Do ultra-poor graduation programs build women's resilience to weather shocks? Evidence from rural Ethiopia

Hirvonen, Kalle (IFPRI); Gilligan, Daniel O. (IFPRI); Leight, Jessica (IFPRI); Tambet, Heleene (IFPRI); Villa, Victor (CIAT)

Abstract

We assess whether layering an ultra-poor graduation program on Ethiopia's Productive Safety Net Programme (PSNP) protects women and their households from localized weather shocks. Using panel data from a large cluster-randomized trial matched with detailed weather records, we identify community-level exposure to dryness shocks. Women who are exposed to dryness shocks experience deteriorating nutritional status (in conjunction with declining household food security and assets) and a significant increase in the risk of intimate partner violence. However, access to the graduation program fully buffers women from any increase in intimate partner violence, and partially buffers women and their households against other adverse effects. Further analysis suggests that improved household savings likely drives the program's protective impact.

Key words: resilience; weather shocks; climate change; graduation model programs; social safety nets

JEL Codes: I38, O12, O15

Acknowledgements: This research was funded through the CGIAR Research Program on Harnessing Gender and Social Equality for Resilience in Agrifood Systems (HER+). The original trial was funded by the United States Agency for International Development (USAID) through a subaward to World Vision under Cooperative Agreement No AID-FFP-A-16-00008 (2016-2021). We thank World Vision, CARE, and ORDA for their contributions to the original impact evaluation. The ethical approvals for the trial were obtained from Hawassa University in Ethiopia and from IFPRI. We thank the editor, Lucie Gadenne, and three anonymous reviewers for excellent feedback. We have also benefitted from discussions with and comments from Jenny Aker, Leah Bevis, Martina Björkman Nyqvist, Ben D'Exelle, Taryn Dinkelman, John Giles, Melissa Hidrobo, Namrata Kala, David McKenzie, Michael Mulford, Shalini Roy, Carly Trachtman, Kelsey Wright, and participants at the IPA and the Global Poverty Action Lab's 2024 Annual Researcher Gathering, 2023 DIME-KDI conference, 2023 Nordic Conference in Development Economics in Gothenburg, and UNU-WIDER brownbag seminar for useful comments. The opinions expressed here belong to the authors and do not necessarily reflect those of the funders or the CGIAR. An earlier version of this paper circulated under the title "Do ultra-poor graduation programs build resilience against droughts? Evidence from rural Ethiopia."

1. Introduction

Global warming is increasing the frequency and severity of harmful weather events such as droughts, floods, and tropical cyclones (Seneviratne et al., 2021). A large literature shows that these shocks remain largely uninsured across low- and middle-income countries,[1] particularly where formal insurance markets and banking services are limited or absent (Collins et al., 2009). In many such settings, gendered patterns of resource access, labour allocation, and health-related vulnerabilities shape how shocks are absorbed within households and render women more vulnerable to their adverse consequences (Björkman-Nyqvist, 2013; Corno et al., 2020; Dercon and Krishnan, 2000; van Daalen et al., 2020). Recent research links adverse weather events to heightened risk of intimate partner violence, a form of violence experienced almost exclusively by women (Abiona and Koppensteiner, 2018; Dehingia et al., 2024; Díaz and Saldarriaga, 2023; Epstein et al., 2020; Hertzog et al., 2025; Wang et al., 2025). As climate change intensifies, a central policy question is how to design mechanisms that reduce the burdens women disproportionately shoulder during weather shocks (World Bank, 2015) – an area where evidence remains limited despite extensive research on the consequences of these shocks in settings with weak formal risk-management institutions.

In such contexts, poor women and households depend heavily on their own strategies to manage risk – accumulating liquid savings or other assets, or diversifying their income sources – whereas safety net programs are designed to complement these self-insurance strategies by providing a stable source of consumption support that is not tied to local weather conditions. These programs have become increasingly common in low- and middle-income countries (Beegle et al., 2018; World Bank, 2018). Cash transfers can be effective in reducing poverty and improving food security (Andrews et al., 2018; Crosta et al., 2023; Hidrobo et al., 2018; Leight et al., 2024), and can also play a role in reducing intimate partner violence (Buller et al., 2018) and increasing women's empowerment (Hidrobo et al., 2024; Peterman et al., 2020). However, cash transfers rarely have long-term effects (Leight et al., 2024), and this has prompted interest in policy strategies that tackle deeper constraints on households' economic choices. Graduation model interventions, developed to help ultra-poor households overcome structural constraints that keep

[1] Seminal empirical papers in this area include, but are not limited to: Fafchamps et al. (1998), Dercon (2004), Alderman et al. (2006), and Maccini and Yang (2009).

them in low-return activities, address capital and skills barriers while promoting saving through group-based platforms. These program components can also strengthen the same self-insurance mechanisms that shape how households cope with weather shocks, making graduation programs a potentially important complement to standard safety nets in shock-prone settings. While multi-faceted graduation model programs show promise in improving livelihood outcomes and in some cases, women's welfare (Balboni et al., 2022; Bandiera et al., 2017; Banerjee et al., 2015; Bossuroy et al., 2022; Brune et al., 2022), evidence around their ability to protect women and their households from adverse effects of shocks remains limited (Kala et al., 2023).

We explore this question in the context of Ethiopia's Productive Safety Net Programme (PSNP), a large-scale social protection program in a country with a long history of devastating droughts (De Waal, 2017). Our analysis draws on data from a multi-arm cluster randomized controlled trial (RCT) that evaluated the Strengthen PSNP4 Institutions and Resilience (SPIR) program; a five-year, gender-sensitive graduation model including multiple livelihoods interventions layered onto the PSNP. Compared to well-known BRAC-designed graduation models evaluated in Banerjee et al. (2015) and Bandiera et al. (2017), the SPIR livelihoods interventions were more limited in scope and involved smaller transfer amounts. Prior evidence showed that SPIR increased savings (+30 percentage points) and access to credit (+8-10 percentage points), effects that are comparable to BRAC-implemented programs (Bandiera et al., 2017; Banerjee et al., 2015), but had no detectable impact on consumption or poverty (Leight et al., 2026).

Our identification relies on exogenous variation in program exposure induced by the cluster RCT coupled with both longitudinal and cross-sectional variation in relative dryness over the three survey rounds conducted between 2016 and 2021.[2] Our primary measure of dryness is constructed using the Standardised Precipitation-Evapotranspiration Index (SPEI) [3] developed by Vicente-Serrano et al. (2010). The indicator captures relative dryness experienced during the cropping season as any negative deviation relative to the long-term mean in the same locality. This allows

[2] Related work in this area has primarily relied on cross-sectional variation in weather or other type of shocks (Emerick et al., 2016; Karlan et al., 2017; Macours et al., 2022; Pomeranz and Kast, 2024), but given high spatial correlation in weather shocks, an analysis relying solely on cross-sectional variation may confront limited power. Accordingly, we argue that exploiting substantial longitudinal variation is a meaningful strength of our empirical design.
[3] We follow a number of recent papers in using the SPEI as a primary measure to capture weather shocks, including Couttenier and Soubeyran (2014); Harari and La Ferrara (2018); Maystadt et al. (2015); Webb (2023).

us to focus on moderate fluctuations in dryness that represent typical variation across cropping seasons, rather than rare extreme weather anomalies.

We find that these moderate changes in relative dryness during the main cropping season lead to statistically significant and meaningful fluctuations in women's welfare and in broader household well-being. To quantify the effect of an average dryness shock during the study period, we evaluate the effects of a 0.25 standard deviation (SD) increase in relative dryness defined relative to the locality-specific long-term mean; this corresponds roughly to the median increase in relative dryness observed in dryness-affected areas, 0.23 SD. Among households receiving only the PSNP (the control arm), a dryness shock of this magnitude leads to modest declines in women's dietary diversity and BMI, and a substantially increased risk of intimate partner violence: a 0.25 SD increase in relative dryness increases IPV by 21 percent, a magnitude broadly consistent with previous literature (Abiona and Koppensteiner, 2018; Díaz and Saldarriaga, 2023; Epstein et al., 2020). Dryness shocks also generate a marked rise in stress among male household members and erode women's decision-making authority over productive income, consistent with the hypothesis that financial strain triggered by weather shocks can intensify within-household conflict and weaken women's bargaining position (Aizer, 2010; Fox et al., 2002). At the household level, the same increase in relative dryness leads to a 36 percent increase in self-reported food insecurity (the food gap), a 35 percent decrease in aggregated livestock holdings, and a five percent decline in consumption (using endline data only).

Strikingly, the corresponding effects of these dryness shocks are either partially or completely muted for women residing in clusters randomly assigned to receive SPIR in addition to the PSNP. Women living in clusters served by SPIR do not experience any decline in diet diversity or BMI linked to dryness, and they are similarly fully buffered from any shift in IPV. We also find that the rise in stress among male household members and the erosion of women's decision-making authority seen during dryness shocks in control communities are significantly smaller in treated clusters. Household-level outcomes show a similar pattern: the increase in the household food gap in dryness-affected clusters served by SPIR is roughly half the magnitude of the increase observed in dryness-affected clusters in the control arm, while the decline in livestock holdings is less than a third of the corresponding decline in dryness-affected control clusters and the decline in consumption is not statistically significant. These findings suggest that the livelihood interventions protected women and their households during dryness shocks, over and above the support received through the existing government safety net program, the PSNP.

Further analysis suggests little evidence of heterogeneity across the different SPIR treatment arms. Moreover, we find no indication that SPIR led households to adjust their farming practices or diversify into non-agricultural activities (though there was some shift toward livestock production). The primary mechanism for the effects described above appears to be a substantial relative increase in savings, attributable to the establishment of VESAs and observed consistently among all households served by SPIR (including those who did not receive one-time cash or poultry transfers). In the face of a dryness shock, treated households then use this enhanced stock of savings to mitigate the impact of the shock, allowing women (and other household members) to better sustain their consumption levels and reducing poverty-related stress that can trigger violence. Together, these results suggest that while extending the national safety net to incorporate a graduation program did not lead to higher levels of aggregate consumption or income, it facilitated the establishment of a buffer stock of savings that safeguarded women and their households during dryness shocks.

We contribute to several strands of literature. First, we bridge two strands of research on intimate partner violence. One rapidly expanding literature shows that adverse rainfall shocks raise the risk of IPV (Abiona and Koppensteiner, 2018; Dehingia et al., 2024; Díaz and Saldarriaga, 2023; Epstein et al., 2020; Hertzog et al., 2025; Wang et al., 2025), while another finds that positive income shocks significantly reduce IPV (Buller et al., 2018; Haushofer et al., 2019; Heath et al., 2020; Roy et al., 2019). We bring these strands together by showing that a gender-sensitive graduation program can fully offset the rise in IPV observed during weather shocks, potentially by weakening the widely discussed pathway through which financial strain increases IPV (Aizer, 2010; Fox et al., 2002). We also provide rare direct empirical evidence on the pathways linking weather shocks to IPV: dryness shocks raise stress levels among male household members and erode women's decision-making authority over productive income, and exposure to the SPIR program weakens both relationships.

Second, we add to the literature examining the consequences of adverse weather events and the emerging evidence on policy instruments that build household resilience. While existing work shows that safety net programs can partly soften the impact of weather shocks and natural disasters

(Adhvaryu et al., 2024; Asfaw et al., 2017; Christian et al., 2019; De Janvry et al., 2006; Hou, 2010; Knippenberg and Hoddinott, 2017; Premand and Stoeffler, 2020),[4] we show that a lighter-touch graduation model embedded within a safety net can contribute to buffer women from the adverse effects of dryness.[5] These results underscore that graduation-style interventions can strengthen self-insurance mechanisms that often fall outside the scope of standard safety net programs.

Our work also builds on Macours et al. (2022), who study a 'cash-plus' program that offered vocational training or a $200 investment grant to short-term safety net program beneficiaries in rural Nicaragua. As in their study, simple cash transfers (offered to the control arm via the PSNP itself) are insufficient to buffer households against weather shocks, while a bundled model of interventions effectively assists households in coping with shocks. The mechanisms, however, differ: in Nicaragua, complementary interventions – vocational training or productive grants – facilitated household diversification into non-agricultural activities, and thus protected them against the negative impacts of weather shocks on agricultural income. In our setting, diversification into non-agricultural activities is virtually non-existent: less than five percent of the households in the control arm report any non-agricultural business or regular wage employment at endline, and there is no evidence of any meaningful treatment effect on these outcomes (Leight et al., 2026). (There is some increase in livestock income, though it is not uniform across arms.) The primary mechanism driving increased resilience here appears to be buffer stock savings.

Overall, these findings suggest that even relatively light-touch interventions (without larger asset or cash transfers) can promote resilience within the context of an existing safety net program, and minimal off-farm income-generating opportunities. Previous literature has generally suggested that enhanced resilience, as well reduced intimate partner violence, are effects observed in conjunction with large treatment-induced increases on consumption, income and other economic

---

[4] Consistent with Knippenberg and Hoddinott (2017), evidence from the control households provided in this paper indicates that the PSNP is not able to fully insure poor households against droughts.

[5] Other active areas of policy experimentation relevant to buffering households against adverse weather effects include the promotion of climate-smart agriculture policies (e.g., drought/flood resistant crop varieties) (Emerick et al., 2016; Lipper et al., 2017), weather index insurance programs targeted to agricultural households (Carter et al., 2017; Jensen and Barrett, 2017; Karlan et al., 2014), anticipatory cash transfers to areas forecasted to experience weather shocks (Pople et al., 2021) and emergency loans to affected households (Lane, 2024).

outcomes.[6] By contrast, these findings suggest that even an intervention that has smaller effects on average can have substantial protective effects. This highlights an often-underrecognized dimension of livelihoods interventions that may not be intensive enough to lift large numbers of households out of poverty, but still have important benefits in terms of resilience.

2. Context: PSNP and SPIR

The Productive Safety Net Programme (PSNP)

This analysis draws on data from a trial conducted in the context of the PSNP, a safety net program launched in 2005 and designed to provide a more sustainable response mechanism in areas of Ethiopia that have been historically vulnerable to droughts, as opposed to recurring *ad hoc* emergency appeals for food aid and famine relief (Wiseman et al., 2010). Within the districts served by the PSNP, poor and chronically food-insecure households receive payments (in the form of cash and/or food) for six months in exchange for performing labor-intensive public works, while similarly poor and food-insecure households with limited labor capacity receive unconditional, direct transfers. Communities themselves select beneficiaries by applying a categorical targeting strategy. The program reaches eight million people, rendering it one of the largest safety net programs in Africa and, outside India, the largest public works program in the world (Beegle et al., 2018). PSNP is implemented by the government of Ethiopia, and largely funded by its international partners.

Evaluations based on quasi-experimental methods show that the PSNP has been successful in improving household food security and creating community assets (Hirvonen et al., 2022). However, PSNP households remain vulnerable to droughts, despite recovering from these shocks more rapidly when compared to poor non-PSNP households (Knippenberg and Hoddinott, 2017). In addition, there has been negligible exit from poverty for PSNP beneficiaries (Sabates-Wheeler et al., 2021).

[6] For reduction of partner violence, increased economic security is a major pathway through which cash or cash plus programs reduce IPV (Buller et al., 2018); and similarly, increased income is hypothesized as a major pathway through which social protection systems (on which SPIR is built) can enhance climate resilience (Hidrobo et al., 2026).

The SPIR program

The SPIR Development Food Security Activity (DFSA) in Ethiopia was a five-year program (2016–2021) providing complementary livelihood, nutrition, and gender activities intended to strengthen the PSNP. Funded by USAID and implemented by World Vision, CARE and ORDA in close collaboration with the Government of Ethiopia, SPIR was organized around a core set of livelihood and nutrition activities delivered to nearly 500,000 beneficiaries. The core livelihood activities (L) included the formation of Village Economic and Social Associations (VESAs), the primary platform supporting financial literacy training and the promotion of new income generation activities. Both women and men were encouraged to join VESAs. They were also used as a platform for the core nutrition activities (N), including behavior change communication (BCC).

SPIR also introduced enhanced models of livelihoods and nutrition activities. The enhanced livelihoods model (L*) included all core livelihood activities, supplemented with a lump-sum transfer (delivered as either cash or an equivalent value poultry package) and targeted to the poorest 10 out of 18 households in each kebele (sub-district).[7] The enhanced nutrition activities (N*) included a more targeted program of nutrition BCC and supplemental interventions targeting male engagement and mental health. These packages were combined into multisectoral graduation model programs and randomized into four treatment arms: T1: L*+N*; T2: L*+N; T3: L+N*; and T4: PSNP only, as summarized in Figure A1 in the appendix.

Average treatment effects

The impacts of the SPIR program on a set of pre-specified economic outcomes have been documented in Leight et al. (2026), and we summarize the key findings here. (A separate paper also documented that the intervention led to a shift toward intrahousehold gender equity in task allocation, but did not assess effects on intimate partner violence (Alderman et al., 2025).) The enhanced livelihoods interventions, which included asset transfers targeted at extremely poor households, led to moderate improvements in financial inclusion and increased cash income from livestock. However, households achieved only modest asset accumulation and some increases in

[7] Administratively, Ethiopia is divided into regions, zones, woredas (districts) and kebeles (sub-districts). Kebele is the lowest-level administrative unit and the community targeting of the PSNP as well as many key government services such as agricultural and health extension are organized and provided at this level.

income from livestock, and there were no statistically significant effects on consumption. Notably, there were few meaningful differences in outcomes between households that received poultry transfers and those that received an equivalent cash transfer.[8]

The most pronounced effects were observed in financial inclusion. Access to credit increased by 8–10 percentage points compared to a control mean of 45 percent, while the probability of reporting any savings rose by more than 30 percentage points from a baseline level of 40 percent. Moreover, both poultry and cash recipients experienced a 30 percent increase in past-year income from livestock. However, no significant changes were detected in total consumption or food security, a pattern that could be consistent with a low propensity to consume out of the marginal earnings or pronounced seasonality in the income from livestock.

Among extremely poor households who received only savings groups and financial literacy training but no poultry or cash transfers, the only measurable impact was an increase in reported savings. The same pattern emerged among less poor households who participated in savings groups and training across all treatment arms but received no additional transfers. In both cases, participation in the program increased the likelihood of reporting any savings by approximately 30 percentage points.

Importantly, there is no evidence that the SPIR program influenced participation in non-agricultural activities. Across both midline and endline rounds, engagement in non-agricultural employment remained minimal: only three percent of households reported operating a non-agricultural business (rising to six percent among less poor households), approximately four percent engaged in regular wage labor, and around 25 percent participated in casual wage labor. These low participation rates remained stable throughout the study period, with no shift toward non-farm economic activities.

[8] Asset accumulation was notably larger among households who received a cash transfer and were not exposed to intensive nutrition programming; households that were exposed to intensive nutrition programming showed reduced asset accumulation, and seemed to invest in child nutrition instead.

In this paper, we assess how these treatment effects evolve in response to a shock, and whether shifts in women's and households' economic welfare driven by SPIR lead to changes in the vulnerability to these shocks.[9]

3. Data

Surveys

The study took place in 13 woredas (districts) and 192 kebeles across the Amhara and Oromia regions of Ethiopia (Figure A2 in Appendix A). The randomization into treatment was stratified at the woreda level and the treatments were randomly assigned at the level of clusters, or kebeles. At the household level, eligibility criteria required being current PSNP beneficiaries, having at least one child aged 0-35 months at baseline, and having the mother or primary female caregiver of the same child also living in the household at baseline.

Three rounds of data were collected, with a baseline survey between February and April 2018, a midline survey between July and October 2019, and an endline survey originally planned for 2020 but delayed due to COVID-19 and undertaken 36 months after the baseline between February and April 2021. The baseline sample included 3,314 households with a child under three years of age, or just over 17 PSNP beneficiary households in each kebele. The midline survey achieved an overall sample of 3,220 of the original households, yielding an attrition rate of 2.8 percent. Of the 3,246 households eligible for the endline survey (after removing those who had permanently moved), 3,094 were able to be located and interviewed, leading to an attrition rate of 4.7 percent relative to the target sample. A large portion of the attrition (80 households) at the endline was due to conflict in northern Amhara (linked to the emerging conflict in Tigray), where four kebeles were rendered permanently inaccessible to the survey team.

Outcome variables

The variables of interest for this analysis include measures of well-being for both women and their households that reflect important dimensions of destitution among poor Ethiopian households

[9] The original SPIR RCT was pre-registered, but the pre-analysis plan did not specify the weather shock-focused analyses or the emphasis on heterogeneity by weather variability. Given that strengthening households' capacity to manage shocks was a core program objective, estimating whether SPIR buffered households during actual weather shocks is a natural extension of the registered evaluation. Nonetheless, the absence of a dedicated pre-analysis plan for this component should be considered a limitation.

(including nutrition, food security, assets and freedom from violence), and are plausibly responsive to dryness shocks. We generally focus on variables that were measured consistently in all three survey rounds, but report selected supplementary results for outcomes that were not collected in all survey rounds. All variables measured for adult women were captured for the primary female in the household, who was identified as the mother or primary caregiver of an index child aged 0-35 months, and we restrict to women aged 15-49.

