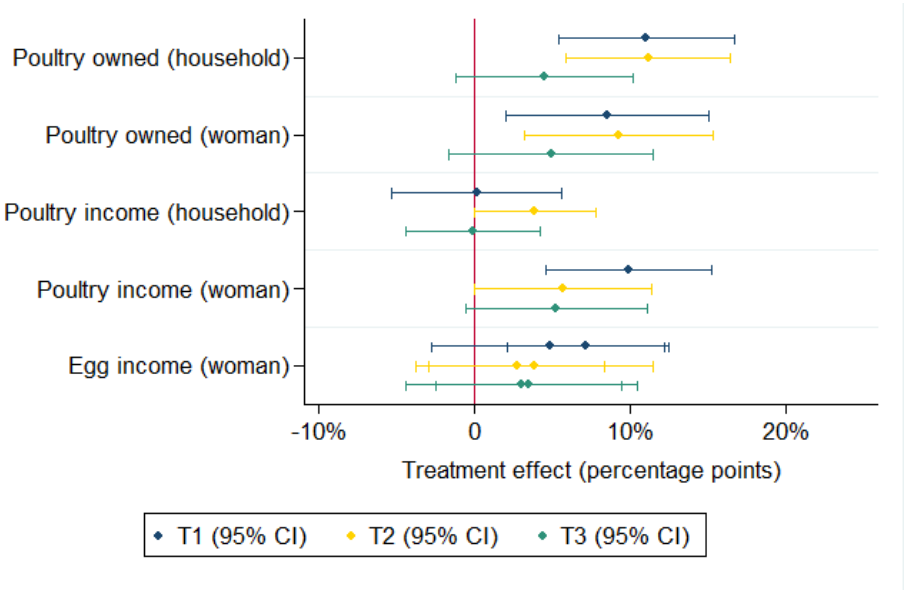
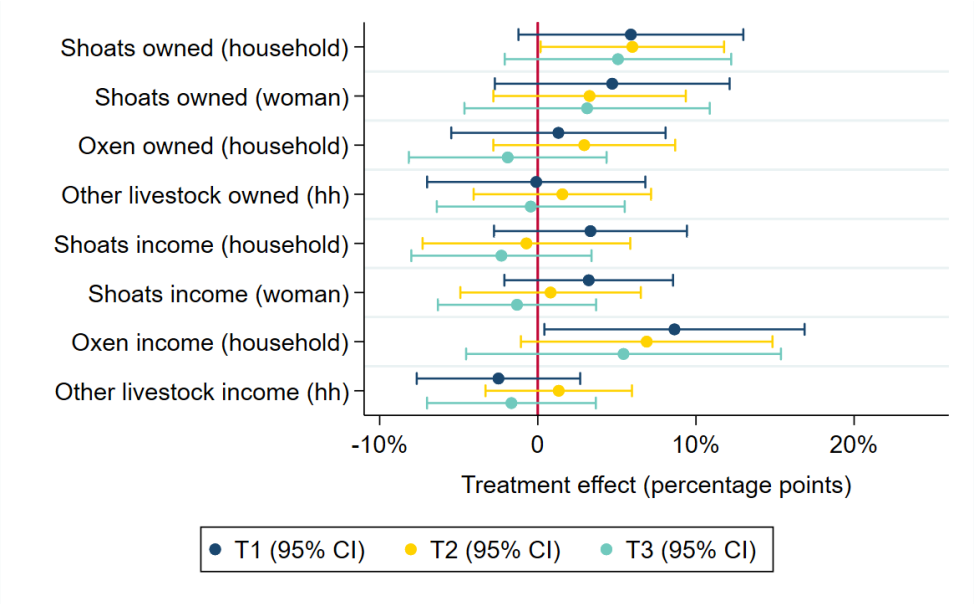


Table 7.5a and 7.6a then report additional results linked to ownership and sales of other livestock (in Table 7.5a) and income and costs of livestock production as well as cropping (in Table 7.6a), again in the pooled sample. In Table 7.5a, there is generally little evidence of any significant effects on ownership and sales of sheep, goats, and oxen. Households in T2 show an increase of 6 percentage points in the probability of owning any sheep or goats (though this coefficient is not robust to correction for multiple hypothesis testing), and households in T1 and T2 show around a 7–9 percentage point increase in the probability of reporting any income from oxen sales (though again, this coefficient is not robust for correction for MHT). These effects are presented in Figure 7.2.

Figure 7.2: Household livestock production (secondary outcomes): Full sample



In Table 7.6a, we analyze the effects on overall income and net income from livestock production as well as cropping; continuous measures of income and net income are transformed using an inverse hyperbolic sine. This method of transformation can be considered analogous to a log transformation (and thus reduces the influence of outliers), but also allows for values that are negative or zero. We can interpret the coefficients (using a simple transformation) as semi-elasticities.[25] The results suggest that assignment to a SPIR treatment arm generates an increase of around 30 percent in net income from sales of livestock products, and 90 percent in income from livestock sales; the increase in sales of livestock products is observed in all three arms, while the increase in net income from livestock sales is observed only in T1

[25] More specifically, the elasticity is calculated as follows, using the coefficient β: exp (β-var(β))-1.

and T2.[26] There is also some shift in costs associated with livestock, however, and thus the increase in net income from livestock production is only marginally significant, and insignificant when corrected for multiple hypothesis testing.

We also report parallel measures for agricultural cropping in columns (5) and (6); cropping activities were generally not a focus of SPIR livelihoods activities, and thus as expected we observe null effects here.

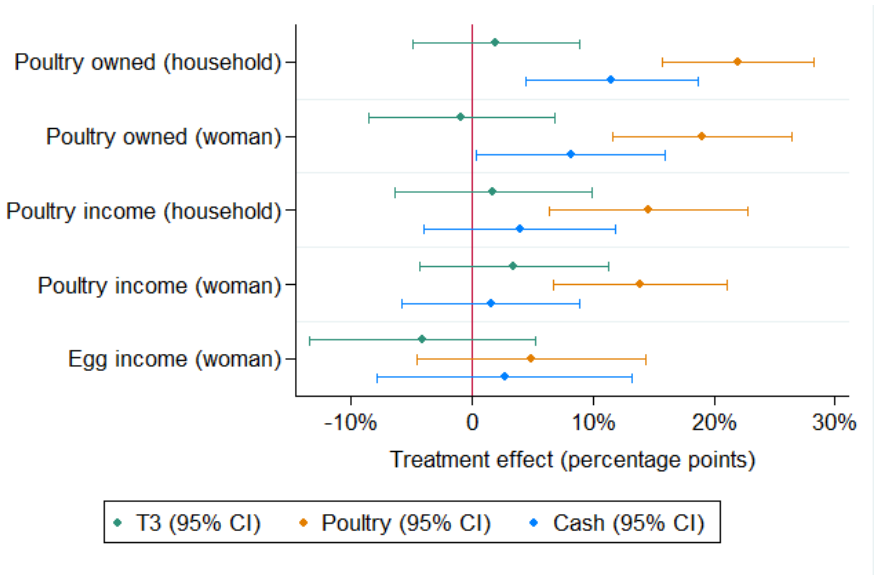
7.3.2 Extremely poor sample

Table 7.4b reports effects for the extremely poor sample; here, we will focus on the average effects of the poultry and cash transfers that targeted only this sample. We can observe in columns (1) and (2) that poultry households remain about 20 percentage points more likely to own poultry vis-à-vis control households, where 60 percent report ownership of poultry; for cash households, the corresponding increase in poultry ownership is about 11 percentage points. (Corresponding effects are observed for women reporting ownership of own poultry.) We can observe in columns (3) and (4) that poultry recipient households own around 2 more poultry on average, while cash recipient households own around 0.5 additional chickens.

In columns (5) through (7), poultry recipient households are much more likely to report any income from sales of poultry (an increase of 14 percentage points, relative to a probability of 30 percent in the control arm), but again there is no effect on egg sales. One interpretation of this pattern is that following the sale of the eight male chickens designated for sale within six months of the transfer, households retained eight chickens (as observed at midline) and over time experienced further decline in the flock due to mortality and sales as chickens passed the productive egg-production age: hence we observe a significant effect on reported income from poultry sales, but not egg sales. It is also important to note that the effects on poultry-related production outcomes are almost uniformly larger for poultry recipients vis-à-vis cash recipients, perhaps unsurprisingly. These effects are captured graphically in Figure 7.3.

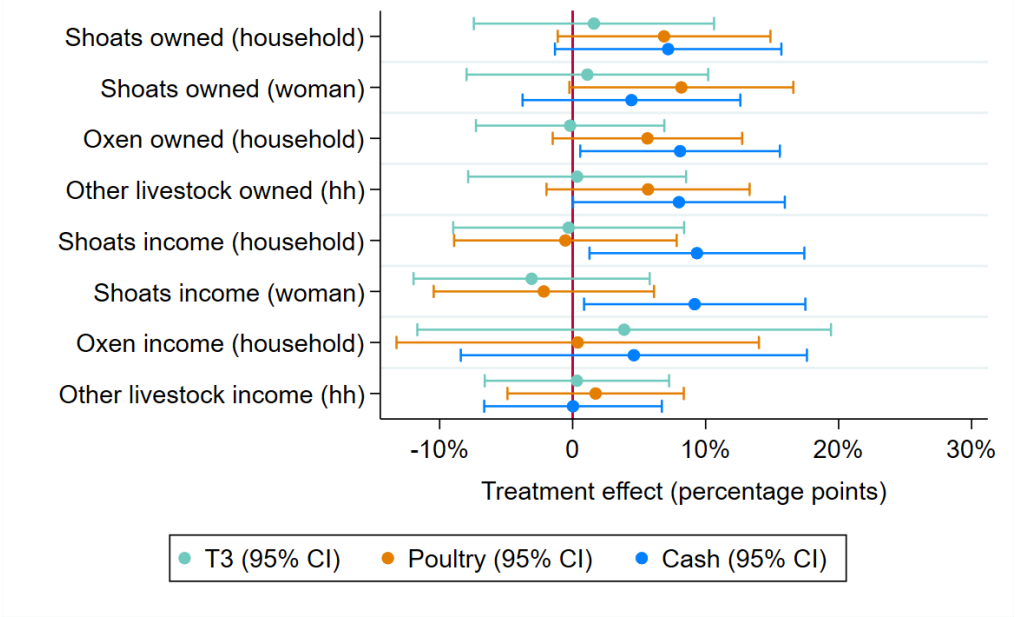
[26] Again, the estimated semi-elasticities are calculated using the formula exp (β-var(β))-1. For estimated treatment coefficients that are smaller in magnitude, the estimated semi-elasticity is roughly similar to the estimated coefficient. However, for estimated coefficients that are larger (i.e., the estimated coefficients on livestock sales are around .7 for T1 and T2), the estimated semi-elasticity is itself even larger (.9). This is consistent with the observation in Bellemare and Wichman (2020) that directly interpreting the coefficient estimated in a regression using a dependent variable with an inverse hyperbolic sine transformation as a semi-elasticity can lead to a significant underestimate of the true semi-elasticity.

Figure 7.3: Household livestock production: Extremely poor sample



Parallel results for other livestock production and income are reported in Table 7.5b and Table 7.6b. Here, we generally see weak evidence that poultry-recipient households report any additional ownership of sheep or goats; only two estimated coefficients are marginally significant, and they are not robust to correction for multiple hypothesis testing. However, cash-recipient households are more likely to report household ownership of sheep or goats, oxen, and other livestock; and more likely to report income from sales of sheep or goats. The estimated coefficients are around 7–9 percentage points in magnitude, relative to a mean probability of ownership in the control arm of 57 percent for sheep or goats, 22 percent for oxen, and 61 percent for other livestock; thus, the proportional effect is particularly large for oxen. These estimates are generally significant at the 10 percent level when corrected for multiple hypothesis testing. These effects are presented in Figure 7.4.

Figure 7.4: Household livestock production (secondary outcomes): Extremely poor sample



In Table 7.6b, we again observe evidence of a substantial increase in net income from sales of livestock products for both poultry and cash recipients, suggesting an increase of nearly 90 percent for poultry recipients and 25 percent for cash recipients. There is also an increase in total income from sale of livestock for both poultry and cash recipients of nearly 200 percent. The cost of livestock raising also increases, and thus an increase in net income from livestock raising is statistically significant only for cash recipients, and the coefficient is marginally significant when corrected for multiple hypothesis testing. However, the estimated coefficients are extremely large in magnitude (75 percent increase in net income for cash recipients, and 170 percent increase in net income for poultry recipients), and the hypothesis that the effects are equal in magnitude comparing across poultry and cash recipients cannot be rejected.

7.3.3 Less poor sample

Table 7.4c reports effects for poultry production for the less poor sample. Here, there is only weak evidence of any significant effects for these households, none of whom received cash or poultry transfers. All three arms show an increase in the probability of any poultry ownership of between 6 and 8 percentage points, and an increase in the number of poultry owned of around 0.5 chickens; however, these coefficients are not statistically significant when corrected for multiple hypothesis testing. The estimated coefficients are captured graphically in Figure 7.5.

Figure 7.5: Household livestock production: Less poor sample

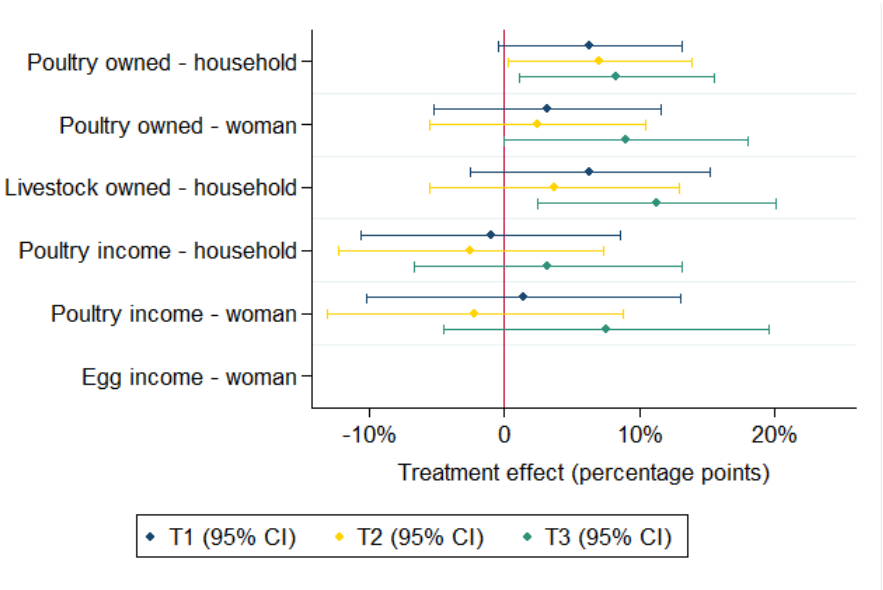


Table 7.5c reports effects for other livestock production. Again, for this less poor sample, we observe no significant effects on ownership or sales of sheep or goats, oxen, or other livestock. Table 7.6c reports effects for income from livestock raising and cropping. There is no robust evidence of any significant and positive treatment effects here; in fact, income from cropping may be weakly declining, but the estimated p-values are not robust to corrections for multiple hypothesis testing.

7.3.4 Comparison to observed effects at midline

When we compare these estimated coefficients for variables linked to poultry with the corresponding estimates in the midline survey, conducted approximately 18 months earlier, we can observe that there has been significant attenuation in the estimated treatment effects. This pattern is consistent with other evidence from the literature in which significant decay in the positive effects of cash or asset transfers has been observed over time, described in more detail in the concluding section. In the pooled sample, the increase in the probability of ownership of any poultry in the T1 and T2 arms attenuated by roughly 50 percent across waves. Similarly, for the extremely poor sample, the increase in the probability of ownership of any poultry attenuated by 50 percent (the coefficient decreasing from around .4 to .2). At midline, poultry-recipient households owned 8 more chickens and cash-recipient households owned 1.4 more chickens; at endline, these numbers are 2 and 0.5, respectively.

By contrast, the estimated effects for ownership and sales of other livestock exhibit a very different time pattern. Focusing on the estimated effects for the extremely poor sample, at midline both cash and poultry households reported an increase in the ownership of sheep or goats (around 12 percentage points), and some decline in the probability of reporting any income from sales of livestock of various types. This

observed decline in the probability of income would be consistent with both poultry- and cash-recipient households withholding animals from market sale in order to participate in fattening activities.

By endline, these effects on poultry households have attenuated to zero: these households are no more or less likely to report ownership of or income from other types of livestock vis-à-vis households in the control arm. It does not seem that poultry households have used the income to expand into higher-cost and higher-reward livestock fattening activities including sheep, goats, or oxen. However, cash-recipient households have maintained significantly higher ownership of sheep or goats (with minimal attenuation) and also report increased ownership of oxen and other livestock, as well as increased income from ownership of sheep or goats. These households appear to have successfully expanded their livestock production activities over time.[27]

The estimated treatment effects for the less poor sample are largely consistent over time comparing across the midline and endline sample. This consistency is logical, given that the less poor sample was not exposed to any one-time transfer, but rather was participating in broader SPIR value chain promotion activities that were expanding in coverage over time.

7.4 Non-agricultural household businesses and wage work

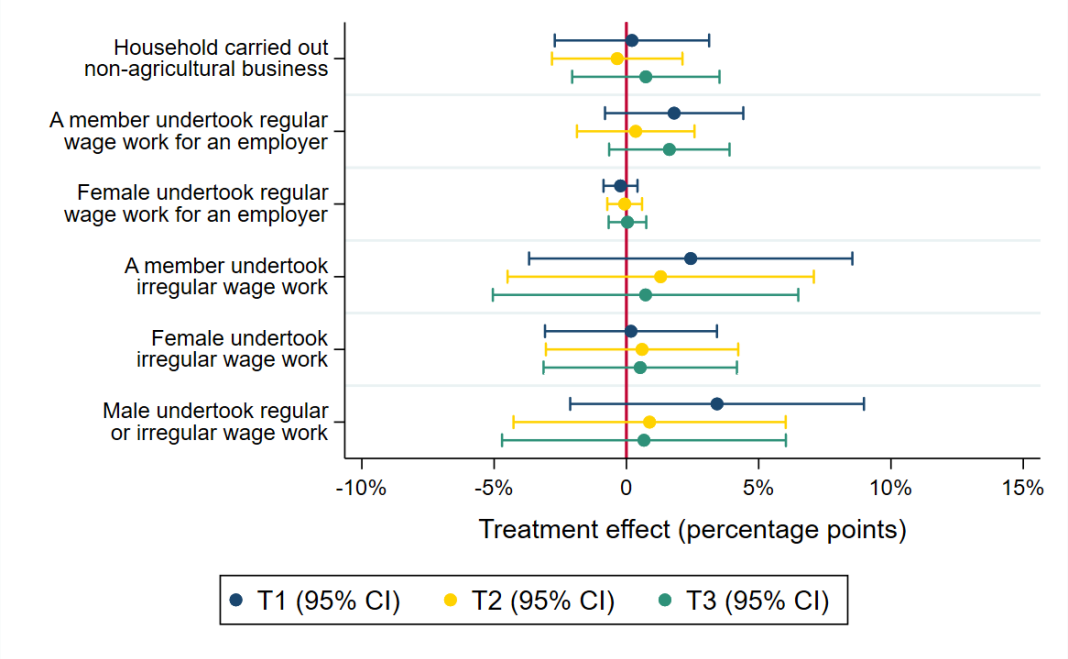
We also analyze a set of variables capturing whether the household reports any non-agricultural business, and whether any members (including the primary female and male) were engaged in regular wage work or irregular/casual wage work. Variables capturing wage work were also reported and analyzed at midline.

7.4.1 Pooled sample

The results reported in Table 7.7a suggest there is no evidence of any meaningful shift in household engagement in non-agricultural businesses or work, as the coefficients of interest are uniformly small in magnitude and statistically insignificant. It is also important to note that the average level of engagement in any non-agricultural activity is extremely low in the control arm. Less than 5 percent of households in the control arm report any non-agricultural household business, and only 3 percent report that any member was engaged in regular wage work. Engagement in irregular/casual wage work is more common, as nearly a quarter of households report that at least one member was engaged in this form of work at some point in the past year; 10 percent of primary female respondents and 16 percent of primary male respondents report past-year engagement in casual labor. These rates are also broadly similar to the midline survey, in which around a quarter of households also reported that at least one member was engaged in casual labor. The estimated effects for the pooled sample are captured in Figure 7.6.

[27] Data on income from and costs of livestock production were not collected at midline.

Figure 7.6: Business wage work: Full sample



7.4.2 Extremely poor sample

The results reported in Table 7.7b are generally consistent and again suggest there is no evidence of any meaningful shift in non-agricultural activities for extremely poor households. Households in T3 do show evidence of an increase in the probability of reporting any non-agricultural business, but this estimated coefficient is not robust to correction for multiple hypothesis testing.

7.4.3 Less poor sample

Again, the results reported in Table 7.7c are generally consistent and suggest there is little evidence of a meaningful shift. Households in T2 and T3 are here somewhat less likely to report any non-agricultural business, but again these estimated coefficients are not robust to correction for multiple hypothesis testing.

7.4.4 Comparison to midline

These findings are parallel to the findings at midline suggesting that there was no effect on non-agricultural businesses and outside labor.

7.5 Savings

We next report variables capturing savings for both the household as a whole (reported by the primary male respondent) and for women (as reported by the primary female respondent on her own behalf).

7.5.1 Pooled sample

Table 7.8a reports the estimated treatment effects for household-level savings for the pooled sample. We can observe in columns (1) and (2) that households in all three SPIR treatment arms are significantly more likely to report any savings: an increase of around 30 percentage points, relative to a probability of 47 percent in the control arm. This corresponds to an increase in savings of around 400 birr relative to a control mean of nearly 800 birr, or an increase of around 50 percent; this increase is largest in magnitude in T1 and smallest in magnitude in T3, but the hypothesis that the effects are consistent across arms cannot be rejected.

Conditional on reporting any savings, we observe in columns (3) through (7) that households are significantly more likely to report that they utilize a VESA/VSLA for savings, and less likely to report use of another financial institution. They are also significantly less likely to report that they save informally at home (a decline of 10 percentage points). Finally, in columns (8) through (10) we observe that households in all three treatment arms are around 60 percentage points more likely to report they are members of a VESA/VSLA, relative to a mean probability of only 13 percent in the control arm. There is no shift in the probability of membership in other financial institutions. These effects are also presented in Figure 7.7.

Figure 7.7: Savings, reported by male for the household: Full sample

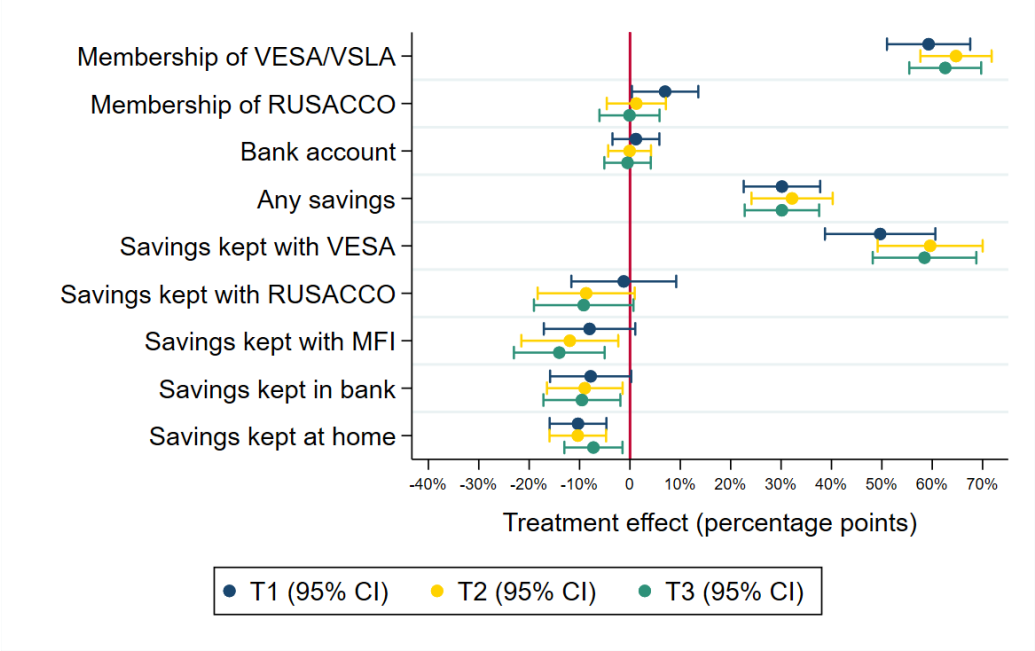
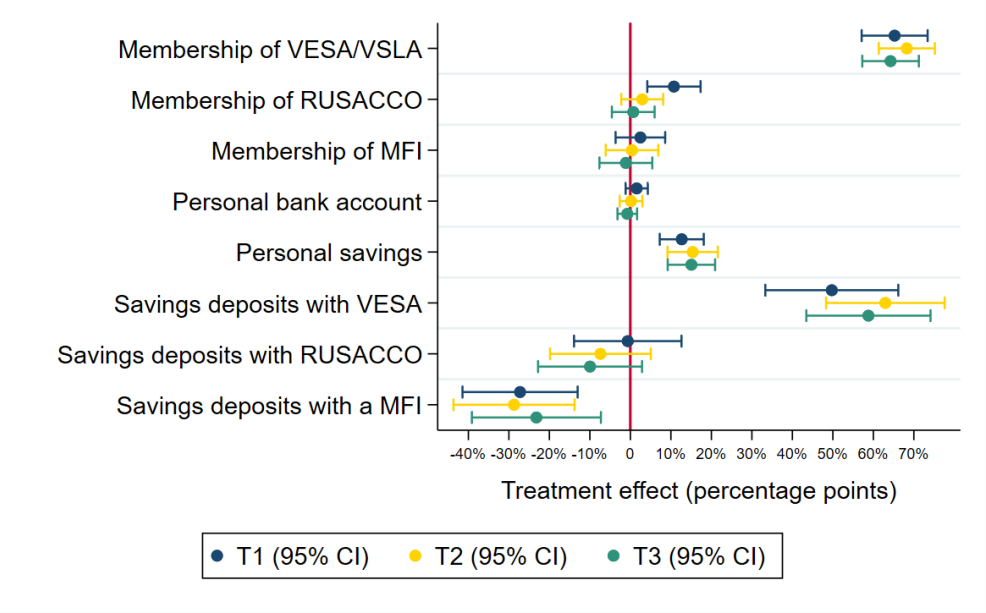


Table 7.9a reports parallel results for female savings. One unique feature of SPIR programming is that both spouses in each household were invited to become VESA members, and thus both can engage in savings. Again, we observe in column (1) that women are significantly more likely to report that they have savings of their own: an increase of between 13 and 15 percentage points relative to a mean in the control arm of only 10 percentage points. The probability of women reporting any savings has more than doubled. The continuous amount of savings, reported in column (2), is weakly higher but the difference is generally not statistically significant, suggesting that the marginal woman who is induced to save by the interventions amasses an amount of savings that is somewhat lower than the average level for women saving in the control arm. We can observe in columns (3) through (5) that, in parallel to the results reported for household savings, women are significantly more likely to report that they deposit their savings in a VESA/VSLA and significantly less likely to utilize other financial institutions. Again, in columns (6) through (8), women are much more likely to report membership in a VESA/VSLA, but there has been no shift in membership in other financial institutions. Again, there is no evidence of any heterogeneity across treatment arms in these effects. The effects are captured graphically in Figure 7.8.

Figure 7.8: Savings, reported by female for herself: Full sample



7.5.2 Extremely poor sample

Tables 7.8b and 7.9b report the estimated effects for household's and women's savings for the extremely poor sample. The estimated coefficients are generally consistent. The increase in the probability of household-level savings is even larger (around 34 percentage points, relative to a mean of 40 percent in the control arm), and the increase in the amount of savings remains around 350 birr relative to a mean in the control arm of 670 birr. For women, the estimated effects are again consistent with the full sample. There is no evidence of heterogeneity across treatment arms, or heterogeneity when comparing poultry recipient households and cash recipient households.

7.5.3 Less poor sample

Tables 7.8c and 7.9c report the estimated effects for the less poor sample. Again, we observe a generally consistent pattern. The increase in the amount of reported savings at the household level is larger, particularly in T1 and T2 (more than 700 birr, relative to a mean of 700 birr in the control arm), suggesting that savings has approximately doubled. In addition, the continuous amount of reported savings for women in fact shows a significant increase in the T1 and T3 arms of between 150 and 200 birr, relative to a mean of around 120 birr in the control arm: accordingly, women's savings have more than doubled. Again, the hypothesis that the effects are equal across arms can generally not be rejected.

7.5.4 Comparison to observed effects at midline

At midline, only savings as reported by the primary female was reported and analyzed. When we compare these estimated treatment effects to those reported at midline, we observe that the effects are also somewhat attenuated: the positive coefficient on female savings has decreased from around 40 percentage points to 12 percentage points. This may reflect the prevalence of adverse shocks since midline.

7.6 Credit access

We next report variables capturing credit access for both the household as a whole (reported by the primary male respondent) and for women (as reported by the primary female respondent on her own behalf).

7.6.1 Pooled sample

Tables 7.10a and 7.11a report estimated effects for credit access for the pooled sample. In Table 7.10a, we observe that there is little evidence of any significant shift in credit access, though there is an increase in the probability of accessing a production loan of 6 percentage points relative to a mean in the control arm of 26 percent that is marginally significant when corrected for multiple hypothesis testing. Conditional on reporting a productive loan, however, households in all three treatment arms are significantly more likely (around 10–15 percentage points) to obtain a loan from a VESA/VSLA.

In Table 7.11a, again we see little evidence of any significant effects on credit access for women. Conditional on reporting any productive loan, women are significantly more likely to access a loan from a VESA/VSLA (between 15 and 20 percentage points).

These results for both household-level credit and female credit are also presented in Figures 7.9 and 7.10.

Figure 7.9: Credit, reported by male for the household: Full sample

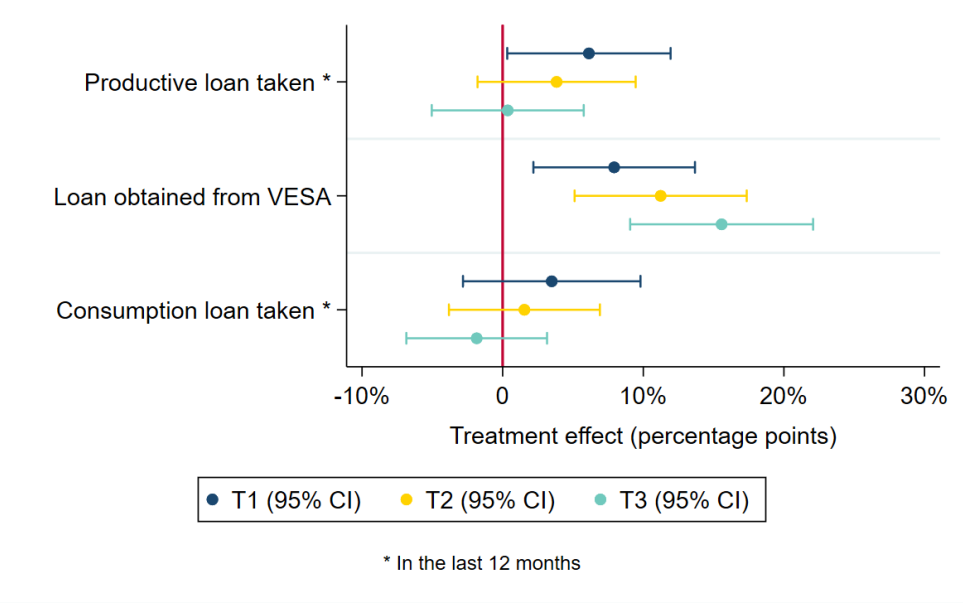
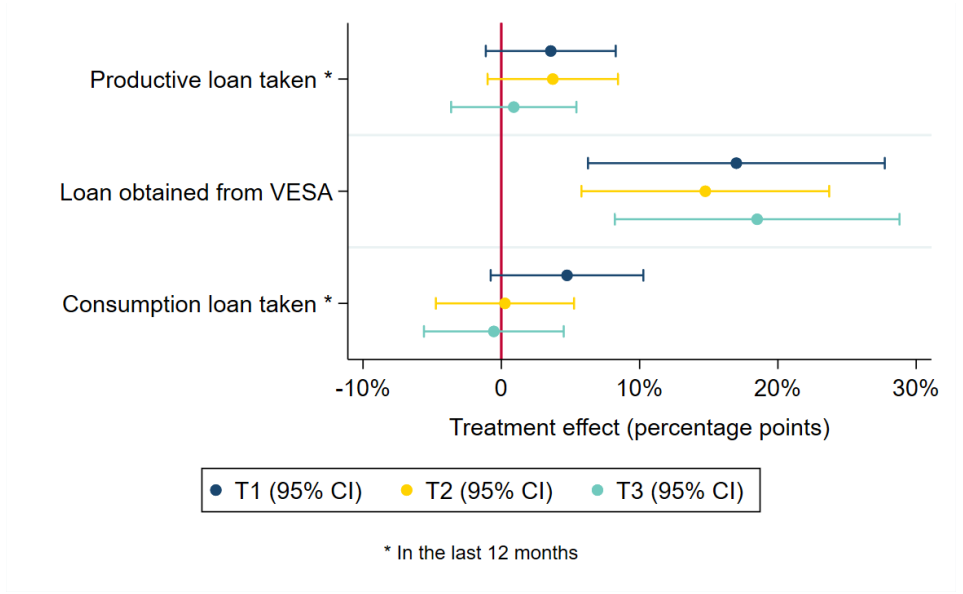


Figure 7.10: Credit, reported by female for herself: Full sample



7.6.2 Extremely poor sample

Tables 7.10b and 7.11b report estimated effects for the extremely poor sample. Here, there is somewhat more robust evidence of an increase in the probability of accessing a productive loan (around 7 percentage points) for poultry and cash recipient households in T1 and T2, though these estimates are not robust to multiple hypothesis testing. There is no strong evidence that households are more likely to access loans from VESAs/VSLAs, however, and no clear evidence of any significant treatment effect for women's credit access.

7.6.3 Less poor sample

Tables 7.10c and 7.11c report estimated effects for the less poor sample. Here, there is no evidence of any shift in the probability of accessing a productive loan. However, conditional on reporting a productive loan, households are more likely to report accessing a loan from a VESA/VSLA.

7.6.4 Comparison to midline results

When compared to midline results, in general the pattern is consistent; however, the increase in the probability of receiving a productive loan was not statistically significant at midline, and this effect has thus amplified over time. (Only credit access as reported by the primary male was analyzed at midline.)

7.7 Assets

We next report a series of variables capturing household assets. These include four asset indices (capturing productive assets, consumer durables, livestock assets, and total household assets) that are constructed using principal component analysis. We also report two additional livestock-related indices capturing the estimated value of livestock assets owned by the household and by the woman herself, valued using prices at the market for each kebele. For the livestock value variables, we again report results from a specification employing an inverse hyperbolic sine transformation.

To further probe effects on specific types of assets, we analyze a series of binary variables capturing ownership of the most common forms of durable goods and productive assets; and a series of variables capturing housing-related investments. Asset-related variables were not measured or analyzed at midline. Importantly, if we compare binary variables for the most common forms of durable goods and productive assets at endline vis-à-vis the baseline survey, we see evidence of a significant increase in asset ownership in general. At baseline, on average 54 percent of households reported owning a sickle (up to nearly 80 percent at endline), 40 percent an axe (over 70 percent at endline), 5 percent a spade or shovel (over 60 percent at endline), 39 percent a solar panel (over 70 percent at endline), 56 percent a blanket (roughly constant at endline), and less than 1 percent a flashlight or torch (over 60 percent at endline).

7.7.1 Pooled sample

The results reported in Table 7.12a suggest there is no evidence of any significant effects on any form of assets for the pooled sample. We similarly observe null effects for binary variables capturing assets in Table 7.13a, and for variables capturing housing characteristics in Table 7.14a. These effects are captured graphically in Figures 7.11, 7.12 and 7.13.

Figure 7.11: Household's aggregate assets: Full sample

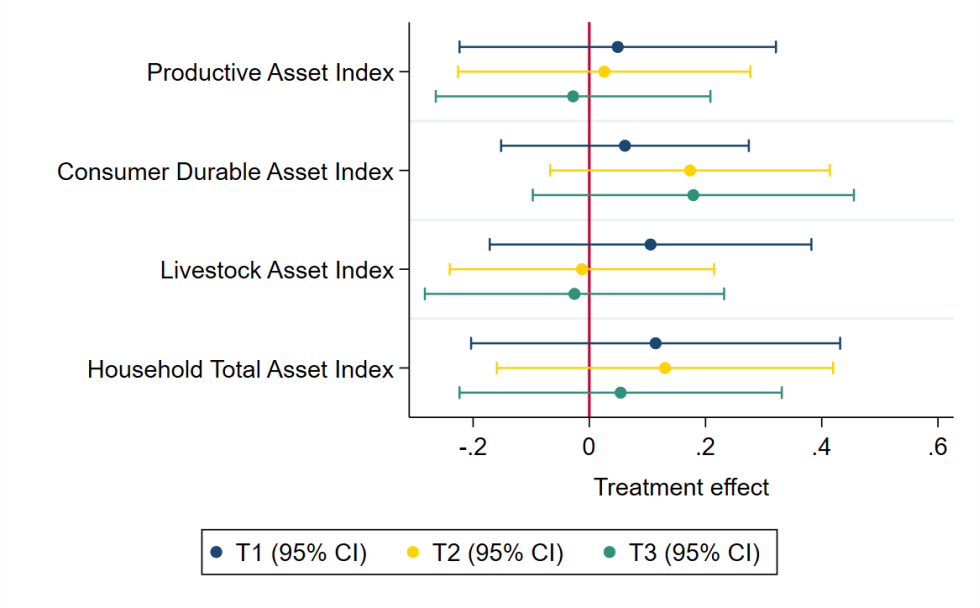


Figure 7.12: Ownership of most common productive assets and consumer durables: Full sample

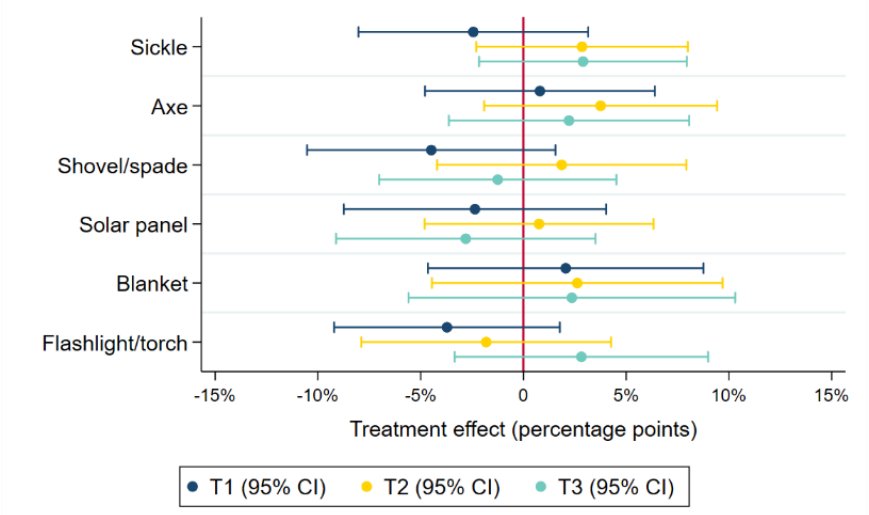
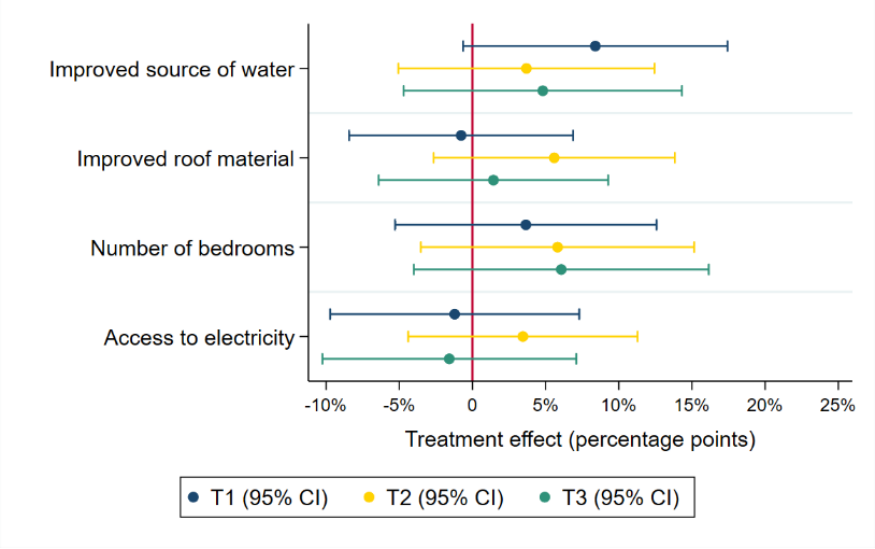


Figure 7.13: Household characteristics: Full sample



7.7.2 Extremely poor sample

The results reported in Table 7.12b suggest that there is a substantial increase in the estimated value of total livestock (130 percent) for poultry recipient households that is precisely estimated, as observed in column (1). The corresponding estimate for cash recipient households is still extremely large (80 percent), but not precisely estimated except for the T2 sample. There is also some evidence of an increase in the household livestock asset index and thus in the total asset index as observed in columns (3) and (6), particularly for poultry transfer recipients and for cash recipients in T2.

However, we observe no evidence of any significant effects on durable goods and productive assets or housing characteristics in Tables 7.13b or 7.14b.

7.7.3 Less poor sample

The results reported in Table 7.12c, Table 7.13c and Table 7.14c uniformly suggest there is no significant evidence of any increase in assets for the less poor sample. There is some weak evidence of a decline in livestock value for less poor households, but the estimates are not statistically significant when corrected for multiple hypothesis testing.

7.8 Consumption

7.8.1 Pooled sample

Table 7.15a reports effects on consumption for the pooled sample: total value of food, nonfood, and total consumption monthly per adult equivalent, and calories of food consumed daily per adult equivalent. It is

evident that there are no significant effects on consumption. Consumption data were not collected at midline.

