IFPRI Discussion Paper 02212

December 2023

Men Can Cook: Effectiveness of a Light-Touch Men's Engagement Intervention to Change Attitudes and Behaviors in Rural Ethiopia

Harold Alderman

Daniel O. Gilligan

Melissa Hidrobo

Jessica Leight

Michael Mulford

Heleene Tambet

Poverty, Gender, and Inclusion Unit

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Harold Alderman (h.alderman@cgiar.org) is a senior research fellow in the Poverty, Gender, and Inclusion (PGI) unit at the International Food Policy Research Institute (IFPRI).

Daniel O. Gilligan (d.gilligan@cgiar.org) is the director of the PGI unit at IFPRI.

Melissa Hidrobo (m.hidrobo@cgiar.org) is a senior research fellow in the PGI unit at IFPRI.

Jessica Leight (j.leight@cgiar.org) is a research fellow in the PGI unit at IFPRI.

Michael Mulford (Michael_mulford@wvi.org) is the Chief of Party of SPIR at World Vision. Addis Ababa, Ethiopia

Heleene Tambet (h.tambet@cgiar.org) is a senior research analyst in the PGI unit at IFPRI.

Notices

[1] IFPRI Discussion Papers contain preliminary material and research results and are circulated in order to stimulate discussion and critical comment. They have not been subject to a formal external review via IFPRI's Publications Review Committee. Any opinions stated herein are those of the author(s) and are not necessarily representative of or endorsed by IFPRI.

[2] The boundaries and names shown and the designations used on the map(s) herein do not imply official endorsement or acceptance by the International Food Policy Research Institute (IFPRI) or its partners and contributors.

ABSTRACT

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

Key words: Men's engagement, gender norms, poverty

ACKNOWLEDGMENTS

We gratefully acknowledge funding for the impact evaluation from USAID under Cooperative Agreement No AID-FFP-A-16-00008. We also thank our partners at World Vision, CARE, and ORDA for their contributions to the study and overall support, and Laterite for careful data collection. We also gratefully acknowledge project management support by Lucy Billings. This work is part of the CGIAR Research Initiative on Gender Equality. We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund.

I. Introduction

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

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

provide an opportunity for men to critically reflect on cultural gender norms and explore the positive and perceived negative effects of male involvement in tasks traditionally assigned to women.

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

Our findings contribute to the recent literature on the importance of men's engagement strategies in achieving improved gender equity and related outcomes (Glinski et al., 2018) including RCTs evaluating the impact of these interventions on gender attitudes (Dhar et al., 2022; Nguyen & Tarp, 2022), intimate partner violence (Abramsky et al., 2014; Christofides et al., 2020; Doyle et al., 2018; Hossain et al., 2014; Sharma et al., 2020; Vaillant et al., 2020), HIV-prevention (Jewkes et al., 2008; Sharma et al., 2020) and family planning (Raj et al., 2016; Shattuck et al., 2011). We contribute to this literature in three notable ways. First, in contrast to the studies mentioned above, we study a men's engagement intervention in the context of a multicomponent anti-poverty program. Second, while gender equitable approaches were mainstreamed into different program components, the men's

engagement activity is lighter touch than most of those mentioned above and implemented at a larger scale.[1] Lastly, there are only a few experimental studies that include outcomes related to men's involvement in household tasks and even fewer that include both women's and men's reports.[2]

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

II. Experimental design

a) Context

The study takes place in rural Ethiopia where gender norms are patriarchal and traditional, with men considered heads of family and the main decisionmakers, while women are expected to be submissive

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

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

to men (Alemu 2007; Mabsout et al 2010). Qualitative work on the study sample reveals that men and women find physical violence acceptable as a form of disciplinary action when women do not complete their household responsibilities or obey their husband (Ranganathan et al 2022). Moreover a third of men and more than one half of women in the study sample believe a husband is justified beating his wife under some circumstances, including perceived negligence of domestic duties such as burning food or neglecting children (Alderman et al 2021).

In terms of gender roles, women are responsible for domestic chores and men responsible for farm activities, to which women and children also contribute (Tefera, 2020). According to the Ethiopian Central Statistical Agency (2014), 94 percent of rural women were involved in domestic activities, spending an average of about 5 to 6 hours per day. While most men are accepting of men contributing to domestic chores, men significantly underestimate the extent to which their peers hold this view (Mccullough et al., 2022). Still, only one third of men in rural Ethiopia are reported by their spouses to help with household chores, and when they do, this is a rare occurrence (Central Statistical Agency and ICF 2016). In our sample, only one fifth of women report that their spouses help with household chores (Alderman et al 2021).

b) Interventions

i. PSNP and SPIR

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

livelihood, nutrition, and gender equity activities to strengthen the program and expand its impacts. SPIR provided support to over 150,000 PNSP households in 15 food insecure districts in Amhara and Oromia regions.