We use three indicators to measure women's exposure to and welfare impacts of dryness shocks. First, we assess dietary diversity among women: in this context, this is one of the simplest indicators that can be employed to capture the quality of a person's diet.[10] Following the FAO and FHI 360 (2016) guidelines, we construct a diet diversity index that counts the number of food groups consumed by the primary female in the 24 hours prior to the interview.[11] At baseline, the average primary female in control households consumed food from only 2.1 out of 10 food groups. As a second measure of women's individual-level welfare, we use data collected on the height and weight of the primary female to construct body-mass index (BMI), computed by dividing weight in kilograms by the square of body height in meters. BMI is considered a reasonable proxy for adult health risks (Fogel, 1994; Waaler, 1984) and has been shown to be responsive to weather shocks (Dercon and Krishnan, 2000; Hoddinott and Kinsey, 2000). The mean BMI among primary females in control households at the baseline is 20.2, and 26 percent of women can be categorized as underweight (BMI<18.5).[12]

As a third individual-level variable, we also analyze women's recent experience of intimate partner violence, motivated in part by recent research examining the influence of weather shocks on the risk of IPV (Abiona and Koppensteiner, 2018; Dehingia et al., 2024; Díaz and Saldarriaga, 2023; Epstein et al., 2020; Hertzog et al., 2025; Wang et al., 2025). All three survey rounds contained an IPV module, administered to the primary female in accordance with the WHO protocol on ethical

[10] There is also some previous evidence that droughts reduce dietary diversity measured at the household level (Carpena, 2019).
[11] In this index foods with similar nutritional contents have been categorized into 10 food groups. The age restriction is imposed since the indicator has only been validated for women who are 15 to 49 years of age; after this restriction, the final data has 9,087 diet diversity observations across the three survey rounds, excluding 16 percent of observations.
[12] For the BMI analysis, we restrict the sample to women who are 15 to 49 years and exclude women with implausible BMI values (BMI below 15 or above 50). After these restrictions, we are left with 8,100 BMI observations across the three survey sounds, excluding 16 percent of observations.

guidelines for conducting research on IPV (WHO, 2016). As the module was administered only if the respondent reported living with her husband in the last 12 months and if she was alone or with a child less than 36 months at the time of the interview,[13] the sample size is smaller than for the other outcomes.[14] At baseline, 17 percent of the women in the control clusters reported to have experienced physical, sexual and/or emotional violence in the previous 12 months. While this pooled variable is our primary indicator relevant to IPV, we report our results separately for the three forms of IPV that were collected in the surveys (see Appendix B for definitions).

At the household level, we focus on indicators that are typically employed to analyze the effects of weather-related shocks on economic welfare: food security and assets (Carpena, 2019; Kazianga and Udry, 2006; Macours et al., 2022).[15] First, as a measure of household level food security, we use a construct called the "food gap". Employed as the main food security indicator in PSNP evaluations, the food gap is measured by asking survey participants to report the number of months, out of the preceding 12 months, that they had "problems satisfying the food needs of the household".[16] Knippenberg and Hoddinott (2017) show how a self-reported drought shock increases the food gap by 1.6 months among PSNP households. In our sample, the mean food gap in control households at the baseline was 2.1 months.

As a second measure of household economic welfare, we use an index of livestock assets. Livestock in this context has several purposes. First, large livestock (bulls, oxen) can be used as draft power during ploughing and threshing (Mekuriaw and Harris-Coble, 2021). Second, livestock products (e.g., dairy, eggs, meat) provide additional income to agrarian households. Third, as access to formal savings institutions remains highly limited in rural areas, livestock is an important form of savings that can be liquidated during negative income shocks (Fafchamps et al.,

[13] Very few eligible women refused to respond to the IPV questions.
[14] At the baseline, 2,136 women responded to the IPV module, at the midline 1,648 women, and at the endline 2,161 women. An error in the CAPI program in the midline survey led to the IPV questions being excluded from the survey administered to about 600 women. All women who reported violence were given the option to be referred to the Women's Affairs Committee in her district. Table B1 shows that SPIR treatment does not predict selection into the IPV sample.
[15] We also consider household per capita consumption based on standard household food and non-food consumption expenditure modules (Deaton and Grosh, 2000). However, these modules were only administered at baseline and endline, not at midline, and thus we cannot directly estimate our preferred specification.
[16] A month in which the household had "problems satisfying food needs" is defined as one where the household experienced hunger for five or more days.

1998). Our measure of livestock holdings is based on tropical livestock units (TLUs).[17] The average household in the control cluster owned 0.97 TLUs at the baseline, including on average 0.6 heads of cattle (bull, oxen, or cow), 0.3 pack animals (donkey or mule), 0.5 sheep/goat and 1.5 chickens.

Dryness indicator

The drought literature distinguishes between meteorological drought — a deficit in water availability relative to long-term norms — and socioeconomic drought, which captures the welfare consequences of such deficits (Van Loon, 2015). A severe meteorological drought does not translate into a socioeconomic drought if households can absorb the shock. Conversely, for households living near subsistence with limited access to insurance and credit markets, even moderate dryness can have serious welfare consequences. This distinction is relevant for our analysis, which studies a population where nearly two-thirds of households fall below the extreme poverty line.

Our primary dryness builds on the Standardised Precipitation-Evapotranspiration Index (SPEI) developed by Vicente-Serrano et al. (2010). The SPEI quantifies changes in climatic water balance over a given period considering the temporal difference between precipitation and potential evapotranspiration influenced by temperature. The SPEI has been widely used by economists to analyze socio-economic impacts of dryness in various settings (e.g., Couttenier and Soubeyran, 2014; Harari and La Ferrara, 2018; Maystadt et al., 2015; Webb, 2023) and has been shown to be a more accurate predictor of crop yield than other similar indicators (Vicente-Serrano et al., 2012).[18] We will subsequently demonstrate that our primary findings are robust to alternative dryness definitions.

The SPEI is a standardized variable with a zero mean and standard deviation of one, quantifying the water balance in terms of standard deviations from the long-run average (here 1990-2020) in the locality, and thus a negative SPEI value indicates a deficit in water availability relative to the

---

[17] The standard measure of a TLU is one cattle with a body weight of 250 kg (Jahnke, 1982). TLU are expressed as ratios relative to this standard unit, where the ratios are based on metabolic weights. So, for example, six sheep have the same energy requirements as one cattle and so six sheep translate into one TLU. Consequently, 1 sheep equals 0.15 TLU while 1 chicken is 0.01 TLU.

[18] We computed the SPEI for our study clusters using a user-written R-routine (Beguería and Vicente-Serrano, 2017). As inputs we used monthly precipitation accumulation (PR) and potential evapotranspiration (PET) from TerraClimate (Abatzoglou et al., 2018), a database with a timespan from 1990 to the present and a 4-km spatial resolution.

long-term average. A negative SPEI value indicates drier-than-normal conditions, with more negative values reflecting greater relative dryness. We focus on the main cropping season (meher) in Ethiopia, spanning between June and September.[19] We therefore compute the 4-month lag SPEI (SPEI-4) for each cluster at the end of each September, calculating the location specific changes in the climatic water balance from the long-run average during this period.

Figure A3 shows the distribution of SPEI during the meher seasons in the past 16 years in the sampled communities, between 2005 and 2020. The 2016 meher season was characterized by an extreme drought in the study localities with the median SPEI value close to -2 standard deviations – a rare event that is expected to occur approximately once every 44 years in normally distributed data. The exposure to the 2016 drought shows limited variation in this sample: the inter-quartile range is extremely narrow (–1.92 to –1.83), and the full range extends only from –1.98 to –1.68. In other years (and in the 2017—2020 study period), the average (median) SPEI value fluctuates between -1 and +1 standard deviations. About one-third of study clusters experienced negative SPEI during the 2019 meher, while only 5 percent did so during the 2020 meher. Figure E4 presents a map capturing the variation across the full set of study clusters in the underlying (raw) SPEI variable by year; there is considerable variation across regions and a steady shift from relatively dry to relatively wet conditions over the study period.

To isolate periods of relative dryness in our econometric analysis, we create a positive rectified dryness shock variable by setting positive SPEI values to zero and multiplying by -1, so that larger positive values indicate worsening dryness conditions. Conditional on being non-zero, the median value of this dryness shock variable is 0.23 SD and the maximum is 0.47 SD in the 2019 and 2020 meher seasons. We will interpret our results at a benchmark of 0.25 SD. By standard meteorological classifications (McKee et al., 1993), dryness shocks of this magnitude fall within the "mild" range (SPEI between -1 and 0). However, for the ultra-poor population in our sample, these shocks are economically consequential, as we demonstrate with two pieces of supporting evidence. A 0.25 SD increase in relative dryness reduces the value of cereal crop harvests by

[19] Meher is by far the most important agricultural season in Ethiopia with more than 90 percent of the total cereal production in the country taking place during this season (Taffesse et al., 2012). In our sample of households, the belg harvest contributed only about 5% to the total annual crop production (in terms of total value measured at the baseline).

approximately \$27 PPP, or six percent (Table C1 in Appendix C).[20] A 0.25 SD increase in our dryness indicator raises the probability of households reporting a drought shock by 10 percentage points, or roughly 20 percent relative to the sample mean (Table C2 in Appendix C).[21] This suggests that the moderate dryness captured by our measure corresponds to conditions that households in these communities perceive as drought shocks. Moreover, as we will show, these moderate shocks translate into sizable welfare effects: a 36 percent increase in the food gap and a 35 percent decline in livestock holdings among control households.

Our econometric analyses measure the impact of dryness shocks on outcomes of interest measured at the midline and endline. The endline took place in the post-harvest period in February-April 2021. For the endline observations, we therefore define the dryness indicator based on the previous meher season (June to September 2020). The midline took place during the 2019 meher (between July and October). It is therefore not *a priori* clear whether the 2019 or 2018 dryness conditions are most relevant. To address this, we seek guidance from our data by restricting the sample to control clusters and then regressing the outcome variables separately on two types of dryness indicators. The first indicator defines dryness using the 2019 meher (i.e., contemporaneous) conditions for the midline and the second indicator defines dryness using the 2018 meher (i.e., previous year's meher) conditions for the midline. The regression results reported in Table C3 in Appendix C indicate that the majority of the outcomes respond to the dryness conditions during the ongoing meher season in the midline. The only clear exception is women's BMI for which the previous meher season in the midline seems a stronger predictor. Therefore, when we measure the impacts on women's body mass, we define the dryness variable using the 2018 meher conditions if the measurements were taken in the midline survey. For all other indicators measured at the midline, we consider the 2019 meher that overlapped with the midline survey round. However, we

[20] Each survey round included a harvest module, but its structure was slightly different across rounds. At baseline and midline, the modules asked the respondent to report on all crops harvested by the household during the most recent harvest season. At endline, the module asked about the five most important crops by area. To maximize comparability, we focus on cereal crops, which are the most important crop category in this context. At baseline, nearly 80 percent of the total harvest value consisted of cereal crops.

[21] We use the self-reported shock data solely to validate that our SPEI-based dryness indicator captures dryness that households perceive as consequential. We do not use self-reported shocks in the main analysis for three reasons: unlike the SPEI, they are potentially endogenous to household characteristics and coping capacity; the shock modules differ in structure across survey rounds; and at endline, the module was administered only to households where the primary male respondent was available.

will also subsequently demonstrate that our primary findings are robust to including both contemporaneous and lagged dryness variables simultaneously into the estimated specification.

Balance

The statistical tests reported in Table D1 in Appendix D show that the sample is balanced across the study arms based on these outcome variables observed at baseline. The table also shows that the households are equally exposed to dryness shocks across the four study arms: the SPEI values for each meher season between 2016 and 2020 are balanced across the study arms. This is expected given that the randomization was stratified at the woreda level and weather outcomes for a single season are highly spatially correlated within a woreda (as we show later).

Another potential source of bias in this analysis would be a high level of serial correlation in dryness exposure, such that areas identified as experiencing dryness during the study period were already characterized by higher levels of poverty driven by past exposure to dryness. We can assess this hypothesis by regressing the dryness shock indicator for midline and endline on various household welfare indicators measured at the baseline together with woreda and survey round fixed effects, as in the main specifications reported below. None of the coefficients are statistically significant at the 5%-level and all are close to zero, indicating that there is no correlation between baseline poverty levels and dryness exposure during the study period (Table D2 in Appendix D). We also present a map (Figure E5) capturing the residualized SPEI values by cluster and by round; shifts geographically across rounds, from the southern and eastern clusters at midline to the northern clusters at endline.

Synthesis

The dryness variation we study occupies the mild end of the meteorological drought spectrum. However, for the ultra-poor households in our sample, this degree of dryness is consequential, as the crop income losses, self-reported drought perceptions, and welfare effects documented below confirm. Following the distinction between meteorological and socioeconomic drought drawn by Van Loon (2015) and discussed above, we refer to our measured variation as "dryness shocks" throughout the remainder of the paper to distinguish it from the severe meteorological droughts that standard classifications reserve for SPEI values below -1.

4. Econometric approach

We regress the outcomes of interest (women's diet diversity, women's body mass index, intimate partner violence, household-level food security, and household-level livestock holdings,) for household i located in kebele k and woreda d at the time t ($Y_{i,k,d,t}$) on a variable capturing dryness conditions in the most relevant meher season ($Dry_{k,\tilde{t}}$):

$$(1) \qquad Y_{i,k,d,t} = \beta Dry_{k,\tilde{t}} + \vartheta T_{k,d} + \delta\left(Dry_{k,\tilde{t}} * T_{k,d}\right) + R_t + \theta_d + \eta Y_{i,k,d,t=0} + \varepsilon_{i,k,d,t},$$

where t=1 if the household is observed in the midline survey and 2 if in the endline survey. $Dry_{k.\tilde{t}}$ is based on the SPEI measured in the relevant meher season: for the endline survey in 2021, the relevant season is 2020. For the midline survey, conducted during the 2019 meher season, we generally define the dryness shock based on the contemporaneous shock, but employ the 2018 season for women's BMI. To understand how SPIR mitigated the adverse impacts of dryness, we interact the dryness variable with a binary variable (T) capturing the exposed to treatment, and also include the non-interacted treatment variable. Initially, we pool the three treatment arms (implying that 75 percent of the sample clusters were exposed to SPIR, while 25 percent are control clusters), but we will later consider impacts by treatment arm.

Given this specification, the impact of dryness in control localities is quantified by β, while δ and ϑ represent intention-to-treat (ITT) effects of SPIR. Again, the dryness variable has been rectified so a higher value corresponds to dryness: we hypothesize β>0 for dryness and food gap, and β<0 for all other outcomes. If SPIR treatments mitigate the adverse impacts of dryness, δ should be of opposite sign to β: the net impact of dryness in treatment clusters is quantified by β+ δ. Additional control variables in Equation (1) include the binary variable R_t equal to one for the endline survey and zero for the midline survey, woreda fixed effects θ_d, and the outcome level at baseline. For the IPV outcomes, we replaced any missing baseline observation with zero and appended the estimated equation with a binary variable capturing these households for which the baseline value was missing.

The recommendation in the analyses of RCTs is to cluster the standard errors at the level of the unit of treatment assignment (Abadie et al., 2022). Here, however, the use of weather data creates strong spatial dependencies across study clusters (for more details, see Appendix E) in which case clustered standard errors are no longer valid (Barrios et al., 2012). Therefore, we compute Conley

(1999) standard errors that are robust to both spatial autocorrelation and heteroskedasticity,[22] applying a Bartlett spatial weighting matrix in which the weights linearly decay as distance between the clusters grows (Conley, 1999). We also report sharpened q-values pooling across the 15 coefficients estimated in this table (three coefficients for each of five outcomes), following Benjamini et al. (2006) and Anderson (2008).

The identifying variation in the dryness × treatment interaction derives from the residualized dryness measure — the component of kebele-level dryness that remains after absorbing woreda and survey-round fixed effects. Figure E5 maps this residualized measure separately for midline and endline, revealing meaningful within-woreda variation across kebeles in both rounds. To quantify the source of this identifying variation, we decompose the residualized dryness measure into a persistent cross-sectional component (the kebele's mean residual across rounds, capturing time-invariant micro-climate differences within a woreda) and a temporal component (the within-kebele deviation from that mean, capturing round-to-round weather fluctuations). Approximately 74 percent of the residual variance is temporal — the same kebeles experiencing different weather realizations across midline and endline — while 26 percent reflects within-woreda cross-sectional differences. The correlation of kebele-level dryness residuals between midline and endline is −0.52, confirming that the residualized measure is not proxying for a fixed kebele characteristic but reflects genuine weather variation that reverses sign across survey rounds.

Since the majority of the identifying variation is temporal (the same kebeles experiencing different weather across rounds) time-invariant kebele characteristics have limited scope to confound the dryness coefficient; they can only operate through the smaller cross-sectional component. A separate concern is whether the dryness × treatment interaction could reflect heterogeneous treatment effects along dimensions correlated with dryness rather than genuine dryness buffering. Table D2 shows that the residualized dryness measure is uncorrelated with observable household characteristics, and in Appendix G we show that augmenting the specification with interactions between baseline characteristics and the treatment indicator leaves the dryness × SPIR coefficients unchanged.

[22] We used user-written *acreg* command in Stata to estimate these regressions (Colella et al., 2019).

5. Results

Main results

The regression results based on the estimation of Equation (1) are provided in Table 1. Figure 1 summarizes these results for women's dietary diversity (Panel A), BMI (Panel B) and IPV (Panel C). In each panel, the top part of the graph provides the average effect of dryness in control clusters (i.e., β in Equation (1)) while the corresponding estimate in treatment clusters (i.e., $\beta+\delta$) is presented below.[23] The difference in the dryness impact between control and SPIR households is reported and labeled δ. The regression estimates have been divided by the baseline mean in the control group and then multiplied by 100, thus capturing the dryness impacts in terms of percent of the baseline mean in the control group. As noted above, we quantify the impact of dryness shocks in terms of a 0.25 SD increase in relative dryness, roughly equivalent to the median dryness shock within the sample areas experiencing any dryness during the study period.[24]

[Table 1 & Figure 1 here]

Columns 1 to 2 in Table 1 and Panels A and B of Figure 1 summarize the results on women's diet diversity and body-mass index, respectively. A 0.25 SD increase in relative dryness decreases women's diet diversity by 9 percent (0.18 food groups), on average, though this estimate is noisy (p-value = 0.164). The estimated effect of dryness for treated households is positive but not statistically different from zero (p = 0.232); however, the hypothesis that the effects of dryness are equivalent for treatment and control households can be rejected at the one percent level (column 1 in Table 1). A very similar pattern is observed for BMI, where we measure a weakly significant negative effect of dryness in control clusters, but a statistically significant positive difference when compared to the effect estimated for control households (column 2 in Table 1). Overall, dryness seem to exert a small negative impact on women's diets and BMI in control households, and there is some evidence that women in treated households were insulated from these negative dryness impacts. In general, we observe a high level of consistency between the reported p-values and q-values in this table; all statistically significant estimates are robust to the use of sharpened q-values.

---

[23] The confidence intervals for the joint estimates were calculated using the *lincom* command in Stata 18.

[24] As noted previously, the median dryness shock within dryness-affected areas is 0.23 standard deviations.

Column 3 of Table 1 and Panel C of Figure 1 report the results on the impact of dryness on risk of IPV in the past 12 months. In line with the previous literature (Abiona and Koppensteiner, 2018; Dehingia et al., 2024; Díaz and Saldarriaga, 2023; Epstein et al., 2020; Hertzog et al., 2025; Wang et al., 2025), dryness increase the risk of IPV in control clusters. Here, a 0.25 SD increase in relative dryness increases the risk of women experiencing intimate partner violence by 21 percent (or 4 percentage points). The corresponding impacts in treated households are considerably smaller in magnitude, and we cannot reject the null hypothesis that dryness have no impact on IPV in treated clusters. The difference between the estimated dryness impacts between control and treated households is statistically significant at the one percent level, suggestive of a meaningful effect of SPIR in buffering the effects of dryness on IPV. In Table F1, we examine the impacts of dryness and SPIR programming on different forms of IPV and find that sexual violence is particularly sensitive to dryness shocks in control households. The dryness impacts on physical and emotional violence are positive, but somewhat smaller and less precisely estimated. For all forms of IPV, the estimated dryness impacts in SPIR households are not statistically significant.

Broadening our focus to household-level outcomes, the evidence in Column 4 of Table 1 and Panel A of Figure 2 suggests that control households exposed to dryness report a considerably higher food gap: a 0.25 SD increase in relative dryness leads to a 0.75 month increase in the food gap, on average ($p<0.01$), or a 36% increase relative to the control mean. The corresponding dryness impact in treated clusters is less than half of this: 0.32 months or 15 percent, and this effect ($\beta+\delta$) is only weakly statistically significant ($p=0.06$). The coefficient on the interaction term (δ) reported in column 4 of Table 1 shows the positive impact of the graduation program on mitigating the dryness effect on the food gap, and this impact is statistically significant ($p<0.01$).

Moving on to Column 5 of Table 1 and Panel B of Figure 2, a 0.25 SD increase in relative dryness decreases livestock holdings by 35 percent, on average. While the percentage decline in livestock appears large, baseline holdings are modest: the estimated loss corresponds to one-third of a tropical livestock unit on average, or roughly two goats and four chickens. These magnitudes are consistent with partial herd liquidation rather than complete collapse. The corresponding impact in SPIR clusters is negative and statistically significant, indicating that treated households are also negatively affected by dryness shocks. However, the average dryness impact is one third of this magnitude in treated households, where a 0.25 SD increase in relative dryness decreases livestock holdings only by nine percent (0.09 TLU), on average. The difference between control and treated households, again captured by the interaction term δ, is statistically different from zero (p<0.01) (see Column 5 of Table 1).

[Figure 2 here]

Table 2 then shows parallel findings, including both the contemporaneous (same year) and lagged (one year) dryness measures in the specification. We can observe that the findings are largely consistent: as previously noted, for women's BMI, dryness as observed in the previous year has more adverse effects, while for diet diversity, intimate partner violence, the food gap, and livestock, dryness as observed in the current year has more adverse effects. (The effect of a current-year shock on intimate partner violence is less precisely estimated when controlling for lagged dryness, but comparable in magnitude).

[Table 2 here]

Theoretical and empirical work on poverty dynamics emphasizes that when households operate close to subsistence thresholds, even relatively modest income or production shortfalls can generate sharp welfare losses through constrained coping mechanisms, including reductions in consumption and distress sales of productive assets (Barrett and Carter, 2013). This is especially relevant in our setting, where nearly two-thirds of households live below the extreme poverty line. Consistent with this framework, prior studies show that rainfall shocks lead to disproportionately large consumption losses among asset-poor households, reflecting limited access to savings, credit, and insurance (Amare et al., 2018; Janzen and Carter, 2019). These dynamics provide a plausible explanation for why moderate dryness shocks in our data translate into relatively large changes in food security and livestock holdings, despite more modest effects on crop income.