7.8.2 Extremely poor sample

Table 7.15b reports parallel results for the extremely poor sample. Again, there is no evidence of any significant treatment effects, even for the sample of cash and poultry recipient households.

7.8.3 Less poor sample

Table 7.15c reports parallel results for the less poor sample, and again we observe no significant effects for consumption. Some estimated coefficients are negative and significant; however, this pattern is not robust to correction for multiple hypothesis testing.

7.9 Conclusion

This chapter presents evidence about the medium-term effects of an integrated nutrition-sensitive social protection program on a range of livelihoods-related outcomes, measured approximately five years following the initiation of programming, and two years following one-time poultry and cash transfers targeted to the poorest households in the sample.

For extremely poor households (who were transfer recipients), we see evidence of persistent and large increases in livestock assets and engagement in livestock production: these effects are concentrated in poultry for poultry recipients and concentrated in nonpoultry livestock for cash recipients. Extremely poor households also show evidence of substantial increases in membership in VESAs and the probability of reporting any savings. There is, however, no robust evidence of any increase in ownership of other durable goods (in a context in which ownership of these goods is rapidly increasing across the sample), or any increased consumption two years post-transfer.

For less poor households who did not receive transfers but were exposed to SPIR programming, we largely do not observe any substantial shifts in livestock assets or production. However, these households also show substantial increases in savings as well as some enhanced access to credit, and there is some weak evidence of improvement in housing characteristics.

Focusing on the results for extremely poor households vis-à-vis the existing literature, these results join a larger literature analyzing the medium-term effects of graduation model programs. Banerjee et al. (2015) analyzed the effects of BRAC's graduation model implemented across six countries (one of which was Ethiopia) approximately three years following an asset transfer. They found significant and positive effects on a range of indicators including consumption, assets, and income (all between .1 and .3 standard deviations). However, the BRAC program entailed much larger transfers (valued at between $400 in India and $1,200 in Ethiopia, based on PPP estimates) compared to the transfer analyzed here ($200).

Similarly, recent work by Bandiera et al. (2017) analyzes the Targeting the Ultra Poor program in Bangladesh in which households received a package of assets and skills training valued at around $1,100. The authors find evidence of substantial increases in income (21 percent higher), consumption expenditure (11 percent higher), and the value of household durables (57 percent higher) four years post-

transfer. In both these cases, the interventions entail a much larger transfer, and seem to generate a more persistent effect on a range of outcomes. One important caveat in analyzing our findings vis-à-vis these earlier papers, however, is that here the entire sample of households analyzed (including the control arm households) are receiving basic consumption support in the form of PSNP transfers. By contrast, previous papers compared households receiving a full graduation model package vis-à-vis households who received no consumption support of any kind.

Our results are broadly similar to some other findings of medium- or long-term effects of one-time transfers. In Kenya, an RCT of unconditional cash transfers offered by GiveDirectly (valued at around $700) found positive effects only on assets three years post-transfer, comparing households who received the transfer to comparable households in other villages in order to abstract from intravillage-spillovers, which appear to be substantial in this context (Haushofer and Shapiro 2018). Blattman et al. (2020) report the long-term effects of cash grants ($400) 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. In both these papers, the existing transfers are substantially larger (between two and four times greater) than the value of SPIR transfers, but they were provided in the absence of any broader graduation model program or associated services. Despite these differences, the general pattern of persistent effects primarily on asset stocks is broadly consistent with what we observe here.

There is also a growing literature on transfers of animal assets, but primarily focused on effects on nutritional or food security (e.g., 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 ($1,600) 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. Mullaly et al. (2021) analyze the effects of another substantial in-kind transfer of chickens and associated materials (valued at $500) in Guatemala and find no significant effects on livelihood-related outcomes, on average, about a year post-transfer. Again, both these transfers are meaningfully larger in value than the transfers implemented under SPIR, but they nonetheless show rather heterogeneous effects, even over a short time horizon.

Finally, this project is one of the first to conduct a systematic comparison of a cash and in-kind transfer in the framework of a randomized trial. Unsurprisingly, at endline we observe poultry households continuing to show a higher level of poultry production, while cash households show a higher level of engagement in nonpoultry livestock production. That being said, there is no robust evidence of any significant difference in income from livestock production, savings, assets, or consumption that would be suggestive of a meaningful welfare difference comparing across the two transfer modalities.

Moving on to the results for less poor households, here there is some weak evidence of shifts in some variables linked to livestock production, though no substantial effects on average. These households do appear to have shown significant increases in savings and credit access, suggestive of greater financial inclusion.

It is also important to note that this evaluation unfolded in the context of widespread adverse shocks affecting rural Ethiopian households, particularly in 2020 and 2021. These include the onset of the COVID-19 pandemic and associated closures and travel restrictions; the wide spread of desert locusts in

2020; and rising political unrest and violence, particularly in 2021 and particularly in Amhara province. Further work will seek to explore in more detail the effects of these shocks on livelihood outcomes, and whether these shocks interacted with the effects of the interventions analyzed here.

To sum up, SPIR had a range of positive effects, particularly on livestock-related production (particularly for cash and poultry households), and particularly for financial inclusion (for all households). That being said, the value of the cash and poultry households was meaningfully lower than a range of other graduation model or transfer programs, and this is plausibly consistent with the absence of medium-term effects on income or consumption. There are, however, also larger transfer programs that have shown similarly minimal effects in the medium or long term. Future programming and research may productively explore whether there is a higher level of asset transfer or income at which rural Ethiopian households can reach a sustainably higher trajectory for income and consumption.

Table 7.2a: Summary table, part 1

	(1) Poultry production	(2) Other livestock production	(3) Revenue from livestock production	(4) Revenue from crop production	(5) Labor and wage work
T1	0.173***	0.063**	0.060*	-0.044	0.038
	(0.044)	(0.032)	(0.036)	(0.056)	(0.035)
T2	0.175***	0.059**	0.079**	0.006	0.007
	(0.041)	(0.029)	(0.034)	(0.062)	(0.036)
T3	0.094**	0.015	0.023	-0.035	0.031
	(0.044)	(0.030)	(0.038)	(0.071)	(0.038)
Test: T1 = T2	0.959	0.889	0.622	0.391	0.399
Test: T2 = T3	0.052	0.146	0.153	0.571	0.530
Test: T1 = T3	0.082	0.147	0.372	0.898	0.865
N	3,812	3,812	3,804	3,804	3,809

Notes Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 7.2b: Summary table, part 1: Subsample of extremely poor households

	(1) Poultry production	(2) Other livestock production	(3) Revenue from livestock production	(4) Revenue from crop production	(5) Labor and wage work
T1 x Poultry	0.405***	0.116**	0.079*	0.091	0.001
	(0.065)	(0.055)	(0.047)	(0.080)	(0.053)
T1 x Cash	0.100*	0.114**	0.048	-0.022	0.038
	(0.060)	(0.048)	(0.048)	(0.082)	(0.057)
T2 x Poultry	0.381***	0.024	0.116*	0.033	0.003
	(0.066)	(0.042)	(0.066)	(0.086)	(0.056)
T2 x Cash	0.179***	0.154***	0.172***	0.135	-0.003
	(0.065)	(0.041)	(0.056)	(0.089)	(0.058)
T3	0.024	0.010	0.008	0.024	0.088*
	(0.051)	(0.040)	(0.053)	(0.101)	(0.051)
Average effect of T1	0.245***	0.113***	0.062	0.033	0.020

	(0.050)	(0.040)	(0.038)	(0.065)	(0.044)
Average effect of T2	0.287***	0.089**	0.146***	0.085	-0.000
	(0.054)	(0.036)	(0.049)	(0.072)	(0.047)
Average effect of poultry	0.392***	0.069*	0.098**	0.061	0.002
	(0.053)	(0.039)	(0.045)	(0.068)	(0.045)
Average effect of cash	0.139***	0.134***	0.110***	0.056	0.018
	(0.051)	(0.037)	(0.041)	(0.070)	(0.047)
Test: T1 = T2	0.415	0.537	0.084	0.444	0.664
Test: T2 = T3	0.000	0.032	0.022	0.558	0.090
Test: T1 = T3	0.000	0.013	0.308	0.929	0.171
Test: Poultry = Cash	0.000	0.095	0.803	0.945	0.728
N	1,771	1,771	1,765	1,765	1,770

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 7.2c: Summary table, part 1: Subsample of less poor households

	(1) Poultry production	(2) Other livestock production	(3) Revenue from livestock production	(4) Revenue from crop production	(5) Labor and wage work
T1	0.049	0.017	0.023	-0.141*	0.104**
	(0.059)	(0.042)	(0.057)	(0.078)	(0.052)
T2	0.054	0.031	0.048	0.023	0.015
	(0.060)	(0.041)	(0.048)	(0.100)	(0.046)
T3	0.144**	0.026	0.036	-0.092	-0.004
	(0.064)	(0.041)	(0.056)	(0.094)	(0.047)
Test: T1 = T2	0.932	0.738	0.639	0.105	0.084*
Test: T2 = T3	0.115	0.897	0.803	0.313	0.664
Test: T1 = T3	0.100*	0.828	0.841	0.557	0.038**
N	1,323	1,323	1,322	1,322	1,323

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 7.3a: Summary table, part 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

	Household savings	Female savings	Household credit	Female credit	Aggregate assets	Individual assets	Housing characteristics	Consumption
T1	0.350***	0.465***	0.155***	0.236***	0.016	-0.037	0.067	-0.043
	(0.035)	(0.041)	(0.045)	(0.068)	(0.067)	(0.034)	(0.070)	(0.057)
T2	0.340***	0.430***	0.161***	0.171***	0.027	0.037	0.111	-0.015
	(0.034)	(0.041)	(0.046)	(0.055)	(0.061)	(0.034)	(0.072)	(0.056)
T3	0.316***	0.383***	0.155***	0.171***	-0.017	0.023	0.070	-0.012
	(0.034)	(0.042)	(0.046)	(0.062)	(0.061)	(0.032)	(0.074)	(0.056)
Test: T1 = T2	0.741	0.284	0.894	0.338	0.860	0.044	0.466	0.617
Test: T2 = T3	0.423	0.182	0.910	1.000	0.440	0.672	0.519	0.952
Test: T1 = T3	0.277	0.015	0.988	0.364	0.617	0.092	0.967	0.566
N	3,804	3,823	3,804	3,704	3,812	3,804	3,775	3,810

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 7.3b: Summary table, part 2: Subsample of extremely poor households

	(1) Female savings	(2) Female credit	(3) Male savings	(4) Male credit	(5) Aggregate assets	(6) Individual assets	(7) Housing characteristics	(8) Consumption
T1 x Poultry	0.486***	0.124*	0.364***	0.091	0.169*	-0.024	0.062	0.011
	(0.053)	(0.067)	(0.041)	(0.060)	(0.095)	(0.051)	(0.082)	(0.090)
T1 x Cash	0.463***	0.237**	0.349***	0.114	-0.048	-0.058	0.093	-0.089
	(0.056)	(0.114)	(0.055)	(0.078)	(0.091)	(0.061)	(0.092)	(0.085)
T2 x Poultry	0.437***	0.111	0.352***	0.164**	0.065	0.063	0.106	0.053
	(0.047)	(0.068)	(0.043)	(0.076)	(0.088)	(0.051)	(0.093)	(0.081)
T2 x Cash	0.456***	0.084	0.385***	0.120*	0.196**	0.094**	0.092	0.019
	(0.047)	(0.075)	(0.047)	(0.067)	(0.081)	(0.042)	(0.085)	(0.073)
T3	0.363***	0.098	0.326***	0.038	-0.047	-0.018	0.080	-0.016
	(0.048)	(0.062)	(0.041)	(0.050)	(0.073)	(0.043)	(0.082)	(0.064)
Average effect of T1	0.466***	0.178***	0.350***	0.101*	0.057	-0.041	0.076	-0.040
	(0.045)	(0.069)	(0.040)	(0.054)	(0.074)	(0.045)	(0.074)	(0.070)
Average effect of T2	0.454***	0.100*	0.375***	0.145**	0.131*	0.080**	0.101	0.037
	(0.041)	(0.059)	(0.039)	(0.057)	(0.072)	(0.040)	(0.078)	(0.066)
Average effect of poultry	0.460***	0.117**	0.358***	0.129**	0.115	0.022	0.085	0.033
	(0.042)	(0.056)	(0.037)	(0.053)	(0.074)	(0.041)	(0.076)	(0.070)
Average effect of cash	0.460***	0.160**	0.367***	0.117**	0.074	0.018	0.092	-0.035
	(0.044)	(0.073)	(0.042)	(0.057)	(0.072)	(0.043)	(0.076)	(0.066)

Test: T1 = T2	0.755	0.287	0.469	0.485	0.297	0.005	0.701	0.244
Test: T2 = T3	0.038	0.975	0.165	0.064	0.009	0.013	0.766	0.367
Test: T1 = T3	0.022	0.285	0.529	0.258	0.162	0.625	0.961	0.711
Test: Poultry = Cash	0.983	0.575	0.798	0.847	0.561	0.926	0.907	0.305
N	1,781	1,723	1,765	1,765	1,771	1,765	1,748	1,770

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 7.3c: Summary table, part 2: Subsample of less poor households

	(1) Female savings	(2) Female credit	(3) Male savings	(4) Male credit	(5) Aggregate assets	(6) Individual assets	(7) Housing characteristics	(8) Consumption
T1	0.508***	0.205*	0.404***	0.232***	-0.045	-0.026	0.132*	-0.104
	(0.056)	(0.108)	(0.041)	(0.079)	(0.084)	(0.046)	(0.079)	(0.068)
T2	0.488***	0.303***	0.405***	0.265***	-0.045	0.012	0.158*	-0.052
	(0.060)	(0.103)	(0.043)	(0.096)	(0.076)	(0.049)	(0.083)	(0.081)
T3	0.472***	0.278**	0.372***	0.491***	-0.028	0.062	0.108	0.001
	(0.054)	(0.113)	(0.040)	(0.113)	(0.075)	(0.046)	(0.081)	(0.075)
Test: T1 = T2	0.665	0.326	0.963	0.720	0.996	0.419	0.687	0.484
Test: T2 = T3	0.718	0.814	0.356	0.072*	0.800	0.261	0.459	0.499
Test: T1 = T3	0.357	0.500	0.371	0.019**	0.835	0.044**	0.701	0.106
N	1,326	1,288	1,322	1,322	1,323	1,322	1,314	1,323

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 7.4a: Poultry production

	(1) Woman reports owning any poultry (solely or jointly owned)	(2) Man reports household owns any poultry	(3) Total number of poultry owned by female	(4) Total number of poultry owned by household (as reported by male)	(5) Woman reports any income from sales of poultry owned (solely or jointly)*	(6) Man reports any income from sales of poultry owned by household*	(7) Woman reports any income from egg sales in last 30 days

T1	0.085**	0.111***	0.616***	0.738***	0.072***	0.099***	0.048
	(0.033)	(0.029)	(0.214)	(0.198)	(0.026)	(0.027)	(0.039)
	[0.015]	[0.001]	[0.009]	[0.001]	[0.009]	[0.001]	[0.244]
T2	0.092***	0.112***	0.803***	0.845***	0.027	0.057*	0.038
	(0.031)	(0.027)	(0.229)	(0.214)	(0.029)	(0.029)	(0.039)
	[0.006]	[0.000]	[0.002]	[0.000]	[0.347]	[0.072]	[0.347]
T3	0.049	0.045	0.378*	0.447**	0.035	0.052*	0.030
	(0.034)	(0.029)	(0.224)	(0.200)	(0.030)	(0.030)	(0.038)
	[0.232]	[0.232]	[0.232]	[0.211]	[0.336]	[0.232]	[0.490]
Test: T1 = T2	(0.825)	(0.957)	(0.440)	(0.648)	(0.102)	(0.146)	(0.792)
	[0.943]	[0.957]	[0.881]	[0.943]	[0.390]	[0.390]	[0.943]
Test: T2 = T3	(0.168)	(0.006)	(0.086)	(0.084)	(0.798)	(0.889)	(0.818)
	[0.269]	[0.051]	[0.172]	[0.172]	[0.889]	[0.889]	[0.889]
Test: T1 = T3	(0.287)	(0.013)	(0.307)	(0.175)	(0.198)	(0.120)	(0.631)
	[0.409]	[0.108]	[0.409]	[0.397]	[0.397]	[0.397]	[0.721]
Mean of control (T4)	0.566	0.660	2.064	1.977	0.304	0.282	0.329
N	3,704	3,803	3,704	3,803	2,488	2,790	2,050

* In last 12 months.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.4b: Poultry production: Subsample of extremely poor households

	(1) Woman reports owning any poultry (solely or jointly owned)	(2) Man reports household owns any poultry	(3) Total number of poultry owned by female	(4) Total number of poultry owned by household (as reported by male)	(5) Woman reports any income from sales of poultry owned (solely or jointly)*	(6) Man reports any income from sales of poultry owned by household*	(7) Woman reports any income from egg sales in last 30 days
T1 x Poultry	0.202***	0.244***	1.636***	1.742***	0.151***	0.148***	0.067
	(0.045)	(0.037)	(0.394)	(0.366)	(0.038)	(0.046)	(0.058)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.249]
T1 x Cash	0.046	0.082**	0.183	0.300	0.052	0.093**	0.006
	(0.043)	(0.041)	(0.259)	(0.235)	(0.047)	(0.046)	(0.060)

	[0.463]	[0.197]	[0.642]	[0.463]	[0.463]	[0.197]	[0.920]
T2 x Poultry	0.180***	0.199***	1.995***	1.779***	0.128**	0.144**	0.033
	(0.046)	(0.036)	(0.466)	(0.405)	(0.052)	(0.055)	(0.056)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.021]	[0.016]	[0.563]
T2 x Cash	0.111**	0.142***	0.590*	0.758**	-0.020	-0.012	0.045
	(0.050)	(0.044)	(0.336)	(0.321)	(0.044)	(0.048)	(0.067)
	[0.057]	[0.005]	[0.130]	[0.051]	[0.751]	[0.809]	[0.671]
T3	-0.009	0.020	0.126	0.158	0.034	0.017	-0.041
	(0.039)	(0.035)	(0.249)	(0.220)	(0.040)	(0.042)	(0.048)
	[0.936]	[0.901]	[0.901]	[0.901]	[0.901]	[0.901]	[0.901]
Average effect of T1	0.120***	0.158***	0.878***	0.987***	0.099***	0.118***	0.035
	(0.036)	(0.033)	(0.257)	(0.235)	(0.034)	(0.038)	(0.049)
	[0.002]	[0.000]	[0.002]	[0.000]	[0.006]	[0.004]	[0.472]
Average effect of T2	0.149***	0.174***	1.331***	1.303***	0.057	0.069	0.039
	(0.040)	(0.035)	(0.320)	(0.286)	(0.039)	(0.043)	(0.052)
	[0.001]	[0.000]	[0.000]	[0.000]	[0.173]	[0.148]	[0.452]
Average effect of poultry	0.190***	0.220***	1.823***	1.761***	0.139***	0.146***	0.049
	(0.038)	(0.032)	(0.328)	(0.291)	(0.037)	(0.042)	(0.048)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.312]
Average effect of cash	0.078**	0.112***	0.386	0.529**	0.016	0.041	0.025
	(0.039)	(0.036)	(0.246)	(0.225)	(0.037)	(0.040)	(0.052)
	[0.087]	[0.015]	[0.189]	[0.080]	[0.653]	[0.409]	[0.653]
Test: T1 = T2	(0.429)	(0.582)	(0.169)	(0.298)	(0.282)	(0.208)	(0.927)
	[0.573]	[0.665]	[0.477]	[0.477]	[0.477]	[0.477]	[0.927]
Test: T2 = T3	(0.000)	(0.000)	(0.000)	(0.000)	(0.605)	(0.214)	(0.071)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.605]	[0.244]	[0.095]
Test: T1 = T3	(0.000)	(0.000)	(0.004)	(0.001)	(0.104)	(0.009)	(0.079)
	[0.002]	[0.000]	[0.009]	[0.002]	[0.119]	[0.014]	[0.105]
Test: Poultry = Cash	(0.003)	(0.000)	(0.000)	(0.000)	(0.002)	(0.008)	(0.611)
	[0.005]	[0.001]	[0.000]	[0.000]	[0.003]	[0.010]	[0.698]
Mean of control (T4)	0.533	0.606	1.821	1.763	0.298	0.297	0.356
N	1,723	1,765	1,723	1,765	1,127	1,245	928

* In last 12 months.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.4c: Poultry production: Subsample of less poor households

	(1) Woman reports owning any poultry (solely or jointly owned)	(2) Man reports household owns any poultry	(3) Total number of poultry owned by female	(4) Total number of poultry owned by household (as reported by male)	(5) Woman reports any income from sales of poultry owned (solely or jointly)*	(6) Man reports any income from sales of poultry owned by household*	(7) Woman reports any income from egg sales in last 30 days
T1	0.032	0.063*	0.227	0.433*	-0.010	0.063	0.014
	(0.043)	(0.035)	(0.270)	(0.261)	(0.049)	(0.045)	(0.059)
	[0.618]	[0.275]	[0.618]	[0.275]	[0.835]	[0.328]	[0.835]
T2	0.025	0.071**	0.447	0.607**	-0.025	0.037	-0.022
	(0.041)	(0.035)	(0.317)	(0.304)	(0.050)	(0.047)	(0.056)
	[0.692]	[0.191]	[0.381]	[0.191]	[0.692]	[0.692]	[0.692]
T3	0.090*	0.083**	0.507	0.595*	0.032	0.113**	0.075
	(0.046)	(0.037)	(0.344)	(0.311)	(0.051)	(0.045)	(0.061)
	[0.114]	[0.102]	[0.228]	[0.114]	[0.526]	[0.102]	[0.294]
Test: T1 = T2	(0.863)	(0.822)	(0.444)	(0.548)	(0.746)	(0.555)	(0.500)
	[0.863]	[0.863]	[0.863]	[0.863]	[0.863]	[0.863]	[0.863]
Test: T2 = T3	(0.119)	(0.718)	(0.864)	(0.970)	(0.212)	(0.083)	(0.058)
	[0.316]	[0.957]	[0.970]	[0.970]	[0.425]	[0.316]	[0.316]
Test: T1 = T3	(0.184)	(0.572)	(0.369)	(0.582)	(0.353)	(0.245)	(0.299)
	[0.492]	[0.582]	[0.492]	[0.582]	[0.492]	[0.492]	[0.492]
Mean of control (T4)	0.605	0.696	2.326	2.179	0.340	0.269	0.345
N	1,288	1,321	1,288	1,321	886	998	728

* In last 12 months.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.5a: Other livestock production

	(1) Woman reports	(2) Man reports household	(3) Man reports household	(4) Man reports household	(5) Woman reports any	(6) Man reports household	(7) Man reports household	(8) Man reports any income

	owning any sheep or goats (solely or jointly owned)	owns any sheep or goats	owns any oxen	owns livestock other than poultry, sheep, goats and oxen	income from sales of sheep/goats owned (solely or jointly)	receives any income from sales of sheep/goats owned	receives any income from sales of oxen owned	from sales of livestock other than poultry, sheep, goats
T1	0.047	0.059	0.013	-0.001	0.032	0.033	0.086**	-0.025
	(0.038)	(0.036)	(0.035)	(0.035)	(0.027)	(0.031)	(0.042)	(0.026)
	[0.457]	[0.426]	[0.812]	[0.977]	[0.457]	[0.457]	[0.328]	[0.462]
T2	0.033	0.060**	0.029	0.016	0.008	-0.007	0.069*	0.013
	(0.031)	(0.030)	(0.029)	(0.029)	(0.029)	(0.034)	(0.041)	(0.024)
	[0.634]	[0.358]	[0.634]	[0.784]	[0.830]	[0.830]	[0.366]	[0.784]
T3	0.031	0.051	-0.019	-0.004	-0.013	-0.023	0.054	-0.017
	(0.040)	(0.037)	(0.032)	(0.030)	(0.026)	(0.029)	(0.051)	(0.027)
	[0.694]	[0.694]	[0.694]	[0.883]	[0.694]	[0.694]	[0.694]	[0.694]
Test: T1 = T2	(0.690)	(0.977)	(0.631)	(0.622)	(0.381)	(0.214)	(0.651)	(0.075)
	[0.789]	[0.977]	[0.789]	[0.789]	[0.789]	[0.789]	[0.789]	[0.601]
Test: T2 = T3	(0.966)	(0.776)	(0.120)	(0.480)	(0.412)	(0.596)	(0.758)	(0.177)
	[0.966]	[0.887]	[0.707]	[0.887]	[0.887]	[0.887]	[0.887]	[0.707]
Test: T1 = T3	(0.711)	(0.832)	(0.387)	(0.921)	(0.057)	(0.047)	(0.495)	(0.743)
	[0.921]	[0.921]	[0.921]	[0.921]	[0.229]	[0.229]	[0.921]	[0.921]
Mean of control (T4)	0.502	0.602	0.322	0.710	0.324	0.552	0.289	0.228
N	3,700	3,802	3,798	3,804	2,112	2,454	1,283	2,714

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.5b: Other livestock production: Subsample of extremely poor households

	(1) Woman reports owning any sheep or goats (solely or jointly owned)	(2) Man reports household owns any sheep or goats	(3) Man reports household owns any oxen	(4) Man reports household owns livestock other than poultry, sheep, goats and oxen	(5) Woman reports any income from sales of sheep/goats owned (solely or jointly)	(6) Man reports household receives any income from sales of sheep/goats owned	(7) Man reports household receives any income from sales of oxen owned	(8) Man reports any income from sales of livestock other than poultry, sheep, goats

T1 x Poultry	0.106*	0.096*	0.074	0.109**	-0.029	0.005	0.107	-0.022
	(0.060)	(0.053)	(0.046)	(0.047)	(0.044)	(0.048)	(0.095)	(0.040)
	[0.206]	[0.206]	[0.227]	[0.170]	[0.663]	[0.925]	[0.415]	[0.663]
T1 x Cash	0.017	0.053	0.033	-0.006	0.161***	0.154***	0.065	-0.033
	(0.056)	(0.057)	(0.051)	(0.058)	(0.052)	(0.043)	(0.077)	(0.042)
	[0.870]	[0.689]	[0.689]	[0.912]	[0.009]	[0.004]	[0.689]	[0.689]
T2 x Poultry	0.060	0.043	0.040	0.008	-0.015	-0.015	-0.092	0.054
	(0.047)	(0.044)	(0.044)	(0.045)	(0.054)	(0.055)	(0.072)	(0.041)
	[0.554]	[0.581]	[0.581]	[0.861]	[0.861]	[0.861]	[0.554]	[0.554]
T2 x Cash	0.072	0.091**	0.129***	0.167***	0.022	0.033	0.027	0.034
	(0.045)	(0.046)	(0.045)	(0.037)	(0.048)	(0.055)	(0.079)	(0.040)
	[0.226]	[0.128]	[0.017]	[0.000]	[0.731]	[0.731]	[0.731]	[0.636]
T3	0.011	0.016	-0.002	0.003	-0.031	-0.003	0.039	0.003
	(0.046)	(0.046)	(0.036)	(0.042)	(0.045)	(0.044)	(0.079)	(0.035)
	[0.958]	[0.958]	[0.958]	[0.958]	[0.958]	[0.958]	[0.958]	[0.958]
Average effect of T1	0.059	0.073	0.052	0.049	0.067	0.079**	0.084	-0.027
	(0.045)	(0.044)	(0.038)	(0.043)	(0.041)	(0.038)	(0.070)	(0.034)
	[0.284]	[0.277]	[0.284]	[0.284]	[0.277]	[0.277]	[0.284]	[0.420]
Average effect of T2	0.067*	0.068*	0.085**	0.087**	0.004	0.009	-0.034	0.045
	(0.039)	(0.039)	(0.037)	(0.037)	(0.044)	(0.045)	(0.066)	(0.034)
	[0.178]	[0.178]	[0.087]	[0.087]	[0.935]	[0.935]	[0.806]	[0.314]
Average effect of poultry	0.082*	0.069*	0.056	0.057	-0.022	-0.005	0.004	0.017
	(0.043)	(0.041)	(0.036)	(0.039)	(0.042)	(0.043)	(0.070)	(0.034)
	[0.294]	[0.294]	[0.294]	[0.294]	[0.813]	[0.956]	[0.956]	[0.813]
Average effect of cash	0.044	0.072*	0.081**	0.080*	0.092**	0.093**	0.046	0.000
	(0.042)	(0.043)	(0.038)	(0.041)	(0.042)	(0.041)	(0.066)	(0.034)
	[0.388]	[0.160]	[0.097]	[0.101]	[0.097]	[0.097]	[0.559]	[0.993]
Test: T1 = T2	(0.862)	(0.892)	(0.398)	(0.312)	(0.087)	(0.079)	(0.058)	(0.024)
	[0.892]	[0.892]	[0.531]	[0.500]	[0.175]	[0.175]	[0.175]	[0.175]
Test: T2 = T3	(0.199)	(0.189)	(0.018)	(0.018)	(0.397)	(0.799)	(0.295)	(0.202)
	[0.324]	[0.324]	[0.074]	[0.074]	[0.453]	[0.799]	[0.394]	[0.324]
Test: T1 = T3	(0.327)	(0.207)	(0.168)	(0.280)	(0.014)	(0.043)	(0.546)	(0.349)
	[0.399]	[0.399]	[0.399]	[0.399]	[0.110]	[0.173]	[0.546]	[0.399]
Test: Poultry = Cash	(0.395)	(0.938)	(0.533)	(0.529)	(0.002)	(0.016)	(0.494)	(0.589)
	[0.673]	[0.938]	[0.673]	[0.673]	[0.018]	[0.065]	[0.673]	[0.673]
Mean of control (T4)	0.464	0.567	0.224	0.611	0.341	0.528	0.309	0.200
N	1,721	1,763	1,763	1,765	932	1,070	464	1,137

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in

parentheses in panel 3. 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. 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 7.5c: Other livestock production: Subsample of less poor households

	(1) Woman reports owning any sheep or goats (solely or jointly owned)	(2) Man reports household owns any sheep or goats	(3) Man reports household owns any oxen	(4) Man reports household owns livestock other than poultry, sheep, goats and oxen	(5) Woman reports any income from sales of sheep/goats owned (solely or jointly)	(6) Man reports household receives any income from sales of sheep/goats owned	(7) Man reports household receives any income from sales of oxen owned	(8) Man reports any income from sales of livestock other than poultry, sheep, goats
T1	0.031	0.049	-0.024	-0.078*	0.022	-0.000	0.102*	-0.019
	(0.042)	(0.040)	(0.044)	(0.043)	(0.042)	(0.051)	(0.058)	(0.040)
	[0.735]	[0.609]	[0.735]	[0.321]	[0.735]	[0.993]	[0.321]	[0.735]
T2	0.011	0.060	-0.017	-0.067*	0.026	-0.025	0.158***	-0.019
	(0.042)	(0.038)	(0.040)	(0.039)	(0.045)	(0.055)	(0.056)	(0.040)
	[0.788]	[0.319]	[0.768]	[0.319]	[0.768]	[0.768]	[0.040]	[0.768]
T3	0.040	0.094**	-0.054	-0.021	0.023	-0.008	0.085	-0.057
	(0.048)	(0.043)	(0.045)	(0.036)	(0.037)	(0.050)	(0.058)	(0.040)
	[0.644]	[0.235]	[0.478]	[0.644]	[0.644]	[0.880]	[0.406]	[0.406]
Test: T1 = T2	(0.645)	(0.760)	(0.875)	(0.799)	(0.945)	(0.609)	(0.312)	(0.985)
	[0.985]	[0.985]	[0.985]	[0.985]	[0.985]	[0.985]	[0.985]	[0.985]
Test: T2 = T3	(0.538)	(0.385)	(0.409)	(0.242)	(0.957)	(0.705)	(0.206)	(0.264)
	[0.717]	[0.654]	[0.654]	[0.654]	[0.957]	[0.805]	[0.654]	[0.654]
Test: T1 = T3	(0.839)	(0.275)	(0.554)	(0.185)	(0.980)	(0.869)	(0.771)	(0.266)
	[0.980]	[0.735]	[0.980]	[0.735]	[0.980]	[0.980]	[0.980]	[0.735]
Mean of control (T4)	0.551	0.638	0.431	0.833	0.309	0.558	0.269	0.262
N	1,287	1,322	1,319	1,322	784	911	563	1,050

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.6a: Income from livestock production and crop cultivation

	(1) Net income from sales of livestock products (eggs, milk, dairy, honey) in past 12 months	(2) Revenue from all livestock sales in past 12 months	(3) Costs associated with livestock production (feed, medicine, outside labor)	(4) Net income from livestock sales (sales of animals minus production costs)	(5) Household earned any revenue from crops cultivated in last Mehr season	(6) Revenue earned from crops cultivated in last Mehr season
T1	0.274**	0.690**	0.051	0.733*	-0.029	-0.255
	(0.117)	(0.278)	(0.219)	(0.389)	(0.030)	(0.265)
	[0.082]	[0.082]	[0.974]	[0.163]	[0.540]	[0.540]
T2	0.330***	0.683**	0.436**	0.661*	-0.026	-0.220
	(0.120)	(0.270)	(0.221)	(0.385)	(0.031)	(0.266)
	[0.049]	[0.049]	[0.132]	[0.159]	[0.410]	[0.410]
T3	0.273**	0.051	0.124	-0.000	-0.057*	-0.498*
	(0.126)	(0.280)	(0.198)	(0.377)	(0.032)	(0.280)
	[0.205]	[0.977]	[0.707]	[0.999]	[0.205]	[0.205]
Test: T1 = T2	(0.651)	(0.980)	(0.084)	(0.857)	(0.899)	(0.879)
	[0.980]	[0.980]	[0.669]	[0.980]	[0.980]	[0.980]
Test: T2 = T3	(0.666)	(0.022)	(0.111)	(0.084)	(0.272)	(0.266)
	[0.666]	[0.087]	[0.177]	[0.167]	[0.311]	[0.311]
Test: T1 = T3	(0.995)	(0.024)	(0.712)	(0.060)	(0.306)	(0.313)
	[0.995]	[0.195]	[0.814]	[0.239]	[0.502]	[0.502]
Mean of control (T4)	0.833	4.887	3.168	2.877	0.325	2.766
N	3,794	3,804	3,802	3,802	3,804	3,804

Note: Estimates from the DFSA SPIR endline survey sample. Inverse hyperbolic sine transformation has been applied to all outcome variables presented in the table. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.6b: Income from livestock production and crop cultivation: Subsample of extremely poor households

	(1) Net income from sales of livestock products	(2) Revenue from all livestock sales in past	(3) Costs associated with livestock	(4) Net income from livestock sales (sales	(5) Household earned any revenue from crops	(6) Revenue earned from crops cultivated in

	(eggs, milk, dairy, honey) in past 12 months	12 months	production (feed, medicine, outside labor)	of animals minus production costs)	cultivated in last Mehr season	last Mehr season
T1 x Poultry	0.622***	1.280***	0.391	1.227**	0.059	0.456
	(0.196)	(0.405)	(0.310)	(0.578)	(0.037)	(0.325)
	[0.007]	[0.007]	[0.209]	[0.070]	[0.149]	[0.185]
T1 x Cash	0.062	1.135***	0.107	1.297**	-0.033	-0.274
	(0.124)	(0.427)	(0.337)	(0.578)	(0.037)	(0.327)
	[0.969]	[0.068]	[0.969]	[0.104]	[0.806]	[0.806]
T2 x Poultry	0.664***	1.082**	1.026***	0.150	0.032	0.234
	(0.233)	(0.455)	(0.364)	(0.567)	(0.045)	(0.376)
	[0.021]	[0.037]	[0.021]	[0.792]	[0.611]	[0.611]
T2 x Cash	0.382**	1.091**	0.972***	0.930	0.087**	0.681**
	(0.184)	(0.433)	(0.357)	(0.680)	(0.039)	(0.343)
	[0.053]	[0.025]	[0.019]	[0.173]	[0.043]	[0.056]
T3	0.111	0.037	0.051	-0.023	-0.043	-0.385
	(0.126)	(0.366)	(0.257)	(0.468)	(0.036)	(0.323)
	[0.610]	[0.961]	[0.961]	[0.961]	[0.610]	[0.610]
Average effect of T1	0.330**	1.184***	0.242	1.239***	0.012	0.082
	(0.129)	(0.334)	(0.259)	(0.457)	(0.031)	(0.273)
	[0.030]	[0.004]	[0.471]	[0.029]	[0.765]	[0.765]
Average effect of T2	0.535***	1.107***	1.018***	0.542	0.060*	0.461
	(0.166)	(0.359)	(0.294)	(0.501)	(0.035)	(0.302)
	[0.004]	[0.004]	[0.003]	[0.280]	[0.114]	[0.146]
Average effect of poultry	0.644***	1.177***	0.721***	0.667	0.045	0.341
	(0.167)	(0.347)	(0.273)	(0.461)	(0.034)	(0.292)
	[0.001]	[0.003]	[0.015]	[0.199]	[0.208]	[0.245]
Average effect of cash	0.222*	1.113***	0.539*	1.114**	0.027	0.202
	(0.129)	(0.349)	(0.280)	(0.500)	(0.032)	(0.283)
	[0.139]	[0.013]	[0.139]	[0.108]	[0.466]	[0.476]
Test: T1 = T2	(0.204)	(0.831)	(0.008)	(0.173)	(0.120)	(0.152)
	[0.233]	[0.831]	[0.064]	[0.230]	[0.230]	[0.230]
Test: T2 = T3	(0.008)	(0.007)	(0.001)	(0.278)	(0.005)	(0.009)
	[0.010]	[0.010]	[0.002]	[0.278]	[0.010]	[0.010]
Test: T1 = T3	(0.077)	(0.002)	(0.460)	(0.009)	(0.089)	(0.107)
	[0.119]	[0.017]	[0.460]	[0.031]	[0.119]	[0.122]

Test: Poultry = Cash	(0.011)	(0.861)	(0.525)	(0.388)	(0.547)	(0.602)
	[0.085]	[0.861]	[0.688]	[0.688]	[0.688]	[0.688]
Mean of control (T4)	0.709	4.338	2.590	2.691	0.279	2.398
N	1,760	1,765	1,763	1,763	1,765	1,765

Note: Estimates from the DFSA SPIR endline survey sample. Inverse hyperbolic sine transformation has been applied to all outcome variables presented in the table. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.6c: Income from livestock production and crop cultivation: Subsample of less poor households

	(1) Net income from sales of livestock products (eggs, milk, dairy, honey) in past 12 months	(2) Revenue from all livestock sales in past 12 months	(3) Costs associated with livestock production (feed, medicine, outside labor)	(4) Net income from livestock sales (sales of animals minus production costs)	(5) Household earned any revenue from crops cultivated in last Mehr season	(6) Revenue earned from crops cultivated in last Mehr season
T1	-0.079	0.106	-0.480	0.404	-0.093**	-0.745**
	(0.182)	(0.397)	(0.332)	(0.575)	(0.042)	(0.365)
	[0.739]	[0.791]	[0.214]	[0.604]	[0.155]	[0.155]
T2	-0.029	0.457	-0.013	0.842	-0.084*	-0.601
	(0.179)	(0.427)	(0.330)	(0.590)	(0.044)	(0.374)
	[0.968]	[0.358]	[0.969]	[0.222]	[0.189]	[0.222]
T3	0.347	0.173	0.157	0.226	-0.079*	-0.672*
	(0.220)	(0.452)	(0.308)	(0.637)	(0.042)	(0.363)
	[0.283]	[0.723]	[0.723]	[0.723]	[0.283]	[0.283]
Test: T1 = T2	(0.775)	(0.315)	(0.168)	(0.410)	(0.823)	(0.643)
	[0.887]	[0.887]	[0.887]	[0.887]	[0.887]	[0.887]
Test: T2 = T3	(0.074)	(0.481)	(0.581)	(0.292)	(0.886)	(0.820)
	[0.740]	[0.886]	[0.886]	[0.886]	[0.886]	[0.886]
Test: T1 = T3	(0.051)	(0.856)	(0.043)	(0.756)	(0.690)	(0.806)
	[0.253]	[0.856]	[0.253]	[0.856]	[0.856]	[0.856]
Mean of control (T4)	1.107	5.519	3.836	3.123	0.369	3.094

N	1,319	1,322	1,322	1,322	1,322	1,322

Note: Estimates from the DFSA SPIR endline survey sample. Inverse hyperbolic sine transformation has been applied to all outcome variables presented in the table. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.7a: Business and wage work, last 12 months

	(1) Household carried out non-agricultural business	(2) A member undertook regular wage work for an employer	(3) Female undertook regular wage work for an employer	(4) A member undertook irregular/casual wage work	(5) Female undertook irregular/casual wage work	(6) Male undertook regular or irregular wage work
T1	0.002	0.018	-0.002	0.024	0.002	0.034
	(0.015)	(0.013)	(0.003)	(0.031)	(0.017)	(0.028)
	[0.917]	[0.684]	[0.745]	[0.745]	[0.917]	[0.684]
T2	-0.003	0.004	-0.001	0.013	0.006	0.009
	(0.013)	(0.011)	(0.003)	(0.030)	(0.019)	(0.026)
	[0.846]	[0.846]	[0.846]	[0.846]	[0.846]	[0.846]
T3	0.007	0.016	0.000	0.007	0.005	0.007
	(0.014)	(0.012)	(0.004)	(0.029)	(0.019)	(0.027)
	[0.913]	[0.913]	[0.913]	[0.913]	[0.913]	[0.913]
Test: T1 = T2	(0.723)	(0.253)	(0.600)	(0.747)	(0.823)	(0.408)
	[0.823]	[0.823]	[0.823]	[0.823]	[0.823]	[0.823]
Test: T2 = T3	(0.459)	(0.247)	(0.750)	(0.862)	(0.973)	(0.942)
	[0.973]	[0.973]	[0.973]	[0.973]	[0.973]	[0.973]
Test: T1 = T3	(0.760)	(0.891)	(0.420)	(0.630)	(0.854)	(0.387)
	[0.891]	[0.891]	[0.891]	[0.891]	[0.891]	[0.891]
Mean of control (T4)	0.046	0.032	0.007	0.238	0.103	0.162
N	2,463	2,462	3,704	2,462	3,703	2,461

