



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)

Midline Report[§]

First Draft: April 1, 2020
Final Draft: October 12, 2020

Harold Alderman, Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight, Gayathri Ramani,
Alemayehu Seyoum Taffesse, and Heleene Tabet

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

§ This version of the midline report has details of some results redacted (on depression and intimate partner violence) to satisfy embargo policies of some peer-reviewed journals, pending journal publication.

* 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 Verdelli and Paul Bolton contributed to the design of the maternal depression sub-study 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 resilience to shocks and livelihoods 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. 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.

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 midline report of the impact evaluation presents evidence from the first 15 months of the project on program performance and the effectiveness of delivery. The midline survey was conducted from July 25 to October 23, 2019 during which time a total of 3,968 households were interviewed out of the target of 4,082 households for the entire study sample.

SPIR Interventions

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 and Purpose 4 on climate resilience, 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. 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
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 are 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) includes integrated nutrition social behavior change communication (SBCC) as well as water, sanitation and health (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.

The enhanced livelihoods (L*) interventions included all livelihoods activities as well as the following. **Social analysis and action (SAA)** enables 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 including US\$200 worth of poultry start-up inputs and training or a large one-time unconditional cash transfers of equivalent value.

The enhanced nutrition (N*) interventions included all nutrition activities as well as the following. In N* kebeles, BCC activities are delivered through a **Timed and Targeted Counseling (TTC)** model, including lessons on IYCF practices and adolescent and maternal nutrition. TTC is conducted at the household level, while the SBCC included in N activities is conducted at community events. TTC also encourages men to support their wives in childcare. **Community-based Participatory Nutrition Promotion (CPNP)** sessions are two-week intensive feeding sessions for acutely malnourished children that include trainings for their mothers on complementary feeding and caring practices. SPIR N* activities are coordinated by a hired and trained **Community Health Facilitator (CHF)** assigned to each of the N* kebeles to support the HEWs to plan, coordinate, and facilitate health and nutrition activities. The CHF supported local HDA volunteers to conduct 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 Groups (IPT-G)** sessions in each of the N* kebeles to address maternal depression.

Evaluation Design

The impact evaluation uses a clustered randomized controlled trial (RCT) design to learn about the effect of different combinations of the SPIR intervention 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 targeted poorer 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 poorer households, and half of the L* kebeles were randomly selected to receive poultry packages targeted to women in the poorer households.

Lastly, 50% 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 aspiration's treatment were balanced such that approximately 25% 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 Midline Survey Sample

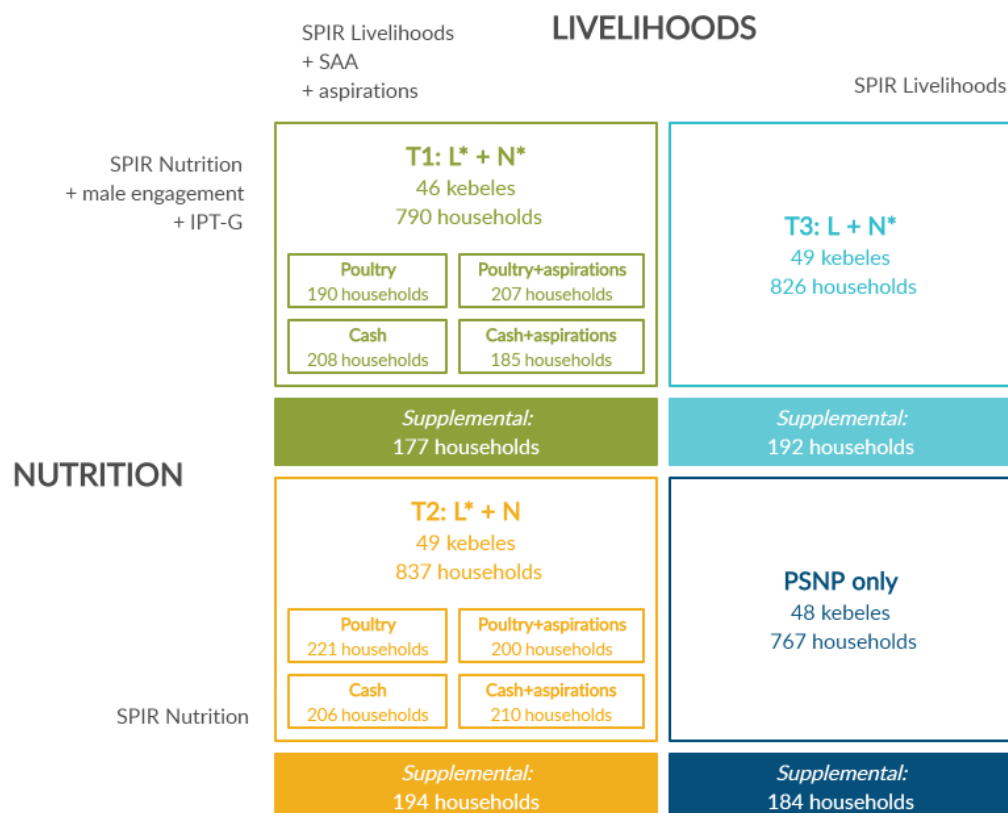
As noted in the Baseline Report, the study takes place in 13 (original) woredas and 192 kebeles across the Amhara and Oromia regions of Ethiopia.¹ The baseline sampling process led to 3,314 households in the sample, or just over 17 households of PSNP4 beneficiaries with at least one child age 0-35 months 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 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 two years. 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 four supplemental households in each kebele (another 768 households) to produce a target midline sample of 4,082 households.

The midline survey enumeration team interviewed a total of 3,968 households out of the target of 4,082 households. Of the 3,314 baseline survey households, 3,220 were able to be located and interviewed at midline, leading to a very low attrition rate of just 2.8 percent. Among the 114 attrited households, the top reasons for attrition included moving outside the survey area (n=44) and refusing consent (n=17). A total

¹ After the baseline survey, two 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.

of 748 supplemental households were added, which is 20 less than planned due to a small number of PSNP4 beneficiary households with children under age 24 months in some kebeles. Balance tests confirmed that the supplemental households were very similar to the original panel households in most characteristics. In total, 3,878 primary female respondents were interviewed as were 3,082 male respondents. Among the Baseline Index Children and Midline Index Children, the number of children in the sample age 0-23 months is 2,075 and the number age 24-60 months is 2,454.

Figure ES.1: SPIR experimental impact evaluation design



Experience with the SPIR program

Midline survey data show that the SPIR project effectively reached households with its complex programming and mostly maintained fidelity to the experimental design, though there are some important gaps. In the main treatment arms (T1, T2, T3), participation in VESA groups was high, but not universal, with 76-79 percent of households reporting that they participated in a VESA group. In the control group (T4), 16 percent of households reported belonging to a VESA group, though these control households may have mistaken common Village Savings and Loan Associations (VSLAs) for the SPIR project VESA groups. Roughly half of women and men in T1-T3 reported attending a VESA meeting in the last 12 months, with women attending 6 meetings on average and men attending 7 meetings on average, suggesting that those who attended at all went to meetings monthly.

Nearly all households identified as poor and assigned to receive the poultry package reported receiving it (93-95), but receipt of the cash livelihood transfers was more limited, with only 60-66 percent of targeted poor households reporting that they received the funds. In the half of L* kebeles that were randomized to receive the aspirations treatment, only 43-48 percent of households reported that a member attended the documentary screening session; respondents may not recall the screening which took place 9 months earlier. Participation in nutrition and WASH training activities varied, but the N* kebeles (T1 and T3) had the highest share of participants; T2 had about two-thirds of the reported participation rate of T1 and T3, and T4 about half of T1 and T3 rate. Conditional on attending a training session, the lessons with highest relative discrepancies between N* and other arms are all related to sanitation: safety of drinking water, handwashing, and proper dishwashing.

Impacts on Livelihood Outcomes

The poultry transfer has had a substantial positive effect on increasing household ownership of and income from poultry. The increase in the probability of owning poultry is 39 percentage points for poultry recipients, compared to 9 percentage points for cash recipients in the comparison treatment. Poultry households also show differentially larger investment in other forms of livestock, showing an increase in the probability of reporting nonpoultry livestock of 8 percentage points. The larger effects of the poultry transfer could be consistent with higher reported receipt rates of the poultry vis-à-vis the cash transfer, though administrative records suggest that rates of receipt were equally high for both transfers. Nonpoor households exposed to enhanced livelihoods programming also increased their engagement in poultry production, showing an increase in the probability of owning poultry of between 8 and 10 percentage points, but there is no effect for households exposed only to the basic livelihoods programming. The latter effect may be suggestive of positive programmatic spillover effects from poor households who are poultry recipients to nonpoor households.

There is a dramatic increase in the probability that women report any savings, driven primarily by women reporting savings in VESAs; this pattern is observed in all three treatment arms, and for both poor and nonpoor women. The increase in the probability of reported female savings is between 36 and 49 percentage points, relative to a mean of 25 percent in the control arm. By contrast, there is little evidence of any shift in credit access. The substantial effect on female savings is consistent with the observed pattern of consistently high participation in VESA groups.

Effects on other dimensions of livelihoods have been minimal. There is little evidence of any effects on engagement in outside labor for the pooled sample or the sample of nonpoor households. For the sample of poor households, there is some evidence of shifts in outside labor. The effects on aspirations outcomes as reported by both men and women are relatively weak. This does not seem to reflect the effects of the targeted cross-randomized aspirations intervention. The aspirations intervention was also characterized by relatively low participation in the data reported by households.

Impacts on Nutrition and Child Welfare Outcomes

The nutrition interventions have shown success in increasing exposure to health services including BCC, food demonstrations, and WASH. The enhanced nutrition programs have led to an average increase in the frequency of measurement of 67 percent. Similarly, communication measures increased significantly, including the provision of feeding advice if a child was weighed (a 25 percent increase over the mean for the control) and participation in camps aimed at addressing acute malnutrition if a child was found to

have a low MUAC (a 300 percent increase over the participation rate for the control). Nevertheless, the frequency of contacts between households and service providers is less than planned and may indicate start-up processes at the time of the midline. Even at the level in June 2019, the SPIR program did have an impact on IYCF, particularly on the provision of colostrum and on breastfeeding. There is a worrisome delay in the introduction of solid or semi-solid foods and an extremely low level of dietary diversity for young children. This is likely a contributing factor to the stunting levels and the decline in stature relative to international norms as children age. At this early stage of the program there is no measurable impact of the improved health-seeking and caregiving behaviors on nutritional status.

Evidence on Gender Norms and Women's Well-being

The SPIR interventions decreased women's depressive symptoms but have no impact on men's depressive symptoms. Treatment arms 2 and 3 lead to significant decreases in the probability that women reported some depressive symptoms. Among poor households, the poultry intervention and Treatment 1 lead to decreases in women's depressive symptoms. Among the nonpoor sample, Treatment 3 lead to a significant decrease in the probability of a woman depressive symptoms compared to Treatment arms 1 and 2, suggesting that for nonpoor households, the particular combination of livelihoods and nutrition interventions had significantly larger impacts than those in other treatment arms.

There are no observed impacts of the SPIR intervention on marital control or intimate partner violence for either the full sample or subsample of non-poor households. However, in the poor subsample where livelihood transfers occurred and where reductions to poverty related stress might be stronger, we find decreases in marital control and IPV as a result of the different interventions. In the poor sample, Treatment 2 leads to significant decreases in physical violence and sexual violence. The effect on physical violence is mainly from the cash sub-intervention, while the effect on sexual violence is mainly from the poultry sub-intervention. Overall, poultry transfers lead to significant reductions in sexual violence, and cash transfers led to significant reductions in marital control.

The SPIR interventions led to significant improvements in the primary male's gender norms, but not in the primary females. In particular, Treatments 1&2 lead to improvements of 7-10 percent on the gender equitable norms scale, and Treatment 2 also improved the probability that the primary male reports that it is not justified to beat his wife by 60 percent and that it is acceptable for a woman to travel alone by 19 percent. Among the poor subsample, impacts are concentrated in the poultry arms, while among the non-poor subsample, impacts occur across all three treatment arms.

Improvements in a woman's decisionmaking with respect to productive activities is observed across all three treatment arms among the full sample. In particular, all three treatment arms lead to significant increases in the probability of a woman having input into most decisions with respect to the use of income generated from productive activities that ranges from 6.1-8.7 percentage points (or 12.8-18.2 percent compared to the control mean).

Table of Contents

EXECUTIVE SUMMARY	I
TABLE OF CONTENTS	VII
LIST OF TABLES	X
LIST OF FIGURES	XII
ACRONYMS.....	XIII
1 INTRODUCTION.....	1
2 THE SPIR INTERVENTIONS.....	4
2.1 L* ENHANCED LIVELIHOOD ACTIVITIES	5
2.2 N* ENHANCED NUTRITION ACTIVITIES.....	6
3 EVALUATION DESIGN	8
3.1 EXPERIMENTAL DESIGN	8
3.1.1 Cash benchmarking of the poultry livelihood intervention.....	9
3.1.2 Sub-study on maternal depression	10
3.2 IMPLEMENTING THE SPIR EVALUATION STUDY DESIGN.....	11
3.3 HOUSEHOLD ELIGIBILITY AND SAMPLING.....	11
3.3.1 Kebele and household eligibility criteria.....	11
3.3.2 Supplemental midline sample.....	12
3.4 EMPIRICAL STRATEGY.....	15
4 MIDLINE DATA COLLECTION	18
4.1 SURVEY INSTRUMENT	18
4.2 ETHICS APPROVAL.....	22
4.3 ENUMERATION TEAMS AND TRAININGS	22
4.4 FIELDWORK EXPERIENCE	22
4.4.1 Achieved sample size and interview completion rate.....	22
4.5 DATA QUALITY AND CLEANING	23
5 EXPERIENCE WITH THE SPIR PROGRAM.....	24
5.1 CORE PROGRAMMING.....	24
5.1.1 Participation in PSNP.....	24
5.1.2 VESA groups	25
5.2 LIVELIHOOD COMPONENTS OF THE PROGRAM	25
5.2.1 Financial literacy and value chains.....	25
5.2.2 Poultry package and cash grant	26
5.2.3 Aspirations treatment.....	29
5.2.4 Social Analysis and Action.....	30
5.3 NUTRITION COMPONENTS OF THE PROGRAM	31
5.3.1 Discussions on health and nutrition.....	31

5.3.2	<i>Male engagement groups</i>	31
5.3.3	<i>Food demonstrations and community health services</i>	32
5.4	CONCLUSIONS	33
6	IMPACTS ON LIVELIHOOD OUTCOMES	43
6.1	INTRODUCTION.....	43
6.1.1	<i>Interpreting tables</i>	43
6.1.2	<i>Participation in livelihood-related SPIR components</i>	44
6.2	LIVESTOCK OWNERSHIP AND SALES	44
6.2.1	<i>Pooled effects by treatment arm</i>	44
6.2.2	<i>Poor households</i>	45
6.2.3	<i>Nonpoor households</i>	46
6.3	SECONDARY OUTCOMES LINKED TO LIVESTOCK OWNERSHIP AND SALES	47
6.3.1	<i>Pooled effects by treatment arm</i>	47
6.3.2	<i>Poor households</i>	47
6.3.3	<i>Nonpoor households</i>	48
6.4	FEMALE SAVINGS	48
6.4.1	<i>Pooled effects by treatment arm</i>	48
6.4.2	<i>Poor households</i>	49
6.4.3	<i>Nonpoor households</i>	49
6.5	MALE CREDIT AND SAVINGS	49
6.5.1	<i>Pooled effects by treatment arm</i>	49
6.5.2	<i>Poor households</i>	50
6.5.3	<i>Nonpoor households</i>	50
6.6	OUTSIDE LABOR	50
6.6.1	<i>Pooled effects by treatment arm</i>	50
6.6.2	<i>Poor households</i>	50
6.6.3	<i>Nonpoor households</i>	51
6.7	EXPENDITURE ON CROP INPUTS	51
6.8	ASPIRATIONS.....	51
6.8.1	<i>Women</i>	51
6.8.2	<i>Men</i>	52
6.9	CONCLUSIONS	52
7	IMPACTS ON NUTRITION AND CHILD WELFARE OUTCOMES	80
7.1	INTRODUCTION.....	80
7.2	ANTHROPOMETRICS	80
7.3	INFANT AND YOUNG CHILD FEEDING (IYCF) PRACTICES	81
7.3.1	<i>IYCF practices</i>	81
7.3.2	<i>Dietary diversity</i>	82
7.4	ACCESS TO AND UTILIZATION OF HEALTH SERVICES	84
7.4.1	<i>Nutrition services</i>	84
7.4.2	<i>Health extension</i>	85
7.5	CHILDCARE AND CHILD DEVELOPMENT	85
7.6	CONCLUSIONS	86

8	EVIDENCE ON GENDER NORMS AND WOMEN’S WELL-BEING	99
8.1	INTRODUCTION.....	99
8.1.1	<i>Interpreting tables.....</i>	99
8.2	DEPRESSION	100
8.2.1	<i>Indicators</i>	100
8.2.2	<i>Pooled effect by treatment arm</i>	100
8.2.3	<i>Poor households.....</i>	100
8.2.4	<i>Nonpoor households</i>	100
8.3	INTIMATE PARTNER VIOLENCE.....	100
8.3.1	<i>Indicators</i>	100
8.3.2	<i>Pooled effect by treatment arm</i>	101
8.3.3	<i>Poor households.....</i>	101
8.3.4	<i>Nonpoor households</i>	101
8.4	GENDER NORMS	101
8.4.1	<i>Indicators</i>	101
8.4.2	<i>Pooled effect by treatment arm</i>	102
8.4.3	<i>Poor households.....</i>	102
8.4.4	<i>Nonpoor households</i>	102
8.5	DECISION-MAKING	103
8.5.1	<i>Indicators</i>	103
8.5.2	<i>Pooled effect by treatment arm</i>	103
8.5.3	<i>Poor households.....</i>	103
8.5.4	<i>Nonpoor households</i>	103
8.6	CONCLUSIONS	104
9	CONCLUSION	115
	REFERENCES.....	118
	APPENDIX A: SPIR MIDLINE SURVEY QUESTIONNAIRE	122
	APPENDIX B: BASELINE BALANCE WITHIN SUBTREATMENT ARMS	123

List of Tables

Table 3.1: Difference in household characteristics at midline: Original vs supplemental sample	14
Table 4.1: Midline Survey Household Questionnaire Modules.....	19
Table 4.2: Midline Survey Female Questionnaire Modules	19
Table 4.3: Midline Survey Male Questionnaire Modules	21
Table 4.4: Summary of midline household sample.....	23
Table 5.1: Participation in PSNP	34
Table 5.2: VESA membership	34
Table 5.3: Financial education and livelihoods (females)	35
Table 5.4: Financial education and livelihoods (males).....	36
Table 5.5: Livelihood transfers and aspirations, only targeted households (females)	36
Table 5.6: Livelihood transfers and aspirations, only targeted households (males)	37
Table 5.7: Social analysis and action (females).....	38
Table 5.8: Social analysis and action (males)	38
Table 5.9: Health and nutrition discussions (females).....	40
Table 5.10: Health and nutrition discussions (males)	41
Table 5.11: Exposure to food demonstrations and health services	42
Table 6.1a: Livestock ownership and sales (primary outcomes)	54
Table 6.1b: Livestock ownership and sales (primary outcomes): Subsample of poor households.....	54
Table 6.1c: Livestock ownership and sales (primary outcomes): Subsample of nonpoor households	56
Table 6.2a: Livestock ownership and sales (secondary outcomes).....	57
Table 6.2b: Livestock ownership and sales (secondary outcomes): sub-sample of poor households.....	57
Table 6.2c: Livestock ownership and sales (secondary outcomes): Sub-sample of nonpoor households ..	59
Table 6.3a: Female's savings.....	60
Table 6.3b: Female's savings: Subsample of poor households	60
Table 6.3c: Female's savings: Subsample of nonpoor households	62
Table 6.4a: Male's credit and savings	63
Table 6.4b: Male's credit and savings: Subsample of poor households	64
Table 6.4c: Male's credit and savings: Subsample of nonpoor households	66
Table 6.5a: Household members' wage work, labor outside the household	67
Table 6.5b: Household members' wage work, labor outside the household: Subsample of poor households	67
Table 6.5c: Household members' wage work/out of farm labor: Subsample of nonpoor households	69
Table 6.6a: Expenditure on crop inputs	71
Table 6.6b: Expenditure on crop inputs: Subsample of poor households.....	71
Table 6.6c: Expenditure on crop inputs: Subsample of nonpoor households	73
Table 6.7a: Female's aspirations	74
Table 6.7b: Female's aspirations: Subsample of poor households.....	74
Table 6.7c: Female's aspirations: Subsample of nonpoor households	75
Table 6.8a: Male's aspirations.....	77
Table 6.8b: Male's aspirations: Subsample of poor households	77
Table 6.8c: Male's aspirations: Subsample of nonpoor households.....	78
Box 7.1: The role of poultry transfers in diet diversity.....	83

Table 7.1a: Anthropometrics	88
Table 7.1b: Anthropometrics	89
Table 7.2: Infant and young child feeding (IYCF) practices	90
Table 7.3a: Children's dietary diversity	91
Table 7.3b: Children's dietary diversity	92
Table 7.4: Women's dietary diversity	93
Table 7.5: Child health history.....	94
Table 7.6: Exposure to health and nutrition services	95
Table 7.7a: Childcare activities in past 3 days.....	96
Table 7.7b: Childcare activities in past 3 days.....	96
Table 7.8: Caregiver Reported Early Development Instruments (CREDI)	98
Table 8.1a: Female's and male's depression outcomes	105
Table 8.1b: Female's and male's depression outcomes: Subsample of poor households	105
Table 8.1c: Female's and male's depression outcomes: Subsample of nonpoor households	106
Table 8.2a: Experience with intimate partner violence (IPV) over past year	107
Table 8.2b: Experience with intimate partner violence (IPV) over past year: Subsample of poor households	107
Table 8.2c: Experience with intimate partner violence (IPV) over past year: Subsample of nonpoor households	108
Table 8.3a: Perceived gender norms.....	109
Table 8.3b: Perceived gender norms: sub-sample of poor households.....	109
Table 8.3c: Perceived gender norms: sub-sample of non-poor households.....	111
Table 8.4a: Female's input into decision making around production.....	112
Table 8.4b: Female's input into decision making around production: sub-sample of poor households....	112
Table 8.4c: Female's input into decision making around production: sub-sample of nonpoor households	114
Table B.1a: Household demographics at baseline: Subsample of poor households	123
Table B.1b: Household demographics at baseline: Subsample of nonpoor households.....	124
Table B.2a: Main livelihood outcomes at baseline: Subsample of poor households.....	126
Table B.2b: Main livelihood outcomes at baseline: Subsample of nonpoor households.....	127
Table B.3a: Main nutrition outcomes at baseline: Subsample of poor households	129
Table B.3b: Main nutrition outcomes at baseline: Subsample of nonpoor households	130
Table B.4a: Main depression outcomes at baseline: Subsample of poor households.....	131
Table B.4b: Main depression outcomes at baseline: Subsample of nonpoor households.....	133
Table B.5a: Main IPV outcomes at baseline: Subsample of poor households.....	134
Table B.5b: Main IPV outcomes at baseline: Subsample of poor households	135

List of Figures

Figure ES.1: SPIR experimental impact evaluation design	iv
Figure 3.1: SPIR Experimental Impact Evaluation.....	8
Figure 5.1: Receipt of the poultry package by households' original assignment	27
Figure 5.2: Reasons for reduction in chicken flock, by gender	28
Figure 5.3: Receipt of the cash grant by households' original assignment.....	29
Figure 5.4: Attendance of aspirations video screenings by gender (only kebeles assigned to treatment) ..	30
Figure 7.2: Treatment effects on probability of egg consumption (poor sample only), children 6-23 months.....	83
Figure 7.3: Treatment effects on probability of egg consumption (poor sample only), primary females...	84
Figure 7.1: HAZ against child age in SPIR baseline and midline surveys and 2016 DHS, by region	87

Acronyms

ANCOVA	Analysis of covariance
BCC	Behavior change communication
CAPI	Computer-assisted personal interview
CHF	Community health facilitator
CPNP	Community-based Participatory Nutrition Promotion
CREDI	Caregiver Reported Early Development Instruments
DFSA	Development Food Security Activity
DHS	Demographic and Health Survey
DID	Difference-in-difference
DS	Direct Support
EC	Ethiopian calendar
ETB	Ethiopian birr
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.² Under funding from USAID's Food for Peace Initiative 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 targets nearly 500,000 PSNP4 beneficiaries in 15 of the most vulnerable woredas in the Amhara and Oromia regions of Ethiopia. The main objectives of SPIR are to enhance resilience to shocks and livelihoods and improve food security and nutrition for rural households vulnerable to food insecurity.

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. The gender-sensitive intervention packages draw from a set of livelihood 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.³

This midline report of the impact evaluation presents evidence from the first 15 months of the project on program performance and the effectiveness of delivery. In addition, this report presents preliminary impact estimates for many of the project's primary and secondary outcomes. Due to the phased roll-out of SPIR interventions and the relatively short life of the project so far, we expect relatively modest impacts on outcomes that require extended duration of exposure. However, other outcomes may have been meaningfully affected already by the substantial investments undertaken so far. It is informative to examine the pattern of impacts at this intermediate stage as a measure of achieved progress and to inform improvements in delivery.⁴ This report briefly summarizes the SPIR interventions, the evaluation study

² 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.

³ In addition, after the midline survey, two study arms will introduce group-based therapy for women screened for depression, in part as a strategy to improve their ability to improve the nutritional status of their children.

⁴ This midline 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.

design, and the process of data collection for the midline survey before presenting information on beneficiaries' experience with the program and evidence of impact. More details on the study design, SPIR interventions and the program context can be found in the SPIR Baseline Report (Alderman et al. 2019).

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, promoting higher aspirations, poverty and mental health, and the effect of cash transfers on intimate partner violence. 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 will be the first study we are aware of that will test the impact of a nutrition-sensitive graduation program. Second, SPIR provides fewer resource transfers (the largest being a one-time poultry or cash grant of US\$200 [ETB equivalent value]), and instead focuses on linking PSNP clients to information and improved public services.

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-sourced 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 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 to this debate by testing a promising poultry start-up package developed by EthioChicken, and benchmarking this package against cash transfers of similar value.⁵

There is growing recognition from economic theorists and empirical researchers that individuals’ capacity to aspire can drive them to invest in their future (Appadurai 2004), and that aspiring too high leads to frustration and underinvestment (Genicot and Ray 2017). Within this literature there is a recognition that poverty can induce an aspirations failure, which causes underinvestment and perpetuates poverty (Janzen et al. 2017; Dalton, Ghosal and Mani 2014). We build on studies from Ethiopia reporting the impact of

⁵ See McIntosh and Zeitlin (2018), for example, who conducted an experiment to benchmark a nutrition intervention in Rwanda against large cash grants.

prompting aspirations on household investment decisions, including in child education (Bernard et al. 2017, 2019; Taffesse and Tadesse 2017). Using the same inspirational documentaries to provide an experimental positive shock to aspirations, we test whether an aspirations intervention expands the impact of the SPIR project.

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 midline survey. Using an intervention implemented by SPIR after the midline survey, we will test at endline whether group-based therapy (Bolton et al. 2007) reduces maternal depression in Ethiopia and enhances the impact of the SPIR project.

Also, recent research has shown that cash transfers and other modalities reduce intimate partner violence (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 Roy forthcoming). Here, we will test whether the SPIR treatment arms provide a comparable reduction in IPV.

This Midline 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 midline survey data collection. Chapter 5 reports on delivery of program components and participants' program exposure. Chapter 6 presents impacts on livelihood outcomes. Chapter 7 presents impacts on nutrition and child welfare outcomes. Chapter 8 presents impacts on depression and IPV at midline. Chapter 9 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, SPIR planned for the delivery of a poultry kit from EthioChicken or an equivalent one-time cash grant of US\$200(ETB equivalent value). Most other benefits of the SPIR project take 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 and Purpose 4 on climate resilience, 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. 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
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 are organized around VESA groups. VESA groups were formed at the beginning of the project, soon after the baseline survey. These groups 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. VESA groups include men and women (often the husband and wife from a single 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 water, sanitation and health (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. Food 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*.⁶

2.1 L* enhanced livelihood activities

Social analysis and action: In the SPIR program, social analysis and action (SAA) is 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.⁷

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.⁸ 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. They are an 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 large one-time unconditional cash transfers of equivalent value. These transfers are targeted toward the poorest SPIR project participants. This project component is 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 (ETB equivalent) or as US\$200 worth of poultry start-up

⁶ See the Baseline Report (Alderman et al. 2019) for more details on these interventions.

⁷ In practice, implementation of the SAA interventions was not as intensive as originally planned.

⁸ See Bernard et al. (2017, 2019) and Taffesse and Tadesse (2017) for the results of recent aspirations experiments conducted in Ethiopia.

inputs and training. These livelihood packages were distributed in April 2019, several months after the screening of the aspirations documentaries.

2.2 N* enhanced nutrition activities

Timed and Targeted Counseling: In the N* treatment, BCC activities are organized under an intervention model referred to as Timed and Targeted Counseling (TTC). Under TTC, community HEWs and HDA leaders provide lessons in health posts at the community level and through household visits on IYCF practices and adolescent and maternal nutrition. As with the primary SPIR Purpose 2 trainings, topics include diversifying diets into sources of nutritious foods (including cooking demonstrations) and promotion of use of health and nutrition services. TTC differs from the SBCC offered in all SPIR communities in that it is conducted at the household level rather than at larger group or community events. TTC also encourages men to support their wives in childcare. TTC endeavors to conduct 11 home-visits in the first 1000 days, including 3 visits during pregnancy.

Community-based Participatory Nutrition Promotion: Community-based Participatory Nutrition Promotion (CPNP) sessions are two-week intensive feeding sessions for acutely malnourished children that include trainings for their mothers on complementary feeding and caring practices. The sessions, as well as many other N* community-level nutrition activities, are led by CHF.

Male engagement in nutrition BCC: Household-level counseling (involving both husband and wife) related to IYCF and maternal nutrition will be conducted using the TTC approach to support shared decision-making. Because TTC conducts nutrition trainings directly in the community, it is more intensive than the SBCC provided in the SPIR nutrition package (N). SPIR hired CHF 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 CHF also support training of CPNP for nutritious food preparation at Growth Monitoring and Promotion sessions.

Male engagement through men champions, men's groups, public awareness campaigns: Male advocates will be identified and trained and will facilitate eight sessions for newly established men's groups in **each of the N* kebeles**. The sessions provide 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 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 7-10 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 was scheduled to be done in the other half of N* kebeles from April to July 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 the midline survey 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 5-8 women in each group) starting in December 2019 in **each of the N* kebeles**.

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

3 Evaluation Design

At midline, the impact evaluation used the 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. See the Pre-Analysis Plan for a complete list of the hypotheses to be tested (Alderman et al. 2020).

3.1 Experimental design

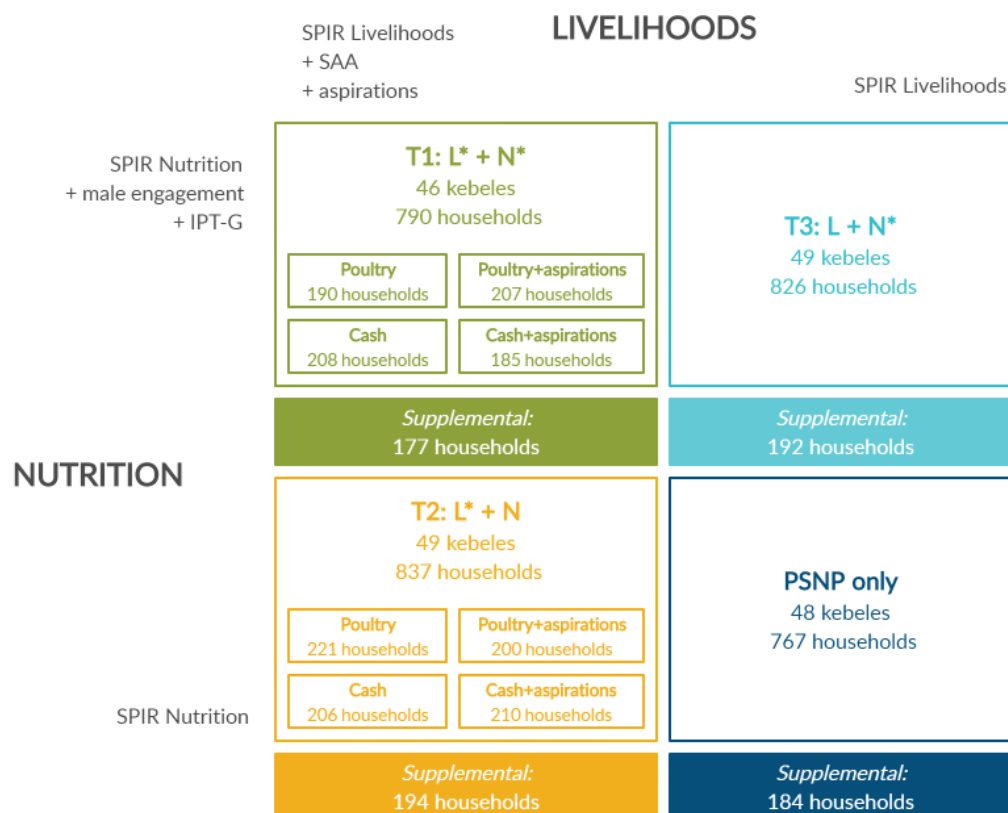
The impact evaluation design compares combinations of these activity packages by randomly assigning kebeles to one of the following four intervention arms (see Figure 3.1):

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 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).⁹ 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 will tell us the additional effect of SAA, aspirations, and livelihood transfers or of male engagement and IPT-G on top of the main integrated SPIR intervention.

Figure 3.1: SPIR Experimental Impact Evaluation

⁹ The IPT-G group therapy intervention was introduced after the midline survey to women who were screened for depression at midline. Its impact as part of N* will be measured in the endline survey.



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 receive 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.¹⁰

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.

¹⁰ 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).

The SPIR midline data provide an opportunity to contribute evidence on 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.

3.1.2 Sub-study on maternal 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 sub-study on the impact of IPT-G on maternal depression and child nutrition outcomes, collaborating with psychologists Lena Verdeli (Teachers College, Columbia University) and Paul Bolton (Johns Hopkins University) for the sub-study related to maternal depression.

During the midline survey, 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 0-27, 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 was 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 survey, we corrected the recall period to two weeks. We also undertook a more careful training of the survey team to improve the approach to enumeration for these sensitive questions. Mean midline PHQ-9 scores are reported in Chapter 8, along with estimates of the impact of SPIR on mean severity of depression symptoms at midline.

3.2 Implementing the SPIR evaluation study design

The experimental study design involves randomly assigning 193 kebeles into these 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) receive the aspirations intervention, stratified across T1 and T2. The livelihood transfer of the poultry start-up package or unconditional cash grant were 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 $\geq$ 8) were invited to enroll in IPT-G in December 2019. 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 – March 2020, and the other half planned for scaling 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.¹¹ 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 Village Economic and Saving Associations (VESA) 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 one 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 Baseline balance

¹¹ After the baseline survey, two 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.

The baseline report presented evidence that key household covariates were balanced at baseline when comparing across the four main treatment arms. In this report, we present additional evidence of balance in baseline covariates 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 tables are presented in Appendix B.

Tables B.1a and B.1b present balance for demographic variables for the poor and nonpoor samples, respectively. In general, there is no robust evidence of any meaningful difference in baseline characteristics; those coefficients that are statistically significant are varying in sign, and significant only at the 10 percent level. In addition, the number of significant coefficients is consistent with what would be expected to arise simply by chance, given the number of outcome variables examined. The only somewhat clear pattern is that poor households in T1 are somewhat older vis-à-vis households in the control arm, and nonpoor households in T1 and T2 are more likely to be female-headed vis-à-vis households in the control arm.

Tables B.2a and B.2b present balance for important variables linked to the livelihoods outcomes of interest discussed in Chapter 6 of this report. In general, again there is no robust evidence of imbalance. Among nonpoor households, households in all three arms are somewhat less likely to report owning livestock vis-à-vis households in the control arm, but the estimated coefficients are small in magnitude.

Tables B.3a and B.3b present evidence for baseline balance for important variables linked to the nutrition outcomes of interest discussed in Chapter 7 of this report. There is no clear evidence of imbalance for any variable presented here.

Tables B.4a and B.4b present evidence for baseline balance for variables around depression and Tables B.5a and B.5b present evidence for baseline balance for variables around IPV as discussed in Chapter 8 of this report. Here, for the subsample of poor households, there is some evidence of imbalance. Men in the cash and aspirations sub-arm of T1 report significantly higher PHQ-9 scores vis-à-vis men in the control arm, and the difference is large (an increase of 2.2 relative to a mean in the control arm of 2.9). By contrast, both men and women in the poultry arm in T2 report somewhat lower PHQ-9 scores at baseline (differences of between 0.7 and 1.0, relative to means of around 3.0), and women also report a lower incidence of emotional IPV.