We can also assess dryness effects on household per capita consumption; however, these data were collected only at baseline and endline. Moreover, because the 2020 meher season prior to the endline had limited dryness exposure (only about 5 percent of the clusters recorded a negative SPEI value), we assess the impact of the 2019 dryness conditions during which about one-third of the study clusters recorded negative SPEI. Figure F1 in the appendix summarizes the impacts on log per capita (adult equivalent) consumption, benchmarked against 0.25 SD increase in relative dryness, and the corresponding results are reported in Table F2. In control households, we see that a dryness in the magnitude of 0.25 SD during the 2019 meher season results in about 5.5% drop in household per capita consumption observed almost two years later in 2021 (p = 0.065). The corresponding impact in SPIR households is close to zero and not statistically significant ($p = 0.427$) while the difference relative to the estimate in the control households is highly significant. These results suggest that control households are in fact forced to cut down their consumption in the face of dryness while treated households are able to smooth.

Thus far, we have assessed the heterogeneity in program impact with respect to dryness exposure by focusing on the signs and magnitudes of β and δ estimates in Equation (1). We could also use our regression results to quantify the program's (pooled midline and endline) ITT effects at the mean level of our dryness indicator using the β, δ and ϑ estimates. Using this approach, the ITT estimates for food gap and livestock reported in Table 1 are not statistically different from zero,[25] in line with the estimated program impacts on livelihood outcomes as reported in Leight et al. (2026).[26] This is all broadly consistent with the stylized fact reported above that only a minority of households are exposed to dryness as defined in this analysis: about one-third of study clusters experienced dryness in 2019 and only 5 percent in 2020, implying that the positive protective effects of SPIR were experienced by a minority of the sample. In the analysis pooling across dryness-affected and non-dryness-affected areas, these positive effects are attenuated toward zero.

Heterogeneity across treatment arms

Next, we explore the heterogeneity across treatment arms to assess whether these protective effects are driven by the more intensive treatment packages, or the livelihood transfers given to the poorest households. Table F3 in Appendix F provides the regression results with β, δ, and ϑ in Equation (1) estimated for each of the three treatment arms. All estimates of δ are of opposite sign to the

---

[25] These are calculated by comparing the dryness shock impacts at the mean level of relative dryness in treated and control households: $[\beta*0.04+ \delta *0.04+ \vartheta] - \beta*0.04$, where 0.04 is the mean level of relative dryness across midline and endline calculated for all households. Using the *lincom* command in Stata we estimate the ITT impacts for food gap as -0.007 ($p = 0.923$) and for TLU as -0.003 ($p = 0.858$).

[26] Comparing the estimates discussed here to those reported in Leight et al. (2026) is not straightforward because the outcome variables are not identical (e.g., the food security indicators are not the same) and because here we pool midline and endline survey rounds while in Leight et al. the treatment effects are estimated separately for midline and endline (as is standard in impact evaluations). Yet, the overall narratives reported in this paragraph and in Leight et al. are similar: the SPIR treatment effects on livestock assets and food security are modest when the full sample of households is considered. Leight et al find positive effects on livestock holdings when the sample is restricted to the poorest households that received the livelihood transfer, but not for the less poor households that did not receive the transfer.

estimated β coefficients indicating that all treatment arms provided at least some protection against dryness. The magnitudes of these coefficients are also broadly similar across the treatment arms.[27]

To understand the role of the livelihood transfers given to the poorest households, we next restrict the sample to eligible households.[28] We then re-estimate the same specification for this sub-sample, and find similar patterns as reported in Table F4. The impacts of dryness on livestock holdings are smaller in absolute terms, but not in relative terms given that baseline livestock holdings are considerably more modest among these households (and livestock holdings were part of the index that determined the eligibility for the livelihood transfers). The corresponding impacts on the IPV outcomes reported in Table F4 show that the dryness estimates for the control households seem larger than those reported in Table F3 indicating that the risk of IPV increases relatively more in poorer households during dryness.[29] Treatment arm T3 did not include a livelihood transfer but was still effective in reducing the negative impact of dryness on sexual and emotional violence. Treatment arm T2 (combining livelihoods training with grants and the core nutrition intervention) seems consistently to have been most effective in protecting women against the increased risk of IPV due to dryness.

Self-reported stress levels and women's bargaining power

Previous research on the impact of weather shocks on the risk of IPV has suggested that increased poverty-related stress levels, particularly among males, are a key factor driving the heightened risk of IPV following such shocks (Díaz and Saldarriaga, 2023; Epstein et al., 2020). However, this literature rarely reports any data on actual self-reported stress. Our endline survey asked both male and female primary respondents to rate their current stress levels using a scale ranging from one to 10.[30] To explore whether dryness increase stress levels and whether SPIR weakens this relationship, we regressed the stress indicator measured in 2021 on a dryness indicator capturing

---

[27] There is some suggestive evidence that the intensive nutrition interventions (N*) were more effective in protecting self-reported food security against dryness shocks. Meanwhile, the intensive livelihood interventions (L*) were more effective in protecting livestock during dryness. In contrast, the Wald tests for diet diversity and BMI show no statistically significant differences across treatment arms.

[28] The eligibility assessment for livelihood transfers was conducted in all treatment clusters, including the control and T3 clusters where none of the households received these transfers.

[29] This was further confirmed with a simple interaction model when the sample was restricted to the control clusters: dryness shocks did not lead to increases in the IPV risk in less poor households. Results available upon request.

[30] This relatively simple stress indicator has been employed in the 'Stress in America' surveys conducted by the American Psychological Association (APA) during the COVID-19 pandemic, see APA (2021).

the weather conditions in the 2019-meher season, the SPIR treatment indicator and their interaction while controlling for woreda fixed effects.[31] The results, reported in Table F5, indicate that dryness increase stress levels for both males (column 1) and females (column 2). The negative and statistically significant ($p<0.10$ for males; $p<0.01$ for females) coefficients on the interaction terms suggest that SPIR weakens this relationship. Specifically, a 0.25-SD dryness is associated with a 30 percent rise in male stress levels in control clusters and 26 percent rise in SPIR clusters. For females, a dryness of the same magnitude is linked to a 27 percent increase in stress levels in control clusters and 22 percent increase in SPIR clusters.[32]

Beyond stress, dryness may also increase IPV by eroding women's bargaining power within the household. Using a measure of women's input into productive decisions (building on Malapit et al., 2019), we find that dryness significantly reduces women's decision-making authority (Column 3 of Table F5). A 0.25-SD dryness reduces the probability that a woman has input into most or all decisions by 13 percentage points (23 percent of the baseline mean) in control clusters, compared to 5 percentage points (9 percent) in SPIR clusters. This identifies a complementary pathway: dryness both heightens male stress and weakens women's economic position, and the SPIR intervention disrupts both links.

Robustness

We conducted an extensive set of robustness checks to assess whether the estimated impacts are sensitive to alternative measurement choices, specification changes, or concurrent shocks. Using a dryness indicator derived from precipitation data yields very similar findings, and the main results remain stable when replacing woreda fixed effects with broader regional controls or the survey dummy with region-specific time fixed effects. Including both contemporaneous and lagged shocks likewise produces consistent patterns. We further account for major covariate shocks that affected Ethiopia during the study period – notably desert locust infestations and

[31] We cannot directly estimate Equation (1) because the outcome data (stress indicators) were only collected at endline. Consequently, we do not include a control for baseline value or survey round fixed effects. In addition, because the meher season prior to endline was relatively good in terms of weather (only 5% of study clusters had negative SPEI), we assess the impact of the 2019 dryness conditions (in which about 30% of study clusters had negative SPEI).

[32] There is also some arguably counterintuitive evidence that the SPIR intervention itself increases stress, particularly for women, a pattern that could be consistent with strain linked to expansion of new livelihoods activities; this suggests some caveats in interpretation may be appropriate, and highlights the need for more empirical evidence around the stress channel in the broader literature.

conflict exposure – and the estimated treatment effects are unchanged. The results are also robust to controlling for households' exposure to the 2016 drought and show no evidence that differential migration or changes in household composition drive the findings. Inference remains stable across a wide range of spatial kernel cut-offs when computing Conley standard errors. To address the concern that the dryness × treatment interaction may reflect heterogeneous treatment effects along dimensions spatially correlated with dryness, we augment the main specification with interactions between baseline household characteristics and the SPIR indicator. The dryness × SPIR coefficients are unchanged. Adding a treatment × endline interaction to allow the average treatment effect to differ across survey rounds likewise leaves the results intact. Appendix G provides the full results and more details on these robustness checks.

6. Mechanisms

The foregoing results demonstrate that SPIR effectively protected women from the risks of declining nutritional status and increased violence during dryness, and also buffered households from declining food security and livestock assets. The underlying pattern suggests households exposed to SPIR are more effectively managing adverse income shocks, and as a result, women's vulnerability is also reduced.

The literature has identified multiple strategies that rural households can use to manage shocks in the absence of formal insurance and credit markets. One strategy is adjusting crop production – either by investing in improved technologies, such as improved seeds, or diversifying the crop mix (Boucher et al., 2024; Emerick et al., 2016; Michler and Josephson, 2017). Another is accumulating liquid buffer stocks, such as cash reserves, that can be drawn upon during financial distress. Productive assets such as livestock can also serve as a buffer (Rosenzweig and Wolpin, 1993). However, in poorly integrated markets asset prices tend to co-move with incomes (Lange and Reimers, 2021), making asset accumulation a highly imperfect form of self-insurance. Finally, households may reduce their dryness risk by diversifying beyond crop agriculture. Some increase their involvement in livestock production, which is less sensitive to rainfall variability than crop production (Toulmin, 1986), while others shift into non-agricultural activities (Macours et al., 2022) that are not directly dependent on rainfall.[33]

Building on these insights, we next examine whether SPIR influenced households' ability to adopt these coping strategies. Like many multifaceted interventions, SPIR likely operates through multiple channels, supporting resilience by easing financial constraints, expanding livelihood opportunities, and strengthening social and economic networks. This aligns with the theory of change in graduation programs, which aims to address the structural barriers that keep households trapped in poverty.

Shifts in cropping patterns

We begin by examining whether SPIR led households to adjust their farming practices. Specifically, we estimate treatment effects on investments in seeds, inorganic fertilizers, pesticides, hired labor, and the number of crops cultivated (crop diversity). To do so, we first restrict the sample to households that cultivated crops in the previous cropping season and use an ANCOVA specification, regressing each outcome on the SPIR treatment indicator, the baseline value of the outcome, the survey round, and woreda fixed effects. We use data from both the midline and endline rounds when available.[34] The treatment effects reported in Table F6 in the appendix are uniformly statistically insignificant and mostly small in magnitude, suggesting that shifts within crop agriculture are not a channel for the overall protective effects of SPIR observed.[35]

Accumulation of buffer stocks

The majority of survey respondents cited protecting themselves against unexpected income losses or health emergencies as their primary reason for saving, highlighting the importance of precautionary savings in this context. Building on this insight as well as on the finding reported in Leight et al. (2026) that SPIR led to increases in savings as well as some increased credit access, we next assess whether SPIR households tap into their savings or draw down additional credit

[33] Even non-agricultural livelihoods may still be affected by rainfall fluctuations through local general equilibrium effects in agrarian economies.
[34] The crop module in the endline questionnaire only asked whether the household spent money on inputs (except seeds) but did not ask how much was spent.
[35] We also estimate a parallel specification including no sample restriction, and also adding as an outcome variable a binary variable equal to one for households that report any cultivation in the meher; the findings reported in Table F7 again consistently show null effects.

when faced with a dryness shock. To explore this, we re-estimate Equation (1) using both total savings (as of the survey date) and any past-year use of credit as the dependent variables;[36] see Columns 1 and 2 of Table 3.

For savings, the coefficient on the uninteracted SPIR variable shows that among households not affected by a dryness shock, SPIR led to a $20 PPP increase in household savings relative to control households. The coefficients on the dryness variables and the interaction terms capturing SPIR households affected by a dryness confirm our hypothesis: treated households draw down their savings in response to shocks, whereas savings levels among control households remain unchanged. A dryness of 0.25 SD reduces savings in treated households by approximately $10 PPP – about half the magnitude of SPIR's overall effect on savings ($20 PPP) in this specification. For credit, the coefficient on the uninteracted SPIR variable shows that SPIR leads to a significant, albeit small increase in reported credit access (5 percentage points). In the context of dryness, households targeted by SPIR do not seem to increase their access to credit, rather it is households in the control arm that show a significant increase in credit use (presumably linked to financial distress), while this effect is fully muted in SPIR clusters. This suggests that SPIR households preferentially rely on savings rather than credit as a cash buffer to cushion the impact of shocks, a choice that may be preferable if credit is costly.

Ultimately, these cash buffers help cushion the impact of shocks, reducing the need to liquidate livestock or cut consumption – coping strategies that can be costly in the long term, as observed among control households (Figure 2, Panel B, and Figure F1, respectively). Column 3 of Table 3 further supports this interpretation, showing that SPIR households sell less livestock after a shock compared to control households.

One possible explanation for the absence of savings drawdown in the control group is simply that most households in this group have no savings to use. At midline and endline, only about 45 percent of control households report holding any savings at all, compared to 77 percent of treated households.

---

[36] We use a binary rather than a continuous measure of credit as we do not have data on the full amount of credit accessed over the past year; the only continuous measure of credit reported is loans outstanding as of the survey date, a variable that also captures patterns of repayment. To avoid any potential confounding linked to the propensity to repay, we focus on the binary variable.

[Table 3 here]

Income diversification

The results reported in Leight et al. (2026) show that SPIR households generate higher cash income from livestock than control households. At the same time, there is no effect on crop incomes, suggesting that livestock began playing a larger role in SPIR households' income portfolios. Unlike Macours et al. (2022) in the context of rural Nicaragua, there is no evidence of meaningful diversification out of agriculture in this setting. At endline, fewer than five percent of households report owning a non-agricultural business, and fewer than five percent report engaging in regular wage labor. This may reflect the limited demand for non-agricultural goods and services in a context where widespread poverty constrains spending largely to food.

7. Cost effectiveness

The marginal cost of the SPIR graduation components above the base PSNP ranges from approximately $150 to $500 per household in 2017 PPP terms, depending on the treatment arm (Leight et al., 2026). This is substantially below the cost of the most successful BRAC-style graduation models (500 to 2,000 $PPP). Critically, the dryness-buffering effects are broadly similar across all SPIR treatment arms, including the arm that received only savings groups and financial literacy training. This suggests that the core graduation programming centering around the savings group platform – the least costly component, at approximately $150 per household – is the primary driver of enhanced resilience.

The cost-effectiveness of this investment is best understood through an insurance lens. The marginal cost of approximately $150 PPP per household represents a one-time premium that builds self-insurance capacity: households accumulate precautionary savings through VESAs that they draw down when dryness occur. The premium is paid once, but the coverage extends to every subsequent dryness episode. Table 4 quantifies this coverage. During a dryness of median magnitude (a 0.25 SD increase in relative dryness), a SPIR household – relative to a PSNP-only household experiencing the same shock – avoids 0.43 additional months of food insecurity (95% CI: [0.21, 0.66]), is buffered from a 4.0 percentage point increase in the risk of intimate partner violence (95% CI: [1.7, 6.4]), retains 0.25 more TLU in livestock (95% CI: [0.20, 0.30]), and experiences no decline in dietary diversity or BMI. Assessing exposure is challenging given that not all households will experience dryness in any given period, and the expected return depends on both dryness frequency and whether the buffering capacity decays with repeated shocks: the illustrative cost-per-unit estimates in Table 4 assume the $150 investment is protective vis-à-vis only a single dryness episode.

[Table 4 here]

The literature assessing the protective impacts of cash and cash-plus programs generally has not reported program cost data,[37] making it impossible to assess relative cost-effectiveness across studies. Even if comparable cost data were available, cross-study comparisons would be challenging to interpret for at least three reasons. First, the magnitude of the dryness shock varies substantially across settings, and comparing the cost of averting a month of food insecurity during a moderate dryness episode against the cost during a severe dryness conflates program effectiveness with shock severity. Second, dryness measurement itself is not standardized: studies variously use binary indicators, continuous rainfall in millimeters, rainfall z-scores, or SPEI indices, rendering the implied 'dose' non-commensurate. Third, studies measure impacts on a wide range of outcomes with often limited overlap, further complicating any attempt at a common cost-effectiveness metric. Nonetheless, we highlight these estimates as indicative of the importance of future cross-context analysis of cost-effectiveness.

8. Conclusions

The combination of high levels of poverty (Beegle et al., 2016), reliance on rainfed agriculture (You et al., 2012), and warm-to-hot climates make sub-Saharan Africa one of the world's most vulnerable regions as climate change intensifies (IMF, 2020). In Ethiopia, even moderate fluctuations in dryness during the cropping season have meaningful consequences for women's well-being. Among PSNP households receiving the base transfers only, dryness reduce women's dietary diversity and body mass, sharply increase food insecurity, deplete livestock holdings, and dramatically increase the risk of intimate partner violence – showing how climate-related shocks can undermine both material welfare and women's safety.

Layering a light-touch graduation model onto the national safety net substantially alters these dynamics. Across all outcomes examined, SPIR households are far less vulnerable to weather

[37] Exception is Macours et al. (2022) that reports costs but not cost-effectiveness calculations: $240 (CCT alone), $470 (CCT + training), or $480 (CCT + productive grant), all in 2005–2006 USD.

shocks: women in treated communities show no decline in diet quality or BMI and face no measurable increase in the risk of intimate partner violence. Treated households also experience smaller increases in food insecurity and much smaller losses in livestock assets during dryness. These findings indicate that even a relatively modest livelihood package, when integrated into an existing safety net, can provide substantial protection against dryness.

The analysis also sheds light on the channels through which women are disproportionately vulnerable to dryness. In control communities, dryness exposure is accompanied by a notable rise in reported stress among male household members and an erosion of women's decision-making authority over productive income. The rollout of a graduation model alters this dynamic: treated households build larger savings and maintain better access to credit, and these financial cushions plausibly ease the pressures that can escalate during dry spells, preserve women's bargaining position, and crucially, minimize any eruption of domestic violence.

Importantly, while our study period was characterized by typical weather fluctuations, it did not include a major dryness, and we should be cautious about generalizing these results outside of the study population (Banerjee et al., 2017) or outside the study period within the same population (Rosenzweig and Udry, 2019). One possible explanation for the magnitude of these impacts is that households may not have fully rebuilt their buffer stocks following the severe 2016 drought. At the same time, it remains uncertain whether the program would have offered similar protection against a major drought.

Climate projections for Ethiopia indicate that exposure to severe drought increases substantially with warming, despite projected increases in precipitation, because rising temperatures increase evaporation (i.e., the rate at which moisture is lost from soils and crops) (Price et al., 2022; Warren et al., 2024). At 3°C of global warming, the probability of severe drought (SPEI $\leq$ -1.5 SD) doubles relative to current conditions, and drought lengths are projected to reach 3–4 years. As severe droughts become more frequent, the moderate dryness episodes studied here will occur more often as well.

The most successful graduation model programs pioneered by BRAC have generated large shifts in household economic welfare that persisted in the long-term (Balboni et al., 2022; Bandiera et al., 2017; Banerjee et al., 2015), but importantly, they require intensive implementation and large asset transfers (in the region of 500 to 2,000 $PPP). The scalability of these programs remains an open question. By contrast, SPIR was designed as a light-touch graduation model program and embedded into the PSNP, one of the largest safety net programs in Africa. Our results provide new evidence on the value of this lighter-touch model, especially in settings exposed to recurrent weather shocks. We show that this model, when embedded in a national safety net, meaningfully reduces the vulnerability of women and their households to localized dryness. This suggests that relatively modest, scalable enhancements to existing safety net systems can play an important role in climate adaptation strategies in low-income settings.

References

Abadie A, Athey S, Imbens GW, Wooldridge J. When should you adjust standard errors for clustering? *Quarterly Journal of Economics* 2022;138; 1–35.

Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific data* 2018;5; 1-12.

Abiona O, Koppensteiner MF. The impact of household shocks on domestic violence: Evidence from Tanzania. IZA Discussion Paper, vol. No. 11992. IZA – Institute of Labor Economics: Bonn, Germany; 2018.

Adhvaryu A, Nyshadham A, Molina T, Tamayo J. Helping children catch up: Early life shocks and the progresa experiment. *The Economic Journal* 2024;134; 1-22.

Aizer A. The gender wage gap and domestic violence. *American Economic Review* 2010;100; 1847-1859.

Alderman H, et al. Men can cook: Effectiveness of a men's engagement intervention to change attitudes and behaviors in rural Ethiopia. *World Development* 2025;185; 106781.

Alderman H, Hoddinott J, Kinsey B. Long term consequences of early childhood malnutrition. *Oxford Economic Papers-New Series* 2006;58; 450-474.

Amare M, Jensen ND, Shiferaw B, Cissé JD. Rainfall shocks and agricultural productivity: Implication for rural household consumption. *Agricultural Systems* 2018;166; 79-89.

Anderson ML. 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* 2008;103; 1481-1495.

Andrews C, Hsiao A, Ralston L. Social Safety Nets Promote Poverty Reduction, Increase Resilience, and Expand Opportunities. In: Beegle K, Coudouel A, Monsalve E (Eds), Realizing the Full Potential of Social Safety Nets in Africa. The World Bank: Washington, D.C.; 2018.

APA. Stress in America January 2021 Stress Snapshot. American Psychological Association (APA): Washington, DC; 2021.

Asfaw S, et al. Cash transfer programmes, weather shocks and household welfare: evidence from a randomised experiment in Zambia. *Journal of Development Effectiveness* 2017;9; 419-442.

Balboni C, et al. Why do people stay poor? *The Quarterly Journal of Economics* 2022;137; 785-844.