Note: Estimates from the DFSA SPIR endline survey sample. Inverse hyperbolic sine transformation has been applied to all outcome variables presented in the table. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.7b: Business and wage work, last 12 months: Subsample of extremely poor households

	(1) Household carried out non-agricultural business	(2) A member undertook regular wage work for an employer	(3) Female undertook regular wage work for an employer	(4) A member undertook irregular/casual wage work	(5) Female undertook irregular/casual wage work	(6) Male undertook regular or irregular wage work
T1 x Poultry	-0.008	-0.011	-0.004	0.059	0.004	0.001
	(0.019)	(0.019)	(0.006)	(0.056)	(0.031)	(0.048)
	[0.986]	[0.986]	[0.986]	[0.986]	[0.986]	[0.986]
T1 x Cash	0.018	0.014	0.003	0.011	-0.017	0.018
	(0.038)	(0.029)	(0.008)	(0.051)	(0.026)	(0.041)
	[0.821]	[0.821]	[0.821]	[0.833]	[0.821]	[0.821]
T2 x Poultry	0.023	-0.010	-0.004	0.007	-0.017	0.002
	(0.024)	(0.022)	(0.006)	(0.057)	(0.030)	(0.042)
	[0.969]	[0.969]	[0.969]	[0.969]	[0.969]	[0.969]
T2 x Cash	0.010	-0.028	-0.010**	0.033	0.016	0.016
	(0.022)	(0.019)	(0.005)	(0.052)	(0.034)	(0.045)
	[0.720]	[0.463]	[0.181]	[0.720]	[0.720]	[0.720]
T3	0.047**	0.013	-0.002	0.047	0.005	0.035
	(0.022)	(0.018)	(0.006)	(0.045)	(0.026)	(0.042)
	[0.225]	[0.737]	[0.836]	[0.737]	[0.836]	[0.737]
Average effect of T1	0.005	0.002	-0.000	0.034	-0.006	0.009
	(0.022)	(0.020)	(0.006)	(0.043)	(0.024)	(0.036)
	[0.941]	[0.941]	[0.941]	[0.941]	[0.941]	[0.941]
Average effect of T2	0.017	-0.019	-0.007	0.020	-0.001	0.009
	(0.018)	(0.018)	(0.005)	(0.046)	(0.026)	(0.037)
	[0.719]	[0.719]	[0.719]	[0.968]	[0.968]	[0.968]
Average effect of poultry	0.008	-0.011	-0.004	0.032	-0.007	0.001
	(0.017)	(0.018)	(0.005)	(0.046)	(0.025)	(0.036)
	[0.942]	[0.942]	[0.942]	[0.942]	[0.943]	[0.973]
Average effect of cash	0.014	-0.007	-0.003	0.022	-0.001	0.017
	(0.023)	(0.020)	(0.006)	(0.043)	(0.025)	(0.036)
	[0.871]	[0.871]	[0.871]	[0.871]	[0.982]	[0.871]
Test: T1 = T2	(0.650)	(0.257)	(0.184)	(0.758)	(0.824)	(0.992)
	[0.989]	[0.770]	[0.770]	[0.989]	[0.989]	[0.992]
Test: T2 = T3	(0.226)	(0.046)	(0.327)	(0.553)	(0.801)	(0.519)
	[0.654]	[0.278]	[0.654]	[0.664]	[0.801]	[0.664]
Test: T1 = T3	(0.144)	(0.565)	(0.752)	(0.764)	(0.624)	(0.534)

	[0.764]	[0.764]	[0.764]	[0.764]	[0.764]	[0.764]
Test: Poultry = Cash	(0.818)	(0.841)	(0.892)	(0.812)	(0.790)	(0.657)
	[0.892]	[0.892]	[0.892]	[0.892]	[0.892]	[0.892]
Mean of control (T4)	0.030	0.043	0.010	0.252	0.124	0.176
N	1,030	1,029	1,723	1,030	1,723	1,029

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.7c: Business and wage work, last 12 months: Subsample of less poor households

	(1) Household carried out non-agricultural business	(2) A member undertook regular wage work for an employer	(3) Female undertook regular wage work for an employer	(4) A member undertook irregular/casual wage work	(5) Female undertook irregular/casual wage work	(6) Male undertook regular or irregular wage work
T1	0.013	0.020	-0.003	0.050	0.027	0.096**
	(0.024)	(0.020)	(0.003)	(0.044)	(0.024)	(0.042)
	[0.588]	[0.400]	[0.400]	[0.400]	[0.400]	[0.142]
T2	-0.036*	-0.001	0.003	0.019	0.024	0.020
	(0.019)	(0.018)	(0.005)	(0.038)	(0.025)	(0.033)
	[0.387]	[0.954]	[0.736]	[0.736]	[0.736]	[0.736]
T3	-0.037*	0.003	-0.003	0.027	0.006	0.028
	(0.020)	(0.018)	(0.003)	(0.041)	(0.024)	(0.035)
	[0.385]	[0.871]	[0.769]	[0.769]	[0.871]	[0.769]
Test: T1 = T2	(0.031)	(0.248)	(0.136)	(0.490)	(0.916)	(0.065)
	[0.185]	[0.371]	[0.273]	[0.588]	[0.916]	[0.195]
Test: T2 = T3	(0.974)	(0.795)	(0.124)	(0.848)	(0.499)	(0.792)
	[0.974]	[0.974]	[0.742]	[0.974]	[0.974]	[0.974]
Test: T1 = T3	(0.030)	(0.363)	(0.809)	(0.614)	(0.417)	(0.113)
	[0.178]	[0.625]	[0.809]	[0.737]	[0.625]	[0.339]
Mean of control (T4)	0.057	0.033	0.003	0.199	0.073	0.133
N	951	951	1,288	951	1,288	951

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.8a: Household's savings

	(1) Household has any savings (with RUSACCO, VESA, VSLA, MFI, bank, at home, with a relative)	(2) Total savings (RUSACCO, VESA, VSLA, MFI, bank, at home, with a relative)	(3) Household has any savings at home or with a relative	(4) Household has any savings in the bank	(5) Household has any savings with MFI	(6) Household has any savings with a VESA/ VSLA	(7) Household has any savings with a RUSACCO	(8) Household is a member of VESA/ VSLA	(9) Household is a member of RUSACCO	(10) A member of the household has a bank account
T1	0.302***	525.716***	-0.103***	-0.078*	-0.080*	0.497***	-0.012	0.593***	0.070**	0.012
	(0.039)	(144.908)	(0.029)	(0.041)	(0.046)	(0.056)	(0.053)	(0.042)	(0.034)	(0.024)
	[0.000]	[0.001]	[0.001]	[0.085]	[0.106]	[0.000]	[0.819]	[0.000]	[0.065]	[0.687]
T2	0.322***	410.731***	-0.103***	-0.090**	-0.119**	0.596***	-0.087*	0.647***	0.013	-0.001
	(0.041)	(143.664)	(0.029)	(0.038)	(0.049)	(0.053)	(0.049)	(0.036)	(0.030)	(0.022)
	[0.000]	[0.009]	[0.001]	[0.029]	[0.027]	[0.000]	[0.099]	[0.000]	[0.747]	[0.976]
T3	0.302***	319.659**	-0.072**	-0.095**	-0.140***	0.585***	-0.092*	0.626***	-0.001	-0.005
	(0.038)	(139.872)	(0.029)	(0.039)	(0.046)	(0.052)	(0.050)	(0.036)	(0.030)	(0.024)
	[0.000]	[0.033]	[0.026]	[0.026]	[0.007]	[0.000]	[0.088]	[0.000]	[0.977]	[0.934]
Test: T1 = T2	(0.519)	(0.455)	(0.976)	(0.683)	(0.233)	(0.005)	(0.023)	(0.116)	(0.043)	(0.574)
	[0.718]	[0.718]	[0.976]	[0.759]	[0.465]	[0.050]	[0.116]	[0.291]	[0.143]	[0.718]
Test: T2 = T3	(0.495)	(0.537)	(0.065)	(0.820)	(0.493)	(0.705)	(0.863)	(0.423)	(0.576)	(0.850)
	[0.863]	[0.863]	[0.647]	[0.863]	[0.863]	[0.863]	[0.863]	[0.863]	[0.863]	[0.863]
Test: T1 = T3	(0.999)	(0.172)	(0.064)	(0.552)	(0.035)	(0.011)	(0.023)	(0.340)	(0.015)	(0.489)
	[0.999]	[0.286]	[0.127]	[0.614]	[0.088]	[0.077]	[0.078]	[0.485]	[0.077]	[0.611]
Mean of control (T4)	0.447	771.886	0.182	0.257	0.489	0.199	0.215	0.130	0.106	0.129
N	3,788	3,788	2,606	2,606	2,606	2,606	2,606	3,794	3,761	3,802

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.8b: Household's savings: Subsample of extremely poor households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

	Household has any savings (with RUSACCO, VESA, VSLA, MFI, bank, at home, with a relative)	Total savings (RUSACCO, VESA, VSLA, MFI, bank, at home, with a relative)	Household has any savings at home or with a relative	Household has any savings in the bank	Household has any savings with MFI	Household has any savings with a VESA/ VSLA	Household has any savings with a RUSACCO	Household is a member of VESA/ VSLA	Household is a member of RUSACCO	A member of the household has a bank account
T1 x Poultry	0.328***	441.008***	-0.167***	-0.079	-0.032	0.552***	-0.061	0.635***	0.025	0.035
	(0.054)	(157.023)	(0.040)	(0.059)	(0.062)	(0.072)	(0.064)	(0.056)	(0.039)	(0.038)
	[0.000]	[0.011]	[0.000]	[0.303]	[0.610]	[0.000]	[0.447]	[0.000]	[0.582]	[0.447]
T1 x Cash	0.311***	363.504*	-0.129***	-0.081	-0.125*	0.499***	0.043	0.561***	0.101**	0.008
	(0.054)	(188.901)	(0.042)	(0.055)	(0.067)	(0.072)	(0.072)	(0.057)	(0.046)	(0.033)
	[0.000]	[0.090]	[0.006]	[0.179]	[0.090]	[0.000]	[0.612]	[0.000]	[0.058]	[0.819]
T2 x Poultry	0.339***	343.009*	-0.135***	-0.068	-0.166**	0.598***	-0.099	0.667***	0.014	0.024
	(0.052)	(175.328)	(0.043)	(0.050)	(0.069)	(0.067)	(0.064)	(0.042)	(0.039)	(0.030)
	[0.000]	[0.087]	[0.005]	[0.217]	[0.035]	[0.000]	[0.172]	[0.000]	[0.723]	[0.471]
T2 x Cash	0.374***	299.598*	-0.121***	-0.095*	-0.098	0.565***	0.003	0.637***	0.067*	0.003
	(0.053)	(167.821)	(0.045)	(0.051)	(0.066)	(0.070)	(0.065)	(0.044)	(0.039)	(0.033)
	[0.000]	[0.126]	[0.020]	[0.126]	[0.171]	[0.000]	[0.969]	[0.000]	[0.127]	[0.969]
T3	0.330***	388.593**	-0.095**	-0.082*	-0.131**	0.570***	-0.103*	0.613***	-0.011	0.012
	(0.043)	(156.756)	(0.042)	(0.047)	(0.059)	(0.066)	(0.062)	(0.042)	(0.034)	(0.027)
	[0.000]	[0.035]	[0.047]	[0.122]	[0.047]	[0.000]	[0.123]	[0.000]	[0.748]	[0.734]
Average effect of T1	0.313***	393.966***	-0.145***	-0.079	-0.078	0.515***	-0.008	0.586***	0.063*	0.020
	(0.045)	(139.268)	(0.039)	(0.048)	(0.058)	(0.066)	(0.061)	(0.046)	(0.035)	(0.028)
	[0.000]	[0.010]	[0.001]	[0.152]	[0.225]	[0.000]	[0.897]	[0.000]	[0.122]	[0.513]
Average effect of T2	0.363***	327.734**	-0.130***	-0.083*	-0.135**	0.593***	-0.050	0.664***	0.040	0.014
	(0.046)	(142.038)	(0.042)	(0.046)	(0.062)	(0.066)	(0.061)	(0.039)	(0.034)	(0.026)
	[0.000]	[0.044]	[0.005]	[0.106]	[0.052]	[0.000]	[0.459]	[0.000]	[0.295]	[0.586]
Average effect of poultry	0.334***	390.081***	-0.150***	-0.074	-0.101*	0.576***	-0.081	0.652***	0.019	0.029
	(0.045)	(136.292)	(0.040)	(0.047)	(0.060)	(0.065)	(0.060)	(0.041)	(0.033)	(0.027)
	[0.000]	[0.009]	[0.001]	[0.174]	[0.154]	[0.000]	[0.220]	[0.000]	[0.562]	[0.310]
Average effect of cash	0.342***	331.619**	-0.125***	-0.088*	-0.112*	0.532***	0.023	0.599***	0.084**	0.005
	(0.046)	(146.761)	(0.041)	(0.047)	(0.060)	(0.067)	(0.063)	(0.043)	(0.036)	(0.027)
	[0.000]	[0.042]	[0.006]	[0.082]	[0.082]	[0.000]	[0.798]	[0.000]	[0.040]	[0.838]
Test: T1 = T2	(0.196)	(0.633)	(0.455)	(0.905)	(0.143)	(0.028)	(0.231)	(0.043)	(0.473)	(0.829)
	[0.461]	[0.792]	[0.675]	[0.905]	[0.461]	[0.216]	[0.461]	[0.216]	[0.675]	[0.905]

Test: T2 = T3	(0.343)	(0.696)	(0.153)	(0.978)	(0.914)	(0.512)	(0.142)	(0.134)	(0.082)	(0.928)
	[0.685]	[0.978]	[0.384]	[0.978]	[0.978]	[0.854]	[0.384]	[0.384]	[0.384]	[0.978]
Test: T1 = T3	(0.627)	(0.972)	(0.021)	(0.927)	(0.143)	(0.134)	(0.015)	(0.516)	(0.021)	(0.771)
	[0.895]	[0.972]	[0.071]	[0.972]	[0.287]	[0.287]	[0.071]	[0.860]	[0.071]	[0.964]
Test: Poultry = Cash	(0.830)	(0.682)	(0.213)	(0.688)	(0.789)	(0.208)	(0.004)	(0.164)	(0.037)	(0.386)
	[0.830]	[0.830]	[0.425]	[0.830]	[0.830]	[0.425]	[0.041]	[0.425]	[0.187]	[0.644]
Mean of control (T4)	0.400	677.288	0.198	0.238	0.459	0.209	0.215	0.118	0.096	0.106
N	1,752	1,752	1,155	1,155	1,155	1,155	1,155	1,764	1,737	1,764

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.8c: Household's savings: Subsample of less poor households

	(1) Household has any savings (with RUSACCO, VESA, VSLA, MFI, bank, at home, with a relative)	(2) Total savings (RUSACCO, VESA, VSLA, MFI, bank, at home, with a relative)	(3) Household has any savings at home or with a relative	(4) Household has any savings in the bank	(5) Household has any savings with MFI	(6) Household has any savings with a VESA/ VSLA	(7) Household has any savings with a RUSACCO	(8) Household is a member of VESA/ VSLA	(9) Household is a member of RUSACCO	(10) A member of the household has a bank account
T1	0.285***	769.249***	-0.088**	-0.077	-0.059	0.504***	-0.007	0.607***	0.078*	0.013
	(0.047)	(217.778)	(0.034)	(0.053)	(0.058)	(0.064)	(0.064)	(0.053)	(0.040)	(0.032)
	[0.000]	[0.001]	[0.021]	[0.210]	[0.386]	[0.000]	[0.911]	[0.000]	[0.088]	[0.764]
T2	0.321***	711.565***	-0.099***	-0.122**	-0.030	0.662***	-0.126**	0.696***	-0.021	-0.017
	(0.049)	(234.966)	(0.034)	(0.050)	(0.057)	(0.057)	(0.059)	(0.043)	(0.037)	(0.031)
	[0.000]	[0.007]	[0.008]	[0.026]	[0.599]	[0.000]	[0.049]	[0.000]	[0.599]	[0.599]
T3	0.301***	338.211*	-0.079**	-0.128**	-0.127**	0.636***	-0.084	0.678***	0.009	-0.018
	(0.045)	(201.402)	(0.033)	(0.050)	(0.057)	(0.057)	(0.058)	(0.043)	(0.036)	(0.032)
	[0.000]	[0.135]	[0.036]	[0.030]	[0.045]	[0.000]	[0.188]	[0.000]	[0.803]	[0.644]
Test: T1 = T2	(0.369)	(0.820)	(0.614)	(0.238)	(0.441)	(0.001)	(0.003)	(0.057)	(0.004)	(0.333)

	[0.527]	[0.820]	[0.682]	[0.476]	[0.551]	[0.010]	[0.015]	[0.143]	[0.015]	[0.527]
Test: T2 = T3	(0.587)	(0.119)	(0.351)	(0.879)	(0.007)	(0.491)	(0.202)	(0.596)	(0.315)	(0.982)
	[0.745]	[0.594]	[0.701]	[0.976]	[0.067]	[0.745]	[0.675]	[0.745]	[0.701]	[0.982]
Test: T1 = T3	(0.652)	(0.056)	(0.674)	(0.179)	(0.057)	(0.006)	(0.046)	(0.134)	(0.041)	(0.326)
	[0.674]	[0.114]	[0.674]	[0.255]	[0.114]	[0.059]	[0.114]	[0.223]	[0.114]	[0.407]
Mean of control (T4)	0.474	714.808	0.182	0.291	0.459	0.155	0.223	0.113	0.115	0.151
N	1,320	1,320	943	943	943	943	943	1,317	1,312	1,321

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.9a: Female's savings

	(1) Primary female has personal savings of her own	(2) Primary female's amount of savings (birr)	(3) Primary female deposits her savings with VESA/VSLA	(4) Primary female deposits savings with RUSACCO in last 12 months	(5) Primary female deposits her savings with MFI	(6) Primary female reports membership of VESA/VSLA	(7) Primary female reports membership of RUSACCO	(8) Primary female reports membership of MFI	(9) Primary female reports having a bank account
T1	0.127***	91.228	0.497***	-0.006	-0.272***	0.653***	0.108***	0.025	0.016
	(0.028)	(58.249)	(0.084)	(0.068)	(0.073)	(0.042)	(0.034)	(0.031)	(0.014)
	[0.000]	[0.178]	[0.000]	[0.924]	[0.001]	[0.000]	[0.003]	[0.484]	[0.339]
T2	0.154***	47.032	0.630***	-0.074	-0.287***	0.683***	0.029	0.004	0.002
	(0.032)	(48.805)	(0.075)	(0.063)	(0.076)	(0.035)	(0.026)	(0.033)	(0.014)
	[0.000]	[0.433]	[0.000]	[0.404]	[0.001]	[0.000]	[0.404]	[0.907]	[0.907]
T3	0.151***	89.197*	0.588***	-0.100	-0.232***	0.642***	0.007	-0.011	-0.008
	(0.030)	(52.590)	(0.078)	(0.066)	(0.081)	(0.036)	(0.027)	(0.033)	(0.012)
	[0.000]	[0.165]	[0.000]	[0.196]	[0.011]	[0.000]	[0.793]	[0.793]	[0.690]
Test: T1 = T2	(0.345)	(0.445)	(0.008)	(0.063)	(0.747)	(0.374)	(0.012)	(0.447)	(0.355)
	[0.503]	[0.503]	[0.055]	[0.189]	[0.747]	[0.503]	[0.055]	[0.503]	[0.503]
Test: T2 = T3	(0.915)	(0.422)	(0.294)	(0.482)	(0.364)	(0.134)	(0.358)	(0.606)	(0.486)
	[0.915]	[0.625]	[0.625]	[0.625]	[0.625]	[0.625]	[0.625]	[0.682]	[0.625]
Test: T1 = T3	(0.363)	(0.973)	(0.085)	(0.025)	(0.463)	(0.764)	(0.002)	(0.190)	(0.076)

	[0.544]	[0.973]	[0.191]	[0.112]	[0.595]	[0.859]	[0.016]	[0.342]	[0.191]
Mean of control (T4)	0.109	147.753	0.194	0.173	0.480	0.121	0.087	0.230	0.047
N	3,703	3,698	804	804	804	3,681	3,644	3,694	3,703

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.9b: Female's savings: Subsample of extremely poor households

	(1) Primary female has personal savings of her own	(2) Primary female's amount of savings (birr)	(3) Primary female deposits her savings with VESA/VSLA	(4) Primary female deposits savings with RUSACCO in last 12 months	(5) Primary female deposits her savings with MFI	(6) Primary female reports membership of VESA/VSLA	(7) Primary female reports membership of RUSACCO	(8) Primary female reports membership of MFI	(9) Primary female reports having a bank account
T1 x Poultry	0.124***	73.111	0.579***	-0.063	-0.256**	0.700***	0.054	0.049	0.025
	(0.039)	(72.820)	(0.100)	(0.077)	(0.107)	(0.053)	(0.042)	(0.039)	(0.019)
	[0.005]	[0.356]	[0.000]	[0.415]	[0.040]	[0.000]	[0.274]	[0.274]	[0.274]
T1 x Cash	0.170***	-43.659	0.475***	0.006	-0.343***	0.630***	0.151***	0.007	0.038
	(0.042)	(61.860)	(0.095)	(0.077)	(0.106)	(0.056)	(0.045)	(0.045)	(0.028)
	[0.000]	[0.619]	[0.000]	[0.940]	[0.003]	[0.000]	[0.002]	[0.940]	[0.251]
T2 x Poultry	0.107***	46.570	0.578***	-0.061	-0.334***	0.700***	0.045	-0.013	0.033
	(0.040)	(71.300)	(0.082)	(0.066)	(0.099)	(0.037)	(0.035)	(0.041)	(0.027)
	[0.019]	[0.579]	[0.000]	[0.459]	[0.003]	[0.000]	[0.321]	[0.746]	[0.321]
T2 x Cash	0.209***	8.076	0.624***	-0.100	-0.309***	0.713***	0.064*	0.016	-0.018
	(0.047)	(65.376)	(0.094)	(0.073)	(0.103)	(0.041)	(0.032)	(0.041)	(0.021)
	[0.000]	[0.902]	[0.000]	[0.261]	[0.007]	[0.000]	[0.090]	[0.781]	[0.519]
T3	0.143***	66.486	0.551***	-0.089	-0.259***	0.633***	0.010	-0.011	-0.020
	(0.034)	(83.737)	(0.084)	(0.070)	(0.099)	(0.039)	(0.029)	(0.036)	(0.013)
	[0.000]	[0.551]	[0.000]	[0.314]	[0.022]	[0.000]	[0.772]	[0.772]	[0.237]
Average effect of T1	0.144***	13.242	0.516***	-0.027	-0.295***	0.652***	0.101***	0.027	0.031*
	(0.033)	(59.425)	(0.084)	(0.068)	(0.096)	(0.043)	(0.033)	(0.035)	(0.018)
	[0.000]	[0.824]	[0.000]	[0.776]	[0.005]	[0.000]	[0.005]	[0.557]	[0.138]
Average effect of T2	0.160***	28.230	0.612***	-0.082	-0.328***	0.719***	0.055*	0.001	0.008
	(0.035)	(60.536)	(0.082)	(0.067)	(0.096)	(0.035)	(0.028)	(0.035)	(0.018)

	[0.000]	[0.732]	[0.000]	[0.337]	[0.002]	[0.000]	[0.095]	[0.972]	[0.732]
Average effect of poultry	0.115***	59.318	0.578***	-0.062	-0.297***	0.700***	0.049	0.017	0.029
	(0.032)	(61.426)	(0.083)	(0.065)	(0.095)	(0.038)	(0.030)	(0.034)	(0.018)
	[0.001]	[0.388]	[0.000]	[0.388]	[0.005]	[0.000]	[0.158]	[0.622]	[0.158]
Average effect of cash	0.189***	-17.846	0.550***	-0.047	-0.326***	0.671***	0.107***	0.012	0.010
	(0.036)	(58.311)	(0.085)	(0.070)	(0.097)	(0.041)	(0.031)	(0.036)	(0.019)
	[0.000]	[0.760]	[0.000]	[0.754]	[0.002]	[0.000]	[0.002]	[0.760]	[0.755]
Test: T1 = T2	(0.657)	(0.741)	(0.091)	(0.160)	(0.537)	(0.060)	(0.162)	(0.411)	(0.308)
	[0.740]	[0.741]	[0.365]	[0.365]	[0.691]	[0.365]	[0.365]	[0.617]	[0.554]
Test: T2 = T3	(0.626)	(0.617)	(0.265)	(0.885)	(0.226)	(0.007)	(0.127)	(0.721)	(0.110)
	[0.805]	[0.805]	[0.477]	[0.885]	[0.477]	[0.067]	[0.382]	[0.811]	[0.382]
Test: T1 = T3	(0.956)	(0.459)	(0.572)	(0.228)	(0.555)	(0.641)	(0.009)	(0.266)	(0.005)
	[0.956]	[0.722]	[0.722]	[0.598]	[0.722]	[0.722]	[0.041]	[0.598]	[0.041]
Test: Poultry = Cash	(0.034)	(0.087)	(0.630)	(0.706)	(0.596)	(0.431)	(0.074)	(0.878)	(0.398)
	[0.260]	[0.260]	[0.794]	[0.794]	[0.794]	[0.775]	[0.260]	[0.878]	[0.775]
Mean of control (T4)	0.110	177.690	0.217	0.174	0.500	0.110	0.077	0.210	0.048
N	1,722	1,719	381	381	381	1,714	1,684	1,720	1,723

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.9c: Female's savings: Subsample of less poor households

	(1) Primary female has personal savings of her own	(2) Primary female's amount of savings (birr)	(3) Primary female deposits her savings with VESA/VSLA	(4) Primary female deposits savings with RUSACCO in last 12 months	(5) Primary female deposits her savings with MFI	(6) Primary female reports membership of VESA/VSLA	(7) Primary female reports membership of RUSACCO	(8) Primary female reports membership of MFI	(9) Primary female reports having a bank account
T1	0.116***	201.852*	0.465***	0.036	-0.231**	0.671***	0.090**	0.029	-0.009
	(0.036)	(107.585)	(0.123)	(0.089)	(0.100)	(0.050)	(0.043)	(0.038)	(0.020)
	[0.004]	[0.093]	[0.001]	[0.686]	[0.052]	[0.000]	[0.066]	[0.561]	[0.686]
T2	0.170***	93.612	0.649***	-0.027	-0.270***	0.729***	-0.008	0.058	-0.010
	(0.040)	(64.826)	(0.113)	(0.084)	(0.101)	(0.043)	(0.037)	(0.044)	(0.021)

	[0.000]	[0.271]	[0.000]	[0.825]	[0.019]	[0.000]	[0.825]	[0.282]	[0.820]
T3	0.177***	144.295**	0.652***	-0.094	-0.212**	0.722***	-0.004	0.011	-0.007
	(0.036)	(56.672)	(0.115)	(0.078)	(0.105)	(0.042)	(0.036)	(0.042)	(0.023)
	[0.000]	[0.026]	[0.000]	[0.351]	[0.082]	[0.000]	[0.912]	[0.894]	[0.894]
Test: T1 = T2	(0.155)	(0.352)	(0.008)	(0.246)	(0.535)	(0.196)	(0.006)	(0.426)	(0.939)
	[0.442]	[0.528]	[0.038]	[0.443]	[0.601]	[0.442]	[0.038]	[0.547]	[0.939]
Test: T2 = T3	(0.844)	(0.463)	(0.956)	(0.086)	(0.414)	(0.835)	(0.878)	(0.236)	(0.881)
	[0.956]	[0.956]	[0.956]	[0.771]	[0.956]	[0.956]	[0.956]	[0.956]	[0.956]
Test: T1 = T3	(0.072)	(0.605)	(0.007)	(0.008)	(0.779)	(0.245)	(0.007)	(0.581)	(0.930)
	[0.161]	[0.778]	[0.024]	[0.024]	[0.877]	[0.440]	[0.024]	[0.778]	[0.930]
Mean of control (T4)	0.103	118.585	0.194	0.161	0.419	0.091	0.114	0.219	0.056
N	1,288	1,286	284	284	284	1,278	1,274	1,282	1,287

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.10a: Household's credit access

	(1) Household has taken out a loan for productive purposes in the last 12 months	(2) Household obtained a productive loan from VESA/VSLA	(3) Amount of production credit taken out	(4) Household has taken out a loan for consumption purposes in the last 12 months
T1	0.061**	0.079***	287.831	0.035
	(0.030)	(0.029)	(381.435)	(0.032)
	[0.080]	[0.030]	[0.451]	[0.373]
T2	0.038	0.112***	33.811	0.015
	(0.029)	(0.031)	(425.083)	(0.027)
	[0.366]	[0.002]	[0.937]	[0.764]
T3	0.004	0.156***	-199.925	-0.018
	(0.028)	(0.033)	(394.386)	(0.026)
	[0.896]	[0.000]	[0.817]	[0.817]
Test: T1 = T2	(0.403)	(0.293)	(0.517)	(0.530)
	[0.530]	[0.530]	[0.530]	[0.530]

Test: T2 = T3	(0.167)	(0.221)	(0.564)	(0.141)
	[0.294]	[0.294]	[0.564]	[0.294]
Test: T1 = T3	(0.028)	(0.025)	(0.176)	(0.068)
	[0.056]	[0.056]	[0.176]	[0.091]
Mean of control (T4)	0.261	0.054	2,577.033	0.263
N	3,804	1,109	3,804	3,804

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.10b: Household's credit access: Subsample of extremely poor households

	(1) Household has taken out a loan for productive purposes in the last 12 months	(2) Household obtained a productive loan from VESA/VSLA	(3) Amount of production credit taken out (birr)	(4) Household has taken out a loan for consumption purposes in the last 12 months
T1 x Poultry	0.087**	-0.011	500.411	0.031
	(0.043)	(0.046)	(453.086)	(0.044)
	[0.176]	[0.815]	[0.542]	[0.646]
T1 x Cash	0.057	0.084	-152.085	0.045
	(0.042)	(0.079)	(468.078)	(0.052)
	[0.515]	[0.515]	[0.746]	[0.515]
T2 x Poultry	0.052	0.152**	564.279	-0.022
	(0.040)	(0.067)	(926.708)	(0.043)
	[0.402]	[0.102]	[0.611]	[0.611]
T2 x Cash	0.086*	-0.006	509.363	0.080**
	(0.044)	(0.044)	(526.563)	(0.038)
	[0.108]	[0.895]	[0.446]	[0.108]
T3	-0.011	0.089*	-309.453	-0.019
	(0.035)	(0.054)	(412.121)	(0.032)
	[0.759]	[0.397]	[0.741]	[0.741]
Average effect of T1	0.070**	0.037	164.160	0.037
	(0.034)	(0.051)	(383.584)	(0.037)
	[0.155]	[0.630]	[0.669]	[0.630]

Average effect of T2	0.070**	0.076	547.381	0.028
	(0.035)	(0.048)	(582.010)	(0.034)
	[0.185]	[0.224]	[0.403]	[0.403]
Average effect of poultry	0.068**	0.074	533.601	0.003
	(0.034)	(0.047)	(565.057)	(0.034)
	[0.179]	[0.239]	[0.462]	[0.920]
Average effect of cash	0.072**	0.039	177.940	0.062*
	(0.035)	(0.052)	(408.664)	(0.037)
	[0.161]	[0.597]	[0.664]	[0.183]
Test: T1 = T2	(0.985)	(0.463)	(0.493)	(0.814)
	[0.985]	[0.985]	[0.985]	[0.985]
Test: T2 = T3	(0.021)	(0.801)	(0.135)	(0.137)
	[0.083]	[0.801]	[0.183]	[0.183]
Test: T1 = T3	(0.019)	(0.356)	(0.195)	(0.124)
	[0.076]	[0.356]	[0.260]	[0.247]
Test: Poultry = Cash	(0.923)	(0.514)	(0.525)	(0.124)
	[0.923]	[0.700]	[0.700]	[0.495]
Mean of control (T4)	0.233	0.119	2,226.037	0.270
N	1,765	479	1,765	1,765

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.10c: Household's credit access: Subsample of less poor households

	(1) Household has taken out a loan for productive purposes in the last 12 months	(2) Household obtained a productive loan from VESA/VSLA	(3) Amount of production credit taken out (birr)	(4) Household has taken out a loan for consumption purposes in the last 12 months
T1	0.075*	0.066**	1,035.443*	-0.012
	(0.043)	(0.027)	(606.286)	(0.046)
	[0.119]	[0.067]	[0.119]	[0.788]
T2	0.055	0.095***	179.771	0.002

	(0.040)	(0.036)	(580.503)	(0.042)
	[0.342]	[0.040]	[0.954]	[0.954]
T3	0.049	0.210***	266.742	-0.042
	(0.038)	(0.045)	(576.658)	(0.038)
	[0.352]	[0.000]	[0.644]	[0.352]
Test: T1 = T2	(0.592)	(0.413)	(0.147)	(0.740)
	[0.740]	[0.740]	[0.588]	[0.740]
Test: T2 = T3	(0.836)	(0.029)	(0.871)	(0.198)
	[0.871]	[0.118]	[0.871]	[0.396]
Test: T1 = T3	(0.457)	(0.002)	(0.185)	(0.449)
	[0.457]	[0.007]	[0.371]	[0.457]
Mean of control (T4)	0.263	0.012	2,456.619	0.272
N	1,322	418	1,322	1,322

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.11a: Female's credit access

	(1) Female has taken out a loan for productive purposes in the last 12 months	(2) Female obtained a productive loan from VESA/VSLA	(3) Amount of production credit taken out (birr)	(4) Female has taken out a loan for consumption purposes in the last 12 months
T1	0.036	0.170***	263.903	0.048*
	(0.024)	(0.055)	(300.411)	(0.028)
	[0.183]	[0.009]	[0.381]	[0.183]
T2	0.037	0.148***	31.844	0.003
	(0.024)	(0.046)	(267.044)	(0.025)
	[0.247]	[0.006]	[0.918]	[0.918]
T3	0.009	0.185***	-98.685	-0.005
	(0.023)	(0.053)	(256.164)	(0.026)
	[0.834]	[0.002]	[0.834]	[0.834]
Test: T1 = T2	(0.950)	(0.690)	(0.427)	(0.115)

	[0.950]	[0.920]	[0.854]	[0.461]
Test: T2 = T3	(0.212)	(0.468)	(0.588)	(0.757)
	[0.757]	[0.757]	[0.757]	[0.757]
Test: T1 = T3	(0.239)	(0.805)	(0.198)	(0.064)
	[0.319]	[0.805]	[0.319]	[0.257]
Mean of control (T4)	0.159	0.070	1,376.452	0.255
N	3,703	670	3,701	3,703

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.11b: Female's credit access: Subsample of extremely poor households

	(1) Female has taken out a loan for productive purposes in the last 12 months	(2) Female obtained a productive loan from VESA/VSLA	(3) Amount of production credit taken out (birr)	(4) Female has taken out a loan for consumption purposes in the last 12 months
T1 x Poultry	0.064	0.034	189.446	0.074
	(0.039)	(0.062)	(411.196)	(0.047)
	[0.233]	[0.646]	[0.646]	[0.233]
T1 x Cash	0.018	0.219*	-17.200	0.089*
	(0.038)	(0.111)	(367.434)	(0.053)
	[0.843]	[0.183]	[0.963]	[0.183]
T2 x Poultry	0.052	0.102	120.544	-0.016
	(0.040)	(0.071)	(328.771)	(0.044)
	[0.390]	[0.390]	[0.721]	[0.721]
T2 x Cash	0.043	0.001	150.658	0.094**
	(0.042)	(0.068)	(410.597)	(0.040)
	[0.614]	[0.991]	[0.952]	[0.085]
T3	0.005	0.124*	18.947	0.015
	(0.028)	(0.072)	(302.007)	(0.032)
	[0.950]	[0.349]	[0.950]	[0.950]
Average effect of T1	0.040	0.126*	82.378	0.080**
	(0.030)	(0.066)	(309.997)	(0.038)

	[0.252]	[0.119]	[0.791]	[0.119]
Average effect of T2	0.049	0.053	137.813	0.039
	(0.033)	(0.058)	(308.339)	(0.034)
	[0.479]	[0.479]	[0.655]	[0.479]
Average effect of poultry	0.058*	0.069	153.640	0.027
	(0.032)	(0.056)	(303.829)	(0.035)
	[0.272]	[0.431]	[0.614]	[0.583]
Average effect of cash	0.030	0.110	66.551	0.091**
	(0.032)	(0.070)	(315.125)	(0.037)
	[0.451]	[0.238]	[0.833]	[0.057]
Test: T1 = T2	(0.797)	(0.300)	(0.856)	(0.305)
	[0.856]	[0.609]	[0.856]	[0.609]
Test: T2 = T3	(0.167)	(0.292)	(0.679)	(0.497)
	[0.584]	[0.584]	[0.679]	[0.663]
Test: T1 = T3	(0.237)	(0.975)	(0.835)	(0.095)
	[0.474]	[0.975]	[0.975]	[0.381]
Test: Poultry = Cash	(0.417)	(0.581)	(0.777)	(0.113)
	[0.775]	[0.775]	[0.777]	[0.451]
Mean of control (T4)	0.157	0.121	1,324.288	0.255
N	1,723	313	1,722	1,722

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.11c: Female's credit access: Subsample of less poor households

	(1) Female has taken out a loan for productive purposes in the last 12 months	(2) Female obtained a productive loan from VESA/VSLA	(3) Amount of production credit taken out (birr)	(4) Female has taken out a loan for consumption purposes in the last 12 months
T1	0.035	0.140*	385.892	-0.014
	(0.036)	(0.073)	(441.247)	(0.037)
	[0.511]	[0.234]	[0.511]	[0.708]

T2	0.062*	0.215***	244.225	-0.015
	(0.034)	(0.074)	(421.866)	(0.038)
	[0.143]	[0.018]	[0.691]	[0.691]
T3	0.033	0.235***	22.944	-0.026
	(0.033)	(0.084)	(365.250)	(0.036)
	[0.614]	[0.024]	[0.950]	[0.635]
Test: T1 = T2	(0.404)	(0.309)	(0.726)	(0.974)
	[0.809]	[0.809]	[0.968]	[0.974]
Test: T2 = T3	(0.339)	(0.810)	(0.501)	(0.782)
	[0.810]	[0.810]	[0.810]	[0.810]
Test: T1 = T3	(0.971)	(0.261)	(0.294)	(0.756)
	[0.971]	[0.587]	[0.587]	[0.971]
Mean of control (T4)	0.143	0.048	1,271.333	0.282
N	1,287	227	1,286	1,288

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.12a: Household's aggregate assets

	(1) Estimated value of all livestock owned by the household[a]	(2) Estimated value of all livestock owned by woman (solely or jointly)[a]	(3) Household Livestock Asset Index	(4) Household Productive Asset Index	(5) Consumer Durable Asset Index	(6) Household Total Asset Index
T1	0.039	-0.010	0.105	0.049	0.061	0.114
	(0.302)	(0.306)	(0.141)	(0.139)	(0.109)	(0.162)
	[0.974]	[0.974]	[0.725]	[0.725]	[0.725]	[0.725]
T2	0.363*	0.217	-0.013	0.026	0.173	0.130
	(0.219)	(0.232)	(0.116)	(0.128)	(0.123)	(0.148)
	[0.159]	[0.410]	[0.954]	[0.954]	[0.954]	[0.954]
T3	-0.315	-0.257	-0.026	-0.028	0.179	0.054
	(0.310)	(0.290)	(0.131)	(0.121)	(0.141)	(0.141)
	[0.603]	[0.603]	[0.846]	[0.846]	[0.411]	[0.846]
Test: T1 = T2	(0.220)	(0.414)	(0.364)	(0.864)	(0.357)	(0.919)

	[0.879]	[0.980]	[0.919]	[0.919]	[0.919]	[0.919]
Test: T2 = T3	(0.014)	(0.069)	(0.912)	(0.654)	(0.969)	(0.576)
	[0.087]	[0.167]	[0.969]	[0.969]	[0.969]	[0.969]
Test: T1 = T3	(0.293)	(0.443)	(0.360)	(0.558)	(0.389)	(0.699)
	[0.502]	[0.590]	[0.699]	[0.699]	[0.699]	[0.699]
Mean of control (T4)	9.094	8.766	-0.013	-0.065	-0.156	-0.150
N	3,804	3,688	3,792	3,800	3,797	3,787

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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.

[a] Inverse hyperbolic sine transformation has been applied to the outcomes.

Table 7.12b: Household's aggregate assets: Subsample of extremely poor households

	(1) Estimated value of all livestock owned by the household[a]	(2) Estimated value of all livestock owned by woman (solely or jointly)[a]	(3) Household Livestock Asset Index	(4) Household Productive Asset Index	(5) Consumer Durable Asset Index	(6) Household Total Asset Index
T1 x Poultry	1.066***	0.837*	0.463**	0.293	-0.022	0.446**
	(0.366)	(0.434)	(0.188)	(0.193)	(0.147)	(0.224)
	[0.011]	[0.089]	[0.087]	[0.261]	[0.883]	[0.143]
T1 x Cash	0.022	0.019	-0.030	-0.078	-0.074	-0.125
	(0.500)	(0.483)	(0.165)	(0.236)	(0.118)	(0.239)
	[0.969]	[0.969]	[0.857]	[0.857]	[0.857]	[0.857]
T2 x Poultry	0.695**	0.872**	0.027	0.084	0.184	0.249
	(0.308)	(0.340)	(0.159)	(0.191)	(0.170)	(0.210)
	[0.040]	[0.029]	[0.918]	[0.918]	[0.840]	[0.840]
T2 x Cash	1.206***	1.021***	0.302**	0.326*	0.158	0.450**
	(0.307)	(0.363)	(0.134)	(0.179)	(0.157)	(0.194)
	[0.001]	[0.019]	[0.077]	[0.105]	[0.316]	[0.077]
T3	-0.518	-0.376	-0.047	-0.086	0.110	-0.041
	(0.400)	(0.382)	(0.136)	(0.148)	(0.146)	(0.173)
	[0.610]	[0.610]	[0.811]	[0.811]	[0.811]	[0.811]
Average effect of T1	0.523	0.411	0.207	0.102	-0.047	0.152
	(0.352)	(0.368)	(0.138)	(0.167)	(0.108)	(0.181)

	[0.279]	[0.425]	[0.733]	[0.733]	[0.733]	[0.733]
Average effect of T2	0.963***	0.962***	0.164	0.206	0.175	0.354**
	(0.278)	(0.310)	(0.126)	(0.154)	(0.130)	(0.171)
	[0.003]	[0.004]	[0.233]	[0.233]	[0.233]	[0.233]
Average effect of poultry	0.873***	0.855***	0.236*	0.185	0.085	0.344*
	(0.291)	(0.327)	(0.138)	(0.155)	(0.125)	(0.175)
	[0.008]	[0.015]	[0.264]	[0.468]	[0.595]	[0.264]
Average effect of cash	0.613*	0.519	0.135	0.124	0.042	0.162
	(0.343)	(0.357)	(0.126)	(0.168)	(0.114)	(0.178)
	[0.139]	[0.198]	[0.557]	[0.557]	[0.715]	[0.557]
Test: T1 = T2	(0.140)	(0.083)	(0.749)	(0.533)	(0.080)	(0.259)
	[0.230]	[0.230]	[0.749]	[0.668]	[0.482]	[0.668]
Test: T2 = T3	(0.000)	(0.000)	(0.096)	(0.050)	(0.684)	(0.019)
	[0.000]	[0.000]	[0.115]	[0.075]	[0.684]	[0.038]
Test: T1 = T3	(0.012)	(0.042)	(0.080)	(0.257)	(0.257)	(0.294)
	[0.031]	[0.084]	[0.239]	[0.294]	[0.294]	[0.294]
Test: Poultry = Cash	(0.389)	(0.301)	(0.447)	(0.724)	(0.732)	(0.314)
	[0.688]	[0.688]	[0.750]	[0.750]	[0.750]	[0.750]
Mean of control (T4)	8.366	7.964	-0.317	-0.429	-0.201	-0.597
N	1,765	1,713	1,762	1,762	1,760	1,758

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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."