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

ii. VESA Gender dialogue

The gender dialogue sessions that were part of VESAs included 6 discussion sessions on 1) workloads of men and women, 2) cooperation and sharing household work; 3) household decision making; 4)

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

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

iii. Men's engagement groups

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

Men's engagement groups were facilitated by men's group facilitators (MGF) and male advocators. The MGFs began by identifying 'male advocators,' men who had the potential to become role models of attitude and behavior change in their communities. If they were willing, these men were

selected into this role through consultation with VESA members and kebele administration leaders. One of the key responsibilities of the male advocators was to support the MGF with the co-facilitation of the men's engagement group, with two male advocators per group. With the goal of establishing 7-8 men's engagement groups in a kebele, between 14-16 male advocators were selected per kebele, and provided with a 3-day orientation and training on the men's engagement methodology and lessons. Afterwards, men's engagement groups were formed, with 20 men in each group, including the male advocators. As PSNP client households only represent the poorest 10-15 percent of households in a community, SPIR intentionally broadened the scope of the men's engagement intervention to include both PSNP and non-PSNP men in an effort to engage the broader community, and included local elders, religious leaders or kebele officials. These groups typically met twice per month. According to the endline survey, about 40 percent of men in N* kebeles reported attending at least one men's engagement session in the last 12 months.[4]

c) Experimental design

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

Treatment 1 (T1): L* + N*

Treatment 2 (T2): L* + N

Treatment 3 (T3): L + N*

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

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

Control (T4): PSNP only

Table 1 shows how the livelihood and nutrition activities map to the intervention arms. For this paper, we are interested in the impact of the three treatment arms (T1, T2, T3) compared to the control group on men's gender equitable attitudes and behavior. All three treatment arms received PSNP and the core nutrition and livelihood interventions which included the VESA gender dialogue. In addition, we are interested in the added impact of N* activities in T1 and T3, which include the men's engagement groups. While most livelihood and nutrition activities started before the midline survey, a few enhanced nutrition (N*) activities (mainly the men's engagement groups and IPT-G) started after the midline survey.[6]

III. Data and empirical strategy

a) Data collection

The SPIR impact evaluation included three rounds of data collection; the baseline survey was conducted from February to April 2018, the midline survey from July to October 2019, and the endline survey from February to April 2021. The sampling frame included 192 study kebeles, from which households were randomly selected based on the following inclusion criteria (1) be a PSNP client household, (2) have at least one child aged 0-35 months (index child), and (3) have the mother or primary female caregiver of the 0-35-month-old child be a member of the household. The last two criteria enabled measurement of maternal and child nutrition, a primary outcome in the overall evaluation. In total 3,314 households were interviewed at baseline, of which 3,220 households were re-interviewed at midline. In addition, 748 supplemental households were added to the midline sample, for a total of 3,968 households at midline. These supplemental households were sampled based on the same criterion as above, with the

[6] In addition, due to budget considerations, men's engagement groups within N* were randomly rolled out, with half of the N* kebeles starting men's engagement groups in December 2019 and the second half of N* kebeles started between June-December 2020. We find no differential effects with respect to the timing of the rollout.

exception that the index child must be under the age of 2 years. The endline survey sought to re-interview all households in the midline sample as well as any baseline household that was not available at midline but deemed likely to be available at endline.[7] Of the 3,996 households in the endline target sample, 3,812 were located and interviewed.

In all rounds, the questionnaire was structured in three parts: a brief household-level questionnaire for identification and household demographics, a male respondent questionnaire and a female respondent questionnaire. The female questionnaire was administered to the mother or primary female caregiver of the index child ("primary female") and the male questionnaire was administered to her husband or partner ("primary male"). In households with a single adult female and no adult male, some of the modules for the primary male respondent were skipped. Each survey respondent was surveyed separately and provided informed oral consent.

b) Outcomes of interest

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

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

We also construct three different indicators of equitable gender attitudes. The first is from a list of four questions asked to the primary male on whether a husband is justified in beating his wife under different circumstances: going out without telling him, burning the food, neglecting the children, and arguing with him. We create a binary indicator that equals one if the respondent says a husband is not justified in beating his wife under any of the four circumstances. The second is from a list of three questions on whether it is acceptable for a woman to travel alone to different places: market, friends/family, health center. We create a binary indicator that equals one if the primary male respondent says it is acceptable for a woman to travel alone to the three specified places. Lastly, we ask the primary male whether they agree with five gender inequitable statements (such as whether "Changing diapers, giving a bath, and feeding kids is only the mother's responsibility" or "A woman should obey her husband in all things") where the responses range from strongly disagree (0) to strongly agree (4).[8] Scores are reversely coded and summed up so that a higher score implies more equitable gender attitudes, ranging from 0-20.