Bandiera O, et al. Labor markets and poverty in village economies. *The Quarterly Journal of Economics* 2017;132; 811-870.

Banerjee A, et al. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science* 2015;348; 1260799.

Banerjee AV, Chassang S, Snowberg E. Decision theoretic approaches to experiment design and external validity. Handbook of Economic Field Experiments, vol. 1. Elsevier; 2017. p. 141-174.

Barrett CB, Carter MR. The economics of poverty traps and persistent poverty: Empirical and policy implications. *The Journal of Development Studies* 2013;49; 976-990.

Barrios T, Diamond R, Imbens GW, Kolesár M. Clustering, spatial correlations, and randomization inference. *Journal of the American Statistical Association* 2012;107; 578-591.

Beegle K, Christiaensen L, Dabalen A, Gaddis I. Poverty in a rising Africa. The World Bank; 2016.

Beegle K, Coudouel A, Monsalve E. Realizing the Full Potential of Social Safety Nets in Africa. The World Bank: Washington D.C.; 2018.

Beguería S, Vicente-Serrano SM. SPEI: calculation of the standardised precipitation-evapotranspiration index. *R package version* 2017;1.

Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 2006;93; 491-507.

Björkman-Nyqvist M. Income shocks and gender gaps in education: Evidence from Uganda. *Journal of Development Economics* 2013;105; 237-253.

Bossuroy T, et al. Tackling psychosocial and capital constraints to alleviate poverty. *Nature* 2022;605; 291-297.

Boucher SR, et al. Bundling Genetic and Financial Technologies for More Resilient and Productive Small-Scale Farmers in Africa. *The Economic Journal* 2024; ueae012.

Brune L, Karlan D, Kurdi S, Udry C. Social protection amidst social upheaval: Examining the impact of a multi-faceted program for ultra-poor households in Yemen. *Journal of Development Economics* 2022;155; 102780.

Buller AM, et al. A mixed-method review of cash transfers and intimate partner violence in low- and middle-income countries. *The World Bank Research Observer* 2018;33; 218-258.

Carpena F. How do droughts impact household food consumption and nutritional intake? A study of rural India. *World Development* 2019;122; 349-369.

Carter M, de Janvry A, Sadoulet E, Sarris A. Index insurance for developing country agriculture: a reassessment. *Annual Review of Resource Economics* 2017;9; 421-438.

Christian P, Kandpal E, Palaniswamy N, Rao V. Safety nets and natural disaster mitigation: evidence from cyclone Phailin in Odisha. *Climatic Change* 2019;153; 141-164.

Colella F, Lalive R, Sakalli SO, Thoenig M. Inference with arbitrary clustering. IZA Discussion Paper, vol. 12584. 2019.

Collins D, Morduch J, Rutherford S, Ruthven O. Portfolios of the Poor: How the World's Poor Live on $2 a Day. Princeton University Press: Princeton, NJ; 2009.

Conley TG. GMM estimation with cross sectional dependence. *Journal of Econometrics* 1999;92; 1-45.

Corno L, Hildebrandt N, Voena A. Age of marriage, weather shocks, and the direction of marriage payments. *Econometrica* 2020;88; 879-915.

Couttenier M, Soubeyran R. Drought and civil war in sub-saharan africa. *The Economic Journal* 2014;124; 201-244.

Crosta T, et al. Unconditional Cash Transfers: A Bayesian Meta-Analysis of 50 Randomized Evaluations in 26 Low and Medium Income Countries. Unpublished manuscript. 2023.

De Janvry A, Finan F, Sadoulet E, Vakis R. Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks? *Journal of Development Economics* 2006;79; 349-373.

De Waal A. Mass starvation: The history and future of famine. Polity Press: Cambridge, UK; 2017.

Deaton A, Grosh M. Consumption. In: Grosh M, Glewwe P (Eds), Designing Household Survey Questionaires for Developing Countries: Lessons from 15 Years of Living Standards Measurement Study, vol. 1. World Bank: Washington D.C.; 2000. p. 91-133.

Dehingia N, et al. Climate and gender: association between droughts and intimate partner violence in India. *American Journal of Epidemiology* 2024;193; 636-645.

Dercon S. Growth and shocks: evidence from rural Ethiopia. *Journal of Development Economics* 2004;74; 309-329.

Dercon S, Krishnan P. In sickness and in health: Risk sharing within households in rural Ethiopia. *Journal of Political Economy* 2000;108; 688-727.

Díaz J-J, Saldarriaga V. A drop of love? Rainfall shocks and spousal abuse: evidence from rural Peru. *Journal of Health Economics* 2023;89; 102739.

Emerick K, De Janvry A, Sadoulet E, Dar MH. Technological innovations, downside risk, and the modernization of agriculture. *American Economic Review* 2016;106; 1537-1561.

Epstein A, et al. Drought and intimate partner violence towards women in 19 countries in sub-Saharan Africa during 2011-2018: a population-based study. *PLoS Medicine* 2020;17; e1003064.

Fafchamps M, Udry C, Czukas K. Drought and saving in West Africa: are livestock a buffer stock? *Journal of Development Economics* 1998;55; 273-305.

FAO, FHI 360. Minimum Dietary Diversity for Women: A Guide to Measurement. Food and Agriculture Organization (FAO) of the United Nations and USAID's Food and Nutrition Technical Assistance III Project (FANTA), managed by FHI 360: Rome; 2016.

Fogel R. Economic growth, population theory, and physiology: the bearing of long-term processes on the making of economic policy. *The American Economic Review* 1994;84; 369-395.

Fox GL, Benson ML, DeMaris AA, Van Wyk J. Economic distress and intimate violence: Testing family stress and resources theories. *Journal of Marriage and Family* 2002;64; 793-807.

Harari M, La Ferrara E. Conflict, Climate and Cells: A disaggregated analysis. *Review of Economics and Statistics* 2018;100; 594-608.

Haushofer J, Ringdal C, Shapiro JP, Wang XY. Income changes and intimate partner violence: Evidence from unconditional cash transfers in Kenya. National Bureau of Economic Research; 2019.

Heath R, Hidrobo M, Roy S. Cash transfers, polygamy, and intimate partner violence: Experimental evidence from Mali. *Journal of Development Economics* 2020;143; 102410.

Hertzog L, et al. Extreme drought and sexual violence against adolescent girls and young women: A multi-country population-based study. *PLOS Global Public Health* 2025;5; e0004752.

Hidrobo M, Hoddinott J, Kumar N, Olivier M. Social Protection, Food Security, and Asset Formation. *World Development* 2018;101; 88-103.

Hidrobo M, et al. Leveraging social protection to strengthen women's and girls' climate-resilience in agrifood systems. *Food Policy* 2026;140; 103066.

Hidrobo M, et al. Social protection and gender: policy, practice and research. IFPRI Discussion Paper vol. 02257. International Food Policy Research Institute (IFPRI): Washington, D.C.; 2024.

Hirvonen K, Machado E, Simons A, Taraz V. Ethiopia's flagship public works program increases tree cover in rural areas and sloped terrains. *Global Environmental Change* 2022;75.

Hoddinott J, Kinsey B. Adult health in the time of drought. FCND Discussion paper, vol. 79. International Food Policy Research Institute (IFPRI): Washington D.C.; 2000.

Hou X. Can drought increase total calorie availability? The impact of drought on food consumption and the mitigating effects of a conditional cash transfer program. *Economic Development and Cultural Change* 2010;58; 713-737.

IMF. Chapter 2. Adapting to Climate Change in Sub-Saharan Africa. In: (IMF) IMF (Ed), Regional Economic Outlook Sub-Saharan Africa. International Monetary Fund (IMF): Washington, D.C.; 2020.

Jahnke HE. Livestock production systems and livestock development in tropical Africa. Kieler Wissenschaftsverlag Vauk Kiel: Germany; 1982.

Janzen SA, Carter MR. After the drought: The impact of microinsurance on consumption smoothing and asset protection. *American Journal of Agricultural Economics* 2019;101; 651-671.

Jensen N, Barrett C. Agricultural index insurance for development. *Applied Economic Perspectives and Policy* 2017;39; 199-219.

Kala N, Balboni C, Bhogale S. Climate adaptation. *VoxDevLit* 2023;7; 1-26.

Karlan D, Osei R, Osei-Akoto I, Udry C. Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics* 2014;129; 597-652.

Karlan D, Savonitto B, Thuysbaert B, Udry C. Impact of savings groups on the lives of the poor. *Proceedings of the National Academy of Sciences* 2017;114; 3079-3084.

Kazianga H, Udry C. Consumption smoothing? Livestock, insurance and drought in rural Burkina Faso. *Journal of Development economics* 2006;79; 413-446.

Knippenberg E, Hoddinott J. Shocks, social protection, and resilience: Evidence from Ethiopia. ESSP Working Paper vol. 109. International Food Policy Research Institute (IFPRI): Addis Ababa; 2017.

Lane G. Adapting to climate risk with guaranteed credit: Evidence from Bangladesh. *Econometrica* 2024;92; 355-386.

Lange S, Reimers M. Livestock as a buffer stock in poorly integrated markets. *Economic Development and Cultural Change* 2021;69; 727-764.

Leight J, et al. Can a light-touch graduation model enhance livelihood outcomes? Evidence from Ethiopia. *Journal of Development Economics* 2026;179.

Leight J, Hirvonen K, Zafar S. The effectiveness of cash and cash plus interventions on livelihoods outcomes: Evidence from a systematic review and meta-analysis. IFPRI Discussion Paper, vol. 2262. International Food Policy Research Institute (IFPRI): Washington D.C.; 2024.

Lipper L, et al. Climate smart agriculture: building resilience to climate change. Springer Nature; 2017.

Maccini S, Yang D. Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. *American Economic Review* 2009;99; 1006-1026.

Macours K, Premand P, Vakis R. Transfers, Diversification and Household Risk Strategies: Can productive safety nets help households manage climatic variability? *The Economic Journal* 2022;132; 2438–2470.

Malapit H, et al. Development of the project-level Women's Empowerment in Agriculture Index (pro-WEAI). *World development* 2019;122; 675-692.

Maystadt J-F, Calderone M, You L. Local warming and violent conflict in North and South Sudan. *Journal of Economic Geography* 2015;15; 649-671.

McKee TB, Doesken NJ, Kleist J. The relationship of drought frequency and duration to time scales. Proceedings of the 8th Conference on Applied Climatology, vol. 17. California; 1993. p. 179-183.

Mekuriaw Z, Harris-Coble L. Ethiopia's Livestock Systems: Overview and Areas of Inquiry. In: Systems FtFILfL (Ed). Gainesville, FL, USA; 2021.

Michler JD, Josephson AL. To specialize or diversify: Agricultural diversity and poverty dynamics in Ethiopia. *World Development* 2017;89; 214-226.

Peterman A, Kumar N, Pereira A, Gilligan DO. Toward Gender Equality: A Critical Assessment of Evidence on Social Safety Nets in Africa. In: Quisumbing AR, Meinzen-Dick R, Suseela N, Jemimah (Eds), 2019 Annual trends and outlook report: Gender equality in rural Africa: From commitments to outcomes. International Food Policy Research Institute (IFPRI): Washington D.C.; 2020.

Pomeranz D, Kast F. Savings accounts to borrow less: experimental evidence from Chile. *Journal of Human Resources* 2024;59; 70-108.

Pople A, Hill R, Dercon S, Brunckhorst B. Anticipatory cash transfers in climate disaster response. CSAEWorkingPaper, vol. WPS/202107. Oxford; 2021.

Premand P, Stoeffler Q. Do cash transfers foster resilience? Evidence from rural Niger. World Bank Policy Research Working Paper. The World Bank: Washington D.C.; 2020.

Price J, et al. Quantification of meteorological drought risks between 1.5° C and 4° C of global warming in six countries. *Climatic Change* 2022;174; 12.

Rosenzweig M, Udry C. External validity in a stochastic world. *The Review of Economic Studies* 2019;87; 343–381.

Rosenzweig MR, Wolpin KI. Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India. *Journal of Political Economy* 1993; 223-244.

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

Sabates-Wheeler R, Lind J, Hoddinott J, Tefera Taye M. Graduation after 10 years of Ethiopia's Productive Safety Net Programme: Surviving but still not thriving. *Development Policy Review* 2021;39; 511-531.

Seneviratne SI, et al. Weather and Climate Extreme Events in a Changing Climate. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, Maycock TK, Waterfield T, Yelekçi O, Yu R, Zhou B (Eds), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA; 2021. p. 1513–1766.

Taffesse AS, Dorosh P, Gemessa SA. Crop Production in Ethiopia. In: Dorosh P, Rashid S (Eds), Food and Agriculture in Ethiopia. University of Pennsylvania Press: Philadelphia; 2012.

Toulmin C. Drought and the farming sector: Loss of farm animals and post-drought rehabilitation. *ILCA ALPAN Network Paper* 1986.

van Daalen K, Jung L, Dhatt R, Phelan AL. Climate change and gender-based health disparities. *The Lancet Planetary Health* 2020;4; e44-e45.

Van Loon AF. Hydrological drought explained. *Wiley Interdisciplinary Reviews: Water* 2015;2; 359-392.

Vicente-Serrano SM, Beguería S, López-Moreno JI. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate* 2010;23; 1696-1718.

Vicente-Serrano SM, et al. Performance of drought indices for ecological, agricultural, and hydrological applications. *Earth Interactions* 2012;16; 1-27.

Waaler HT. Height. Weight and mortality the Norwegian experience. *Acta Medica Scandinavica* 1984;215; 1-56.

Wang P, et al. Drought, Extreme Heat, and Intimate Partner Violence in Low-and Middle-Income Countries. *JAMA Network Open* 2025;8; e2527818-e2527818.

Warren R, et al. Risks associated with global warming of 1.5 to 4 C above pre-industrial levels in human and natural systems in six countries. *Climatic Change* 2024;177; 48.

Webb D. Critical Periods in Cognitive and Socioemotional Development: Evidence from Weather Shocks in Indonesia. *The Economic Journal* 2023;134; 1637–1665.

WHO. 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. World Health Organization (WHO): Geneva; 2016.

Wiseman W, Van Domelen J, Coll-Black S. Designing and implementing a rural safety net in a low income setting: Lessons learned from Ethiopia's Productive Safety Net Program 2005-2009. The World Bank: Washington D.C.; 2010.

World Bank. World Bank Group gender strategy (FY16-23): Gender equality, poverty reduction and inclusive growth. The World Bank Group: Washington D.C.; 2015.

World Bank. The State of Social Safety Nets 2018. World Bank: Washington D.C.; 2018.

You L, Ringler, C., et al. What Is the Irrigation Potential for Africa? A Combined Biophysical and Socioeconomic Approach. *Food Policy* 2012;36; 770-782.

Table 1. Impact of graduation programming and dryness conditions on household and individual level outcomes

	(1)	(2)	(3)	(4)	(5)
Outcome:	Primary woman's diet diversity	BMI of primary woman	Primary woman experienced IPV	Household food gap	Tropical livestock units owned by household
Dryness	-0.706	-0.158*	0.144***	2.997***	-1.375***
	(0.507)	(0.094)	(0.056)	(0.633)	(0.123)
	[0.098]	[0.056]	[0.009]	[0.000]	[0.000]
Dryness X SPIR	1.294***	0.213***	-0.161***	-1.732***	1.015***
	(0.168)	(0.038)	(0.048)	(0.453)	(0.104)
	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]
SPIR	0.033	-0.032	0.007	0.067	-0.046***
	(0.053)	(0.039)	(0.006)	(0.061)	(0.017)
	[0.219]	[0.178]	[0.143]	[0.145]	[0.008]
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome in control group	2.02	20.12	0.17	2.06	0.97
‘Dryness’ + ‘Dryness X SPIR’ = 0	p = 0.232	p = 0.489	p = 0.637	p = 0.064	p < 0.001
Normalized dryness impact in control households	-0.18	-0.04*	0.04***	0.75***	-0.34***
As % of the baseline control mean	*-8.72*	*-0.20**	*20.67****	*36.32****	*-35.31****
Normalized dryness impact in SPIR households	0.15	0.01	-0.00	0.32*	-0.09***
As % of the baseline control mean	*7.26*	*0.07*	*-2.42*	*15.33**	*-9.26****
Number of observations	5,877	4,888	3,808	6,212	6,274

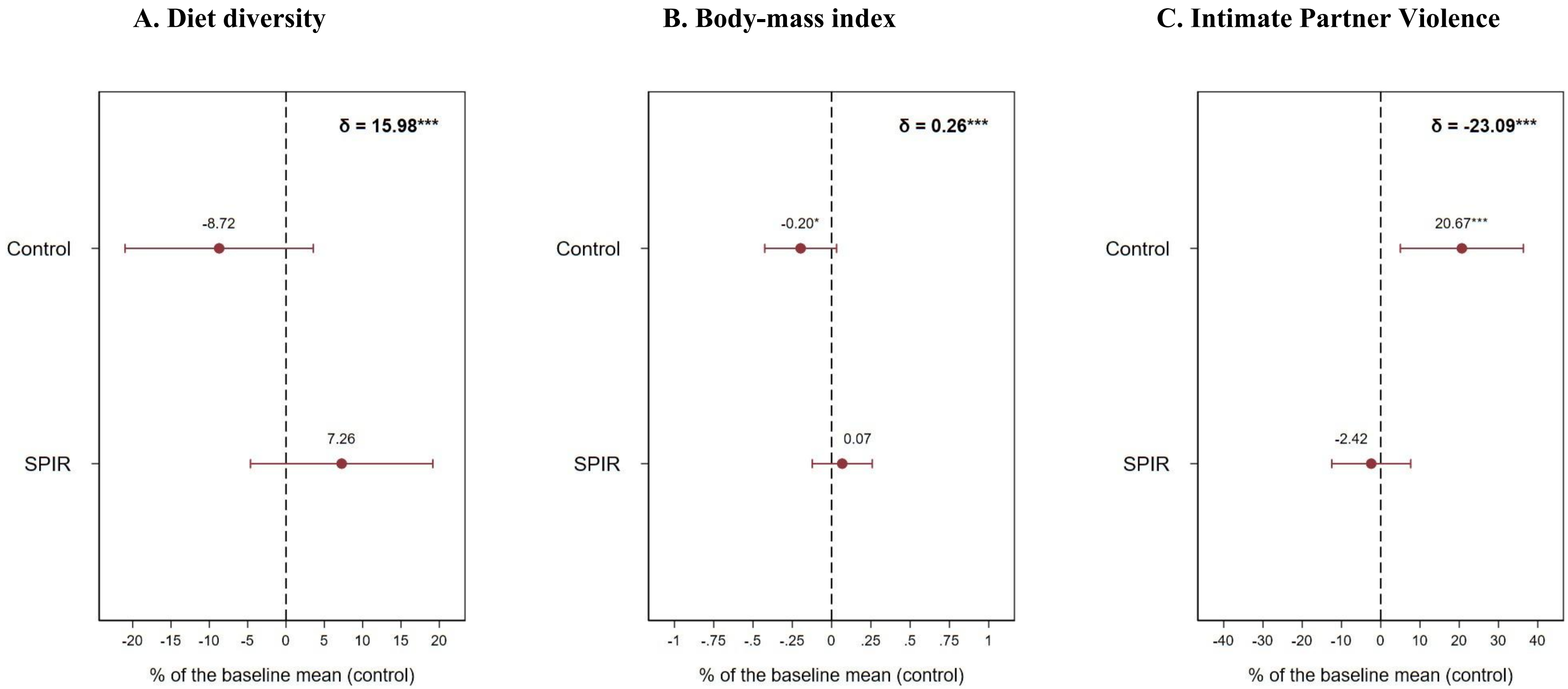
Note: BMI = Body mass index. IPV = Intimate Partner Violence. ‘Dryness’ is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. In column 3, the estimated equation includes a binary variable capturing households for which the baseline value was missing (these missing baseline values were set to zero). The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Q-values corrected for multiple hypothesis testing reported in brackets. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Normalized dryness impact refers to the estimated dryness impact to a 0.25 SD increase in relative dryness.