[a] Inverse hyperbolic sine transformation has been applied to the outcomes.

Table 7.12c: Household's aggregate assets: Subsample of less poor households

	(1) Estimated value of all livestock owned by the household[a]	(2) Estimated value of all livestock owned by woman (solely or jointly)[a]	(3) Household Livestock Asset Index	(4) Household Productive Asset Index	(5) Consumer Durable Asset Index	(6) Household Total Asset Index
T1	-0.571*	-0.563*	-0.084	0.067	0.184	0.138
	(0.314)	(0.334)	(0.179)	(0.171)	(0.142)	(0.205)
	[0.155]	[0.155]	[0.698]	[0.698]	[0.393]	[0.698]

T2	-0.365	-0.612**	-0.220	-0.068	0.324**	0.060
	(0.251)	(0.278)	(0.161)	(0.171)	(0.152)	(0.195)
	[0.222]	[0.145]	[0.262]	[0.761]	[0.105]	[0.761]
T3	-0.445	-0.397	-0.110	0.040	0.248	0.132
	(0.322)	(0.307)	(0.176)	(0.154)	(0.157)	(0.177)
	[0.283]	[0.283]	[0.639]	[0.793]	[0.234]	[0.639]
Test: T1 = T2	(0.500)	(0.887)	(0.371)	(0.444)	(0.380)	(0.704)
	[0.887]	[0.887]	[0.766]	[0.766]	[0.766]	[0.771]
Test: T2 = T3	(0.797)	(0.495)	(0.454)	(0.480)	(0.649)	(0.670)
	[0.886]	[0.886]	[0.711]	[0.711]	[0.711]	[0.711]
Test: T1 = T3	(0.725)	(0.643)	(0.876)	(0.871)	(0.694)	(0.972)
	[0.856]	[0.856]	[0.972]	[0.972]	[0.972]	[0.972]
Mean of control (T4)	10.062	9.817	0.390	0.302	-0.164	0.306
N	1,322	1,282	1,317	1,321	1,320	1,316

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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."

[a] Inverse hyperbolic sine transformation has been applied to the outcomes.

Table 7.13a: Most owned productive assets and consumer durables

	(1) Household owns at least one sickle	(2) Household owns at least one axe	(3) Household owns at least one spade or shovel	(4) Household owns at least one solar panel	(5) Household owns at least one blanket	(6) Household owns at least one flashlight/torch
T1	-0.024	0.008	-0.045	-0.024	0.021	-0.037
	(0.029)	(0.029)	(0.031)	(0.033)	(0.034)	(0.028)
	[0.658]	[0.779]	[0.557]	[0.658]	[0.658]	[0.557]
T2	0.029	0.038	0.019	0.008	0.026	-0.018
	(0.026)	(0.029)	(0.031)	(0.028)	(0.036)	(0.031)
	[0.671]	[0.671]	[0.671]	[0.789]	[0.671]	[0.671]
T3	0.029	0.022	-0.012	-0.028	0.024	0.028
	(0.026)	(0.030)	(0.029)	(0.032)	(0.041)	(0.031)
	[0.672]	[0.672]	[0.672]	[0.672]	[0.672]	[0.672]
Test: T1 = T2	(0.064)	(0.305)	(0.063)	(0.340)	(0.862)	(0.517)
	[0.193]	[0.510]	[0.193]	[0.510]	[0.862]	[0.620]

Test: T2 = T3	(0.985)	(0.603)	(0.340)	(0.260)	(0.946)	(0.149)
	[0.985]	[0.905]	[0.680]	[0.680]	[0.985]	[0.680]
Test: T1 = T3	(0.056)	(0.632)	(0.328)	(0.899)	(0.936)	(0.029)
	[0.169]	[0.936]	[0.656]	[0.936]	[0.936]	[0.169]
Mean of control (T4)	0.774	0.711	0.638	0.713	0.566	0.606
N	3,803	3,804	3,803	3,804	3,803	3,802

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.13b: Most owned productive assets and consumer durables: Subsample of extremely poor households

	(1) Household owns at least one sickle	(2) Household owns at least one axe	(3) Household owns at least one spade or shovel	(4) Household owns at least one solar panel	(5) Household owns at least one blanket	(6) Household owns at least one flashlight/torch
T1 x Poultry	0.003	0.063	-0.074	0.002	-0.029	-0.037
	(0.046)	(0.050)	(0.046)	(0.041)	(0.053)	(0.049)
	[0.962]	[0.625]	[0.625]	[0.962]	[0.885]	[0.885]
T1 x Cash	-0.045	-0.019	-0.058	-0.035	0.014	-0.022
	(0.056)	(0.046)	(0.044)	(0.052)	(0.049)	(0.048)
	[0.776]	[0.776]	[0.776]	[0.776]	[0.776]	[0.776]
T2 x Poultry	0.084*	0.075	-0.011	0.009	0.022	-0.003
	(0.047)	(0.050)	(0.046)	(0.039)	(0.051)	(0.054)
	[0.412]	[0.412]	[0.957]	[0.957]	[0.957]	[0.957]
T2 x Cash	0.053	0.104**	0.075	0.062	0.016	-0.045
	(0.044)	(0.040)	(0.045)	(0.042)	(0.046)	(0.048)
	[0.342]	[0.064]	[0.273]	[0.273]	[0.728]	[0.422]
T3	0.016	0.006	-0.070*	-0.038	0.010	0.023
	(0.038)	(0.038)	(0.039)	(0.037)	(0.043)	(0.038)
	[0.872]	[0.872]	[0.462]	[0.872]	[0.872]	[0.872]
Average effect of T1	-0.021	0.021	-0.065*	-0.016	-0.007	-0.028
	(0.041)	(0.038)	(0.035)	(0.036)	(0.041)	(0.038)
	[0.783]	[0.783]	[0.417]	[0.783]	[0.869]	[0.783]
Average effect of T2	0.070*	0.091**	0.031	0.036	0.019	-0.024
	(0.038)	(0.039)	(0.037)	(0.034)	(0.041)	(0.042)
	[0.199]	[0.128]	[0.597]	[0.572]	[0.640]	[0.640]

Average effect of poultry	0.045	0.069*	-0.041	0.006	-0.002	-0.019
	(0.038)	(0.041)	(0.037)	(0.032)	(0.042)	(0.041)
	[0.520]	[0.520]	[0.520]	[0.953]	[0.953]	[0.953]
Average effect of cash	0.004	0.042	0.008	0.014	0.015	-0.033
	(0.041)	(0.037)	(0.036)	(0.038)	(0.040)	(0.039)
	[0.918]	[0.918]	[0.918]	[0.918]	[0.918]	[0.918]
Test: T1 = T2	(0.024)	(0.056)	(0.012)	(0.151)	(0.516)	(0.910)
	[0.072]	[0.111]	[0.070]	[0.227]	[0.619]	[0.910]
Test: T2 = T3	(0.140)	(0.018)	(0.014)	(0.039)	(0.815)	(0.244)
	[0.210]	[0.053]	[0.053]	[0.078]	[0.815]	[0.293]
Test: T1 = T3	(0.354)	(0.691)	(0.902)	(0.584)	(0.695)	(0.169)
	[0.834]	[0.834]	[0.902]	[0.834]	[0.834]	[0.834]
Test: Poultry = Cash	(0.314)	(0.471)	(0.193)	(0.829)	(0.657)	(0.738)
	[0.829]	[0.829]	[0.829]	[0.829]	[0.829]	[0.829]
Mean of control (T4)	0.723	0.650	0.608	0.691	0.560	0.588
N	1,764	1,765	1,764	1,765	1,764	1,764

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.13c: Most owned productive assets and consumer durables: Subsample of less poor households

	(1) Household owns at least one sickle	(2) Household owns at least one axe	(3) Household owns at least one spade or shovel	(4) Household owns at least one solar panel	(5) Household owns at least one blanket	(6) Household owns at least one flashlight/torch
T1	-0.057*	0.021	-0.032	0.006	0.041	-0.039
	(0.032)	(0.038)	(0.044)	(0.047)	(0.042)	(0.040)
	[0.471]	[0.696]	[0.696]	[0.906]	[0.651]	[0.651]
T2	-0.020	0.014	-0.017	0.009	0.029	0.025
	(0.032)	(0.043)	(0.040)	(0.038)	(0.047)	(0.041)
	[0.816]	[0.816]	[0.816]	[0.816]	[0.816]	[0.816]
T3	0.019	0.049	0.017	0.003	0.034	0.046
	(0.033)	(0.042)	(0.040)	(0.042)	(0.051)	(0.041)
	[0.811]	[0.786]	[0.811]	[0.947]	[0.811]	[0.786]
Test: T1 = T2	(0.243)	(0.835)	(0.735)	(0.943)	(0.736)	(0.116)

	[0.729]	[0.943]	[0.943]	[0.943]	[0.943]	[0.698]
Test: T2 = T3	(0.219)	(0.361)	(0.390)	(0.871)	(0.920)	(0.580)
	[0.780]	[0.780]	[0.780]	[0.920]	[0.920]	[0.870]
Test: T1 = T3	(0.020)	(0.418)	(0.284)	(0.951)	(0.852)	(0.035)
	[0.105]	[0.627]	[0.567]	[0.951]	[0.951]	[0.105]
Mean of control (T4)	0.840	0.772	0.699	0.721	0.587	0.606
N	1,322	1,322	1,322	1,322	1,322	1,321

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.14a: Housing characteristics

	(1) Household has an improved source of water (rainy season)	(2) Household has an improved roof material	(3) Household's number of bedrooms	(4) Household has access to electricity
T1	0.084*	-0.008	0.037	-0.012
	(0.046)	(0.039)	(0.046)	(0.043)
	[0.280]	[0.843]	[0.843]	[0.843]
T2	0.037	0.056	0.058	0.034
	(0.045)	(0.042)	(0.048)	(0.040)
	[0.410]	[0.410]	[0.410]	[0.410]
T3	0.048	0.014	0.061	-0.016
	(0.048)	(0.040)	(0.051)	(0.044)
	[0.646]	[0.721]	[0.646]	[0.721]
Test: T1 = T2	(0.204)	(0.090)	(0.616)	(0.250)
	[0.333]	[0.333]	[0.616]	[0.333]
Test: T2 = T3	(0.785)	(0.271)	(0.958)	(0.226)
	[0.958]	[0.542]	[0.958]	[0.542]
Test: T1 = T3	(0.388)	(0.536)	(0.599)	(0.933)
	[0.799]	[0.799]	[0.799]	[0.933]
Mean of control (T4)	0.559	0.533	1.339	0.827
N	3,775	3,773	3,764	3,772

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 7.14b: Housing characteristics: Subsample of extremely poor households

	(1) Household has an improved source of water (rainy season)	(2) Household has an improved roof material	(3) Household's number of bedrooms	(4) Household has access to electricity
T1 x Poultry	0.093	-0.031	0.063	0.002
	(0.067)	(0.068)	(0.055)	(0.057)
	[0.499]	[0.868]	[0.499]	[0.974]
T1 x Cash	0.082	0.073	0.048	-0.006
	(0.074)	(0.080)	(0.063)	(0.052)
	[0.594]	[0.594]	[0.594]	[0.908]
T2 x Poultry	0.076	0.046	0.068	0.026
	(0.068)	(0.074)	(0.076)	(0.057)
	[0.653]	[0.653]	[0.653]	[0.653]
T2 x Cash	0.014	0.056	0.069	0.043
	(0.066)	(0.075)	(0.063)	(0.055)
	[0.838]	[0.609]	[0.609]	[0.609]
T3	0.080	0.012	0.088	-0.008
	(0.062)	(0.063)	(0.061)	(0.048)
	[0.407]	[0.860]	[0.407]	[0.860]
Average effect of T1	0.086	0.022	0.054	-0.002
	(0.060)	(0.061)	(0.050)	(0.044)
	[0.557]	[0.961]	[0.557]	[0.961]
Average effect of T2	0.046	0.052	0.070	0.035
	(0.060)	(0.063)	(0.058)	(0.046)
	[0.448]	[0.448]	[0.448]	[0.448]
Average effect of poultry	0.084	0.009	0.066	0.014
	(0.059)	(0.060)	(0.055)	(0.046)
	[0.474]	[0.881]	[0.474]	[0.881]
Average effect of cash	0.048	0.064	0.058	0.019

	(0.061)	(0.064)	(0.053)	(0.044)
	[0.574]	[0.574]	[0.574]	[0.673]
Test: T1 = T2	(0.426)	(0.600)	(0.763)	(0.413)
	[0.763]	[0.763]	[0.763]	[0.763]
Test: T2 = T3	(0.516)	(0.500)	(0.759)	(0.375)
	[0.688]	[0.688]	[0.759]	[0.688]
Test: T1 = T3	(0.909)	(0.872)	(0.526)	(0.892)
	[0.909]	[0.909]	[0.909]	[0.909]
Test: Poultry = Cash	(0.464)	(0.339)	(0.882)	(0.925)
	[0.925]	[0.925]	[0.925]	[0.925]
Mean of control (T4)	0.568	0.542	1.297	0.811
N	1,748	1,748	1,742	1,746

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.14c: Housing characteristics: Subsample of less poor households

	(1) Household has an improved source of water (rainy season)	(2) Household has an improved roof material	(3) Household's number of bedrooms	(4) Household has access to electricity
T1	0.133**	0.077	0.059	0.002
	(0.059)	(0.065)	(0.057)	(0.047)
	[0.097]	[0.410]	[0.410]	[0.961]
T2	0.027	0.151**	0.081	0.051
	(0.063)	(0.062)	(0.062)	(0.043)
	[0.665]	[0.062]	[0.316]	[0.316]
T3	0.105	0.062	0.075	-0.013
	(0.065)	(0.062)	(0.061)	(0.048)
	[0.431]	[0.431]	[0.431]	[0.787]
Test: T1 = T2	(0.038)	(0.205)	(0.695)	(0.235)
	[0.150]	[0.314]	[0.695]	[0.314]
Test: T2 = T3	(0.182)	(0.100)	(0.931)	(0.139)

	[0.243]	[0.243]	[0.931]	[0.243]
Test: T1 = T3	(0.594)	(0.799)	(0.760)	(0.742)
	[0.799]	[0.799]	[0.799]	[0.799]
Mean of control (T4)	0.544	0.497	1.360	0.828
N	1,314	1,313	1,310	1,313

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 7.15a: Households' food and nonfood consumption

	(1) Total value of food consumption per month per adult equivalent (birr)	(2) Monthly expenditure on non-food items per adult equivalent (birr)	(3) Total consumption expenditure per month per adult equivalent (birr)	(4) Calories of daily food consumption per adult equivalent (kcal)
T1	-0.062	0.024	-0.049	-0.048
	(0.068)	(0.061)	(0.059)	(0.052)
	[0.539]	[0.698]	[0.539]	[0.539]
T2	-0.027	0.030	-0.015	-0.035
	(0.063)	(0.067)	(0.057)	(0.048)
	[0.797]	[0.797]	[0.797]	[0.797]
T3	-0.085	0.128*	-0.043	-0.021
	(0.070)	(0.069)	(0.059)	(0.039)
	[0.454]	[0.257]	[0.590]	[0.590]
Test: T1 = T2	(0.563)	(0.914)	(0.530)	(0.826)
	[0.914]	[0.914]	[0.914]	[0.914]
Test: T2 = T3	(0.360)	(0.155)	(0.603)	(0.756)
	[0.721]	[0.620]	[0.756]	[0.756]
Test: T1 = T3	(0.735)	(0.096)	(0.915)	(0.593)
	[0.915]	[0.385]	[0.915]	[0.915]
Mean of control (T4)	7.109	5.280	7.329	8.523
N	3,686	3,803	3,680	3,708

Note: 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.

Table 7.15b: Households' food and nonfood consumption: Subsample of extremely poor households

	(1) Total value of food consumption per month per adult equivalent (birr)	(2) Monthly expenditure on non-food items per adult equivalent (birr)	(3) Total consumption expenditure per month per adult equivalent (birr)	(4) Calories of daily food consumption per adult equivalent (kcal)
T1 x Poultry	0.039	-0.023	0.010	0.004
	(0.090)	(0.105)	(0.083)	(0.074)
	[0.961]	[0.961]	[0.961]	[0.961]
T1 x Cash	-0.135	0.053	-0.097	-0.107
	(0.095)	(0.085)	(0.082)	(0.104)
	[0.403]	[0.537]	[0.403]	[0.403]
T2 x Poultry	0.008	0.079	0.020	0.089
	(0.083)	(0.110)	(0.076)	(0.069)
	[0.919]	[0.919]	[0.919]	[0.798]
T2 x Cash	-0.009	0.072	0.003	0.009
	(0.075)	(0.092)	(0.070)	(0.073)
	[0.961]	[0.961]	[0.961]	[0.961]
T3	-0.103	0.145*	-0.064	0.002
	(0.071)	(0.082)	(0.063)	(0.055)
	[0.299]	[0.299]	[0.424]	[0.972]
Average effect of T1	-0.049	0.016	-0.043	-0.052
	(0.074)	(0.080)	(0.065)	(0.071)
	[0.680]	[0.845]	[0.680]	[0.680]
Average effect of T2	-0.000	0.077	0.012	0.051
	(0.068)	(0.087)	(0.062)	(0.059)
	[1.000]	[0.787]	[1.000]	[0.787]
Average effect of poultry	0.023	0.030	0.015	0.048
	(0.071)	(0.089)	(0.064)	(0.059)
	[0.812]	[0.812]	[0.812]	[0.812]
Average effect of cash	-0.072	0.062	-0.047	-0.049
	(0.071)	(0.078)	(0.063)	(0.071)

	[0.491]	[0.491]	[0.491]	[0.491]
Test: T1 = T2	(0.473)	(0.401)	(0.374)	(0.145)
	[0.473]	[0.473]	[0.473]	[0.473]
Test: T2 = T3	(0.115)	(0.353)	(0.210)	(0.355)
	[0.355]	[0.355]	[0.355]	[0.355]
Test: T1 = T3	(0.447)	(0.058)	(0.753)	(0.420)
	[0.596]	[0.234]	[0.753]	[0.596]
Test: Poultry = Cash	(0.165)	(0.671)	(0.323)	(0.168)
	[0.337]	[0.671]	[0.431]	[0.337]
Mean of control (T4)	7.100	5.211	7.315	8.495
N	1,706	1,764	1,701	1,714

Note: Estimates from the DFSA SPIR endline survey sample. Inverse hyperbolic sine transformation has been applied to all outcome variables presented in the table. 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. 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 7.15c: Households' food and nonfood consumption: Subsample of less poor households

	(1) Total value of food consumption per month per adult equivalent (birr)	(2) Monthly expenditure on non-food items per adult equivalent (birr)	(3) Total consumption expenditure per month per adult equivalent (birr)	(4) Calories of daily food consumption per adult equivalent (kcal)
T1	-0.144*	0.012	-0.106	-0.095
	(0.084)	(0.072)	(0.071)	(0.065)
	[0.190]	[0.864]	[0.190]	[0.190]
T2	-0.050	0.057	-0.017	-0.153*
	(0.090)	(0.085)	(0.082)	(0.079)
	[0.775]	[0.775]	[0.839]	[0.213]
T3	-0.048	0.112	-0.004	-0.051
	(0.098)	(0.081)	(0.079)	(0.058)
	[0.836]	[0.673]	[0.957]	[0.761]
Test: T1 = T2	(0.227)	(0.611)	(0.201)	(0.500)
	[0.453]	[0.611]	[0.453]	[0.611]
Test: T2 = T3	(0.981)	(0.552)	(0.871)	(0.167)

	[0.981]	[0.981]	[0.981]	[0.668]
Test: T1 = T3	(0.252)	(0.220)	(0.122)	(0.447)
	[0.335]	[0.335]	[0.335]	[0.447]
Mean of control (T4)	7.132	5.333	7.352	8.600
N	1,291	1,322	1,290	1,298

Note: Estimates from the DFSA SPIR endline survey sample. Inverse hyperbolic sine transformation has been applied to all outcome variables presented in the table. 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. 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."

8 Impacts on nutrition and food security

8.1 Introduction

8.1.1 Sample composition

The primary focus of this chapter is to evaluate the impact of the SPIR interventions related to nutritional services on nutrition, diet, food security, and utilization of health services, with a particular focus on children younger than 24 months. In anticipation of the fact that the index children for this aspect of the study from the baseline survey would not be the core of the sample in the endline survey, the sample was refreshed with additional young children (younger than 24 months) at midline. However, the delay in the endline data collection necessitated by COVID-19 meant that even most of the additional midline sample were older than 24 months by the endline. Fortunately, 1,116 of the baseline index children had younger siblings less than 24 months who were included in the endline sample. However, even so, the sample for the analysis of current feeding practices and interaction with healthcare workers is less than at the baseline and is smaller than the sample for most of the other topics analyzed in this report.[28] Table 8.1 reflects the observations in the anthropometry analysis (not the age distribution of the household). Also, since the households added in the supplemental midline sample do not have comparable poverty rankings as were used in other parts of the analyses, presentations of impacts on outcomes of extremely poor and less poor households are not strictly comparable with those in other chapters.

Table 8.1. Number of children in the anthropometrics sample

	Baseline survey	Endline survey
<12 months	1,020	720
12-23 months	1,059	656
24-35 months	1,235	787
36-47 months	0	1,069
>48 months	0	877
Total	3,314	4,109

8.1.2 Analytical plan

The regression analysis for these children differs from the approach used to investigate caregiver behavior or livelihoods. As the baseline values for the young children in the endline sample are not available, rather than including individual baseline outcomes on the right-hand side of regressions as employed in ANCOVA models, the regressions with observations on children use repeated cross-sectional observations in a DID approach. Regressions in which the sample is based on caregivers or households follow the ANCOVA models discussed elsewhere in this report.

The detailed regression framework utilized for the DID approach is described in Chapter 3.

[28] Table 8.1 indicates the children for whom anthropometry is studied. This sample is slightly larger than the sample of index children, some of whom have young siblings.

8.2 Average standard treatment effects

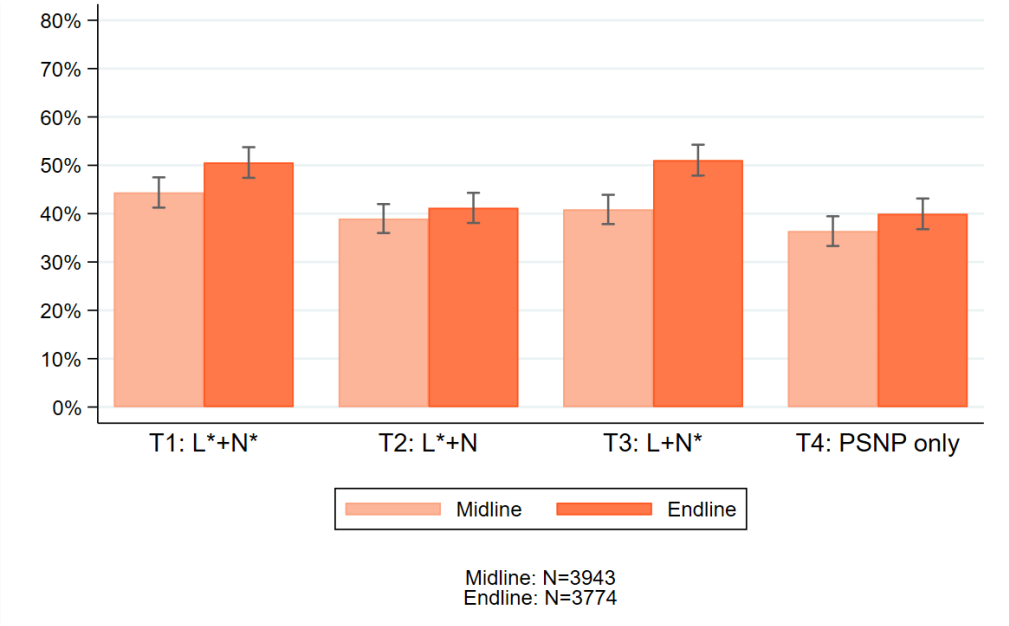
As indicated in Table 8.2, both N and N* treatments had substantial overall average standard treatment effects on access to health services with the N* TTC approach having substantially larger ASTE. These details are discussed further in the following section. The improvement in child health histories, including regular measurement, are more muted, with only the T1 arm showing significant improvements. Moreover, as summarized in Table 8.3, the improved service access did not have discernible effects on anthropometric measurements.

8.3 Service provision and utilization

8.3.1 Healthcare services

While the primary outcome variable of interest regarding N and N* treatments is anthropometric status, it is useful to first explore the provision of services that can mediate nutritional outcomes. As indicated in Table 8.4, the N* program resulted in an increased intensity of interaction of care providers with nutrition services. For example, the N* program increased the probability of meeting with a health extension worker (HEW) by 25 percent above the 40 percent rate of contact for the control (Figure 8.1).

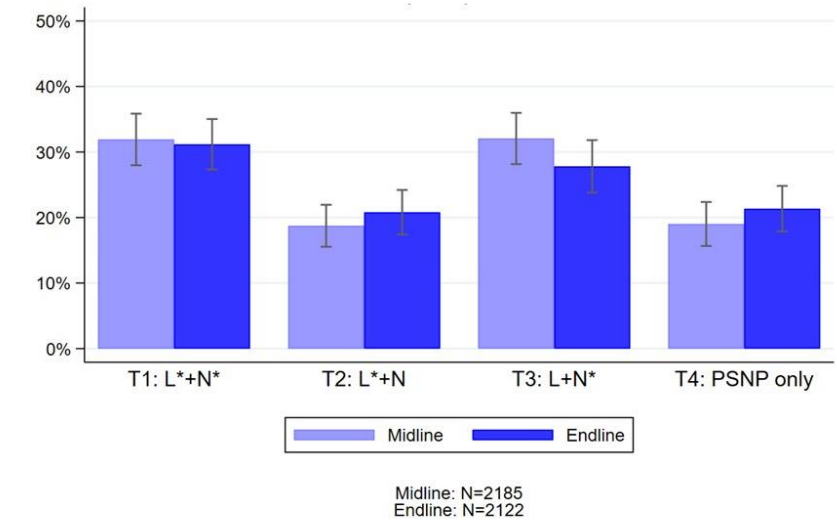
Figure 8.1: Primary female had contact with a HEW in the last 3 months



8.3.2 Child health outcomes

Table 8.5 and Figure 8.2 show that there was a small increase in the share of children younger than 24 months in the N* communities who were weighed in the last 3 months, roughly the same magnitude and proportional increase as the frequency of contact with the HEW at home. There was, however, no significant increase in the share who had their height or mid-upper-arm circumference (MUAC) measured. Nor was there an increase in any measurement in the T2 arm. As less than 40 percent of the age group in any arm of the study had their weight or MUAC recorded in the 3 months prior to the endline survey, the rate of measurement falls short of project goals.

Figure 8.2: Weight of index child (<24mo) was measured in past 6 months



Children who are underweight—below a Z score of weight for age of −2—are encouraged to participate in community-based participatory nutrition promotion (CPNP). Column 6 indicates that the share of children who attend CPNP sessions is significantly higher in N* communities. However, as the number of children who were indicated as underweight is small, few children in the sample attended CPNP sessions. Moreover, the correspondence between underweight status and participation is small; only 23 percent of those whose caregivers reported that the measurements indicate that their child was moderately or severely underweight had attended the sessions and half of caregivers who reported participation did not report that their child was underweight in the previous six months.

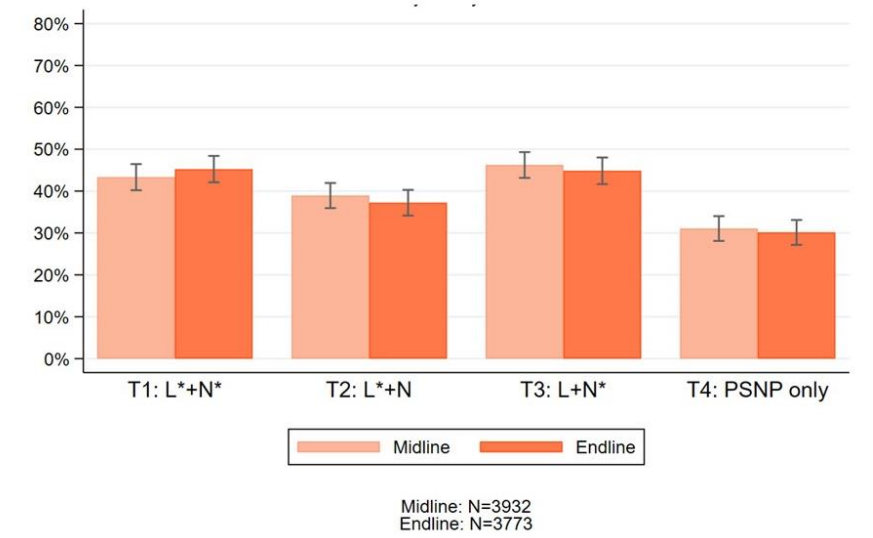
8.3.3 Comparison with observed effects at midline.

While the rate of contact with HEW was significantly higher in the N* communities compared with the control at midline, the difference more than doubled by the endline. Contact with a Health Development Army (HDA) worker also increased substantially in proportional terms; however, this was from a low base. While not all contact with HEWs or HDA workers occurred at home, home visits comprise a large share of the increase in total contacts in N* communities.

The proportional increases for less frequent services were even greater than for contact with the HEW. For example, attendance at BCC sessions more than doubled. While these sessions were part of TTC, the reported increase in attendance is greater than the increase of contact with HDA workers. In addition, participation in community-led total sanitation and hygiene events was 50 percent higher in the N* communities than in the control (Figure 8.3). Indeed, all eight of the services studied in Table 8.2 indicate that the N* services were provided at a statistically significantly higher rate than for the control; moreover, they were more prevalent than they were in communities where the regular nutrition (T2)

services were offered. This improvement in service delivery did not carry over in the area of child measurement. Indeed, the share of children weighed in T1 and T3 communities fell slightly from levels at midline while the share in the control actually increased from 19.0 percent to 21.5 percent in that period.

Figure 8.3: Attended a community-led total sanitation and hygiene event

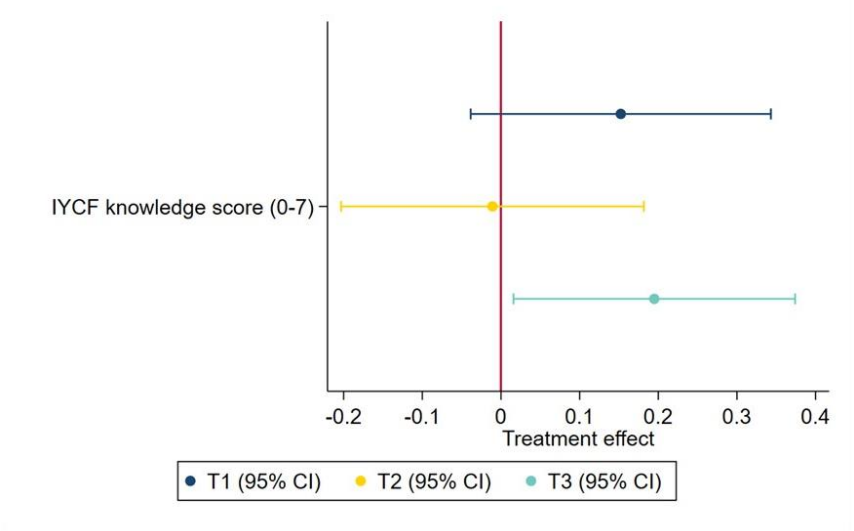


8.4 Infant and young child feeding (IYCF)

8.4.1 Women's nutrition knowledge

As indicated in Table 8.6, as well as Figure 8.4, for women in communities that were randomly assigned to receive the N* interventions (in T1 and T3), there is a statistically significant increase in the overall indicator of nutritional knowledge of 0.174, compared to a control group mean score of 3.93 (a 4 percent increase). This mean score serves in lieu of an ASTE for knowledge. However, the share of correct responses increased for only four of the seven questions, and only three of these increases were statistically significant at even marginal levels ($q < 0.1$). Of note is that less than a third of caregivers in all treatment arms understood the common problem with gruels given during weaning. While a higher share was aware of the importance of animal-source foods during weaning, this share was not associated with participation in any of the treatment arms.

Figure 8.4: Increase in female nutritional knowledge



8.4.2 Children's dietary diversity

Consistent with the absence of improved understanding of weaning, there was no significant increase in diet diversity of children among the pooled N* treatment arms (Table 8.7). However, there was a small impact on diet diversity of 0.199** for the T1 arm.[29] While the recommended minimum diet diversity (MDD) for a child of 6–24 months is to consume at least five of eight food groups daily, including breast milk (WHO and UNICEF, 2017), children in the SPIR sample receive foods from slightly more than two categories on average. 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 did not lead to any increase in this feeding practice.

8.4.3 Women's dietary diversity

In contrast with the results for child diet diversity, caregivers in the N* treatment did improve their own diets, again from a low base of adequacy, defined for women as consumption in the previous 24 hours from at least 5 of 10 food groups, as indicated in Table 8.8. Similar improvements were noted for the subsample of extremely poor households (Table B.8.2). However, the endline survey overlapped with

[29] The significant effect of T1 was not noted for the sample of extremely poor households (see Appendix Table B.8.1). There was, however, an improvement for the T2 in that subsample.

Lent. Thus, the regressions in Table 8.7 and Table 8.8 include a dummy variable defined as 1 if the household was Orthodox and the previous day was a fast day. As anticipated, very few Orthodox women consumed animal-source foods during a fast.[30] (See Table 8.9). Fasts led to a reduction of the already low probability of meat consumption for young children as well as a reduction in egg consumption. However, milk consumption by children was unaffected, 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.

Table 8.9. Consumption of animal-source foods in Orthodox households during the previous day

Category		Mean consumption	
		No fast day	Fast day
Women	Eggs	7.9%	1.2%
	Dairy	4.1%	1.3%
	Meat	13.4%	1.0%
Children	Eggs	5.5%	2.7%
	Dairy	7.2%	7.5%
	Meat	2.1%	0.4%

8.4.4 Child feeding

Table 8.10 provides further insight into IYCF practices in the SPIR sites. Looking first at the children who were 0–5 months at the time of the interview, there is no significant difference in the share exclusively breastfeeding associated with the SPIR programs (column 1). This share is relatively high and only declines slightly as the child ages. The next column indicates that while the N* communities had a slightly smaller share of children currently 6–15 months who continued exclusive breastfeeding up to 6 months at baseline, by endline this gap had closed. The timing of exclusive breastfeeding, however, is a somewhat ambiguous measure; 6 months of exclusive breastfeeding is recommended, but after that a child should receive a more diverse diet. The results in column 3 show that the project has not achieved success in limiting exclusive breastfeeding to an appropriate window. The next two columns of Table 8.10 reveal that less than half of the children 6–8 months were provided solid food (column 4) and even a third of the children 9–12 months had not yet received any foods other than breast milk (column 5).[31] These results are in keeping with the outcomes in diet diversity discussed above; the SPIR program had no significant impact on these important aspects of proper IYCF.

8.4.5 Comparison with observed effects at midline

While the diet diversity score for children is higher in the N* arms than in the control, this difference is smaller than it was at midline (1.25 compared with 1.81); in both cases, children in the T1 arm consumed more food groups than those in T3, albeit not significantly so. Moreover, the comparison with the control masks the fact that the number of food groups consumed in the control declined from 2.61 in the midline to 1.94 by the endline. In contrast, the number of food groups consumed by women in the control group increased from 2.94 to 3.93, and for women in N* communities the increase in diet diversity scores in the

[30] It is not known how many of the women who did not fast were pregnant at the time of the survey.
[31] This has also been noted in other studies in Ethiopia. See K. Hirvonen, W. Abdulazize, A.Laillou, V. Vinci, S. Chitekwe, and K. Baye, "Understanding Delays in the Introduction of Complementary Foods in Rural Ethiopia," *Maternal & Child Nutrition* (2021): e13247.

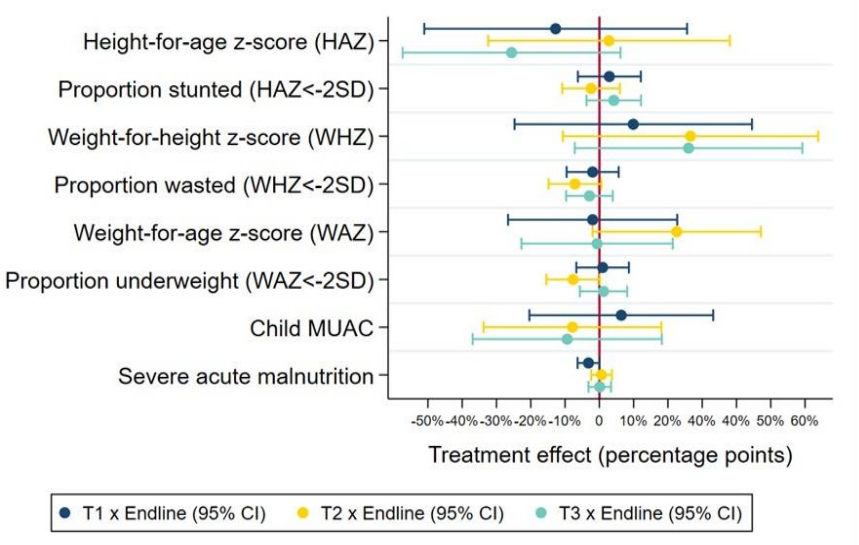
endline was larger than in the midline. The effects of fast days on women's consumption as well as the absence of an effect of fasting on milk consumption by children are roughly similar in the two rounds. As nutritional knowledge was not assessed in the midline, no comparison is possible.

8.5 Child anthropometry and development

8.5.1 Anthropometry

While the SPIR program has made inroads on service delivery and increased nutrition knowledge, this progress has not yet offset other barriers to proper nutrition. Table 8.11a reports the results for anthropometry for all children under 60 months of age. There were no improvements in height-for-age (HAZ) or stunting in N and N* treatment groups (see Figure 8.5). HAZ is a cumulative measure that is often deemed indicative of chronic undernutrition (stunting) in contrast with weight-for-height (WHZ) which is considered indicative of acute malnutrition (wasting). There is a small improvement in weight in the T2 treatment arm, leading to reduced wasting and reduced underweight. This, however, was not significant when considering multiple hypotheses nor was it observed in the more intensive N* arms. In as much as a share of the sample in Table 8.11a had already been weaned when SPIR services were scaled up and because determinants of nutrition are often age and stage dependent (Alderman and Headey, 2018), Table 8.11b reports anthropometry of children older than 5 months but younger than 24 months, who would have been born after SPIR was initiated. The table indicates an effect of T2 on weight for the younger children that is somewhat larger than the corresponding effect for all children under 60 months. These effects are not statistically significant at the 5 percent level after controlling for multiple hypothesis testing, but they are significant at the 10 percent level.

Figure 8.5: Program effects on height-for-age



8.5.2 Childcare Activities

The endline survey also included a subsection on activities related to a broader concept of child development than nutrition alone. The results are reported for females and males in Tables 8.12 and 8.13, with the total number of activities serving in lieu of ASTE. While the SPIR projects did not focus on the promotion of these aspects of childcare, there was a small increase in the number of activities performed by women in T2.

TTC communication does includes active and responsive feeding. As three of the activities are related to meal preparation or child feeding, they are worth highlighting. As indicated in columns 10 and 11, not only do most respondents not eat with their youngest child, but also the child is not always fed by the primary caregivers. Indeed, the primary female caregiver was far more likely to bathe her child than feed her. As parental involvement in feeding a child is no different among the treatment arms, it appears that responsive feeding in which a caregiver introduces and encourages new foods is not effective in IYCF messaging.

Although storytelling may foster vocabulary even prior to the child understanding the story, it is seldom practiced in low-income or low-education households throughout the world; the Ethiopian setting conforms to this generalization. Similarly, naming or drawing is very seldom practiced with these young children. More parents sing songs to their child, although this activity is still practiced by a minority of parents. The majority of respondents reported playing with their child in the previous three days, with a larger share of men than women indicating this activity.

8.5.3 Comparison with observations at midline

As key anthropometry measures of height-for-age and weight-for-age reflect a cumulative impact, these were not assessed at midline, given the short time for some services to have geared up. Thus, no comparisons are presented here.