For the subsample of nonpoor households, there is less evidence of imbalance. However, men in T1 who are not exposed to the aspirations treatment report lower PHQ-9 score at baseline, while men and women in the aspirations sub-arm report generally higher PHQ-9 scores. Again, the magnitudes of the differences are substantial relative to the control arm means.

In general, we can conclude from these results 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, and this pattern will be taken into account when the primary impact results are interpreted.

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 two years. 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 four such households in each kebele to add another 768 households to the sample, for 4,082 households overall in the midline sample.

In order to assure that the supplemental midline survey households were similar to the households in the baseline and midline panel survey, we compared the characteristics of the panel households and supplemental sample households using the midline data (Table 3.1). For each characteristic, we present the means in Amhara and Oromia as well as the mean in the panel sample and the supplemental sample at midline. We present two test statistics for the equality of the mean: the normalized difference and the p-value. The normalized difference is the ratio of the difference in means to the average of the within-treatment-group standard deviation) suggested by Imbens (2015). This test for equality of means is less sensitive to sample size than the more commonly used p-value. Following Imbens and Rubin (2015), we interpret normalized differences below 0.25 as failing to reject equality of the means across the panel and supplemental sample. Similarly, a p-value above 0.05 indicates a failure to reject equality of the means.

The table shows that supplemental sample households are very similar to the panel sample households. Based on the p-value, supplemental sample households have heads and primary females (mother of the index child) that are somewhat less likely to be married and primary females are somewhat older as well. Using baseline data for the panel households and midline data for the supplemental households, we compare housing characteristics and find that supplemental households are about 8 percentage points less likely to have improved roof materials on their homes.

**Table 3.1: Difference in household characteristics at midline
Original vs supplemental sample**

	N	Amhara	Oromia	Original sample	Supplemental sample	Normalized difference	P-value
Age of household head	3,924	40.404 (11.226)	38.556 (9.588)	39.151 (10.164)	39.737 (10.705)	-0.056	0.216
Household head: married, monogamous	3,923	0.772 (0.420)	0.864 (0.343)	0.842 (0.365)	0.804 (0.397)	0.102	0.022
Age of primary female	3,855	31.875 (8.591)	31.572 (7.391)	30.863 (6.448)	31.958 (8.439)	-0.146	0.000
Primary female: married, monogamous	3,822	0.794 (0.404)	0.912 (0.283)	0.870 (0.337)	0.839 (0.368)	0.088	0.044
Primary female has some education	3,825	0.269 (0.443)	0.240 (0.427)	0.250 (0.433)	0.258 (0.438)	-0.019	0.643
Primary female's main activity is crop production	3,821	0.303 (0.460)	0.270 (0.444)	0.280 (0.449)	0.291 (0.454)	-0.025	0.515
Age of primary male	3,330	40.313 (10.371)	38.615 (9.768)	39.012 (9.535)	39.658 (10.269)	-0.065	0.151
Primary male: married, monogamous	3,325	0.970 (0.171)	0.899 (0.301)	0.949 (0.220)	0.935 (0.247)	0.063	0.117
Primary male has some education	3,315	0.348 (0.476)	0.519 (0.500)	0.425 (0.495)	0.427 (0.495)	-0.003	0.943
Primary male's main activity is crop production	3,322	0.915 (0.279)	0.967 (0.178)	0.952 (0.213)	0.936 (0.245)	0.071	0.113
Household has improved source of water - rainy season *	4,040	0.724 (0.447)	0.494 (0.500)	0.636 (0.482)	0.625 (0.484)	0.023	0.655
Household has improved roof material *	4,040	0.411 (0.492)	0.463 (0.499)	0.497 (0.500)	0.419 (0.493)	0.158	0.000

Notes: Estimates from the SPIR midline survey sample. Standard deviations are in parentheses and p-values are from the tests of difference of means between the two samples.

* Data collected at baseline for the original sample and at midline for the supplemental sample.

3.4 Empirical strategy

In this midline report, impacts of the SPIR project are estimated on the baseline and midline 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 using the original panel sample and exclude the supplemental sample in order to be able to control for the baseline outcome. For child nutrition and dietary outcomes measured on children age 0-35 months or 6-23 months, we estimate impacts using a repeated cross-section model on both the original baseline and supplemental sample. In these models we control for baseline kebele mean outcomes since baseline data are not available for most children in the midline sample.

For most outcomes, we estimate three models:

- 1) the average treatment effect of each treatment (T1, T2 and T3) over all its cross-randomized sub-treatments (the 'pooled' treatment effect of that treatment),
- 2) the impact of each treatment and all cross-randomized sub-treatments (poultry/cash and aspirations in T1 and T2) on the sample of 'poor' households, and
- 3) the impact of each treatment with and without the aspirations intervention in the 'nonpoor' sample.¹²

Muralidharan et al. (2020) note that in the presence of cross-randomized interventions, like the aspirations intervention that was randomly assigned to half of the L* kebeles in T1 and T2, it is necessary to estimate the long form of the treatment effects regression model containing interaction terms for all treatment arms and sub-randomizations in order to be able to estimate the pure effect of the higher level treatment (for example, the effect of T1 without aspirations). However, it is also meaningful to report the pooled effect of T1 across all T1 kebeles, averaging the effect of T1 without aspirations and T1 with aspirations, as long as that interpretation is made clear. This is the approach we take with the first pooled model because of its simplicity.

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

¹² The 'poor' sample includes the poorest 10 out of 18 households based on a baseline PCA asset index and the 'nonpoor' sample includes the 8 out of 18 least poor households based on the same index.

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

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 $T \in \{T1, T2, T3\}$ 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 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 poor household sample as follows:

$$Y_{1hv} = \beta_0 + \beta_1 T1_v * P_v + \beta_2 T1_v * P_v * A_{hv} + \beta_3 T1_v * C_v + \beta_4 T1_v * C_v * A_{hv} + \beta_5 T2_v * P_v + \beta_6 T2_v * P_v * A_{hv} + \beta_7 T2_v * C_v + \beta_8 T2_v * C_v * A_{hv} + \beta_9 T3_v + \beta_{10} Y_{0hv} + \beta_{11} X_{0hv} + \mu_d + \varepsilon_{hv} \quad (2)$$

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. $T1_v * P_v * A_{hv}$ is an indicator for a household in a T1 treatment kebele assigned to the poultry transfer that is also assigned to the aspirations treatment. β_1 is an estimate of the average impact of T1 and the poultry transfer without the aspiration intervention relative to the control group. We can test whether the aspirations treatment changes the impact of the T1 with poultry relative to the control group by testing whether β_2 is significantly different from zero. β_3 is an estimate of the average impact of T1 and the cash transfer without the aspiration intervention.

For model 3, we differentiate household impacts of the two L* treatment arms, T1 and T2, by whether the household was randomly assigned to the aspirations documentary treatment or not. This model will be estimated on the sample of households in all four treatment arms as follows:

$$Y_{1hv} = \beta_0 + \beta_1 T1_v + \beta_2 T1_v * A_{hv} + \beta_3 T2_v + \beta_4 T2_v * A_{hv} + \beta_5 T3_v + \beta_6 Y_{0hv} + \beta_7 X_{0hv} + \varepsilon_{hv}, \quad (3)$$

where $T1_v$ indicates that household h from kebele v and woreda d is assigned to treatment T1, and $T1_v * A_{hv}$ is the interaction of the $T1_v$ dummy with an indicator for whether the kebele was assigned to the

aspirations treatment. β_1 is an estimate of the average impact of T1 without the aspiration intervention. We can test whether the aspirations treatment changes the impact of the pooled T1 intervention relative to the control group by testing whether β_2 is significantly different from zero. That is, β_2 measures the additional effect of T1 relative to the control group when the aspirations treatment is added.

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 a large number of hypothesis tests, the probability of rejecting the null hypothesis may be increasing in the number of tests. Approaches to correct for multiple hypothesis testing are generally conservative. These methods include a Bonferroni correction that adjusts the required p-value that indicates rejection of the null hypothesis to be the p-value that corresponds to the significance level divided by the number of tests. Other corrections reduce the number of tests by creating indices of families of outcomes or using step-down methods that adjust p-values to control for the familywise error rate (see Romano and Wolf 2005; Kling et al. 2007; Anderson 2008).

We want the reader to be aware of the potential concern about multiple hypothesis testing but choose not to make adjustments in the results in this midline survey report. Rather, readers should be aware that at $\alpha=0.05$, on average 5 estimates out of 100 would be statistically significant by chance. In addition, correcting for multiple hypothesis testing would reduce the number of these significant results.

4 Midline Data Collection

Fieldwork for the midline survey data collection was completed from July 25 to October 23. Laterite served as the in-country survey partner, leading the fieldwork for the midline data collection in cooperation with the quantitative evaluation team from IFPRI.

4.1 Survey instrument

The midline household questionnaire was developed by the IFPRI team based on the baseline household questionnaire. Like the baseline questionnaire, the midline household questionnaire was structured in three parts: a brief household-level interview for identification and household demographics, a male respondent questionnaire and a female respondent questionnaire. Several changes were made from the baseline questionnaire to make it suitable for the midline. First, we added a substantial module capturing respondents' participation with and exposure to the SPIR project as well as their satisfaction with the project components. We also added more outcomes related to the use of the poultry and cash transfers. To save time, we removed the food and nonfood consumption module and shortened the agricultural production module. GPS coordinates were also collected for each household to assist in tracking households in subsequent rounds. The list of midline survey questionnaire modules is presented in Tables 4.1 – 4.3. The complete baseline questionnaire can be found in Appendix A.

Initially, the interviews for the midline household survey were taking well over two hours and many of the kebeles and households were very remote, sometimes requiring more than two hours walk. After removing another section of the agricultural module and inviting some of the remote households to a central location for interviews in exchange for a transport fee, interviews averaged less than two hours per household and the midline fieldwork schedule got back on track. Primary female and male respondents were retained from the baseline survey in the original sample households when possible. If a younger sibling of the Baseline Index Child under age two years was available for nutrition assessment, that child was assigned to be the Midline Index Child in that household from the original sample. In supplemental households, a child under age two years was selected at random and was assigned to be the Midline Index Child; the mother or primary female caregiver of that child was identified as the primary female respondent.

A separate, specially trained team conducted the anthropometry measurements. The index child, sibling of the index child, and primary female were measured. Children who were found to have a low mid-upper arm circumference indicating possible severe acute malnutrition were referred to the local health post in the kebele.

The midline 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.

Table 4.1: Midline Survey Household Questionnaire Modules

Strengthen PSNP4 Institutions and Resilience (SPIR)
Development Food Security Activity (DFSA)
MIDLINE SURVEY: Household Questionnaire – August 15, 2019

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
Part 2: Education and Activities (age 6-13)

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
-----------------------	----------------	---------------------

Table 4.2: Midline Survey Female Questionnaire Modules

**Strengthen PSNP4 Institutions and Resilience (SPIR)
Development Food Security Activity (DFSA)
MIDLINE SURVEY: Female Questionnaire – August 15, 2019**

DRAFT: For Research Purpose only

Sample Variables *(from household questionnaire)*

Module A: Housing, Water, Sanitation and Hygiene

- Part 1: Housing and water (Supplemental sample only)
- 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
- Part 3: Agricultural extension related to livestock

Module D: Own Business Activities and Wage Employment

- Part 2: Income Apart from Own-Agricultural Activities: Wage Employment

Module F: Access to credit and financial services

- Part 3: Access to savings

Module L: Aspirations & Wellbeing

- Part 3: Experience with depression and emotional wellbeing
- Part 4: Safety protocol

Module H: Dietary Diversity and Food Security

- Part 1: Women's dietary diversity (24-hour recall)
- Part 3: Household food security

Universal Codes *(Include with all CAPI options):*

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

Module I: Household activities, decision-making and empowerment

- Part 1: Decision-making on value chains and market access
- Part 2: Agency
- Part 5: Intrahousehold dynamics and attitudes

Module J: Nutrition, Health, and Care of Child

- Part 1: Pregnancy and participation in PSNP4
- Part 2: Use of antenatal and postnatal services
- Part 3: Infant and young child feeding (IYCF) practices
- Part 4: Child health history
- Part 6: Child care activities
- Part 7: Exposure to health and nutrition services
- Part 8: Anthropometry

Module N: Caregiver Reported Early Development Instruments

- Part 1: 0-5 months
- Part 2: 6-11 months
- Part 3: 12-17 months
- Part 4: 18-23 months
- Part 5: 24-29 months

Module L: Aspirations & Wellbeing

- Part 1: Aspirations

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 4: Disaster resilience and mitigation

Table 4.3: Midline Survey Male Questionnaire Modules

Strengthen PSNP4 Institutions and Resilience (SPIR)
Development Food Security Activity (DFSA)
MIDLINE SURVEY: Male Questionnaire – August 15, 2019

DRAFT: For Research Purpose only

Sample Variables (from household questionnaire)

Module B: Livestock Production

Part 1: Livestock Ownership and Management
Part 4: Agricultural Extension Related to Livestock

Module C: Agriculture

Part 1: Land characteristics and tenure
Part 3a: Crop choice – Mehr Season
Part 3b: Crop inputs and labour – Mehr Season
Part 3c: Crop production, sales, and use – Mehr Season
Part 4: Agricultural Extension Related to Crop Production

Module D: Income Apart from Own-Agricultural Activities

Part 1: Wage Employment

Module E: Youth training

Part 2: Youth training and referrals

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 L: Aspirations and Wellbeing

Part 4: Experience with Depression and emotional wellbeing
Part 5: Safety protocol

Module I: Household decision-making and empowerment

Part 1: Decision-making on value chains and market access
Part 2: Agency
Part 3: Intrahousehold Dynamics and Attitudes
Part 4: Gender norms

Module J: Nutrition, Health, and Care of Child

Part 2: Child care activities
Part 3: Exposure to health and nutrition services

Module L: Aspirations and Wellbeing

Part 1: Aspirations

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
Part 4: Disaster resilience and mitigation

Module L: Aspirations and Wellbeing

Part 2: Experience with shocks
Part 3: Poverty perceptions and wellbeing

Universal Codes (Include with all CAPI options):

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

4.2 Ethics approval

IFPRI received approval from its Institutional Review Board (IRB) at baseline for the SPIR DFSA quantitative evaluation design. This approval was updated for the midline survey. IFPRI also received ethics approval from the Institutional Review Board (IRB) at Hawassa University, which has a national research center on nutrition.

Informed oral consent was collected from all participants prior to the start of the interviews. The entire field team was trained on ethical data collection prior to the start of the field work. Before beginning a household survey, enumerators read the 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 household interview questions. The full written consent can be seen at the beginning of the household survey in Annex A. The enumerator only completed a survey for the household if they received verbal consent to participate in the study from the potential respondent.

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 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 to which only the research team will have access. Anonymized versions of the datasets that exclude these personal identifiers will be the ones made available for public access.

4.3 Enumeration teams and trainings

Laterite, working closely with IFPRI, organized enumeration training for the baseline survey in late July 2019. The two-week training consisted of validating eligibility of households, administering the survey on tablets, and an in-depth training on all modules. Substantial time was taken to train the enumerators on conducting interviews for the PHQ-9 depression symptoms tool and the IPV questions. The training included one day of pretesting the survey, on a community of farmers in a rural PSNP community outside of Addis Ababa. A special four-day anthropometry training was held for anthropometry specialists, which included repeated practice measurements of women and children.

Each survey team consisted of three enumerators, one supervisor, and one anthropometric specialist. Each enumerator was expected to complete, on average, two household interviews per day.

4.4 Fieldwork experience

4.4.1 Achieved sample size and interview completion rate

The midline survey enumeration team interviewed a total of 3,968 households out of the target of 4,082 households for the entire sample (see Table 4.4). This included targeting 3,314 households interviewed at baseline plus 768 supplemental sample households. Among the baseline sample households, 3,220 were able to be located and interviewed at midline, leading to a very low attrition rate of just 2.8 percent. Among the 114 attrited households, the top reasons for attrition included moving outside the survey area (n=44) and refusing consent (n=17). In total, 3,878 primary female respondents were interviewed as were 3,082 male respondents. Among the Baseline Index Children and Midline Index Children, the number of children in the sample age 0-23 months is 2,075 and the number age 24-60 months is 2,454.

Table 4.4: Summary of midline household sample

	Panel sample			Supplemental sample			Total
	Amhara	Oromia	Total	Amhara	Oromia	Total	
Number of EAs targeted	112	80	192	112	80	192	192
Intended households per kebele	--	--	--	4	4	4	--
Completed households per kebele	--	--	--	3.94	3.84	3.90	--
Intended household interviews	1,921	1,393	3,314	448	320	768	4,082
Household not found	45	32	77	--	--	--	77
Consent not given	14	3	17	0	0	0	17
Completed household interviews	1,862	1,358	3,220	441	307	748	3,968
T1: L*+ N*	469	321	790	105	72	177	967
T2: L*+ N	484	353	837	112	82	194	1,031
T3: L + N*	485	341	826	114	78	192	1,018
C: PSNP only	424	343	767	110	75	185	952
Number of primary female respondents	1,796	1,336	3,132	440	306	746	3,878
Number of primary male respondents	1,358	1,124	2,482	341	259	600	3,082
Number of children 0-23 months	678	650	1,328	440	307	747	2,075
Number of children 24-60 months	1,352	1,102	2,454	0	0	0	2,454

4.5 Data quality and cleaning

Data from the midline household survey 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.

Laterite and IFPRI took extensive measures to ensure the quality of the data collection. Team supervisors travelled with the enumeration teams, sat in on interviews, and reviewed the data being collected. Enumerators and supervisors conducted revisits to address any issues that may have come up from the data collection teams. Any issues were discussed, and solutions provided. Main problems included duplicate household identification numbers and missing data in certain modules.

An error in the SurveyCTO program for indirect measures of the prevalence of IPV (list experiment) led to the recall IPV questions not being asked for more than 600 women in the sample. In response, Laterite sent a team of mostly female enumerators back to these households to collect the missing IPV variables.

5 Experience with the SPIR Program

In this chapter, we describe participation in different SPIR interventions as reported by participants in the midline 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.

5.1 Core programming

5.1.1 Participation in PSNP

The base of the PSNP programming, Public Works and Direct Support (DS) were available to all four treatment arms. As shown in Table 5.1, roughly 90 percent of all respondents indicate participating in Public Works during both the year preceding the survey (Hamle 2010 – Sene 2011 in Ethiopian calendar, or July 2018 – June 2019) and the year before that (July 2017 – June 2018). This is slightly higher than the roughly 84 percent participation rate observed at baseline. The participation across treatment arms is similar and ranges between 88.0 percent in T3 to 91.0 percent in T1. Receiving DS payments is reported by 36.0 percent of PSNP households during the year preceding the survey, and by 35.6 percent of households during the year before that, the receipt rates being, again, the lowest in T3 (31.6 percent) and the highest in T1 (38.4 percent).

While no household is supposed to participate in Public Works *and* receive DS at the same time, we see a high overlap with 29.2 percent of households indicating participation in both programs during the year preceding the survey. The most likely explanation to this pattern is a considerable number of these households receiving Temporary Direct Support (TDS)—short-term payments targeted toward pregnant and lactating mothers in Public Works’ households—and confusing this with Permanent DS.

In terms of regional differences, households in Oromia are more likely than households in Amhara to have participated in Public Works over the previous year (94.6 percent vs. 86.1 percent) and less likely to have received DS payments (29.2 percent vs. 40.9 percent). While the pattern of regional differences is consistent with what was observed in the baseline survey a year and a half earlier, the rates of receipt are slightly higher than expected, given that 3.9 percent of beneficiaries in Oromia and 19.8 percent in Amhara reported receiving DS payments over the previous year at baseline. The lower share of DS recipients in Oromia can likely be explained by both the low share of female-headed households in Oromia as well as the much higher proportion of households already receiving payments before the start of PSNP4 in Amhara.

Only 21 out of 3,129 male respondents indicate that they would describe their household as having graduated from the PSNP program at midline, the majority of these men saying that this happened due to their income level. As a comparison, at baseline 6.9 percent of respondents described themselves as having graduated from PSNP4 in the last two years. In addition, at midline 104 men or 3.7 percent of respondents say that they stopped participating in the PSNP within the last two years without having graduated. In terms of form of benefits, 11.6 percent of respondents indicate receiving their PSNP Public Works payments as food, 5 percent as cash, and the remaining 83.4 percent as both food and cash.

5.1.2 VESA groups

In the midline survey, both females and males were asked whether someone in their household is a member of a VESA group. Table 5.2 reports both measures as well as a combined indicator of someone in the household being a member. In treatment groups T1, T2 and T3, 78.7 percent, 78.0 percent and 75.9 percent, respectively, report having a VESA group member in the household. In all groups, females are 6–7 percentage points less likely to report a VESA member, and Oromia has consistently lower rates than Amhara for all three measures of membership.

Contrary to the program design, 16.3 percent of households in the control group report having a member in a VESA group. The reported membership is concentrated in a few specific kebeles in Meket woreda¹³ (where the average membership among T4 kebeles is 42.9 percent), Grawa woreda (22.2 percent), and Sekota woreda (18.5 percent). Notably, a small share of households (5.9 percent) already reported having joined a VESA group at the baseline survey, and already at that time, the two woredas with the highest share were Meket (17.6 percent across the sample) and Sekota (8.3 percent). Grawa, however, had no reported VESA membership at baseline.

5.2 Livelihood components of the program

5.2.1 Financial literacy and value chains

The primary VESA group activities such as trainings on financial literacy, agriculture and livestock value chain development, 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). As discussed above, reported VESA membership at the time of the midline study is not universal, but across these arms nearly 78 percent of households have a member who belongs to a VESA group. Participation in both financial education trainings and value chain trainings is, however, low. Table 5.3 and Table 5.4 show that 25.4 percent of females in T1, 29.1 percent in T2, and 23.3 percent in T3 report participating in financial education trainings or discussions. For males, these figures are 28.0 percent, 27.2 percent, and 26.7 percent, respectively. Among the 640 women and 663 males who do report participating, the average number of financial education trainings attended is 2.4 and 2.3, respectively.

While the goal of VESA groups 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 multi-day trainings on targeted value chains, depending on cost implications, recipients' interest, and perceived capacity to participate in respective value chains. Participation rates in value chain development trainings were thus low among both male and female respondents, and a little lower in T3 than in T1 and T2: roughly 16 percent of females and males report attending in T1 and 15 percent in T2, while in T3, 11 percent of males and 12 percent of females attended. Among males, by far the most frequently attended types of trainings were on poultry production and sheep/goat fattening/rearing, with 64.8 percent and 63.9 percent of respondents who had been to value chain trainings attending these types, respectively. Ox fattening training was attended by 15.0 and vegetable production by 11.6 percent of participating males. Among females, poultry production training was the most popular type by a wide margin, with 64.9 percent of

¹³ Specific kebeles in the control group with high membership rate are Woyra Bere (81%), Kila (80%), Mebel (88%), and Hmusit (60%).

females in value chain trainings attending, while other livestock fattening and marketing was the runner-up with 40.0 percent of participating women participating in this training.¹⁴

In addition to various trainings, SPIR also supports the formation of producer marketing groups associated with SPIR-promoted value chains. However, this activity was only launched at the time of the midline survey, and groups relevant to specific value chains are not available in all locations. Indeed, observed membership in these groups is low within targeted communities—roughly 12 percent of females in T1, 11 percent in T2, and 6 percent in T3 report being a member of a marketing group. Similarly for males, these rates are 12 percent, 12 percent, and 5 percent, respectively. Again, the most common is membership in groups associated with poultry production (69 percent of participating women and 53 percent of men) and sheep/goat fattening (25 percent of participating women and 35 percent of men). Of the 230 females who are part of a producer group, less than 3 percent claim to be part of a bean production, a vegetable production, an ox fattening, or a beekeeping group. Among the 211 men who are part of a group, roughly 4 percent are part of an ox fattening, a bean production, or a vegetable production group, and less than 1 percent belong to a beekeeping group.

5.2.2 Poultry package and cash grant

SPIR livelihood transfers were given to women in the 10 poorest households of each L* kebele, the poverty status being determined by information on assets and land collected during the baseline survey. Half of these L* kebeles were randomized into receiving a cash grant, i.e. a transfer of US\$200 in cash (ETB equivalent), and another half into receiving an improved breed poultry package, i.e. US\$200 worth of poultry start-up inputs including 16 Sasso breed pullets from EthioChicken, feed, materials and cash for chicken coop construction, a feeding trough, and training. The livelihood packages were distributed between April and June 2019.

The midline survey asked respondents about the received transfers. Table 5.5 and Table 5.6 show these answers, restricting the sample to the households who were actually assigned to get the respective transfer: 50 percent of L* kebeles for both cash grant and poultry package, respectively, and 55 percent of households (or 10 out of 18 households) in these kebeles that were identified as poor.

¹⁴ Female's CAPI program had an error for the question about which value chain the respondent participated in, so instead of the original options of *sheep/goat fattening, poultry production, ox fattening, beekeeping, vegetable production* and *bean production*, the choices in the survey were *improved fodder management and feeding practice, multiplication of improved forage seed and seedlings, livestock fattening and marketing, technical training on poultry production* and *market driven agricultural production*. The outcomes reported in Table 5.3 are recovered from these answers.

Figure 5.1: Receipt of the poultry package by households' original assignment

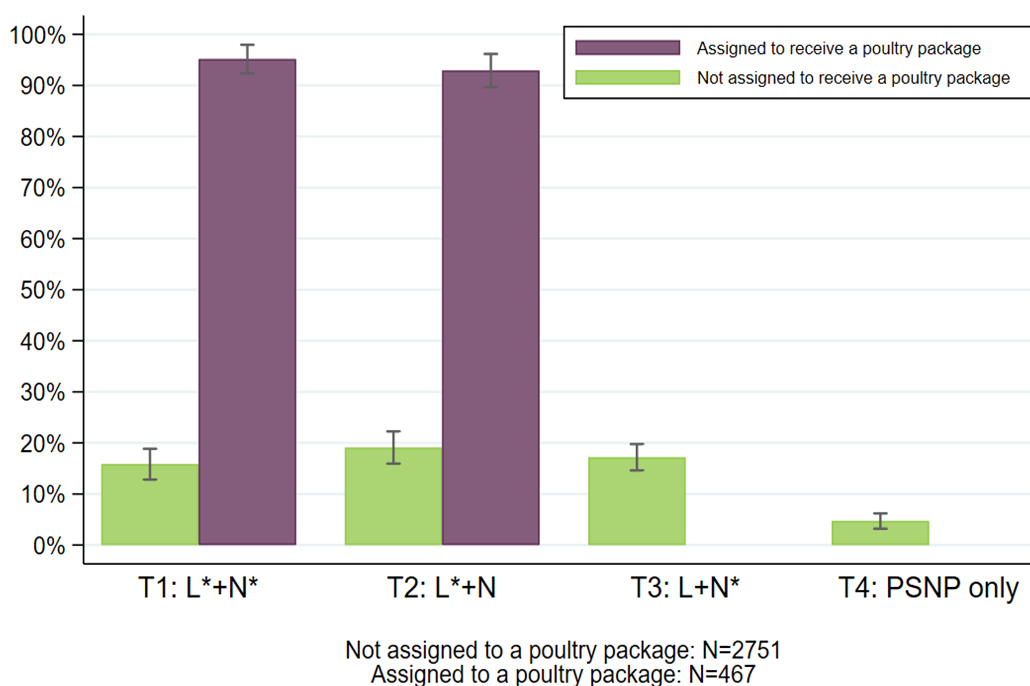
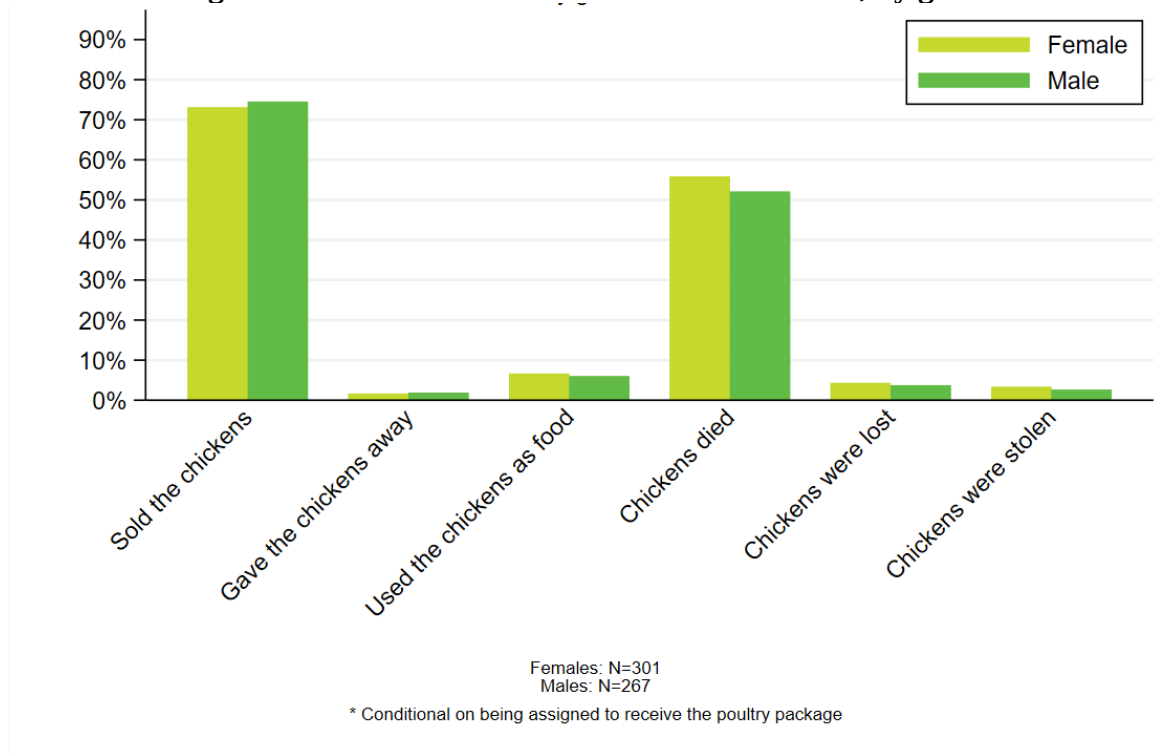


Figure 5.1 combines answers given by female and male in the same household on whether they received the poultry package, and illustrates the very high likelihood that households that were assigned to receive the package did get it—95 percent of assigned households in T1 and 93 percent in T2 report doing so. Both livelihood transfers were targeted toward females, and looking at answers by gender indeed shows higher reported receipt rates among females, but not by a wide margin: the gap is roughly 10 percentage points in both T1 (92 percent of females compared to 81 percent of males) and T2 (89 vs. 80 percent). As seen in Figure 5.1, a significant number of households among those not assigned to get the package (i.e., nontreatment kebeles or households that are not among the poorest 10) also report getting one—the rate ranges from 16 percent in T1 to 19 percent in T2. The most likely explanation for this is another intervention under SPIR programming—cost-sharing provision of poultry to pregnant and lactating mothers. This entailed provision of six chickens, the cost of which the beneficiaries had to partly cover themselves. Indeed, looking at the reported number of chickens received, those assigned to treatment report receiving 15.8 chickens, on average—aligned with the actual package size of 16 pullets—while those not assigned but claiming that they received the package report getting 6.9 (female respondents) or 7.1 (males) pullets, confirming that they are likely referring to the cost-share program.

Figure 5.2: Reasons for reduction in chicken flock, by gender



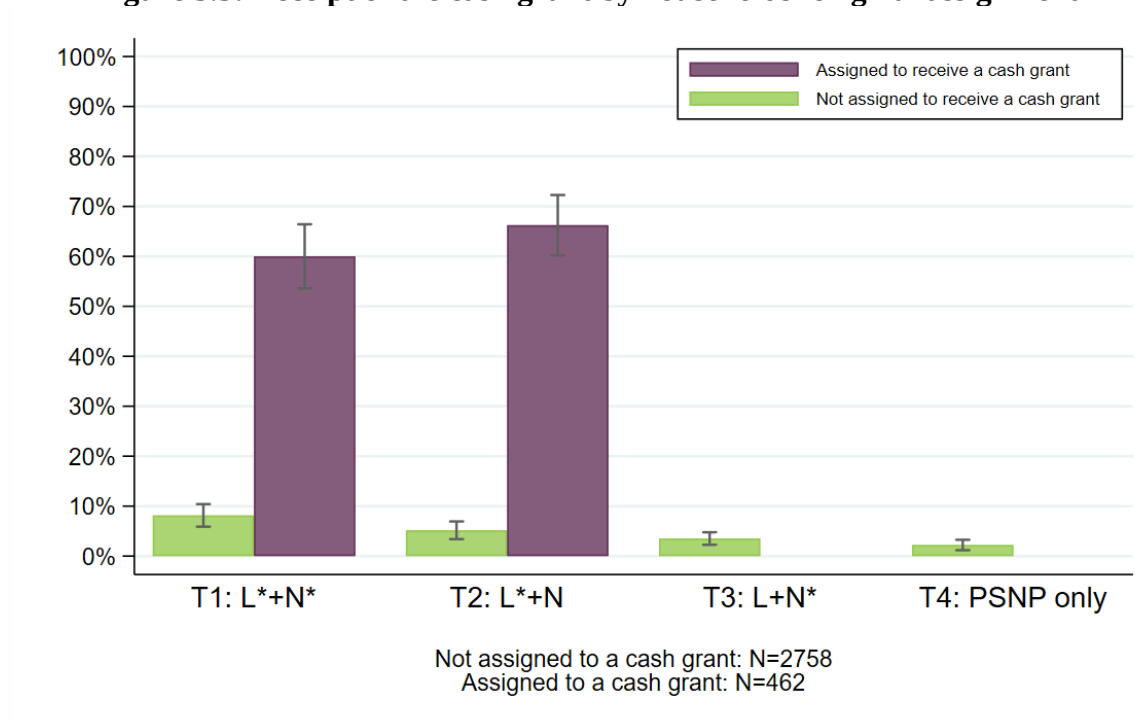
Among those who were assigned to receive the package and actually report doing so, the average number of pullets that are left by the time of the midline study—that is, up to six months after delivery—is 10.1 for both females and males. Both the mean and the median loss were six chickens, somewhat consistent with the expectation of recipients selling the male pullets after 60 days. Only 10 women and 4 men among those who got the package report losing all of the chickens they received—making it less likely that widespread disease was encountered—while reduction in flock seems to be considerably higher in Amhara: in kebeles assigned to poultry, the average reduction was 8.3 in Amhara compared to 2.3 in Oromia for both females and males.¹⁵ Some reasons behind this regional discrepancy might be timing of transfers (that is, earlier maturation of pullets in Amhara) as well as stronger enforcement of the message on the need to sell male chicks before four months of age. Figure 5.2 displays main reasons stated for reductions in chicken flock since the receipt of the package. It shows that few beneficiaries used chickens as food, gave them away, or simply lost them. The most prevailing reasons are sale of chickens (73 percent of women vs. 75 percent of men) or death of chickens (56 vs. 52 percent). Note that the answer options were not exclusive and multiple reasons could have been chosen.

Finally, Table 5.5 and Table 5.6 also show that roughly 90 percent of those targeted by the poultry treatment claim that the chicken breed they got was Sasso—consistent with the intervention—and the

¹⁵ The reduction in chickens is especially noticeable in a few kebeles in Amhara: the flocks of recipients in Aymat, Cherey Etiknu, Menshiwa, Tsata and Yekatit woredas decreased by more than 10 chickens on average (both female and male), while Cherey Etiknu’s average reduction was 15 out of 16. In Oromia, the reduction rates were universally low with the exceptions of conflict-ridden Shelo Belela and Shesha Goyke kebeles in Siraro.

average amount of money they received to cover veterinary and coop building costs is reported as 1,266 ETB (US\$44.4¹⁶) by females and 1,296 ETB (US\$45.4) by males, also consistent with the program.

Figure 5.3: Receipt of the cash grant by households' original assignment



As illustrated in Figure 5.3, reported receipt rates of the cash grant are considerably lower than those of the poultry package. Combining the answers by female and male in the same household, in T1 60 percent and in T2 66 percent of targeted households report getting the grant. Table 5.5 and Table 5.6 show the breakdown by gender, and similar to the start-up package, females are about 10 percentage points more likely to report a receipt: 56.2 percent of women from targeted households in T1 and 52.7 percent in T2 indicate getting the grant, while 43.7 and 47.0 percent of men report getting a grant in the two treatment arms, respectively. The amount of cash that beneficiaries report getting is, on average, lower than the ETB equivalent of US\$200 that they were supposed to get: women indicate receiving 5,453 ETB (\$US191.1¹⁷) and men 5,441 ETB (\$US190.6), on average. On the other hand, the 4.5 percent of households in the sample that weren't targeted by the program but claim getting a cash grant (see Figure 5.3: Receipt of the cash grant by households' original assignment), indicate only getting 2,264 ETB on average, meaning that they likely mistook the cash grant for some other program.