Table 2. Impact of graduation programming and dryness, using both contemporaneous and lagged dryness measures

	(1)	(2)	(3)	(4)	(5)
Outcome:	Primary woman's diet diversity	BMI of primary woman	Primary woman experienced IPV	Household food gap	Tropical livestock units owned by household
Dryness	-0.749	0.292	0.103	2.125**	-1.277***
	(0.460)	(0.756)	(0.119)	(0.891)	(0.165)
Dryness X SPIR	2.000***	-0.531	-0.124	-1.837*	0.989***
	(0.382)	(0.509)	(0.093)	(0.947)	(0.148)
Lagged dryness	-0.099	-0.205**	0.018	0.511	-0.051**
	(0.100)	(0.093)	(0.033)	(0.390)	(0.024)
Lagged dryness X SPIR	-0.281***	0.298***	-0.014	0.077	0.006
	(0.103)	(0.088)	(0.022)	(0.220)	(0.027)
SPIR	0.084	0.006	0.004	0.047	-0.042**
	(0.093)	(0.021)	(0.008)	(0.091)	(0.019)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome in control group	2.02	20.12	0.17	2.06	0.97
Number of observations	5,877	4,888	3,808	6,212	6,274

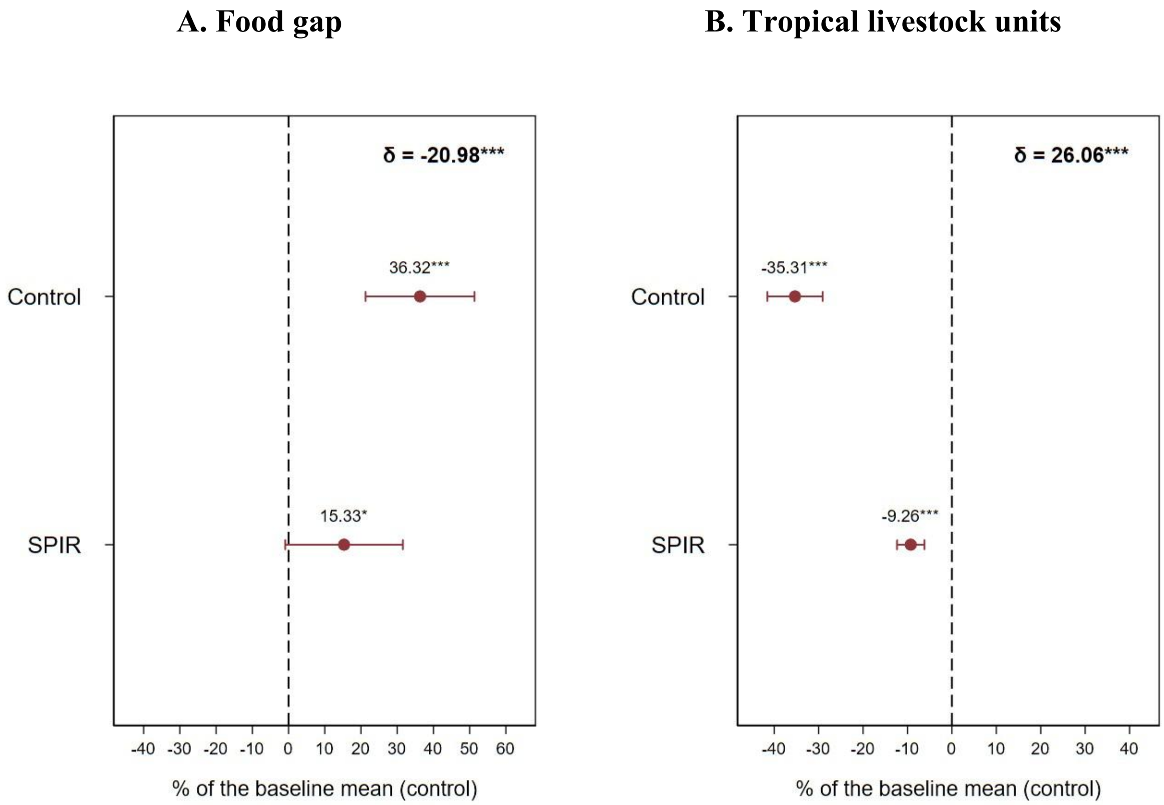
Note: 'Dryness' is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The estimated equation includes a binary variable capturing households for which the baseline value was missing (these missing baseline values were set to zero). The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1. Impact of a 0.25 SD increase in relative dryness on women's outcomes, by treatment status



Note: δ = difference between the estimates (coefficient on the interaction term). Based on Equation (1) with underlying regression results reported in columns 1-3 of Table 1. The capped lines represent 95-% confidence intervals of the point estimates (marked with a solid dot). The number of observations in Panel A is 5,877, in Panel B 4,888 and Panel C 3,808. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2. Impact of a 0.25 SD increase in relative dryness on household level outcomes, by treatment status



Note: δ = difference between the estimates (coefficient on the interaction term). Based on Equation (1) with underlying regression results reported in column 4 and 5 of Table 1. The capped lines represent 95-% confidence intervals of the point estimates (marked with a solid dot). The number of observations in Panel A is 6,212 and in Panel B 6,274. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. Impact of graduation programming and dryness conditions on household savings, credit and livestock sales

	(1)	(2)	(3)
Outcome:	Total amount saved ($PPP)	Any credit	Livestock sales (TLU)
Dryness	14.35	0.229**	0.074
	(16.86)	(0.096)	(0.084)
Dryness X SPIR	-52.89**	-0.316***	-0.290***
	(24.95)	(0.090)	(0.039)
SPIR	20.27***	0.053*	0.018***
	(4.16)	(0.029)	(0.005)
Woreda fixed effects?	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes
Baseline controls?	No	No	Yes, TLU levels
Mean of the outcome in control group	42.53	0.448	0.245
'Dryness' + 'Dryness X SPIR' = 0	p = 0.049	0.356	p = 0.008
Normalized dryness impact in control households	3.59	0.06**	0.02
As % of the control mean	8.44	12.78**	7.59
Normalized dryness impact in SPIR households	-9.64**	-0.02	-0.05***
As % of the control mean	-22.66**	-4.88	-22.07***
Number of observations	5,568	5568	6,274

Note: 'Dryness' is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Normalized dryness impact refers to the estimated dryness impact to a 0.25 SD increase in relative dryness.

Table 4. Cost-effectiveness of SPIR per 0.25 SD dryness episode

Outcome	Protection	95% CI	Unit	Cost/unit ($PPP)
Women's dietary diversity	0.32	[0.24, 0.41]	points	464
Women's BMI	0.05	[0.03, 0.07]	kg/m^2	2,821
Intimate partner violence	4.03	[1.66, 6.39]	percentage points	37
Household food gap	0.43	[0.21, 0.66]	months	346
Livestock holdings	0.25	[0.20, 0.30]	TLU	591

Note: Protection is the Dryness X SPIR interaction coefficient from Table 1, normalized to a 0.25 SD increase in relative dryness. Cost/unit divides the one-time marginal cost of the SPIR savings-group component ($150 per household, 2017 PPP) by the per-episode protection estimate, assuming only a single dryness episode. 95% confidence intervals are based on Conley (1999) standard errors with a 600 km distance cut-off. Cost estimates carry some uncertainty (see Leight et al., 2026). For outcomes where the sum of the Dryness and Dryness X SPIR coefficients is not statistically different from zero (dietary diversity, BMI, IPV), SPIR provides full buffering against the dryness shock.

Online Appendices: Do ultra-poor graduation programs build women's resilience to weather shocks? Evidence from rural Ethiopia

Appendix A. Context

Figure A1. SPIR interventions and surveys

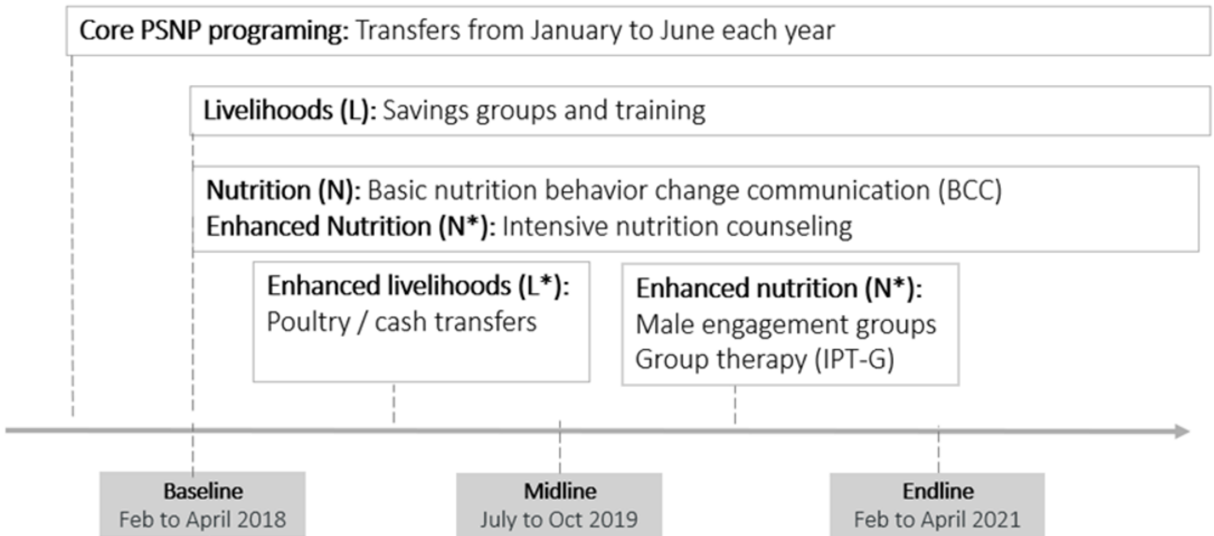
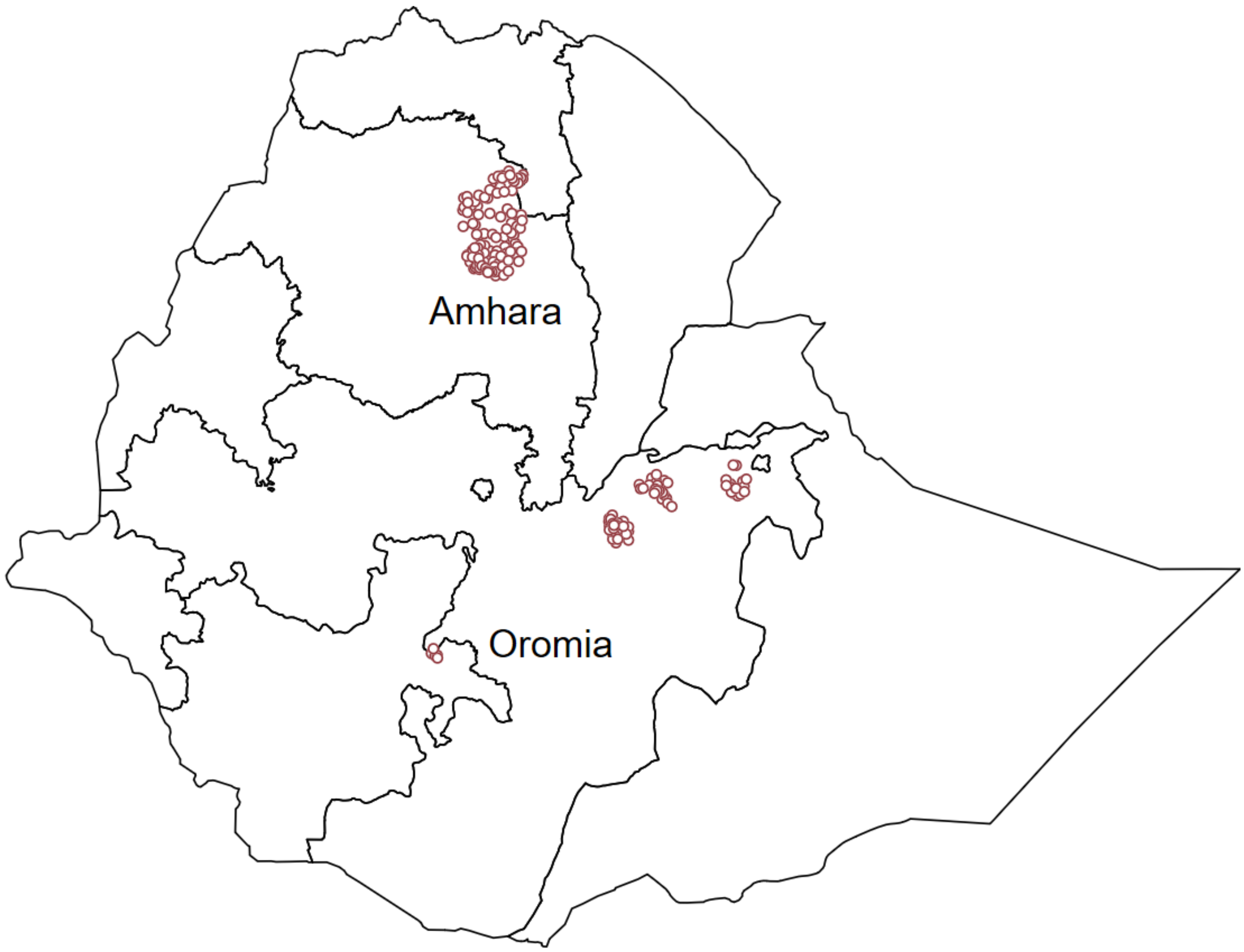
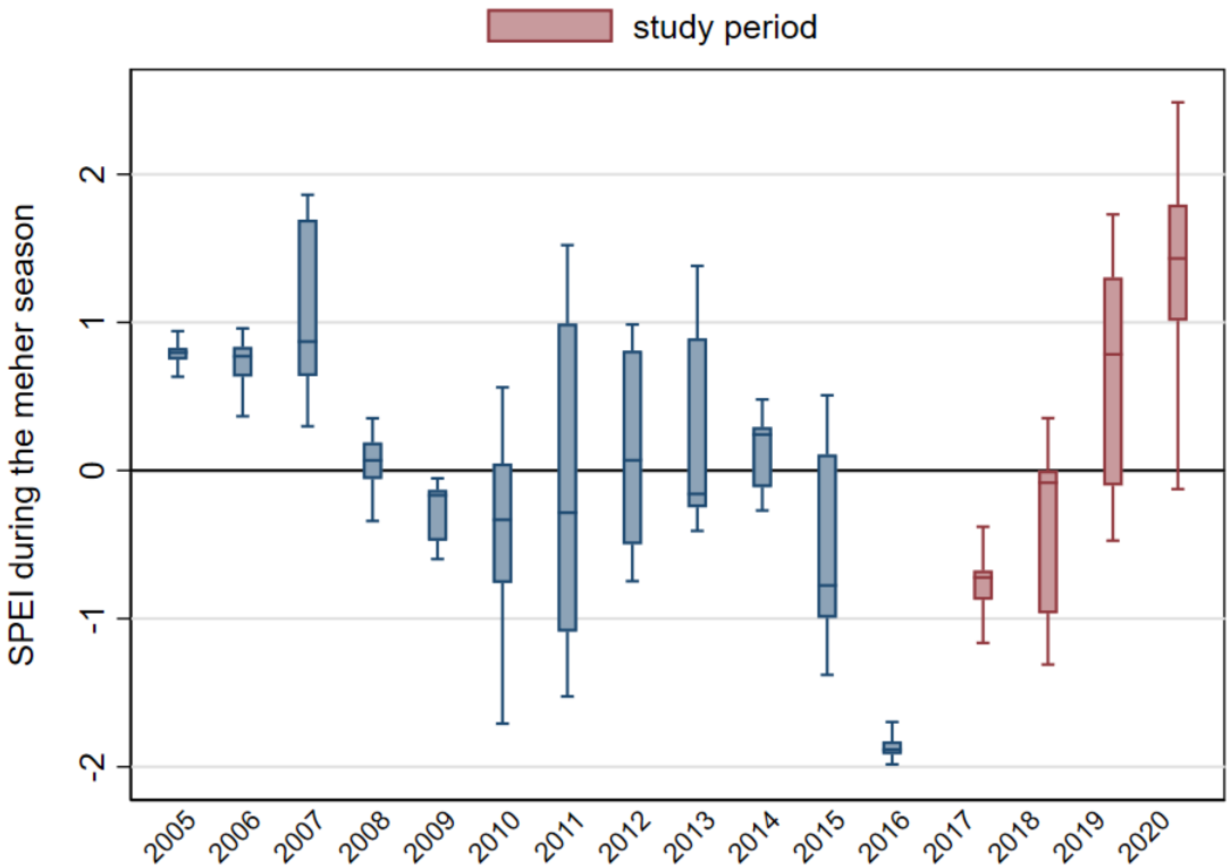


Figure A2. Location of study clusters



Note: 192 study clusters (kebeles): 112 in the Amhara region and 80 in the Oromia region of Ethiopia. Solid black lines represent regional or national boundaries.

Figure A3. SPEI during the meher seasons in 2005–2020



Note: Box plot. The size of the box indicates the difference between the 25th percentile (the bottom part of the box) and the 75th percentile (the top part of the box) of the Standardized Precipitation-Evapotranspiration Index (SPEI) distribution. The bottom and top rule marks the bottom 5th and top 5th percentiles, respectively. The vertical bar rule inside the box shows the median value. N = 192 clusters (kebeles).

Appendix B. IPV definitions and sample selection

Definition of IPV categories

- **Physical spousal violence**: Husband/partner pushed you, shook you, or threw something at you; slapped you; twisted your arm or pulled your hair; punched you with his fist or with something that could hurt you; kicked you, dragged you, or beat you up; tried to choke you or burn you on purpose; or threatened or attacked you with a knife, gun, or any other weapon.
- **Sexual spousal violence**: Husband/partner physically forced you to have sexual intercourse with him even when you did not want to; physically forced you to perform any other sexual acts you did not want to; forced you with threats or in any other way to perform sexual acts you did not want to.
- **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.

Source: Alderman H, Billings L, Gilligan DO, Hidrobo M, Leight J, Taffesse AS, Tambet H. Impact evaluation of the strengthen PSNP4 institutions and resilience (SPIR) development food security activity (DFSA): Endline report. International Food Policy Research Institute (IFPRI): Washington D.C.; 2021.

Table B1. Impact of SPIR treatment on IPV sample selection

Panel A. IPV variable is missing

	(1) Midline	(2) Endline	(3) Both
SPIR	0.032	-0.027	0.003
	(0.034)	(0.029)	(0.023)
Woreda fixed effects	Yes	Yes	Yes
Survey round fixed effects	No	No	Yes
Number of observations	3,012	2,899	5,911
Baseline mean (control group)	0.370	0.370	0.370

Panel B. Lived with husband/partner in the past year

	(1) Midline	(2) Endline	(3) Both
SPIR	0.011	0.031	0.021
	(0.020)	(0.026)	(0.022)
Woreda fixed effects	Yes	Yes	Yes
Survey round fixed effects	No	No	Yes
Number of observations	2,974	2,891	5,865
Baseline mean (control group)	0.813	0.813	0.813

Note: Panel A tests whether SPIR treatment affects the probability that the IPV outcome variable is missing. Panel B tests whether treatment affects the probability the primary woman has lived with a husband/partner in the last 13 months (the first gating question for the IPV module). The sample is restricted to households where the primary woman was surveyed. All specifications include woreda fixed effects. Standard errors (in parentheses) clustered at the kebele level (unit of treatment). Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix C. Additional dryness regressions

Table C1. Impact of dryness on cereal crop production

	(1) Cereal crop production ($PPP)	(2) Cereal crop production ($PPP)
Sample:	*Control clusters*	*SPIR clusters*
Dryness	-105.95**	-107.50***
	(41.63)	(27.17)
Woreda fixed effects?	Yes	Yes
Survey round fixed effects?	Yes	Yes
Outcome variable at the baseline?	Yes	Yes
Mean of the outcome variable	445.06	465.07
Normalized dryness impact	-26.50	-26.88
Number of observations	1,221	3,943

Note: ANCOVA specification with midline and endline survey rounds. The sample is restricted to households that reported cultivating crops during the cropping season. The variable 'Dryness' is positive rectified SPEI (Standardised Precipitation-Evapotranspiration Index), multiplied by -1 so that larger positive values reflect increases in relative dryness (i.e., worse dryness conditions). The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Normalized dryness impact refers to the estimated dryness impact to a 0.25 SD increase in relative dryness.

Table C2. Relationship between SPEI-based dryness measure and self-reported drought shocks

	(1)	(2)	(3)
Sample:	*Control clusters*	*SPIR clusters*	*All clusters*
Dryness	0.40***	0.41***	0.41***
	(0.06)	(0.05)	(0.05)
Woreda fixed effects?	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes
Mean of the outcome variable	0.535	0.505	0.512
Normalized dryness impact	0.100	0.102	0.102
Number of observations	1,201	3,966	5,167

Note: ANCOVA specification with midline and endline survey rounds. At endline, the sample was restricted to households where the primary male respondent was present and available for interview, as the shock module was administered only to this respondent. The variable 'Dryness' is positive rectified SPEI (Standardised Precipitation-Evapotranspiration Index), multiplied by -1 so that larger positive values reflect increases in relative dryness (i.e., worse dryness conditions). The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Normalized dryness impact refers to the estimated dryness impact to a 0.25 SD increase in relative dryness.

Table C3. Impact of dryness on household and women's outcomes, control clusters only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Primary woman's diet diversity		BMI of primary female		Experienced IPV		Household food gap		Tropical livestock units	
Dryness, ongoing meher for the midline	-0.515**		-0.421		0.104*		2.285***		-1.346***	
	(0.201)		(0.587)		(0.062)		(0.409)		(0.256)	
Dryness, previous meher for the midline		-0.056		-0.311***		0.026**		0.678**		-0.277***
		(0.068)		(0.070)		(0.012)		(0.289)		(0.046)
Observations	1392	1392	1168	1168	905	905	1503	1503	1505	1505

Note: 'Dryness' is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The regressions include baseline value of the outcome variable, survey round dummy and woreda fixed effects. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix D. Balance checks

Table D1. Balance in baseline outcomes and dryness conditions across study arms

Variable	Study arm: (1) T1 Mean/*N*	(2) T2 Mean/*N*	(3) T3 Mean/*N*	(4) Control Mean/*N*	t-test p-val (1)-(2)	t-test p-val (1)-(3)	t-test p-val (1)-(4)	t-test p-val (2)-(3)	t-test p-val (2)-(4)
Primary woman's diet diversity	2.103	1.944	2.091	2.024	0.079*	0.905	0.457	0.110	0.390
	775	*817*	*827*	*757*					
BMI of primary female	20.169	20.024	19.993	20.117	0.334	0.276	0.724	0.825	0.475
	766	*814*	*821*	*755*					
Experienced intimate partner violence (IPV)	0.153	0.195	0.182	0.174	0.234	0.422	0.490	0.746	0.554
	524	*564*	*549*	*499*					
Household food gap	2.094	2.200	2.386	2.063	0.692	0.292	0.907	0.502	0.602
	810	*836*	*850*	*791*					
Tropical livestock units owned by household	0.912	0.912	0.884	0.973	0.997	0.735	0.488	0.705	0.438
	812	*857*	*853*	*791*					
Dryness during 2016 meher season [a)]	1.873	1.866	1.865	1.868	0.638	0.550	0.762	0.903	0.897
	812	*858*	*853*	*791*					
Dryness during 2017 meher season [a)]	0.798	0.829	0.809	0.822	0.545	0.821	0.651	0.696	0.903
	812	*858*	*853*	*791*					
Dryness during 2018 meher season [a)]	0.389	0.453	0.436	0.454	0.531	0.650	0.531	0.868	0.989
	812	*858*	*853*	*791*					
Dryness during 2019 meher season [a)]	0.083	0.065	0.081	0.078	0.489	0.946	0.861	0.552	0.606
	812	*858*	*853*	*791*					
Dryness during 2020 meher season [a)]	0.008	0.009	0.008	0.006	0.903	0.989	0.830	0.889	0.749
	812	*858*	*853*	*791*					

Note: The value displayed for t-tests are p-values based on standard errors clustered at the cluster (kebele) level. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The numbers in italics are the number of non-missing observations. T1, T2 and T3 are SPIR treatment arms where T1 received the intensive livestock (L*) and intensive nutrition intervention (N*), T2 received the intensive livestock intervention (L*) and standard nutrition intervention (N), and T3 received standard livestock (L) and intensive nutrition intervention (N*). Control are households located in clusters that did not receive any SPIR interventions.
[a)] 'Dryness' is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions.