The number of childcare activities declined substantially from those reported at midline. This was driven primarily by a decrease in the number of times a female reported that the child was taken outside the home from 0.61 at midline to only 0.27 for the control and 0.34 in N* communities. Males reported similar decreases from 0.63 at midline. This is likely attributable to the COVID-19 pandemic. Conversely, both males and females reported increasing the frequency with which they played with the child at home, with the unadjusted means reported by women and men of 0.6 and 0.7 more than twice those in the midline. A small share of this difference may reflect the ages of the children as the midline report indicates that there was a significant, but small, increase of the probability that a female caregiver played with their child in the last three days of 0.004 (.001). The corresponding increase per month of the child's age for male caregivers was 0.03 (0.004).

8.6 Food security

8.6.1 Food Insecurity Experience Scale

Table 8.14 indicates the experience of food insecurity using the Food Insecurity Experience Scale (FIES), which aggregates results from eight questions on access to food over the previous four weeks. The table also indicates the share who experienced severe food insecurity (defined as a score of 7 or 8) or moderate plus severe insecurity (defined as a score of 6 or greater).

As indicated, there are no significant differences in insecurity among treatment arms. This is also the case for the subset of extremely poor households (see Appendix Table B.8.3). However, in general the total score is higher in the treatment arms than in the control. The average in the severely poor subsample is 3.50 compared with 3.24 for the less poor. There is also a large regional difference, with respondents in Oromia reporting a FIES score of 4.38 compared to 2.58 in Amhara.

In as much as the FIES was included in the 2020 COVID-19 phone surveys, a comparison of the endline reports of food insecurity by household who were in the phone survey target sample and those who were not provides a measure of the difference between the experience of the population with phone access and those who do not have access. While those selected for the COVID-19 phone survey reported an average of FIES score of 3.0, those who were not had a score of 3.6.[32] Of the former group (phone survey sample), 13.8 percent were severely food insecure and 39.8 percent were moderately or severely food insecure. The corresponding measures for the general population were 19.8 percent and 48.0 percent. Thus, as expected, those with phone access seem to be less food insecure than those who do not own a phone.

8.6.2 Food gap

The female respondent was also asked in how many months the household had difficulty satisfying its food requirements. As with other food security indicators, there were no differences among treatment groups, with the coefficients and the standard errors of the difference between T1, T2, and T3 respectively and the control being 0.23 (0.20), 0.16 (0.21) and 0.07 (0.20). Siraro woreda reported the highest amount of food insecurity with nearly five months of difficulty meeting requirements. The difference between the number of months of food insecurity reported by the severely poor and the rest of the sample, however, was modest, with a difference only 0.16 (0.09) months. The coefficient of share of crops damaged by locusts was significant at 0.80 (0.25). Although mean damage for the sample was only 0.087, 84 percent of the respondents had no damage, implying that the mean, conditional on damage, was over half the crop. Thus, the households with locust damage had an extra 0.4 months of food insecurity on average.

8.6.3 Comparison of food gaps with baseline.

On average, the sample experienced 2.19 months of food insecurity in the 12 months prior to the baseline survey. This may reflect the fact that this year in the Ethiopian calendar (EC 2009-2010) was not a very difficult year for food security on average. The reported gap increased to 3.20 months of insecurity in the year prior to the endline with 9 of the 13 woredas indicating an increase of food insecurity. Food gaps were not measured at midline.

8.7 Conclusion

At midline, it was clear that the N* —and to a lesser degree the N— interventions had the capacity to increase exposure to health services including BCC, food demonstrations, and WASH. This service

[32] The FIES reported in the endline survey cannot be directly compared with the reports in the phone surveys as all respondents in the endline were women while the majority of respondents in the phone surveys were men. Moreover, the FIES was not included in the baseline or midline survey, so no direct comparisons of trends are possible.

delivery has continued to 2021 despite the intervening COVID-19 strain on healthcare resources and the decreased mobility that the pandemic imposed. These gains are summarized in terms of average standardized treatment effects reported in Table 8.2. But this is almost literally a “glass half-filled” story; for no indicator of access to health services studied shows more than half the target population participating. Moreover, the BCC has not been able to improve key measures of IYCF, such as the age at which semi-solid or solid foods are introduced or child diet diversity. Since proper complementary feeding is an essential element of nutritional care, this barrier likely contributes to the stagnating stunting rate. As summarized in Table 8.3, we cannot reject that the average standardized treatment effects for anthropomorphic outcomes are zero. The SPIR project has, however, made modest inroads in responding to underweight when it is identified. But, again, with child weighing apparently infrequent, this improved service delivery does not fully cover the eligible population. Thus, identifying the gaps in coverage as well as improving the messaging on weaning appear to be ways that the initial progress in intensified nutritional service delivery can achieve greater progress in improving nutritional outcomes.

Table 8.2: Summary table of health outcomes

	(1) Access to health services	(2) Child health history
T1	0.355***	0.136**
	(0.054)	(0.066)
T2	0.086*	-0.047
	(0.048)	(0.070)
T3	0.370***	0.006
	(0.057)	(0.068)
Test: T1 = T2	0.000	0.020
Test: T2 = T3	0.000	0.509
Test: T1 = T3	0.822	0.067
N	3,775	1,058

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 8.3: Summary table of anthropometrics

	(1) HAZ, WAZ, WHZ, MUAC	(2) Binary indicators of stunted, wasted, underweight, malnourished
T1 x Endline	-0.010	-0.038
	(0.063)	(0.056)
T2 x Endline	0.066	-0.089
	(0.063)	(0.058)
T3 x Endline	-0.008	0.004
	(0.061)	(0.057)
Test: T1 x Endline = T2 x Endline	0.201	0.372
Test: T2 x Endline = T3 x Endline	0.192	0.111
Test: T1 x Endline = T3 x Endline	0.982	0.454
N	7,318	7,321

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 8.4: Access to health services

	(1) Has had contact with the HEW in past 3 months	(2) Has been visited by the HEW at home in past 3 months	(3) Has had contact with the HDA/leader in past 3 months	(4) Has been visited by the HDA/leader at home in past 3 months	(5) Has attended a food demonstration in her community in last 3 months	(6) Has attended a BCC session in past 3 months	(7) Has attended a community-led total sanitation and hygiene (CLTSH) event	(8) Has participated in an open defecation free (ODF) event
T1	0.110***	0.079***	0.082***	0.062***	0.135***	0.181***	0.160***	0.130***
	(0.028)	(0.024)	(0.020)	(0.015)	(0.024)	(0.029)	(0.039)	(0.041)
	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]
T2	0.003	0.030	0.007	0.022*	0.021	0.041*	0.062*	0.054
	(0.033)	(0.024)	(0.015)	(0.011)	(0.017)	(0.024)	(0.035)	(0.038)
	[0.917]	[0.297]	[0.736]	[0.220]	[0.297]	[0.220]	[0.220]	[0.297]
T3	0.103***	0.073***	0.093***	0.073***	0.160***	0.179***	0.140***	0.138***
	(0.030)	(0.024)	(0.019)	(0.015)	(0.023)	(0.026)	(0.037)	(0.040)
	[0.001]	[0.003]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]
Pooled effect of N*: T1 or T3	0.107***	0.076***	0.087***	0.068***	0.147***	0.180***	0.150***	0.134***
	(0.025)	(0.021)	(0.016)	(0.012)	(0.018)	(0.022)	(0.032)	(0.035)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Test: T1 = T2	(0.002)	(0.054)	(0.000)	(0.011)	(0.000)	(0.000)	(0.014)	(0.066)
	[0.004]	[0.061]	[0.001]	[0.017]	[0.000]	[0.000]	[0.018]	[0.066]
Test: T2 = T3	(0.007)	(0.081)	(0.000)	(0.001)	(0.000)	(0.000)	(0.038)	(0.035)
	[0.010]	[0.081]	[0.000]	[0.002]	[0.000]	[0.000]	[0.044]	[0.044]
Test: T1 = T3	(0.808)	(0.814)	(0.647)	(0.529)	(0.412)	(0.931)	(0.608)	(0.849)
	[0.931]	[0.931]	[0.931]	[0.931]	[0.931]	[0.931]	[0.931]	[0.931]
Mean of control (T4)	0.400	0.163	0.099	0.038	0.061	0.108	0.301	0.334
N	3,774	3,773	3,763	3,762	3,771	3,773	3,773	3,774

Note: Estimates from the DFSA SPIR endline survey sample, estimated as an ANCOVA model at the household level. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.5: Child health history

	(1) Received dose of vitamin A in past 6 months (6–23 months)	(2) Child's weight was measured in past 3 months (<24 months)	(3) Child's height was measured in past 3 months (<24 months)	(4) Child's MUAC was measured in past 3 months (6–23 months)	(5) Child identified as severely or moderately underweight in past 6 months (<24 months)	(6) Caregiver and child participated in a 2-week cooking demonstration and feeding session (CPNP) (<24 months)	(7) Child identified as severely malnourished in past 6 months (6–23 months)	(8) Received a specific food or milk as treatment for severe malnutrition (6–23 months)
T1	0.023	0.096**	0.051*	0.030	0.033	0.081***	-0.030	0.054
	(0.053)	(0.041)	(0.027)	(0.050)	(0.029)	(0.023)	(0.071)	(0.173)
	[0.757]	[0.077]	[0.153]	[0.757]	[0.513]	[0.004]	[0.757]	[0.757]
T2	-0.012	-0.037	-0.013	-0.096**	0.008	0.012	-0.005	-0.070
	(0.051)	(0.032)	(0.023)	(0.045)	(0.029)	(0.017)	(0.067)	(0.203)
	[0.937]	[0.937]	[0.937]	[0.272]	[0.937]	[0.937]	[0.945]	[0.937]
T3	0.052	0.060	0.009	-0.012	0.037	0.048**	0.067	-0.346*
	(0.053)	(0.038)	(0.024)	(0.046)	(0.031)	(0.021)	(0.085)	(0.183)
	[0.521]	[0.320]	[0.788]	[0.788]	[0.467]	[0.202]	[0.577]	[0.256]
Pooled effect of N*: T1 or T3	0.037	0.078**	0.030	0.009	0.035	0.065***	0.018	-0.146
	(0.044)	(0.032)	(0.021)	(0.040)	(0.026)	(0.018)	(0.064)	(0.157)
	[0.537]	[0.061]	[0.357]	[0.828]	[0.357]	[0.003]	[0.828]	[0.537]
Test: T1 = T2	(0.550)	(0.002)	(0.017)	(0.021)	(0.399)	(0.005)	(0.738)	(0.559)
	[0.639]	[0.014]	[0.042]	[0.042]	[0.639]	[0.021]	[0.738]	[0.639]
Test: T2 = T3	(0.259)	(0.016)	(0.368)	(0.082)	(0.348)	(0.128)	(0.405)	(0.203)
	[0.405]	[0.128]	[0.405]	[0.329]	[0.405]	[0.341]	[0.405]	[0.405]
Test: T1 = T3	(0.612)	(0.434)	(0.123)	(0.435)	(0.889)	(0.214)	(0.282)	(0.021)
	[0.700]	[0.580]	[0.494]	[0.580]	[0.889]	[0.564]	[0.564]	[0.167]
Mean of control (T4)	0.449	0.215	0.102	0.343	0.135	0.038	0.262	0.688
N	709	1,046	1,044	714	1,028	1,058	246	63

Note: Estimates from the DFSA SPIR endline survey sample, estimated as an ANCOVA model at the household level. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.6: Women's IYCF knowledge

	(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
T1	0.152	0.023	0.022	-0.002	0.050*	-0.018	0.033	0.047*
	(0.097)	(0.016)	(0.021)	(0.034)	(0.029)	(0.025)	(0.023)	(0.026)
		[0.285]	[0.415]	[0.953]	[0.285]	[0.548]	[0.285]	[0.285]
T2	-0.011	0.002	-0.003	-0.048	0.030	-0.044*	0.014	0.040
	(0.098)	(0.018)	(0.022)	(0.033)	(0.033)	(0.025)	(0.025)	(0.026)
		[0.904]	[0.904]	[0.341]	[0.621]	[0.341]	[0.818]	[0.341]
T3	0.195**	0.039**	0.032*	-0.059*	0.086***	0.000	0.018	0.078***
	(0.091)	(0.017)	(0.019)	(0.031)	(0.029)	(0.026)	(0.023)	(0.027)
		[0.052]	[0.134]	[0.105]	[0.014]	[0.996]	[0.492]	[0.014]
Pooled effect of N*: T1 or T3	0.174**	0.031**	0.027	-0.031	0.068***	-0.009	0.026	0.062***
	(0.085)	(0.015)	(0.018)	(0.029)	(0.025)	(0.022)	(0.020)	(0.023)
		[0.092]	[0.227]	[0.335]	[0.027]	[0.688]	[0.296]	[0.027]
Test: T1 = T2	(0.072)	(0.191)	(0.234)	(0.161)	(0.546)	(0.275)	(0.397)	(0.799)
		[0.482]	[0.482]	[0.482]	[0.637]	[0.482]	[0.556]	[0.799]
Test: T2 = T3	(0.015)	(0.025)	(0.069)	(0.707)	(0.082)	(0.088)	(0.833)	(0.161)
		[0.154]	[0.154]	[0.825]	[0.154]	[0.154]	[0.833]	[0.226]
Test: T1 = T3	(0.605)	(0.255)	(0.557)	(0.067)	(0.199)	(0.466)	(0.481)	(0.244)
		[0.447]	[0.557]	[0.447]	[0.447]	[0.557]	[0.557]	[0.447]
Mean of control (T4)	3.929	0.872	0.834	0.272	0.602	0.287	0.693	0.371
N	3,704	3,703	3,704	3,704	3,704	3,702	3,703	3,702

Note: Estimates from the DFSA SPIR endline survey sample, estimated as an ANCOVA model at the household level. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.7: Children's dietary diversity:

Age 6–23 months

	(1) Number of food groups (of 8) consumed	(2) Child meets the minimum meal frequency (MMF)	(3) Child consumed grains, roots or tubers	(4) Child consumed legumes or nuts	(5) Child consumed dairy	(6) Child consumed fish or meat	(7) Child consumed eggs	(8) Child consumed vitamin A–rich fruits or vegetables	(9) Child consumed other fruits or vegetables
Child age in months	0.022***	0.019***	0.026***	0.010***	0.002***	0.000**	0.001***	0.002***	0.000
	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male child	0.022	0.001	0.016	-0.004	0.005	-0.005	0.007	-0.003	-0.005
	(0.027)	(0.019)	(0.012)	(0.011)	(0.010)	(0.003)	(0.006)	(0.005)	(0.003)
T1	-0.016	-0.004	0.001	-0.006	0.003	-0.001	-0.004	-0.003	-0.003
	(0.020)	(0.022)	(0.013)	(0.008)	(0.005)	(0.001)	(0.003)	(0.004)	(0.002)
T2	-0.030	-0.011	-0.011	-0.012	-0.002	-0.001	-0.002	-0.005	-0.002
	(0.022)	(0.024)	(0.015)	(0.008)	(0.005)	(0.002)	(0.003)	(0.004)	(0.002)
T3	-0.034	-0.015	-0.015	-0.009	-0.002	-0.002	-0.001	-0.004	-0.001
	(0.022)	(0.022)	(0.015)	(0.008)	(0.005)	(0.001)	(0.003)	(0.004)	(0.002)
Endline	0.193**	0.317***	0.101***	-0.010	0.066**	-0.005	-0.003	-0.029**	-0.011
	(0.082)	(0.055)	(0.037)	(0.024)	(0.032)	(0.008)	(0.012)	(0.013)	(0.010)
T1 x Endline	0.199**	-0.003	0.012	0.000	0.064	-0.002	0.034*	0.037**	0.005
	(0.098)	(0.073)	(0.047)	(0.034)	(0.039)	(0.010)	(0.017)	(0.019)	(0.011)
		[0.993]	[0.993]	[0.993]	[0.278]	[0.993]	[0.212]	[0.212]	[0.993]
T2 x Endline	0.170	0.001	0.048	0.006	0.037	0.002	0.028	0.030	-0.002
	(0.103)	(0.073)	(0.046)	(0.036)	(0.038)	(0.011)	(0.020)	(0.019)	(0.011)
		[0.993]	[0.665]	[0.993]	[0.665]	[0.993]	[0.665]	[0.665]	[0.993]
T3 x Endline	0.052	-0.049	-0.005	-0.010	0.039	0.003	0.019	0.015	-0.009
	(0.110)	(0.082)	(0.049)	(0.037)	(0.041)	(0.011)	(0.020)	(0.020)	(0.012)
		[0.875]	[0.914]	[0.910]	[0.875]	[0.910]	[0.875]	[0.875]	[0.875]
Pooled T1 x Endline and T3 x Endline	0.125	-0.026	0.004	-0.005	0.052	0.001	0.027*	0.026	-0.002
	(0.093)	(0.066)	(0.042)	(0.030)	(0.036)	(0.009)	(0.016)	(0.016)	(0.011)
		[0.947]	[0.947]	[0.947]	[0.399]	[0.947]	[0.399]	[0.399]	[0.947]
Test: T1xEndline = T2xEndline	(0.738)	(0.959)	(0.389)	(0.884)	(0.427)	(0.684)	(0.774)	(0.696)	(0.261)
		[0.959]	[0.959]	[0.959]	[0.959]	[0.959]	[0.959]	[0.959]	[0.959]
Test: T2xEndline = T3xEndline	(0.238)	(0.536)	(0.225)	(0.701)	(0.957)	(0.908)	(0.701)	(0.489)	(0.410)
		[0.935]	[0.935]	[0.935]	[0.957]	[0.957]	[0.935]	[0.935]	[0.935]
	(0.123)	(0.570)	(0.692)	(0.799)	(0.500)	(0.595)	(0.459)	(0.291)	(0.118)

Test: T1xEndline = T3xEndline		[0.791]	[0.791]	[0.799]	[0.791]	[0.791]	[0.791]	[0.791]	[0.791]
Mean of control (T4)	1.941	0.445	0.729	0.267	0.094	0.018	0.048	0.058	0.023
N	4,341	2,146	4,341	4,341	4,341	4,341	4,341	4,341	4,341

Note: Estimates from the DFSA SPIR endline survey sample, estimated as a difference-in-difference model at the child level. The model included a dummy variable defined as 1 if the household was Orthodox and the previous day was a fast day. Standard errors (in parentheses below treatment effects in panels 1 and 2) are clustered at the kebele level. All models control for woreda-level fixed effects and the cluster mean of the baseline value of the outcome. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.8: Women's dietary diversity: Full sample

	(1) Women's dietary diversity score (1–10)	(2) Met minimum dietary diversity for women (MDD-W)	(3) Primary female consumed tubers and grains	(4) Primary female consumed pulses	(5) Primary female consumed nuts and seeds	(6) Primary female consumed dairy	(7) Primary female consumed meat, fish, poultry	(8) Primary female consumed eggs	(9) Primary female consumed green leafy veg-etables	(10) Primary female consumed vitamin Arich fruits and veg-etables	(11) Primary female consumed other veg-etables	(12) Primary female consumed other fruits
T1	0.175*	0.047***	-0.016	-0.016	0.074**	0.009	0.017	0.011	0.033**	0.014	0.054	-0.001
	(0.090)	(0.014)	(0.012)	(0.024)	(0.032)	(0.017)	(0.013)	(0.012)	(0.014)	(0.018)	(0.039)	(0.010)
		[0.015]	[0.354]	[0.610]	[0.084]	[0.659]	[0.354]	[0.509]	[0.083]	[0.584]	[0.354]	[0.956]
T2	0.024	0.014	-0.017	-0.053**	0.024	0.008	0.014	0.001	0.016	0.002	0.022	0.005
	(0.084)	(0.015)	(0.012)	(0.021)	(0.029)	(0.019)	(0.011)	(0.010)	(0.013)	(0.019)	(0.038)	(0.011)
		[0.737]	[0.639]	[0.128]	[0.737]	[0.836]	[0.639]	[0.939]	[0.639]	[0.939]	[0.836]	[0.836]
T3	0.253***	0.045**	-0.006	-0.027	0.068**	0.032	0.011	0.018	0.017	0.024	0.093***	0.030**
	(0.089)	(0.017)	(0.009)	(0.023)	(0.027)	(0.021)	(0.013)	(0.012)	(0.014)	(0.019)	(0.034)	(0.013)
		[0.045]	[0.480]	[0.282]	[0.045]	[0.237]	[0.412]	[0.237]	[0.282]	[0.282]	[0.045]	[0.053]
Pooled effect of N*: T1 or T3	0.214***	0.046***	-0.011	-0.022	0.071***	0.020	0.014	0.015	0.025**	0.019	0.073**	0.015
	(0.076)	(0.013)	(0.009)	(0.020)	(0.025)	(0.017)	(0.011)	(0.010)	(0.011)	(0.016)	(0.032)	(0.010)
		[0.007]	[0.263]	[0.267]	[0.032]	[0.263]	[0.263]	[0.249]	[0.081]	[0.263]	[0.078]	[0.249]
Test: T1 = T2	(0.101)	(0.034)	(0.914)	(0.147)	(0.125)	(0.939)	(0.822)	(0.342)	(0.234)	(0.518)	(0.435)	(0.601)
		[0.370]	[0.939]	[0.540]	[0.540]	[0.939]	[0.939]	[0.753]	[0.642]	[0.814]	[0.798]	[0.827]
Test: T2 = T3	(0.015)	(0.097)	(0.318)	(0.268)	(0.107)	(0.211)	(0.796)	(0.141)	(0.955)	(0.287)	(0.062)	(0.052)
		[0.294]	[0.389]	[0.389]	[0.294]	[0.387]	[0.875]	[0.310]	[0.955]	[0.389]	[0.294]	[0.294]
Test: T1 = T3	(0.415)	(0.919)	(0.381)	(0.681)	(0.841)	(0.188)	(0.672)	(0.627)	(0.274)	(0.631)	(0.306)	(0.011)
		[0.919]	[0.832]	[0.832]	[0.919]	[0.832]	[0.832]	[0.832]	[0.832]	[0.832]	[0.832]	[0.117]
Mean of	2.637	0.052	0.981	0.579	0.195	0.119	0.040	0.040	0.047	0.088	0.512	0.036

control (T4)												
N	3,704	3,704	3,704	3,704	3,704	3,704	3,704	3,704	3,704	3,704	3,704	3,704

Note: Estimates from the DFSA SPIR endline survey sample, estimated as an ANCOVA model at the household level. The model included a dummy variable defined as 1 if the household was Orthodox and the previous day was a fast day. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.10: Infant and young child feeding (IYCF) practices

	(1) Infants 0–5 months who are fed exclusively with breast milk	(2) Children 6–15 months: stopped exclusive breastfeeding at 6 months or later	(3) Children 6–15 months: stopped exclusive breastfeeding in month 6, 7 or 8	(4) Children 6–8 months of age who receive solid, semi-solid or soft foods	(5) Children 9–12 months who receive solid, semi-solid or soft foods
Child age in months	-0.062***				
	(0.009)				
Male child	0.007	-0.024	-0.043**	0.008	-0.024
	(0.032)	(0.018)	(0.021)	(0.044)	(0.035)
T1	-0.048	-0.088*	-0.087*	-0.056	0.033
	(0.074)	(0.045)	(0.048)	(0.092)	(0.064)
T2	0.104	-0.091*	-0.084	0.043	0.018
	(0.070)	(0.050)	(0.053)	(0.097)	(0.066)
T3	0.004	-0.088*	-0.091*	-0.077	0.044
	(0.072)	(0.045)	(0.049)	(0.088)	(0.061)
Endline	0.110*	0.079**	0.035	-0.071	0.001
	(0.064)	(0.037)	(0.042)	(0.113)	(0.071)
T1 x Endline	-0.062	0.116*	0.073	0.095	0.005
	(0.093)	(0.060)	(0.068)	(0.142)	(0.103)
	[0.841]	[0.393]	[0.360]	[0.945]	[0.945]
T2 x Endline	-0.140	0.111*	0.095	0.010	0.014
	(0.091)	(0.063)	(0.065)	(0.142)	(0.102)
	[0.205]	[0.205]	[0.620]	[0.620]	[0.913]
T3 x Endline	-0.083	0.094	0.049	0.097	0.012
	(0.090)	(0.057)	(0.068)	(0.143)	(0.106)
	[0.445]	[0.002]	[0.861]	[0.001]	[0.025]
Pooled T1 x Endline and T3 x Endline	-0.073	0.105**	0.061	0.096	0.008
	(0.079)	(0.049)	(0.057)	(0.128)	(0.089)

	[0.568]	[0.083]	[0.568]	[0.568]	[0.927]
Test: T1xEndline = T2xEndline	(0.389)	(0.943)	(0.760)	(0.503)	(0.927)
	[0.839]	[0.943]	[0.943]	[0.839]	[0.943]
Test: T2xEndline = T3xEndline	(0.516)	(0.804)	(0.531)	(0.475)	(0.981)
	[0.981]	[0.981]	[0.884]	[0.981]	[0.981]
Test: T1xEndline = T3xEndline	(0.818)	(0.734)	(0.752)	(0.987)	(0.950)
	[0.987]	[0.987]	[0.987]	[0.987]	[0.987]
Mean of control (T4)	0.722	0.806	0.798	0.455	0.653
N	744	1,513	1,318	458	691

Note: Estimates from the DFSA SPIR endline survey sample, estimated as a difference-in-difference model at the child level. Standard errors (in parentheses below treatment effects in panels 1 and 2) are clustered at the kebele level. All models control for woreda-level fixed effects and the cluster mean of the baseline value of the outcome. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.11a: Anthropometrics: Children up to 60 months of age

	(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)
Child age in months	-0.020***	0.003***	-0.006***	-0.001***	-0.017***	0.001***	0.038***	-0.003***
	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Male child	-0.491***	0.129***	-0.372***	0.061***	-0.513***	0.139***	-0.015	-0.002
	(0.044)	(0.012)	(0.032)	(0.007)	(0.031)	(0.010)	(0.029)	(0.005)
T1	0.065	-0.008	-0.044	0.005	0.023	-0.002	-0.005	0.014*
	(0.091)	(0.021)	(0.086)	(0.017)	(0.061)	(0.018)	(0.069)	(0.008)
T2	-0.017	0.015	-0.116	0.029	-0.088	0.037**	-0.009	-0.005
	(0.084)	(0.019)	(0.092)	(0.020)	(0.055)	(0.017)	(0.067)	(0.008)
T3	0.022	0.007	-0.143*	0.018	-0.050	0.018	-0.002	0.002
	(0.083)	(0.019)	(0.081)	(0.015)	(0.051)	(0.016)	(0.072)	(0.008)
Endline	0.125	-0.000	0.037	0.002	0.153	0.004	0.248***	0.007
	(0.139)	(0.033)	(0.140)	(0.027)	(0.097)	(0.029)	(0.094)	(0.011)
T1 x Endline	-0.144	0.034	0.082	-0.018	-0.040	0.013	0.044	-0.031*
	(0.196)	(0.047)	(0.176)	(0.039)	(0.126)	(0.039)	(0.140)	(0.016)
	[0.753]	[0.753]	[0.753]	[0.753]	[0.753]	[0.753]	[0.753]	[0.469]
T2 x Endline	0.032	-0.022	0.242	-0.068*	0.211*	-0.071*	-0.094	0.007
	(0.180)	(0.043)	(0.190)	(0.039)	(0.126)	(0.039)	(0.136)	(0.015)
	[0.860]	[0.762]	[0.406]	[0.253]	[0.253]	[0.253]	[0.762]	[0.762]

T3 x Endline	-0.235	0.042	0.233	-0.028	-0.011	0.014	-0.118	0.001
	(0.160)	(0.040)	(0.167)	(0.034)	(0.112)	(0.034)	(0.143)	(0.017)
	[0.655]	[0.655]	[0.655]	[0.655]	[0.953]	[0.914]	[0.655]	[0.953]
Pooled T1 x Endline and T3 x Endline	-0.190	0.038	0.158	-0.023	-0.026	0.014	-0.037	-0.015
	(0.160)	(0.038)	(0.157)	(0.032)	(0.109)	(0.033)	(0.120)	(0.014)
	[0.656]	[0.656]	[0.656]	[0.755]	[0.815]	[0.815]	[0.815]	[0.656]
Test: T1xEndline = T2xEndline	(0.327)	(0.199)	(0.336)	(0.203)	(0.025)	(0.024)	(0.330)	(0.014)
	[0.336]	[0.324]	[0.336]	[0.324]	[0.068]	[0.068]	[0.336]	[0.068]
Test: T2xEndline = T3xEndline	(0.060)	(0.083)	(0.953)	(0.255)	(0.022)	(0.010)	(0.870)	(0.720)
	[0.161]	[0.166]	[0.953]	[0.408]	[0.088]	[0.078]	[0.953]	[0.953]
Test: T1xEndline = T3xEndline	(0.575)	(0.848)	(0.290)	(0.770)	(0.770)	(0.982)	(0.277)	(0.055)
	[0.969]	[0.969]	[0.772]	[0.969]	[0.969]	[0.982]	[0.772]	[0.442]
Mean of control (T4)	-1.416	0.362	-0.363	0.114	-1.085	0.221	13.356	0.058
N	7,162	7,162	7,146	7,146	7,262	7,262	7,052	7,056

Notes: Estimates from the DFSA SPIR endline survey sample, estimated as a difference-in-difference model at the child level. Standard errors (in parentheses below treatment effects in panels 1 and 2) are clustered at the kebele level. All models control for woreda-level fixed effects and the cluster mean of the baseline value of the outcome. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.11b: Anthropometrics: Children 6–23 months

	(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)
Child age in months	-0.081***	0.017***	-0.007	-0.003**	-0.034***	0.005***	0.037***	-0.005***
	(0.007)	(0.002)	(0.005)	(0.001)	(0.005)	(0.002)	(0.005)	(0.001)
Male child	-0.609***	0.153***	-0.345***	0.067***	-0.540***	0.155***	-0.022	-0.005
	(0.068)	(0.019)	(0.054)	(0.013)	(0.056)	(0.017)	(0.051)	(0.009)
T1	0.013	-0.002	-0.038	0.002	-0.012	0.009	0.014	0.001
	(0.080)	(0.019)	(0.066)	(0.017)	(0.047)	(0.015)	(0.058)	(0.007)
T2	0.023	0.012	-0.118	0.036*	-0.066	0.028*	-0.005	0.001
	(0.082)	(0.019)	(0.076)	(0.018)	(0.048)	(0.016)	(0.062)	(0.008)
T3	0.018	0.017	-0.112*	0.018	-0.062	0.026*	-0.026	0.009
	(0.072)	(0.017)	(0.060)	(0.014)	(0.044)	(0.014)	(0.061)	(0.008)
Endline	-0.106	0.002	-0.145	0.014	-0.125	0.058*	0.040	0.017
	(0.169)	(0.038)	(0.152)	(0.036)	(0.107)	(0.033)	(0.112)	(0.022)
T1 x Endline	-0.160	0.083	0.234	-0.010	0.043	0.003	0.146	-0.016

	(0.267)	(0.059)	(0.208)	(0.056)	(0.162)	(0.049)	(0.179)	(0.030)
	[0.937]	[0.937]	[0.937]	[0.956]	[0.956]	[0.956]	[0.937]	[0.937]
T2 x Endline	-0.021	0.025	0.492**	-0.104*	0.364**	-0.115**	0.143	-0.012
	(0.246)	(0.058)	(0.235)	(0.054)	(0.161)	(0.051)	(0.162)	(0.028)
	[0.933]	[0.758]	[0.100]	[0.114]	[0.100]	[0.100]	[0.609]	[0.758]
T3 x Endline	-0.223	0.026	0.373*	-0.052	0.193	-0.050	0.087	-0.031
	(0.214)	(0.053)	(0.199)	(0.051)	(0.154)	(0.054)	(0.220)	(0.032)
	[0.478]	[0.692]	[0.478]	[0.478]	[0.478]	[0.478]	[0.692]	[0.478]
Pooled T1 x Endline and T3 x Endline	-0.192	0.054	0.303*	-0.031	0.118	-0.024	0.117	-0.023
	(0.208)	(0.048)	(0.179)	(0.046)	(0.135)	(0.043)	(0.162)	(0.026)
	[0.565]	[0.565]	[0.565]	[0.565]	[0.565]	[0.587]	[0.565]	[0.565]
Test: T1xEndline = T2xEndline	(0.608)	(0.353)	(0.259)	(0.113)	(0.060)	(0.028)	(0.983)	(0.889)
	[0.810]	[0.564]	[0.519]	[0.301]	[0.241]	[0.222]	[0.983]	[0.983]
Test: T2xEndline = T3xEndline	(0.363)	(0.998)	(0.591)	(0.346)	(0.298)	(0.266)	(0.804)	(0.531)
	[0.727]	[0.998]	[0.788]	[0.727]	[0.727]	[0.727]	[0.919]	[0.788]
Test: T1xEndline = T3xEndline	(0.800)	(0.320)	(0.474)	(0.460)	(0.366)	(0.350)	(0.802)	(0.637)
	[0.802]	[0.759]	[0.759]	[0.759]	[0.759]	[0.759]	[0.802]	[0.802]
Mean of control (T4)	-1.457	0.371	-0.440	0.127	-1.150	0.236	13.282	0.042
N	2,557	2,557	2,557	2,557	2,589	2,589	2,600	2,602

Note: Estimates from the DFSA SPIR endline survey sample, estimated as a difference-in-difference model at the child level. Standard errors (in parentheses below treatment effects in panels 1 and 2) are clustered at the kebele level. All models control for woreda-level fixed effects and the cluster mean of the baseline value of the outcome. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.12: Female's childcare activities in past 3 days

	(1) Number of activities that primary female carried out (0–10)	(2) Told stories to index child	(3) Sang songs to or with index child	(4) Took index child outside the home	(5) Played with index child	(6) Named, counted or drew things with or for index child	(7) Gave index child a bath	(8) Cared for the index child when they were sick	(9) Prepared food for index child (if not exclusively breastfed)	(10) Physically fed index child (if not exclusively breastfed)	(11) Ate a meal together with index child (if not exclusively breastfed)
T1	0.160	0.006	0.030	0.036	0.036	0.023	0.007	0.007	0.037	-0.076	-0.034
	(0.150)	(0.014)	(0.033)	(0.045)	(0.044)	(0.026)	(0.029)	(0.025)	(0.052)	(0.063)	(0.047)
		[0.810]	[0.684]	[0.684]	[0.684]	[0.684]	[0.810]	[0.810]	[0.684]	[0.684]	[0.684]
T2	0.408***	0.003	0.057**	0.067	0.045	0.010	0.056**	0.041	0.062	0.028	-0.034

	(0.145)	(0.010)	(0.027)	(0.043)	(0.041)	(0.027)	(0.025)	(0.026)	(0.064)	(0.055)	(0.050)
		[0.782]	[0.187]	[0.308]	[0.533]	[0.782]	[0.187]	[0.308]	[0.549]	[0.768]	[0.719]
T3	0.209	-0.004	0.069**	0.108**	-0.013	0.002	0.052*	0.023	-0.044	-0.006	-0.040
	(0.147)	(0.009)	(0.031)	(0.044)	(0.043)	(0.026)	(0.026)	(0.031)	(0.053)	(0.063)	(0.055)
		[0.914]	[0.138]	[0.138]	[0.947]	[0.947]	[0.168]	[0.782]	[0.782]	[0.947]	[0.782]
Pooled effect of N*: T1 or T3	0.185	0.001	0.049*	0.072*	0.011	0.012	0.029	0.015	-0.003	-0.041	-0.037
	(0.129)	(0.010)	(0.027)	(0.040)	(0.038)	(0.024)	(0.024)	(0.024)	(0.044)	(0.053)	(0.044)
		[0.944]	[0.356]	[0.356]	[0.944]	[0.862]	[0.763]	[0.862]	[0.944]	[0.862]	[0.862]
Test: T1 = T2	(0.087)	(0.823)	(0.401)	(0.455)	(0.815)	(0.586)	(0.070)	(0.190)	(0.717)	(0.098)	(0.989)
		[0.914]	[0.910]	[0.910]	[0.914]	[0.914]	[0.490]	[0.635]	[0.914]	[0.490]	[0.989]
Test: T2 = T3	(0.155)	(0.448)	(0.671)	(0.291)	(0.145)	(0.733)	(0.856)	(0.557)	(0.130)	(0.545)	(0.913)
		[0.913]	[0.913]	[0.913]	[0.726]	[0.913]	[0.913]	[0.913]	[0.726]	[0.913]	[0.913]
Test: T1 = T3	(0.739)	(0.428)	(0.257)	(0.086)	(0.261)	(0.346)	(0.100)	(0.590)	(0.164)	(0.295)	(0.917)
		[0.536]	[0.492]	[0.492]	[0.492]	[0.495]	[0.492]	[0.655]	[0.492]	[0.492]	[0.917]
Mean of control (T4)	2.981	0.015	0.192	0.274	0.598	0.075	0.880	0.143	0.570	0.537	0.329
N	1,064	1,064	1,064	1,064	1,064	1,064	1,064	1,064	600	600	600

Note: Estimates from the DFSA SPIR endline survey sample, estimated as an ANCOVA model at the household level. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.13: Male's childcare activities in past 3 days

	(1) Number of activities that primary female carried out (0–10)	(2) Told stories to index child	(3) Sang songs to or with index child	(4) Took index child outside the home	(5) Played with index child	(6) Named, counted or drew things with or for index child	(7) Gave index child a bath	(8) Cared for the index child when they were sick	(9) Prepared food for index child (if not exclusively breastfed)	(10) Physically fed index child (if not exclusively breastfed)	(11) Ate a meal together with index child (if not exclusively breastfed)
T1	-0.004	-0.017	0.038	0.057	-0.044	-0.011	0.007	0.024	0.005	-0.023	-0.046
	(0.203)	(0.024)	(0.045)	(0.060)	(0.049)	(0.028)	(0.042)	(0.032)	(0.033)	(0.062)	(0.055)
		[0.789]	[0.789]	[0.789]	[0.789]	[0.869]	[0.869]	[0.789]	[0.869]	[0.869]	[0.789]
T2	0.225	-0.017	0.069	0.066	-0.049	-0.001	-0.039	0.023	0.017	0.117**	0.027
	(0.173)	(0.024)	(0.042)	(0.053)	(0.053)	(0.025)	(0.043)	(0.031)	(0.035)	(0.057)	(0.059)

		[0.680]	[0.503]	[0.680]	[0.680]	[0.984]	[0.680]	[0.680]	[0.712]	[0.432]	[0.712]
T3	-0.001	-0.036*	0.049	0.039	-0.037	-0.004	-0.043	0.009	0.021	0.030	-0.059
	(0.182)	(0.020)	(0.045)	(0.058)	(0.048)	(0.025)	(0.048)	(0.034)	(0.039)	(0.057)	(0.054)
		[0.753]	[0.756]	[0.756]	[0.756]	[0.878]	[0.756]	[0.878]	[0.756]	[0.756]	[0.756]
Pooled effect of N*: T1 or T3	-0.002	-0.027	0.043	0.048	-0.040	-0.007	-0.018	0.016	0.013	0.003	-0.052
	(0.173)	(0.021)	(0.039)	(0.052)	(0.043)	(0.023)	(0.038)	(0.028)	(0.032)	(0.050)	(0.048)
		[0.717]	[0.717]	[0.717]	[0.717]	[0.839]	[0.839]	[0.839]	[0.839]	[0.948]	[0.717]
Test: T1 = T2	(0.166)	(0.995)	(0.456)	(0.864)	(0.930)	(0.709)	(0.315)	(0.996)	(0.715)	(0.050)	(0.177)
		[0.996]	[0.996]	[0.996]	[0.996]	[0.996]	[0.996]	[0.996]	[0.996]	[0.504]	[0.885]
Test: T2 = T3	(0.126)	(0.190)	(0.622)	(0.584)	(0.802)	(0.890)	(0.945)	(0.697)	(0.929)	(0.180)	(0.109)
		[0.634]	[0.945]	[0.945]	[0.945]	[0.945]	[0.945]	[0.945]	[0.945]	[0.634]	[0.634]
Test: T1 = T3	(0.984)	(0.190)	(0.808)	(0.741)	(0.864)	(0.808)	(0.322)	(0.683)	(0.654)	(0.430)	(0.794)
		[0.864]	[0.864]	[0.864]	[0.864]	[0.864]	[0.864]	[0.864]	[0.864]	[0.864]	[0.864]
Mean of control (T4)	2.404	0.051	0.174	0.287	0.736	0.079	0.309	0.084	0.054	0.330	0.241
N	753	753	753	753	753	753	753	753	457	457	457

Note: Estimates from the DFSA SPIR endline survey sample, estimated as an ANCOVA model at the household level. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 8.14: Household's food security

	(1) Raw score from 8 FIES questions	(2) Household is moderately or severely food insecure based on FIES score	(3) Household is severely food insecure based on FIES score	(4) Household has worried about not having enough food to eat because of lack of resources	(5) Household has been unable to eat healthy and nutritious food because of lack of resources	(6) Household has eaten only a few kinds of foods because of lack of resources	(7) Household has had to skip a meal because there were not enough resources to get food	(8) Household has eaten less than they thought they should because of lack of resources	(9) Household has run out of food because of lack of resources	(10) Household has been hungry but not eaten because there were not enough resources for food	(11) Household has gone without eating for a whole day because of lack of resources
T1	0.289	0.045	0.043	0.031	0.020	0.052*	0.059*	0.035	0.022	0.030	0.040*
	(0.191)	(0.035)	(0.026)	(0.031)	(0.032)	(0.031)	(0.030)	(0.036)	(0.034)	(0.027)	(0.022)
		[0.410]	[0.259]	[0.413]	[0.533]	[0.259]	[0.259]	[0.413]	[0.533]	[0.413]	[0.259]
T2	0.195	0.026	0.030	0.005	0.030	0.055*	0.002	0.028	0.036	0.029	0.010
	(0.184)	(0.035)	(0.027)	(0.028)	(0.028)	(0.029)	(0.028)	(0.035)	(0.035)	(0.029)	(0.018)

		[0.634]	[0.634]	[0.946]	[0.634]	[0.598]	[0.947]	[0.634]	[0.634]	[0.634]	[0.709]
T3	0.158	0.030	0.016	0.036	0.008	0.038	0.004	0.032	0.022	0.008	0.005
	(0.172)	(0.033)	(0.024)	(0.028)	(0.027)	(0.029)	(0.028)	(0.033)	(0.032)	(0.026)	(0.018)
		[0.835]	[0.835]	[0.835]	[0.853]	[0.835]	[0.883]	[0.835]	[0.835]	[0.853]	[0.853]
Pooled effect of N*: T1 or T3	0.223	0.038	0.030	0.033	0.014	0.045*	0.031	0.034	0.022	0.019	0.022
	(0.156)	(0.030)	(0.023)	(0.024)	(0.025)	(0.026)	(0.025)	(0.030)	(0.029)	(0.024)	(0.017)
		[0.355]	[0.355]	[0.355]	[0.567]	[0.355]	[0.355]	[0.370]	[0.506]	[0.506]	[0.355]
Test: T1 = T2	(0.635)	(0.618)	(0.611)	(0.455)	(0.781)	(0.899)	(0.059)	(0.856)	(0.666)	(0.978)	(0.122)
		[0.978]	[0.978]	[0.978]	[0.978]	[0.978]	[0.587]	[0.978]	[0.978]	[0.978]	[0.609]
Test: T2 = T3	(0.837)	(0.909)	(0.523)	(0.320)	(0.463)	(0.528)	(0.936)	(0.904)	(0.630)	(0.434)	(0.722)
		[0.936]	[0.936]	[0.936]	[0.936]	[0.936]	[0.936]	[0.936]	[0.936]	[0.936]	[0.936]
Test: T1 = T3	(0.481)	(0.680)	(0.217)	(0.866)	(0.720)	(0.633)	(0.061)	(0.939)	(0.987)	(0.391)	(0.067)
		[0.987]	[0.723]	[0.987]	[0.987]	[0.987]	[0.335]	[0.987]	[0.987]	[0.976]	[0.335]
Mean of control (T4)	3.296	0.441	0.161	0.570	0.699	0.610	0.303	0.452	0.310	0.247	0.109
N	3,775	3,775	3,775	3,772	3,769	3,770	3,767	3,770	3,770	3,767	3,772

Note: Estimates from the DFSA SPIR endline survey sample, estimated as an ANCOVA model at the household level. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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.