5.2.3 Aspirations treatment

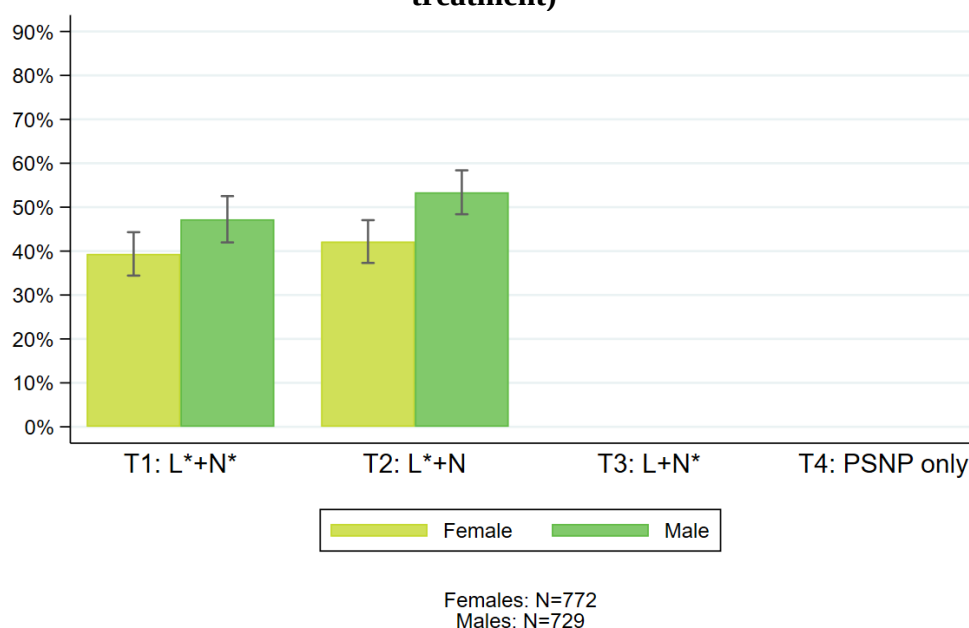
In a randomly chosen 50 percent of L* kebeles, livelihood transfers were complemented with an aspirations treatment—documentaries designed to motivate individuals and incentivize them to undertake forward-looking actions. In kebeles randomly assigned to receive the aspirations treatment, adult men and women in sample households were invited to attend the movie screenings, which took place in December

¹⁶ Using exchange rates from April 2019.

¹⁷ Using exchange rates from April 2019.

2018, before the distribution of livelihood transfers. The midline survey asked respondents about their attendance, and the rates can be seen in Table 5.5 and Table 5.6 as well as in Figure 5.4. The latter shows that attendance was low in both treatment arms, averaging 43 percent in T1 and 48 percent in T2. Furthermore, females were significantly less likely to attend across treatment arms and regions: on average, only 40.8 percent of invited females recall attending the screening, compared to 50.5 percent of males. Very low attendance was, again, concentrated in specific kebeles, some of which reported an attendance rate of zero.¹⁸ As the midline survey took place nearly half a year after the intervention, it is likely that respondents also couldn't recall the screenings, exemplified by the answers given to the question about the date of the attended screening that range from July 2017 to July 2019.

Figure 5.4: Attendance of aspirations video screenings by gender (only kebeles assigned to treatment)



5.2.4 Social Analysis and Action

A final part of the L* specific livelihood activities is the social analysis and action (SAA) intervention. Its goal is to address constraints on women's role in intrahousehold decision-making, mobility, and choice of livelihood activities. Table 5.7 and

Table 5.8 report answers to questions on SAA-related activities given by females and males, respectively.

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. Thus, attendance at VESA meetings is a precondition for being part of any SAA-related discussions. As the two tables show, roughly half of females and males have attended

¹⁸ Due to an implementation error, females were originally not invited to screenings in some kebeles in Meket and Dehena woreda, and while they eventually got the invitations, female attendance rate is ~40 percent lower in the aforementioned kebeles.

VESA group meetings in the last 12 months: among females, the shares are 48.3, 45.1 and 45.2 percent for T1, T2 and T3, respectively. The share of men who indicate attendance is slightly higher in all arms, with 49.7, 54.4 and 47.4 percent attending in the three arms, respectively. In the control group, roughly 8 percent of females and 11 percent of males report having attended a meeting. Across the sample, females recall participating in 5.9 meetings in last 12 months, and males in 6.8 meetings, on average. Once excluding respondents who never attended a single meeting, both females and males participated in roughly one meeting per month, on average.

Less than half of both females and males say that gender topics or couples' relationship questions were discussed at least once at VESA meetings, the rate being a little higher for males (44 percent vs. 42 percent for females). On average, females recall these topics coming up 1.9 times and males recall 2.1 times over the course of previous year. Those who indicate discussing gender or couples issues at the meetings were asked about coverage of specific subtopics. The most broadly covered topic appears to be workload sharing—72 percent of females and 75 percent of males recalls discussing it. Other topics that were widely recalled are assigning gender equitable tasks to boys and girls (69 percent of females and 72 percent of males remember discussing it), respect or mutual understanding (62 percent and 61 percent), and male engagement in childcare (61 percent and 58 percent). Lastly, respondents were asked if they have observed any changes in their spouse's behavior or actions as a result of these discussions. Roughly 56 percent of females and 70 percent of males who were part of the gender discussions indicate observing changes in their husband's or wife's behavior, respectively, meaning a considerable gender gap of roughly 14 percentage points between what has been perceived.

5.3 Nutrition components of the program

5.3.1 Discussions on health and nutrition

Another central component of VESA group meetings are discussions on hygiene and health, nutrition, child feeding and childcare. Table 5.9 and Table 5.10 show the exposure to nutrition- and hygiene-related topics as reported by females and males, respectively.

As indicated above, females in treatment arms T1 – T3 report attending 5.7 to 6.3 VESA meetings in the past year, on average, while males in these arms report attending 6.3 to 7.5 meetings on average. Females who did attend at least one meeting report the number of times that health and nutrition topics were discussed being 3.0 in T1, 3.4 in T2, and 2.9 in T3. For males, these numbers are a little higher at 3.2, 3.4 and 3.6 times for T1, T2 and T3, respectively. Both genders report very similar coverage of topics: 50 percent of males and 48 percent of females report covering some hygiene topics—the aggregate category for mentions of handwashing, food hygiene, safe water and waste management, and latrine use—while the most widely recalled individual topic was food hygiene, with roughly 35 percent of the sample mentioning it. A somewhat lower share, roughly 43 percent of both females and males, report discussing some nutrition topics—aggregate category for mentions of IYCF practices, adolescent girls' nutrition, maternal nutrition, and nutritious cooking—with the individual topic most commonly mentioned here being IYCF practices, recalled by roughly 30 percent. In regard to differences between treatment arms, the coverage of most topics is reported with somewhat higher likelihood by respondents in T3 group.

5.3.2 Male engagement groups

While male engagement groups had not been formally launched by the time of the midline survey, the questionnaire did ask male survey respondents about their exposure to topics relevant to such groups. Yet,

as seen in Table 5.10, 15 to 21 percent of males in T1 – T3 report already participating in a male group, which is likely due to respondents confusing the groups with VESA gender discussions, covered above. Respondents’ answers to follow-up questions on topics covered therefore shed more light on their recall of gender discussions. Nearly 61 percent of men who responded indicate covering gender roles and division of labor, roughly 45 percent recall caregiving and related division of labor, and 37 percent indicate discussing control of income and productive assets within household. This aligns with what was observed about VESA gender discussions (see Table 5.8).

5.3.3 Food demonstrations and community health services

The SPIR programming goal of improved access to health services is covered in the following section as well as in Chapter 7, which discusses respondents’ interaction with Health Extension Workers (HEWs) and the Health Development Army (HDA). Table 5.11: Exposure to food demonstrations and health services and the following section present reported health post visits and participation in behavior change communication (BCC) sessions, sanitation-related events, and community food demonstrations.

As shown in the table, 80.5 percent of primary females indicate having visited a health post at least once, with near negligible differences of maximum 3.5 percentage points among the four treatment arms. Of those who have visited a health post, roughly 47 percent did so in the previous three months for reasons related to their child or themselves. Again, the differences among treatment arms are marginal. Nearly 49 percent of everyone who had visited a health post indicate receiving advice or information about breastfeeding, child feeding or nutrition at their last visit, and here the share of receivers is slightly higher among N* communities: the percentage of those who recall receiving such information is roughly 54 percent in T1, 46 percent in T2, 51 percent in T3, and 44 percent in T4.

The table also shows that 38 percent of female respondents have attended a BCC session to talk about maternal nutrition, breastfeeding, complementary feeding, or child nutrition at a public work site or health post, the share ranging from 27.1 percent in the control group to 42.3 percent in T1. 17.9 percent of females, or a little less than half of those who indicate attending anytime, have done so in the previous 3 months, with the highest share, again, within T1 and T3 treatment arms. Roughly 76 percent of participating women said that the leader of the BCC session was a HEW, while 11 percent said it was a SPIR Health Community Facilitator.

All women were asked whether they attended a community-led total sanitation and hygiene (CLTSH) triggering event, meant to enhance the skills of the participants in triggering local collective action related to sanitation and hygiene. Table 5.11 shows a participation rate of roughly 41 percent in the full sample, with T1 and T3 arms’ attendance being about 5 percentage points higher than that of T2, and about 13 percentage points higher than T4. A CLTSH follow-up known as Open Defecation Free (ODF) event was less attended, with 23 percent of females or roughly two times less than in case of CLTSH saying that they participated. Again, T1 and T3 had the highest share of participants, T2 had about two-thirds of the reported participation rate of T1 and T3, and T4 about half of T1 and T3 rate. Finally, roughly 17 percent of female respondents recalls attending a Growth Monitoring and Promotion session. The differences among treatment arms are similar to those for previous questions in direction and magnitude. Finally, SPIR also organized cooking demonstrations within communities. The sessions were conducted using food ingredients that are offered and currently available in local markets. At midline survey, 30.4 percent of women indicates attending a food demonstration in their community, and reported attendance is higher in N* communities: the respective percentages by treatment arm are 40.6, 27.7, 35.0, and 17.8 percent. Women who report participation were asked about the advice or messages they received at the last food

demonstration they attended. As seen in Table 5.11, the most widely recalled lessons are on how to make porridge for children (mentioned by 81 percent of those who attended), how to feed children (67 percent), and locally available nutritious foods (60 percent). While the differences among treatment arms are small conditional on participation, the lessons with highest relative discrepancies between N* and other arms are all related to sanitation: safety of drinking water, handwashing, and proper dishwashing (see Table 5.11).

5.4 Conclusions

This chapter summarized results from the midline 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 effectively reached households with its complex programming and mostly maintained fidelity to the experimental design, though there are some important gaps. In the main treatment arms (T1, T2, T3), participation in VESA groups was high, but not universal, with 76-79 percent of households reporting that they participated in a VESA group. In the control group (T4), 16 percent of households reported belonging to a VESA group, though these control households may have mistaken common Village Savings and Loan Associations (VSLAs) for the SPIR project VESA groups. Roughly half of women and men in T1-T3 reported attending a VESA meeting in the last 12 months, with women attending 6 meetings on average and men attending 7 meetings on average, suggesting that those who attended at all went to meetings monthly.

Nearly all households identified as poor and assigned to receive the poultry package reported receiving it (93-95), but receipt of the cash livelihood transfers was more limited, with only 60-66 percent of targeted poor households reporting that they received the funds. In the half of L* kebeles that were randomized to receive the aspirations treatment, only 43-48 percent of households reported that a member attended the documentary screening session; respondents may not recall the screening which took place 9 months earlier. Participation in nutrition and WASH training activities varied, but the N* kebeles (T1 and T3) had the highest share of participants; T2 had about two-thirds of the reported participation rate of T1 and T3, and T4 about half of T1 and T3 rate. Conditional on attending a training session, the lessons with highest relative discrepancies between N* and other arms are all related to sanitation: safety of drinking water, handwashing, and proper dishwashing.

Table 5.1: Participation in PSNP

	N	All	Amhara	Oromia	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Household participated in PNSP Public Works activities between HAMLE 2009 and SENE 2010	3,174	0.895 (0.306)	0.865 (0.342)	0.937 (0.243)	0.910 (0.286)	0.899 (0.301)	0.880 (0.326)	0.891 (0.311)
Household participated in PNSP Public Works activities between HAMLE 2010 and SENE 2011	3,172	0.897 (0.305)	0.861 (0.346)	0.946 (0.226)	0.900 (0.300)	0.898 (0.303)	0.893 (0.309)	0.895 (0.306)
Household received Direct Support payments between HAMLE 2009 and SENE 2010	3,071	0.356 (0.479)	0.415 (0.493)	0.273 (0.446)	0.365 (0.482)	0.369 (0.483)	0.330 (0.471)	0.357 (0.480)
Household received Direct Support payments between HAMLE 2010 and SENE 2011	3,098	0.360 (0.480)	0.409 (0.492)	0.292 (0.455)	0.384 (0.487)	0.384 (0.487)	0.316 (0.465)	0.353 (0.478)
Male describes the household as having 'graduated' from the PSNP	3,129	0.007 (0.082)	0.003 (0.057)	0.011 (0.106)	0.001 (0.036)	0.012 (0.110)	0.006 (0.079)	0.007 (0.082)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.2: VESA membership

	N	All	Amhara	Oromia	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Someone in the household is a member of a VESA group	3,215	0.630 (0.483)	0.657 (0.475)	0.592 (0.492)	0.787 (0.410)	0.780 (0.415)	0.759 (0.428)	0.163 (0.370)
Primary female or her household member is a member of a VESA group	3,190	0.536 (0.499)	0.589 (0.492)	0.464 (0.499)	0.688 (0.463)	0.676 (0.468)	0.642 (0.480)	0.115 (0.320)
Primary male or his household member is a member of a VESA group	3,174	0.592 (0.492)	0.624 (0.485)	0.547 (0.498)	0.741 (0.438)	0.739 (0.439)	0.714 (0.452)	0.142 (0.350)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.3: Financial education and livelihoods (females)

	N	All	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Participated in SPIR financial education trainings or discussions	3,078	0.209 (0.406)	0.254 (0.435)	0.291 (0.454)	0.233 (0.423)	0.043 (0.203)
Number of financial education trainings attended	640	2.388 (2.309)	2.373 (2.190)	2.198 (1.937)	2.527 (2.335)	3.065 (4.449)
Participated in SPIR value chain trainings	3,074	0.112 (0.316)	0.157 (0.364)	0.152 (0.359)	0.122 (0.328)	0.010 (0.098)
Participated in fodder management and feeding practice training	345	0.174 (0.380)	0.151 (0.360)	0.172 (0.379)	0.216 (0.414)	0.000 (0.000)
Participated in livestock fattening and marketing training	345	0.400 (0.491)	0.353 (0.480)	0.402 (0.492)	0.454 (0.500)	0.429 (0.535)
Participated in poultry production training	345	0.649 (0.478)	0.731 (0.445)	0.689 (0.465)	0.515 (0.502)	0.429 (0.535)
Participated in market-driven agricultural production training	345	0.168 (0.375)	0.160 (0.368)	0.107 (0.310)	0.247 (0.434)	0.286 (0.488)
Member of a producer marketing group associated with a SPIR promoted value chain	3,065	0.075 (0.264)	0.115 (0.319)	0.114 (0.318)	0.062 (0.241)	0.004 (0.065)
Producer marketing group is associated with sheep/goat fattening	230	0.252 (0.435)	0.310 (0.465)	0.198 (0.401)	0.245 (0.434)	0.333 (0.577)
Producer marketing group is associated with poultry production	230	0.687 (0.465)	0.575 (0.497)	0.769 (0.424)	0.735 (0.446)	0.667 (0.577)
Participated in a training on natural resource management (FMNR, climate-smart agriculture)	3,834	0.043 (0.203)	0.057 (0.232)	0.054 (0.227)	0.038 (0.190)	0.022 (0.147)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.4: Financial education and livelihoods (males)

	N	All	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Participated in SPIR financial education trainings or discussions	2,943	0.225 (0.418)	0.280 (0.449)	0.272 (0.445)	0.267 (0.443)	0.076 (0.265)
Number of financial education trainings attended	663	2.333 (2.482)	2.422 (3.090)	2.244 (2.260)	2.313 (2.097)	2.426 (2.133)
Participated in SPIR value chain trainings	2,936	0.111 (0.315)	0.159 (0.366)	0.150 (0.357)	0.108 (0.310)	0.025 (0.158)
Participated in sheep/goat fattening or rearing training	327	0.639 (0.481)	0.540 (0.501)	0.704 (0.458)	0.728 (0.448)	0.444 (0.511)
Participated in poultry production training	327	0.648 (0.478)	0.699 (0.461)	0.678 (0.469)	0.580 (0.497)	0.444 (0.511)
Participated in ox fattening training	327	0.150 (0.357)	0.133 (0.341)	0.070 (0.256)	0.235 (0.426)	0.389 (0.502)
Participated in beekeeping/honey training	327	0.024 (0.155)	0.018 (0.132)	0.026 (0.160)	0.025 (0.156)	0.056 (0.236)
Participated in haricot bean production training	327	0.086 (0.280)	0.097 (0.298)	0.070 (0.256)	0.074 (0.264)	0.167 (0.383)
Participated in vegetable production training	327	0.116 (0.321)	0.097 (0.298)	0.078 (0.270)	0.173 (0.380)	0.222 (0.428)
Member of a producer marketing group associated with a SPIR promoted value chain	2,905	0.073 (0.260)	0.115 (0.320)	0.115 (0.319)	0.050 (0.217)	0.010 (0.099)
Producer marketing group is associated with sheep/goat fattening	211	0.351 (0.478)	0.407 (0.494)	0.279 (0.451)	0.432 (0.502)	0.143 (0.378)
Producer marketing group is associated with poultry production	211	0.526 (0.501)	0.469 (0.502)	0.628 (0.486)	0.486 (0.507)	0.143 (0.378)
Participated in a training on natural resource management (FMNR, climate smart agriculture)	3,661	0.111 (0.314)	0.129 (0.335)	0.140 (0.348)	0.085 (0.279)	0.089 (0.285)
Number of natural resource management trainings attended	405	2.309 (1.850)	2.482 (2.062)	2.295 (1.937)	2.152 (1.468)	2.237 (1.730)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.5: Livelihood transfers and aspirations, only targeted households (females)

	N	All	T1: L* + N*	T2: L* + N
Received an improved breed poultry start-up package	448	0.902 (0.298)	0.918 (0.275)	0.886 (0.318)
Number of chickens received from start-up package	403	15.742 (1.698)	15.507 (2.055)	15.975 (1.207)
Number of chickens left at the time of survey	404	10.099 (4.954)	9.682 (4.820)	10.512 (5.062)
Received chickens were of Sasso breed	387	0.894 (0.308)	0.896 (0.306)	0.892 (0.311)
Amount of cash received from the poultry start-up package	401	1,265.980 (611.942)	1,239.608 (629.967)	1,291.960 (594.070)
Received an unconditional cash grant from an NGO (ORDA, CARE, World Vision)	450	0.562 (0.497)	0.527 (0.500)	0.596 (0.492)
Amount of cash received as the unconditional cash grant	252	5,452.937 (791.563)	5,476.207 (868.842)	5,433.088 (721.835)
Attended an aspirations video screening	772	0.408 (0.492)	0.394 (0.489)	0.422 (0.494)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.6: Livelihood transfers and aspirations, only targeted households (males)

	N	All	T1: L* + N*	T2: L* + N
Received an improved breed poultry start-up package	416	0.800 (0.400)	0.805 (0.397)	0.796 (0.404)
Number of chickens received from start-up package	333	15.832 (1.247)	15.709 (1.546)	15.952 (0.846)
Number of chickens left at the time of survey	333	10.066 (4.524)	9.952 (4.274)	10.179 (4.767)
Received chickens were of Sasso breed	322	0.882 (0.323)	0.867 (0.341)	0.896 (0.306)
Amount of cash received from the poultry start-up package	333	1,296.210 (769.287)	1,313.636 (941.941)	1,279.095 (551.914)
Received an unconditional cash grant from an NGO (ORDA, CARE, World Vision)	452	0.454 (0.498)	0.437 (0.497)	0.470 (0.500)
Amount of cash received as the unconditional cash grant	205	5,440.634 (805.248)	5,537.938 (726.713)	5,353.241 (863.791)
Attended an aspirations video screening	729	0.505 (0.500)	0.472 (0.500)	0.534 (0.500)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.7: Social analysis and action (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	3,048	0.373 (0.484)	0.483 (0.500)	0.451 (0.498)	0.452 (0.498)	0.077 (0.267)
Number of VESA group meetings/discussions attended in past 12 months	3,048	4.749 (12.774)	5.654 (9.748)	6.252 (20.053)	5.923 (10.225)	0.748 (3.549)
Gender topics/couples' relationships were discussed at least once at VESA meetings	1,096	0.418 (0.493)	0.379 (0.486)	0.457 (0.499)	0.427 (0.495)	0.365 (0.486)
Number of times gender topics/couples' relationships were discussed at VESA meetings	1,096	1.892 (4.475)	1.407 (3.660)	2.296 (5.092)	2.000 (4.468)	1.885 (5.098)
Workload sharing was covered at VESA meetings	458	0.716 (0.451)	0.713 (0.454)	0.744 (0.438)	0.694 (0.462)	0.684 (0.478)
Male engagement in childcare was covered at VESA meetings	458	0.607 (0.489)	0.574 (0.496)	0.628 (0.485)	0.619 (0.487)	0.579 (0.507)
Respect or mutual understanding was covered at VESA meetings	458	0.620 (0.486)	0.625 (0.486)	0.622 (0.487)	0.605 (0.490)	0.684 (0.478)
Assigning gender equitable tasks to boys and girls was covered at VESA meetings	458	0.694 (0.461)	0.647 (0.480)	0.699 (0.460)	0.701 (0.460)	0.947 (0.229)
Shared decision-making on household finance was covered at VESA meetings	458	0.362 (0.481)	0.397 (0.491)	0.321 (0.468)	0.361 (0.482)	0.474 (0.513)
Violence against women was covered at VESA meetings	458	0.341 (0.474)	0.382 (0.488)	0.340 (0.475)	0.286 (0.453)	0.474 (0.513)
Has observed changes in spouse's behavior or action	401	0.559 (0.497)	0.575 (0.496)	0.514 (0.502)	0.603 (0.491)	0.455 (0.522)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.8: Social analysis and action (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	3,059	0.413 (0.492)	0.497 (0.500)	0.544 (0.498)	0.474 (0.500)	0.110 (0.313)
Number of VESA group meetings/discussions attended in past 12 months	3,059	5.457 (12.953)	6.301 (10.140)	7.534 (20.142)	6.561 (10.267)	1.046 (3.903)
Gender topics/couples' relationships were discussed at least once at VESA meeting	1,237	0.440 (0.497)	0.456 (0.499)	0.450 (0.498)	0.438 (0.497)	0.316 (0.468)
Number of times gender topics/couples' relationships were discussed at VESA meetings	1,237	2.100 (4.661)	1.748 (3.666)	2.571 (5.286)	1.975 (4.616)	1.848 (5.397)
Workload sharing was covered at VESA meetings	544	0.746 (0.436)	0.715 (0.453)	0.751 (0.433)	0.785 (0.412)	0.680 (0.476)

Male engagement in childcare was covered at VESA meetings	544	0.581 (0.494)	0.628 (0.485)	0.524 (0.501)	0.608 (0.490)	0.520 (0.510)
Respect or mutual understanding was covered at VESA meetings	544	0.614 (0.487)	0.593 (0.493)	0.635 (0.483)	0.601 (0.491)	0.680 (0.476)
Assigning gender equitable tasks to boys and girls was covered at VESA meetings	544	0.717 (0.451)	0.703 (0.458)	0.714 (0.453)	0.722 (0.450)	0.800 (0.408)
Shared decision-making on household finance was covered at VESA meetings	544	0.393 (0.489)	0.372 (0.485)	0.397 (0.491)	0.386 (0.488)	0.560 (0.507)
Violence against women was covered at VESA meetings	544	0.342 (0.475)	0.331 (0.472)	0.344 (0.476)	0.329 (0.471)	0.480 (0.510)
Has observed changes in spouse's behavior or action	516	0.703 (0.457)	0.701 (0.459)	0.711 (0.455)	0.709 (0.456)	0.625 (0.495)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.9: Health and nutrition discussions (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	3,048	0.373 (0.484)	0.483 (0.500)	0.451 (0.498)	0.452 (0.498)	0.077 (0.267)
Number of VESA group meetings/discussions attended in past 12 months	3,048	4.749 (12.774)	5.654 (9.748)	6.252 (20.053)	5.923 (10.225)	0.748 (3.549)
Number of time health and nutrition topics were discussed in VESA group	1,100	3.088 (5.619)	2.956 (6.302)	3.419 (5.677)	2.925 (4.789)	2.902 (5.375)
Nutrition topics (IYCF practices, girls' or maternal nutrition, nutritious cooking) were covered at VESA meetings	1,135	0.430 (0.495)	0.430 (0.496)	0.398 (0.490)	0.460 (0.499)	0.444 (0.502)
Hygiene topics (handwashing, food hygiene, safe water management, latrine use) were covered at VESA meetings	1,135	0.475 (0.500)	0.478 (0.500)	0.440 (0.497)	0.511 (0.501)	0.444 (0.502)
Saving money for health emergency was covered at VESA meetings	1,135	0.206 (0.405)	0.216 (0.412)	0.193 (0.395)	0.212 (0.409)	0.185 (0.392)
Healthy timing and spacing of pregnancy were covered at VESA meetings	1,135	0.121 (0.326)	0.122 (0.327)	0.118 (0.323)	0.119 (0.324)	0.148 (0.359)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.10: Health and nutrition discussions (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	3,059	0.413 (0.492)	0.497 (0.500)	0.544 (0.498)	0.474 (0.500)	0.110 (0.313)
Number of VESA group meetings/discussions attended in past 12 months	3,059	5.457 (12.953)	6.301 (10.140)	7.534 (20.142)	6.561 (10.267)	1.046 (3.903)
Number of time health and nutrition topics were discussed in VESA group	1,232	3.391 (5.412)	3.195 (5.066)	3.609 (5.610)	3.396 (5.350)	3.114 (6.177)
Nutrition topics (IYCF practices, girls' or maternal nutrition, nutritious cooking) were covered at VESA meetings	1,263	0.428 (0.495)	0.404 (0.491)	0.413 (0.493)	0.478 (0.500)	0.380 (0.488)
Hygiene topics (handwashing, food hygiene, safe water management, latrine use) were covered at VESA meetings	1,263	0.500 (0.500)	0.512 (0.501)	0.499 (0.501)	0.489 (0.501)	0.506 (0.503)
Saving money for health emergency was covered at VESA meetings	1,263	0.226 (0.419)	0.206 (0.405)	0.249 (0.433)	0.218 (0.413)	0.241 (0.430)
Healthy timing and spacing of pregnancy were covered at VESA meetings	1,263	0.112 (0.316)	0.106 (0.308)	0.109 (0.311)	0.116 (0.320)	0.152 (0.361)
Attended a men's engagement group in past 12 months	2,903	0.150 (0.357)	0.187 (0.390)	0.153 (0.360)	0.206 (0.405)	0.050 (0.218)
Number of men's engagement group meetings attended in past 12 months	434	3.839 (5.167)	3.542 (4.774)	4.435 (5.742)	3.765 (5.167)	3.314 (4.594)
Gender roles and division of labor were covered at engagement groups	435	0.605 (0.489)	0.682 (0.468)	0.583 (0.495)	0.582 (0.495)	0.486 (0.507)
Caregiving and division of caregiving were covered at engagement groups	435	0.448 (0.498)	0.485 (0.502)	0.452 (0.500)	0.444 (0.499)	0.314 (0.471)
Understanding nutrition was covered at engagement groups	435	0.287 (0.453)	0.235 (0.426)	0.374 (0.486)	0.275 (0.448)	0.257 (0.443)
Household decision-making was covered at engagement groups	435	0.278 (0.449)	0.303 (0.461)	0.322 (0.469)	0.235 (0.426)	0.229 (0.426)
Control of income, productive assets, and resources were covered at engagement groups	435	0.368 (0.483)	0.311 (0.465)	0.383 (0.488)	0.359 (0.481)	0.571 (0.502)
Gender norms, beliefs and expectations were covered at engagement groups	435	0.320 (0.467)	0.326 (0.470)	0.348 (0.478)	0.301 (0.460)	0.286 (0.458)
Participation in leadership and public spaces were covered at engagement groups	435	0.154 (0.361)	0.152 (0.360)	0.148 (0.356)	0.150 (0.359)	0.200 (0.406)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

Table 5.11: Exposure to food demonstrations and health services

	N	All	T1: L* + N*	T2: L* + N	T3: L + N*	T4: PSNP4 only
Has visited the health post at least once	3,205	0.805 (0.397)	0.824 (0.381)	0.798 (0.402)	0.808 (0.394)	0.789 (0.408)
Visited the health post for reasons related to child/herself in past 3 months	2,578	0.474 (0.499)	0.489 (0.500)	0.481 (0.500)	0.459 (0.499)	0.468 (0.499)
At last health post visit, received advice/information about breastfeeding, child feeding or nutrition	2,550	0.487 (0.500)	0.539 (0.499)	0.463 (0.499)	0.505 (0.500)	0.438 (0.497)
Primary female has attended a BCC session at a public work site/health post	3,184	0.380 (0.485)	0.423 (0.494)	0.368 (0.483)	0.449 (0.498)	0.271 (0.445)
Has attended a BCC session at public work site/health post in past 3 months	3,182	0.179 (0.384)	0.217 (0.413)	0.154 (0.361)	0.233 (0.423)	0.109 (0.312)
The last BCC session was led by a HEW	1,160	0.759 (0.428)	0.753 (0.432)	0.785 (0.412)	0.720 (0.449)	0.804 (0.398)
The last BCC session was led by a SPIR Health Community Facilitator/Health Animator	1,160	0.110 (0.313)	0.136 (0.343)	0.069 (0.255)	0.141 (0.349)	0.072 (0.259)
Primary female has attended a community-led total sanitation and hygiene (CLTSH)	3,188	0.408 (0.491)	0.447 (0.498)	0.403 (0.491)	0.457 (0.498)	0.319 (0.466)
Primary female has participated in a CLTSH follow-up or Open Defecation Free (ODF) event	3,185	0.230 (0.421)	0.288 (0.453)	0.209 (0.407)	0.270 (0.444)	0.149 (0.357)
Primary female has attended a Growth Monitoring and Promotion session	3,143	0.172 (0.378)	0.203 (0.403)	0.150 (0.357)	0.211 (0.408)	0.123 (0.329)
Primary female has attended a food demonstration in her community	3,180	0.304 (0.460)	0.406 (0.491)	0.277 (0.448)	0.350 (0.477)	0.178 (0.383)
Locally available beneficial foods were covered at last food demonstration	968	0.601 (0.490)	0.596 (0.492)	0.588 (0.493)	0.631 (0.483)	0.575 (0.496)
How to make porridge for children was covered at last food demonstration	968	0.808 (0.394)	0.837 (0.370)	0.763 (0.426)	0.812 (0.392)	0.806 (0.397)
How to feed children was covered at last food demonstration	968	0.666 (0.472)	0.680 (0.467)	0.636 (0.482)	0.679 (0.468)	0.657 (0.477)
How to cook nutritious meals was covered at last food demonstration	968	0.431 (0.495)	0.436 (0.497)	0.425 (0.495)	0.422 (0.495)	0.448 (0.499)
How to make drinking water safe was covered at last food demonstration	968	0.237 (0.425)	0.263 (0.441)	0.184 (0.389)	0.251 (0.434)	0.231 (0.423)
Handwashing was covered at last food demonstration	968	0.269 (0.443)	0.292 (0.455)	0.241 (0.429)	0.272 (0.446)	0.254 (0.437)
Washing dishes was covered at last food demonstration	968	0.198 (0.399)	0.232 (0.423)	0.167 (0.373)	0.195 (0.397)	0.179 (0.385)

Notes: Estimates from the DFSA SPIR Midline Survey sample. Standard deviations are in parentheses.

6 Impacts on Livelihood Outcomes

6.1 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 relatively poor households and the aspirations intervention.

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 poultry transfer and the salience of livestock production value chains in SPIR programming. We also report effects on savings (particularly via Village Economic and Savings Associations, or VESAs) and credit access. In order to analyze whether households are substituting into new sectors, we analyze engagement in labor outside the household; unfortunately, data on engagement in nonagricultural household businesses are extremely limited. Finally, we analyze reported individual aspirations for children's education, income and assets in order to assess whether households' economic aspirations have shifted.

6.1.1 Interpreting tables

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

The tables reporting 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 tables reporting effects for the 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). The tables then report coefficients for the effects of poultry, poultry in conjunction with aspirations, cash, and cash in conjunction with aspirations; each of these four coefficients is reported separately for T1 and T2, the two arms in which the cash, poultry, and aspirations interventions were cross-randomized. (Again, every poor household in T1 and T2 was randomized to receive either poultry or cash, with or without the additional aspirations intervention.) 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. The latter reports the average effects for poultry and cash abstracting from the additional estimated effect of the aspirations treatment. We again report the same tests of equality across treatment arms.

The tables reporting effects for the nonpoor 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 and T2, both in conjunction with the aspirations treatment, as well as T3. (No households in the nonpoor sample were assigned to receive

poultry or cash transfers.) We again report linear combinations corresponding to the average effects of T1 and T2.

As noted in Section 3.4, the reported p-values do not correct for multiple hypothesis testing, adjustments to p-values that account for estimation of multiple models for multiple outcomes of interest.

6.1.2 Participation in livelihood-related SPIR components

In order to interpret the results, it is useful to summarize the key observations from Chapter 5 around household participation in the livelihood-related components of SPIR programming. In general, participation in VESA groups is consistently high across all three arms, at nearly 80 percent. Participation in financial and value chain trainings is lower, ranging between 20 and 30 percent, and this pattern also reflects the more limited targeting of these trainings to households who are engaging in related household productive activities.

For the additional livelihood-related programming rolled out in the L* arms, reported receipt of the poultry package among intended package recipients exceeds 90 percent, while receipt of the cash transfers exceeds 70 percent. (Administrative records suggest higher rates of receipt, and it is possible that some households may not recall the transfer or may not attribute it accurately to SPIR.) Some households not targeted for transfers also report their receipt, and may reflect the misreporting of other transfers received through other programmatic channels. Attendance at screenings of the aspirations videos was significantly lower, below 50 percent, and was lower for women than for men. Further analysis will compare these reported rates to reported attendance data collected by SPIR itself.

6.2 Livestock ownership and sales

In this subsection, we analyze a series of variables capturing the household ownership and sales of livestock: whether the woman or household reports owning any poultry, and whether the household reports owning any other form of livestock; the total number of poultry owned by the woman and by the household; and whether the woman and the household report any income from sales of poultry. Note that for all variables, women report information about poultry that they own, while men report information about poultry owned at the household level.

6.2.1 Pooled effects by treatment arm

Table 6.1a reports pooled effects for the three treatment arms. 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 16 and 19 percentage points in columns (1) and (2), relative to a base probability of 57 percent for households in the control arm. There is, however, no shift in the probability that the household owns livestock other than poultry, as observed in column (3). The total number of poultry reported owned by the woman herself and the household increases by between 3 and 3.4, as reported in columns (4) and (5), relative to a mean in the control arm of around 2, suggesting the average size of the poultry flock has more than doubled.

However, there is no evidence of any shift in the probability that the woman or the household reports income from sales of poultry (columns 6 and 7). The probability that the woman reports any income from

egg sales, by contrast, has increased across all three treatment arms by 8 to 11 percentage points, relative to a mean in the control arm of 20 percent.

Both treatment arms T1 and T2 include L* interventions as well as transfers of the poultry package and cash to eligible (poor) households. Accordingly, these coefficients should be interpreted as an average effect for the full set of interventions. We also report the statistical test $T1=T2$, and it is evident that the hypothesis that the effects are equal across these two treatment arms cannot be rejected.

In the third treatment arm T3, only L interventions were rolled out. Here, there is evidence of a weakly significant increase in the probability that the woman reports owning any livestock (an increase of 7 percentage points on a base probability of 57 percentage points), and an increase in the total number of poultry owned by the household (an increase of 0.9 chickens, relative to a mean of 1.8 chickens). There is no evidence of any significant effect on the other outcomes of interest related to poultry and livestock. In general, the hypothesis that the effects of interest are equal in T3 vis-à-vis T1 and T2 can be rejected in columns (1) through (7), as the experimental effects are significantly smaller in T3. However, for the increased probability of income from egg sales, the hypothesis that these effects are equal across all three treatment arms cannot be rejected.