Table D2. Balance in baseline characteristics with respect to dryness shocks prior midline and endline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log household per capita consumption at the baseline	0.000						-0.001
	(0.001)						(0.001)
Household size at the baseline		-0.002*					-0.002*
		(0.001)					(0.001)
Household head had some education at the baseline			0.003*				0.003*
			(0.001)				(0.001)
Household head's age at the baseline				-0.000			0.000*
				(0.000)			(0.000)
Food gap (months) at the baseline					0.000		0.000
					(0.000)		(0.000)
Tropical livestock units owned at the baseline						-0.000	0.000
						(0.000)	(0.001)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of the outcome variable	0.042	0.042	0.042	0.042	0.042	0.042	0.042
Observations	6550	6628	6622	6624	6574	6626	6526

Note: OLS regression. Outcome variable is dryness shock prior midline and endline. The dryness shock variable is based on positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix E. Spatial correlation and identifying variation in dryness

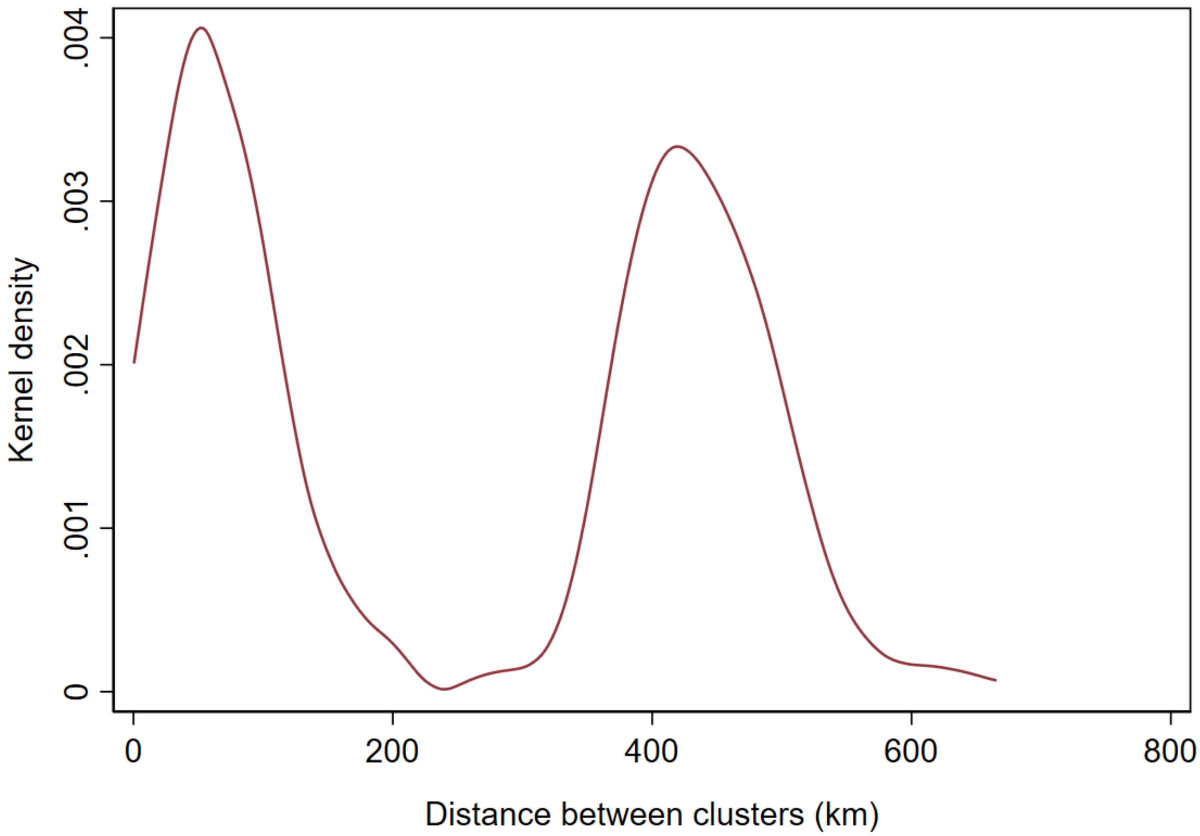
Spatial correlation across study clusters

To explore spatial dependencies in our data, we first calculated the Euclidean distances between all cluster pairs. The mean distance is 256 km, and the median is 308 km. However, these averages hide the fact that the density distribution of the distances is bimodal (Figure E1). Given that our study clusters are in two regions, the cluster pairs are located either between 0 and 200 km or between 300 and 600 km from each other.

We expect that the weather outcomes in each agricultural season are highly correlated between clusters that are located close to each other.[1] To verify this, we computed the SPEI for each meher season between 1990 and 2021 for all clusters and used these data to calculate the correlation coefficient of the meher season SPEI between all cluster-pairs. We then used these dyadic data to explore the relationship between these correlation coefficients and distances between clusters. Figure E2 shows the results based on a local polynomial regression that regresses the correlation coefficient on the distance variable. As expected, the correlation coefficient is close to 1 for clusters located very close to each other. As we move to clusters that are farther apart, the correlation decreases. For clusters located 600 km apart from each other, the correlation coefficient is 0.2, on average, suggesting a relatively weak spatial correlation in meher season weather outcomes.

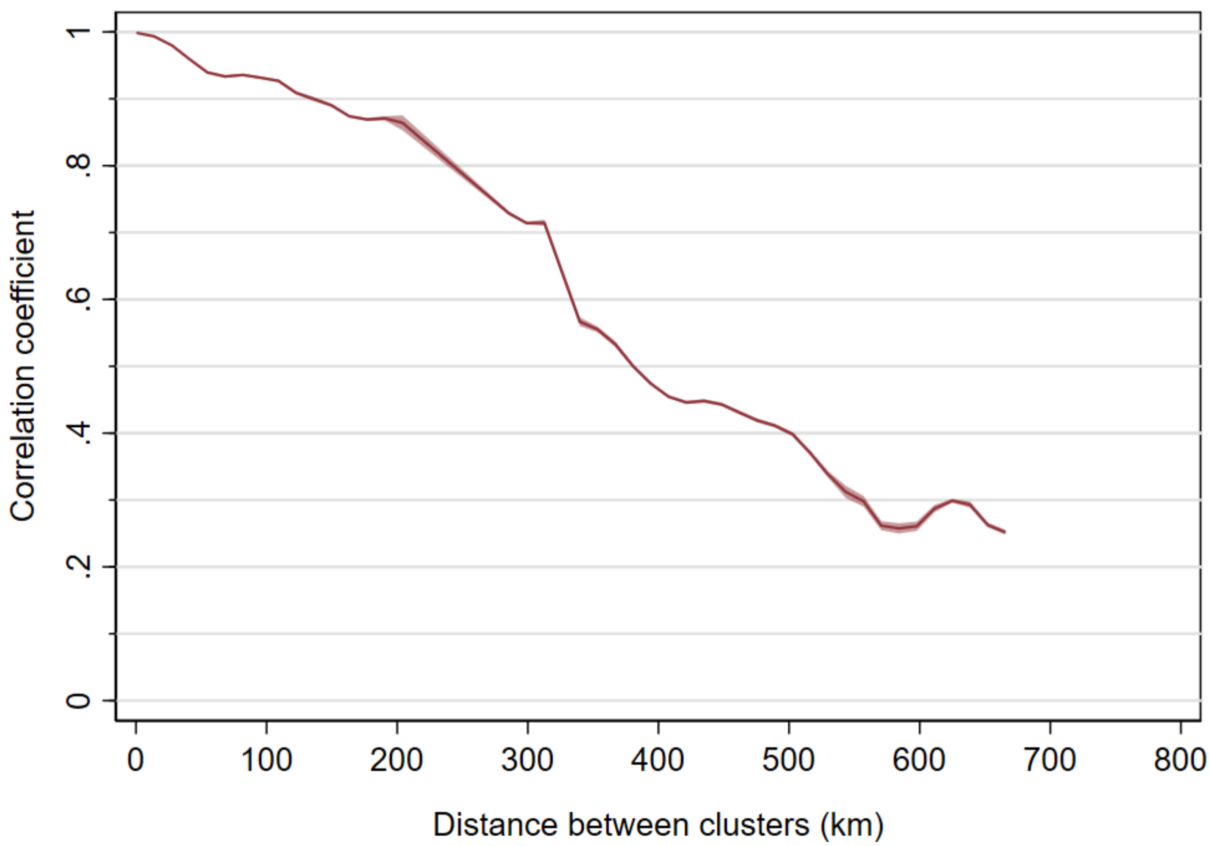
[1] Day-to-day weather outcomes can vary within small geographical areas and with high resolution climate data one may be able to detect this variation. However, when aggregated over an entire agricultural season, the spatial autocorrelation tends to be very high.

Figure E1. Density distribution of distances between study clusters



Note: Kernel density estimator. Based on dyadic data with 18,336 observations from 192 clusters (192*191/2=18,336). The vertical axis measures kernel density. The horizontal axis measures the Euclidean distance in km between the cluster dyads.

Figure E2. Relationship between meher season SPEI correlation coefficients and distance between study clusters



Note: Local polynomial regression. Shaded areas represent 95-% confidence intervals. Based on dyadic data with 18,336 observations from 192 clusters (192*191/2=18,336). The vertical axis measures the correlation coefficient in meher season SPEI in 1990-2021 between all possible cluster dyads. The horizontal axis measures the Euclidean distance in km between the cluster dyads.

Identifying variation in dryness

Figure E3: SPEI (absolute values) by cluster and year

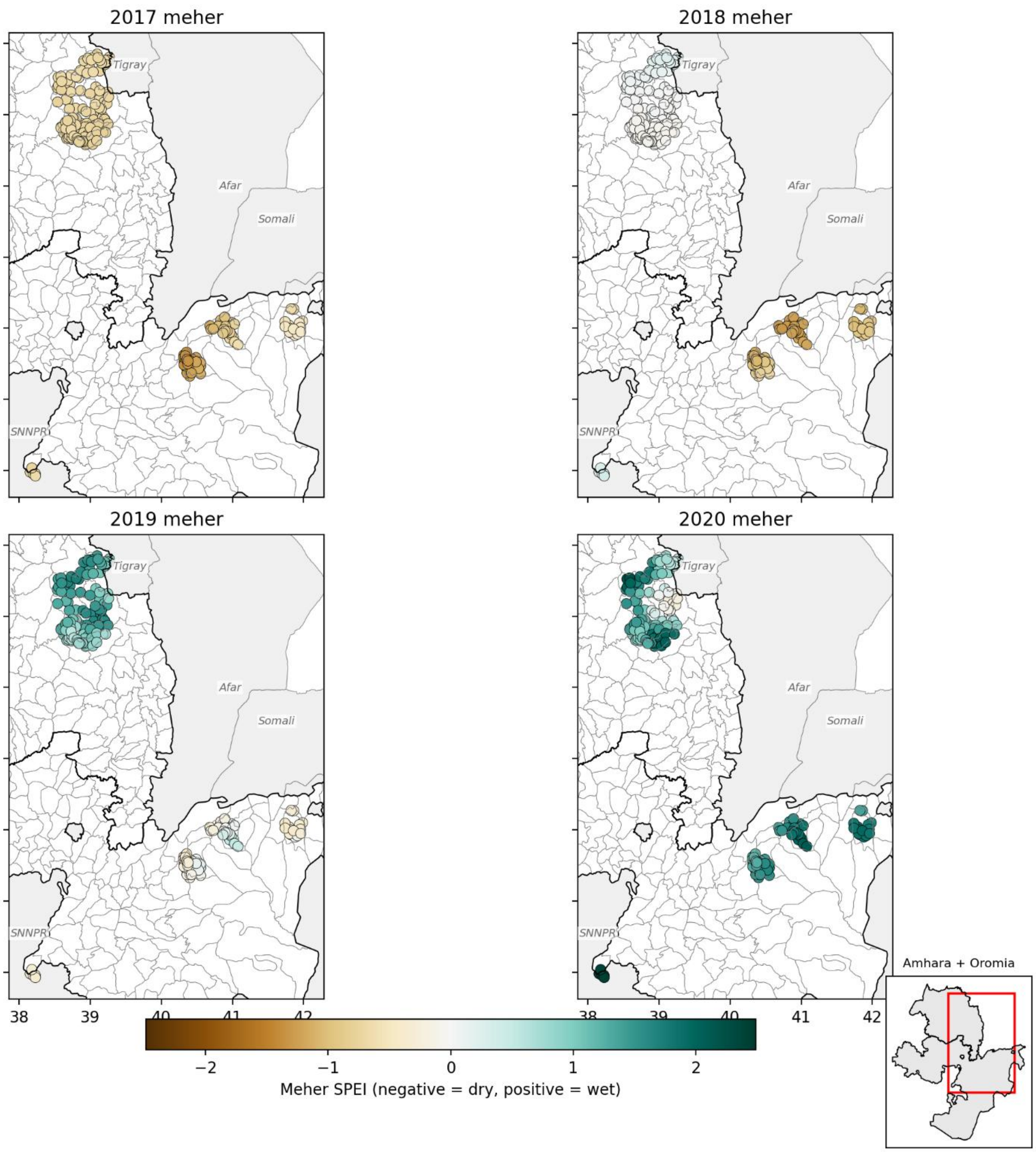
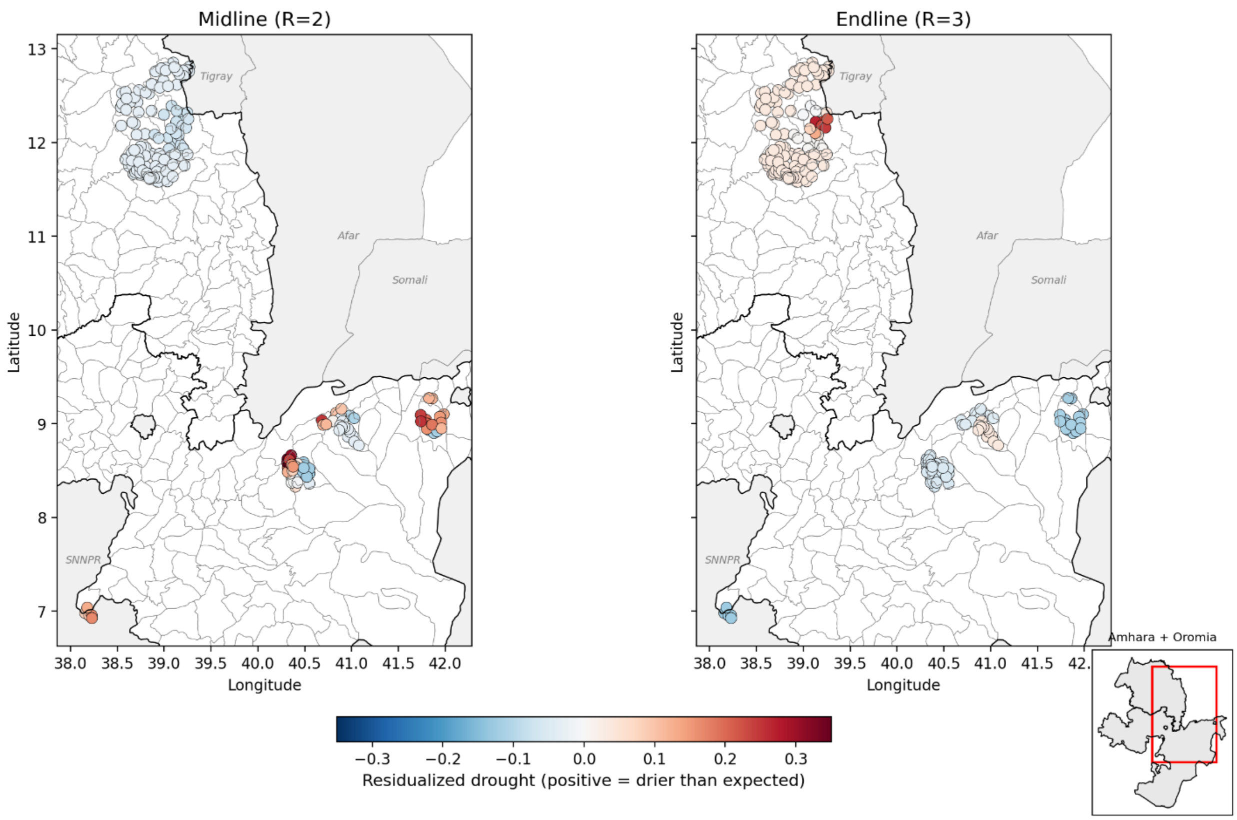


Figure E4: SPEI (residualized values) by cluster and survey round



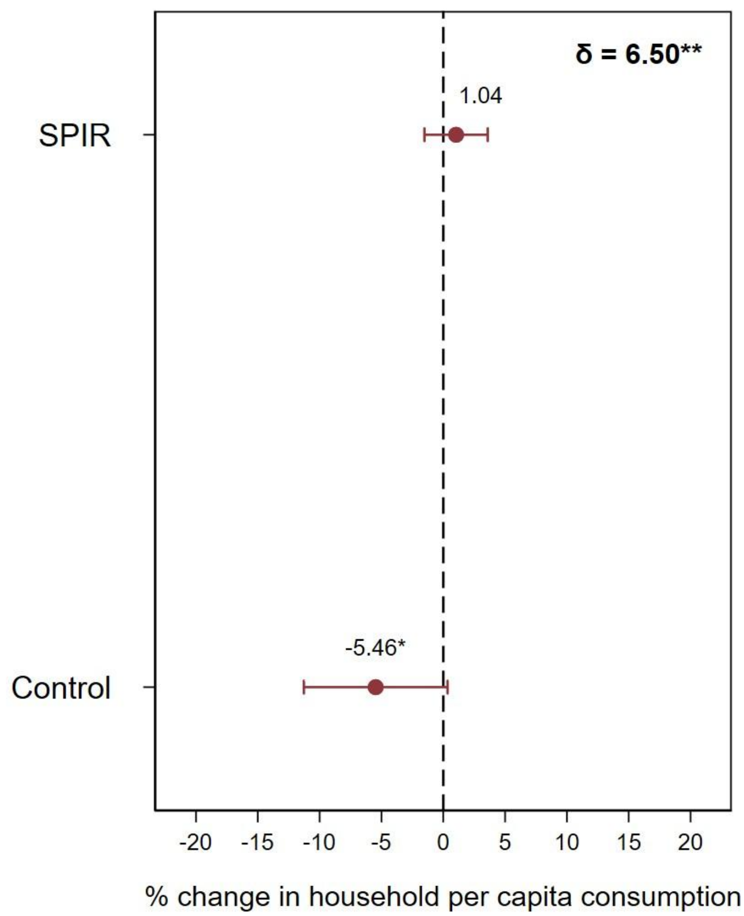
Appendix F. Extensions

Table F1. Impact of graduation programming and dryness conditions on different forms of IPV

	(1)	(2)	(3)	(4)
Outcome:	Any IPV	Physical	Sexual	Emotional
Dryness	0.144***	0.068	0.128***	0.086*
	(0.056)	(0.044)	(0.020)	(0.046)
Dryness X SPIR	-0.161***	-0.080*	-0.099***	-0.144**
	(0.048)	(0.044)	(0.028)	(0.061)
SPIR	0.007	-0.003	0.003	0.012**
	(0.006)	(0.004)	(0.002)	(0.006)
Woreda fixed effects?	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes
Baseline mean of the outcome in control group	0.17	0.06	0.04	0.13
‘Dryness’ + ‘Dryness X SPIR’ = 0	p = 0.637	p = 0.374	p = 0.139	p = 0.150
Normalized dryness impact in control households	0.04***	0.02	0.03***	0.02*
Normalized dryness impact in SPIR households	-0.00	-0.00	0.01	-0.01
Number of observations	3,808	3,808	3,808	3,808

Note: ‘Dryness’ is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The estimated equation includes a binary variable capturing households for which the baseline value was missing (these missing baseline values were set to zero). The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Normalized dryness impact refers to the estimated dryness impact to a 0.25 SD increase in relative dryness.

Figure F1. Impact of a 0.25 SD increase in relative dryness in 2019-meher on household per capita consumption in 2021, by treatment status



Note: δ = difference between the estimates (coefficient on the interaction term). The underlying regression results are reported in Table F2. The outcome variable is log household consumption in adult equivalents. The capped lines represent 95-% confidence intervals of the point estimates (marked with a solid dot). Endline data only: the number of observations is 2,962. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F2. Impact of graduation programming and dryness conditions on household per capita consumption

	(1)
Outcome:	Log household consumption in adult equivalents
Dryness (in 2019)	-0.219*
	(0.119)
Dryness X SPIR	0.260**
	(0.114)
SPIR	-0.042
	(0.029)
Woreda fixed effects?	Yes
Survey round fixed effects?	No
Outcome variable at the baseline?	Yes
‘Dryness’ + ‘Dryness X SPIR’ = 0	p = 0.427
Normalized dryness impact in control households	-0.055*
Normalized dryness impact in SPIR households	0.010
Number of observations	2,962

Note: ANCOVA specification without midline data. ‘Dryness’ is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Normalized dryness impact refers to the estimated dryness impact to a 0.25 SD increase in relative dryness.