9 Evidence on men's and women's mental health, relationship dynamics, agency, and gender equitable attitudes and roles

9.1 Introduction

In this chapter, we report impacts of the different SPIR treatments on men's and women's mental health, marital dynamics, women's agency, and gender equitable attitudes and roles. All four livelihood and nutrition interventions—L, L*, N, and N*—have the potential to improve beneficiaries' mental health, marital relationships, women's agency, and gender attitudes. The VESA and BCC activities under L and N bring women and men together to provide them with knowledge and training over a wide range of topics, thereby improving social capital and catalyzing women's empowerment. The L* and N* interventions go a step further by directly addressing constraints on women's role in intrahousehold decision-making, mobility, and choice of livelihood activities, as well as restrictions on access to markets that derive from cultural and social norms (through SAA) (L*); targeting women for the poultry and cash transfers (L*); including men in the nutrition BCC (N*); creating male engagement groups (N*); and providing IPT-G for women screened for depression (N*). Positive social interaction and social support and reduced poverty-related stress as a result of higher incomes (through the poultry and cash grants or other livelihood activities) are pathways through which SPIR may lead to improvements in mental health and decreases in IPV (Buller et al., 2018; Lund et al., 2018).

9.1.1 Interpreting tables

Similar to the tables shown in Chapter 7 on livelihoods, we present three sets of tables for each set of outcomes related to mental health, marital dynamics, agency, and gender attitudes: the pooled effect of T1 (L*+N*), T2 (L*+N), and T3 (L+N*) on the full sample, and then impacts of subtreatments for the extremely poor and less poor samples. We present the impacts separately for the primary male and primary female. Here, we briefly summarize again how to interpret the second and third tables that look at the restricted samples.

The tables reporting effects for the extremely poor sample restrict the sample across all four treatment arms to the households that were identified as eligible for cash or poultry transfers (the poorest 10 out of 18 households in each kebele). Every extremely poor household in T1 and T2 was randomized to receive either poultry or cash. The tables report the coefficients for the effects of the poultry and cash intervention in T1 or T2 compared to the control arm. In other words, the coefficient for *T1*poultry* represents the effect of the group randomized to receive the combination of T1 and poultry compared to the control arm. Using these coefficients, we calculate linear combinations (means) to estimate the average effect of T1 or T2 across poultry and cash, and the average effect of poultry or cash across T1 and T2. The tables also report the coefficient for T3. At the bottom of the tables, we report the tests of equality of impacts across treatment arms and across poultry and cash.

The tables reporting effects for the less poor sample restrict the sample across all four treatment arms to the households that were **not** eligible for cash or poultry transfers (the richest 8 out of 18 households in each kebele). The tables report coefficients for the effects of T1, T2, and T3 and, at the bottom, the tests of equality of impacts across treatment arms.

In all tables, we adjust the estimation for multiple hypothesis testing and report the resulting q-values in brackets. While the full sample includes the supplemental midline sample, the extremely poor and less poor samples do not, as they were not part of the baseline sample and thus were not randomized to the cash or poultry transfer intervention.

9.2 Impacts on summary indices

We first present impacts on summary indices that represent the average standardized treatment effect for a family of outcomes, and then present impacts on each subcomponent that makes up the summary indices. We create summary indices for a family of outcomes related to mental health, relationship dynamics, agency, and gender attitudes and roles. We create these separately for the primary female and primary male. Following Kling (2007), we either positively or negatively code all subcomponents that make up the summary indices, so that each subcomponent goes in the same direction. For example, in the mental health summary index, we negatively code each subcomponent so that higher values equal worse outcomes for all subcomponents. We then jointly estimate treatment effects on all subcomponents in a seemingly unrelated regression framework, and standardize the resulting effects with respect to the standard deviation of each component's control mean. The average treatment effect is then the equally weighted average of the standardized effects on all subcomponents. The subcomponents that make up each summary index are detailed in Table 9.1 below.

Table 9.1: Summary indices

	Primary Female	**Primary Male**
Mental health	Negatively coded summary index composed of the following indicators: • PHQ-9 (range from 0-27) • PHQ-9>8 • PHQ-9>10 • Stress (range from 0-10) • Unhappiness	Negatively coded summary index composed of the following indicators: • PHQ-9 (range from 0–27) • PHQ-9>8 • PHQ-9>10 • Stress (range from 0–10) • Unhappiness
Relationship dynamics	Negatively coded summary index composed of the following indicators: • Controlling behaviors by husband • Physical violence • Emotional violence • Sexual violence	Positively coded summary index composed of the following indicators: • Respects his wife • Wife respects him • Trusts his wife • Feels comfortable telling her when he disagrees with her
Female agency	Positively coded summary index composed of the following indicators: • Locus of control (1–10) • Decision-making with respect to productive inputs that should be used • Decision-making with respect to quantity of the output that should be sold or consumed at home • Decision-making with respect to use of income generated from the productive activity	N/A

Gender attitudes	Positively coded summary index composed of the following indicators: • Husband is not justified in beating his wife • Acceptable for a woman to travel alone	Positively coded summary index composed of the following indicators: • Husband is not justified in beating his wife • Acceptable for a woman to travel alone • Gender equitable attitudes score (0–20) • Gender equitable attitudes score>median
Gender roles	NA	Positively coded summary index composed of the following indicators: • Spouse helped with household chores (female report) • Spouse helped with cooking or meal preparation (female report) • Spouse helped with collecting firewood and water (female report) • Helped with household chores (male report) • Helped with cooking or meal preparation (male report) • Helped with collecting firewood and water (male report) • Helped the children with their homework for school (male report) • Helped the children prepare for school in the morning (male report)

Figure 9.1 (accompanying Table 9.2a) reveals that for the full sample, there are no impacts on the primary female's summary indices across any treatment arm. Coefficients for all summary indices are close to zero and not significant. Disaggregating by the extremely poor and less poor samples reveals similar results of null impacts (Table 9.2b and Table 9.2c)

Figure 9.1: Impacts on primary female's summary indices, full sample

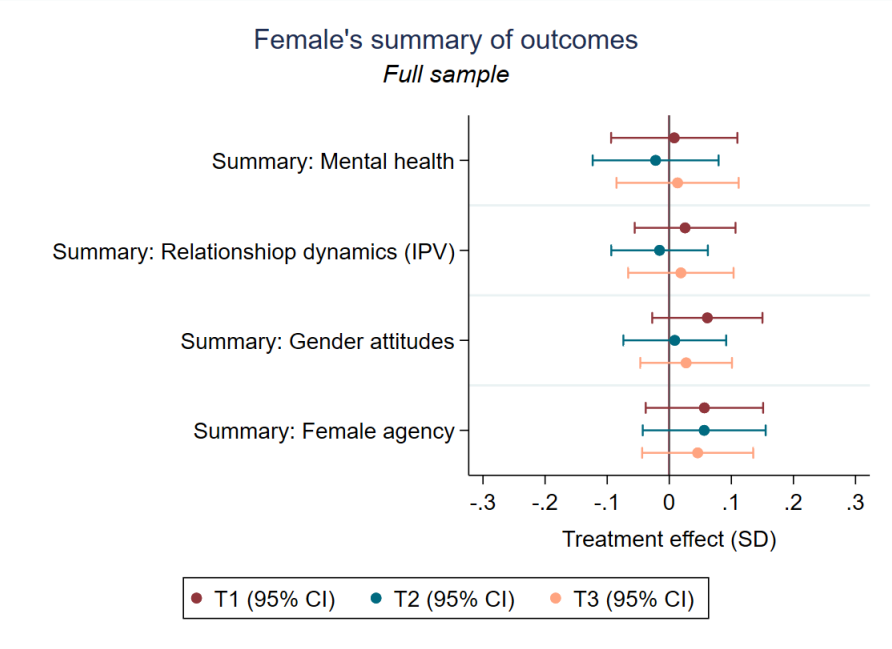


Figure 9.2 (accompanying Table 9.3a) reveals similar null effects for the primary male in terms of mental health and relationship dynamics, but large and significant impacts on their perceived gender equitable attitudes and roles. Impacts are concentrated in the T1 (L*+N*) and T3 (L+N*) arms and are significantly different from those of T2 (L*+N), indicating that N* is needed for impacts on gender equitable attitudes and roles.

Figure 9.2: Impacts on primary male's summary indices, full sample

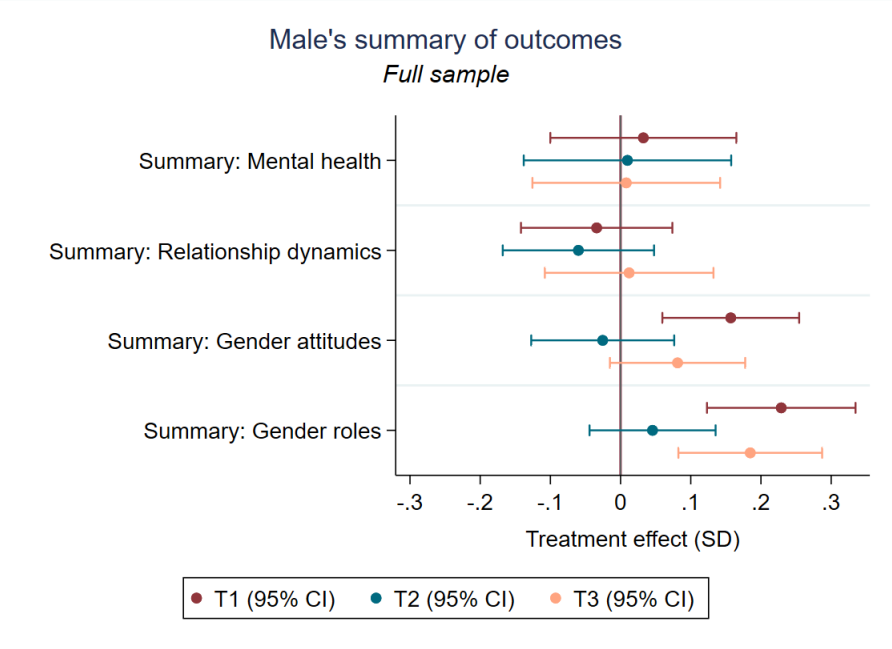


Table 9.3b and Table 9.3c reveal similar results for the extremely poor and less poor subsamples. In the extremely poor sample, improvements in gender equitable attitudes range from 0.129 (T3) to 0.149 (T1) standard deviations and improvements in gender roles range from 0.219 (T3) to 0.252 (T1) standard deviations. In the less poor sample, only T1 leads to significant improvements in gender equitable attitudes, but both T1 and T3 lead to significant improvements in gender roles that range from 0.218 (T1) to 0.245 (T3) standard deviations.

9.3 Mental Health

9.3.1 Indicators

As indicated in Table 9.1, the indicators we analyze related to mental health are depression using the PHQ-9 instrument, unhappiness from the World Value survey, and stress. These survey questions were administered separately for the primary male and female in each household.

The PHQ-9—a nine-item depression diagnostic instrument—was used to assess depressive symptoms of both the primary male and female. The module asks respondents to indicate how frequently they experienced a set of depressive symptoms in the past two weeks, rating these on a scale of zero (never) to three (nearly every day). The PHQ-9 is then scored by adding up the nine responses, leading to a scale of zero to 27. There are various cutoffs for depression suggested in the literature. One that is commonly used defines having mild depression as reporting a PHQ-9 score between 5 and 9, moderate depression between 10 and 14, moderately severe between 15 and19, and an individual with a score 20 and above is deemed to have symptoms of severe depression (Kroenke et al., 2001). For recruitment into the IPT-G groups, we used a cutoff score of 8 in order to increase the potential sample participating in the groups. In the tables below, we show impact estimates on depressive symptoms using both the continuous PHQ-9 scale and binary indicators representing mild-to-severe depressive symptoms using 8 as the cutoff, and moderate-to-severe depressive symptoms using 10 as the cutoff.

In addition to the PHQ-9, we analyze self-reported measures of happiness from the World Value survey. The question asks, "Taking all things together, would you say you are… a) very happy, b) rather happy, c) not very happy, d) not at all happy. To be consistent with the measures of depression above where higher values indicate worse outcomes, we create a binary indicator that is negatively coded and represents responses of not very or not at all happy. Lastly, we analyze a question that asks respondent to rate their current levels of stress on a scale of 1 (not stressed at all) to 10 (extremely stressed).

9.3.2 Pooled effect by treatment arm

The second-to-last rows of Tables 9.4a and 9.5a report the means for the control group for each outcome of interest and reveal that the mean PHQ-9 score is 2.85 for women and 3.0 for men. Approximately 11.4–11.8 percent of males and females reveal mild-to-severe levels of depressive symptoms (PHQ-9>=8) and 4.6–6.2 percent moderate-to-severe levels (PHQ-9>=10) of depressive symptoms. These means are lower than at midline (mean PHQ-9 of 4.23 for women and 4.22 for men), revealing a secular trend of improvements in depressive symptoms. More men than women state they are not happy (30 percent versus 26 percent), and men have slightly higher average stress scores than women (5.3 versus 5.1).

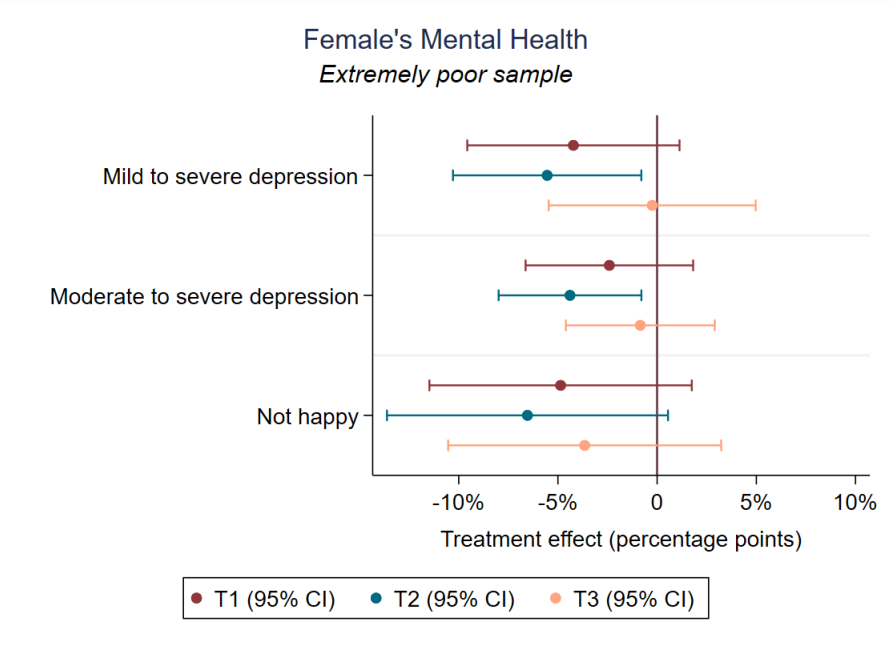
For both the primary male and primary female, we find no impacts across any treatment arm on any mental health indicator (Table 9.4a and Table 9.5a).

9.3.3 Extremely poor households

Table 9.4b and Table 9.5b report the experimental effects observed among extremely poor households. The coefficients in the first part of the table represent impacts of T1 with poultry or cash, T2 with poultry or cash, and T3 compared to extremely poor households in the control arm. The second part of the table calculates the average effect of T1 and T2 for extremely poor households and the average effect of poultry and cash.

Table 9.4b reveals significant decreases in the probability of a woman reporting mild or moderate-to-severe depressive symptoms and unhappiness from the combination of T1 and poultry and the combination of T2 and cash. The average impact of T2 across poultry and cash is significant and represents a decrease of 5.5 percentage points in the probability that the primary female reports mild-to-severe depressive symptoms and 4.4 percentage points in the probability that she reports moderate-to-severe depressive symptoms. These are large impacts and represent decreases of 37–50 percent compared to the control group mean and are significantly different from the impacts of T3. T2 also leads to a significant decrease in the probability that a woman reports feeling unhappy by 6.5 percentage points, or a 20 percent decrease compared to the control group mean. These effects are captured graphically in Figure 9.3.

Figure 9.3: Impacts on mental health, extremely poor sample



The poultry intervention across T1 and T2 also leads to significant decreases of 5.5 percentage points in the probability that a woman reports mild-to-severe depressive symptoms and of 7.5 percentage points in the probability that she reports being unhappy. While the average effects of cash are slightly smaller than the average effects of poultry and only marginally significant in one case, we cannot reject the null hypothesis that the impacts of cash and poultry are equal.

In contrast to the results for the primary female, we do not see any impact on the primary male's mental health indicators across any treatment arm (Table 9.5b).

9.3.4 Less poor households

Table 9.4c and Table 9.5c report the experimental effects of T1, T2, and T3 observed among nonpoor households; these households were not targeted for poultry or cash transfers. Both tables reveal that there are no impacts of SPIR on depressive symptoms or stress among men and women in the less poor sample. Surprisingly, we find positive impacts on women's self-reported unhappiness, however, these positive impacts are not robust to multiple hypothesis testing (q-value is greater than 10 percent), and thus should be interpreted with caution.

9.3.5 Comparison to midline results

At midline, we only collected information on depression and not stress or happiness and found decreases in the probability that the primary female (but not male) reports mild-to-severe depressive symptoms in the full sample of households and the subsamples of extremely poor and less poor households. In particular, we found decreases from T1, T2, and poultry in the subsample of extremely poor households, and from T3 in the subsample of less poor households. While impacts from T1 and T3 have disappeared at endline, impacts from T2 and poultry in the subsample of extremely poor households persist, albeit the magnitude is slightly lower than midline.

9.4 Relationship dynamics

9.4.1 Indicators

The primary female and primary male were administered two different survey instruments to better understand the dynamics of their relationship. The primary female was administered the WHO Violence Against Women Instrument to measure IPV, while the primary male was administered a module from the Women's Empowerment in Agriculture Index (WEAI) to measure marital capital. The latter was measured through questions that ask whether the primary male respects his wife, whether she respects him, whether he trusts her, and whether he feels comfortable telling her when he disagrees with her. The questions had four response categories that range from "never" to "most of the time." We create four binary indicators that correspond to male responding "most of the time."

IPV indicators were collected in accordance with the WHO protocol on ethical guidelines for conducting research on IPV (WHO 2016); only the primary female was administered the IPV module, she had to be alone or with a child less than 36 months, and if she reported any violence she was given the option to be referred to the Women's Affairs Committee in her woreda. We also restricted the module to only women who reported living with their husbands in the last 13 months. From a sample of 3,704 primary females at endline, only 3,038 (82.0 percent) lived with their partners in the last 13 months, and 2,683 (72.4 percent) were also alone or with a child less than 36 months. Of the 2,683 eligible women, 2,676 agreed to be administered the WHO module. For the three types of violence—emotional, physical, and sexual—multiple behaviorally specific questions were administered in order to reduce underreporting. We asked if the woman had ever experienced an act of violence and if she had experienced it in the last 13 months. We analyze the latter in the subsequent tables. The three types of violence are defined as follows.

Emotional spousal violence: Husband/partner said or did something to humiliate you in front of others; threatened to hurt or harm you or someone close to you; insulted you or made you feel bad about yourself.

Physical spousal violence: Husband/partner pushed you, shook you, or threw something at you; slapped you; twisted your arm or pulled your hair; punched you with his fist or with something that could hurt you; kicked you, dragged you, or beat you up; tried to choke you or burn you on purpose; or threatened or attacked you with a knife, gun, or any other weapon.

Sexual spousal violence: Husband/partner physically forced you to have sexual intercourse with him even when you did not want to; physically forced you to perform any other sexual acts you did not want to; forced you with threats or in any other way to perform sexual acts you did not want to.

In addition to the three types of violence, we administered questions on **marital control** as defined as husband/partner demonstrating at least one of the following controlling behaviors: is jealous or angry if she talks to other men; frequently accuses her of being unfaithful; does not permit her to meet her female friends; tries to limit her contact with her family; and insists on knowing where she is at all times.

9.4.2 Pooled effect by treatment arm

The second to last rows of Table 9.6a and Table 9.7a reveal high levels of relationship capital as reported by the primary male (that range from 71 percent of men feeling comfortable disagreeing with their spouse to 93 percent feeling that their spouse trusts them), and low levels IPV as reported by the primary female. While about 43 percent of women have experienced marital control by their husbands/partners, the rates

of violence in the last 13 months are low; 8 percent of women have experienced emotional violence in the last 13 months, 7 percent physical violence, and 3 percent sexual violence. These rates are lower than midline and lower than those reported in Ethiopia's 2016 DHS, which reports 20.3 percent emotional violence in the last 12 months, 16.9 percent physical violence, and 9.3 percent sexual violence.

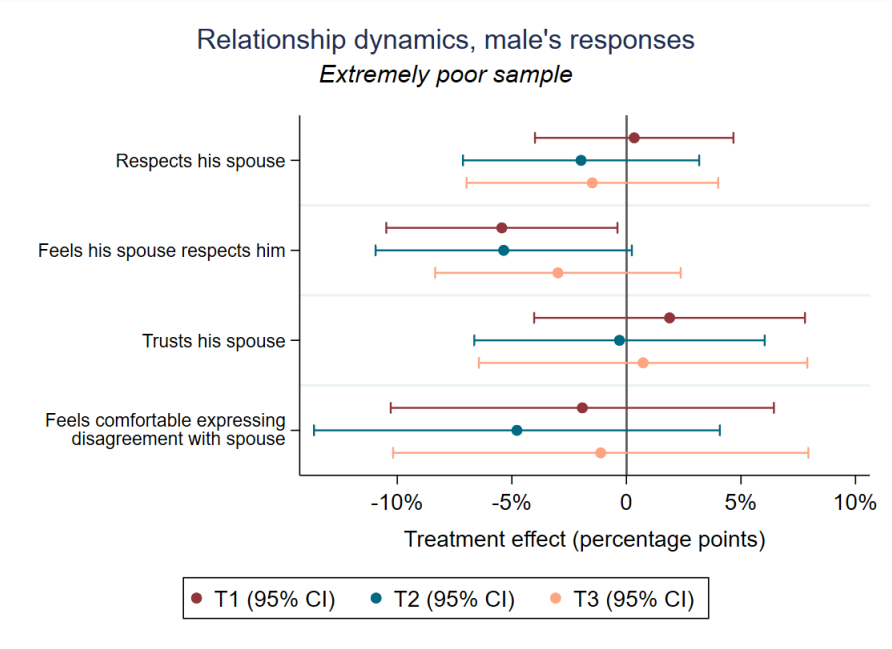
There are no average treatment effects of the SPIR intervention (T1, T2, T3) on marital control or IPV in the last 13 months (Table 9.6a). The coefficients are small in magnitude and statistically insignificant. Except for a couple of marginally significant decreases (that are not robust to multiple hypothesis testing), there are also no impacts of the SPIR intervention on marital capital (Table 9.7a).

9.4.3 Extremely poor households

Although there are no average impacts on IPV among the full sample of women, there are some impacts among the extremely poor sample of households (Table 9.6b). In particular, the combination of T2 and poultry increases emotional IPV by 5 percentage points and increases physical IPV by 6 percentage points. While there are no average treatment effects of T1 or T2 across poultry and cash, the poultry intervention across T1 and T2 also leads to marginally significant increases in emotional and physical violence of 3.7–3.9 percentage points. However, these positive impacts of poultry do not hold up to multiple hypothesis testing.

Consistent with the increases in IPV reported by the primary female, we see decreases in marital capital reported by the primary male (Table 9.7b). In particular, the T1 poultry and T2 poultry interventions decrease the probability that the primary male reports feeling that his spouse respects him most of the time. While the average effects of T1 and T2 across poultry and cash are negative and significant (see Figure 9.4), they do not hold up to multiple hypothesis testing. However, the average effect of poultry across T1 and T2 is large and robust to multiple hypothesis testing. In particular, the poultry intervention across T1 and T2 leads to a decrease of 7.2 percentage points in the probability that the primary male feels his spouse respects him most of the time. The impact of cash on this indicator is also negative but not significant, and while smaller in magnitude, it is not significantly different to that of poultry.

Figure 9.4: Impacts on relationship dynamics, extremely poor sample



9.4.4 Less poor households

Table 9.6c and Table 9.7c reveal that for less poor households, there are no average treatment effects of the SPIR intervention (T1, T2, T3) on marital capital, marital control, or IPV in the last 13 months. In general, the coefficients are small in magnitude and statistically insignificant. The couple of exceptions where we see decreases in physical violence from T2 or decreases in trust from T1, the results are not robust to multiple hypothesis testing.

9.4.5 Comparison to midline results

While we did not analyze impacts on marital capital as reported by the primary male at midline, we did analyze impacts on IPV. At midline reports of IPV from the control arm were higher than those reported at endline, but still lower than those reported in the DHS. Similar to the endline, at midline we found no impacts of any SPIR treatment arms on IPV among the full sample or less poor sample. However, among the extremely poor sample, we found decreases in IPV from the T2 or poultry interventions at midline. The decrease in IPV found in the T2 arm at midline has completely dissipated by endline, and the decrease in IPV found from the poultry intervention, while negative and sometimes significant at midline, is now positive and sometimes significant. Given that the poultry and cash transfers were a one-time transfer that occurred in April 2019, a few months before the midline survey, these finding suggest that the beneficial impacts on relationship dynamics found immediately after the transfer were not sustained almost two years later, and in fact, the poultry transfer may have caused more tensions once the economic benefits of the transfer dissipated.

9.5 Female agency

9.5.1 Indicators

We analyze impacts of SPIR on women's agency using questions on locus of control and decision-making. Locus of control was assessed using a 10-step ladder (similar to the Cantril (1965) ladder used to assess life satisfaction), where respondents are asked to place themselves on the ladder; on the first step are those who are totally unable to change their lives, while on step 10 are those who have full control over their own lives. Decision-making was assessed using an adapted version of the pro-WEAI module on decision-making around production and income that included three activities: crop farming, large livestock, and poultry and other small animals. For each activity, we ask three questions on decision-making specific to the quantity of inputs that should be used; the quantity of the outputs that should be sold or consumed at home; and the use of income generated from the activity. We create three binary indicators that equal one if a woman reports having input into most decisions for all activities in which she participates.

9.5.2 Pooled effect by treatment arm

As can be seen by the mean in the control arm, about half of all women have input into most decisions related to quantity of inputs to be used, quantity of outputs to be sold or consumed, and use of generated income (Table 9.8a). While coefficients are all positive, we find no impacts of SPIR on any female agency indicator.

9.5.3 Extremely poor households

Within the subsample of extremely poor households, we find no impacts of any SPIR treatment on female agency (Table 9.8b). Coefficients are mostly positive, but not significant at conventional levels.

9.5.4 Less poor households

Among the subsample of less poor households, we again see no impacts of any SPIR treatment arm on female agency (Table 9.8c).

9.5.5 Comparison from midline results

While we did not analyze impacts on locus of control at midline, we did analyze impacts on decision-making. At midline we asked the same decision-making questions across four domains related to horticulture, large livestock, small livestock, and poultry and found large impacts across all three arms, especially with respect to the use of generated income. In the subsample of extremely poor households, we also found large average impacts of poultry on decision-making. In the subsample of less poor households, impacts at midline were concentrated in T2 (L*+N). By endline, impacts across all samples and treatment arms have faded. There are many potential explanations such as shocks—COVID-19, fall armyworm, locusts—diminishing any previous impacts. Among the extremely poor sample, the economic benefits of the livelihood transfers may have also dissipated by endline, leading to attenuation of women's agency. Moreover, we may not have captured all domains of decision-making in our survey or indicators. Unfortunately, it is beyond the scope of this report to investigate the reasons for impacts fading over time.

9.6 Gender equitable attitudes

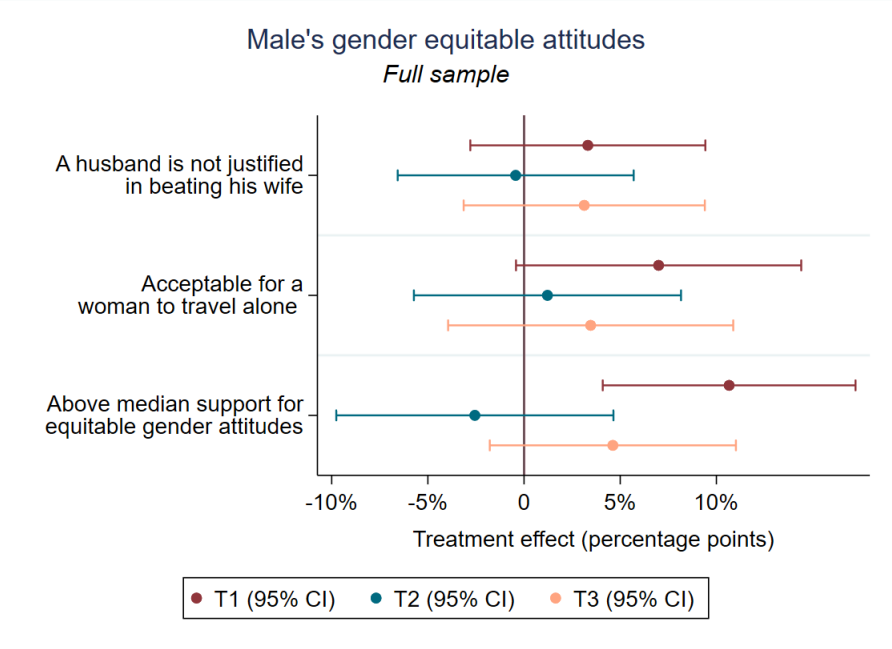
9.6.1 Indicators

We construct three different indicators of equitable gender attitudes. The first is from a list of five questions asked separately to the primary male and female on whether a husband is justified in beating his wife under different circumstances (going out without telling him, burning the food, neglecting the children, arguing with him, refusing to have sex 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 five circumstances. The second is from a list of three questions, again asked separately to the primary male and female, 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 respondent says it is acceptable for a woman to travel alone to the three specified places. Both these indicators on whether the husband is justified in beating his wife and whether it is acceptable for a woman to travel alone are created for the primary male and female. Lastly, for the primary male only, we ask five questions on gender inequitable attitudes where the responses range from strongly disagree (0) to strongly agree (4). Scores are reversely coded and summed up so that a higher score implies more equitable gender attitudes. We analyze both the continuous score and a binary indicator that equals one if scores are above the median.

9.6.2 Pooled effect by treatment arm

Table 9.9a reveals that there are no impacts on the primary female's gender equitable attitudes. In contrast, Figure 9.5 (and accompanying Table 9.10a) reveals that T1 improves the primary male's gender equitable attitudes. In particular, T1 leads to an increase in the probability that the primary male reports it is acceptable for his wife to travel alone of 7 percentage points compared to the control mean of 42.3 percent. It also leads to improvements in the primary male's index of support for gender equitable norms of 0.84 points (compared to the control mean of 10.31) and an increase in the probability that he is above the median score by 10.7 percentage points. These impacts are significantly different from those of T2, indicating that L* is not sufficient in changing attitudes, and N* is also needed for impacts.

Figure 9.5: Impacts on men's gender equitable attitudes, full sample



9.6.3 Extremely poor households

Impacts among the subsample of extremely poor households reveal a similar pattern; we find no impacts on women's attitudes (Table 9.9b), but large impacts from T1 among men (Table 9.10b). In particular, T1 leads to increases in the probability that the primary male reports it is acceptable for his wife to travel alone by 11.8 percentage points compared to the control mean of 37.6 percent. It also leads to improvements in the primary male's index of support for gender equitable norms of 0.73 points (compared to the control mean of 10.29) and an increase in the probability that he is above the median score by 10.2 percentage points. These impacts are again significantly different from those of T2, indicating that L* is not sufficient in changing attitudes, and N* is also needed for impacts. T3 and the combined impact of cash across T1 and T2 also improves the probability that the primary male reports it is acceptable for a woman to travel alone by 7.8 and 11.2 percentage points, respectively.

9.6.4 Less poor households

Similar to the impacts above, we find no impacts on women's attitudes (Table 9.9c), but large impacts from T1 among men (Table 9.10c). T1 leads to improvements in the primary male's gender equitable attitudes by 0.83 points and an increase in the probability that he is above the median score by 10.4 percentage points. These impacts are significantly different from those of T2, indicating that L* is not sufficient in changing attitudes, and N* is also needed for impacts.

9.6.5 Comparison from midline results

At midline, both T1 and T2 led to improvements in the primary male's gender equitable attitudes. Among the extremely poor subsample, there were also large improvements from the poultry intervention across T1 and T2. By endline both the impacts from T2 and the poultry intervention have faded, but impacts

from T1 remain strong, even increasing in magnitude for indicators related to finding it acceptable for women to travel alone and being above the median on the gender equitable norms index.

9.7 Gender equitable roles

9.7.1 Indicators

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 five activities in the last three days: household chores, meal preparation and cooking, collecting firewood and water, children's homework, and preparing child for school in the morning. Women were asked to report their spouse's involvement for the first three activities. We create five binary indicators that equal one if the primary male reports being involved in the specific activity and three binary indicators that equal one if the primary female reports her spouse was engaged in the specific activity.

9.7.2 Pooled effect by treatment arm

The second to last row of Table 9.11a reveals differences in men's and women's reporting of men's engagement in household activities, with men tending to report higher levels of engagement than women. Although men tend to report higher levels of involvement, impacts of T1 and T3 are remarkably similar across male and female responses. In particular, T1 leads to significant increases in men's involvement in household chores, meal preparation and cooking, and collecting firewood and water that range from 0.090–0.184 for female reports and 0.105–0.172 for male reports. Similarly, T3 leads to large increases in men's involvement in these three activities that range from 0.079–0.137 for female reports and 0.112–0.124 for male reports. Impacts of both T1 and T3 are significantly different to those of T2, indicating that N* is needed for improvements in men's engagement in household activities. Similar to impacts on men's involvement in childcare reported in Table 8.13, we find no impacts on men's involvement on helping children with schoolwork or preparing to go to school.

9.7.3 Extremely poor households

Impacts among the subsample of extremely poor households reveals a similar pattern (Table 9.11b). In particular, both T1 and T3 lead to large improvements in men's engagement in household tasks as reported by the primary male and female. Within the T1 arm, both the poultry and cash sub-interventions lead to significant improvements in men's involvement in household chores, cooking and meal preparation, and fetching firewood and water as reported by the primary female and/or male. Average effects of T1 across poultry and cash on the non-childcare activities range from 0.055 to 0.252 percentage points. T3 also leads to significant improvements in non-childcare related activities that range from 0.094 to 0.161 percentage points. Except for fetching firewood and water, impacts of T2 are small and not significant.

9.7.4 Less poor households

Similar to the impacts above, we find that T1 and T3 lead to large improvements in men's involvement in household activities as reported by the primary female and male (Table 9.11c). Impacts are concentrated on non-childcare activities related to household chores, cooking and meal preparation, and fetching firewood and water. Impacts of T1 and T3 are similar, ranging from 0.10 to 0.15 percentage points for T1,

and 0.108 to 0.142 percentage points for T3. Impacts from T1 and T3 are significantly different from T2, suggesting that N* is needed for improvements in men's engagement in household activities.

9.7.5 Comparison with midline results

We did not analyze impacts on men's engagement in household activities at midline, and thus are unable to compare whether impacts are similar. However, male engagement groups that were part of N* were rolled out after the midline survey, and thus impacts would only be captured at endline.

9.8 Conclusion

Among the full sample and less poor sample of households, we find no evidence that SPIR improved the mental health of the primary male or primary female at endline. However, among the subsample of extremely poor households, the combination of T1 and poultry and the combination of T2 and cash lead to decreases in women's (but not men's) depressive symptoms as measured by the probability of a woman reporting mild or moderate-to-severe depressive symptoms and to decreases in reported unhappiness. These impacts are with respect to a control group that receives only PSNP. The average impact of T2 across poultry and cash is large and significant, as is the average impact of poultry across T1 and T2. Impacts of T2 (L*+N) are significantly different from T3 (L+N*), suggesting that L* is needed for improvements in women's mental health among the extremely poor subsample. This is consistent with the literature that links improved economic well-being with improved mental health (Lund 2010). It is important to note that while IPT-G groups were part of N*, these groups were a very small portion of the sample, and thus impacts from IPT-G are likely not captured in this analysis. A separate substudy on impacts of IPT-G groups is planned for fall 2021.

In terms of marital dynamics and IPV, we find no impacts of the SPIR intervention (T1, T2, T3) on these outcomes for either the full sample or subsample of less poor households. For the subsample of extremely poor households, there is some indication that marital relations are worse, especially among poultry households where IPV has increased, and the primary male is less likely to report that his spouse respects him.

Unlike at midline, we find no impacts on female agency of the SPIR intervention in any sample. Any gains made at midline in terms of female agency may have disappeared due to the multiple shocks (COVID-19, locusts, fall armyworm, conflict) households faced between midline and endline, or due to the economic benefits of the one-time livelihood transfer dissipating. Unfortunately, it is beyond the scope of this report to investigate this further.

Although we do not find any gains in female agency at endline, we do find that men's (but not women's) gender equitable attitudes have improved. In particular, across all samples, T1 leads to improvements of men's gender equitable norms scale that range from 0.72 –0.83 points. T1 also increases the probability that the primary male reports that it is acceptable for a woman to travel alone by 7.0 in the full sample and 11.8 percentage points in the subsample of extremely poor households. Impacts of T1 (L*+N*) are significantly different from impacts of T2 (L*+N), suggesting that N* is needed for impacts.

Consistent with improvements in men's gender equitable attitudes, we find large and significant impacts of T1 and T3 on men's engagement in household tasks, as reported by both the primary female and male. Impacts occur across all subsamples and are significantly different from T2, again suggesting that N* is needed for impacts.

A couple of caveats are worth noting. First, we did not stratify our sample across poor and nonpoor households, and therefore, we may not be balanced at baseline on a few indicators in the subsamples. Second, rates of IPV are much lower than those reported in the DHS. A couple of potential explanations are that the rates of IPV are decreasing over time or that violence among PSNP beneficiaries is lower than that of the sample captured by the DHS. Alternatively, households in our sample may be underreporting violence. Although at midline we experimentally compared rates of reporting via indirect methods (using a list experiment) or direct face-to-face methods and found no difference, both methods may not have properly captured violence in the household. In future studies we plan to improve our methods of collecting IPV data by using audio computer-assisted self interviews.

Despite these limitations, the handful of positive impacts across T1 and T3 for men's gender equitable attitudes and role in household tasks and T2 for women's mental health are promising, given that some outcomes might take more time to change, such as personal attitudes regarding gender. However, the change from positive impacts at midline to negative impacts at endline on marital dynamics, especially from the poultry intervention, highlights that positive impacts on marital dynamics from a livelihood transfer targeted to women will not be sustained if the economic benefits are not sustained.