6.2.2 Poor households

Table 6.1b reports the experimental effects observed among poor households. All poor households in T1 and T2 were randomized to receive a cash transfer or a poultry package, and half were also cross-randomized to receive the aspirations intervention.

We observe in the first row that households receiving a poultry package in the first treatment arm show a positive effect on every dimension of reported livestock production. They are more likely to report owning poultry, more likely to report owning nonpoultry livestock, more likely to report owning a larger number of poultry, and more likely to report income from sales of poultry. The probability that the household owns any poultry increases by 43 percentage points relative to a mean of 54 percent in the control arm, and the size of the poultry flock increases by 8.3 relative to a mean of 1.7 in the control arm. The observed increase in flock size is consistent with the program's stated goals, in that households received 16 chickens but were expected to sell the 8 male chickens for meat within months, yielding a net increase of 8 chickens.

In general, the coefficients are parallel for reports of ownership by women and by men for the entire household, suggesting that the targeting of the transfer to women was effective: women are reporting the poultry as female owned. While both men and women are more likely to report that the household earned income from sales of poultry, the increase in the reported probability of sales is larger for men than for women (27 percentage points compared to 18 percentage points in T1, and 28 percentage points compared to 13 percentage points in T2). This suggests that men are somewhat more likely to benefit from increased commercialization opportunities following receipt of the poultry package. There is also a large increase (23 percentage points) in the probability women report any income from egg sales. In the second row of the table, there is no evidence of heterogeneity in these effects for households exposed to the aspirations treatment.

In the third row of the table, we observe that for households assigned to receive a cash transfer, there is some evidence of an increase in poultry production, but the effects observed are much smaller. There is a

small increase in the probability the woman or the household reports owning any poultry (between 13 and 15 percentage points), and an increase in the total number of poultry reported owned (of 1.4 chickens). While reported income from poultry sales does not increase, there is an increase in the probability of reporting income from egg sales that is narrowly significant (14 percentage points). There is no evidence of any effects on the other variables of interest, and again little evidence of any heterogeneity for households who were also exposed to the aspirations treatment in T1.

We now analyze parallel effects for poor households in T2. For households receiving the poultry transfer, the effects are qualitatively similar though slightly smaller in magnitude, and there is no evidence of any increase in the probability of owning nonpoultry livestock. However, for households receiving cash in T2, there is limited evidence of any positive effects of the cash transfer for poultry and livestock-related outcomes. Households receiving cash are less likely to report income from sales of poultry and livestock (consistent with the hypothesis they do not face any cash constraint that would require liquidating assets), but show no other positive effects. There is little evidence of any heterogeneity of the effects of the poultry package with respect to the aspirations treatment. For the cash transfer, the observed pattern in which households who receive cash are less likely to report income from sales of poultry or other livestock is reversed for households who were exposed to the aspirations intervention, suggesting that these households are continuing to commercialize livestock assets.

The linear combinations at the base of the table sum up the effects as observed on average. In general, the positive effects are slightly larger in T1, but the hypothesis that the experimental effects are equal across T1 and T2 can generally not be rejected. The effects of poultry are clearly significantly more positive for these outcomes than the effects of the cash transfer, suggesting that households who received cash are not primarily using the cash to invest in poultry or livestock-related outcomes.

6.2.3 Nonpoor households

Table 6.1c reports the experimental effects observed among nonpoor households; these households were not targeted for poultry or cash transfers, and accordingly the effects can be interpreted as simply the effect of the base L and L* interventions.

Here, we observe only some significant effects of the interventions of interest on livestock-related outcomes, but the effects are smaller in magnitude, and evident only for a subset of outcomes (primarily, poultry owned). In both T1 and T2, households are slightly more likely to report owning any poultry (in column (2)); we observe an experimental effect of 8 to 10 percentage points relative to a mean in the control arm of 62 percent. In T2 only, there is evidence of a significant increase in the number of poultry owned by the household, but a decline in the probability the woman or the household reports any income from sales of poultry, suggesting that households are retaining poultry rather than selling them. There is no evidence of any shift in the reported probability of income from egg sales.

If we examine the interaction effects with the aspirations treatment, again in general the coefficients of interest are insignificant. In T1, women who have been exposed to the aspirations treatment are less likely to report income from sales of poultry (a decline of 19 percentage points relative to a mean in the control arm of 37 percentage points), while in T1, women exposed to the aspirations treatment are weakly more likely to report income from sales of poultry. The channel for this different pattern is not immediately obvious.

Finally, if we examine the effects of T3, we observe only an increase in the number of poultry owned in column (5), an increase of one chicken relative to a mean in the control arm of 2.1. However, given the imprecision in the estimated coefficients, we cannot reject the hypotheses that the experimental effects are consistent across T1, T2, and T3.

6.3 Secondary outcomes linked to livestock ownership and sales

In this subsection, we analyze additional secondary outcomes linked to livestock ownership and sales: whether the woman and the household reports owning any “shoats” (a term meaning “sheep or goats”), whether the household reports owning any oxen, and any other type of livestock (other than poultry, shoats, or oxen); whether the woman or the household reports income from shoats; and whether the household reports income from oxen or other livestock.

6.3.1 Pooled effects by treatment arm

In Table 6.2a, we observe that women and households in T1 report an increase in the probability of owning any shoats, of 10 percentage points relative to a mean in the control arm of 50 percent. There is, however, a decline in the probability that women report any income from the sale of shoats, and a decline in the probability that the household reports any income from sales of other livestock. The latter effect is also evident in T2. There is no other evidence of meaningful experimental effects; the estimated effects for T3 are null, and the hypothesis that the effects in T1 and T3 are equal can generally be rejected (though the parallel test for effects in T2 and T3 fails to reject).

6.3.2 Poor households

In Table 6.2b, we observe that for households in T1, both poultry and cash recipients report an increase in the probability of owning any shoats of roughly equal magnitude, as observed in columns (1) and (2) in rows (1) and (3). (Both women and households report effects of equal magnitude.) Poultry recipients are also more likely to report ownership of any other livestock type, and less likely to report income from shoats. Cash recipients show no other effects other than a decline in the probability of earning income from oxen.

For households in T2 receiving poultry, the pattern is meaningfully different. These households show no increase in the probability of owning shoats, a decrease in the probability of owning oxen, and a decrease in the probability of reporting household income from sales of shoats or other livestock. By contrast, cash transfer households show an increase in the probability of owning shoats (parallel to what was observed in T1) and a decrease in the probability of income from shoats, oxen, or other livestock. For households exposed to the aspirations treatment, there is some evidence of heterogeneity in this effect, as the interaction effects are positive and significant (in column 7), suggesting that these households are not less likely to commercialize shoats or other livestock. This pattern is parallel to the pattern observed for sales of poultry in the previous subsection. Given that shoat raising may be a relatively new activity for these households, these patterns will certainly be examined again at endline.

The estimated effects for T3 are null, and the hypothesis that the effects in T1 vis-à-vis T3, and T2 vis-à-vis T3, are equal can generally be rejected. If we examine the linear combinations at the base of the table, it is evident that on average both poultry and cash recipient households report a parallel increase in the probability of owning shoats. Cash households report significant declines in the probability of income

from shoats, oxen or other livestock, though there is some more modest evidence of these effects for poultry-recipient households as well.

6.3.3 Nonpoor households

In Table 6.2c, we observe that there is generally no evidence of any experimental effects on these additional livestock variables for nonpoor households. The only exceptions are a weakly positive increase in the probability that the household reports any income from oxen sales in households also exposed to the aspirations treatment in both T1 and T2 (column (7), rows (2) and (4)). It does not seem that nonpoor households' ownership of other livestock types was significantly affected by the interventions.

6.4 Female savings

In this subsection, we analyze variables linked to female savings: whether the primary female in the household reports any personal savings, whether she deposits her savings in a VESA or Village Savings and Loan Association (VSLA), whether she deposits her savings with a microfinance institution (MFI), and the amount of savings in birr saved through any mechanism (a VESA, another financial institution, or an informal mechanism at home or with friends or family). The latter three variables are only reported if the woman reports any savings.

6.4.1 Pooled effects by treatment arm

We can observe in Table 6.3a that the SPIR interventions led to uniform shifts in savings behavior across all three treatment arms. There is a substantial increase in the probability that women report any savings, and that they deposit their savings with a VESA/VSLA. There is a decrease in the probability that women deposit savings with an MFI, and a decrease in the reported amount of savings. The latter effect is consistent with the hypothesis that the marginal woman induced to save by inclusion in SPIR programming saves a somewhat lower amount vis-à-vis a woman who would be saving in the absence of SPIR.

In this respect, it is important to note that given the differential savings rates, the number of women reporting any savings is dramatically different in the control arm (234 reporting any savings) vis-à-vis the three treatment arms (between 600 and 700 women reporting any savings in each treatment arm). If a substantial number of the “new savers” induced to save by the treatment are saving relatively small amounts, then this will generate a negative estimated treatment effect for the amount of savings, and this is exactly what is observed.

Moreover, the magnitude of the effects is large. The increase in the probability of reporting any female savings (reported in column (1)) is between 36 and 49 percentage points, relative to a probability of 25 percent in the control arm. Here, the effect observed is significantly smaller in T3. The increase in the probability of depositing savings with a VESA/VSLA ranges between 52 and 58 percentage points relative to a mean in the control arm of 34 percent, and this effect is consistent across treatment arms; similarly, the decrease in the probability of depositing savings with an MFI is around 25 percentage points, and consistent across treatment arms. The decline in the reported magnitude of savings on net ranges between 475 and 530 birr, a substantial decline relative to reported mean savings of 909 birr in the

control arm.¹⁹ The tests reported in the bottom rows of the table confirm that with the exception of the first binary variable capturing any female savings, the hypothesis that the observed effects are consistent across treatment arms cannot be rejected.

6.4.2 Poor households

In Table 6.3b, we observe that the effects for poor households are broadly similar to those reported in the pooled sample. The increase in the probability of savings is roughly consistent for households who receive poultry and cash in T1 and T2, and the other experimental effects similarly do not vary. There is no evidence of any heterogeneity with respect to the aspirations treatment. Again, similar effects are observed in T3, though the increase in the probability of female savings is somewhat smaller in magnitude.

6.4.3 Nonpoor households

In Table 6.3c, we again observe that the effects for nonpoor households are again broadly similar, with some exceptions. In T2, there is some evidence of heterogeneity with respect to aspirations, in that households exposed to the aspirations treatment are somewhat less likely to report increased savings in a VESA (as well as in an MFI), suggesting that these households may be using other savings mechanisms. In addition, the declines in the quantity of savings are noisily estimated, suggesting that the pattern in which the marginal saver reports a lower amount of savings is not as pronounced among nonpoor households.

In general, however, it seems that we can conclude that the effects of SPIR interventions on female savings are consistent both across arms and across the poor and nonpoor sample.

6.5 Male credit and savings

In this subsection, we analyze variables linked to credit and savings behavior as reported by the primary male of the household. The variables of interest include whether the man reports receipt of a loan for productive purpose, and, conditional on reporting any loan, whether the loan was received from a VESA and the total value of the loan. We also analyze a variable for whether the primary male reports a bank account, and whether he deposited money with a rural savings and credit cooperative (RuSACCO) in the last 12 months (conditional on membership in a RuSACCO).

6.5.1 Pooled effects by treatment arm

The results in Table 6.4a suggest there is little robust evidence that the SPIR interventions shaped credit access. There is no shift in the overall probability of reporting a loan for productive purposes; though the coefficients of interest are positive, they are statistically insignificant. Conditional on reporting a loan, there is a significant increase in the probability the loan is from a VESA; the increase is between 9 and 13 percentage points relative to a mean probability of 4 percent in the control arm, and it is consistent across arms. There is weakly significant evidence of a decline in the value of productive loans, but no evidence of any shift in the use of other financial services.

¹⁹ These effects are also consistent when estimated using the log of female savings as a dependent variable.

6.5.2 Poor households

In general, the effects estimated for poor households are somewhat varied (Table 6.4b). There is an increase in the probability of reporting any productive loan for households in T3. The increase in the probability of receiving a productive loan from VESA is observed for households receiving the poultry package and the aspirations treatment in T1; for households receiving the poultry package only in T2; and for households receiving the cash transfer and the aspirations treatment in T2. (For households receiving a poultry package in conjunction with the aspirations treatment in T2, the probability of a productive loan from a VESA declines.) There is some evidence of an increase in deposits in a RuSACCO for households receiving poultry in T2, but not for households also exposed to aspirations programming. These effects should be interpreted cautiously, however, as the confidence intervals do not allow us to reject the hypothesis that the effects are consistent across treatment arms.

6.5.3 Nonpoor households

In general, the effects estimated for nonpoor households are again somewhat heterogeneous and noisily estimated. There is evidence of an increase in the probability of receiving a productive loan in T2, and an increase in the probability that the loan is from a VESA in all three treatment arms. There is also some evidence of a decline in the probability of reporting a bank account, though only in T1. Again, these effects should be interpreted cautiously, as the confidence intervals do not allow us to reject the hypothesis that the effects are consistent across treatment arms.

6.6 Outside labor

In this subsection, we analyze variables capturing households' engagement in outside or wage labor. We analyze the following variables: Two variables capture whether any household member reported regular wage work or casual/irregular work in the last 12 months.²⁰ One variable captures whether the primary male's main economic activity is crop production. Two variables capture whether the primary female reported regular wage work or casual/irregular work in the last 12 months, and one variable captures whether the primary female's main economic activity is crop production.

6.6.1 Pooled effects by treatment arm

We can observe in Table 6.5a that there is little evidence of any significant effect of the SPIR interventions on wage labor in the pooled sample. The coefficients are small in magnitude and generally statistically insignificant.

6.6.2 Poor households

For poor households, the pattern of effects for wage labor is complex. Households receiving poultry transfers in T1 are less likely to report any casual/irregular wage work at the household level or for the primary female over the last year, as evident in the first row of the table; however, those also exposed to the aspirations treatment show a countervailing positive effect for regular wage work as well as casual

²⁰ In the survey, respondents receive the following clarifying examples to define these terms: examples of wage work include working as mechanic, truck driver, or government workers such as a teacher or a nurse. Examples of casual work include agricultural day labor and work herding cattle.

wage work. There is little evidence of effects for cash recipients in T1, though there is some noisy evidence that cash recipients also exposed to aspirations programming may be more likely to engage in outside labor.

In T2, households who receive poultry transfers are somewhat more likely to report any casual or irregular wage work at the household level, while those also exposed to the aspirations treatment show a countervailing negative effect. The channel for this different pattern in T2 is not clear. Again, there is little evidence of any significant effects for cash recipients, but there is some weak evidence that recipients also exposed to aspirations programming show an increase in engagement in outside labor.

When we examine the linear combinations reporting average effects in T1 and T2 and among poultry and cash recipients, these estimated linear combinations are again uniformly insignificant. This suggests that while there may be some movement (both increases and decreases) in engagement in wage work in different intervention sub-cells, on average the SPIR interventions have not stimulated any large shifts in wage work.

6.6.3 Nonpoor households

For nonpoor households, the results are broadly similar to the pooled sample in that there is little evidence of any meaningful shift in wage work. The only notable effect is a decrease in the probability that the primary female's main economic activity is crop production in T2, and an increase in T3. While the former effect would be consistent with nonpoor households substituting into new economic sectors as a result of L* interventions implemented in T2, the channel for the latter effect is unclear.

6.7 Expenditure on crop inputs

In this subsection, we analyze variables capturing households' investment in crop inputs. This includes the total amount reported paid for seed, the total amount paid for inorganic fertilizer, the total amount paid for pesticides, and the total amount paid for hired agricultural labor, all reported in birr in the last Mehr season.

Here, in all three specifications—the pooled sample, the sample of poor households, and the sample of nonpoor households—there is no evidence of any significant effect on crop expenditure. The coefficients of interest are varied in sign and generally not statistically significant.

6.8 Aspirations

In this subsection, we analyze variables capturing households' stated aspirations. All variables are reported in parallel by both women and men. The variables of interest include binary variables for the aspired level of education for the eldest child (no education, primary education, secondary education, or postsecondary education), and the aspired and expected level of income in 10 years. Men additionally report the expected and aspired level of assets in 10 years.

6.8.1 Women

The results reported for women's aspirations in Tables 6.7a through 6.7c generally suggest there is no significant effect of the interventions implemented on reported aspirations. The only exception is some weak evidence that the aspired level of income is higher in T1, as evident in column (5) of Table 6.7a. In

Table 6.7c, it seems that this effect is primarily driven by nonpoor households, who show a significant increase in aspired level of income in T1. It is important to note that there is no evidence that this effect is driven by the cross-randomized aspirations treatment per se; instead, it may reflect households' updated aspirations for future income given entry or expected entry into new economic activities as a result of the interventions rolled out in T1.

6.8.2 Men

The results reported for men's aspirations in Tables 6.8a through 6.8c again generally suggest there are limited significant effects of the interventions implemented on reported aspirations. In Table 6.8a reporting on the pooled sample results, there is some evidence of an increase in the probability respondents report a stated aspiration for postsecondary education (in T2 and T3) and, parallel to the results observed for women, an increase in the aspired level of income in T1.

In examining Tables 6.8b and 6.8c, it seems that both effects are primarily driven by the sample of nonpoor households.²¹ Within the set of nonpoor households, we observe an increase in the probability of reporting postsecondary education among households in arms T2 and T3, and an increase in the aspired level of income in T1. Again, there is little evidence that these patterns reflect the effect of the cross-randomized aspirations intervention per se.

6.9 Conclusions

This chapter has provided evidence about the effects of a range of SPIR interventions on livelihoods-related outcomes. While the poultry transfer has had a substantial positive effect on increasing household ownership of and income from poultry, comparable investments by households who received the one-time cash transfer have been much smaller in magnitude; the increase in the probability of owning poultry is 39 percentage points for poultry recipients, compared to 9 percentage points for cash recipients. (The probability of owning poultry for households in the control arm is 54 percent.) Poultry households also show differentially larger investment in other forms of livestock vis-à-vis cash recipient households, showing an increase in the probability of reporting nonpoultry livestock of 8 percentage points, relative to a mean in the control arm of 51 percent. The larger effects of the poultry transfer could be consistent with higher reported receipt rates of the poultry vis-à-vis the cash transfer, though administrative records suggest that rates of receipt were equally high for both transfers.

Nonpoor households exposed to base L* programming also increase their engagement in poultry production, showing an increase in the probability of owning poultry of between 8 and 10 percentage points, but there is no effect for households exposed only to L. The latter effect may be suggestive of positive programmatic spillover effects from poor households who are poultry recipients to nonpoor households.

For savings and credit, by contrast, the effects are consistent across all three treatment arms. There is a dramatic increase in the probability that women report any savings, driven primarily by women reporting savings in VESAs; this pattern is observed in all three treatment arms, and for both poor and nonpoor women. The increase in the probability of reported female savings is between 36 and 49 percentage

²¹ The increased probability of reporting post-secondary educational aspirations is also observed for poor households among cash recipients in T2.

points, relative to a mean of 25 percent in the control arm. (The amount of reported savings, however, decreased as a result of the interventions as a large number of new savers were reported in the intervention arms; this presumably reflects the fact that the marginal woman induced to save by SPIR programming save relatively little in absolute terms.) By contrast, there is little evidence of any shift in credit access. The substantial effect on female savings is consistent with the observed pattern of consistently high participation in VESA groups.

In evaluating outside labor, we find little evidence of any effects on engagement in outside labor for the pooled sample or the sample of nonpoor households. For the sample of poor households, there is some evidence of shifts in outside labor; however, the effects are varying for households in the two L* treatment arms (T1 and T2), and for households who are and are not exposed to aspirations programming. Finally, the effects on aspirations outcomes as reported by both men and women are relatively weak, though households in T1 show some evidence of an increase in aspired income (an increase in 13,000 birr relative to a mean of 36,000 birr in the control arm). This does not seem to reflect the effects of the targeted cross-randomized aspirations intervention. The aspirations intervention was also characterized by relatively low participation (less than 50 percent) in the data reported by households, though this pattern will be explored further using the SPIR project's administrative data.

In general, it seems accurate to conclude that the poultry transfer has stimulated significantly greater short-term transformation in productive activities vis-à-vis the cash transfer for the subset of outcomes measured in the midline survey. (We cannot evaluate potentially welfare-enhancing effects of the cash transfer along other dimensions, such as consumption or durable assets.) L* interventions have also been effective in increasing poultry production. The increase in female savings is notable across all treatment arms, but effects on other dimensions of livelihoods have been minimal.

**Table 6.1a: Livestock ownership and sales (primary outcomes)
by SPIR treatment arms**

	(1) Woman reports owning any poultry (solely or jointly owned)	(2) Man reports household owns any poultry	(3) Man reports household owns any livestock other than poultry	(4) Total number of poultry owned by female	(5) Total number of poultry owned by household (as reported by male)	(6) Woman reports any income from sales of poultry owned (solely or jointly) [§]	(7) Man reports any income from sales of poultry owned by household [§]	(8) Woman reports any income from egg sales in last 30 days
T1	0.172*** (0.040)	0.189*** (0.035)	0.022 (0.031)	3.145*** (0.596)	3.297*** (0.543)	0.010 (0.035)	0.052 (0.041)	0.114*** (0.032)
T2	0.163*** (0.037)	0.178*** (0.032)	0.030 (0.025)	2.936*** (0.591)	3.400*** (0.565)	-0.041 (0.035)	0.033 (0.040)	0.083*** (0.030)
T3	0.071* (0.037)	0.054 (0.033)	-0.005 (0.025)	0.580 (0.366)	0.895*** (0.311)	0.004 (0.033)	-0.008 (0.031)	0.076** (0.030)
Test: T1 = T2	0.803	0.743	0.782	0.769	0.884	0.147	0.676	0.368
Test: T2 = T3	0.004***	0.000***	0.120	0.000***	0.000***	0.185	0.250	0.829
Test: T1 = T3	0.006***	0.000***	0.348	0.000***	0.000***	0.855	0.105	0.273
Mean of control (T4)	0.573	0.574	0.829	2.098	1.859	0.347	0.374	0.206
N	3,127	3,189	3,189	3,220	3,220	2,204	2,244	3,098

Notes: [§]In the last 12 months. Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: Man reports household owns any poultry, Total number of poultry owned by male. * p<0.1 ** p<0.05 *** p<0.01

**Table 6.1b: Livestock ownership and sales (primary outcomes): Subsample of poor households
by SPIR subtreatment arms**

	(1) Woman reports owning any poultry (solely or jointly	(2) Man reports household owns any poultry	(3) Man reports household owns any livestock other than poultry	(4) Total number of poultry owned by female	(5) Total number of poultry owned by household (as reported	(6) Woman reports any income from sales of poultry owned (solely or	(7) Man reports any income from sales of poultry owned by household [§]	(8) Woman reports any income from egg sales in last 30 days
--	---	--	---	--	---	--	--	--

	owned)				by male)	jointly) §		
T1 x Poultry	0.418*** (0.047)	0.428*** (0.043)	0.083* (0.043)	8.530*** (1.192)	8.332*** (1.117)	0.182** (0.077)	0.272*** (0.066)	0.231*** (0.067)
T1 x Poultry x Aspirations ^a	0.013 (0.050)	0.005 (0.049)	-0.033 (0.057)	1.819 (1.601)	1.796 (1.543)	-0.041 (0.093)	0.036 (0.092)	-0.093 (0.092)
T1 x Cash	0.129* (0.076)	0.147** (0.070)	0.059 (0.073)	1.186 (0.917)	1.389* (0.837)	-0.064 (0.061)	-0.046 (0.075)	0.137* (0.071)
T1 x Cash x Aspirations ^a	-0.144 (0.093)	-0.147* (0.086)	-0.082 (0.101)	-1.202 (0.978)	-1.310 (0.965)	0.012 (0.080)	-0.006 (0.086)	-0.062 (0.079)
T2 x Poultry	0.375*** (0.044)	0.344*** (0.045)	-0.005 (0.042)	7.928*** (1.132)	8.190*** (1.023)	0.129** (0.053)	0.282*** (0.065)	0.140** (0.070)
T2 x Poultry x Aspirations ^a	0.026 (0.046)	0.085* (0.047)	0.073 (0.055)	1.369 (1.533)	1.761 (1.435)	-0.043 (0.069)	0.083 (0.074)	-0.091 (0.086)
T2 x Cash	0.024 (0.063)	0.017 (0.045)	0.061 (0.052)	0.268 (0.592)	0.499 (0.525)	-0.205*** (0.054)	-0.140* (0.078)	0.071 (0.050)
T2 x Cash x Aspirations ^a	0.101 (0.080)	0.097* (0.057)	0.082 (0.057)	0.557 (0.668)	0.516 (0.643)	0.140** (0.069)	0.066 (0.086)	0.097 (0.078)
T3	0.072* (0.043)	0.045 (0.038)	-0.023 (0.032)	0.451 (0.433)	0.605* (0.357)	0.049 (0.040)	0.035 (0.045)	0.087** (0.035)
Linear combination: effect of T1	0.239*** (0.040)	0.251*** (0.036)	0.042 (0.036)	4.980*** (0.569)	4.952*** (0.520)	0.050 (0.042)	0.119** (0.046)	0.144*** (0.037)
Linear combination: effect of T2	0.234*** (0.039)	0.228*** (0.033)	0.065** (0.028)	4.639*** (0.558)	4.972*** (0.485)	-0.011 (0.039)	0.111** (0.044)	0.108*** (0.037)
Linear combination: effect of poultry	0.397*** (0.038)	0.387*** (0.036)	0.038 (0.033)	8.247*** (0.851)	8.282*** (0.777)	0.156*** (0.051)	0.278*** (0.052)	0.185*** (0.051)
Linear combination: effect of cash	0.077 (0.053)	0.083* (0.045)	0.060 (0.047)	0.732 (0.603)	0.948* (0.527)	-0.133*** (0.046)	-0.092 (0.059)	0.104** (0.046)
Test: T1 = T2	0.879	0.452	0.513	0.578	0.973	0.114	0.862	0.376
Test: T2 = T3	0.000***	0.000***	0.005***	0.000***	0.000***	0.100*	0.055*	0.591
Test: T1 = T3	0.000***	0.000***	0.085*	0.000***	0.000***	0.971	0.047**	0.146
Mean of control (T4)	0.545	0.542	0.770	1.839	1.682	0.329	0.356	0.190
N	1,800	1,825	1,825	1,847	1,847	1,279	1,287	1,782

Notes: §In the last 12 months. Estimates from the 'poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: Man reports household owns any poultry, Total number of poultry owned by male. * p<0.1 ** p<0.05 *** p<0.01

Table 6.1c: Livestock ownership and sales (primary outcomes): Subsample of nonpoor households by SPIR subtreatment arms

	(1) Woman reports owning any poultry (solely or jointly owned)	(2) Man reports household owns any poultry	(3) Man reports household owns any livestock other than poultry	(4) Total number of poultry owned by female	(5) Total number of poultry owned by household (as reported by male)	(6) Woman reports any income from sales of poultry owned (solely or jointly) *	(7) Man reports any income from sales of poultry owned by household *	(8) Woman reports any income from egg sales in last 30 days
T1	0.028 (0.057)	0.081* (0.042)	0.000 (0.043)	0.069 (0.529)	0.527 (0.340)	0.016 (0.054)	-0.061 (0.060)	0.071 (0.048)
T1 x Aspirations ^a	0.079 (0.057)	0.025 (0.053)	-0.016 (0.040)	0.723 (0.462)	0.441 (0.468)	-0.187*** (0.063)	-0.031 (0.069)	-0.012 (0.054)
T2	0.030 (0.050)	0.097** (0.049)	-0.002 (0.037)	0.272 (0.538)	0.853** (0.421)	-0.176*** (0.047)	-0.155*** (0.046)	0.025 (0.046)
T2 x Aspirations ^a	0.067 (0.051)	0.017 (0.053)	-0.024 (0.041)	0.667 (0.638)	0.687 (0.693)	0.106* (0.059)	0.044 (0.061)	0.044 (0.060)
T3	0.061 (0.046)	0.060 (0.042)	0.018 (0.028)	0.589 (0.512)	1.048*** (0.350)	-0.057 (0.043)	-0.064 (0.041)	0.059 (0.040)
Linear combination: effect of T1	0.068 (0.046)	0.093** (0.039)	-0.008 (0.032)	0.430 (0.487)	0.748** (0.297)	-0.077* (0.043)	-0.077* (0.045)	0.065 (0.040)
Linear combination: effect of T2	0.063 (0.043)	0.106*** (0.039)	-0.014 (0.030)	0.599 (0.535)	1.190*** (0.381)	-0.123*** (0.042)	-0.134*** (0.044)	0.047 (0.042)
Test: T1 = T2	0.901	0.736	0.837	0.653	0.268	0.274	0.199	0.643
Test: T2 = T3	0.961	0.239	0.186	0.981	0.748	0.110	0.090*	0.760
Test: T1 = T3	0.871	0.402	0.322	0.654	0.414	0.635	0.758	0.867
Mean of control (T4)	0.613	0.618	0.912	2.458	2.106	0.371	0.396	0.228
N	1,327	1,364	1,364	1,373	1,373	925	957	1,316

Notes: [§]In the last 12 months. Estimates from the ‘non-poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: Man reports household owns any poultry, Total number of poultry owned by male.

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

**Table 6.2a: Livestock ownership and sales (secondary outcomes)
by SPIR treatment arms**

	(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 or goats owned (solely or jointly)	(6) Man reports household receives any income from sales of sheep or 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	0.100** (0.039)	0.100** (0.039)	-0.010 (0.036)	0.022 (0.031)	-0.053* (0.029)	-0.042 (0.034)	-0.050 (0.045)	-0.066** (0.032)
T2	0.045 (0.030)	0.021 (0.031)	-0.022 (0.032)	0.030 (0.025)	-0.035 (0.031)	-0.035 (0.035)	-0.059 (0.049)	-0.072** (0.033)
T3	0.017 (0.036)	0.015 (0.037)	-0.054 (0.033)	-0.005 (0.025)	0.011 (0.030)	0.009 (0.034)	0.030 (0.049)	-0.023 (0.030)
Test: T1 = T2	0.105	0.023**	0.740	0.782	0.531	0.854	0.826	0.877
Test: T2 = T3	0.357	0.827	0.310	0.120	0.120	0.227	0.058*	0.142
Test: T1 = T3	0.036**	0.034**	0.233	0.348	0.021**	0.150	0.065*	0.191
Mean of control (T4)	0.502	0.526	0.272	0.829	0.292	0.509	0.369	0.569
N	3,129	3,189	3,189	3,189	1,779	1,951	999	2,761

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

**Table 6.2b: Livestock ownership and sales (secondary outcomes): sub-sample of poor households
by SPIR sub-treatment arms**

	(1) Woman reports owning any sheep or goats (solely or	(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,	(5) Woman reports any income from sales of sheep or goats owned	(6) Man reports household receives any income from sales of sheep or	(7) Man reports household receives any income from sales of oxen	(8) Man reports any income from sales of livestock other than
--	--	--	---	--	---	--	--	---

	jointly owned)			sheep, goats and oxen	(solely or jointly)	goats owned	owned	poultry, sheep, goats
T1 x Poultry	0.176** (0.076)	0.189*** (0.072)	0.073 (0.055)	0.083* (0.043)	-0.122*** (0.042)	0.008 (0.068)	-0.077 (0.081)	-0.000 (0.059)
T1 x Poultry x Aspirations ^a	-0.052 (0.086)	-0.022 (0.082)	-0.107 (0.070)	-0.033 (0.057)	0.023 (0.050)	-0.079 (0.078)	0.146 (0.118)	-0.026 (0.074)
T1 x Cash	0.131* (0.079)	0.147* (0.077)	0.051 (0.068)	0.059 (0.073)	-0.022 (0.057)	-0.027 (0.060)	-0.186** (0.075)	-0.056 (0.058)
T1 x Cash x Aspirations ^a	-0.025 (0.091)	-0.063 (0.090)	-0.161* (0.087)	-0.082 (0.101)	-0.025 (0.072)	-0.177** (0.076)	0.036 (0.099)	-0.189*** (0.069)
T2 x Poultry	0.060 (0.046)	0.062 (0.052)	-0.083* (0.050)	-0.005 (0.042)	-0.067 (0.056)	-0.157*** (0.060)	-0.040 (0.088)	-0.147*** (0.048)
T2 x Poultry x Aspirations ^a	0.003 (0.054)	-0.049 (0.062)	0.081 (0.059)	0.073 (0.055)	-0.001 (0.073)	0.273*** (0.089)	0.107 (0.120)	0.175** (0.079)
T2 x Cash	0.148*** (0.043)	0.111** (0.047)	0.011 (0.056)	0.061 (0.052)	-0.054 (0.054)	-0.135** (0.063)	-0.257*** (0.075)	-0.183*** (0.067)
T2 x Cash x Aspirations ^a	-0.032 (0.062)	0.032 (0.063)	0.048 (0.074)	0.082 (0.057)	-0.072 (0.072)	0.010 (0.077)	0.175** (0.079)	0.027 (0.080)
T3	0.013 (0.036)	0.017 (0.039)	-0.070** (0.035)	-0.023 (0.032)	-0.001 (0.042)	0.003 (0.044)	0.013 (0.072)	-0.041 (0.042)
Linear combination: effect of T1	0.134*** (0.039)	0.146*** (0.040)	-0.006 (0.037)	0.042 (0.036)	-0.071** (0.034)	-0.074* (0.039)	-0.086 (0.058)	-0.082** (0.038)
Linear combination: effect of T2	0.096*** (0.031)	0.082** (0.034)	-0.005 (0.034)	0.065** (0.028)	-0.079** (0.037)	-0.077* (0.041)	-0.077 (0.058)	-0.115*** (0.041)
Linear combination: effect of poultry	0.117** (0.047)	0.125** (0.048)	-0.006 (0.041)	0.038 (0.033)	-0.094** (0.040)	-0.076 (0.050)	-0.058 (0.068)	-0.075* (0.044)
Linear combination: effect of cash	0.139*** (0.048)	0.129*** (0.049)	0.031 (0.047)	0.060 (0.047)	-0.037 (0.043)	-0.080* (0.047)	-0.220*** (0.062)	-0.118** (0.048)
Test: T1 = T2	0.311	0.085*	0.987	0.513	0.824	0.942	0.866	0.360
Test: T2 = T3	0.014**	0.071*	0.051*	0.005***	0.055*	0.063*	0.168	0.063*
Test: T1 = T3	0.004***	0.002***	0.080*	0.085*	0.072*	0.063*	0.148	0.277
Mean of control (T4)	0.456	0.476	0.207	0.770	0.327	0.508	0.393	0.558
N	1,801	1,825	1,825	1,825	978	1,073	432	1,494

Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available). * p<0.1 ** p<0.05 *** p<0.01

Table 6.2c: Livestock ownership and sales (secondary outcomes): Sub-sample of nonpoor households by SPIR sub-treatment arms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Woman reports owning any sheep or goats (solely or jointly owned)	Man reports household owns any sheep or goats	Man reports household owns any oxen	Man reports household owns livestock other than poultry, sheep, goats and oxen	Woman reports any income from sales of sheep or goats owned (solely or jointly)	Man reports household receives any income from sales of sheep or goats owned	Man reports household receives any income from sales of oxen owned	Man reports any income from sales of livestock other than poultry, sheep, goats
T1	0.021 (0.067)	0.022 (0.067)	-0.029 (0.052)	0.000 (0.043)	-0.048 (0.051)	0.022 (0.056)	-0.078 (0.065)	-0.029 (0.050)
T1 x Aspirations ^a	0.059 (0.067)	0.026 (0.068)	0.013 (0.066)	-0.016 (0.040)	0.019 (0.050)	-0.086 (0.055)	0.108* (0.061)	-0.068 (0.056)
T2	-0.032 (0.050)	-0.069 (0.048)	-0.008 (0.053)	-0.002 (0.037)	-0.037 (0.059)	0.008 (0.062)	-0.097 (0.070)	-0.051 (0.048)
T2 x Aspirations ^a	0.005 (0.054)	0.018 (0.054)	-0.073 (0.056)	-0.024 (0.041)	0.110 (0.069)	0.015 (0.075)	0.155* (0.083)	0.058 (0.059)
T3	0.021 (0.049)	0.011 (0.048)	-0.036 (0.045)	0.018 (0.028)	0.022 (0.044)	0.006 (0.048)	0.047 (0.067)	-0.005 (0.039)
Linear combination: effect of T1	0.050 (0.049)	0.035 (0.048)	-0.022 (0.045)	-0.008 (0.032)	-0.039 (0.042)	-0.021 (0.045)	-0.024 (0.058)	-0.064 (0.040)
Linear combination: effect of T2	-0.030 (0.044)	-0.060 (0.042)	-0.043 (0.043)	-0.014 (0.030)	0.017 (0.047)	0.015 (0.052)	-0.021 (0.064)	-0.023 (0.041)
Test: T1 = T2	0.055*	0.024**	0.629	0.837	0.159	0.444	0.956	0.319
Test: T2 = T3	0.216	0.082*	0.855	0.186	0.899	0.854	0.258	0.652
Test: T1 = T3	0.533	0.615	0.780	0.322	0.095*	0.534	0.182	0.144
Mean of control (T4)	0.568	0.596	0.363	0.912	0.250	0.510	0.350	0.583
N	1,328	1,364	1,364	1,364	801	878	567	1,267

Notes: Estimates from the 'non-poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

**Table 6.3a: Female's savings
by SPIR treatment arms**

	(1)	(2)	(3)	(4)
	Primary female has personal savings of her own	Primary female deposits her savings with VESA/VSLA	Primary female deposits her savings with an MFI	Primary female's amount of savings (birr)
T1	0.490*** (0.040)	0.551*** (0.075)	-0.245*** (0.057)	-498.405** (197.584)
T2	0.448*** (0.040)	0.530*** (0.076)	-0.238*** (0.060)	-474.758** (192.577)
T3	0.362*** (0.038)	0.576*** (0.074)	-0.236*** (0.060)	-530.096*** (195.534)
Test: T1 = T2	0.238	0.490	0.796	0.745
Test: T2 = T3	0.010**	0.106	0.960	0.422
Test: T1 = T3	0.000***	0.356	0.757	0.655
Mean of control (T4)	0.248	0.346	0.379	909.172
N	3,128	1,808	1,808	1,765

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

**Table 6.3b: Female's savings: Subsample of poor households
by SPIR subtreatment arms**

	(1)	(2)	(3)	(4)
	Primary female has personal savings of her own	Primary female deposits her savings with VESA/VSLA	Primary female deposits her savings with an MFI	Primary female's amount of savings (birr)
T1 x Poultry	0.472*** (0.051)	0.507*** (0.082)	-0.194*** (0.066)	-384.628** (164.217)
T1 x Poultry x Aspirations ^a	0.071	-0.068	-0.026	51.951

	(0.066)	(0.057)	(0.052)	(127.213)
T1 x Cash	0.474***	0.484***	-0.212***	-487.071***
	(0.086)	(0.093)	(0.071)	(162.667)
T1 x Cash x Aspirations ^a	0.019	-0.040	-0.057	212.463
	(0.110)	(0.066)	(0.056)	(195.552)
T2 x Poultry	0.458***	0.473***	-0.138*	-462.696***
	(0.058)	(0.088)	(0.077)	(154.226)
T2 x Poultry x Aspirations ^a	-0.031	-0.074	-0.098	93.154
	(0.075)	(0.083)	(0.069)	(129.969)
T2 x Cash	0.513***	0.448***	-0.243***	-469.548***
	(0.051)	(0.089)	(0.069)	(150.954)
T2 x Cash x Aspirations ^a	0.002	-0.044	0.016	129.670
	(0.066)	(0.078)	(0.062)	(123.270)
T3	0.349***	0.487***	-0.195***	-419.160***
	(0.041)	(0.079)	(0.064)	(155.130)
Linear combination: effect of T1	0.496***	0.468***	-0.224***	-370.244**
	(0.044)	(0.079)	(0.060)	(151.859)
Linear combination: effect of T2	0.478***	0.431***	-0.209***	-411.601***
	(0.040)	(0.082)	(0.063)	(145.164)
Linear combination: effect of poultry	0.466***	0.491***	-0.166**	-425.427***
	(0.044)	(0.081)	(0.065)	(148.610)
Linear combination: effect of cash	0.492***	0.465***	-0.227***	-477.142***
	(0.055)	(0.084)	(0.064)	(147.204)
Test: T1 = T2	0.647	0.302	0.600	0.564
Test: T2 = T3	0.000***	0.093*	0.664	0.924
Test: T1 = T3	0.000***	0.515	0.334	0.565
Mean of control (T4)	0.245	0.415	0.349	846.010
N	1,800	1,046	1,046	1,023

Notes: Estimates from the 'poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available). 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'.