Table F3. Impact of graduation programming and dryness conditions on household and individual level outcomes, by treatment arm

	(1)	(2)	(3)	(4)	(5)
Outcome:	Diet diversity	Body-mass index	IPV	Food gap in months	Tropical livestock units
Dryness	-0.709	-0.158*	0.143**	3.007***	-1.366***
	(0.503)	(0.094)	(0.056)	(0.632)	(0.119)
Dryness X T1: L* + N*	1.571***	0.281***	-0.216***	-1.918***	1.259***
	(0.346)	(0.062)	(0.037)	(0.402)	(0.156)
Dryness X T2: L* + N	0.988***	0.182***	-0.141	-0.891	1.048***
	(0.175)	(0.069)	(0.093)	(0.788)	(0.174)
Dryness X T3: L + N*	1.209***	0.200***	-0.145**	-2.201***	0.774***
	(0.179)	(0.064)	(0.058)	(0.454)	(0.128)
T1: L* + N*	0.055	0.029	0.019**	0.180***	0.003
	(0.039)	(0.049)	(0.008)	(0.050)	(0.026)
T2: L* + N	-0.052	-0.065	-0.010**	0.018	-0.052**
	(0.077)	(0.089)	(0.005)	(0.065)	(0.020)
T3: L + N*	0.097*	-0.062*	0.016*	-0.000	-0.090***
	(0.051)	(0.036)	(0.009)	(0.143)	(0.019)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome variable in the control group	2.02	20.12	0.17	2.06	0.97
‘Dryness X T1’ = ‘Dryness X T2’	p = 0.177	p = 0.212	p = 0.386	p = 0.060	p = 0.211
‘Dryness X T1’ = ‘Dryness X T3’	p = 0.219	p = 0.388	p = 0.123	p = 0.585	p = 0.030
‘Dryness X T2’ = ‘Dryness X T3’	p = 0.267	p = 0.850	p = 0.963	p = 0.024	p = 0.065
Number of observations	5,877	4,888	3,808	6,212	6,274

Note: ‘Dryness’ is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. T1, T2 and T3 are SPIR treatment arms where T1 received the intensive livestock (L*) and intensive nutrition intervention (N*), T2 received the intensive livestock intervention (L*) and standard nutrition intervention (N), and T3 received standard livestock (L) and intensive nutrition intervention (N*). The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F4. Impact of graduation programming and dryness conditions on household and individual level outcomes in poorest households eligible for livelihood transfers, by treatment arm

	(1)	(2)	(3)	(4)	(5)
Outcome:	Diet diversity	Body-mass index	IPV	Food gap in months	Tropical livestock units
Dryness	-0.763	-0.120	0.359***	3.050***	-0.972***
	(0.598)	(0.092)	(0.050)	(0.770)	(0.149)
Dryness X T1: L* + N*	1.422***	0.371***	-0.235***	-2.439***	0.855***
	(0.423)	(0.118)	(0.077)	(0.549)	(0.208)
Dryness X T2: L* + N	1.130***	0.173*	-0.435***	-1.356*	0.737***
	(0.245)	(0.091)	(0.062)	(0.731)	(0.195)
Dryness X T3: L + N*	1.005***	0.281***	-0.270***	-2.207***	0.407***
	(0.181)	(0.042)	(0.102)	(0.669)	(0.060)
T1: L* + N*	0.092	0.007	0.025**	0.303***	0.077*
	(0.082)	(0.094)	(0.012)	(0.087)	(0.042)
T2: L* + N	-0.020	-0.028	0.006	0.006	0.015
	(0.129)	(0.086)	(0.010)	(0.131)	(0.016)
T3: L + N*	0.062	-0.056	0.026**	0.074	-0.110***
	(0.040)	(0.036)	(0.013)	(0.191)	(0.018)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome variable in the control group	1.95	20.18	0.18	2.13	0.65
‘Dryness X T1’ = ‘Dryness X T2’	p = 0.582	p = 0.087	p = 0.004	p = 0.057	p = 0.622
‘Dryness X T1’ = ‘Dryness X T3’	p = 0.294	p = 0.457	p = 0.608	p = 0.608	p = 0.035
‘Dryness X T2’ = ‘Dryness X T3’	p = 0.685	p = 0.278	p = 0.011	p = 0.078	p = 0.077
Number of observations	3,374	2,826	2,015	3,553	3,590

Note: Sample restricted to the poorest 55 % of households at the baseline that were eligible for the livelihood transfer. ‘Dryness’ is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. T1, T2 and T3 are SPIR treatment arms where T1 received the intensive livestock (L*) and intensive nutrition intervention (N*), T2 received the intensive livestock intervention (L*) and standard nutrition intervention (N), and T3 received standard livestock (L) and intensive nutrition intervention (N*). The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F5. Impact of graduation programming and dryness conditions on male and female stress levels and women's bargaining power

	(1)	(2)	(3)
Outcome:	Male stress levels	Female stress levels	Women's input in productive decisions
Dryness	6.331***	5.532***	-0.503***
	(0.910)	(1.035)	(0.071)
Dryness X SPIR	-0.725*	-0.976***	0.309***
	(0.432)	(0.370)	(0.058)
SPIR	0.135*	0.158***	0.035***
	(0.075)	(0.044)	(0.009)
Woreda fixed effects?	Yes	Yes	Yes
Survey round fixed effects?	No	No	Yes
Outcome variable at the baseline?	No	No	Yes
Mean of the outcome in control group	5.32	5.19	0.55
'Dryness' + 'Dryness X SPIR' = 0	p<0.001	p<0.001	p<0.001
Normalized dryness impact in control households	1.583***	1.383***	-0.13***
As % of the control mean	29.78***	26.67***	-22.68***
Normalized dryness impact in SPIR households	1.402***	1.139***	-0.05***
As % of the control mean	26.37***	21.96***	-8.74***
Number of observations	1,981	3,011	4,882

Note: Columns 1 and 2: Endline data only; dryness variable is based on 2019 meher. The number of observations in column 1 is lower because males were not always present in the household during the interview. Column 3: The outcome is a binary indicator equal to 1 if the woman reports having input into most or all decisions about use of income generated from productive activities (crop cultivation, livestock raising, or poultry raising). The indicator equals 1 if the woman meets the threshold in at least one activity she participates in. 'Dryness' is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F6. Impact of graduation programming on agricultural investments and practices

	(1)	(2)	(3)	(4)	(5)
Outcome:	Seeds (0/1)	Inorganic fertilizer (0/1)	Pesticides (0/1)	Hired labor (0/1)	Crop diversity (count)
SPIR	0.003	0.031	-0.028	-0.004	-0.035
	(0.032)	(0.031)	(0.027)	(0.018)	(0.056)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey rounds	ML	ML & EL	ML & EL	ML & EL	ML & EL
Survey round fixed effects?	No	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome variable in the control group	0.229	0.357	0.037	0.078	2.06
Number of observations	2,574	5,077	2517	5,087	5,124

	(6)	(7)	(8)	(9)
Outcome:	Seeds ($PPP)	Inorganic fertilizer ($PPP)	Pesticides ($PPP)	Hired labor ($PPP)
SPIR	1.67	-0.12	-0.50	-3.16
	(1.71)	(2.19)	(0.54)	(11.14)
Woreda fixed effects?	Yes	Yes	Yes	Yes
Survey rounds	ML	ML	ML	ML
Survey round fixed effects?	No	No	No	No
Outcome variable at the baseline?	Yes	Yes	Yes	Yes
Baseline mean of the outcome variable in the control group	7.27	21.77	0.44	3.85
Number of observations	2574	2560	2517	2576

Note: An ANCOVA specification. The sample is restricted to households that cultivated crops during the previous cropping season. The outcome variable in columns 1 to 4 is binary, obtaining value 1 if the household reported expenditures on the input category in the previous cropping season, zero otherwise. The outcome variable in column 5 is a count capturing the number of different crops cultivated by the household in the previous cropping season. The outcome variables in columns 6 to 9 are continuous, capturing the input expenditure amounts in 2017 $PPP. These outcomes were not measured at the endline. The standard errors are reported in parentheses and clustered at the kebele level (unit of treatment). ML = Midline survey; EL = Endline survey. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F7. Impact of graduation programming on agricultural investments and practices (all households)

	(1)	(2)	(3)	(4)	(5)
Outcome:	Seeds (0/1)	Inorganic fertilizer (0/1)	Pesticides (0/1)	Hired labor (0/1)	Crop diversity (count)
SPIR	0.014	0.006	0.029	-0.023	-0.001
	(0.021)	(0.029)	(0.028)	(0.022)	(0.018)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey rounds	ML	ML & EL	ML & EL	ML & EL	ML & EL
Survey round fixed effects?	No	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome variable in the control group	0.779	0.178	0.278	0.029	0.061
Number of observations	6,236	3,154	6,227	3,095	6,239

	(6)	(7)	(8)	(9)
Outcome:	Seeds ($PPP)	Inorganic fertilizer ($PPP)	Pesticides ($PPP)	Hired labor ($PPP)
SPIR	1.39	-0.43	-0.41	-2.45
	(1.51)	(1.88)	(0.42)	(9.07)
Woreda fixed effects?	Yes	Yes	Yes	Yes
Survey rounds	ML	ML	ML	ML
Survey round fixed effects?	No	No	No	No
Outcome variable at the baseline?	Yes	Yes	Yes	Yes
Baseline mean of the outcome variable in the control group	5.66	16.95	0.34	3.00
Number of observations	3,154	3,140	3,095	3,156

Note: Standard errors (in parentheses) clustered at the kebele level. The sample is all households. The outcome variable in columns 1 to 4 is binary, obtaining value 1 if the household reported expenditures on the input category in the previous cropping season, zero otherwise. The outcome variable in column 5 is a count capturing the number of different crops cultivated by the household in the previous cropping season. The outcome variables in columns 6 to 9 are continuous, capturing the input expenditure amounts in 2017 $PPP. These outcomes were not measured at the endline. The standard errors are reported in parentheses and clustered at the kebele level (unit of treatment). ML = Midline survey; EL = Endline survey. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix G. Robustness checks

We performed a range of robustness checks to explore the sensitivity of these findings. First, previous work examining the impacts of droughts in Ethiopia and elsewhere in Africa have often quantified droughts in terms of deviation from long-term precipitation during the cropping season (Hirvonen et al., 2020; Shively, 2017; Thiede, 2014). We obtained monthly precipitation data from the TerraClimate database (Abatzoglou et al., 2018) and computed Z-score deviations of precipitation (rainfall) during the meher season. As with the SPEI indicator, we created a rainfall shock variable by rectifying positive Z-score values to zero and multiplied this positive rectified rainfall variable with -1 so that larger positive values indicate worsening conditions. Table G1 shows that this alternative dryness indicator yields qualitatively similar results: rainfall shocks are highly damaging in control clusters and less so in treated clusters.

Second, a potential concern in including the woreda fixed effects in the estimated equation is that they may absorb a large amount of variation in the dryness indicator (Fisher et al., 2012). To check this, we regressed the dryness indicator on the woreda dummies and the survey round dummy. This regression yields an R^2 of 0.36, indicating that there remains considerable variation in the dryness indicator after controlling for woreda and survey round fixed effects. In Table G2, we further show that our results are robust to replacing woreda fixed effects with a binary indicator capturing the region in which the study cluster is located.

Third, the survey round dummy aims to control for shocks that are common to all households surveyed at the same time (e.g., the COVID-19 pandemic). However, in Ethiopian context, covariate shocks can be highly region-specific raising a concern that our dryness shock variable could be capturing the impact of some other covariate shocks instead. To address this, we replace the survey round dummy with region-specific time fixed effects. As before, the results are robust to this adjustment (Table G3).

Fourth, a concern with the dryness × treatment interaction is that it may reflect heterogeneous treatment effects along dimensions that happen to correlate spatially with dryness, rather than genuine dryness buffering. To address this, we augment the main specification with a vector of baseline household characteristics (log consumption per capita, household size, head's education, head's age, food gap, and livestock holdings) and their interactions with the SPIR treatment indicator. Table G4 shows that the dryness × SPIR coefficients are virtually unchanged, even though the baseline characteristic × treatment interactions are jointly significant ($p < 0.001$) in all columns.

Fifth, since approximately three-quarters of the identifying variation in dryness is temporal (see Section 4: Econometric approach in the main text), the dryness × treatment interaction could in principle capture a treatment effect that evolves over time rather than a differential response to weather. We test this by adding a SPIR × endline interaction to the main specification. Table G5 shows that the dryness × SPIR coefficients remain significant across all outcomes, while the SPIR × endline coefficient is mostly insignificant.

Sixth, the 2020–2021 period was highly volatile, with the COVID-19 pandemic, desert locust infestations that devastated harvests, and a two-year civil war beginning in November 2020 that severely affected SPIR sites in the Amhara region (Alderman et al., 2021). If dryness shocks are correlated with these other covariate shocks, then our dryness impact estimates may be biased. To explore this, we merged the household survey data with remote sensed data on locust invasions (FAO, 2023) and conflict (Raleigh et al., 2010). Table G6 then replicates the main regressions, adding controls for locust and conflict shocks. The estimates are nearly identical to those reported in Table 1, indicating that the story documented here is not driven by other major shocks that occurred in these localities during the study period.

Seventh, previous research has found that dryness shocks shape migration patterns in rural Ethiopia (Gray and Mueller, 2012). To explore whether changes in household composition (e.g., through in- or out-migration) could explain our findings, we re-estimated equation (1) using household size as the dependent variable. As Table G7shows, we cannot reject the null hypothesis that dryness shocks do not change household composition in control clusters, when measured in terms of number of members (column 1) or in terms of adult equivalents (column 3). Moreover, the coefficient on the SPIR interaction term is not statistically different from zero (columns 2 and 4), indicating no differential dryness impacts between treatment and control clusters.

Eighth, we used a 600 km distance cut-off to calculate Conley (1999) standard errors. Our main findings – that dryness have a negative effect, particularly on livestock holdings and food security, and increase the risk of IPV, while SPIR helps to mitigate these effects – remain robust to alternative distance cut-offs (50, 100, …, 700 km). Panels A to E in Figure G1 report the p-values for the key estimates based on these alternative cut-offs.

Finally, the 2016 drought was particularly severe in our study clusters, raising a question of whether our results are driven by households' exposure to the 2016 drought through serially correlated shocks. Below we show that the dryness shocks that occurred during the study period in 2019-2021 are only weakly correlated with the variation in household's exposure to the 2016 drought (Figures G2 to G4). Moreover, our results are robust to controlling for drought conditions in the 2016 meher season (Table G8). It remains possible that the 2016 drought forced households to draw down buffer stocks, reducing their resilience to later shocks. Unfortunately, because exposure to the 2016 drought exhibits very limited cross-sectional variation in our sample (see Figure A3), we cannot assess heterogeneity along this dimension.

Table G1. Replicating Table 1, but using rainfall Z-score instead of SPEI to define dryness

	(1)	(2)	(3)	(4)	(5)
Outcome:	Primary woman's diet diversity	BMI of primary woman	Primary woman experienced IPV	Household food gap	Tropical livestock units owned by household
Dryness	-1.106***	-0.149	0.327***	2.460***	-1.399***
	(0.318)	(0.099)	(0.092)	(0.281)	(0.478)
Dryness X SPIR	0.963***	0.197***	-0.328***	-1.135	1.105***
	(0.360)	(0.052)	(0.114)	(0.723)	(0.347)
SPIR	0.062	-0.032	0.009	0.021	-0.035*
	(0.060)	(0.040)	(0.006)	(0.067)	(0.018)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome in control group	2.02	20.12	0.17	2.06	0.97
‘Dryness’ + ‘Dryness X SPIR’ = 0	p = 0.824	p = 0.530	p = 0.985	p = 0.037	p = 0.048
Number of observations	5,877	4,888	3,808	6,212	6,274

Note: ‘Dryness’ is rainfall z-score multiplied by -1 and then negative rectified at zero so that larger (positive) values reflect worse rainfall conditions. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table G2. Replicating Table 1, but using region fixed effects instead of woreda fixed effects

	(1)	(2)	(3)	(4)	(5)
Outcome:	Primary woman's diet diversity	BMI of primary woman	Primary woman experienced IPV	Household food gap	Tropical livestock units owned by household
Dryness	-0.908**	-0.075	0.189***	3.743***	-1.290***
	(0.405)	(0.077)	(0.065)	(0.548)	(0.148)
Dryness X SPIR	1.186***	0.216***	-0.140***	-1.685***	0.991***
	(0.229)	(0.036)	(0.035)	(0.510)	(0.114)
SPIR	0.052	0.015	0.002	0.039	-0.046***
	(0.057)	(0.046)	(0.007)	(0.050)	(0.016)
Region fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome in control group	2.02	20.12	0.17	2.06	0.97
‘Dryness’ + ‘Dryness X SPIR’ = 0	p = 0.218	p = 0.055	p = 0.299	p < 0.001	p = 0.001
Number of observations	5,877	4,888	3,808	6,212	6,274

Note: ‘Dryness’ is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table G3. Replicating Table 1, but using region-by-survey round fixed effects instead of survey round fixed effects

	(1)	(2)	(3)	(4)	(5)
Outcome:	Primary woman's diet diversity	BMI of primary woman	Primary woman experienced IPV	Household food gap	Tropical livestock units owned by household
Dryness	-0.386	-0.093	0.173**	1.530**	-1.369***
	(0.610)	(0.336)	(0.068)	(0.661)	(0.136)
Dryness X SPIR	1.286***	0.213***	-0.162***	-1.716***	1.015***
	(0.165)	(0.037)	(0.049)	(0.436)	(0.104)
SPIR	0.032	-0.032	0.007	0.071	-0.046***
	(0.053)	(0.038)	(0.006)	(0.062)	(0.017)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Region X survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome in control group	2.02	20.12	0.17	2.06	0.97
‘Dryness’ + ‘Dryness X SPIR’ = 0	$p = 0.119$	$p = 0.699$	$p = 0.806$	$p = 0.782$	$p < 0.001$
Number of observations	5877	4888	3808	6212	6274

Note: ‘Dryness’ is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table G4. Main results controlling for baseline characteristics x treatment interactions

	(1) Primary woman's diet diversity	(2) BMI of primary woman	(3) Primary woman experienced IPV	(4) Household food gap	(5) Tropical livestock units owned by household
Dryness	-0.809*	-0.185**	0.121*	3.095***	-1.257***
	(0.467)	(0.074)	(0.069)	(0.587)	(0.161)
Dryness X SPIR	1.367***	0.257***	-0.127**	-1.778***	0.902***
	(0.105)	(0.033)	(0.063)	(0.412)	(0.150)
SPIR	0.384	-0.739**	-0.096	-0.026	-0.516***
	(0.375)	(0.366)	(0.064)	(0.607)	(0.141)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Baseline outcome?	Yes	Yes	Yes	Yes	Yes
Baseline X and X x SPIR controls?	Yes	Yes	Yes	Yes	Yes
Joint p-value: X x SPIR = 0	0.000	0.000	0.000	0.000	0.000
Observations	5828	4851	3762	6179	6192

Note: Each column augments the corresponding Table 1 specification with a vector of baseline household characteristics (log consumption per capita, household size, head's education, head's age, food gap, Tropical Livestock Units; TLU) and their interactions with the SPIR treatment indicator. For food gap and TLU, the own-baseline variable is excluded from the added vector (already included as baseline outcome control). Joint p-value tests the null that all baseline household characteristics x SPIR interaction coefficients are jointly zero. Standard errors (in parentheses) based on Conley (1999) with a 600 km distance cut-off. Statistical significance denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table G5. Main results with treatment x survey round interaction

	(1) Primary woman's diet diversity	(2) BMI of primary woman	(3) Primary woman experienced IPV	(4) Household food gap	(5) Tropical livestock units owned by household
Dryness	-0.973***	-0.218**	0.117*	3.108***	-1.445***
	(0.315)	(0.092)	(0.071)	(0.545)	(0.153)
Dryness X SPIR	1.630***	0.287***	-0.126*	-1.874***	1.104***
	(0.382)	(0.045)	(0.065)	(0.514)	(0.141)
SPIR	-0.080	-0.108**	-0.004	0.115**	-0.076***
	(0.052)	(0.043)	(0.009)	(0.048)	(0.024)
SPIR X Endline	0.197	0.120*	0.018	-0.086	0.053***
	(0.147)	(0.071)	(0.012)	(0.079)	(0.019)
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Baseline outcome?	Yes	Yes	Yes	Yes	Yes
SPIR X Endline interaction?	Yes	Yes	Yes	Yes	Yes
Observations	5877	4888	3808	6212	6274

Note: Each column augments the corresponding Table 1 specification with an interaction between the SPIR treatment indicator and the endline round dummy. This allows the average treatment effect to differ between midline and endline, absorbing any time-varying treatment effect that could confound the dryness x treatment interaction when dryness variation is partly temporal. Standard errors (in parentheses) based on Conley (1999) with a 600 km distance cut-off. Statistical significance denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table G6. Replicating Table 1, but controlling for locust and conflict shocks

	(1)	(2)	(3)	(4)	(5)
Outcome:	Primary woman's diet diversity	BMI of primary woman	Primary woman experienced IPV	Household food gap	Tropical livestock units owned by household
Dryness	-0.590	0.137**	0.126***	3.045***	-1.397***
	(0.483)	(0.056)	(0.019)	(0.618)	(0.115)
Dryness X SPIR	1.266***	-0.165***	-0.100***	-1.737***	1.022***
	(0.168)	(0.050)	(0.028)	(0.453)	(0.106)
SPIR	0.035	0.007	0.003	0.067	-0.046***
	(0.056)	(0.006)	(0.002)	(0.060)	(0.017)
Controls for locust and conflict shocks?	Yes	Yes	Yes	Yes	Yes
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome in control group	2.02	20.12	0.17	2.06	0.97
'Dryness' + 'Dryness X SPIR' = 0	p = 0.149	p = 0.370	p = 0.481	p = 0.050	p < 0.000
Number of observations	5,877	4,888	3,808	6,212	6,274

Note: 'Dryness' is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The controls for locust and conflict are binary variables obtaining value 1 if there were conflict events or locust swarms within 20 km distance from the community, and zero otherwise. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * p < 0.10, ** p < 0.05, *** p < 0.01.