Table 9.2a: Female's summary of all outcome areas

	(1) Mental health (negatively coded)	(2) Intimate partner violence	(3) Gender attitudes	(4) Agency
T1	0.008	0.026	0.061	0.057
	(0.052)	(0.041)	(0.045)	(0.048)
T2	-0.022	-0.016	0.009	0.056
	(0.052)	(0.040)	(0.042)	(0.051)
T3	0.013	0.019	0.027	0.046
	(0.050)	(0.043)	(0.038)	(0.046)
Test: T1 = T2	0.569	0.320	0.274	0.994
Test: T2 = T3	0.487	0.421	0.648	0.817
Test: T1 = T3	0.917	0.878	0.430	0.798
N	3,704	2,676	3,702	3,699

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 9.2b: Female's summary of all outcome areas: Subsample of extremely poor households

	(1) Mental health (negatively coded)	(2) Intimate partner violence	(3) Gender attitudes	(4) Agency
T1 x Poultry	-0.120*	0.026	0.071	0.113
	(0.062)	(0.074)	(0.075)	(0.077)
T1 x Cash	-0.015	0.008	0.023	0.084
	(0.094)	(0.080)	(0.068)	(0.079)
T2 x Poultry	-0.067	0.195*	0.011	0.072
	(0.081)	(0.100)	(0.065)	(0.067)
T2 x Cash	-0.111*	-0.020	0.062	0.030
	(0.063)	(0.075)	(0.071)	(0.079)
T3	-0.013	0.088	0.010	0.061
	(0.060)	(0.073)	(0.049)	(0.060)
Average effect of T1	-0.065	0.016	0.046	0.096
	(0.063)	(0.065)	(0.056)	(0.064)
Average effect of T2	-0.090	0.092	0.037	0.052
	(0.061)	(0.074)	(0.056)	(0.063)
Average effect of poultry	-0.092	0.114	0.040	0.091

	(0.060)	(0.073)	(0.055)	(0.061)
Average effect of cash	-0.063	-0.006	0.043	0.057
	(0.064)	(0.067)	(0.056)	(0.066)
Test: T1 = T2	0.695	0.246	0.874	0.446
Test: T2 = T3	0.197	0.959	0.600	0.874
Test: T1 = T3	0.413	0.260	0.490	0.516
Test: Poultry = Cash	0.643	0.065	0.966	0.544
N	1,723	1,153	1,722	1,722

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 9.2c: Female's summary of all outcome areas: Subsample of less poor households

	(1) Mental health (negatively coded)	(2) Intimate partner violence	(3) Gender attitudes	(4) Agency
T1	0.075	0.057	0.067	-0.045
	(0.063)	(0.069)	(0.058)	(0.061)
T2	0.082	-0.059	-0.047	0.024
	(0.067)	(0.056)	(0.057)	(0.061)
T3	0.048	0.019	0.039	0.041
	(0.060)	(0.061)	(0.058)	(0.061)
Test: T1 = T2	0.926	0.085*	0.054*	0.279
Test: T2 = T3	0.623	0.194	0.148	0.786
Test: T1 = T3	0.669	0.584	0.646	0.173
N	1,288	1,008	1,288	1,288

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 9.3a: Male's summary of all outcome areas

	(1) Emotional well-being (negatively coded)	(2) Relationship dynamics	(3) Perceived gender attitudes	(4) Male's involvement in household tasks [a]
T1	0.032	-0.034	0.157***	0.229***
	(0.068)	(0.055)	(0.050)	(0.054)
T2	0.010	-0.060	-0.025	0.046
	(0.075)	(0.055)	(0.052)	(0.046)

T3	0.008	0.012	0.081*	0.185***
	(0.068)	(0.061)	(0.049)	(0.052)
Test: T1 = T2	0.705	0.630	0.000	0.000
Test: T2 = T3	0.978	0.231	0.014	0.004
Test: T1 = T3	0.642	0.446	0.064	0.429
N	2,463	2,352	2,460	3,213

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.
[a] The set of outcomes also includes females' reports about their spouse's involvement in domestic tasks.

Table 9.3b: Male's summary of all outcome areas: Subsample of extremely poor households

	(1) Emotional well-being (negatively coded)	(2) Relationship dynamics	(3) Perceived gender attitudes	(4) Male's involvement in household tasks [a]
T1 x Poultry	0.017	-0.094	0.149**	0.189***
	(0.092)	(0.073)	(0.071)	(0.066)
T1 x Cash	0.027	-0.017	0.154**	0.322***
	(0.108)	(0.081)	(0.069)	(0.090)
T2 x Poultry	0.067	-0.188*	-0.088	0.114
	(0.117)	(0.114)	(0.077)	(0.073)
T2 x Cash	-0.025	-0.022	0.055	0.077
	(0.116)	(0.081)	(0.068)	(0.073)
T3	0.027	-0.049	0.129**	0.219***
	(0.092)	(0.084)	(0.059)	(0.066)
Average effect of T1	0.022	-0.054	0.149**	0.252***
	(0.087)	(0.063)	(0.059)	(0.062)
Average effect of T2	0.022	-0.108	-0.019	0.097
	(0.100)	(0.080)	(0.062)	(0.061)
Average effect of poultry	0.043	-0.143*	0.026	0.150***
	(0.092)	(0.077)	(0.062)	(0.057)
Average effect of cash	0.001	-0.019	0.104*	0.199***
	(0.096)	(0.067)	(0.059)	(0.065)
Test: T1 = T2	0.996	0.459	0.002	0.016
Test: T2 = T3	0.954	0.519	0.006	0.067
Test: T1 = T3	0.942	0.954	0.693	0.641
Test: Poultry = Cash	0.603	0.096	0.141	0.434
N	1,030	977	1,029	1,372

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.
[a] The set of outcomes also includes females' reports about their spouse's involvement in domestic tasks.

Table 9.3c: Male's summary of all outcome areas: Subsample of less poor households

	(1) Emotional well-being (negatively coded)	(2) Relationship dynamics	(3) Perceived gender attitudes	(4) Male's involvement in household tasks [a]
T1	0.100	-0.127	0.144**	0.218***
	(0.081)	(0.078)	(0.072)	(0.068)
T2	0.039	-0.056	-0.036	0.079
	(0.083)	(0.083)	(0.068)	(0.061)
T3	0.018	0.010	0.049	0.245***
	(0.080)	(0.076)	(0.071)	(0.069)
Test: T1 = T2	0.470	0.415	0.006***	0.029**
Test: T2 = T3	0.792	0.431	0.175	0.009***
Test: T1 = T3	0.301	0.093*	0.161	0.697
N	951	918	949	1,219

Note: Estimates from the DFSA SPIR endline survey sample. All 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. Standard errors are presented in parenthesis. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.
[a] The set of outcomes also includes females' reports about their spouse's involvement in domestic tasks.

Table 9.4a: Female's mental health

	(1) Depression severity score (PHQ-9, 0–27)	(2) Depression severity score is 8 or higher	(3) Depression severity score is 10 or higher	(4) Stress score (1–10)	(5) Respondent is not very happy or not at all happy
T1	0.072	-0.013	-0.003	0.096	0.013
	(0.241)	(0.018)	(0.015)	(0.185)	(0.027)
	[0.835]	[0.835]	[0.835]	[0.835]	[0.835]
T2	0.041	-0.022	-0.016	0.068	-0.007
	(0.242)	(0.018)	(0.013)	(0.181)	(0.028)
	[0.864]	[0.519]	[0.519]	[0.864]	[0.864]
T3	0.119	0.008	-0.004	0.096	-0.005
	(0.235)	(0.019)	(0.012)	(0.179)	(0.027)
	[0.843]	[0.843]	[0.843]	[0.843]	[0.843]

Test: T1 = T2	(0.899)	(0.606)	(0.324)	(0.876)	(0.493)
	[0.899]	[0.899]	[0.899]	[0.899]	[0.899]
Test: T2 = T3	(0.740)	(0.110)	(0.252)	(0.875)	(0.952)
	[0.952]	[0.548]	[0.631]	[0.952]	[0.952]
Test: T1 = T3	(0.841)	(0.281)	(0.961)	(0.997)	(0.523)
	[0.997]	[0.997]	[0.997]	[0.997]	[0.997]
Mean of control (T4)	2.849	0.118	0.062	5.144	0.260
N	3,703	3,703	3,703	3,704	3,699

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.4b: Female's mental health: Subsample of extremely poor households

	(1) Depression severity score (PHQ-9, 0-27)	(2) Depression severity score is 8 or higher	(3) Depression severity score is 10 or higher	(4) Stress score (1-10)	(5) Respondent is not very happy or not at all happy
T1 x Poultry	-0.428	-0.061**	-0.036	-0.117	-0.082**
	(0.317)	(0.028)	(0.025)	(0.225)	(0.041)
	[0.223]	[0.124]	[0.223]	[0.604]	[0.124]
T1 x Cash	0.124	-0.026	-0.013	0.096	-0.019
	(0.515)	(0.040)	(0.029)	(0.295)	(0.043)
	[0.810]	[0.810]	[0.810]	[0.810]	[0.810]
T2 x Poultry	0.090	-0.050	-0.028	0.133	-0.069
	(0.396)	(0.031)	(0.024)	(0.246)	(0.047)
	[0.820]	[0.361]	[0.410]	[0.738]	[0.361]
T2 x Cash	-0.400	-0.059**	-0.059***	0.041	-0.059
	(0.318)	(0.026)	(0.018)	(0.274)	(0.039)
	[0.262]	[0.056]	[0.008]	[0.882]	[0.231]
T3	0.108	-0.002	-0.008	0.064	-0.037
	(0.308)	(0.027)	(0.019)	(0.202)	(0.035)
	[0.926]	[0.926]	[0.926]	[0.926]	[0.926]
Average effect of T1	-0.144	-0.042	-0.024	-0.008	-0.049
	(0.335)	(0.027)	(0.022)	(0.212)	(0.034)
	[0.836]	[0.377]	[0.442]	[0.970]	[0.377]
Average effect of T2	-0.153	-0.055**	-0.044**	0.089	-0.065*
	(0.302)	(0.024)	(0.018)	(0.215)	(0.036)

	[0.678]	[0.058]	[0.058]	[0.678]	[0.120]
Average effect of poultry	-0.159	-0.055**	-0.032	0.013	-0.075**
	(0.297)	(0.024)	(0.020)	(0.196)	(0.036)
	[0.742]	[0.094]	[0.181]	[0.948]	[0.094]
Average effect of cash	-0.137	-0.043	-0.036*	0.069	-0.039
	(0.341)	(0.027)	(0.020)	(0.228)	(0.034)
	[0.764]	[0.293]	[0.293]	[0.764]	[0.428]
Test: T1 = T2	(0.978)	(0.616)	(0.316)	(0.658)	(0.642)
	[0.978]	[0.823]	[0.823]	[0.823]	[0.823]
Test: T2 = T3	(0.374)	(0.033)	(0.036)	(0.902)	(0.432)
	[0.540]	[0.091]	[0.091]	[0.902]	[0.540]
Test: T1 = T3	(0.449)	(0.161)	(0.450)	(0.728)	(0.727)
	[0.728]	[0.728]	[0.728]	[0.728]	[0.728]
Test: Poultry = Cash	(0.948)	(0.628)	(0.839)	(0.797)	(0.304)
	[0.948]	[0.948]	[0.948]	[0.948]	[0.948]
Mean of control (T4)	3.176	0.148	0.088	5.340	0.319
N	1,723	1,723	1,723	1,723	1,721

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.4c: Female's mental health: Subsample of less poor households

	(1) Depression severity score (PHQ-9, 0–27)	(2) Depression severity score is 8 or higher	(3) Depression severity score is 10 or higher	(4) Stress score (1-10)	(5) Respondent is not very happy or not at all happy
T1	0.120	0.013	0.019	0.104	0.067*
	(0.268)	(0.022)	(0.017)	(0.221)	(0.038)
	[0.656]	[0.656]	[0.656]	[0.656]	[0.399]
T2	0.215	0.011	0.009	0.200	0.073**
	(0.298)	(0.026)	(0.017)	(0.224)	(0.037)
	[0.660]	[0.660]	[0.660]	[0.660]	[0.245]
T3	0.247	0.017	0.001	0.157	0.014
	(0.269)	(0.025)	(0.015)	(0.241)	(0.036)
	[0.858]	[0.858]	[0.924]	[0.858]	[0.874]
Test: T1 = T2	(0.756)	(0.961)	(0.606)	(0.640)	(0.867)
	[0.961]	[0.961]	[0.961]	[0.961]	[0.961]

Test: T2 = T3	(0.914)	(0.824)	(0.667)	(0.851)	(0.107)
	[0.914]	[0.914]	[0.914]	[0.914]	[0.534]
Test: T1 = T3	(0.644)	(0.840)	(0.320)	(0.810)	(0.156)
	[0.840]	[0.840]	[0.799]	[0.840]	[0.782]
Mean of control (T4)	2.591	0.096	0.043	4.983	0.206
N	1,288	1,288	1,288	1,288	1,286

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.5a: Male's mental health

	(1) Depression severity score (PHQ-9, 0-27)	(2) Depression severity score is 8 or higher	(3) Depression severity score is 10 or higher	(4) Stress score (1-10)	(5) Respondent is not very happy or not at all happy
T1	0.210	0.008	0.016	0.069	-0.015
	(0.298)	(0.025)	(0.014)	(0.223)	(0.036)
	[0.757]	[0.757]	[0.757]	[0.757]	[0.757]
T2	0.002	0.005	0.007	0.160	-0.029
	(0.331)	(0.028)	(0.015)	(0.219)	(0.039)
	[0.995]	[0.995]	[0.995]	[0.995]	[0.995]
T3	0.154	-0.002	0.007	0.053	-0.023
	(0.294)	(0.026)	(0.013)	(0.217)	(0.037)
	[0.939]	[0.939]	[0.939]	[0.939]	[0.939]
Test: T1 = T2	(0.463)	(0.900)	(0.517)	(0.642)	(0.660)
	[0.825]	[0.900]	[0.825]	[0.825]	[0.825]
Test: T2 = T3	(0.586)	(0.756)	(0.990)	(0.574)	(0.850)
	[0.990]	[0.990]	[0.990]	[0.990]	[0.990]
Test: T1 = T3	(0.819)	(0.609)	(0.464)	(0.935)	(0.801)
	[0.935]	[0.935]	[0.935]	[0.935]	[0.935]
Mean of control (T4)	2.998	0.114	0.046	5.302	0.302
N	2,460	2,460	2,460	2,463	2,458

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.5b: Male's mental health: Subsample of extremely poor households

	(1) Depression severity score (PHQ-9, 0-27)	(2) Depression severity score is 8 or higher	(3) Depression severity score is 10 or higher	(4) Stress score (1-10)	(5) Respondent is not very happy or not at all happy
T1 x Poultry	-0.092	0.002	0.024	0.003	-0.013
	(0.452)	(0.035)	(0.025)	(0.311)	(0.050)
	[0.993]	[0.993]	[0.993]	[0.993]	[0.993]
T1 x Cash	0.226	0.006	0.011	0.029	-0.007
	(0.571)	(0.047)	(0.032)	(0.332)	(0.060)
	[0.929]	[0.929]	[0.929]	[0.929]	[0.929]
T2 x Poultry	0.175	0.023	0.030	0.278	-0.005
	(0.572)	(0.050)	(0.032)	(0.286)	(0.061)
	[0.931]	[0.931]	[0.885]	[0.885]	[0.931]
T2 x Cash	0.020	0.002	0.008	-0.111	-0.073
	(0.551)	(0.043)	(0.029)	(0.370)	(0.053)
	[0.971]	[0.971]	[0.971]	[0.971]	[0.861]
T3	0.272	0.005	0.011	-0.054	0.015
	(0.446)	(0.038)	(0.024)	(0.280)	(0.053)
	[0.906]	[0.906]	[0.906]	[0.906]	[0.906]
Average effect of T1	0.069	0.004	0.017	0.016	-0.010
	(0.440)	(0.036)	(0.024)	(0.257)	(0.047)
	[0.950]	[0.950]	[0.950]	[0.950]	[0.950]
Average effect of T2	0.101	0.013	0.019	0.089	-0.039
	(0.481)	(0.040)	(0.026)	(0.271)	(0.050)
	[0.834]	[0.834]	[0.834]	[0.834]	[0.834]
Average effect of poultry	0.047	0.013	0.027	0.146	-0.009
	(0.452)	(0.037)	(0.025)	(0.248)	(0.049)
	[0.917]	[0.917]	[0.917]	[0.917]	[0.917]
Average effect of cash	0.123	0.004	0.009	-0.041	-0.040
	(0.474)	(0.038)	(0.026)	(0.283)	(0.049)
	[0.910]	[0.910]	[0.910]	[0.910]	[0.910]
Test: T1 = T2	(0.935)	(0.789)	(0.921)	(0.784)	(0.456)
	[0.935]	[0.935]	[0.935]	[0.935]	[0.935]
Test: T2 = T3	(0.667)	(0.807)	(0.720)	(0.615)	(0.234)
	[0.807]	[0.807]	[0.807]	[0.807]	[0.807]
Test: T1 = T3	(0.569)	(0.995)	(0.784)	(0.800)	(0.566)
	[0.995]	[0.995]	[0.995]	[0.995]	[0.995]

Test: Poultry = Cash	(0.848)	(0.789)	(0.445)	(0.493)	(0.446)
	[0.848]	[0.848]	[0.821]	[0.821]	[0.821]
Mean of control (T4)	3.107	0.132	0.056	5.406	0.318
N	1,029	1,029	1,029	1,030	1,028

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.5c: Male's mental health: Subsample of less poor households

	(1) Depression severity score (PHQ-9, 0-27)	(2) Depression severity score is 8 or higher	(3) Depression severity score is 10 or higher	(4) Stress score (1-10)	(5) Respondent is not very happy or not at all happy
T1	0.456	0.037	0.033	0.149	0.003
	(0.371)	(0.030)	(0.020)	(0.283)	(0.044)
	[0.366]	[0.366]	[0.366]	[0.749]	[0.941]
T2	-0.028	0.036	0.004	0.189	-0.003
	(0.367)	(0.034)	(0.017)	(0.290)	(0.045)
	[0.947]	[0.947]	[0.947]	[0.947]	[0.947]
T3	0.224	0.020	0.001	0.181	-0.046
	(0.351)	(0.031)	(0.019)	(0.292)	(0.041)
	[0.671]	[0.671]	[0.937]	[0.671]	[0.671]
Test: T1 = T2	(0.205)	(0.978)	(0.128)	(0.885)	(0.890)
	[0.513]	[0.978]	[0.513]	[0.978]	[0.978]
Test: T2 = T3	(0.482)	(0.641)	(0.887)	(0.976)	(0.312)
	[0.976]	[0.976]	[0.976]	[0.976]	[0.976]
Test: T1 = T3	(0.519)	(0.578)	(0.116)	(0.909)	(0.230)
	[0.723]	[0.723]	[0.574]	[0.909]	[0.574]
Mean of control (T4)	2.735	0.081	0.033	5.218	0.275
N	949	949	949	951	949

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.6a: Intimate partner violence (IPV)

	(1) Any marital control by husband/partner	(2) Experienced emotional violence in the past 13 months	(3) Experienced physical violence in the past 13 months	(4) Experienced sexual violence in the past 13 months
T1	0.003	0.025	0.002	-0.001
	(0.031)	(0.016)	(0.014)	(0.010)
	[0.927]	[0.473]	[0.927]	[0.927]
T2	-0.004	-0.013	-0.007	0.004
	(0.031)	(0.013)	(0.013)	(0.010)
	[0.903]	[0.903]	[0.903]	[0.903]
T3	0.020	0.014	-0.006	0.001
	(0.034)	(0.016)	(0.014)	(0.010)
	[0.911]	[0.911]	[0.911]	[0.926]
Test: T1 = T2	(0.815)	(0.017)	(0.511)	(0.620)
	[0.815]	[0.066]	[0.815]	[0.815]
Test: T2 = T3	(0.485)	(0.075)	(0.905)	(0.747)
	[0.905]	[0.302]	[0.905]	[0.905]
Test: T1 = T3	(0.618)	(0.540)	(0.618)	(0.856)
	[0.824]	[0.824]	[0.824]	[0.856]
Mean of control (T4)	0.430	0.080	0.072	0.030
N	2,676	2,676	2,676	2,675

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.6b: Intimate partner violence (IPV): Subsample of extremely poor households

	(1) Any marital control by husband/partner	(2) Experienced emotional violence in the past 13 months	(3) Experienced physical violence in the past 13 months	(4) Experienced sexual violence in the past 13 months
T1 x Poultry	-0.040	0.027	0.011	0.004
	(0.056)	(0.024)	(0.020)	(0.021)
	[0.781]	[0.781]	[0.781]	[0.857]
T1 x Cash	-0.001	0.030	0.010	-0.022

	(0.063)	(0.027)	(0.026)	(0.015)
	[0.987]	[0.537]	[0.937]	[0.537]
T2 x Poultry	0.048	0.051**	0.062*	0.034
	(0.061)	(0.025)	(0.031)	(0.023)
	[0.437]	[0.102]	[0.102]	[0.191]
T2 x Cash	0.008	-0.022	0.004	-0.003
	(0.056)	(0.020)	(0.021)	(0.018)
	[0.883]	[0.883]	[0.883]	[0.883]
T3	0.011	0.038*	0.024	0.011
	(0.054)	(0.022)	(0.022)	(0.016)
	[0.845]	[0.354]	[0.584]	[0.682]
Average effect of T1	-0.020	0.028	0.010	-0.009
	(0.048)	(0.021)	(0.019)	(0.015)
	[0.684]	[0.684]	[0.684]	[0.684]
Average effect of T2	0.029	0.015	0.034	0.016
	(0.050)	(0.019)	(0.022)	(0.017)
	[0.563]	[0.563]	[0.484]	[0.563]
Average effect of poultry	0.006	0.039*	0.037*	0.019
	(0.049)	(0.020)	(0.021)	(0.017)
	[0.909]	[0.165]	[0.165]	[0.355]
Average effect of cash	0.004	0.004	0.007	-0.013
	(0.050)	(0.020)	(0.019)	(0.014)
	[0.943]	[0.943]	[0.943]	[0.943]
Test: T1 = T2	(0.281)	(0.515)	(0.254)	(0.130)
	[0.375]	[0.515]	[0.375]	[0.375]
Test: T2 = T3	(0.713)	(0.274)	(0.663)	(0.748)
	[0.748]	[0.748]	[0.748]	[0.748]
Test: T1 = T3	(0.542)	(0.638)	(0.521)	(0.198)
	[0.638]	[0.638]	[0.638]	[0.638]
Test: Poultry = Cash	(0.966)	(0.077)	(0.145)	(0.047)
	[0.966]	[0.155]	[0.194]	[0.155]
Mean of control (T4)	0.404	0.060	0.053	0.026
N	1,153	1,153	1,153	1,152

Notes: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.6c: Intimate partner violence (IPV): Subsample of less poor households

	(1)	(2)	(3)	(4)

	Any marital control by husband/partner	Experienced emotional violence in the past 13 months	Experienced physical violence in the past 13 months	Experienced sexual violence in the past 13 months
T1	0.025	0.018	0.005	0.014
	(0.042)	(0.028)	(0.028)	(0.016)
	[0.727]	[0.727]	[0.863]	[0.727]
T2	-0.000	-0.033	-0.039*	0.002
	(0.042)	(0.022)	(0.021)	(0.013)
	[0.998]	[0.285]	[0.260]	[0.998]
T3	0.044	0.019	-0.020	-0.001
	(0.041)	(0.030)	(0.024)	(0.012)
	[0.699]	[0.699]	[0.699]	[0.916]
Test: T1 = T2	(0.554)	(0.048)	(0.070)	(0.456)
	[0.554]	[0.139]	[0.139]	[0.554]
Test: T2 = T3	(0.304)	(0.069)	(0.315)	(0.789)
	[0.420]	[0.277]	[0.420]	[0.789]
Test: T1 = T3	(0.653)	(0.968)	(0.345)	(0.333)
	[0.871]	[0.968]	[0.691]	[0.691]
Mean of control (T4)	0.435	0.096	0.083	0.022
N	1,008	1,008	1,008	1,008

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.7a: Relationship dynamics, reported by primary male

	(1) Respects his spouse most of the time*	(2) Feels his spouse respects him most of the time*	(3) Trusts his spouse most of the time*	(4) Feels comfortable expressing disagreement most of the time*
T1	0.005	-0.035*	-0.006	0.003
	(0.017)	(0.020)	(0.024)	(0.031)
	[0.932]	[0.325]	[0.932]	[0.932]
T2	-0.002	-0.025	-0.001	-0.057*

	(0.018)	(0.018)	(0.023)	(0.032)
	[0.974]	[0.333]	[0.974]	[0.319]
T3	0.006	-0.013	0.019	0.013
	(0.020)	(0.019)	(0.025)	(0.033)
	[0.759]	[0.759]	[0.759]	[0.759]
Test: T1 = T2	(0.686)	(0.614)	(0.827)	(0.045)
	[0.827]	[0.827]	[0.827]	[0.181]
Test: T2 = T3	(0.701)	(0.544)	(0.399)	(0.028)
	[0.701]	[0.701]	[0.701]	[0.112]
Test: T1 = T3	(0.951)	(0.303)	(0.312)	(0.734)
	[0.951]	[0.624]	[0.624]	[0.951]
Mean of control (T4)	0.918	0.933	0.875	0.712
N	2,351	2,350	2,352	2,347

* Questions were asked as 'Do you [respect] your spouse?' with the answer options being most of the time, sometimes, rarely, and never.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.7b: Relationship dynamics, reported by primary male: Subsample of extremely poor households

	(1) Respects his spouse most of the time*	(2) Feels his spouse respects him most of the time*	(3) Trusts his spouse most of the time*	(4) Feels comfortable expressing disagreement most of the time*
T1 x Poultry	-0.017	-0.061*	-0.005	-0.012
	(0.026)	(0.033)	(0.037)	(0.048)
	[0.900]	[0.256]	[0.900]	[0.900]
T1 x Cash	0.023	-0.050	0.042	-0.027
	(0.027)	(0.035)	(0.038)	(0.053)
	[0.535]	[0.528]	[0.528]	[0.607]
T2 x Poultry	-0.042	-0.083**	-0.023	-0.062
	(0.035)	(0.040)	(0.042)	(0.054)
	[0.341]	[0.158]	[0.580]	[0.341]
T2 x Cash	0.004	-0.021	0.018	-0.031
	(0.029)	(0.032)	(0.036)	(0.049)
	[0.899]	[0.815]	[0.815]	[0.815]
T3	-0.015	-0.030	0.007	-0.011

	(0.028)	(0.027)	(0.037)	(0.046)
	[0.842]	[0.842]	[0.842]	[0.842]
Average effect of T1	0.003	-0.054**	0.019	-0.019
	(0.022)	(0.026)	(0.030)	(0.043)
	[0.878]	[0.144]	[0.870]	[0.870]
Average effect of T2	-0.020	-0.054*	-0.003	-0.048
	(0.026)	(0.029)	(0.032)	(0.045)
	[0.604]	[0.249]	[0.926]	[0.583]
Average effect of poultry	-0.030	-0.072**	-0.014	-0.038
	(0.025)	(0.028)	(0.032)	(0.044)
	[0.485]	[0.047]	[0.659]	[0.519]
Average effect of cash	0.013	-0.036	0.030	-0.029
	(0.023)	(0.026)	(0.030)	(0.044)
	[0.569]	[0.569]	[0.569]	[0.569]
Test: T1 = T2	(0.317)	(0.978)	(0.474)	(0.444)
	[0.632]	[0.978]	[0.632]	[0.632]
Test: T2 = T3	(0.865)	(0.458)	(0.778)	(0.355)
	[0.865]	[0.865]	[0.865]	[0.865]
Test: T1 = T3	(0.473)	(0.405)	(0.742)	(0.833)
	[0.833]	[0.833]	[0.833]	[0.833]
Test: Poultry = Cash	(0.068)	(0.239)	(0.148)	(0.817)
	[0.274]	[0.319]	[0.296]	[0.817]
Mean of control (T4)	0.932	0.950	0.864	0.740
N	977	976	977	975

* Questions were asked as "Do you [respect] your spouse?" with the answer options being most of the time, sometimes, rarely, and never.
Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.7c: Relationship dynamics, reported by primary male: Subsample of less poor households

	(1) Respects his spouse most of the time*	(2) Feels his spouse respects him most of the time*	(3) Trusts his spouse most of the time*	(4) Feels comfortable expressing disagreement most of the time*
T1	-0.017	-0.025	-0.062**	-0.043
	(0.026)	(0.031)	(0.031)	(0.047)

	[0.504]	[0.504]	[0.175]	[0.504]
T2	0.003	0.006	-0.008	-0.079
	(0.027)	(0.028)	(0.030)	(0.052)
	[0.896]	[0.896]	[0.896]	[0.509]
T3	0.010	0.009	0.002	0.006
	(0.024)	(0.029)	(0.029)	(0.047)
	[0.957]	[0.957]	[0.957]	[0.957]
Test: T1 = T2	(0.475)	(0.277)	(0.078)	(0.459)
	[0.475]	[0.475]	[0.311]	[0.475]
Test: T2 = T3	(0.804)	(0.904)	(0.736)	(0.076)
	[0.904]	[0.904]	[0.904]	[0.305]
Test: T1 = T3	(0.317)	(0.252)	(0.033)	(0.271)
	[0.317]	[0.317]	[0.132]	[0.317]
Mean of control (T4)	0.920	0.920	0.915	0.715
N	917	917	918	916

* Questions were asked as "Do you [respect] your spouse?" with the answer options being most of the time, sometimes, rarely, and never.
Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.8a: Female's agency and input into decision-making around production

	(1) Locus of control (1–10)	(2) Input into most decisions about how much inputs to be used	(3) Input into most decisions about how much output to be consumed or sold	(4) Input into most decisions about use of generated income
T1	0.085	0.030	0.036	0.029
	(0.155)	(0.028)	(0.029)	(0.031)
	[0.583]	[0.479]	[0.479]	[0.479]
T2	0.110	0.026	0.027	0.042
	(0.148)	(0.030)	(0.030)	(0.031)
	[0.457]	[0.457]	[0.457]	[0.457]
T3	0.141	0.014	0.033	0.017
	(0.141)	(0.028)	(0.030)	(0.031)
	[0.618]	[0.618]	[0.618]	[0.618]
Test: T1 = T2	(0.880)	(0.897)	(0.744)	(0.670)
	[0.897]	[0.897]	[0.897]	[0.897]
Test: T2 = T3	(0.837)	(0.632)	(0.846)	(0.409)

	[0.846]	[0.846]	[0.846]	[0.846]
Test: T1 = T3	(0.723)	(0.513)	(0.896)	(0.695)
	[0.896]	[0.896]	[0.896]	[0.896]
Mean of control (T4)	4.194	0.511	0.511	0.537
N	3,693	3,073	3,073	3,073

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.8b: Female's agency and input into decision-making: Subsample of extremely poor households

	(1) Locus of control (1-10)	(2) Input into most decisions about how much inputs to be used	(3) Input into most decisions about how much output to be consumed or sold	(4) Input into most decisions about use of generated income
T1 x Poultry	0.396	0.045	0.058	0.031
	(0.241)	(0.045)	(0.049)	(0.048)
	[0.409]	[0.431]	[0.431]	[0.518]
T1 x Cash	-0.273	0.079	0.063	0.085
	(0.222)	(0.055)	(0.054)	(0.053)
	[0.247]	[0.247]	[0.247]	[0.247]
T2 x Poultry	0.159	0.027	0.052	0.023
	(0.196)	(0.044)	(0.046)	(0.045)
	[0.611]	[0.611]	[0.611]	[0.611]
T2 x Cash	0.037	0.012	0.003	0.056
	(0.231)	(0.048)	(0.048)	(0.047)
	[0.947]	[0.947]	[0.947]	[0.940]
T3	0.076	0.019	0.043	0.050
	(0.192)	(0.041)	(0.039)	(0.041)
	[0.692]	[0.692]	[0.553]	[0.553]
Average effect of T1	0.054	0.061	0.059	0.057
	(0.182)	(0.041)	(0.042)	(0.042)
	[0.769]	[0.225]	[0.225]	[0.225]
Average effect of T2	0.102	0.020	0.028	0.040
	(0.177)	(0.040)	(0.041)	(0.040)
	[0.610]	[0.610]	[0.610]	[0.610]
Average effect of poultry	0.273	0.036	0.055	0.027

	(0.176)	(0.038)	(0.040)	(0.039)
	[0.346]	[0.469]	[0.346]	[0.500]
Average effect of cash	-0.118	0.046	0.033	0.071*
	(0.184)	(0.043)	(0.042)	(0.042)
	[0.522]	[0.522]	[0.522]	[0.376]
Test: T1 = T2	(0.792)	(0.272)	(0.414)	(0.625)
	[0.792]	[0.792]	[0.792]	[0.792]
Test: T2 = T3	(0.895)	(0.971)	(0.670)	(0.755)
	[0.971]	[0.971]	[0.971]	[0.971]
Test: T1 = T3	(0.910)	(0.280)	(0.644)	(0.849)
	[0.910]	[0.910]	[0.910]	[0.910]
Test: Poultry = Cash	(0.035)	(0.780)	(0.556)	(0.210)
	[0.139]	[0.780]	[0.741]	[0.420]
Mean of control (T4)	4.093	0.551	0.545	0.555
N	1,721	1,369	1,369	1,369

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.8c: Female's agency and input into decision-making: Subsample of less poor households

	(1) Locus of control (1-10)	(2) Input into most decisions about how much inputs to be used	(3) Input into most decisions about how much output to be consumed or sold	(4) Input into most decisions about use of generated income
T1	0.013	-0.019	-0.032	-0.044
	(0.239)	(0.037)	(0.038)	(0.041)
	[0.957]	[0.807]	[0.787]	[0.787]
T2	-0.079	0.049	0.001	0.019
	(0.204)	(0.039)	(0.040)	(0.041)
	[0.933]	[0.838]	[0.977]	[0.933]
T3	0.204	0.029	0.032	-0.017
	(0.192)	(0.038)	(0.041)	(0.041)
	[0.595]	[0.595]	[0.595]	[0.685]
Test: T1 = T2	(0.708)	(0.063)	(0.400)	(0.140)
	[0.708]	[0.251]	[0.534]	[0.281]

Test: T2 = T3	(0.151)	(0.590)	(0.475)	(0.408)
	[0.590]	[0.590]	[0.590]	[0.590]
Test: T1 = T3	(0.408)	(0.182)	(0.117)	(0.523)
	[0.523]	[0.363]	[0.363]	[0.523]
Mean of control (T4)	4.452	0.467	0.500	0.537
N	1,284	1,106	1,106	1,106

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.9a: Female's gender attitudes

	(1) A husband is not justified in beating his wife in any of these situations*	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends
T1	0.040	0.019
	(0.025)	(0.033)
	[0.205]	[0.565]
T2	-0.004	0.012
	(0.023)	(0.030)
	[0.863]	[0.863]
T3	0.037	-0.010
	(0.025)	(0.028)
	[0.285]	[0.731]
Test: T1 = T2	(0.067)	(0.833)
	[0.133]	[0.833]
Test: T2 = T3	(0.089)	(0.462)
	[0.178]	[0.462]
Test: T1 = T3	(0.905)	(0.378)
	[0.905]	[0.756]
Mean of control (T4)	0.438	0.360
N	3,693	3,693

* 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 refuses to have sex with husband.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.9b: Female's gender attitudes: Subsample of extremely poor households

	(1) A husband is not justified in beating his wife in any of these situations*	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends
T1 x Poultry	0.026	0.039
	(0.046)	(0.051)
	[0.567]	[0.567]
T1 x Cash	0.029	-0.006
	(0.040)	(0.049)
	[0.898]	[0.898]
T2 x Poultry	0.009	-0.002
	(0.039)	(0.044)
	[0.970]	[0.970]
T2 x Cash	-0.003	0.060
	(0.041)	(0.050)
	[0.948]	[0.463]
T3	0.003	0.005
	(0.034)	(0.034)
	[0.918]	[0.918]
Average effect of T1	0.027	0.016
	(0.034)	(0.038)
	[0.687]	[0.687]
Average effect of T2	0.003	0.029
	(0.034)	(0.038)
	[0.927]	[0.884]
Average effect of poultry	0.017	0.018
	(0.035)	(0.037)
	[0.632]	[0.632]
Average effect of cash	0.013	0.027
	(0.033)	(0.039)
	[0.691]	[0.691]

Test: T1 = T2	(0.474)	(0.743)
	[0.743]	[0.743]
Test: T2 = T3	(0.991)	(0.507)
	[0.991]	[0.991]
Test: T1 = T3	(0.480)	(0.779)
	[0.779]	[0.779]
Test: Poultry = Cash	(0.908)	(0.823)
	[0.908]	[0.908]
Mean of control (T4)	0.448	0.336
N	1,719	1,719

* 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 refuses to have sex with husband.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.9c: Female's gender attitudes: Subsample of less poor households

	(1) A husband is not justified in beating his wife in any of these situations*	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends
T1	0.034	0.032
	(0.035)	(0.041)
	[0.444]	[0.444]
T2	-0.022	-0.020
	(0.034)	(0.040)
	[0.609]	[0.609]
T3	0.054	-0.011
	(0.037)	(0.039)
	[0.292]	[0.784]

Test: T1 = T2	(0.104)	(0.212)
	[0.208]	[0.212]
Test: T2 = T3	(0.038)	(0.808)
	[0.075]	[0.808]
Test: T1 = T3	(0.589)	(0.307)
	[0.589]	[0.589]
Mean of control (T4)	0.435	0.363
N	1,285	1,284

*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 refuses to have sex with husband.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.10a: Male's gender attitudes

	(1) A husband is not justified in beating his wife in any of these situations*	(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: all questions	(4) Above median support for equitable gender norms
T1	0.033	0.070*	0.839**	0.107***
	(0.031)	(0.038)	(0.339)	(0.034)
	[0.289]	[0.088]	[0.028]	[0.007]
T2	-0.004	0.012	-0.228	-0.026
	(0.031)	(0.035)	(0.348)	(0.037)
	[0.890]	[0.890]	[0.890]	[0.890]
T3	0.031	0.035	0.409	0.046
	(0.032)	(0.038)	(0.318)	(0.033)
	[0.361]	[0.361]	[0.361]	[0.361]
Test: T1 = T2	(0.195)	(0.073)	(0.000)	(0.000)
	[0.195]	[0.098]	[0.000]	[0.000]

Test: T2 = T3	(0.217)	(0.481)	(0.014)	(0.024)
	[0.289]	[0.481]	[0.048]	[0.048]
Test: T1 = T3	(0.949)	(0.310)	(0.086)	(0.033)
	[0.949]	[0.413]	[0.172]	[0.133]
Mean of control (T4)	0.667	0.423	10.311	0.519
N	2,450	2,458	2,460	2,460

* 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 refuses to have sex with husband.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 9.10b: Male's gender attitudes: Subsample of extremely poor households

	(1) A husband is not justified in beating his wife in any of these situations*	(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: all questions	(4) Above median support for equitable gender norms
T1 x Poultry	-0.057	0.144**	0.912*	0.104*
	(0.053)	(0.058)	(0.503)	(0.058)
	[0.286]	[0.054]	[0.102]	[0.102]
T1 x Cash	0.034	0.099*	0.573	0.103*
	(0.052)	(0.059)	(0.452)	(0.054)
	[0.518]	[0.189]	[0.276]	[0.189]
T2 x Poultry	-0.055	0.005	-0.417	-0.060
	(0.047)	(0.052)	(0.551)	(0.065)
	[0.601]	[0.927]	[0.601]	[0.601]
T2 x Cash	0.021	0.125**	-0.201	-0.019
	(0.053)	(0.056)	(0.480)	(0.053)
	[0.727]	[0.111]	[0.727]	[0.727]
T3	0.032	0.078*	0.604	0.070
	(0.041)	(0.046)	(0.404)	(0.046)

	[0.438]	[0.182]	[0.182]	[0.182]
Average effect of T1	-0.010	0.118**	0.725*	0.102**
	(0.042)	(0.046)	(0.409)	(0.046)
	[0.808]	[0.046]	[0.104]	[0.059]
Average effect of T2	-0.018	0.065	-0.317	-0.041
	(0.042)	(0.046)	(0.451)	(0.050)
	[0.660]	[0.641]	[0.644]	[0.644]
Average effect of poultry	-0.056	0.072	0.221	0.019
	(0.041)	(0.045)	(0.448)	(0.050)
	[0.341]	[0.341]	[0.713]	[0.713]
Average effect of cash	0.027	0.112**	0.187	0.043
	(0.043)	(0.047)	(0.411)	(0.046)
	[0.650]	[0.076]	[0.650]	[0.650]
Test: T1 = T2	(0.847)	(0.220)	(0.004)	(0.001)
	[0.847]	[0.293]	[0.007]	[0.005]
Test: T2 = T3	(0.203)	(0.753)	(0.007)	(0.010)
	[0.270]	[0.753]	[0.019]	[0.019]
Test: T1 = T3	(0.303)	(0.365)	(0.690)	(0.425)
	[0.567]	[0.567]	[0.690]	[0.567]
Test: Poultry = Cash	(0.047)	(0.370)	(0.922)	(0.584)
	[0.187]	[0.740]	[0.922]	[0.779]
Mean of control (T4)	0.693	0.376	10.291	0.517
N	1,025	1,029	1,029	1,029

* 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 refuses to have sex with husband.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.10c: Male's gender attitudes: Subsample of less poor households

	(1) A husband is not justified in beating his wife in any of these	(2) It is acceptable for a woman to travel alone to	(3) Index of support for equitable gender norms: all	(4) Above median support for equitable gender norms

	situations*	market, health center, and to visit friends	questions	
T1	0.059	0.018	0.834*	0.104*
	(0.043)	(0.055)	(0.447)	(0.054)
	[0.227]	[0.747]	[0.127]	[0.127]
T2	0.014	-0.008	-0.311	-0.031
	(0.042)	(0.052)	(0.443)	(0.052)
	[0.882]	[0.882]	[0.882]	[0.882]
T3	0.042	-0.000	0.313	0.018
	(0.046)	(0.056)	(0.415)	(0.051)
	[0.903]	[0.997]	[0.903]	[0.971]
Test: T1 = T2	(0.208)	(0.576)	(0.006)	(0.006)
	[0.278]	[0.576]	[0.013]	[0.013]
Test: T2 = T3	(0.469)	(0.864)	(0.097)	(0.277)
	[0.625]	[0.864]	[0.389]	[0.554]
Test: T1 = T3	(0.665)	(0.716)	(0.175)	(0.075)
	[0.716]	[0.716]	[0.350]	[0.300]
Mean of control (T4)	0.651	0.455	10.289	0.517
N	944	948	949	949

* 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 refuses to have sex with husband.

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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. 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 9.11a: Male's involvement in domestic tasks in the past 3 days

	(1) Female reports that spouse helped with household chores (e.g., cleaning)	(2) Female reports that spouse helped with cooking or meal preparation	(3) Female reports that spouse 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) Male reports that he helped the children with their homework for school	(8) Male reports that he helped the children prepare for school in

								the morning
T1	0.106***	0.090***	0.184***	0.127***	0.105***	0.172***	0.005	0.023
	(0.035)	(0.031)	(0.033)	(0.038)	(0.030)	(0.032)	(0.034)	(0.036)
	[0.005]	[0.005]	[0.000]	[0.002]	[0.001]	[0.000]	[0.874]	[0.600]
T2	-0.003	0.012	0.055*	0.028	0.002	0.101***	-0.012	-0.009
	(0.032)	(0.025)	(0.029)	(0.034)	(0.027)	(0.033)	(0.032)	(0.037)
	[0.955]	[0.955]	[0.223]	[0.955]	[0.955]	[0.018]	[0.955]	[0.955]
T3	0.079**	0.087***	0.137***	0.124***	0.118***	0.112***	0.005	-0.014
	(0.034)	(0.026)	(0.036)	(0.036)	(0.029)	(0.036)	(0.031)	(0.036)
	[0.029]	[0.002]	[0.001]	[0.002]	[0.000]	[0.004]	[0.868]	[0.788]
Test: T1 = T2	(0.001)	(0.005)	(0.000)	(0.007)	(0.000)	(0.008)	(0.616)	(0.381)
	[0.003]	[0.009]	[0.000]	[0.011]	[0.002]	[0.011]	[0.616]	[0.435]
Test: T2 = T3	(0.011)	(0.001)	(0.014)	(0.006)	(0.000)	(0.748)	(0.581)	(0.877)
	[0.021]	[0.004]	[0.022]	[0.016]	[0.000]	[0.854]	[0.774]	[0.877]
Test: T1 = T3	(0.424)	(0.931)	(0.197)	(0.938)	(0.664)	(0.050)	(0.995)	(0.282)
	[0.847]	[0.995]	[0.753]	[0.995]	[0.995]	[0.399]	[0.995]	[0.753]
Mean of control (T4)	0.198	0.154	0.367	0.295	0.228	0.553	0.450	0.699
N	3,070	3,070	3,069	2,452	2,450	2,451	2,145	2,149

Note: 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.