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

Table 6.3c: Female's savings: Subsample of nonpoor households by SPIR sub treatment arms

	(1)	(2)	(3)	(4)
	Primary female has personal savings of her own	Primary female deposits her savings with VESA/VSLA	Primary female deposits her savings with an MFI	Primary female's amount of savings (birr)
T1	0.504*** (0.051)	0.671*** (0.082)	-0.278*** (0.081)	-642.346* (368.819)
T1 x Aspirations ^a	-0.049 (0.057)	-0.003 (0.045)	0.012 (0.049)	10.001 (152.901)
T2	0.382*** (0.056)	0.712*** (0.079)	-0.220** (0.089)	-531.186 (394.337)
T2 x Aspirations ^a	0.048 (0.064)	-0.078* (0.047)	-0.101* (0.055)	6.272 (175.311)
T3	0.379*** (0.048)	0.699*** (0.077)	-0.288*** (0.078)	-662.500* (384.172)
Linear combination: effect of T1	0.479*** (0.046)	0.669*** (0.079)	-0.272*** (0.077)	-637.346* (384.721)
Linear combination: effect of T2	0.406*** (0.048)	0.673*** (0.079)	-0.270*** (0.080)	-528.110 (377.879)
Test: T1 = T2	0.080*	0.908	0.956	0.349
Test: T2 = T3	0.542	0.364	0.647	0.196
Test: T1 = T3	0.018**	0.315	0.667	0.794
Mean of control (T4)	0.252	0.250	0.421	997.600
N	1,328	762	762	742

Notes: Estimates from the 'non-poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

**Table 6.4a: Male's credit and savings
by SPIR treatment arms**

	(1) Primary male received a loan for productive purposes in past 12 months	(2) Primary male received a productive loan from VESA	(3) Total value of productive loans received by male	(4) Primary male has a bank account	(5) Primary male deposited money with a RuSACCO in past 12 months
T1	0.051 (0.033)	0.123*** (0.040)	-415.218 (856.370)	-0.019 (0.020)	0.054 (0.061)
T2	0.038 (0.033)	0.079** (0.038)	-1,529.718* (879.680)	-0.002 (0.021)	0.034 (0.061)
T3	0.046 (0.032)	0.090** (0.044)	-864.845 (923.049)	-0.004 (0.019)	0.011 (0.068)
Test: T1 = T2	0.663	0.297	0.168	0.405	0.700
Test: T2 = T3	0.787	0.825	0.442	0.925	0.706
Test: T1 = T3	0.844	0.462	0.606	0.396	0.468
Mean of control (T4)	0.207	0.041	9,144.385	0.097	0.809
N	2,481	626	627	2,479	343

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: Primary male took out a loan for productive purposes in past 12 months, Total value of productive loans primary male took out, Primary male has a bank account, Primary male deposited money with a RuSACCO in past 12 months.

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

Table 6.4b: Male's credit and savings: Subsample of poor households by SPIR subtreatment arms

	(1) Primary male received a loan for productive purposes in past 12 months	(2) Primary male received a productive loan from VESA	(3) Total value of productive loans received by male	(4) Primary male has a bank account	(5) Primary male deposited money with a RuSACCO in past 12 months
T1 x Poultry	0.052 (0.042)	0.017 (0.069)	393.719 (1,603.457)	0.014 (0.026)	0.073 (0.132)
T1 x Poultry x Aspirations ^a	0.049 (0.054)	0.208** (0.092)	-429.712 (1,487.903)	-0.024 (0.036)	-0.073 (0.184)
T1 x Cash	0.058 (0.069)	0.103 (0.074)	-2,169.350* (1,213.195)	-0.047 (0.031)	0.011 (0.162)
T1 x Cash x Aspirations ^a	-0.029 (0.093)	0.002 (0.103)	1,832.286 (1,977.636)	0.059 (0.056)	-0.040 (0.219)
T2 x Poultry	0.039 (0.069)	0.168* (0.087)	-2,957.874*** (1,125.050)	0.070 (0.046)	0.186*** (0.065)
T2 x Poultry x Aspirations ^a	-0.074 (0.076)	-0.230** (0.097)	2,048.341 (1,594.264)	-0.026 (0.065)	-0.199* (0.119)
T2 x Cash	-0.020 (0.049)	-0.004 (0.054)	-1,283.885 (2,146.811)	-0.025 (0.030)	-0.055 (0.112)
T2 x Cash x Aspirations ^a	0.115 (0.073)	0.169* (0.090)	-2,393.159 (2,268.375)	0.015 (0.041)	0.160 (0.125)
T3	0.067* (0.038)	0.064 (0.058)	-1,179.251 (1,148.975)	0.015 (0.024)	0.136 (0.086)
Linear combination: effect of T1	0.060 (0.038)	0.115** (0.050)	-558.866 (1,023.496)	-0.009 (0.023)	0.013 (0.093)
Linear combination: effect of T2	0.020 (0.038)	0.069 (0.052)	-2,220.494** (1,106.468)	0.021 (0.026)	0.058 (0.077)
Linear combination: effect of poultry	0.045 (0.045)	0.094 (0.062)	-1,311.721 (1,165.253)	0.042 (0.028)	0.131 (0.081)
Linear combination: effect of cash	0.020 (0.047)	0.051 (0.051)	-1,728.817 (1,349.927)	-0.036 (0.025)	-0.021 (0.108)
Test: T1 = T2	0.273	0.364	0.068*	0.253	0.531
Test: T2 = T3	0.182	0.929	0.309	0.816	0.268
Test: T1 = T3	0.854	0.350	0.531	0.298	0.145

Mean of control (T4)	0.197	0.066	8,790.082	0.078	0.822
<i>N</i>	1,296	313	313	1,295	173

Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: Primary male took out a loan for productive purposes in past 12 months, Total value of productive loans primary male took out, Primary male has a bank account, Primary male deposited money with a RuSACCO in past 12 months. * p<0.1 ** p<0.05 *** p<0.01

Table 6.4c: Male's credit and savings: Subsample of nonpoor households by SPIR sub-treatment arms

	(1) Primary male received a loan for productive purposes in past 12 months	(2) Primary male received a productive loan from VESA	(3) Total value of productive loans received by male	(4) Primary male has a bank account	(5) Primary male deposited money with a RuSACCO in past 12 months
T1	-0.003 (0.048)	0.151* (0.080)	-808.345 (1,141.264)	-0.050* (0.030)	0.046 (0.096)
T1 x Aspirations ^a	0.084 (0.056)	-0.050 (0.091)	1,225.984 (1,497.958)	0.041 (0.038)	0.019 (0.094)
T2	0.078* (0.045)	0.075* (0.042)	-1,101.650 (1,188.066)	-0.048 (0.032)	-0.100 (0.143)
T2 x Aspirations ^a	-0.051 (0.059)	-0.026 (0.053)	807.699 (1,679.376)	0.044 (0.037)	0.172 (0.150)
T3	0.022 (0.039)	0.111** (0.055)	-455.859 (1,075.912)	-0.026 (0.028)	-0.112 (0.105)
Linear combination: effect of T1	0.039 (0.040)	0.126** (0.054)	-195.353 (1,103.069)	-0.029 (0.030)	0.055 (0.083)
Linear combination: effect of T2	0.053 (0.039)	0.063* (0.036)	-705.504 (1,214.461)	-0.027 (0.030)	-0.016 (0.099)
Test: T1 = T2	0.738	0.242	0.679	0.933	0.409
Test: T2 = T3	0.410	0.358	0.826	0.966	0.375
Test: T1 = T3	0.649	0.803	0.806	0.893	0.077*
Mean of control (T4)	0.218	0.017	9,498.689	0.118	0.795
N	1,185	313	314	1,184	170

Notes: Estimates from the 'non-poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: Primary male took out a loan for productive purposes in past 12 months, Total value of productive loans primary male took out, Primary male has a bank account, Primary male deposited money with a RuSACCO in past 12 months. * p<0.1 ** p<0.05 *** p<0.01

**Table 6.5a: Household members' wage work, labor outside the household
by SPIR treatment arms**

	(1) A household member undertook regular wage work for an employer in last 12 months	(2) A household member undertook casual/irregular wage work in last 12 months	(3) Primary male's main activity is crop production	(4) Primary female undertook regular wage work for an employer in last 12 months	(5) Primary female undertook casual/irregular wage work in last 12 months	(6) Primary female's main activity is crop production
T1	0.012 (0.013)	-0.011 (0.039)	0.002 (0.017)	0.004 (0.005)	-0.001 (0.022)	-0.014 (0.039)
T2	0.001 (0.012)	0.006 (0.036)	0.002 (0.014)	0.003 (0.004)	-0.019 (0.021)	-0.038 (0.031)
T3	0.004 (0.011)	0.043 (0.036)	0.007 (0.016)	-0.003 (0.004)	0.003 (0.021)	-0.056* (0.033)
Test: T1 = T2	0.382	0.619	0.989	0.795	0.354	0.499
Test: T2 = T3	0.846	0.211	0.698	0.107	0.208	0.557
Test: T1 = T3	0.493	0.109	0.742	0.132	0.844	0.270
Mean of control (T4)	0.034	0.273	0.935	0.007	0.100	0.317
N	2,479	2,477	2,671	3,127	3,128	3,078

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: A household member undertook regular wage work for an employer in last 12 months, A household member undertook casual/irregular wage work in last 12 months, Primary male's main activity is crop production, Primary female's main activity is crop production.

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

**Table 6.5b: Household members' wage work, labor outside the household: Subsample of poor households
by SPIR subtreatment arms**

	(1) A household member undertook regular wage work for an	(2) A household member undertook casual/irregular	(3) Primary male's main activity is crop production	(4) Primary female undertook regular wage work for an	(5) Primary female undertook casual/irregular wage work in last 12 months	(6) Primary female's main activity is crop production
--	--	---	--	--	--	--

	employer in last 12 months	wage work in last 12 months		employer in last 12 months		
T1 x Poultry	-0.013 (0.022)	-0.138** (0.060)	-0.012 (0.036)	0.010 (0.016)	-0.091*** (0.032)	0.022 (0.072)
T1 x Poultry x Aspirations ^a	0.058** (0.029)	0.143** (0.065)	0.046 (0.039)	-0.019 (0.016)	0.102*** (0.039)	-0.063 (0.082)
T1 x Cash	-0.001 (0.029)	-0.116 (0.080)	0.023 (0.041)	0.011 (0.015)	-0.017 (0.035)	0.048 (0.087)
T1 x Cash x Aspirations ^a	0.041 (0.040)	0.177* (0.094)	-0.061 (0.059)	0.007 (0.022)	0.100* (0.056)	-0.149 (0.094)
T2 x Poultry	0.005 (0.024)	0.152** (0.065)	0.005 (0.035)	-0.001 (0.008)	0.043 (0.047)	-0.074* (0.042)
T2 x Poultry x Aspirations ^a	-0.029 (0.023)	-0.259*** (0.079)	0.010 (0.038)	0.013 (0.014)	-0.121** (0.053)	0.124* (0.065)
T2 x Cash	-0.012 (0.025)	0.009 (0.044)	-0.003 (0.027)	-0.009 (0.006)	-0.034 (0.031)	-0.079* (0.046)
T2 x Cash x Aspirations ^a	0.038 (0.035)	-0.001 (0.056)	-0.002 (0.034)	0.030** (0.012)	0.003 (0.037)	0.069 (0.060)
T3	0.008 (0.018)	0.038 (0.046)	0.007 (0.026)	-0.001 (0.006)	-0.026 (0.025)	-0.049 (0.036)
Linear combination: effect of T1	0.018 (0.017)	-0.046 (0.044)	0.002 (0.025)	0.008 (0.007)	-0.002 (0.026)	-0.018 (0.041)
Linear combination: effect of T2	-0.001 (0.016)	0.019 (0.041)	0.003 (0.022)	0.006 (0.006)	-0.023 (0.027)	-0.030 (0.034)
Linear combination: effect of poultry	-0.004 (0.018)	0.009 (0.051)	-0.004 (0.029)	0.004 (0.009)	-0.023 (0.032)	-0.027 (0.046)
Linear combination: effect of cash	-0.006 (0.021)	-0.054 (0.051)	0.010 (0.028)	0.001 (0.009)	-0.025 (0.028)	-0.014 (0.053)
Test: T1 = T2	0.221	0.078*	0.975	0.790	0.362	0.768
Test: T2 = T3	0.598	0.629	0.840	0.258	0.866	0.535
Test: T1 = T3	0.584	0.050*	0.834	0.233	0.250	0.425
Mean of control (T4)	0.036	0.304	0.912	0.007	0.123	0.301
N	1,295	1,294	1,408	1,799	1,800	1,764

Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: A household member

undertook regular wage work for an employer in last 12 months, A household member undertook casual/irregular wage work in last 12 months, Primary male's main activity is crop production, Primary female's main activity is crop production. * p<0.1 ** p<0.05 *** p<0.01

Table 6.5c: Household members' wage work/out of farm labor: Subsample of nonpoor households by SPIR sub treatment arms

	(1) A household member undertook regular wage work for an employer in last 12 months	(2) A household member undertook casual/irregular wage work in last 12 months	(3) Primary male's main activity is crop production	(4) Primary female undertook regular wage work for an employer in last 12 months	(5) Primary female undertook casual/irregular wage work in last 12 months	(6) Primary female's main activity is crop production
T1	-0.009 (0.019)	0.059 (0.060)	0.012 (0.023)	0.000 (0.007)	-0.020 (0.027)	-0.008 (0.065)
T1 x Aspirations ^a	0.039 (0.027)	-0.033 (0.061)	-0.036 (0.026)	-0.001 (0.009)	0.057* (0.033)	-0.019 (0.064)
T2	0.004 (0.021)	0.078 (0.055)	-0.010 (0.023)	-0.001 (0.007)	0.009 (0.020)	-0.093* (0.047)
T2 x Aspirations ^a	0.002 (0.024)	-0.150*** (0.058)	0.013 (0.026)	0.001 (0.008)	-0.042* (0.024)	0.089 (0.055)
T3	0.002 (0.015)	0.056 (0.042)	0.006 (0.014)	-0.006 (0.004)	0.047* (0.024)	-0.066 (0.047)
Linear combination: effect of T1	0.011 (0.016)	0.042 (0.046)	-0.006 (0.018)	-0.000 (0.006)	0.008 (0.023)	-0.017 (0.048)
Linear combination: effect of T2	0.006 (0.016)	0.005 (0.045)	-0.004 (0.017)	-0.001 (0.006)	-0.012 (0.019)	-0.049 (0.044)
Test: T1 = T2	0.760	0.384	0.880	0.965	0.324	0.459
Test: T2 = T3	0.830	0.182	0.541	0.161	0.008***	0.674
Test: T1 = T3	0.595	0.736	0.452	0.188	0.120	0.289
Mean of control (T4)	0.032	0.239	0.960	0.007	0.066	0.339
N	1,184	1,183	1,263	1,328	1,328	1,314

Notes: Estimates from the 'non-poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal

impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available) except for outcomes: A household member undertook regular wage work for an employer in last 12 months, A household member undertook casual/irregular wage work in last 12 months, Primary male's main activity is crop production, Primary female's main activity is crop production. * p<0.1 ** p<0.05 *** p<0.01

**Table 6.6a: Expenditure on crop inputs
by SPIR treatment arms**

	(1)	(2)	(3)	(4)
	Total amount paid for seed in last Mehr season (birr)	Total amount paid for inorganic fertilizer in last Mehr season (birr)	Total amount paid for pesticides in last Mehr season (birr)	Total amount paid for hired agricultural labor in last Mehr season (birr)
T1	4.397 (34.376)	11.239 (41.275)	-4.078 (8.198)	214.661 (1,201.911)
T2	42.364 (31.429)	43.047 (44.325)	-3.099 (7.787)	-252.524 (964.865)
T3	26.361 (33.382)	-50.046 (36.110)	-5.755 (8.201)	-161.564 (919.451)
Test: T1 = T2	0.268	0.487	0.898	0.656
Test: T2 = T3	0.625	0.023**	0.719	0.894
Test: T1 = T3	0.540	0.103	0.833	0.709
Mean of control (T4)	200.050	286.683	38.771	1,608.739
N	3,189	3,189	3,189	3,189

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

**Table 6.6b: Expenditure on crop inputs: Subsample of poor households
by SPIR sub treatment arms**

	(1)	(2)	(3)	(4)
	Total amount paid for seed in last Mehr season (birr)	Total amount paid for inorganic fertilizer in last Mehr season (birr)	Total amount paid for pesticides in last Mehr season (birr)	Total amount paid for hired agricultural labor in last Mehr season (birr)
T1 x Poultry	76.484	39.474	-4.329	4,570.598

	(55.398)	(55.067)	(9.523)	(5,147.476)
T1 x Poultry x Aspirations ^a	-80.531	-24.970	-5.187	-6,526.422
	(57.179)	(63.965)	(11.611)	(5,191.722)
T1 x Cash	-33.928	-19.543	4.928	-971.181
	(37.449)	(86.775)	(13.545)	(1,518.009)
T1 x Cash x Aspirations ^a	-43.885	-71.971	-26.434*	3,103.573
	(39.359)	(117.439)	(14.169)	(3,109.185)
T2 x Poultry	-22.670	-37.366	0.799	-622.617
	(41.866)	(52.026)	(10.280)	(1,292.664)
T2 x Poultry x Aspirations ^a	78.073	-1.865	-0.307	2,585.306
	(64.350)	(58.588)	(9.795)	(3,182.175)
T2 x Cash	76.779	63.157	-7.636	-909.511
	(57.284)	(73.036)	(9.221)	(1,137.024)
T2 x Cash x Aspirations ^a	-32.040	-13.424	15.458	-124.217
	(63.782)	(117.578)	(13.781)	(1,199.572)
T3	13.324	-72.491*	-3.523	-1,149.650
	(31.779)	(42.788)	(10.428)	(1,173.029)
Linear combination: effect of T1	-11.094	-14.760	-7.557	855.972
	(30.154)	(48.476)	(7.286)	(1,757.410)
Linear combination: effect of T2	37.394	8.238	0.362	-161.946
	(33.168)	(48.646)	(7.064)	(1,372.628)
Linear combination: effect of poultry	26.207	0.457	-1.730	1,938.860
	(38.715)	(45.424)	(8.166)	(2,662.940)
Linear combination: effect of cash	20.504	21.102	-1.252	-938.263
	(37.634)	(61.752)	(9.047)	(1,219.376)
Test: T1 = T2	0.079*	0.608	0.191	0.530
Test: T2 = T3	0.400	0.029**	0.674	0.247
Test: T1 = T3	0.339	0.135	0.680	0.162
Mean of control (T4)	176.144	265.206	30.776	1,725.923
N	1,825	1,825	1,825	1,825

Notes: Estimates from the 'poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available). 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'.

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

Table 6.6c: Expenditure on crop inputs: Subsample of nonpoor households by SPIR sub treatment arms

	(1)	(2)	(3)	(4)
	Total amount paid for seed in last Mehr season (birr)	Total amount paid for inorganic fertilizer in last Mehr season (birr)	Total amount paid for pesticides in last Mehr season (birr)	Total amount paid for hired agricultural labor in last Mehr season (birr)
T1	35.217 (86.019)	92.921 (57.371)	20.792 (19.443)	-229.402 (726.713)
T1 x Aspirations ^a	-28.320 (104.060)	-98.295 (70.111)	-43.874** (20.098)	-779.164 (756.032)
T2	56.125 (56.595)	111.119 (68.972)	-21.847* (11.757)	-694.710 (510.073)
T2 x Aspirations ^a	-18.110 (68.396)	-43.023 (110.179)	29.504 (19.773)	476.715 (828.395)
T3	41.547 (59.499)	-17.981 (44.344)	-8.611 (11.856)	1,104.000 (1,075.807)
Linear combination: effect of T1	21.057 (58.891)	43.773 (44.173)	-1.145 (13.299)	-618.983 (552.085)
Linear combination: effect of T2	47.243 (45.821)	90.017 (59.899)	-7.377 (12.918)	-460.899 (600.560)
Test: T1 = T2	0.655	0.460	0.644	0.768
Test: T2 = T3	0.923	0.080*	0.915	0.136
Test: T1 = T3	0.769	0.190	0.543	0.097*
Mean of control (T4)	233.158	316.426	49.842	1,446.456
N	1,364	1,364	1,364	1,364

Notes: Estimates from the ‘non-poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available). * p<0.1 ** p<0.05 *** p<0.01

**Table 6.7a: Female's aspirations
by SPIR treatment arms**

	(1)	(2)	(3)	(4)	(5)	(6)
	Aspired level of child's education: no education	Aspired level of child's education: 1st-8th grade	Aspired level of child's education: 9th-12th grade	Aspired level of child's education: college / university	Aspired level of income ('000 birr)	Expected level of income in 10 years ('000 birr)
T1	0.003 (0.008)	-0.013 (0.008)	0.020 (0.037)	-0.003 (0.037)	13.664* (6.930)	10.328 (9.765)
T2	0.002 (0.007)	-0.000 (0.010)	-0.006 (0.036)	0.012 (0.035)	-0.020 (5.305)	-0.776 (8.676)
T3	-0.007 (0.007)	-0.012 (0.008)	0.022 (0.035)	0.000 (0.034)	7.680 (6.140)	11.174 (10.942)
Test: T1 = T2	0.779	0.151	0.486	0.699	0.046**	0.287
Test: T2 = T3	0.147	0.202	0.433	0.743	0.208	0.294
Test: T1 = T3	0.102	0.952	0.964	0.928	0.416	0.945
Mean of control (T4)	0.016	0.038	0.486	0.449	36.274	55.031
N	2,911	2,911	2,911	2,911	2,931	2,730

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

**Table 6.7b: Female's aspirations: Subsample of poor households
by SPIR sub treatment arms**

	(1)	(2)	(3)	(4)	(5)	(6)
	Aspired level of child's education: no education	Aspired level of child's education: 1st-8th grade	Aspired level of child's education: 9th-12th grade	Aspired level of child's education: college / university	Aspired level of income ('000 birr)	Expected level of income in 10 years ('000 birr)
T1 x Poultry	-0.007 (0.007)	-0.019 (0.016)	0.037 (0.073)	-0.012 (0.063)	10.000 (11.174)	-12.251 (19.867)
T1 x Poultry x Aspirations ^a	0.004 (0.011)	0.017 (0.019)	-0.004 (0.081)	-0.004 (0.077)	-0.619 (14.005)	23.059 (32.788)

T1 x Cash	0.024 (0.017)	-0.020 (0.014)	0.108* (0.060)	-0.096 (0.060)	12.186 (13.222)	-3.531 (10.839)
T1 x Cash x Aspirations ^a	-0.027 (0.019)	0.011 (0.017)	-0.125 (0.090)	0.136 (0.094)	-10.601 (13.259)	-7.039 (15.586)
T2 x Poultry	0.004 (0.012)	0.021 (0.025)	-0.063 (0.063)	0.050 (0.061)	7.435 (6.357)	25.953 (20.441)
T2 x Poultry x Aspirations ^a	-0.001 (0.012)	-0.040 (0.029)	0.097 (0.077)	-0.056 (0.076)	1.858 (12.682)	-2.527 (24.195)
T2 x Cash	0.006 (0.019)	0.001 (0.018)	-0.020 (0.077)	0.026 (0.079)	0.659 (6.179)	-20.909 (13.648)
T2 x Cash x Aspirations ^a	0.004 (0.021)	0.041 (0.028)	0.039 (0.094)	-0.092 (0.101)	-12.000 (7.879)	13.955 (13.689)
T3	0.003 (0.007)	-0.007 (0.011)	0.033 (0.043)	-0.024 (0.042)	7.440 (7.995)	8.441 (10.570)
Linear combination: effect of T1	0.003 (0.008)	-0.012 (0.009)	0.041 (0.043)	-0.022 (0.042)	8.309 (6.287)	-3.649 (10.392)
Linear combination: effect of T2	0.005 (0.007)	0.011 (0.013)	-0.009 (0.043)	0.002 (0.044)	1.631 (5.512)	5.745 (9.964)
Linear combination: effect of poultry	-0.001 (0.008)	0.001 (0.016)	-0.014 (0.053)	0.020 (0.049)	8.721 (7.096)	7.168 (14.872)
Linear combination: effect of cash	0.015 (0.013)	-0.010 (0.013)	0.045 (0.053)	-0.036 (0.053)	6.495 (7.820)	-12.051 (9.705)
Test: T1 = T2	0.801	0.052*	0.236	0.573	0.261	0.383
Test: T2 = T3	0.760	0.167	0.302	0.537	0.450	0.799
Test: T1 = T3	0.976	0.602	0.858	0.971	0.913	0.291
Mean of control (T4)	0.010	0.035	0.476	0.466	34.639	52.052
N	1,680	1,680	1,680	1,680	1,685	1,570

Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

Table 6.7c: Female's aspirations: Subsample of nonpoor households by SPIR subtreatment arms

	(1)	(2)	(3)	(4)	(5)	(6)
	Aspired level of child's education: no education	Aspired level of child's education: 1st-8th grade	Aspired level of child's education: 9th-12th grade	Aspired level of child's education: college / university	Aspired level of income (‘000 birr)	Expected level of income in 10 years (‘000 birr)
T1	0.022 (0.019)	-0.013 (0.015)	0.002 (0.057)	-0.004 (0.059)	24.029* (14.121)	52.516 (33.209)
T1 x Aspirations ^a	-0.040** (0.018)	-0.000 (0.016)	-0.028 (0.071)	0.066 (0.070)	-7.618 (14.652)	-54.335 (35.911)
T2	-0.003 (0.014)	-0.013 (0.014)	-0.022 (0.048)	0.045 (0.046)	-5.766 (6.571)	-32.034* (18.303)
T2 x Aspirations ^a	-0.000 (0.013)	-0.005 (0.016)	0.042 (0.061)	-0.043 (0.061)	8.029 (12.307)	38.389* (19.840)
T3	-0.019* (0.012)	-0.019 (0.014)	0.005 (0.047)	0.034 (0.045)	8.148 (7.279)	14.815 (21.557)
Linear combination: effect of T1	0.002 (0.014)	-0.013 (0.013)	-0.012 (0.048)	0.029 (0.048)	20.220** (9.657)	25.348 (19.486)
Linear combination: effect of T2	-0.003 (0.012)	-0.016 (0.013)	-0.002 (0.044)	0.024 (0.043)	-1.827 (8.374)	-13.205 (14.727)
Test: T1 = T2	0.676	0.808	0.822	0.907	0.027**	0.037**
Test: T2 = T3	0.074*	0.780	0.887	0.818	0.206	0.172
Test: T1 = T3	0.056*	0.616	0.727	0.920	0.179	0.658
Mean of control (T4)	0.025	0.043	0.500	0.425	38.655	59.332
N	1,231	1,231	1,231	1,231	1,246	1,160

Notes: Estimates from the ‘non-poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available). * p<0.1 ** p<0.05 *** p<0.01

**Table 6.8a: Male's aspirations
by SPIR treatment arms**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Aspired level of child's education: no education	Aspired level of child's education: 1st-8th grade	Aspired level of child's education: 9th-12th grade	Aspired level of child's education: college / university	Aspired level of income (‘000 birr)	Expected level of income in 10 years (‘000 birr)	Aspired level of assets (‘000 birr)	Expected level of assets in 10 years (‘000 birr)
T1	-0.006 (0.009)	-0.006 (0.008)	-0.008 (0.038)	0.017 (0.040)	20.167* (11.512)	-1.159 (9.991)	28.083 (27.117)	28.704 (24.159)
T2	-0.007 (0.009)	-0.004 (0.008)	-0.066* (0.035)	0.076** (0.036)	4.148 (11.038)	2.052 (11.471)	16.103 (28.480)	28.468 (25.720)
T3	-0.011 (0.008)	-0.012 (0.008)	-0.050 (0.037)	0.071* (0.039)	4.482 (9.758)	7.606 (11.675)	13.551 (26.124)	5.978 (20.325)
Test: T1 = T2	0.895	0.802	0.117	0.123	0.153	0.774	0.666	0.993
Test: T2 = T3	0.626	0.323	0.656	0.903	0.972	0.655	0.923	0.359
Test: T1 = T3	0.525	0.430	0.272	0.176	0.117	0.450	0.561	0.318
Mean of control (T4)	0.023	0.025	0.422	0.530	57.804	78.966	170.437	157.586
N	2,347	2,347	2,347	2,347	2,425	2,322	2,396	2,288

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

**Table 6.8b: Male's aspirations: Subsample of poor households
by SPIR sub-treatment arms**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Aspired level of child's education: no education	Aspired level of child's education: 1st-8th grade	Aspired level of child's education: 9th-12th grade	Aspired level of child's education: college / university	Aspired level of income (‘000 birr)	Expected level of income in 10 years (‘000 birr)	Aspired level of assets (‘000 birr)	Expected level of assets in 10 years (‘000 birr)
T1 x Poultry	-0.005 (0.008)	0.001 (0.014)	0.001 (0.058)	0.003 (0.061)	39.130 (27.697)	19.232 (24.380)	63.196 (86.578)	-23.731 (42.937)

T1 x Poultry x Aspirations ^a	0.013 (0.024)	-0.007 (0.013)	0.116 (0.073)	-0.122 (0.081)	-28.310 (29.933)	-19.388 (36.063)	-58.907 (88.403)	46.148 (69.832)
T1 x Cash	0.012 (0.015)	-0.014 (0.015)	0.027 (0.100)	-0.025 (0.102)	-1.068 (28.889)	-0.864 (18.365)	64.066 (67.560)	92.739 (56.399)
T1 x Cash x Aspirations ^a	-0.005 (0.021)	0.016 (0.025)	-0.042 (0.111)	0.031 (0.109)	39.808 (32.308)	-20.041 (25.965)	-73.207 (67.008)	-136.670** (66.772)
T2 x Poultry	0.008 (0.015)	0.015 (0.019)	-0.066 (0.058)	0.044 (0.057)	-5.124 (16.854)	6.281 (18.239)	77.587 (58.825)	35.682 (49.736)
T2 x Poultry x Aspirations ^a	0.006 (0.018)	-0.023 (0.019)	-0.038 (0.059)	0.055 (0.059)	25.443 (28.285)	39.935 (27.415)	-18.177 (81.283)	131.203 (87.834)
T2 x Cash	-0.016* (0.009)	0.021 (0.016)	-0.094** (0.042)	0.089** (0.043)	3.624 (15.585)	-38.806** (15.596)	-14.639 (46.594)	-59.329 (41.140)
T2 x Cash x Aspirations ^a	0.025 (0.023)	-0.029 (0.021)	0.111* (0.058)	-0.106 (0.068)	-16.913 (23.417)	29.075 (18.556)	3.189 (52.937)	69.897 (44.367)
T3	0.003 (0.009)	-0.006 (0.010)	-0.049 (0.042)	0.052 (0.043)	0.936 (15.002)	1.525 (13.235)	55.257 (42.196)	7.033 (29.973)
Linear combination: effect of T1	0.006 (0.010)	-0.004 (0.010)	0.033 (0.045)	-0.035 (0.047)	21.392 (16.091)	-0.927 (12.683)	30.244 (40.687)	13.116 (30.418)
Linear combination: effect of T2	0.004 (0.009)	0.005 (0.010)	-0.062* (0.037)	0.053 (0.039)	1.255 (15.136)	1.016 (11.770)	28.656 (38.460)	38.196 (32.617)
Linear combination: effect of poultry	0.002 (0.010)	0.008 (0.013)	-0.033 (0.046)	0.024 (0.047)	16.705 (18.185)	12.690 (16.173)	70.696 (56.191)	6.455 (35.983)
Linear combination: effect of cash	-0.002 (0.010)	0.003 (0.012)	-0.032 (0.059)	0.031 (0.060)	1.238 (18.639)	-19.485 (13.140)	25.260 (46.542)	17.845 (37.724)
Test: T1 = T2	0.864	0.356	0.014**	0.029**	0.156	0.875	0.965	0.437
Test: T2 = T3	0.934	0.316	0.709	0.973	0.980	0.969	0.476	0.333
Test: T1 = T3	0.795	0.864	0.059*	0.052*	0.139	0.863	0.537	0.841
Mean of control (T4)	0.010	0.017	0.408	0.564	67.676	77.482	165.789	155.003
N	1,228	1,228	1,228	1,228	1,266	1,212	1,250	1,190

Notes: Estimates from the 'poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

Table 6.8c: Male's aspirations: Subsample of nonpoor households by SPIR subtreatment arms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Aspired level of child's education: no education	Aspired level of child's education: 1st-8th grade	Aspired level of child's education: 9th-12th grade	Aspired level of child's education: college / university	Aspired level of income (‘000 birr)	Expected level of income in 10 years (‘000 birr)	Aspired level of assets (‘000 birr)	Expected level of assets in 10 years (‘000 birr)
T1	-0.009 (0.015)	0.006 (0.020)	-0.059 (0.053)	0.054 (0.054)	12.583 (10.850)	-16.710 (14.944)	14.539 (27.137)	59.418 (44.003)
T1 x Aspirations ^a	-0.018 (0.014)	-0.026 (0.019)	-0.007 (0.069)	0.048 (0.072)	12.260 (13.024)	23.957 (21.210)	8.237 (35.526)	-42.578 (55.568)
T2	-0.019 (0.019)	-0.005 (0.017)	-0.108* (0.057)	0.123** (0.059)	0.684 (7.449)	-21.762 (14.489)	-20.342 (27.702)	-25.219 (24.057)
T2 x Aspirations ^a	0.000 (0.018)	-0.013 (0.017)	0.062 (0.074)	-0.041 (0.074)	14.305 (14.193)	43.693** (21.082)	44.739 (39.230)	71.008** (32.363)
T3	-0.023* (0.013)	-0.018 (0.013)	-0.054 (0.047)	0.091* (0.047)	7.734 (7.885)	13.013 (19.088)	-32.071 (24.444)	5.810 (24.226)
Linear combination: effect of T1	-0.018 (0.014)	-0.007 (0.015)	-0.062 (0.047)	0.079 (0.048)	18.713** (9.145)	-4.731 (14.490)	18.658 (27.082)	38.129 (33.873)
Linear combination: effect of T2	-0.019 (0.015)	-0.012 (0.014)	-0.077 (0.048)	0.103** (0.048)	7.700 (9.349)	-0.332 (15.268)	1.601 (28.059)	9.608 (23.855)
Test: T1 = T2	0.953	0.730	0.767	0.625	0.273	0.747	0.513	0.356
Test: T2 = T3	0.668	0.554	0.633	0.802	0.997	0.460	0.152	0.862
Test: T1 = T3	0.584	0.369	0.854	0.806	0.208	0.326	0.025**	0.298
Mean of control (T4)	0.037	0.034	0.437	0.493	46.835	80.563	175.481	160.361
N	1,119	1,119	1,119	1,119	1,159	1,110	1,146	1,098

Notes: Estimates from the ‘non-poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. Models do not control for the baseline value of the outcome (no baseline data available).