Table G7. Impact of graduation programming and dryness conditions on household composition

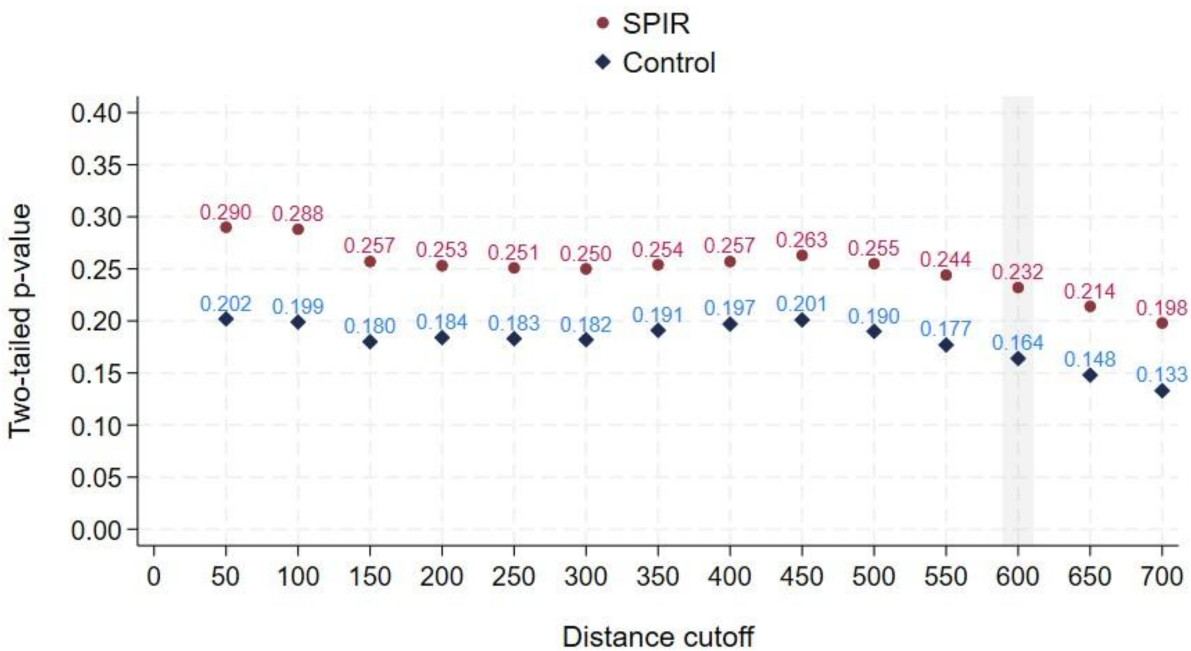
	(1)	(2)	(3)	(4)
Outcome:	Household size	Household size	Household size in adult equivalent units	Household size in adult equivalent units
Sample:	*Control households*	*All households*	*Control households*	*All households*
Dryness	-0.039	-0.121	0.075	-0.132
	(0.545)	(0.357)	(0.404)	(0.255)
Dryness X SPIR		-0.080		0.107
		(0.205)		(0.126)
SPIR		0.076		0.057*
		(0.053)		(0.032)
Woreda fixed effects?	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes
Baseline mean of the outcome variable in the control group	5.75	5.75	4.42	4.42
Number of observations	1,517	6,314	1,517	6,312

Note: 'Dryness' is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

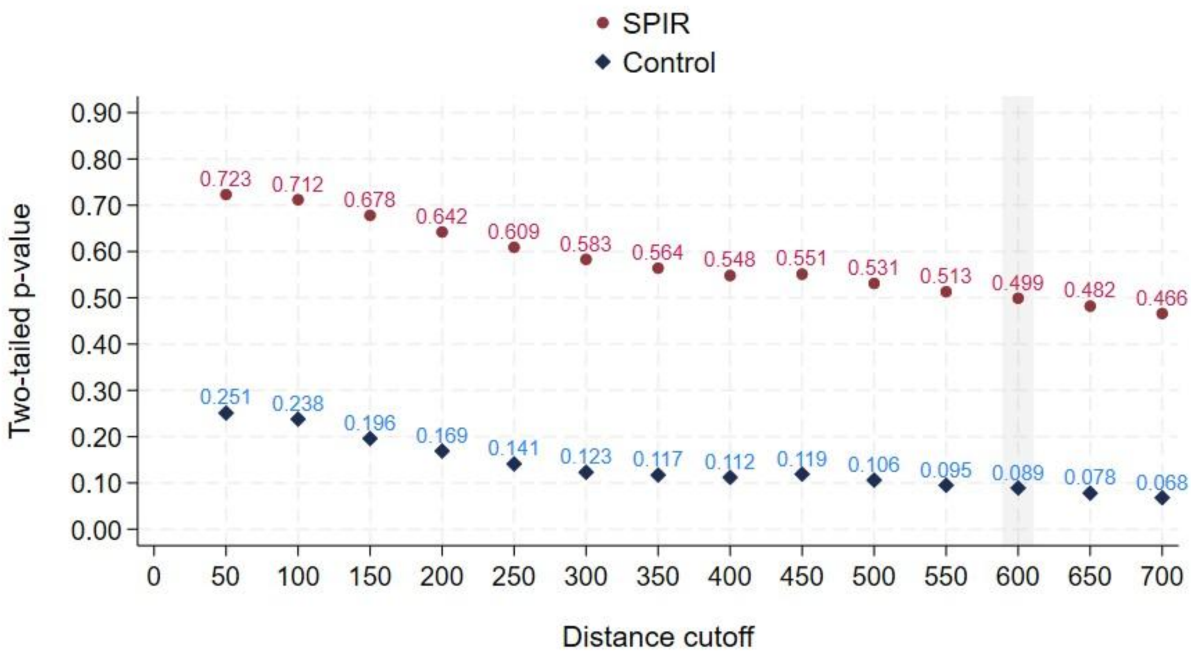
Figure G1. Impact of increase in relative dryness on household and individual level outcomes, by treatment status: p-values based on Conley (1999) standard errors with different distance cut-offs

Note: 'SPIR' is the p-value for the test of $\beta + \delta = 0$ in Eq. (1), capturing the effect of dryness in SPIR households based on a joint hypothesis test. 'Control' is the p-value of $\beta = 0$ in Eq. (1), capturing the effect of dryness in control households based on a single coefficient test. The shaded distance cut-off, 600 km, is used throughout the paper.

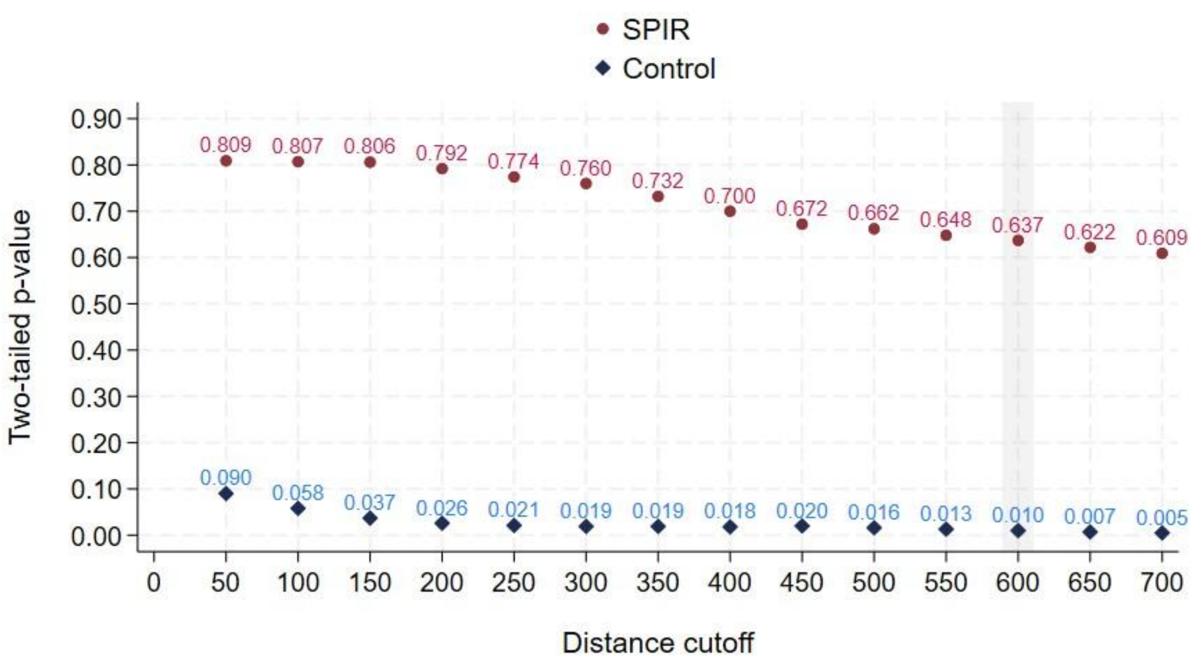
A) Primary woman's diet diversity



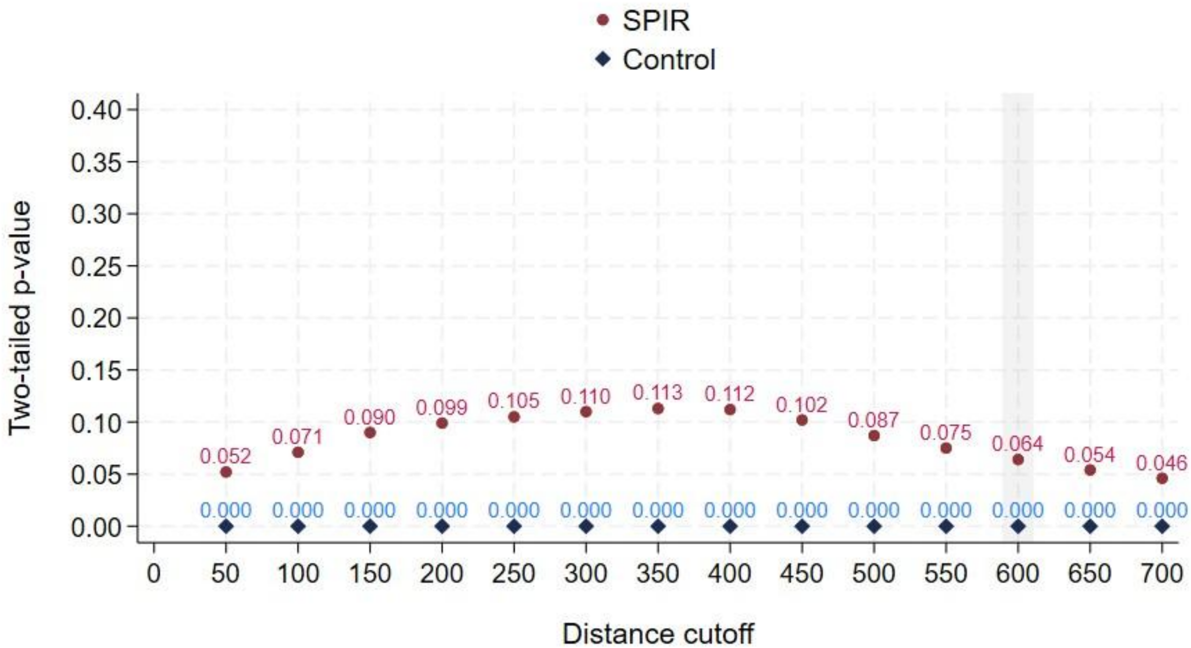
B) BMI of primary woman



C) Primary woman experienced IPV



D) Household food gap



E) Tropical livestock units owned by household

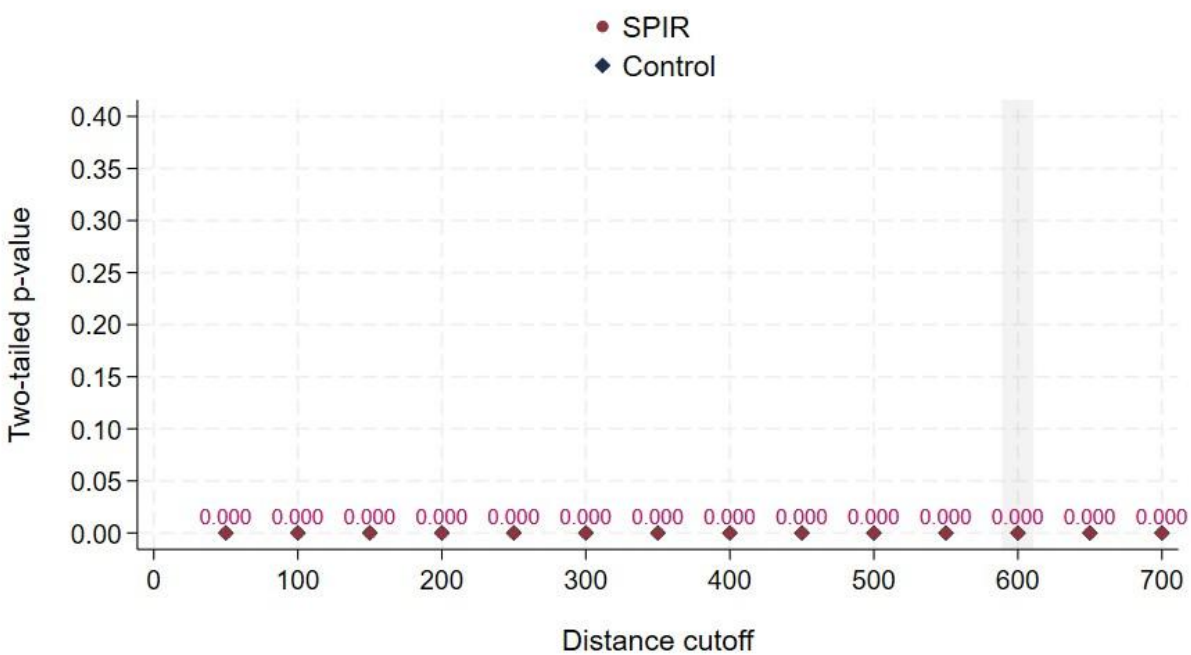


Figure G2. The relationship between SPEI in 2018 meher and SPEI in 2016 meher

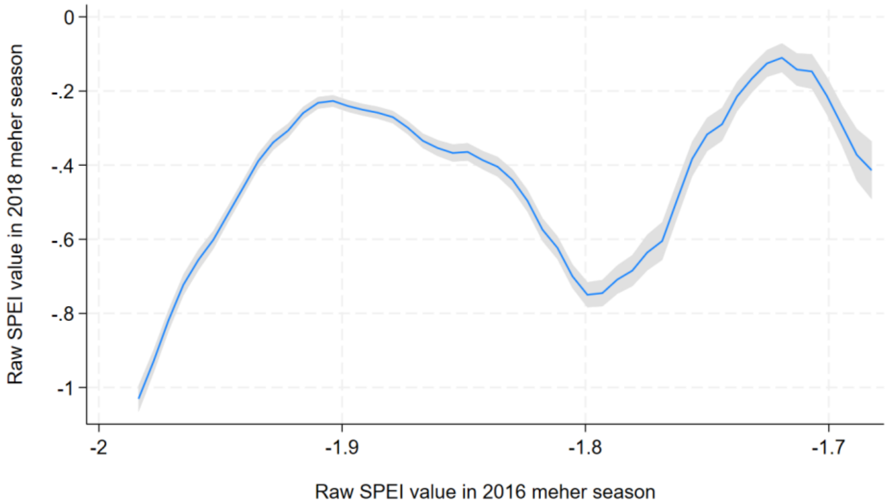


Figure G3. The relationship between SPEI in 2019 meher and SPEI in 2016 meher

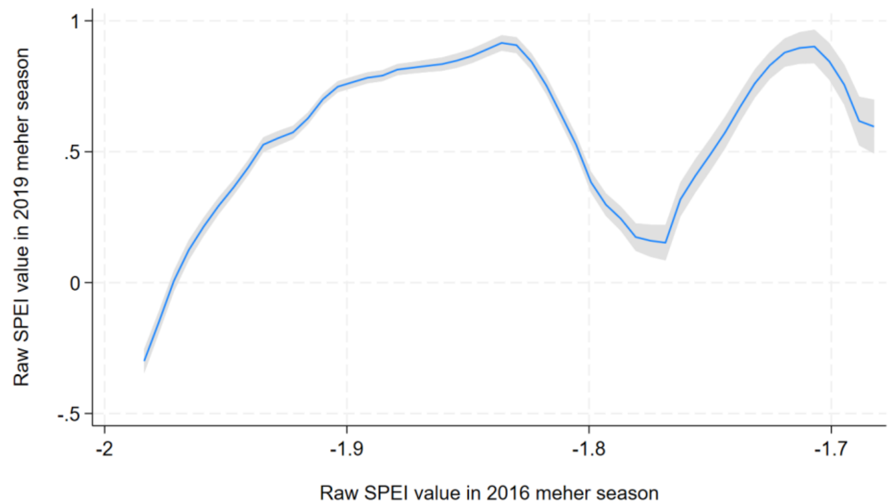


Figure G4. The relationship between SPEI in 2020 meher and SPEI in 2016 meher

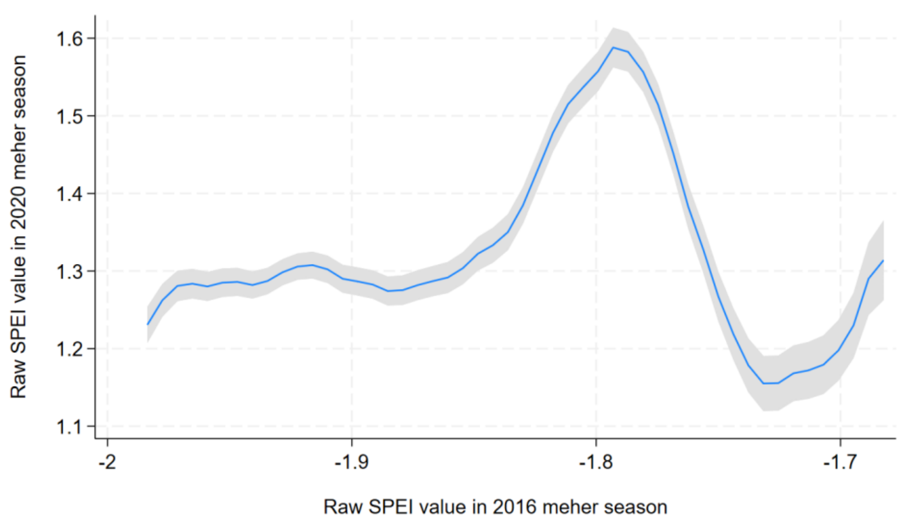


Table G1. Replicating Table 1 but controlling for 2016 dryness conditions

	(1)	(2)	(3)	(4)	(5)
Outcome:	Primary woman's diet diversity	BMI of primary woman	Primary woman experienced IPV	Household food gap	Tropical livestock units owned by household
Dryness	-1.025**	-0.144*	0.144**	3.443***	-1.294***
	(0.411)	(0.086)	(0.059)	(0.582)	(0.074)
Dryness X SPIR	1.333***	0.207***	-0.161***	-1.786***	1.003***
	(0.162)	(0.039)	(0.049)	(0.418)	(0.098)
SPIR	0.034	-0.031	0.007	0.068	-0.045**
	(0.055)	(0.041)	(0.006)	(0.066)	(0.018)
Drought in 2016?	Yes	Yes	Yes	Yes	Yes
Woreda fixed effects?	Yes	Yes	Yes	Yes	Yes
Survey round fixed effects?	Yes	Yes	Yes	Yes	Yes
Outcome variable at the baseline?	Yes	Yes	Yes	Yes	Yes
Baseline mean of the outcome in control group	2.02	20.12	0.17	2.06	0.97
'Dryness' + 'Dryness X SPIR' = 0	p = 0.450	p = 0.389	p = 0.649	p = 0.005	p = 0.001
Number of observations	5,877	4,888	3,808	6,212	6,274

Note: IPV = Intimate Partner Violence. 'Dryness' is positive rectified SPEI multiplied by -1 so that larger (positive) values reflect worse dryness conditions. In column 5, the estimated equation includes a binary variable capturing households for which the baseline value was missing (these missing baseline values were set to zero). The standard errors are reported in parentheses and based on Conley (1999) with a 600 km distance cut-off. Statistical significance is denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix references

Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific data* 2018;5; 1-12.

Alderman H, et al. Impact evaluation of the strengthen PSNP4 institutions and resilience (SPIR) development food security activity (DFSA): Endline report. International Food Policy Research Institute (IFPRI): Washington D.C.; 2021.

FAO. FAO Locust Hub FAO's Desert Locust Information Service. Food and Agriculture Organization of the United Nations (FAO): Rome; 2023.

Fisher AC, Hanemann WM, Roberts MJ, Schlenker W. The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: comment. *The American Economic Review* 2012;102; 3749-3760.

Gray CL, Mueller V. Drought and Population Mobility in Rural Ethiopia. *World Development* 2012;40; 134-145.

Hirvonen K, Sohnesen TP, Bundervoet T. Impact of Ethiopia's 2015 drought on child undernutrition. *World Development* 2020;131.

Raleigh C, Linke A, Hegre H, Karlsen J. Introducing ACLED: an armed conflict location and event dataset: special data feature. *Journal of Peace Research* 2010;47; 651-660.

Shively GE. Infrastructure mitigates the sensitivity of child growth to local agriculture and rainfall in Nepal and Uganda. *Proceedings of the National Academy of Sciences* 2017;114; 903-908.

Thiede BC. Rainfall shocks and within-community wealth inequality: evidence from rural Ethiopia. *World Development* 2014;64; 181-193.