Table 9.11b: Male's involvement in domestic tasks in the past 3 days: Subsample of extremely poor households

	(1) Female reports that spouse helped with household chores (e.g., cleaning)	(2) Female reports that spouse helped with cooking or meal preparation	(3) Female reports that spouse 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) Male reports that he helped the children with their homework for school	(8) Male reports that he helped the children prepare for school in the morning
T1 x Poultry	0.105**	0.080*	0.222***	0.041	0.066	0.189***	-0.076	0.027
	(0.044)	(0.042)	(0.048)	(0.049)	(0.051)	(0.055)	(0.054)	(0.050)
	[0.047]	[0.114]	[0.000]	[0.462]	[0.260]	[0.003]	[0.254]	[0.594]

T1 x Cash	0.112**	0.109**	0.289***	0.156**	0.045	0.225***	0.102*	0.115*
	(0.055)	(0.051)	(0.062)	(0.072)	(0.064)	(0.060)	(0.059)	(0.064)
	[0.068]	[0.067]	[0.000]	[0.067]	[0.481]	[0.001]	[0.095]	[0.095]
T2 x Poultry	0.063	0.076**	0.065	0.042	0.029	0.126**	-0.001	-0.012
	(0.046)	(0.039)	(0.050)	(0.056)	(0.049)	(0.055)	(0.059)	(0.062)
	[0.385]	[0.197]	[0.385]	[0.729]	[0.735]	[0.192]	[0.984]	[0.973]
T2 x Cash	0.010	0.014	0.103*	0.077	-0.005	0.163***	-0.021	-0.053
	(0.049)	(0.037)	(0.055)	(0.058)	(0.048)	(0.062)	(0.057)	(0.067)
	[0.921]	[0.921]	[0.256]	[0.509]	[0.921]	[0.073]	[0.921]	[0.866]
T3	0.119***	0.094***	0.155***	0.161***	0.130***	0.154***	-0.012	-0.049
	(0.045)	(0.035)	(0.048)	(0.049)	(0.043)	(0.053)	(0.046)	(0.054)
	[0.012]	[0.011]	[0.006]	[0.006]	[0.008]	[0.008]	[0.802]	[0.411]
Average effect of T1	0.107***	0.093**	0.252***	0.098**	0.055	0.203***	0.014	0.070
	(0.040)	(0.036)	(0.044)	(0.049)	(0.045)	(0.049)	(0.045)	(0.047)
	[0.021]	[0.021]	[0.000]	[0.072]	[0.258]	[0.000]	[0.749]	[0.183]
Average effect of T2	0.038	0.047	0.085*	0.060	0.013	0.147***	-0.011	-0.032
	(0.040)	(0.032)	(0.044)	(0.048)	(0.041)	(0.052)	(0.048)	(0.053)
	[0.558]	[0.381]	[0.209]	[0.422]	[0.820]	[0.041]	[0.820]	[0.725]
Average effect of poultry	0.083**	0.078**	0.141***	0.041	0.047	0.157***	-0.037	0.007
	(0.038)	(0.032)	(0.041)	(0.044)	(0.041)	(0.049)	(0.047)	(0.047)
	[0.057]	[0.044]	[0.006]	[0.465]	[0.402]	[0.006]	[0.489]	[0.887]
Average effect of cash	0.061	0.062*	0.197***	0.117**	0.020	0.194***	0.041	0.031
	(0.042)	(0.035)	(0.047)	(0.052)	(0.045)	(0.052)	(0.047)	(0.053)
	[0.232]	[0.161]	[0.000]	[0.070]	[0.651]	[0.001]	[0.515]	[0.641]
Test: T1 = T2	(0.084)	(0.199)	(0.000)	(0.429)	(0.344)	(0.158)	(0.584)	(0.038)
	[0.224]	[0.319]	[0.001]	[0.490]	[0.459]	[0.317]	[0.584]	[0.150]
Test: T2 = T3	(0.063)	(0.156)	(0.137)	(0.040)	(0.007)	(0.868)	(0.989)	(0.755)
	[0.168]	[0.250]	[0.250]	[0.159]	[0.053]	[0.989]	[0.989]	[0.989]
Test: T1 = T3	(0.782)	(0.981)	(0.047)	(0.215)	(0.117)	(0.243)	(0.561)	(0.018)
	[0.894]	[0.981]	[0.187]	[0.388]	[0.313]	[0.388]	[0.748]	[0.147]
Test: Poultry = Cash	(0.574)	(0.645)	(0.206)	(0.126)	(0.550)	(0.349)	(0.100)	(0.624)
	[0.645]	[0.645]	[0.548]	[0.502]	[0.645]	[0.645]	[0.502]	[0.645]
Mean of control (T4)	0.175	0.135	0.337	0.299	0.231	0.526	0.440	0.678
N	1,304	1,304	1,304	1,026	1,025	1,024	886	889

Note: 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. 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 9.11c: Male's involvement in domestic tasks in the past 3 days: Subsample of less poor households

	(1) Female reports that spouse helped with household chores (e.g., cleaning)	(2) Female reports that spouse helped with cooking or meal preparation	(3) Female reports that spouse 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) Male reports that he helped the children with their homework for school	(8) Male reports that he helped the children prepare for school in the morning
T1	0.123**	0.100***	0.122***	0.125***	0.123***	0.150***	0.004	-0.008
	(0.048)	(0.038)	(0.046)	(0.044)	(0.042)	(0.044)	(0.048)	(0.046)
	[0.015]	[0.015]	[0.015]	[0.013]	[0.013]	[0.006]	[0.940]	[0.940]
T2	0.022	0.013	0.034	0.029	0.022	0.086*	0.038	0.046
	(0.046)	(0.033)	(0.042)	(0.043)	(0.040)	(0.045)	(0.050)	(0.046)
	[0.695]	[0.695]	[0.695]	[0.695]	[0.695]	[0.467]	[0.695]	[0.695]
T3	0.099**	0.108***	0.134***	0.132***	0.142***	0.132***	0.056	0.039
	(0.045)	(0.036)	(0.046)	(0.045)	(0.043)	(0.047)	(0.046)	(0.046)
	[0.039]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.259]	[0.392]
Test: T1 = T2	(0.034)	(0.016)	(0.052)	(0.034)	(0.008)	(0.104)	(0.483)	(0.220)
	[0.068]	[0.064]	[0.084]	[0.068]	[0.060]	[0.139]	[0.483]	[0.251]
Test: T2 = T3	(0.082)	(0.005)	(0.027)	(0.024)	(0.002)	(0.282)	(0.706)	(0.882)
	[0.131]	[0.019]	[0.054]	[0.054]	[0.014]	[0.377]	[0.807]	[0.882]
Test: T1 = T3	(0.594)	(0.828)	(0.816)	(0.876)	(0.646)	(0.667)	(0.266)	(0.276)
	[0.876]	[0.876]	[0.876]	[0.876]	[0.876]	[0.876]	[0.876]	[0.876]
Mean of control (T4)	0.175	0.139	0.383	0.275	0.194	0.540	0.413	0.689
N	1,173	1,173	1,172	947	947	948	841	844

Note: 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. 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."

10 Limitations

In this chapter, we summarize the key limitations of the SPIR impact evaluation and analysis presented in this report. Every impact evaluation suffers from some limitations of breadth, which is required to give the evaluation focus. In addition, we faced constraints to what could be captured in the survey rounds to stay within the survey budget and to keep interviews from becoming so long that the accuracy of responses would be significantly affected. Other limitations arose from disruptions to plans for data collection. Despite the limitations listed here, the impact evaluation was successfully completed without substantial disruptions or interference.

- Despite SPIR having Purpose 4 of its implementation structure devoted to investments in natural resource management, this study did not evaluate natural resource management or watershed outcomes. In addition, the research design did not allow for analysis of the effects of community assets generated by PSNP public works programming. These topics are not addressed because it was necessary to develop a focused impact evaluation design. The agreed-upon topics to cover were related to livelihoods, nutrition, gender dimensions of well-being, and mental health.
- The SPIR evaluation analyzes bundles of interventions implemented jointly, and thus identifying the contributory effect of separate interventions is challenging. For example, while we suspect that the male engagement groups in N* were influential to improving gender equitable attitudes and roles, we are unable to disentangle the male engagement groups from other N* activities.
- There was a modest amount of cross-over or contamination of the control group during the SPIR study because sample households in control communities were able to access some components of the SPIR programming, including the aspirations film screenings, some VESA sessions, and some nutrition interventions. Overall, the extent of the cross-over is quite limited for a randomized evaluation of a program of this size and complexity.
- Households in the sample were exposed to a range of shocks linked to pests (locusts and armyworm), the COVID-19 pandemic, and conflict that may have attenuated impacts, rendering it more challenging to identify the effects of SPIR programming.
- COVID-19 and to civil unrest caused some disruptions in the timing of survey data collection for the evaluation, which mainly affected interpersonal group therapy (IPT-G) outcomes. For example, we planned a follow-up survey in March 2020 for households with members who were eligible for the first round of IPT-G intervention, but we were unable to conduct in-person interviews because of the start of the COVID-19 pandemic. After a delay of a few weeks, we were able to interview most of the intended respondents by phone, but survey responses about topics like experience with depressive symptoms may have been affected by asking the questions in a phone interview. Similarly, plans for a limited follow-up survey after the second round of IPT-G sessions had to be dropped in Amhara region because of unrest. We were able to complete the interviews in Oromia, but our ability to measure the impact of the IPT-G sessions on depression was hampered by the smaller sample.

- Also, the endline survey was postponed from June 2020 to March 2021 because of the pandemic. This delay may have reduced measured impacts of SPIR for some program components that were delivered earlier or were more intensive earlier in the project.
- In terms of measuring women's decision-making, we were constrained in survey length at endline, so we only included decision-making with respect to productive decisions in agriculture and not other domains such as childcare or nutrition that may have been affected as a result of SPIR.
- The endline survey did not determine if the respondents were pregnant, so we are unable to determine whether women fast during pregnancy. Similarly, the sample had relatively few women who were lactating, so fasting for these women was also not analyzed.

11 Key Results and Recommendations

11.1 Key Results

- Financial inclusion
    - SPIR programming is having large positive effects on engagement in savings for households and women.
    - Extremely poor households also show evidence of substantial increases in membership in VESAs and the probability of reporting any savings.
    - There is some evidence that women's savings was declining over time (in all arms), and there are no substantial effects on credit access.
- Diversifying livelihoods
    - SPIR had a range of positive effects, particularly on livestock-related production (particularly for cash and poultry households), and financial inclusion (for all households).
    - Medium-size cash and asset transfers ($200) had substantial medium-term positive effects on household livelihood assets and income from livestock value chains.
    - Engagement in non-agricultural businesses or wage work is consistently low for households in the sample, and this remains consistent over time and following the implementation of SPIR interventions.
    - For extremely poor households (who received cash or poultry transfers), we see evidence of persistent and large increases in livestock assets and engagement in livestock production: these effects are concentrated in poultry for poultry recipients, and are concentrated in nonpoultry livestock for cash recipients.
    - There is no clear evidence that either poultry or cash transfers were differentially effective compared to the other forms of transfer in terms of effects on revenue and consumption.
    - However, there is no robust evidence that SPIR led to any increase in ownership of other durable goods or any increased consumption two years post-transfer.
    - Less poor households who did not receive transfers largely do not show any substantial shifts in livestock assets or production.
    - However, these less poor households also show substantial increases in savings as well as some enhanced access to credit, and there is some weak evidence of improvement in housing characteristics.

- Poverty reduction
    - Despite the positive effects on savings and on some measures of engagement in livestock production, consumption remained consistent for households participating in SPIR vis-à-vis the control arm, as did ownership of other types of household goods and assets.
    - Transfers did not seem to be large enough to allow households to graduate from poverty.
- Nutrition
    - The TTC component of SPIR led to increased contact with healthcare workers and had measurable success in promoting participation in CPNP when the need was recognized. But the frequency of weighing and of direct contact with caregivers was below the target, limiting the impact on nutritional outcomes.
    - While most Orthodox women follow fasting customs, this does not translate to reduction in milk for children since the supply of milk is not easily deferred. Young children do not consume meat when adults fast, but since they seldom consume meat at any time this has little effect on child dietary patterns. Fasting has mixed effects on egg consumption. There is some reduction of egg consumption for children when adults fast, but this is not complete. Egg marketing is also reduced during fast periods, but it is not clear if the eggs are sold later or allowed to hatch.
    - Proper IYCF is seriously hindered by late introduction of semi-solid or solid foods. Less than half of the children 6–8 months were provided solid food and even a third of the children 9–12 months had not yet received any foods other than breastmilk.
- Gender dynamics and women's well-being
    - Male engagement succeeded in motivating increased male participation in a range of domestic tasks from a low base and improved men's gender equitable attitudes.
    - SPIR programming did not lead to sustained improvements in marital relationship dynamics nor women's empowerment. Impacts that appeared at midline, in part due to the livelihood transfers, have dissipated by endline (two years after the distribution of the livelihood transfer). Marital relations have in fact worsened for the extremely poor households that had received the poultry transfer two years prior.
    - Among the extremely poor households, SPIR programming combined with livelihood transfers decreased women's symptoms of depression.

11.2 Recommendations

- Livelihoods
    - Recommendations for future programming include continuing cash transfers (given that they are more cost-effective to implement, vis-à-vis poultry transfers) and exploring the feasibility of a transfer large enough to enable households to access a sustainable path out of poverty.

- VESAs offer a valuable and widely adapted platform for savings, but future programs could explore how to stimulate more access to credit financing.
- Interventions targeting enhanced aspirations were largely not effective and should not be scaled up.

- Nutrition
  - Renewed efforts are needed to test tailored messages from trusted sources regarding the timing and composition of diets to assist in improving weaning.

- Gender dynamics and women's well-being
  - Engaging men through N* activities is a promising strategy for improving their attitudes and roles and should be continued.
  - Renewed focus is needed on activities to sustainably improve women's empowerment, while ensuring that men are also sensitized in order to not create tensions or backlash.
  - The combination of T1 and T2 with livelihood transfers is a promising strategy for improving women's mental health among extremely poor households.

12 Conclusion

SPIR is an ambitious graduation model program seeking to expand the evidence on integrated strategies to reduce poverty by improving financial inclusion, livelihoods, mental health, gender norms, empowerment for women, and nutrition for their children. Within the evidence on the impact of graduation model programs, SPIR's approach relies less on large cash or asset transfers, and focused more on strengthening a broad set of services related to financial inclusion, business development, health (including mental health), maternal and child nutrition, and changing gender norms around women's agency and access to markets and men's roles in household tasks. In this sense, SPIR was designed as a gender- and nutrition-sensitive approach, particularly compared to the graduation model Targeting the Ultra Poor (TUP) program developed by BRAC and tested in the six-country study reported in Banerjee et al. (2015). Another substantial difference in design between the TUP program and SPIR is that asset transfers in TUP were roughly $1,200 per household above the monthly consumption support, whereas only the poorest half of households in two treatment arms in SPIR received cash or in-kind transfers of $200 (and all received six months of annual consumption support). The results of this endline survey analysis provide evidence about whether SPIR's approach of improving service delivery and better supporting transformation of outcomes beyond the economic sphere of the household has the potential to lead to greater improvements in child nutritional status, women's empowerment, or mental health.

The endline survey documented that households across the SPIR study area faced numerous significant shocks, particularly in the last 1.5 years of the project, including the COVID-19 pandemic and associated restrictions, pest infestations (desert locusts and fall armyworm), droughts and flooding and, in some kebeles, civil unrest. Some of these shocks, particularly COVID-19, led to some disruption in program delivery, while others led to income losses for study households. Nonetheless, estimates show that the prevalence of these shocks are relatively balanced across study treatment arms.

The endline survey also adds to the evidence from the midline survey that a defining feature of the implementation of the SPIR project is phased rollout of project components, including a cascade of training topics covered in VESA group meetings over the years of the project and phased rollouts of CHF training, group therapy to address depression, and male engagement sessions to promote male participation in household tasks. This phased approach is programmatically practical for a program that relies on regular trainings with project households, but it was not known before now whether the phased approach would produce an accumulation of knowledge and service access that contribute to a growing stream of benefits or would perhaps lead to smaller impacts because services and resources are spread over a longer period. In addition, some key program components, such as maternal nutrition and IYCF counseling and male engagement trainings reached large shares of participants but were not nearly universal.

The results of this endline analysis from the SPIR impact evaluation shows that SPIR had significant impacts on a broad set of livelihood, diet, mental health, and empowerment outcomes, though these effects were somewhat piecemeal in many of these outcome families. The livelihood results show that SPIR had a range of positive effects, particularly on livestock-related production (particularly for cash and poultry households), and on financial inclusion (for all households). For extremely poor households (who were the poultry and cash-transfer recipients), we see evidence of persistent and large increases in livestock assets and engagement in livestock production: these effects are concentrated in poultry for poultry recipients and in nonpoultry livestock for cash recipients. Extremely poor households also show

evidence of substantial increases in membership in VESAs and the probability of reporting any savings. There is, however, no robust evidence of any increase in ownership of other durable goods (in a context in which ownership of these goods is rapidly increasing across the sample), or any increased consumption two years post-transfer. For less poor households who did not receive transfers but were exposed to SPIR programming, we largely do not observe any substantial shifts in livestock assets or production, but there is an increase in savings as well as some enhanced access to credit.

Impacts on diets, nutrition and health are mixed. The SPIR program increased access to health services including visits by Health Development Army volunteers, BCC exposure, food demonstrations, and WASH. This despite the intervening COVID-19 strain on healthcare resources and the decreased mobility that the pandemic imposed. However, no indicator of access to health services studied indicated more than half the target population participating. Moreover, despite the innovating Timed and Targeted Counseling, key measures of IYCF such as the age at which semi-solid or solid foods are introduced or child diet diversity have not improved in the communities where the intervention has been prioritized. Since proper complementary feeding is an essential element of nutritional care, this barrier likely contributes to the stagnating stunting rate. There are no indications of improvement in anthropometric outcomes, other than an improvement in weight-for-age in selected treatment arms. Thus, SPIR made modest inroads in responding to underweight when it is identified. But, again, with child weighing apparently infrequent, this improved service delivery does not fully cover the eligible population. Thus, identifying the gaps in coverage as well as improving the messaging on weaning appear to be ways that the initial progress in intensified nutritional service delivery can achieve progress in improving nutritional outcomes.

The study also investigated the impacts of the SPIR intervention on men's and women's mental health, marital dynamics, women's agency, and gender equitable attitudes and roles. We find that the SPIR intervention, and in particular T1 and T3, improved men's gender equitable attitudes and roles. These impacts are significantly different from T2, indicating that N* was needed for these transformative changes. However, we do not see any improvements in other dimensions of empowerment particularly related to women's decision-making or self-efficacy.

Impacts on mental health were mixed. We find no evidence that the SPIR intervention improved the mental health of the primary male or primary female at endline for the full sample or less poor sample. However, among the subsample of extremely poor households, the combination of T1 and poultry and the combination of T2 and cash transfers led to decreases in women's (but not men's) depressive symptoms as measured by the probability of a woman reporting mild or moderate-to-severe depressive symptoms, and to decreases in reported unhappiness. The average impact of T2 across poultry and cash is large and significant, as is the average impact of poultry across T1 and T2. Impacts of T2 (L*+N) are significantly different from T3 (L+N*), suggesting that L* is needed for improvements in women's mental health among the extremely poor subsample.

In terms of marital dynamics and intimate partner violence (IPV), we find no impacts of the SPIR intervention (T1, T2, T3) on these outcomes for either the full sample or subsample of less poor households. For the subsample of extremely poor households, there is some indication that marital relations are worse, especially among poultry households where IPV has increased, and the primary male is less likely to report that his spouse respects him.

Overall, these results show a number of positive impacts of SPIR across with important gains in livestock assets and financial inclusion, weight gain in children, men's gender equitable attitudes and roles, some dimensions of mental health, and male participation in some household tasks. Many of these results reflect a broadening of impacts into diets, nutrition, mental health, and gender equitable norms that were missing in other graduation model programs. However, the limited impacts on improvements in assets and consumption probably result from not providing more substantial resource transfers. This likely limited the potential of SPIR to contribute to poverty reduction or poverty graduation. Nonetheless, we expect that the lessons from this impact evaluation will contribute to improvements in future programming conducted by World Vision, CARE and ORDA and will provide lessons for the government and other implementing partners and stakeholders, and may contribute to improvements in the implementation of the fifth phase of Ethiopia's Productive Safety Net Programme and related complementary interventions.

References

Adewuya A. O., Ola B. A., and O.O. Afolabi. 2006. Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students. *J Affect Disord.* 96:89–93.

Alderman, Harold, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Heleene Tambet and Alemayehu Seyoum Taffesse. 2020a. Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA): Midline Report. Washington DC: International Food Policy Research Institute.

Alderman, Harold, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Michael Mulford and Alemayehu Seyoum Taffesse. 2020b. A Randomized Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Graduation Program in Ethiopia: Pre-Analysis Plan. Washington DC: International Food Policy Research Institute.

Alderman, Harold, Fantu Bachewe, Daniel Gilligan, Melissa Hidrobo, Natasha Ledlie, Gayathri Ramani and Alemayehu Seyoum Taffesse. 2019. Impact Evaluation of the Strengthen PSNP4 Institutions and Resilience (SPIR) Development Food Security Activity (DFSA): Baseline Report. Washington DC: International Food Policy Research Institute.

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

Angelucci, Manuela. 2020. A New Hope: The Economic Impact of Mental Health Care in India. Presented at IFPRI, March 2020.

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.

Banerjee, Abhijit, Esther Duflo, Raghabendra Chattopadhyay, and Jeremy Shapiro. 2016. "Long Term Impact of a Livelihood Intervention: Evidence from West Bengal." Working Paper.

Banerjee, Abhijit, Dean Karlan, Robert Osei, Hannah Trachtman and Christopher Udry. 2019. Unpacking a Multi-Faceted Program to Build Sustainable Income for the Very Poor. NBER Working Paper No. 24271.

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.

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): 824-859.

Bass J, Neugebauer R, Clougherty KF, Verdeli H, Wickramaratne P, Ndogoni L, Speelman L, Weissman M, and P Bolton. 2006. Group interpersonal psychotherapy for depression in rural Uganda: 6-month outcomes. *The British Journal of Psychiatry* 188, no. 6 (2006): 567-573.

Bernard T., Dercon S., Orkin K., Taffesse A. S. 2019. Parental Aspirations for Children's Education: Is There a "Girl Effect"? Experimental Evidence from Rural Ethiopia. *American Economics Association Papers and Proceedings* 109: 127-132.

Bernard T., Dercon S., Orkin K., Taffesse A. S. 2017. The Future in Mind: Long-Run Impact of an Aspirations Intervention in Rural Ethiopia. Washington DC: International Food Policy Research Institute.

Black M.M., Baqui A.H., Zaman K., El Arifeen S., Black R.E. 2009. Maternal depressive symptoms and infant growth in rural Bangladesh. *American Journal of Clinical Nutrition* 89(3):951S–957S.

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.

Bolton, P, Bass, J, Betancourt, T, Speelman, L, Onyango, G, Clougherty, KF, Neugebauer, R, Murray, L and H Verdeli. 2007. Interventions for depression symptoms among adolescent survivors of war and displacement in northern Uganda: a randomized controlled trial. *JAMA* 298(5): 519-527.

Bolton, P, Bass, J, Betancourt, T, Speelman, L, Onyango, G, Clougherty, KF, Neugebauer, R, Murray, L and H Verdeli. 2003. Group interpersonal psychotherapy for depression in rural Uganda: A randomized controlled trial. *JAMA* 289(23): 3117-3124.

Buller, A. M., Peterman, A., Ranganathan, M., Bleile, A., Hidrobo, M., & Heise, L. 2018. A mixed-method review of cash transfers and intimate partner violence in low- and middle-income countries. *World Bank Research Observer*.

Cantril, H. 1965. The pattern of human concerns. New Brunswick, Rutgers University Press.

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

Christian, C, Hansel, L. and C Roth. 2019. Income shocks and suicides: Causal evidence from Indonesia. *Review of Economics and Statistics* 101(5): 905-920.

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.

EPHI and ICF, 2019, Ethiopia Mini Demographic and Health Survey, Rockville, Maryland, Ethiopian Public Health Institute, ICF, https://dhsprogram.com/pubs/pdf/PR120/PR120.pdf.

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.

Gelayea, B, MA Williams, S Lemma, N Deyessa, Y Bahretibeb, T Shibre, D Wondimagegn, A Lemenhe, J Fann, AV Stoep, and X-H A Zhou. 2013. Validity of the Patient Health Questionnaire-9 for Depression Screening and Diagnosis in East Africa. *Psychiatry Research* 210(2):1-23.

Haushofer J, R Mudida and J Shapiro. 2019. The Comparative Impact of Cash Transfers and Psychotherapy on Psychological and Economic Well-being.

Haushofer, J., and J. Shapiro. 2016. The short-term impacts of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. *Quarterly Journal of Economics* 131(4): 1973-2042.

Haushofer, J., and J. Shapiro. 2018. The long-term impact of unconditional cash transfers: experimental evidence from Kenya. Busara Center for Behavioral Economics, Nairobi, Kenya.

Heath R, M Hidrobo and S. Roy. Forthcoming. Cash transfers, polygamy, and intimate partner violence: Experimental evidence from Mali. *Journal of Development Economics*. forthcoming.

Hidrobo, M, A Peterman and L Heiss. 2016. The effect of cash, vouchers and food transfers on intimate partner violence: Evidence from a randomized experiment in Northern Ecuador. *American Economic Journal: Applied Economics* 8(3): 284–303.

Jodlowski, M., Winter-Nelson, A., Baylis, K. and Goldsmith, P.D., 2016. Milk in the data: food security impacts from a livestock field experiment in Zambia. World Development, 77, pp.99-114.

Kling, J., Liebman, J., & Katz, L. 200). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. 2001. The PHQ-9. *Journal of General Internal Medicine*, *16*, 606–613.

Lund, C., A. Breen, A. J. Flisher, R. Kakuma, J. Corrigall, J. A. Joska, L. Schwartz, and V. Patel. 2010. "Poverty and CommonMentalDisorders in Lowand Middle Income Countries:ASystematicReview." Social Science & Medicine 71 (3): 517–28.

Lund, C., Brooke-Sumner, C., Baingana, F., Baron, E. C., Breuer, E., Chandra, P., … Saxena, S. 2018. Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. *The Lancet Psychiatry*, *5*(4), 357–369. https://doi.org/10.1016/S2215-0366(18)30060-9

McIntosh, Craig and Andrew Zeitlin. 2018. Benchmarking a child nutrition program against cash: Experimental evidence from Rwanda. IPA Working Paper.

McKenzie, D. 2012. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics* 99(2): 210-221.

Mullally, C., Rivas, M. and McArthur, T. 2021. Using Machine Learning to Estimate the Heterogeneous Effects of Livestock Transfers. Amer. J. Agr. Econ., 103: 1058-1081.

Nguyen, Phuong Hong, Jed Friedman, Mohini Kak, Purnima Menon and Harold Alderman. 2018. Maternal depressive symptoms are negatively associated with child growth and development: Evidence from rural India. *Maternal and Child Nutrition* 14:e12621.

Ong, Q, W Theseira and I.Y.H. Ng. 2019. Reducing debt improves psychological functioning and changes decision-making in the poor. *Proceedings of the National Academy of Sciences* 116(15): 7244-7249.

Phadera, L., Michelson, H., Winter-Nelson, A. and Goldsmith, P., 2019. Do asset transfers build household resilience? Journal of Development Economics, 138, pp.205-227.

Rawlins, R., S. Pimkina, C. B. Barrett, S. Pedersen, and B. Wydick. 2014. Got milk? The impact of

Heifer International's livestock donation programs in Rwanda on nutritional outcomes. Food Policy 44, 202-213.

Roy, S, M Hidrobo, J Hoddinott and A Ahmed. 2019. Transfers, behavior change communication, and intimate partner violence: Post-program evidence from rural Bangladesh. *Review of Economics and Statistics*. 101(5): 865-877.

Taffesse, Alemayehu Seyoum, and Fanaye Tadesse. 2017. Pathways Less Explored—Locus of Control and Technology Adoption," *Journal of African Economies* 1–37.

Wachs, T.D., Black, M.M. and Engle, P.L. 2009. Maternal depression: a global threat to children's health, development, and behavior and to human rights. Child Development Perspectives, 3(1), pp.51-59. WHO and UNICEF 2017. Global Nutrition Monitoring Framework. Operational Guidelines for meeting Targets for 2025. WHO Geneva.

WHO. 2016. Ethical and safety recommendations for intervention research on violence against women. Building on lessons from the WHO publication *Putting women first: ethical and safety recommendations for research on domestic violence against women.* Geneva: World Health Organization.

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

Appendix A: SPIR Endline Survey Questionnaire

The endline survey questionnaire is not included in this version of the endline report in order to keep the file size manageable. The complete version of this report, including the questionnaire, is available upon request.

Appendix B: Estimated Impacts on Dietary Diversity and Food Security, Extremely Poor Sample

Table B.8.1 Children's dietary diversity: Extremely poor households
Children of age 6-23 months

	(1) Number of food groups (of 8) consumed	(2) Child meets the minimum meal frequency (MMF)	(3) Child consumed grains, roots or tubers	(4) Child consumed legumes or nuts	(5) Child consumed dairy	(6) Child consumed fish or meat	(7) Child consumed eggs	(8) Child consumed vitamin A-rich fruits or vegetables	(9) Child consumed other fruits or vegetables
Child age in months	0.018***	0.020***	0.024***	0.009***	0.002***	0.001*	0.001***	0.001***	0.000
	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Male child	0.064*	-0.015	0.025	-0.006	0.002	-0.004	0.012*	0.013*	-0.009*
	(0.037)	(0.025)	(0.016)	(0.015)	(0.013)	(0.004)	(0.007)	(0.007)	(0.005)
T1	-0.017	0.016	0.005	-0.008	0.001	-0.001	-0.001	-0.004	-0.001
	(0.021)	(0.024)	(0.015)	(0.009)	(0.005)	(0.002)	(0.002)	(0.003)	(0.002)
T2	-0.031	0.005	-0.018	-0.013	-0.005	0.000	-0.000	-0.002	-0.002
	(0.022)	(0.024)	(0.018)	(0.008)	(0.005)	(0.003)	(0.003)	(0.004)	(0.001)
T3	-0.034	0.012	-0.022	-0.010	-0.002	-0.001	-0.001	-0.002	-0.001
	(0.024)	(0.023)	(0.017)	(0.009)	(0.005)	(0.002)	(0.003)	(0.004)	(0.002)
Endline	0.286***	0.378***	0.127***	0.023	0.062*	-0.000	0.020	-0.013	-0.018**
	(0.093)	(0.062)	(0.046)	(0.031)	(0.037)	(0.012)	(0.019)	(0.017)	(0.008)
T1 x Endline	0.180	-0.021	0.013	-0.007	0.079*	-0.007	0.012	0.021	0.017
	(0.131)	(0.090)	(0.067)	(0.043)	(0.047)	(0.016)	(0.024)	(0.023)	(0.010)
		[0.878]	[0.878]	[0.878]	[0.413]	[0.878]	[0.878]	[0.878]	[0.413]
T2 x Endline	0.249*	-0.004	0.081	-0.010	0.123**	-0.002	-0.006	0.025	0.011
	(0.130)	(0.084)	(0.063)	(0.051)	(0.049)	(0.019)	(0.030)	(0.025)	(0.009)
		[0.966]	[0.543]	[0.966]	[0.112]	[0.966]	[0.966]	[0.651]	[0.543]
T3 x Endline	0.038	-0.096	-0.020	-0.037	0.069	-0.002	-0.014	0.002	-0.003
	(0.139)	(0.087)	(0.072)	(0.048)	(0.058)	(0.017)	(0.025)	(0.023)	(0.012)
		[0.930]	[0.930]	[0.930]	[0.930]	[0.930]	[0.930]	[0.930]	[0.930]
Pooled T1 x Endline	0.109	-0.058	-0.003	-0.022	0.074	-0.005	-0.001	0.012	0.007
	(0.116)	(0.076)	(0.059)	(0.038)	(0.045)	(0.015)	(0.022)	(0.020)	(0.010)

and T3 x Endline		[0.913]	[0.972]	[0.913]	[0.830]	[0.972]	[0.972]	[0.913]	[0.913]
Test: T1xEndline = T2xEndline	(0.606)	(0.843) [0.945]	(0.298) [0.945]	(0.945) [0.945]	(0.332) [0.945]	(0.759) [0.945]	(0.495) [0.945]	(0.879) [0.945]	(0.497) [0.945]
Test: T2xEndline = T3xEndline	(0.134)	(0.271) [0.598]	(0.155) [0.598]	(0.637) [0.850]	(0.342) [0.598]	(0.973) [0.973]	(0.766) [0.876]	(0.374) [0.598]	(0.182) [0.598]
Test: T1xEndline = T3xEndline	(0.309)	(0.405) [0.811]	(0.655) [0.848]	(0.540) [0.848]	(0.848) [0.848]	(0.747) [0.848]	(0.204) [0.811]	(0.406) [0.811]	(0.088) [0.700]
Mean of control (T4)	1.846	0.392	0.708	0.236	0.074	0.017	0.032	0.045	0.026
N	2,390	1,150	2,390	2,390	2,390	2,390	2,390	2,390	2,390

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) are clustered at the kebele level. All models control for woreda-level fixed effects and the cluster mean of the baseline value of the outcome. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 B.8.2: Women's dietary diversity: Extremely poor households

	(1) Women's dietary diversity score (1-10)	(2) Met minimum dietary diversity for women (MDD-W)	(3) Primary female consumed tubers and grains	(4) Primary female consumed pulses	(5) Primary female consumed nuts and seeds	(6) Primary female consumed dairy	(7) Primary female consumed meat, fish, poultry	(8) Primary female consumed eggs	(9) Primary female consumed green leafy veg-etables	(10) Primary female consumed vitamin A- rich fruits and veg-etables	(11) Primary female consumed other vegetables	(12) Primary female consumed other fruits
T1	0.198*	0.032*	-0.014	-0.008	0.065*	0.007	0.021	0.004	0.022	0.019	0.082*	0.002
	(0.101)	(0.019)	(0.013)	(0.029)	(0.036)	(0.024)	(0.014)	(0.015)	(0.018)	(0.020)	(0.046)	(0.012)
		[0.336]	[0.546]	[0.840]	[0.336]	[0.840]	[0.405]	[0.840]	[0.516]	[0.558]	[0.336]	[0.840]
T2	0.047	0.027	-0.025*	-0.050**	0.002	0.030	0.014	0.001	0.014	0.026	0.021	0.014
	(0.107)	(0.021)	(0.015)	(0.025)	(0.034)	(0.025)	(0.012)	(0.012)	(0.018)	(0.024)	(0.046)	(0.017)
		[0.490]	[0.465]	[0.465]	[0.953]	[0.490]	[0.490]	[0.953]	[0.593]	[0.490]	[0.780]	[0.593]
T3	0.176	0.042*	-0.011	-0.021	0.029	0.026	0.015	0.018	0.018	0.004	0.078*	0.020
	(0.110)	(0.023)	(0.012)	(0.025)	(0.033)	(0.026)	(0.015)	(0.016)	(0.019)	(0.021)	(0.042)	(0.013)
		[0.364]	[0.437]	[0.437]	[0.437]	[0.437]	[0.437]	[0.437]	[0.437]	[0.841]	[0.364]	[0.437]
Pooled effect of N*:	0.187**	0.037**	-0.013	-0.014	0.047	0.017	0.018	0.011	0.020	0.011	0.080**	0.011
T1 or T3	(0.090)	(0.018)	(0.011)	(0.022)	(0.030)	(0.022)	(0.013)	(0.013)	(0.016)	(0.018)	(0.039)	(0.011)
		[0.216]	[0.437]	[0.517]	[0.425]	[0.517]	[0.425]	[0.517]	[0.437]	[0.517]	[0.216]	[0.472]
Test: T1 = T2	(0.161)	(0.805)	(0.517)	(0.195)	(0.081)	(0.321)	(0.597)	(0.859)	(0.649)	(0.752)	(0.174)	(0.465)
		[0.859]	[0.859]	[0.716]	[0.716]	[0.859]	[0.859]	[0.859]	[0.859]	[0.859]	[0.716]	[0.859]
Test: T2 = T3	(0.279)	(0.527)	(0.378)	(0.301)	(0.397)	(0.898)	(0.937)	(0.261)	(0.835)	(0.378)	(0.173)	(0.722)
		[0.828]	[0.728]	[0.728]	[0.728]	[0.937]	[0.937]	[0.728]	[0.937]	[0.728]	[0.728]	[0.937]
Test: T1 = T3	(0.841)	(0.631)	(0.847)	(0.678)	(0.313)	(0.410)	(0.716)	(0.405)	(0.824)	(0.515)	(0.929)	(0.170)
		[0.929]	[0.929]	[0.929]	[0.929]	[0.929]	[0.929]	[0.929]	[0.929]	[0.929]	[0.929]	[0.929]
Mean of control (T4)	2.610	0.057	0.979	0.569	0.207	0.110	0.036	0.045	0.052	0.086	0.495	0.031
N	1,723	1,723	1,723	1,723	1,723	1,723	1,723	1,723	1,723	1,723	1,723	1,723

Note: Estimates from the DFSA SPIR endline survey sample, estimated as an ANCOVA model at the household level. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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 B.8.3 Household's food security: Extremely poor households

	(1) Raw score from 8 FIES questions	(2) Household is moder-ately or severely food insecure based on FIES score	(3) Household is severely food insecure based on FIES score	(4) Household has worried about not having enough food to eat because of a lack of resources	(5) Household has been unable to eat healthy and nutritious food because of a lack of resources	(6) Household has eaten only a few kinds of foods because of a lack of resources	(7) Household has had to skip a meal because there were not enough resources to get food	(8) Household has eaten less than they thought they should because of a lack of resources	(9) Household has run out of food because of a lack of resources	(10) Household has been hungry but not eaten because there were not enough resources for food	(11) Household has gone without eating for a whole day because of a lack of resources
T1	0.403*	0.076*	0.061*	0.062*	0.013	0.068*	0.077**	0.086*	0.023	0.042	0.034
	(0.220)	(0.041)	(0.034)	(0.035)	(0.037)	(0.039)	(0.037)	(0.043)	(0.039)	(0.035)	(0.026)
		[0.134]	[0.134]	[0.134]	[0.716]	[0.134]	[0.134]	[0.134]	[0.620]	[0.290]	[0.270]
T2	0.236	0.041	0.024	0.024	0.026	0.052	-0.004	0.068	0.031	0.041	0.006
	(0.193)	(0.041)	(0.028)	(0.031)	(0.035)	(0.036)	(0.031)	(0.043)	(0.039)	(0.031)	(0.023)
		[0.579]	[0.579]	[0.579]	[0.579]	[0.579]	[0.896]	[0.579]	[0.579]	[0.579]	[0.867]
T3	0.231	0.049	0.041	0.063**	-0.006	0.020	0.047	0.044	0.018	0.040	0.004
	(0.176)	(0.037)	(0.028)	(0.032)	(0.032)	(0.032)	(0.030)	(0.039)	(0.036)	(0.031)	(0.025)
		[0.393]	[0.393]	[0.393]	[0.879]	[0.776]	[0.393]	[0.436]	[0.776]	[0.393]	[0.879]
Pooled effect of N*: T1 or T3	0.317*	0.063*	0.051*	0.063**	0.004	0.044	0.062**	0.065*	0.020	0.041	0.019
	(0.169)	(0.034)	(0.027)	(0.028)	(0.029)	(0.031)	(0.028)	(0.036)	(0.033)	(0.028)	(0.022)
		[0.152]	[0.152]	[0.152]	[0.893]	[0.223]	[0.152]	[0.152]	[0.592]	[0.223]	[0.489]
Test: T1 = T2	(0.456)	(0.421)	(0.242)	(0.300)	(0.748)	(0.679)	(0.030)	(0.692)	(0.832)	(0.975)	(0.231)
		[0.843]	[0.750]	[0.750]	[0.924]	[0.924]	[0.301]	[0.924]	[0.924]	[0.975]	[0.750]
Test: T2 = T3	(0.980)	(0.839)	(0.504)	(0.235)	(0.367)	(0.308)	(0.096)	(0.513)	(0.725)	(0.964)	(0.910)
		[0.964]	[0.855]	[0.855]	[0.855]	[0.855]	[0.855]	[0.855]	[0.964]	[0.964]	[0.964]
Test: T1 = T3	(0.413)	(0.487)	(0.527)	(0.981)	(0.612)	(0.175)	(0.405)	(0.286)	(0.903)	(0.945)	(0.233)
		[0.875]	[0.875]	[0.981]	[0.875]	[0.875]	[0.875]	[0.875]	[0.981]	[0.981]	[0.875]
Mean of control (T4)	3.353	0.437	0.166	0.572	0.724	0.624	0.291	0.438	0.333	0.252	0.122
N	1,748	1,748	1,748	1,746	1,743	1,746	1,744	1,744	1,746	1,744	1,746

Note: Estimates from the DFSA SPIR endline survey sample. Standard errors (in parentheses below treatment effects in panels 1 and 2) 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. P-values on t-tests of equality of treatment effects across arms are presented in parentheses in panel 3. 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.