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

7 Impacts on Nutrition and Child Welfare Outcomes

7.1 Introduction

Ethiopia has made major strides in improving nutrition in the past two decades; the prevalence of stunting decreased considerably, from 58 percent in 2000 to 38 percent in 2016 and further to 37 percent in 2018 (EPHI and ICF 2019). While there is no consensus on the underlying causes of this improvement—although substantial increases in income and education surely contributed—there is consensus that more must be done to maintain the momentum. In particular, a number of infant and young child feeding (IYCF) practices can be improved. While breastfeeding initiation is nearly universal, many children are not exclusively breastfed until they are 6 months old, in accord with the recommended practice. Similarly, few children aged 6-23 months meet the minimum acceptable dietary standards. This chapter will discuss outcomes for a few key IYCF practices in the SPIR project communities.

7.2 Anthropometrics

Stunting rates in the SPIR project area are roughly similar to those observed in the Amhara and Oromia regions in 2016 DHS survey (DHS 2016). Stunting²² (low height-for-age) measures chronic undernutrition, which is cumulative; in low-income settings, children often fall behind international growth norms in the period up to 24 months of age. This is the case for the samples in both Amhara and Oromia as indicated in Figure 7.1. Despite being drawn from a rural, poorer PSNP sample, the average height-for-age in the study areas in the baseline is fairly close to the DHS height data in their respective regions at all ages. The pattern of decline in the first 18 months is also consistent in both regions and in the SPIR sample as well as the DHS.

Stunting rates for children up to 60 months in the midline were 48.3 for Amhara and 36.2 for Oromia, compared with DHS results of 46.3 and 36.5, respectively. The baseline survey collected anthropometric data on children less than 36 months; given the age patterns in stature indicated in Figure 7.1, a direct comparison of the averages in that sample with the means from the DHS that cover children up to 60 months would be misleading. However, regression analysis indicates that, controlling for age, gender and parental education, stunting rates in the midline were 4.4 percentage points higher in both regions than at baseline. It is not clear what lies behind this trend, whether it is due to worsening food security, feeding practices, or sanitation and hygiene. Drilling down a bit, this increase in stunting is only statistically significant in either region for children older than 24 months. Indeed, rates of stunting are lower in the midline than the baseline for children age 0-11 months in Oromia. The regression analysis also confirms that—as in most of Africa—boys are significantly more likely to be stunted than are girls in both regions.

Table 7.1a reports the differences in mean stature (height-for-age) and prevalence of stunting, mean weight-for-height and prevalence of wasting (acute undernutrition), and mean mid-upper arm circumference (MUAC) and prevalence of severe acute malnutrition (MUAC<11.5 cm) for children age 0 to 35 months between the treatment arms, controlling for age and gender. Table 7.1b reports the same, but only for children age 6-23 months. There are no differences in stunting prevalence across treatment arms for children age 0-35 months or for the younger subsample. However, there is a surprising—and currently unexplained—worsening in acute nutritional status in T1 and T3 compared to T4 (control) in the subset of children 6-23 months. For children in this age range, the mean WHZ is 0.200 and 0.214 standard deviations lower in T1 and T3, respectively; this results in a significant

²² Stunting is a measure of growth faltering in children that indicates that the child's recumbent length is two standard deviations below an international norm for their age and gender. The standardized measure of the child's length is height-for-age z-score (HAZ); a child is stunted when HAZ<-2.

increase in wasting for the T3 arm.²³ This pattern of effects is surprising because T1 and T3 are the N* treatment arms, which include the most intensive nutrition interventions. The size of these negative effects is moderate, but they deserve attention. Global experience with MUAC indicates that screening for acute malnutrition using this measure does not always coincide with results using WHZ. MUAC results from the midline provide partial support to the concern with T3 in that the average MUAC declines, albeit not significantly so. However, MUAC measures in column (5) are slightly larger for T1 and T2 compared to the control despite the negative signs for WHZ in column (3). These results suggest a need for the SPIR project to determine how information from different measures will be utilized as well as a need to strengthen the implementation of the N* interventions. It is possible that the implementation of the N* interventions was still scaling up at the time of the midline survey and that, with further attention to implementation, this pattern of effects could be reversed by the endline.

7.3 Infant and young child feeding (IYCF) practices

7.3.1 IYCF practices

IYCF includes both recommended practices for breastfeeding and feeding practices for children 6-23 months who are older than the age for which exclusive breastfeeding is recommended. Between the ages of 6 and 8 months, solid foods should be introduced. Table 7.2 indicates key IYCF practices for breastfeeding and weaning. Most mothers put their child to the breast shortly after delivery, hence over 82.8 percent of control group mothers in the midline indicate that their child received colostrum (column (1)). The table shows that T1 and T3, the two N* arms, had a significant impact on the probability of offering colostrum to the child. Going a step further, column (2) shows significant impacts of T1 and T3 on whether the mother initiates breastfeeding within an hour of birth in accord with the current recommendation. The tables also report two approaches to assessing exclusive breastfeeding. In column (3) exclusive breastfeeding is indicated if a mother reported breastfeeding their child and also did not indicate any fluids or solid foods in the section on recall of the child's diet on the previous day. This is defined only over the set of children 0-5 months (inclusive) and—unlike other outcomes in this table—the response is age dependent. Thus, age is included in the regressions. As indicated, the older the child, the lower the probability of exclusive breastfeeding. By this measure, women in T1 have a weakly significantly higher probability of exclusively breastfeeding children in this age group. Moreover, the average effect in T1 and T3 is to significantly increase the probability of exclusive breastfeeding by 11 percentage points. This raises the average from 67 to 78 percent.

Columns (4) and (5) are based on the age that the caregiver reports she stopped exclusive breastfeeding. This is recorded only for the women who reported that the child was no longer exclusively breastfed as of the date of the survey and covers children 6-15 months. This period was chosen since it covers the period for which exclusive breastfeeding should be augmented with solid foods. It also avoids overlaps between survey rounds, which were 15 months apart. Virtually all women reported that they waited at least 6 months to cease exclusive breastfeeding. This is at odds with the behavior used to define exclusive breastfeeding in column (3) and may reflect social desirability bias. Conversely, or additionally, it may reflect a confusion as to what constitutes exclusive breastfeeding; that is, respondents may not be clear as to whether or not it covers the inclusion of other liquids, including water. This is an issue that might be considered when designing BCC messages. Both columns indicate a slight tendency to cease exclusive breastfeeding earlier for boys. There is no indication in columns (4) and (5) nor in columns (6) and (7) that the interventions influenced breastfeeding at the time these caregivers were considering weaning of these children, who are somewhat older than those who form the basis of the results in column (3).

²³ We note that at baseline, there was a weakly statistically significant difference in WHZ, with mean WHZ 0.249 standard deviations lower in T3 than in T4 (control) for children at 0-35 months, so it is possible that some of this midline effect reflects this meaningful difference in mean WHZ at baseline for this outcome. Still, this pattern would not explain the other midline effects for WHZ between T1 and T4.

The evidence on the introduction of solid food or semi-solid soft foods in columns (6) and (7) point to a challenge for nutrition in Ethiopia; less than half of the caregivers for children 6-8 months report that the child has received solid or semi-solid food (again, this is in partial contrast to the result in columns (4) and (5)). Tables 7.3a and 7.3b look further at the diets of young children aged 6-23 months. Very few children in the sample meet the minimum dietary diversity cutoff of five food groups. Only 2.7 percent of control group children report achieving this modest benchmark in the midline compared to 0.6 percent in the baseline. For children in the 6-8-month age range, the share with five food groups is only 0.8 percent. On average, children who are this young consume only 1.7 food groups in a day. Three-quarters of the sample meet the recommendation for meal frequency, with T3 showing a significant increase of more than 10 percent over the control mean. Nevertheless, the small percentage of children who meet the minimum dietary diversity (MDD) cutoff is the limiting factor for dietary adequacy. As indicated in Table 7.3a, both diet diversity and meal frequency increase slightly with age.

7.3.2 Dietary diversity

Promotion of IYCF also aims at achieving a minimum acceptable diet which is a combination of the MDD and minimum meal frequency. While MDD has been defined as including 4 of 7 food groups in the previous day, after a 2017 expert consultation, this guideline was updated to reflect inclusion of breastmilk as an eighth food group. Therefore, the cutoff for MDD changed from 4 of 7 to 5 of 8 groups (WHO and UNICEF 2017). Adding an additional category while raising the threshold for MDD will clearly raise the average number of food groups recommended but, at most, it will keep the percentage consuming at or above the MDD constant or (more likely) will lower this percentage. Minimum meal frequency is defined as the proportion of children aged 6-23 months; who receive solid, semi-solid, or soft foods at least two times for children aged 6-8 months and three times for children aged 9-23 months.

Table 7.3b reports impact estimates for child dietary diversity and specific food groups consumed in the previous day by children aged 6-23 months.²⁴ T1 improved young child dietary diversity, increasing the number of foods consumed by 0.24 food groups. This effect is statistically significant. Also, on average T1 and T3 significantly increased the number of food groups consumed by children age 6-23 months, by 0.18 food groups. With the exception of eggs and legumes, there are few significant differences in dietary patterns among children by treatment groups. Egg consumption increases in all treatment groups, but the difference is larger in T1 and T2 where there was also a poultry transfer. The impact of the transfer is further discussed in Box 7.1. Table 7.4 presents impacts on consumption across food groups by primary caregivers. As in Table 7.3b, these tables also report the previous 24 hours but differ in that women's dietary adequacy is defined over 10 groups, which are reported in the table. There is far higher consumption of vegetables in the three groups of fruits and vegetables for women compared to the two corresponding categories for children. Also, there is a statistically significant increase in the proportion of women who meet the threshold for MDD (consumed at least 5 out of 10 food groups in the last 24 hours) in the T3 treatment group. T3 increased this measure of women's dietary diversity by 3.1 percentage points, which represents a 27 percent increase over the control group mean of 11.5 percent of women meeting this threshold.

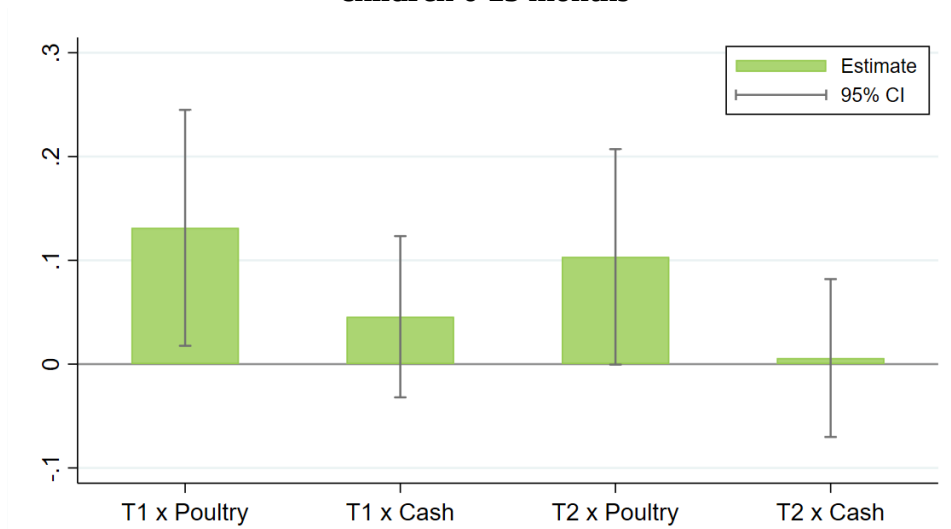
²⁴ These regressions control for the fact that the previous day may have been a fast day. The coefficients on fast days (not reported) indicate that although children are not expected to fast, children 6-23 months were reported to have consumed flesh foods as well as eggs less frequently by a statistically significant 5 percentage points. The corresponding declines of reported consumption of meat and eggs on fast days for women were 8 and 7 percentage points, respectively. However, children had no decrease in consumption of milk and milk products on fast days, while women reported a 9 percentage point decrease.

Box 7.1: The role of poultry transfers in diet diversity

There is evidence, albeit mixed, that regular egg consumption augments child growth (citation?). There is also evidence that home egg consumption is associated with poultry ownership (citation?). While this appears to be logical, households often focus on raising birds for meat and for sale. In a textbook model of farm households, production decisions and consumption decisions are viewed as separable. The randomization of poultry transfers matched with cash transfers of similar value in the L* kebeles provides a simple and clear test of this separability assumption.

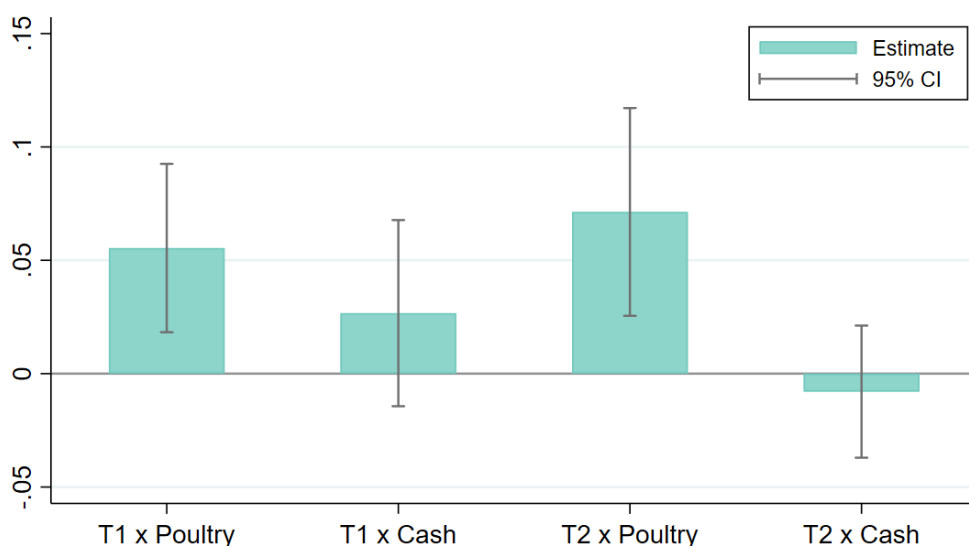
To verify this, regressions were run using the sub-sample of the poorest 10 households in each kebele following the approach discussed in chapter 6. Children in households that received the poultry transfers in L* communities had a statistically significant increase in the probability of consuming eggs in the previous day. The increase of 0.1 more than doubled egg consumption compared to the control mean of 0.06. The corresponding increase among those receiving cash transfers was significantly smaller than for the poultry transfer. While positive, it was not, however, significantly different from the control group. There was also no increase in the category of meats, a category that included a separate inquiry about poultry consumption. There was, however, a significant increase in legume consumption in the T2 treatment, an increase that was apparent for both the cash and the poultry transfer recipients. The increase of legumes in T1 was, however, not significantly different than the control.

Figure 7.2: Treatment effects on probability of egg consumption (poor sample only), children 6-23 months



Women who received the poultry transfers also increased their consumption of eggs significantly with the probability rising between 0.05 and 0.07, again more than a 100 percent increase over the control mean of only 0.04. This increase was significantly larger than the effect of the cash transfer in the sample communities. While there was no offsetting reducing reduction in consumption in for meats for children, the increase in egg consumption by women who receive poultry was matched by a similar decrease in the probability of consuming organ meats. This reduction, however, was also observed for the cash recipients in both the T1 and T2 sub-samples.

Figure 7.3: Treatment effects on probability of egg consumption (poor sample only), primary females



7.4 Access to and utilization of health services

7.4.1 Nutrition services

Table 7.5 explores aspects of nutrition services for young children. While children aged 6–59 months should receive vitamin A supplements every six months, the first column of Table 7.5 indicates that this is seldom the case and that there is no difference among treatment groups. Another recommendation is that children up to 36 months should be weighed monthly. However, since in woredas where community-based nutrition (CBN) is implemented the recommended age for regular weighing is up to 2 years, column (2) reports on children up to 24 months. As indicated in column (2), there is a significant increase in the probability that a child in this age group will be weighed among those in N* communities (T1 and T3). There is no corresponding recommended frequency for measuring height. Nevertheless, N* communities also take height measurements significantly more frequently than other communities, even though the percentage of children in N* communities that have their height measured is roughly only a third as large as the share who have their weight recorded (column (3)). The recommendation for measuring mid-upper arm circumference (MUAC)—a measure of wasting—is for monthly measurement from 6 months to 59 months. As indicated in column (4), N* communities also significantly increase the probability that a child has their MUAC measured.

Conditional on being measured, caregivers in both N and N* communities are significantly more likely to provide information on feeding (column (5)) even though there was no difference in the likelihood that the child was malnourished (column (6)). Furthermore, conditional on their child being deemed underweight, roughly half of caregivers in N* communities attended two-week cooking demonstration (Community Participatory Nutrition Promotion, or CPNP) sessions (column (7)). Since these sessions are a feature of the SPIR program in N* kebeles, the fact that a small number of caregivers in the control or T2 reported participating in such camps appears to be a

moderate amount of cross-over. Column (8) indicates the share of children deemed acutely malnourished among the children who had their MUAC measured. Unlike the risk of being underweight, the chance that a child has a small MUAC declines slightly with age. There is an apparent increase in acute malnutrition reported in N* kebeles conditional on being measured, but this may reflect the increased attention to measurement or the subsample composition.

7.4.2 Health extension

Table 7.6 indicates the contacts with HEWs and HDAs. By the various measures in the first six columns, caregivers in N* communities have more contact with both HEWs and HDAs, in general, as well as in the 3 months prior to the survey. Still, even with the increase of contact with health service providers in the N* sites, less than half of the families had contact with a HEW in the previous 3 months, and less than a fifth had a home visit. Home visits by HDAs are even rarer. This implies that the implementation of the scheduled visits for the TTC had not yet fully scaled up by the time of the midline survey. Of the 341 women who indicated they have been visited by an HDA member at home, 38.7 percent said their husband participated in the discussion during the visit, and a similar number of women indicated their husbands supported them while they tried the behaviors discussed. While the share of households having an HDA visit is significantly higher in the N* kebeles, there is no clear pattern on the participation of husbands, conditional on such a visit.

In addition to the CPNP feeding camps, which last for two weeks and are targeted to caregivers of acutely malnourished children, SPIR also organizes food demonstrations that are open to all members of the community. As indicated in column (7) of Table 7.6, nearly a fifth of the control group attended such demonstrations as did a far greater share of the N* and kebeles. The remainder of the columns in Table 7.6 reinforce the results on community nutrition, showing that the N and N* treatments increase participation in BCC as well as WASH activities, with caregivers in N* kebeles more likely to participate than those in T2, even though participation by families in the latter generally exceeds that of the control group. Nevertheless, at its highest, participation in any event still covers less than half of the target population.

7.5 Childcare and child development

Nurturing care goes beyond nutrition and health services; nurturing involves stimulation as well as feeding and protecting a child from illness. Table 7.7a and b indicate the activities undertaken with the index child in the past 3 days for female and male caregivers respectively. Comparing means of these bivariate responses in the control group, it is not surprising that meal preparation is primarily a female responsibility. However, there is also a large difference in the number of times a man reported having counted or drawn with their child compared to women. Conversely, there is little difference in the number of times a primary caregiver reported having told stories, or sung, or played with their child. Moreover, apparently it is virtually never the case that either parent reports sharing mealtime with their child. The nutrition intervention from SPIR appears to have a small but statistically significant impact on the number of times a respondent reported having spent time preparing meals by both caregivers as well as the verbal interactions indicated in column (5). These, however, remain small, and do not yet appear to be a focus of the project. Female education has a significantly positive impact in 7 of the 10 activities in Table 7.7a and three of the regressions in 7.7b. Similarly, male education has a positive role in six of the activities in 7.7b and two of the activities in 7.7a.

Table 7.8 reports on a measure of a child's overall developmental status, the Caregiver Reported Early Development Instruments (CREDI), a relatively new, open-source instrument for assessing early child development.²⁵ CREDI has both a short form for overall child development and a longer form that covers

²⁵ <https://sites.sph.harvard.edu/credi/credi-materials/>

separate domains of motor, cognitive, language and socioemotional development. The short form is used here. The instrument is designed for children from birth up to 36 months, although the questions the caregiver is asked depend on the age of the child. The CREDI website does not recommend standardizing responses with a functional equivalent of the z-scores used in assessing nutritional status. Thus, the mean values of the scores reported in Table 7.8 differ by child age. At this time, there are no significant differences across treatment groups in the development scores in the CREDI. The observed age, gender and education effects are consistent with similar results in the general literature.

7.6 Conclusions

The N*—and to a lesser degree the N—interventions have shown success in increasing exposure to health services including BCC, food demonstrations, and WASH. In the case of N* programs, these have led to an average increase in the frequency of measurement of 67 percent. Similarly, communication measures increased significantly, including the provision of feeding advice if a child was weighed (a 25 percent increase over the mean for the control) and participation in camps aimed at addressing acute malnutrition if a child was found to have a low MUAC (a 300 percent increase over the participation rate for the control). These increases reflect the goals of the SPIR design. Nevertheless, the frequency of contacts between households and service providers is less than planned and may indicate start-up processes at the time of the midline. Even at the level in June 2019, the SPIR program did have an impact on IYCF, particularly on the provision of colostrum and on breastfeeding. There is, however, a worrisome delay in the introduction of solid or semi-solid foods and an extremely low level of dietary diversity for young children. This is likely a contributing factor to the stunting levels and the decline in stature relative to international norms as children age. At this early stage of the program there is no measurable impact of the improved health-seeking and caregiving behaviors on nutritional status.

Figure 7.1: HAZ against child age in SPIR baseline and midline surveys and 2016 DHS, by region

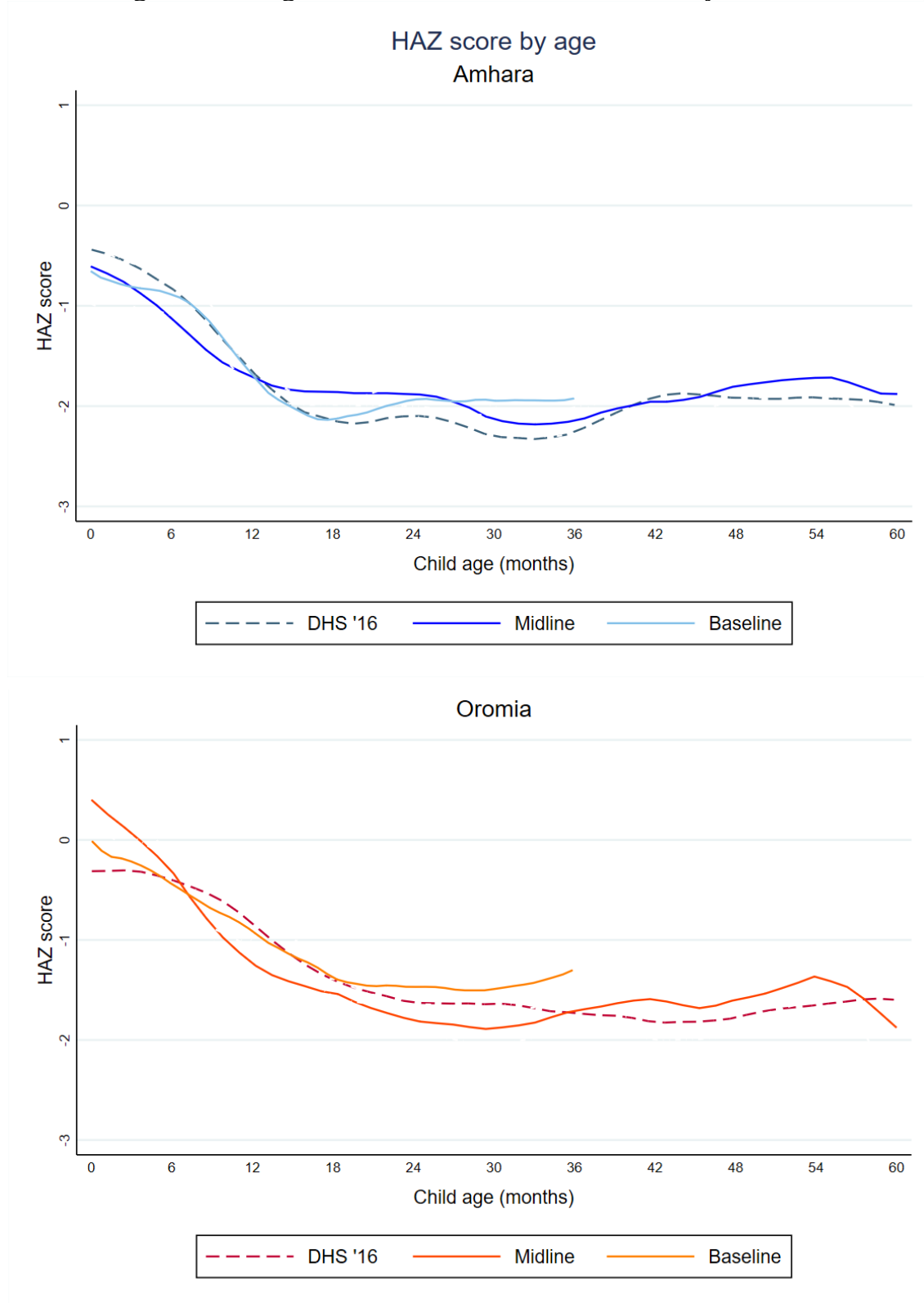


**Table 7.1a: Anthropometrics
Children under 36 months of age at midline**

	(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) Mid-upper arm circumference (MUAC)	(6) Proportion of severe acute malnutrition (MUAC<11.5 cm)
Child age in months	-0.020*** (0.002)	0.004*** (0.001)	0.001 (0.002)	-0.002*** (0.000)	0.014*** (0.004)	-0.001*** (0.000)
Male child	-0.104** (0.041)	0.047*** (0.014)	-0.034 (0.034)	0.020*** (0.008)	0.080 (0.127)	-0.009 (0.006)
T1	0.001 (0.079)	-0.000 (0.026)	-0.096 (0.066)	0.005 (0.012)	0.025 (0.158)	-0.010 (0.008)
T2	0.019 (0.076)	-0.012 (0.024)	-0.028 (0.061)	0.011 (0.013)	0.052 (0.154)	-0.014* (0.009)
T3	-0.075 (0.082)	0.013 (0.025)	-0.051 (0.061)	0.018 (0.012)	0.084 (0.165)	-0.000 (0.010)
Test: T1 = T2	0.814	0.607	0.220	0.616	0.839	0.505
Test: T2 = T3	0.240	0.249	0.638	0.554	0.823	0.070*
Test: T1 = T3	0.358	0.580	0.419	0.240	0.690	0.172
Mean of control (T4)	-1.572	0.422	-0.406	0.075	13.758	0.044
N	4,436	4,436	4,448	4,448	4,359	4,359

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.1b: Anthropometrics
Children 6-23 months of age at midline**

	(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) Mid-upper arm circumference (MUAC)	(6) Proportion of severe acute malnutrition (MUAC<11.5 cm)
Child age in months	-0.043*** (0.007)	0.016*** (0.002)	0.007 (0.006)	-0.003** (0.001)	0.038** (0.017)	-0.006*** (0.001)
Male child	-0.198*** (0.071)	0.086*** (0.024)	-0.120* (0.064)	0.019 (0.017)	-0.014 (0.228)	-0.018 (0.012)
T1	0.046 (0.114)	0.013 (0.037)	-0.200** (0.091)	0.018 (0.021)	0.176 (0.190)	-0.011 (0.014)
T2	0.063 (0.109)	-0.010 (0.034)	-0.069 (0.090)	0.018 (0.022)	0.085 (0.114)	-0.004 (0.015)
T3	-0.114 (0.107)	0.053 (0.034)	-0.214** (0.090)	0.048** (0.024)	-0.157 (0.161)	0.019 (0.017)
Test: T1 = T2	0.880	0.489	0.100*	0.984	0.559	0.599
Test: T2 = T3	0.072*	0.032**	0.060*	0.198	0.095*	0.120
Test: T1 = T3	0.134	0.224	0.865	0.175	0.154	0.033**
Mean of control (T4)	-1.523	0.392	-0.508	0.096	13.331	0.053
N	1,643	1,643	1,647	1,647	1,659	1,660

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.2: Infant and young child feeding (IYCF) practices
Midline**

	(1) Mother gave child colostrum	(2) Children born in the last 24 months who were put to the breast within one hour	(3) Infants 0-5 months of age who are fed exclusively with breast milk	(4) Children 6- 15 months: stopped exclusive breastfeeding at 6 months or later	(5) Month at which exclusive breastfeeding stopped (for children 6-15 months)	(6) Children 6-8 months of age who receive solid, semi- solid or soft foods	(7) Children 9-12 months of age who receive solid, semi- solid or soft foods
Child age in months			-0.089*** (0.015)				
Male child	-0.023 (0.016)	-0.001 (0.020)	0.009 (0.042)	-0.035* (0.018)	-0.337** (0.135)	-0.061 (0.066)	0.027 (0.034)
T1	0.057** (0.027)	0.083** (0.035)	0.124* (0.064)	-0.002 (0.027)	-0.255 (0.198)	0.062 (0.086)	-0.027 (0.045)
T2	0.007 (0.029)	0.009 (0.038)	0.085 (0.057)	-0.029 (0.024)	-0.139 (0.212)	0.060 (0.094)	-0.091 (0.057)
T3	0.057** (0.027)	0.082** (0.038)	0.095 (0.060)	-0.010 (0.027)	0.013 (0.206)	0.085 (0.084)	0.008 (0.052)
Linear combination: effect of N*	0.057** (0.024)	0.082** (0.032)	0.110** (0.053)	-0.006 (0.022)	-0.121 (0.179)	0.074 (0.075)	-0.010 (0.043)
Test: T1 = T2	0.053*	0.038**	0.526	0.343	0.563	0.984	0.200
Test: T2 = T3	0.051*	0.051*	0.859	0.494	0.450	0.784	0.069*
Test: T1 = T3	0.997	0.979	0.651	0.798	0.152	0.774	0.460
Mean of control (T4)	0.828	0.749	0.667	0.934	6.156	0.407	0.877
N	1,601	1,563	403	795	529	236	307

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.3a: Children's dietary diversity
Children of age 6-23 months at midline**

	(1) Children 6- 23 months of age who meet the minimum dietary diversity (5 of 8 food groups)	(2) Children 6- 23 months of age who meet the minimum meal frequency	(3) Children 6- 23 months of age who meet the minimum acceptable diet
Child age in months	0.002*** (0.001)	0.021*** (0.002)	0.002*** (0.001)
Male child	0.012 (0.009)	-0.045** (0.020)	0.012 (0.009)
T1	0.014 (0.012)	0.036 (0.032)	0.014 (0.012)
T2	0.003 (0.012)	0.032 (0.029)	0.001 (0.011)
T3	0.020* (0.012)	0.080*** (0.027)	0.013 (0.012)
Linear combination: effect of N*	0.017 (0.010)	0.058** (0.025)	0.013 (0.010)
Test: T1 = T2	0.378	0.916	0.269
Test: T2 = T3	0.142	0.105	0.250
Test: T1 = T3	0.601	0.166	0.967
Mean of control (T4)	0.027	0.716	0.027
N	1,672	1,668	1,668

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.3b: Children's dietary diversity
Children of age 6-23 months at midline**

	(1) Number of food groups (of 8) consumed	(2) Child consumed grains, roots or tubers	(3) Child consumed legumes or nuts	(4) Child consumed dairy	(5) Child consumed fish or meat	(6) Child consumed eggs	(7) Child consumed vitamin A rich fruits or vegetables	(8) Child consumed other fruits or vegetables
Child age in months	0.049*** (0.006)	0.030*** (0.002)	0.019*** (0.003)	0.001 (0.003)	0.003*** (0.001)	0.003* (0.002)	0.007*** (0.002)	0.001 (0.001)
Male child	-0.064 (0.065)	-0.006 (0.019)	-0.074*** (0.028)	0.002 (0.029)	0.015 (0.011)	0.008 (0.019)	0.016 (0.019)	0.003 (0.005)
T1	0.240** (0.110)	0.009 (0.030)	0.112** (0.044)	0.053 (0.051)	0.001 (0.012)	0.086*** (0.028)	0.008 (0.033)	-0.009 (0.007)
T2	0.083 (0.096)	0.011 (0.030)	0.105** (0.043)	-0.048 (0.046)	0.005 (0.015)	0.040* (0.023)	-0.012 (0.028)	-0.008 (0.008)
T3	0.123 (0.099)	0.013 (0.027)	0.012 (0.045)	0.019 (0.048)	0.032 (0.023)	0.038* (0.021)	0.015 (0.030)	-0.001 (0.010)
Linear combination: effect of N*	0.181** (0.090)	0.011 (0.025)	0.062 (0.038)	0.036 (0.045)	0.016 (0.015)	0.062*** (0.020)	0.012 (0.027)	-0.005 (0.008)
Test: T1 = T2	0.143	0.957	0.886	0.022**	0.710	0.116	0.516	0.840
Test: T2 = T3	0.675	0.931	0.037**	0.092*	0.201	0.934	0.320	0.317
Test: T1 = T3	0.270	0.878	0.026**	0.436	0.119	0.069*	0.843	0.227
Mean of control (T4)	2.612	0.876	0.318	0.331	0.021	0.041	0.128	0.012
N	1,009	1,009	1,009	1,009	1,009	1,009	1,009	1,009

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.4: Women's dietary diversity
Midline**

	(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 dairy	(6) Primary female consumed meat, fish, poultry	(7) Primary female consumed organ meat	(8) Primary female consumed eggs	(9) Primary female consumed green leafy vegetables	(10) Primary female consumed vitamin A rich fruits and vegetables	(11) Primary female consumed other vegetables	(12) Primary female consumed nuts and seeds
T1	0.096 (0.085)	0.016 (0.017)	0.000 (0.002)	0.057** (0.027)	0.037* (0.022)	0.000 (0.021)	-0.043** (0.018)	0.020 (0.012)	0.046* (0.025)	0.013 (0.023)	-0.029 (0.033)	0.002 (0.010)
T2	0.022 (0.078)	-0.005 (0.017)	-0.004 (0.003)	0.034 (0.025)	0.016 (0.023)	-0.008 (0.020)	-0.038** (0.018)	0.022* (0.012)	0.028 (0.024)	-0.004 (0.020)	-0.023 (0.033)	-0.002 (0.011)
T3	0.138* (0.079)	0.031** (0.015)	0.000 (0.002)	0.013 (0.029)	0.035 (0.024)	0.028 (0.023)	-0.017 (0.021)	0.007 (0.011)	0.034 (0.024)	0.027 (0.021)	0.001 (0.031)	0.016 (0.012)
Linear combination: effect of N*	0.117 (0.071)	0.024* (0.014)	0.000 (0.002)	0.035 (0.024)	0.036* (0.019)	0.014 (0.019)	-0.030* (0.018)	0.014 (0.010)	0.040* (0.022)	0.020 (0.018)	-0.014 (0.028)	0.009 (0.010)
Test: T1 = T2	0.360	0.227	0.181	0.388	0.355	0.661	0.724	0.914	0.393	0.455	0.869	0.683
Test: T2 = T3	0.122	0.032**	0.219	0.465	0.436	0.085*	0.199	0.237	0.748	0.172	0.412	0.093*
Test: T1 = T3	0.607	0.384	0.924	0.135	0.914	0.209	0.139	0.306	0.576	0.581	0.315	0.202
Mean of control (T4)	2.943	0.115	0.997	0.618	0.239	0.196	0.081	0.046	0.123	0.140	0.469	0.035
N	3,856	3,856	3,856	3,856	3,856	3,856	3,856	3,856	3,856	3,856	3,856	3,856

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.5: Child health history
Midline**

	(1) Received dose of vitamin A in past 6 months (6-59 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 (6-59 months)	(5) Feeding information was given at the time of measurement (<24 months)	(6) Child identified as severely or moderately underweight in past 6 months (<24 months)	(7) Child participated in a 2-week cooking demonstra- tion and feeding session (CPNP) (<24 months)	(8) Child identified as severely malnourished in past 6 months (6-59 months)	(9) Received a specific food or milk as treatment for severe malnutrition (6-59 months)
Child age in months	-0.001*** (0.000)	0.007*** (0.001)	0.005*** (0.001)	-0.005*** (0.001)	0.007* (0.003)	0.007** (0.003)	0.001 (0.006)	-0.003*** (0.001)	-0.001 (0.002)
Male child	-0.013 (0.013)	-0.017 (0.019)	-0.010 (0.013)	0.005 (0.011)	-0.018 (0.033)	-0.015 (0.038)	-0.011 (0.064)	-0.040 (0.026)	-0.099* (0.053)
T1	0.030 (0.027)	0.142*** (0.030)	0.046** (0.020)	0.098*** (0.025)	0.151*** (0.053)	0.069 (0.060)	0.384*** (0.080)	0.092** (0.036)	0.026 (0.083)
T2	0.018 (0.028)	-0.009 (0.029)	0.008 (0.019)	0.009 (0.024)	0.122** (0.053)	-0.005 (0.061)	-0.003 (0.066)	0.009 (0.035)	0.087 (0.089)
T3	0.035 (0.027)	0.116*** (0.028)	0.059*** (0.019)	0.078*** (0.023)	0.141*** (0.048)	0.012 (0.053)	0.391*** (0.077)	0.059 (0.036)	0.046 (0.076)
Linear combination: effect of N*	0.033 (0.023)	0.129*** (0.026)	0.053*** (0.016)	0.088*** (0.020)	0.146*** (0.046)	0.041 (0.052)	0.388*** (0.064)	0.076** (0.031)	0.036 (0.074)
Test: T1 = T2	0.670	0.000***	0.070*	0.001***	0.545	0.141	0.000***	0.020**	0.429
Test: T2 = T3	0.573	0.000***	0.014**	0.006***	0.654	0.682	0.000***	0.168	0.550
Test: T1 = T3	0.877	0.364	0.551	0.405	0.817	0.205	0.931	0.354	0.738
Mean of control (T4)	0.277	0.190	0.074	0.226	0.612	0.313	0.119	0.215	0.673
N	4,096	2,026	2,026	4,089	704	700	231	1,263	331

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.6: Exposure to health and nutrition services
Reported by primary female at midline**

	(1) Has had contact with the HEW	(2) Has had contact with the HEW in past 3 months	(3) Has been visited by the HEW at home in past 3 months	(4) Has had contact with the HDA/leader	(5) Has had contact with the HDA/leader in past 3 months	(6) Has been visited by the HDA/leader at home in past 3 months	(7) Has attended a food demonstra- tion in her community	(8) Has attended a BCC session at a public work site/health post	(9) Has attended a BCC session in past 3 months	(10) Has attended a community led total sanitation and hygiene (CLTSH) event	(11) Has participated in an Open Defecation Free (ODF) event
T1	0.048* (0.025)	0.089*** (0.029)	0.036* (0.021)	0.109*** (0.031)	0.100*** (0.025)	0.039** (0.017)	0.214*** (0.029)	0.146*** (0.031)	0.103*** (0.022)	0.136*** (0.033)	0.126*** (0.020)
T2	0.013 (0.026)	0.016 (0.030)	0.020 (0.023)	0.060** (0.028)	0.028 (0.020)	-0.003 (0.015)	0.102*** (0.028)	0.098*** (0.027)	0.033* (0.019)	0.079** (0.031)	0.043** (0.021)
T3	0.038* (0.023)	0.039 (0.025)	0.038* (0.021)	0.091*** (0.026)	0.061*** (0.021)	0.029* (0.015)	0.169*** (0.028)	0.171*** (0.028)	0.119*** (0.021)	0.153*** (0.030)	0.110*** (0.020)
Linear combination: effect of N*	0.043** (0.020)	0.064*** (0.024)	0.037** (0.018)	0.100*** (0.025)	0.080*** (0.019)	0.034** (0.013)	0.192*** (0.025)	0.158*** (0.026)	0.111*** (0.018)	0.144*** (0.027)	0.118*** (0.017)
Test: T1 = T2	0.229	0.023**	0.489	0.117	0.004***	0.016**	0.000***	0.093*	0.002***	0.085*	0.000***
Test: T2 = T3	0.336	0.396	0.417	0.224	0.115	0.035**	0.018**	0.005***	0.000***	0.014**	0.003***
Test: T1 = T3	0.712	0.072*	0.913	0.541	0.120	0.571	0.113	0.390	0.519	0.590	0.459
Mean of control (T4)	0.684	0.364	0.125	0.207	0.128	0.072	0.180	0.269	0.112	0.311	0.155
N	3,915	3,906	3,886	3,700	3,698	3,726	3,900	3,890	3,887	3,932	3,929

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.7a: Childcare activities in past 3 days
Reported by female**

	(1) Told stories to index child	(2) Sang songs to or with index child	(3) Took index child outside the home	(4) Played with index child	(5) Named, counted or drew things with or for index child	(6) Prepared food for index child who is not exclusively breastfed	(7) Physically fed index child who is not exclusively breasted	(8) Gave index child a bath	(9) Cared for the index child when they were sick	(10) Ate a meal together with index child
Child age in months	0.000 (0.000)	-0.000 (0.000)	0.001* (0.000)	0.004*** (0.000)	0.007*** (0.001)	-0.002*** (0.001)	-0.004*** (0.000)	-0.001** (0.000)	0.010*** (0.001)	-0.001*** (0.000)
Child age squared	-0.000 (0.000)	0.000 (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000 (0.000)	-0.000*** (0.000)	0.000*** (0.000)
Male child	-0.020 (0.013)	-0.028* (0.014)	-0.025* (0.014)	0.013 (0.012)	0.015 (0.013)	0.008 (0.015)	-0.003 (0.012)	0.013 (0.011)	-0.021 (0.015)	-0.011** (0.005)
Primary female has some education	0.069*** (0.016)	0.036* (0.020)	0.024 (0.017)	0.086*** (0.016)	0.033** (0.014)	0.071*** (0.018)	0.070*** (0.014)	0.013 (0.015)	0.044** (0.018)	-0.004 (0.006)
Primary male has some education	0.018 (0.015)	0.037** (0.018)	0.005 (0.017)	0.019 (0.014)	0.026* (0.015)	-0.011 (0.018)	0.005 (0.014)	0.004 (0.014)	0.005 (0.015)	-0.003 (0.005)
T1	0.007 (0.022)	0.033 (0.023)	0.008 (0.021)	0.018 (0.020)	0.049** (0.020)	0.044* (0.026)	0.031 (0.022)	0.012 (0.021)	0.036 (0.022)	-0.004 (0.010)
T2	0.017 (0.022)	0.003 (0.023)	0.000 (0.021)	0.020 (0.018)	0.003 (0.022)	0.057** (0.029)	0.006 (0.024)	-0.015 (0.020)	0.041* (0.023)	-0.018** (0.009)
T3	0.011 (0.021)	0.020 (0.025)	0.016 (0.022)	-0.005 (0.019)	0.024 (0.022)	0.069*** (0.026)	0.022 (0.022)	-0.016 (0.019)	0.021 (0.022)	-0.013 (0.009)
Linear combination: effect of N*	0.009 (0.019)	0.026 (0.021)	0.012 (0.019)	0.007 (0.017)	0.037** (0.018)	0.056** (0.023)	0.027 (0.018)	-0.002 (0.017)	0.029 (0.020)	-0.008 (0.008)
Test: T1 = T2	0.641	0.183	0.697	0.954	0.031**	0.629	0.328	0.225	0.793	0.092*
Test: T2 = T3	0.760	0.462	0.485	0.167	0.347	0.677	0.523	0.960	0.355	0.492
Test: T1 = T3	0.835	0.610	0.744	0.225	0.236	0.304	0.709	0.190	0.495	0.250
Mean of control (T4)	0.256	0.393	0.610	0.246	0.686	0.580	0.760	0.180	0.604	0.042
N	4,637	4,637	4,637	4,637	4,637	4,344	4,344	4,637	4,637	4,637

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

Table 7.7b: Childcare activities in past 3 days

Reported by male

	(1) Told stories index child	(2) Sang songs to or with index child	(3) Took index child outside the home	(4) Played with index child	(5) Named, counted or drew things with or for index child	(6) Prepared food for index child who is not exclusively breastfed	(7) Physically fed index child who is not exclusively breasted	(8) Gave index child a bath	(9) Cared for the index child when they were sick	(10) Ate a meal together with index child
Child age in months	0.023*** (0.004)	0.041*** (0.005)	0.048*** (0.005)	0.030*** (0.004)	0.013*** (0.003)	0.066*** (0.006)	0.014** (0.007)	0.013*** (0.003)	0.060*** (0.004)	-0.058*** (0.004)
Child age squared	-0.000*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.000*** (0.000)	-0.000** (0.000)	-0.002*** (0.000)	-0.000 (0.000)	-0.000*** (0.000)	-0.001*** (0.000)	0.001*** (0.000)
Male child	-0.005 (0.019)	0.022 (0.020)	0.031 (0.020)	0.008 (0.017)	-0.008 (0.012)	0.028 (0.021)	-0.005 (0.021)	0.011 (0.014)	0.005 (0.020)	-0.012 (0.014)
Primary female has some education	0.036 (0.022)	0.008 (0.023)	0.002 (0.024)	0.048** (0.021)	0.027 (0.018)	0.081*** (0.025)	0.040 (0.026)	0.034* (0.018)	0.024 (0.024)	-0.017 (0.019)
Primary male has some education	0.060*** (0.020)	0.071*** (0.020)	0.058*** (0.021)	0.062*** (0.018)	0.033** (0.014)	0.086*** (0.022)	0.061*** (0.023)	0.018 (0.014)	0.028 (0.019)	-0.038** (0.016)
T1	0.034 (0.023)	0.047* (0.028)	-0.002 (0.027)	0.017 (0.022)	0.020 (0.020)	0.077*** (0.028)	0.018 (0.032)	-0.011 (0.025)	0.043 (0.030)	-0.004 (0.023)
T2	0.024 (0.022)	-0.012 (0.033)	-0.006 (0.030)	-0.011 (0.024)	-0.024 (0.018)	0.019 (0.029)	0.011 (0.030)	-0.037* (0.021)	0.020 (0.031)	0.002 (0.026)
T3	0.021 (0.024)	0.010 (0.028)	-0.002 (0.028)	-0.022 (0.023)	0.036** (0.018)	0.034 (0.025)	0.034 (0.030)	-0.057*** (0.022)	-0.015 (0.030)	-0.032 (0.024)
Linear combination: effect of N*	0.028 (0.020)	0.028 (0.023)	-0.002 (0.024)	-0.002 (0.019)	0.028* (0.016)	0.055** (0.023)	0.026 (0.027)	-0.034* (0.021)	0.014 (0.026)	-0.018 (0.021)
Test: T1 = T2	0.687	0.090*	0.897	0.274	0.025**	0.044**	0.818	0.231	0.471	0.777
Test: T2 = T3	0.902	0.510	0.890	0.681	0.001***	0.555	0.403	0.247	0.256	0.118
Test: T1 = T3	0.625	0.219	0.994	0.118	0.454	0.097*	0.603	0.037**	0.064*	0.149
Mean of control (T4)	0.240	0.361	0.628	0.273	0.102	0.457	0.241	0.159	0.468	0.173
N	2,170	2,170	2,170	2,170	2,170	1,920	1,920	2,170	2,170	2,170

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

**Table 7.8: Caregiver Reported Early Development Instruments (CREDI)
by age group at midline**

	(1) CREDI score (0-17 months)	(2) CREDI score (18-29 months)
Child age in months	0.654*** (0.112)	-0.082 (0.425)
Child age squared	-0.018*** (0.006)	0.010 (0.009)
Male child	-0.612** (0.256)	-0.349 (0.221)
Primary female has some education	0.075 (0.306)	0.537** (0.240)
Primary male has some education	0.416 (0.287)	0.137 (0.220)
T1	0.071 (0.425)	0.314 (0.255)
T2	-0.111 (0.370)	-0.089 (0.274)
T3	-0.375 (0.394)	-0.130 (0.258)
Linear combination: effect of N*	-0.152 (0.350)	0.092 (0.221)
Test: T1 = T2	0.654	0.153
Test: T2 = T3	0.478	0.883
Test: T1 = T3	0.297	0.092*
Mean of control (T4)	5.520	4.355
N	1,305	1,292

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and the cluster mean of the baseline value of the outcome. * p<0.1 ** p<0.05 *** p<0.01

8 Evidence on Gender Norms and Women's Well-being

8.1 Introduction

In this chapter we report impacts of the different SPIR treatments on gender norms, women's decision-making, intimate partner violence (IPV) and depression. All four interventions – L, L*, N, N*- have the potential to improve gender norms and women's well-being. 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); targeting women for the poultry and cash transfers; including men in the nutrition BCC and creating male engagement groups; and providing IPT-G for women and men screened for depression. 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 also pathways through which SPIR may lead to reductions in maternal and paternal depression, and IPV (Lund et al. 2018; Buller et al. 2018).

At the time of the midline survey, however, the IPT-G groups had not yet begun and the male's engagement had not been fully rolled out.

8.1.1 Interpreting tables

Similar to the tables shown in Chapter 6 on livelihoods, we present three sets of tables for each set of outcomes related to depression, IPV, gender norms, and decisionmaking: 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 poor and nonpoor samples. Here, we briefly summarize again how to interpret the second and third table that look at the restricted samples.

The tables reporting effects for the 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 poor household in T1 and T2 was randomized to receive either poultry or cash, with or without the additional aspirations intervention. For both T1 and T2, the tables report the coefficients for the effects of the poultry and cash intervention compared to the control arm and then the additional effects of the aspirations treatment. In other words, the coefficient for *T1*poultry* represents the effect of the poultry intervention compared to the control arm in the T1 treatment arm without aspirations, and the coefficient for *T1*poultry*aspirations* represents the additional impact of the aspirations treatment for households in the T1 poultry group. The table also reports the coefficient for T3.

Using the coefficients on the cash and poultry interventions with and without aspirations, we calculate linear combinations (means) to estimate the average effect of T1 and T2 for poor households, and the average effect of poultry and cash. The latter reports the average effects for poultry and cash abstracting from the additional estimated effect of the aspirations treatment. We again report the same tests of equality across treatment arms.

The tables reporting effects for the nonpoor 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 and T2 without aspirations, and the additional effect of adding the aspirations in each arm. The table also reports the coefficient for T3. We again report linear combinations corresponding to the average effects of T1 and T2.

8.2 Depression

8.2.1 Indicators

The PHQ-9—a nine-item depression diagnostic instrument—was used to assess depressive symptoms of both the primary male and female. The respondent was only administered the instrument if she/he was alone or with a child younger than 36 months. Of the 3,132 primary females, 2,645 (or 84 percent) were alone or with a child younger than 36 months and thus eligible to be administered the instrument, and of the 2,482 primary males, 2,055 (or 83 percent) were eligible to be administered the instrument. 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 and 19, 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 no depressive symptoms, mild to severe depressive symptoms using 8 as the cutoff, and moderate to severe depressive symptoms using 10 as a cutoff.

8.2.2 Pooled effect by treatment arm

[Text redacted pending publication of results]

8.2.3 Poor households

[Text redacted pending publication of results]

8.2.4 Nonpoor households

[Text redacted pending publication of results]

8.3 Intimate partner violence

8.3.1 Indicators

Indicators of internationally validated standardized IPV measures from the WHO Violence Against Women Instrument were administered in the baseline and midline surveys and included three types of violence (physical, sexual, emotional). Violence 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 3,132 primary females at midline, only 2,566 lived with their partners in the last 13 months. Due to CAPI errors in the questionnaire (see Chapter 4 for more detail), only 1,922 of the 2,566 women were administered the IPV module, of whom 1,660 were also alone or with a child <36 months. Of the 1,660 eligible women, 1,648 agreed to be administered the WHO module.

For the three types of violence, multiple behaviorally specific questions were administered in order to reduce underreporting. We asked if the women had ever experienced the 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 attack 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.

8.3.2 Pooled effect by treatment arm

[Text redacted pending publication of results]

8.3.3 Poor households

[Text redacted pending publication of results]

8.3.4 Nonpoor households

[Text redacted pending publication of results]

8.4 Gender norms

8.4.1 Indicators

We construct three different indicators of equitable gender norms. The first is from a list of four questions asked separately to the primary male and female on whether a husband is justified in beating his wife under different circumstances. We create a binary indicator that equals one if the respondent says a husband is not justified in beating his wife under any of the four circumstances. The second is from a list of three questions, again asked separately to the primary male and female, on whether it is acceptable for a woman to travel alone to different places. 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 him a list of 10 questions on gender inequitable norms where the responses are scored as strongly agrees (=0), somewhat agrees (=1), or does not agree (=2). Scores are summed up so that a higher score implies more equitable gender norms. We analyze both the continuous score and a binary indicator that equals one if scores are above the median.

8.4.2 Pooled effect by treatment arm

Table 8.3a reveals that T1 and T2 improve the primary male's gender norms but there are no impacts on the primary female's gender norms. In particular, T1 leads to improvements in the primary male's index of support for gender equitable norms of .86 points which represents a 10 percent improvement with respect to the control groups mean of 8.6 points; it also increases the probability that a man is above the median score by 5.9 percentage points. T2 leads to increases in the probability that a man reports it is not justified to beat his wife by 5.9 percentage points (compared to a control mean of 60 percent); the probability that a man reports it is acceptable for a wife to travel alone by 6.2 percent (compared to a control mean of 33 percent); and increases the gender equitable norms scale by .6 points.

8.4.3 Poor households

Among poorer households, the T1 and T2 poultry interventions lead to improvements in the primary male's gender equitable norms scale (Table 8.3b). There is some indication of heterogeneous impacts with respect to the aspirations treatment in the T2 arm; the coefficient on the interaction is negative implying that the positive impact of the T2 poultry intervention disappears with the aspirations treatment. The T2 cash intervention also leads to a significant improvement in the probability that the primary male reports that it is not justified to beat a wife by 12.2 percentage points (compared to a mean in the control group of 61 percent). Overall, impacts of T1, T2, and T3 are not significant with the exception of T1 which marginally improves the primary male's gender equitable norms score. The combined impact of the poultry intervention across T1 and T2 improves the probability that the primary male reports it is acceptable for a woman to travel alone by 8.2 percentage points (compared to a mean in the control group of 35 percent) and improves the primary male's gender equitable norms scale by 1.36 points (compared to a mean in the control group of 8.69 points). However, this result should be interpreted with caution, as at baseline, the indicator for the primary male reporting it was acceptable for a woman to travel alone was not balanced across the poultry and control arms.

8.4.4 Nonpoor households

Similar to the impacts on poorer households, T1 leads to marginal improvements in the primary male's gender equitable norms score by .89 points (Table 8.3c). T2 leads to marginal increases in the probability that the primary male reports it is acceptable for a woman to travel alone by 7 percentage points and increases his gender equitable norms score by .78 points. T3 leads to marginal increases in the probability that the primary male reports it is not justified for a husband to beat his wife and that he is above the median on the gender equitable norms scale. There are no differential impacts with respect to the aspirations treatment.

8.5 Decision-making

8.5.1 Indicators

We analyze impacts of SPIR on women's decision-making with respect to input into productive decisions across 7 domains that include crop farming, horticulture, large livestock, small livestock, poultry, large household purchases and small household purchases. For each activity we ask 4 questions on decisionmaking. The first question is more general about how much input the respondent has in making decisions about the activity. The second through fourth questions – which are only asked about horticulture, large livestock, small livestock, and poultry - are more specific to the quantity of inputs that should be used; the quantity of the output that should be sold or consumed at home; and the use of income generated from the activity. We create 4 binary indicators that equal one if a woman reports having input into most decisions for all activities in which she participates. We use most decisions instead of some decisions because a high percentage of women already have some decisionmaking power (in the control group, 84-92 percent of women have input into some decisions for all the productive activities in which they participate).

8.5.2 Pooled effect by treatment arm

As can be seen by the mean in the control arm, 40-55 percent of women have input into most decisions for all the productive activities in which they participate (Table 8.4a). All three arms lead to large increases in the probability of having input into most decisions with respect to the use of income generated from productive activities that ranges from 6.1-8.7 percentage points. T2 also leads to a marginally significant increase in decisionmaking on how much of the output of the activities should be consumed or sold.

8.5.3 Poor households

Within the subsample of poorer households, the T2 poultry intervention consistently leads to large increases in decisionmaking across the three latter indicators (Table 8.4b). Impacts on the probability that a woman has input into most decisions range from 9.6-15.9 percentage points for decisions with respect to input use, if output should be consumed or sold, and the use of income generated from the productive activities. There is some indication that there are negative heterogeneous effects of the aspirations intervention in the T2 poultry arm. Overall, T1 and T2 lead to large impacts on decisionmaking with respect to the use of income generated from the productive activities. Impacts of T1 are mainly from the cash sub-randomization, while impacts of T2 are mainly from the poultry sub-randomization. The poultry intervention -across all sub treatments - also leads to large impacts on the probability that a woman makes most decisions with respect to input use and the use of income generated from productive activities.

8.5.4 Nonpoor households

Among the subsample of nonpoor households, T1 and T3 have no impacts on women's decisionmaking, but T2 has large impacts (Table 8.4c). In particular, T2 increases the probability that a woman has input into most decisions with respect to general decisionmaking (column 1), input use (column 2), and the use of income generated from productive activities (column 4). Impacts range from 6.9 to 10.8 percentage points and are significantly different from T1.

8.6 Conclusions

[Text redacted pending publication of results]

We also find significant improvements in the primary male's gender norms, but not in the primary females. In particular, among the full sample, T1 and T2 lead to improvements of .61-.86 points (or 7-10 percent) on the gender equitable norms scale, and T2 also improves the probability that the primary male reports that it is not justified to beat his wife by 5.9 percentage points (or 60 percent) and that it is acceptable for a woman to travel alone by 6.2 percentage points (or 19 percent). Among the poor subsample, impacts are concentrated in the poultry arms, while among the non-poor subsample, impacts occur across all three treatment arms (T1, T2, T3).

Lastly, we find significant improvements in a woman's decisionmaking with respect to productive activities across all three arms among the full sample. In particular, all three arms lead to significant increases in the probability of a woman having input into most decisions with respect to the use of income generated from productive activities that ranges from 6.1-8.7 percentage points (or 12.8-18.2 percent compared to the control mean). Among the poor sample, impacts are concentrated in the T1 and T2 arms and in the poultry sub randomization, while among the non-poor sample, impacts are concentrated in T2.

A couple of caveats are worth noting. The first is that we recognize that the nature of this report is to analyze impacts on multiple outcomes and across multiple samples. While informative, this leads to issues of multiple hypothesis testing, which we plan to address in future analysis. Second, is that we did not stratify our sample across poor and non-poor households, and therefore, we may not be balanced at baseline on a few indicators in the subsamples. This is the case for depression outcomes among the poor sample.

Despite these limitations, the handful of positive impacts across T1, T2, and T3 are promising given that not all activities had been rolled out by midline and some outcomes might take more time to change, such as personal norms regarding gender.

**Table 8.1a: Female's and male's depression outcomes
by SPIR treatment arms**

[Table redacted pending publication of results]

**Table 8.1b: Female's and male's depression outcomes: Subsample of poor households
by SPIR subtreatment arms**

[Table redacted pending publication of results]

**Table 8.1c: Female's and male's depression outcomes: Subsample of nonpoor households
by SPIR subtreatment arms**

[Table redacted pending publication of results]

**Table 8.2a: Experience with intimate partner violence (IPV) over past year
by SPIR treatment arms**

[Table redacted pending publication of results]

**Table 8.2b: Experience with intimate partner violence (IPV) over past year: Subsample of poor households
by SPIR subtreatment arms**

[Table redacted pending publication of results]

Table 8.2c: Experience with intimate partner violence (IPV) over past year: Subsample of nonpoor households by SPIR subtreatment arms

[Table redacted pending publication of results]

**Table 8.3a: Perceived gender norms
by SPIR subtreatment arms**

	(1) Female: A husband is not justified in beating his wife in any of these situations [§]	(2) Male: A husband is not justified in beating his wife in any of these situations [§]	(3) Female: Acceptable for a woman to travel alone to market, health center, and to visit friends	(4) Male: Acceptable for a woman to travel alone to market, health center, and to visit friends	(5) Male: Index of support for equitable gender norms: all questions	(6) Male: Above median support for equitable gender norms
T1	0.015 (0.032)	0.021 (0.037)	0.010 (0.027)	0.044 (0.030)	0.862** (0.337)	0.059* (0.035)
T2	-0.015 (0.028)	0.059* (0.034)	-0.012 (0.027)	0.062** (0.029)	0.609* (0.354)	0.035 (0.035)
T3	0.028 (0.031)	0.055 (0.034)	0.032 (0.025)	0.037 (0.030)	0.604 (0.367)	0.049 (0.035)
Test: T1 = T2	0.324	0.268	0.393	0.541	0.404	0.445
Test: T2 = T3	0.134	0.890	0.063*	0.383	0.988	0.650
Test: T1 = T3	0.680	0.329	0.368	0.826	0.425	0.755
Mean of control (T4)	0.416	0.600	0.285	0.334	8.577	0.445
N	3,093	2,473	3,116	2,478	2,480	2,480

Notes: [§] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband. Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and baseline values of the outcomes that are included as ANCOVA models. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

**Table 8.3b: Perceived gender norms: sub-sample of poor households
by SPIR subtreatment arms**

	(1) Female: A husband is not justified in beating his wife in any of these situations [§]	(2) Male: A husband is not justified in beating his wife in any of these situations [§]	(3) Female: Acceptable for a woman to travel alone to market, health center,	(4) Male: Acceptable for a woman to travel alone to market, health center,	(5) Male: Index of support for equitable gender norms: all questions	(6) Male: Above median support for equitable gender norms
--	---	---	---	---	---	---

			and to visit friends	and to visit friends		
T1 x Poultry	-0.017 (0.062)	0.026 (0.060)	0.021 (0.043)	0.089 (0.065)	1.288* (0.699)	0.082 (0.065)
T1 x Poultry x Aspirations	0.012 (0.080)	-0.045 (0.082)	-0.028 (0.053)	-0.120 (0.083)	-0.426 (0.839)	-0.009 (0.078)
T1 x Cash	0.080 (0.054)	0.021 (0.079)	0.021 (0.055)	0.031 (0.058)	0.185 (0.782)	-0.028 (0.075)
T1 x Cash x Aspirations	-0.075 (0.075)	0.046 (0.099)	-0.092 (0.069)	0.035 (0.079)	0.975 (1.067)	0.084 (0.104)
T2 x Poultry	0.048 (0.049)	0.042 (0.078)	-0.039 (0.043)	0.075 (0.047)	1.425** (0.609)	0.128** (0.059)
T2 x Poultry x Aspirations	-0.112* (0.066)	-0.061 (0.090)	-0.041 (0.057)	-0.062 (0.064)	-1.482* (0.829)	-0.129* (0.075)
T2 x Cash	0.038 (0.053)	0.122*** (0.034)	-0.006 (0.046)	0.069 (0.063)	0.695 (0.655)	0.016 (0.067)
T2 x Cash x Aspirations	-0.061 (0.069)	-0.054 (0.064)	-0.018 (0.063)	-0.015 (0.071)	-0.862 (0.725)	-0.058 (0.071)
T3	0.038 (0.040)	0.034 (0.036)	0.020 (0.030)	0.037 (0.041)	0.585 (0.428)	0.025 (0.044)
Linear combination: effect of T1	0.017 (0.042)	0.023 (0.041)	-0.009 (0.031)	0.037 (0.039)	0.862* (0.466)	0.045 (0.046)
Linear combination: effect of T2	0.001 (0.039)	0.053 (0.039)	-0.037 (0.031)	0.053 (0.036)	0.494 (0.428)	0.027 (0.043)
Linear combination: effect of poultry	0.016 (0.045)	0.034 (0.052)	-0.010 (0.034)	0.082* (0.043)	1.361*** (0.515)	0.106** (0.049)
Linear combination: effect of cash	0.059 (0.044)	0.070 (0.047)	0.008 (0.039)	0.050 (0.047)	0.435 (0.560)	-0.007 (0.056)
Test: T1 = T2	0.650	0.482	0.353	0.656	0.389	0.666
Test: T2 = T3	0.270	0.608	0.040**	0.666	0.811	0.947
Test: T1 = T3	0.575	0.779	0.296	0.999	0.516	0.630
Mean of control (T4)	0.415	0.609	0.298	0.350	8.696	0.456
N	1,780	1,291	1,792	1,293	1,295	1,295

Notes: ^s Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband. Estimates from the 'poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the subrandomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm subrandomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and baseline values of the outcomes that are included as ANCOVA models. * p<0.1 ** p<0.05 *** p<0.01

**Table 8.3c: Perceived gender norms: sub-sample of non-poor households
by SPIR sub-treatment arms**

	(1) Female: A husband is not justified in beating his wife in any of these situations [§]	(2) Male: A husband is not justified in beating his wife in any of these situations [§]	(3) Female: Acceptable for a woman to travel alone to market, health center, and to visit friends	(4) Male: Acceptable for a woman to travel alone to market, health center, and to visit friends	(5) Male: Index of support for equitable gender norms: all questions	(6) Male: Above median support for equitable gender norms
T1	0.000 (0.047)	0.023 (0.054)	0.004 (0.050)	0.056 (0.051)	0.656 (0.574)	0.038 (0.054)
T1 x Aspirations	0.022 (0.057)	-0.006 (0.065)	0.059 (0.055)	-0.003 (0.061)	0.460 (0.683)	0.073 (0.063)
T2	-0.032 (0.043)	0.035 (0.049)	0.002 (0.042)	0.049 (0.049)	0.499 (0.483)	0.030 (0.047)
T2 x Aspirations	-0.010 (0.054)	0.056 (0.047)	0.037 (0.046)	0.044 (0.057)	0.577 (0.572)	0.031 (0.056)
T3	0.015 (0.044)	0.078* (0.047)	0.047 (0.036)	0.040 (0.036)	0.642 (0.477)	0.073* (0.044)
Linear combination: effect of T1	0.011 (0.039)	0.020 (0.048)	0.033 (0.039)	0.054 (0.039)	0.886* (0.485)	0.074 (0.045)
Linear combination: effect of T2	-0.037 (0.038)	0.062 (0.043)	0.021 (0.036)	0.070* (0.038)	0.782* (0.441)	0.045 (0.043)
Test: T1 = T2	0.217	0.290	0.719	0.700	0.812	0.481
Test: T2 = T3	0.227	0.662	0.402	0.425	0.742	0.488
Test: T1 = T3	0.930	0.201	0.695	0.726	0.606	0.967
Mean of control (T4)	0.417	0.589	0.266	0.318	8.446	0.432
N	1,313	1,182	1,324	1,185	1,185	1,185

Notes: [§] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband. Estimates from the 'non-poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects and baseline values of the outcomes that are included as ANCOVA models. * p<0.1 ** p<0.05 *** p<0.01

**Table 8.4a: Female's input into decision making around production
by SPIR sub-treatment arms**

	(1) Female had input into most decisions about all activities she participates in	(2) Female had input into most inputs' decisions about all activities she participates in	(3) Female had input into most outputs' decisions about all activities she participates in	(4) Female had input into most decisions about generated income for all activities she participates in
T1	0.020 (0.033)	0.037 (0.036)	0.020 (0.034)	0.076** (0.033)
T2	0.002 (0.029)	0.053 (0.035)	0.057* (0.033)	0.087*** (0.031)
T3	0.016 (0.031)	0.011 (0.036)	0.044 (0.034)	0.061* (0.034)
Test: T1 = T2	0.518	0.585	0.224	0.708
Test: T2 = T3	0.576	0.131	0.673	0.382
Test: T1 = T3	0.893	0.394	0.430	0.637
Mean of control (T4)	0.396	0.546	0.503	0.478
N	3,113	2,429	2,429	2,429

Notes: Estimates from the DFSA SPIR midline survey sample. Coefficients for T1, T2 and T3 are impacts of each pooled treatment arm (including sub-randomized aspirations or poultry and cash for L* in T1 and T2) on the outcome relative to control (T4). Standard errors (in parentheses) are clustered at the kebele level. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

**Table 8.4b: Female's input into decision making around production: sub-sample of poor households
by SPIR sub-treatment arms**

	(1) Female had input into most decisions about all activities she participates in	(2) Female had input into most inputs' decisions about all activities she participates in	(3) Female had input into most outputs' decisions about all activities she participates in	(4) Female had input into most decisions about generated income for all activities she participates in
T1 x Poultry	0.016 (0.051)	0.021 (0.054)	-0.003 (0.035)	0.056 (0.036)

T1 x Poultry x Aspirations	-0.052 (0.062)	0.033 (0.057)	0.088* (0.047)	0.019 (0.049)
T1 x Cash	0.032 (0.069)	0.112 (0.073)	0.088 (0.067)	0.153** (0.073)
T1 x Cash x Aspirations	0.032 (0.076)	-0.059 (0.099)	-0.034 (0.098)	-0.061 (0.097)
T2 x Poultry	-0.039 (0.038)	0.159*** (0.050)	0.096* (0.049)	0.141*** (0.047)
T2 x Poultry x Aspirations	-0.121** (0.053)	-0.138** (0.057)	-0.067 (0.073)	-0.073 (0.072)
T2 x Cash	0.017 (0.047)	-0.099* (0.057)	-0.039 (0.069)	-0.028 (0.065)
T2 x Cash x Aspirations	-0.022 (0.055)	0.086 (0.071)	0.121 (0.079)	0.109 (0.073)
T3	0.022 (0.043)	-0.000 (0.043)	0.056 (0.039)	0.060 (0.043)
Linear combination: effect of T1	0.019 (0.041)	0.061 (0.043)	0.057 (0.039)	0.095** (0.040)
Linear combination: effect of T2	-0.046 (0.039)	0.020 (0.040)	0.043 (0.038)	0.067* (0.040)
Linear combination: effect of poultry	-0.012 (0.039)	0.091** (0.043)	0.047 (0.035)	0.100*** (0.036)
Linear combination: effect of cash	0.025 (0.048)	0.008 (0.051)	0.025 (0.051)	0.064 (0.052)
Test: T1 = T2	0.033**	0.256	0.705	0.441
Test: T2 = T3	0.034**	0.555	0.723	0.867
Test: T1 = T3	0.931	0.116	0.979	0.385
Mean of control (T4)	0.445	0.596	0.534	0.524
N	1,791	1,376	1,376	1,376

Notes: Estimates from the 'poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. * p<0.1 ** p<0.05 *** p<0.01

Table 8.4c: Female's input into decision making around production: sub-sample of nonpoor households by SPIR sub treatment arms

	(1) Female had input into most decisions about all activities she participates in	(2) Female had input into most inputs' decisions about all activities she participates in	(3) Female had input into most outputs' decisions about all activities she participates in	(4) Female had input into most decisions about generated income for all activities she participates in
T1	-0.013 (0.048)	-0.036 (0.055)	-0.039 (0.053)	0.051 (0.055)
T1 x Aspirations	0.085 (0.056)	0.076 (0.059)	0.019 (0.062)	-0.009 (0.066)
T2	0.059 (0.041)	0.093 (0.057)	0.051 (0.055)	0.111** (0.052)
T2 x Aspirations	0.020 (0.057)	-0.017 (0.058)	0.044 (0.058)	-0.005 (0.053)
T3	0.010 (0.039)	0.021 (0.049)	0.034 (0.046)	0.062 (0.045)
Linear combination: effect of T1	0.030 (0.041)	0.002 (0.050)	-0.030 (0.048)	0.047 (0.048)
Linear combination: effect of T2	0.069* (0.040)	0.084* (0.048)	0.072 (0.045)	0.108** (0.042)
Test: T1 = T2	0.317	0.047**	0.015**	0.142
Test: T2 = T3	0.105	0.112	0.325	0.241
Test: T1 = T3	0.606	0.652	0.125	0.727
Mean of control (T4)	0.324	0.485	0.464	0.422
N	1,322	1,053	1,053	1,053

Notes: Estimates from the 'non-poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. * p<0.1 ** p<0.05 *** p<0.01

9 Conclusion

As a graduation model program working to expand the evidence on integrated strategies to reduce poverty by improving livelihoods, wellbeing and empowerment for women and nutrition for their children, SPIR as an ambitious project. SPIR project partners from World Vision, CARE and ORDA used their expertise to design and implement core gender-sensitive interventions in Amhara and Oromia in livelihoods (L – including formation of VESA groups, financial literacy, agriculture and livestock value chain development and home gardening and forage production) and nutrition (N – including nutrition social behavior change communication (SBCC), WASH activities including CLTSH, delivered through a cascade training model starting with health extension worker supervisors). To then developed enhanced livelihoods models (L* – adding Social Analysis and Action (SAA) to improve women’s access to markets, aspirations documentaries, and poultry or cash livelihood transfers for poor women) and enhanced nutrition models (N* – adding Timed and Targeted Counseling, CPNP, and male engagement in BCC supported by Community Health Facilitators) to test as part of the evaluation field experiment. The evaluation design combines these into four treatment arms of packages of graduation model interventions as T1: L* + N*, T2: L* + N, T3: L + N*, and T4: control.

Implementation of a large number of SPIR components rolled out over more than a year of implementation, engaging PSNP beneficiaries in the SPIR project in new value chain activities and lessons about improving child nutrition and caring practices. For some activities, such as male engagement groups and the IPT-G therapy groups not rolled out substantially or at all before the midline survey. The SAA interventions to promote changes in norms around women’s access to markets and supply of labor were also not implemented with the intensity that was originally planned. The poultry and cash livelihoods transfers were delivered roughly 7 months before the midline survey, which would have been enough time to see impacts on holdings of poultry and sale and consumption of eggs, but maybe not yet for longer term effects on women’s empowerment and control over their livelihoods and other aspects of decision-making or status within the household.

This midline report summarizes information from the midline survey conducted in July-October 2019, roughly 14 months after the formation of VESA groups, the start of implementation activity, following the baseline survey. Topics covered include participation and exposure to the SPIR project components as well as impact of the treatment arms on outcomes in livelihoods, nutrition, child caring practices, health care utilization, maternal depression and intimate partner violence.

Results show that delivery of the SPIR interventions was effective across most activities, with some exceptions, and that the SPIR partners were successful at maintaining fidelity to the randomized study design. A high share of respondents report belonging VESA groups in SPIR intervention communities. While attendance at VESA group meetings varied, those that were engaged attended regularly. Poultry kits were received by a very high share of those assigned to the poultry livelihoods intervention. However, respondents assigned to receive the cash livelihoods transfer or invited to the aspirations documentaries report relatively low levels of compliance. Households in N* kebeles do report more interaction with health service providers.

Impact estimates show that the poultry transfer had a large impact on increasing household ownership of poultry, particularly for women and increasing income from poultry. The cash livelihood transfer had no comparable effect on poultry holdings or earnings. The larger effects of the poultry transfer could be

consistent with higher reported receipt rates of the poultry vis-à-vis the cash transfer, though administrative records suggest that rates of receipt were equally high for both transfers. Nonpoor households exposed to base L* programming also increase participation in poultry production, plausibly demonstrating a spillover effect of the poultry transfers to poor households. Also notable is an increase in savings, particularly for women, through savings in VESAs, across all three treatment arms and for poor and nonpoor households. The SPIR program had only limited effects on the probability of women engaging in labor outside their farms, with effects spread unevenly among poor households.

For nutrition outcomes, results show that the N* interventions, and to a lesser degree the N interventions, have increased exposure to health services including BCC, food demonstrations, and WASH. Screening of children for malnutrition has increased in N* kebeles as has provision of nutrition messages. The frequency of contacts between households and service providers is less than planned, but this may reflect continuation of start-up processes at the time of the midline. SPIR did have an impact on IYCF practices, particularly on the provision of colostrum and on breastfeeding. Results show a worrisome delay in the introduction of solid or semi-solid foods and an extremely low level of dietary diversity for young children. Finally, there is a surprising negative effect of N* interventions on measures of acute nutritional status (WHZ for T1 and T3; wasting for T3), indicating a need for more effort to improve delivery of services in kebeles operating the enhanced nutrition program.

SPIR showed promising effects on women's mental health by reducing depressive symptoms in some treatment arms, but there were no effects on depression in men. In the poor sample, women in T1 households who also received the poultry intervention had the largest reductions in depression, while these effects were largest for women in T3 households in the nonpoor sample. For intimate partner violence, we find no impacts of the SPIR intervention on these outcomes for either the full sample or subsample of non-poor households. In the poor subsample, T2 leads to substantial decreases in physical violence and sexual violence. The effect on physical violence of T2 is mainly from the cash sub-intervention, while the effect on sexual violence is mainly from the poultry sub-intervention. Overall in L*, poultry transfers lead to significant reductions in sexual violence, and cash transfers led to significant reductions in marital control.

These findings indicate a number of areas in which the SPIR project is having substantial impacts and is improving the wellbeing of women and their families. In other areas, such as acute malnutrition for children under age 2, the report identifies modest negative effects that suggest the need for further attention to quality of delivery. These findings should be of use to the SPIR partners as they consider how to complete the current phase of the project over the next 18 months.

As this report was being finalized, the COVID-19 crisis interrupted all aspects of SPIR project functioning and threatened to interfere with delivery of the PSNP transfers. The IFPRI research team was about to field a brief follow-up survey following completion of the first 12-week group therapy sessions for women screened for depression. This follow-up survey had to be canceled. Plans for the endline survey, which was scheduled to start June 1, are likely to be postponed. Initially, many SPIR programming activities have been suspended while the team focused on implementing a double food transfer in March-April. When possible, SPIR partners hope to pivot their activities to those that are most relevant for COVID-19 response, such as those related to WASH activities, which is appropriate to the seriousness of this crisis. Although this suspension of activities and possible redirection of resources interrupts the experiment, a large number of activities have been completed under the experimental design, including in the six month period after the midline survey and before the COVID-19 crisis. When

it is safe and appropriate to send survey teams to the sample kebeles for the endline survey, we will learn whether the SPIR interventions helped bolster household resilience to the health and economic shock. This question will be an added focus of the endline survey and analysis.

References

- Adewuya AO, Ola BA, Afolabi OO. 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.
- Anderson, ML. 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry preschool, and early training projects. *Journal of the American Statistical Association* 103(484): 1481–1495.
- 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 (DFSAs): Baseline Report. Washington DC: International Food Policy Research Institute.
- Angelucci, Manuela. 2020. A New Hope: The Economic Impact of Mental Health Care in India. Presented at IFPRI, March 2020.
- Aizer, Anna. 2011. Poverty, Violence, and Health: The Impact of Domestic Violence During Pregnancy on Newborn Health. *The Journal of Human Resources* 46 (3): 518-538.
- Appadurai, A. 2004. The Capacity to Aspire. in *Culture and Public Action*, ed. by M. Walton and V. Rao. Stanford, CA: Stanford University Press.
- Banerjee A, 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, Raghavendra 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.
- 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.
- Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., ... Bhutta, Z. A. 2017. Nurturing care: promoting early childhood development. *The Lancet*, 389(10064), 91–102.
- Broaddus-Shea, E.T., Manohar, S., Thorne-Lyman, A.L., Bhandari, S., Nonyane, B.A., Winch, P.J. and West, K.P. 2020. Small-Scale Livestock Production in Nepal Is Directly Associated with Children's In-creased Intakes of Eggs and Dairy, But Not Meat. *Nutrients*, 12(1), p.252.
- 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*.
- 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.
- Dalton, Patricio S., Sayantan Ghosal and Anandi Mani. 2014. Poverty and Aspirations Failure. *The Economic Journal*. 126:165-188.
- Ellsberg, Mary, Henrica AFM Jansen, Lori Heise, Charlotte H Watts, and Claudia Garcia-Moreno. 2008. Intimate Partner Violence and Women's Physical and Mental Health in the 0 Multi-Country Study on Women's Health and Domestic Violence: An Observational Study. *The Lancet* 371(9619): 1165–72.
- 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 Shibire, 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.
- Genicot G and D Ray. 2017. Aspirations and Inequality. *Econometrica* 85(2): 489-519.

- 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.
- Heath R, M Hidrobo and S. Roy. Cash transfers, polygamy, and intimate partner violence: Experimental evidence from Mali. *Journal of Development Economics*. forthcoming.
- Herba, C. M., Glover, V., Ramchandani, P. G., and Rondon, M. B. 2016. Maternal depression and mental health in early childhood: an examination of underlying mechanisms in low-income and middle-income countries. *The Lancet Psychiatry*, 3(10), 983–992. [https://doi.org/10.1016/S2215-0366\(16\)30148-1](https://doi.org/10.1016/S2215-0366(16)30148-1)
- 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.
- Iannotti, L.L., Lutter, C.K., Stewart, C.P., Riofrío, C.A.G., Malo, C., Reinhart, G., Palacios, A., Karp, C., Chapnick, M., Cox, K. and Waters, W.F., 2017. Eggs in early complementary feeding and child growth: a randomized controlled trial. *Pediatrics*, 140(1), p.e20163459.
- Imbens, G. W. 2015. Matching Methods in Practice: Three Examples. *Journal of Human Resources*.
- Imbens, G. W. and Rubin, D. B. 2015. 'Causal inference: For statistics, social, and biomedical sciences an introduction' in *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*.
- Janzen, Sarah A., Nicholas Magnan, Sudhindra Sharma, William M. Thompson. 2017. Aspirations failure and formation in rural Nepal. *Journal of Economic Behavior & Organization* 139, 1-25.
- Kling, JR., Liebman, JB., and LF Katz. 2007. 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., 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](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.
- 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.
- Romano, JP and M Wolf. 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73(4), 1237–1282.
- 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.
- Stewart, C.P., Caswell, B., Iannotti, L., Lutter, C., Arnold, C.D., Chipatala, R., Prado, E.L. and Maleta, K., 2019. The effect of eggs on early child growth in rural Malawi: the Mazira Project randomized controlled trial. *The American journal of clinical nutrition*, 110(4), pp.1026-1033.
- 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 Midline Survey Questionnaire

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

Appendix B: Baseline Balance Within Subtreatment Arms

Table B.1a: Household demographics at baseline: Subsample of poor households by SPIR subtreatment arms

	(1) Household size	(2) Female- headed household	(3) Age of primary female	(4) Primary female is married (monogamous)	(5) Primary female has some education	(6) Age of primary male	(7) Primary male is married (monogamous)	(8) Primary male has some education
T1 x Poultry	0.303 (0.251)	-0.104** (0.049)	-0.020 (0.886)	0.086* (0.047)	-0.011 (0.034)	0.847 (1.239)	0.006 (0.006)	-0.011 (0.053)
T1 x Poultry x Aspirations ^a	0.289 (0.337)	0.064 (0.055)	1.743* (1.031)	-0.073 (0.050)	-0.042 (0.043)	1.794 (1.358)	-0.011 (0.009)	-0.020 (0.069)
T1 x Cash	0.065 (0.295)	0.004 (0.065)	0.945 (0.688)	-0.024 (0.063)	-0.090* (0.048)	-0.205 (0.953)	-0.012 (0.013)	-0.127** (0.052)
T1 x Cash x Aspirations ^a	-0.054 (0.310)	0.112 (0.107)	0.232 (1.062)	-0.028 (0.105)	0.076 (0.050)	1.232 (1.348)	0.017 (0.013)	0.147** (0.069)
T2 x Poultry	0.087 (0.151)	0.014 (0.069)	0.845 (0.917)	-0.027 (0.045)	0.029 (0.057)	-0.616 (0.784)	-0.008 (0.010)	-0.001 (0.063)
T2 x Poultry x Aspirations ^a	-0.081 (0.218)	-0.126 (0.081)	-0.839 (1.262)	0.092 (0.066)	-0.015 (0.077)	0.421 (1.335)	-0.002 (0.015)	-0.030 (0.088)
T2 x Cash	0.056 (0.174)	-0.043 (0.054)	1.369 (0.936)	-0.016 (0.054)	-0.048 (0.062)	1.416 (1.288)	-0.019 (0.014)	-0.069 (0.062)
T2 x Cash x Aspirations ^a	0.466 (0.310)	-0.147*** (0.056)	-0.823 (1.156)	0.148** (0.058)	0.046 (0.072)	-0.552 (1.608)	0.019 (0.013)	0.018 (0.072)
T3	-0.062 (0.139)	0.018 (0.038)	0.419 (0.538)	-0.031 (0.035)	-0.014 (0.033)	-0.047 (0.607)	-0.009 (0.007)	0.005 (0.038)
Linear combination: T1	0.243 (0.155)	-0.005 (0.040)	0.975* (0.524)	0.005 (0.038)	-0.043 (0.029)	1.083* (0.641)	-0.002 (0.005)	-0.038 (0.033)
Linear combination: T2	0.166 (0.136)	-0.080** (0.037)	0.696 (0.573)	0.037 (0.033)	-0.002 (0.036)	0.349 (0.674)	-0.009 (0.006)	-0.037 (0.037)
Linear combination: Poultry	0.194 (0.163)	-0.044 (0.046)	0.420 (0.693)	0.029 (0.037)	0.009 (0.038)	0.104 (0.796)	-0.001 (0.007)	-0.006 (0.045)
Linear combination: Cash	0.060 (0.186)	-0.019 (0.046)	1.150* (0.630)	-0.020 (0.045)	-0.069 (0.042)	0.591 (0.844)	-0.015 (0.010)	-0.098** (0.044)

Test: T1 x Poultry x Aspirations = T4	0.049**	0.042**	0.394	0.110	0.423	0.108	0.656	0.679
Test: T1 x Cash x Aspirations = T4	0.828	0.460	0.094*	0.535	0.088*	0.794	0.476	0.061*
Test: T2 x Poultry x Aspirations = T4	0.658	0.616	0.445	0.939	0.614	0.511	0.258	0.831
Test: T2 x Cash x Aspirations = T4	0.214	0.034**	0.133	0.450	0.482	0.206	0.205	0.187
Mean of control (T4)	5.439	0.278	29.770	0.780	0.219	37.042	0.997	0.338
N	1,847	1,847	1,830	1,813	1,815	1,391	1,388	1,389

Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for

T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4).^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table B.1b: Household demographics at baseline: Subsample of nonpoor households

by SPIR sub treatment arms

	(1) Household size	(2) Female- headed household	(3) Age of primary female	(4) Primary female is married (monogamous)	(5) Primary Female has some education	(6) Age of primary male	(7) Primary male is married (monogamous)	(8) Primary male has some education
T1	0.054 (0.165)	0.061* (0.031)	-0.661 (0.588)	-0.011 (0.019)	0.037 (0.042)	-0.914 (0.839)	0.009* (0.005)	0.042 (0.054)
T1 x Aspirations ^a	-0.154 (0.234)	-0.005 (0.047)	-0.139 (1.063)	-0.004 (0.026)	0.030 (0.051)	1.054 (1.203)	-0.006 (0.008)	0.027 (0.063)
T2	0.188 (0.195)	0.043 (0.038)	0.684 (0.693)	-0.024 (0.025)	0.052 (0.043)	0.024 (0.973)	-0.004 (0.009)	0.029 (0.058)
T2 x Aspirations ^a	-0.454* (0.243)	0.028 (0.057)	-1.300* (0.767)	-0.057 (0.052)	-0.003 (0.055)	-0.300 (1.204)	0.001 (0.011)	0.034 (0.069)
T3	0.000 (0.163)	0.015 (0.026)	0.110 (0.583)	-0.028 (0.021)	0.019 (0.028)	0.246 (0.645)	0.003 (0.006)	0.022 (0.042)
Linear combination: T1	-0.022 (0.163)	0.058* (0.030)	-0.730 (0.659)	-0.013 (0.018)	0.052 (0.033)	-0.387 (0.788)	0.006 (0.006)	0.055 (0.043)
Linear combination: T2	-0.034 (0.161)	0.057* (0.031)	0.046 (0.523)	-0.052** (0.026)	0.051 (0.034)	-0.124 (0.774)	-0.004 (0.007)	0.046 (0.045)
Mean of control (T4)	6.268	0.056	31.089	0.968	0.161	39.148	0.990	0.313
N	1,373	1,372	1,348	1,340	1,339	1,287	1,284	1,285

Notes: Estimates from the ‘non-poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

**Table B.2a: Main livelihood outcomes at baseline: Subsample of poor households
by SPIR sub treatment arms**

	(1) Household reports owning poultry	(2) Household reports owning livestock	(3) Male reports taking a productive loan	(4) Primary female's main occupation agriculture	(5) Primary male's main occupation agriculture	(6) Female: aspires secondary or higher education for eldest child	(7) Male: aspires secondary or higher education for eldest child
T1 x Poultry	0.038 (0.071)	0.067 (0.069)	-0.007 (0.022)	0.019 (0.050)	0.014 (0.070)	-0.059 (0.055)	-0.006 (0.047)
T1 x Poultry x Aspirations ^a	-0.022 (0.087)	-0.025 (0.079)	0.027 (0.050)	0.030 (0.072)	0.083 (0.074)	-0.005 (0.075)	-0.037 (0.087)
T1 x Cash	0.050 (0.074)	0.012 (0.064)	0.015 (0.039)	0.038 (0.045)	-0.046 (0.105)	-0.020 (0.083)	0.008 (0.076)
T1 x Cash x Aspirations ^a	-0.073 (0.089)	-0.071 (0.082)	-0.011 (0.051)	-0.032 (0.060)	0.101 (0.111)	0.006 (0.089)	-0.013 (0.096)
T2 x Poultry	0.080 (0.063)	0.056 (0.041)	0.025 (0.059)	-0.012 (0.044)	-0.100 (0.064)	0.104* (0.053)	0.059 (0.049)
T2 x Poultry x Aspirations ^a	-0.042 (0.082)	0.014 (0.062)	-0.015 (0.070)	0.013 (0.062)	0.088 (0.073)	-0.244*** (0.070)	-0.041 (0.059)
T2 x Cash	0.055 (0.057)	0.060 (0.050)	-0.018 (0.038)	0.002 (0.033)	0.090* (0.050)	-0.085 (0.073)	-0.047 (0.064)
T2 x Cash x Aspirations ^a	-0.026 (0.079)	-0.010 (0.059)	0.021 (0.045)	0.016 (0.038)	0.039 (0.056)	0.028 (0.075)	0.048 (0.088)
T3	0.041 (0.033)	0.024 (0.034)	0.002 (0.024)	0.020 (0.026)	0.036 (0.045)	0.003 (0.044)	0.027 (0.043)
Linear combination: T1	0.020 (0.036)	0.015 (0.038)	0.008 (0.025)	0.029 (0.031)	0.030 (0.046)	-0.039 (0.042)	-0.012 (0.044)
Linear combination: T2	0.051 (0.034)	0.059* (0.033)	0.006 (0.027)	0.003 (0.027)	0.024 (0.040)	-0.042 (0.041)	0.009 (0.041)
Linear combination: Poultry	0.060 (0.049)	0.061 (0.044)	0.009 (0.034)	0.004 (0.036)	-0.044 (0.052)	0.024 (0.044)	0.027 (0.040)
Linear combination: Cash	0.052 (0.048)	0.036 (0.045)	-0.001 (0.030)	0.021 (0.031)	0.021 (0.063)	-0.052 (0.059)	-0.019 (0.054)
Test: T1 x Poultry x Aspirations = T4	0.556	0.262	0.936	0.473	0.426	0.206	0.685
Test: T1 x Cash x Aspirations = T4	0.634	0.814	0.717	0.448	0.869	0.767	0.950

Test: T2 x Poultry x Aspirations = T4	0.170	0.109	0.635	0.844	0.171	0.588	0.279
Test: T2 x Cash x Aspirations = T4	0.290	0.164	0.711	0.784	0.019**	0.176	0.556
Mean of control (T4)	0.348	0.742	0.082	0.153	0.730	0.577	0.571
N	1,847	1,847	1,379	1,814	1,388	1,312	1,321

Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for

T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. * $p < 0.1$ **

$p < 0.05$ *** $p < 0.01$

Table B.2b: Main livelihood outcomes at baseline: Subsample of nonpoor households

by SPIR sub treatment arms

	(1) Household reports owning poultry	(2) Household reports owning livestock	(3) Male reports taking a productive loan	(4) Primary female's main occupation agriculture	(5) Primary male's main occupation agriculture	(6) Female: aspires secondary or higher education for eldest child	(7) Male: aspires secondary or higher education for eldest child
T1	-0.074 (0.055)	-0.032 (0.030)	-0.018 (0.025)	0.033 (0.044)	-0.015 (0.062)	-0.062 (0.064)	-0.029 (0.057)
T1 x Aspirations ^a	0.060 (0.060)	-0.025 (0.037)	0.059* (0.035)	-0.052 (0.046)	-0.041 (0.063)	0.038 (0.069)	0.004 (0.065)
T2	-0.066 (0.054)	-0.032 (0.025)	0.087** (0.037)	0.048 (0.043)	0.053 (0.049)	-0.001 (0.053)	0.024 (0.053)
T2 x Aspirations ^a	0.024 (0.062)	-0.012 (0.039)	-0.038 (0.048)	-0.018 (0.050)	-0.141* (0.073)	0.021 (0.064)	-0.032 (0.060)
T3	0.003 (0.044)	-0.031* (0.016)	0.026 (0.027)	-0.003 (0.031)	0.016 (0.042)	0.006 (0.041)	-0.004 (0.036)
Linear combination: T1	-0.044 (0.044)	-0.044** (0.022)	0.011 (0.027)	0.007 (0.033)	-0.036 (0.045)	-0.043 (0.043)	-0.027 (0.041)
Linear combination: T2	-0.054 (0.045)	-0.038* (0.020)	0.069** (0.031)	0.039 (0.033)	-0.016 (0.046)	0.009 (0.040)	0.008 (0.039)
Mean of control (T4)	0.511	0.972	0.113	0.135	0.797	0.553	0.549
N	1,373	1,373	1,287	1,340	1,285	1,216	1,227

Notes: Estimates from the 'non-poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

**Table B.3a: Main nutrition outcomes at baseline: Subsample of poor households
by SPIR sub treatment arms**

	(1) Number of food groups (of 10) female consumed over previous day	(2) Number of food groups (of 8) index child consumed over previous day	(3) Weight-for- age z-score (WAZ)	(4) Weight-for- height z-score (WHZ)	(5) Female's IYCF knowledge score (percent)	(6) Male's IYCF knowledge score (percent)
T1 x Poultry	0.025 (0.094)	0.049 (0.081)	0.123 (0.162)	0.022 (0.180)	-0.906 (2.548)	1.772 (2.090)
T1 x Poultry x Aspirations ^a	-0.283* (0.149)	-0.297*** (0.087)	-0.127 (0.206)	0.129 (0.315)	-1.349 (2.857)	-0.541 (2.665)
T1 x Cash	0.135 (0.154)	-0.114 (0.081)	0.014 (0.164)	-0.147 (0.206)	-0.181 (2.067)	0.811 (2.971)
T1 x Cash x Aspirations ^a	-0.106 (0.196)	-0.048 (0.102)	0.290 (0.236)	-0.012 (0.254)	0.418 (3.043)	-0.288 (3.500)
T2 x Poultry	-0.197 (0.126)	-0.110 (0.092)	-0.054 (0.186)	-0.301 (0.196)	-0.977 (2.746)	1.980 (2.080)
T2 x Poultry x Aspirations ^a	0.216 (0.161)	0.069 (0.108)	-0.044 (0.250)	0.045 (0.235)	0.582 (3.526)	1.767 (2.618)
T2 x Cash	-0.200 (0.124)	0.029 (0.102)	-0.297* (0.151)	-0.169 (0.258)	-0.781 (2.348)	1.725 (1.625)
T2 x Cash x Aspirations ^a	-0.040 (0.174)	-0.103 (0.116)	-0.008 (0.192)	-0.248 (0.290)	-8.340* (4.486)	-5.636 (3.823)
T3	-0.023 (0.082)	-0.045 (0.078)	-0.078 (0.114)	-0.195 (0.139)	-1.915 (1.677)	1.565 (1.511)
Linear combination: T1	-0.019 (0.083)	-0.122** (0.060)	0.108 (0.117)	-0.033 (0.154)	-0.782 (1.575)	1.075 (1.435)
Linear combination: T2	-0.155* (0.081)	-0.050 (0.066)	-0.186 (0.121)	-0.286* (0.152)	-2.777 (1.836)	0.909 (1.412)
Linear combination: Poultry	-0.088 (0.089)	-0.032 (0.071)	0.033 (0.140)	-0.142 (0.156)	-0.945 (2.035)	1.883 (1.619)
Linear combination: Cash	-0.030 (0.106)	-0.044 (0.074)	-0.139 (0.126)	-0.157 (0.182)	-0.475 (1.761)	1.258 (1.810)
Test: T1 x Poultry x Aspirations = T4	0.404	0.431	0.565	0.710	0.496	0.354
Test: T1 x Cash x Aspirations = T4	0.404	0.060*	0.429	0.375	0.981	0.742

Test: T2 x Poultry x Aspirations = T4	0.212	0.263	0.660	0.091*	0.724	0.127
Test: T2 x Cash x Aspirations = T4	0.037**	0.955	0.022**	0.224	0.111	0.951
Mean of control (T4)	1.933	1.673	-1.091	-0.363	53.075	47.822
N	1,837	1,837	1,803	1,758	1,847	1,394

Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for

T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4).^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. * $p < 0.1$ **

$p < 0.05$ *** $p < 0.01$

Table B.3b: Main nutrition outcomes at baseline: Subsample of nonpoor households

by SPIR sub treatment arms

	(1) Number of food groups (of 10) female consumed over previous day	(2) Number of food groups (of 8) index child consumed over previous day	(3) Weight-for- age z-score (WAZ)	(4) Weight-for-height z-score (WHZ)	(5) Female's IYCF knowledge score (percent)	(6) Male's IYCF knowledge score (percent)
T1	0.221 (0.146)	-0.134 (0.086)	-0.073 (0.128)	-0.125 (0.144)	2.754 (2.202)	3.955** (1.962)
T1 x Aspirations ^a	-0.175 (0.146)	0.083 (0.083)	0.067 (0.133)	-0.073 (0.184)	-0.469 (2.844)	-1.697 (2.260)
T2	-0.051 (0.133)	-0.074 (0.096)	-0.207 (0.137)	-0.069 (0.175)	0.710 (2.015)	0.829 (1.761)
T2 x Aspirations ^a	0.055 (0.132)	0.115 (0.107)	-0.010 (0.129)	-0.279 (0.196)	-3.506 (3.316)	0.569 (2.241)
T3	0.109 (0.127)	-0.055 (0.070)	-0.081 (0.128)	-0.239* (0.138)	-1.826 (1.667)	0.526 (1.569)
Linear combination: T1	0.134 (0.118)	-0.092 (0.066)	-0.039 (0.120)	-0.161 (0.135)	2.520 (1.865)	3.107* (1.617)
Linear combination: T2	-0.024 (0.114)	-0.018 (0.075)	-0.212* (0.121)	-0.206 (0.142)	-1.010 (1.937)	1.108 (1.627)
Mean of control (T4)	2.171	1.776	-1.075	-0.372	50.757	46.154
N	1,365	1,365	1,341	1,313	1,373	1,289

Notes: Estimates from the 'non-poor' sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table B.4a: Main depression outcomes at baseline: Subsample of poor households

by SPIR sub treatment arms

	(1) Female's depression severity score (PHQ-9)	(2) Female's severity score is 8 or higher	(3) Female's severity score is 10 or higher	(4) Male's depression severity score (PHQ-9)	(5) Male's severity score is 8 or higher	(6) Male's severity score is 10 or higher
T1 x Poultry	0.763 (0.546)	0.010 (0.042)	0.024 (0.036)	0.337 (0.403)	0.028 (0.044)	0.011 (0.031)
T1 x Poultry x Aspirations ^a	-0.271 (0.693)	0.021 (0.066)	-0.023 (0.042)	1.021 (0.687)	0.077 (0.062)	0.076 (0.052)
T1 x Cash	-0.355 (0.707)	-0.055 (0.068)	-0.029 (0.043)	-0.765 (0.692)	-0.086* (0.046)	-0.025 (0.035)
T1 x Cash x Aspirations ^a	1.052 (0.791)	0.085 (0.072)	0.022 (0.047)	2.275*** (0.793)	0.193*** (0.060)	0.096* (0.056)
T2 x Poultry	-1.027** (0.469)	-0.101*** (0.031)	-0.057** (0.028)	-0.722* (0.370)	-0.033 (0.027)	-0.042 (0.027)
T2 x Poultry x Aspirations ^a	0.071 (0.717)	-0.013 (0.043)	-0.002 (0.036)	0.520 (0.885)	0.026 (0.069)	0.035 (0.050)
T2 x Cash	-0.615 (0.414)	-0.109*** (0.034)	-0.052** (0.023)	0.215 (0.473)	-0.009 (0.034)	-0.006 (0.030)
T2 x Cash x Aspirations ^a	1.410* (0.801)	0.124* (0.069)	0.080 (0.058)	0.907 (0.862)	0.092 (0.073)	0.073 (0.061)
T3	-0.128 (0.393)	-0.037 (0.031)	-0.004 (0.024)	-0.231 (0.370)	-0.010 (0.031)	-0.010 (0.024)
Linear combination: T1	0.389 (0.399)	0.004 (0.034)	-0.003 (0.025)	0.608 (0.371)	0.039 (0.031)	0.036 (0.025)
Linear combination: T2	-0.463 (0.406)	-0.078** (0.031)	-0.035 (0.026)	0.087 (0.403)	0.007 (0.034)	0.002 (0.026)
Linear combination: Poultry	-0.146 (0.416)	-0.046 (0.030)	-0.017 (0.026)	-0.201 (0.329)	-0.003 (0.030)	-0.016 (0.024)
Linear combination: Cash	-0.482 (0.465)	-0.081* (0.042)	-0.040 (0.028)	-0.282 (0.458)	-0.048 (0.032)	-0.016 (0.026)
Test: T1 x Poultry x Aspirations = T4	0.145	0.649	0.594	0.084*	0.148	0.200
Test: T1 x Cash x Aspirations = T4	0.990	0.595	0.519	0.982	0.550	0.806
Test: T2 x Poultry x Aspirations = T4	0.025**	0.001***	0.029**	0.188	0.451	0.251
Test: T2 x Cash x Aspirations = T4	0.711	0.052*	0.363	0.262	0.574	0.571
Mean of control (T4)	3.146	0.171	0.091	2.949	0.132	0.075

N 1,792 1,792 1,792 1,387 1,387 1,387

Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for

T1xPoultry, T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table B.4b: Main depression outcomes at baseline: Subsample of nonpoor households by SPIR sub treatment arms

	(1) Female's depression severity score (PHQ-9)	(2) Female's severity score is 8 or higher	(3) Female's severity score is 10 or higher	(4) Male's depression severity score (PHQ-9)	(5) Male's severity score is 8 or higher	(6) Male's severity score is 10 or higher
T1	-0.420 (0.452)	0.006 (0.036)	-0.029 (0.029)	-0.856** (0.425)	-0.034 (0.029)	-0.040 (0.026)
T1 x Aspirations ^a	0.907* (0.472)	0.051 (0.037)	0.047 (0.032)	1.308*** (0.426)	0.048 (0.035)	0.047 (0.031)
T2	-0.373 (0.405)	-0.003 (0.032)	-0.031 (0.029)	-0.515 (0.388)	-0.060* (0.031)	-0.044* (0.023)
T2 x Aspirations ^a	-0.180 (0.499)	-0.032 (0.040)	-0.000 (0.034)	0.009 (0.466)	-0.020 (0.035)	0.010 (0.024)
T3	-0.257 (0.369)	0.019 (0.032)	-0.005 (0.025)	-0.120 (0.406)	-0.023 (0.033)	-0.002 (0.025)
Linear combination: T1	0.033 (0.364)	0.032 (0.028)	-0.005 (0.024)	-0.202 (0.366)	-0.010 (0.028)	-0.016 (0.024)
Linear combination: T2	-0.461 (0.365)	-0.018 (0.029)	-0.031 (0.026)	-0.510 (0.390)	-0.070** (0.028)	-0.039* (0.023)
Mean of control (T4)	2.891	0.103	0.071	2.910	0.131	0.074
N	1,333	1,333	1,333	1,287	1,287	1,287

Notes: Estimates from the ‘non-poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table B.5a: Main IPV outcomes at baseline: Subsample of poor households by SPIR sub treatment arms

	(1) Experienced emotional violence in the past 13 months	(2) Experienced physical violence in the past 13 months	(3) Experienced sexual violence in the past 13 months
T1 x Poultry	-0.001 (0.068)	-0.025 (0.027)	-0.021 (0.033)
T1 x Poultry x Aspirations ^a	-0.053 (0.072)	0.028 (0.027)	0.062 (0.041)
T1 x Cash	-0.072 (0.056)	-0.011 (0.042)	0.021 (0.044)
T1 x Cash x Aspirations ^a	0.083 (0.061)	0.025 (0.049)	-0.037 (0.043)
T2 x Poultry	-0.079** (0.039)	-0.027 (0.030)	0.032 (0.032)
T2 x Poultry x Aspirations ^a	0.005 (0.054)	-0.018 (0.035)	-0.025 (0.036)
T2 x Cash	0.050 (0.054)	0.017 (0.032)	-0.024 (0.017)
T2 x Cash x Aspirations ^a	0.057 (0.079)	0.084 (0.056)	0.063 (0.043)
T3	0.038 (0.040)	0.026 (0.030)	-0.012 (0.017)
Linear combination: T1	-0.030 (0.034)	-0.005 (0.025)	0.007 (0.019)
Linear combination: T2	0.000 (0.036)	0.011 (0.026)	0.014 (0.018)
Linear combination: Poultry	-0.040 (0.042)	-0.026 (0.025)	0.006 (0.024)
Linear combination: Cash	-0.012 (0.043)	0.003 (0.030)	-0.001 (0.025)
Test: T1 x Poultry x Aspirations = T4	0.701	0.552	0.981
Test: T1 x Cash x Aspirations = T4	0.294	0.923	0.784

Test: T2 x Poultry x Aspirations = T4	0.034**	0.224	0.318
Test: T2 x Cash x Aspirations = T4	0.140	0.159	0.859
Mean of control (T4)	0.139	0.078	0.029
N	1,051	1,051	1,051

* Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband.
Notes: Estimates from the ‘poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where the 10 poorest households out of 18 in each kebele are classified as poor. Coefficients for T1xPoultry,

T1xCash, T2xPoultry and T2xCash are impacts of the sub-randomization of the treatment arm with poultry or cash on the outcome relative to control (T4). ^a Coefficients for T1xPoultryxAspirations, T1xCashxAspirations, T2xPoultryxAspirations and T2xCashxAspirations are the additional effect of adding aspirations to that treatment arm sub-randomized with poultry or cash. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm or of poultry or cash. Tests report p-values on the null hypothesis of equal impacts between treatment arms. All models control for woreda level fixed effects. * p<0.1 ** p<0.05 *** p<0.01

Table B.5b: Main IPV outcomes at baseline: Subsample of poor households by SPIR sub-treatment arms

	(1) Experienced emotional violence in the past 13 months	(2) Experienced physical violence in the past 13 months	(3) Experienced sexual violence in the past 13 months
T1	-0.018 (0.042)	0.030 (0.035)	0.012 (0.029)
T1 x Aspirations ^a	-0.034 (0.038)	-0.023 (0.038)	-0.019 (0.029)
T2	0.003 (0.037)	0.040 (0.030)	-0.004 (0.026)
T2 x Aspirations ^a	0.019 (0.047)	0.012 (0.043)	-0.017 (0.033)
T3	-0.015 (0.031)	-0.005 (0.022)	-0.007 (0.021)
Linear combination: T1	-0.035 (0.030)	0.019 (0.026)	0.003 (0.022)
Linear combination: T2	0.012 (0.032)	0.046* (0.026)	-0.012 (0.024)
Mean of control (T4)	0.130	0.053	0.053
N	1,041	1,041	1,041

Notes: Estimates from the ‘non-poor’ sample of households in each kebele from the DFSA SPIR midline survey sample. Sample of non-poor households is determined by ranking households within kebeles based on a land and asset index constructed at baseline, where 8 least poor households out of 18 in each kebele are classified as non-poor. Coefficients

for T1, T2 and T3 are impacts of the treatment arm without aspirations on the outcome relative to control (T4). ^a Coefficients for T1xAspirations and T2xAspirations are the additional effect of adding aspirations to that treatment arm. Standard errors (in parentheses) are clustered at the kebele level. Linear combinations report the pooled impact estimate of that treatment arm. Tests report p-values on the null hypothesis of equal impacts between treatment arms. * p<0.1 ** p<0.05 *** p<0.01