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

c) Sample, attrition and balance

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

The sample we use in the analysis are households with a primary male as the main outcomes are specific to men. For the main analysis we use the unbalanced panel sample that includes all primary males available at each time period, as this gives us the most power to detect impacts. However, we also conduct the analysis on the balanced sample of primary males that were available at midline and endline as a robustness check. Figure 1 illustrates the sample of primary males surveyed across all three rounds of data collection. At baseline there are 2,813 primary males, of which 2,398 are re-interviewed at midline in addition to 684 primary males that were added to the sample (600 from the supplemental sample and 84 from households without a primary male at baseline) for a midline sample of 3,082 primary males. Of the 3,082 primary males at midline, 2,235 were resurveyed at endline, plus 127 primary males from baseline that were not available at midline, for a total sample of 2,362.

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

d) Estimation

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

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

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

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

value (as stars in the tables) and the sharpened q-value (in brackets), while clustered standard errors are reported in parentheses.

IV. Results

a) Main impacts

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

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

across all three indicators, consistent with the pattern on the summary index. Increased point estimates for the T1 and T3 arms at endline range from 12.7 to 20.8 percentage point increase over control means of 22.7 to 55.1 percent.

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

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

b) Robustness

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

One potential alternative reason for the difference in impact estimates between midline and endline is that the sample of men changes across rounds. To ensure our results are not driven by the different sample of men, we run the same analysis for the sample of men who were included in both the midline and endline rounds. Appendix Table 4 and 5, reveal that our results are robust to this panel sample.

Another potential explanation is that men's behavior did not actually change, and instead their responses were subject to social desirability bias. If this was more likely to occur in the treatment groups with N*, then this could bias our estimates. To investigate whether social desirability bias is likely to be an issue, we estimate impacts on women's reports of their partners' behavior. Appendix Table 6 reveals that at endline, impacts from women's reports of their husbands' behavior are similar to impacts from men's self-reports, and in particular the pattern of large impacts across T1 and T3, but not T2, holds.

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

into IPT-G. Appendix tables 7 and 8 reveal that results are robust to excluding households with women who were eligible to participate in IPT-G.

c) Other outcomes

While involving men in childcare activities was an integral part of the curriculum in both the VESA gender dialogues and the men's engagement groups, we do not include men's childcare activities in the main analysis as questions regarding childcare were only asked on the subsample of men in each time period who had a child less than 24 months, which is only about a quarter of households by endline. However, appendix table 9 reveals that the magnitude of impact on men's childcare activities is generally small and close to zero suggesting that regardless of the smaller sample, there are no impacts on these indicators. Thus, while SPIR was able to change men's involvement in chores and their attitudes about gender equitable norms, it does not appear to have led to greater participation in childcare activities. One potential reason for why SPIR was not able to change men's behavior with respect to childcare, is that most men were already engaging in childcare activities (80 percent of men in the control group at midline and 98 percent at endline took part in at least one childcare activity in the last 3 days), which suggests there was just less room for impact.

V. Conclusion

In this paper we investigate the impact on men's gender equitable attitudes and behaviors of a graduation model program that seeks to simultaneously "push" households out of poverty and improve gender equitable norms. We find that at midline all treatment arms lead to improvements in men's gender equitable attitudes and their involvement in domestic tasks. Unfortunately, we cannot disentangle what component or activities lead to the improvements but given that impacts at midline do not vary by arm, we conclude that it is likely not the L* and N* components but instead, the core N and L components that were similar across all arms. By endline, only the treatment arms with the men's

engagement groups continue to have significant impacts which are large and range from 0.26-0.29 SD for men's reported involvement in household tasks. Impacts at endline on gender equitable attitudes are smaller in magnitude and range from 0.07-0.14 SD. Across both sets of outcomes, the magnitudes of impact in the treatment arms receiving the men's engagement intervention (T1 and T3) are significantly different than that of the treatment arm that did not receive this intervention (T2). Thus, we attribute the sustained or larger impacts at endline in T1 and T3 to the men's engagement groups which were rolled out after the midline survey. Results are robust to using women's report of their husband's behaviors, a balanced sample, and a smaller sample that excludes households where women were also treated for depression.

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

References

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

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

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

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

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

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

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." Journal of the American Statistical Association 103: 1481–95. https://doi.org/10.1198/016214508000000841.

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

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

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

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

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

Christofides, N. J., Hatcher, A. M., Rebombo, D., McBride, R. S., Munshi, S., Pino, A., Abdelatif, N., Peacock, D., Levin, J., & Jewkes, R. K. (2020). Effectiveness of a multi-level intervention to reduce men's perpetration of intimate partner violence: A cluster randomised controlled trial. *Trials*, *21*(1). https://doi.org/10.1186/s13063-020-4185-7.

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

Dhar, D., Jain, T., & Jayachandran, S. (2022). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review*, *112*(3), 899–927. https://doi.org/10.1257/AER.20201112.

Doyle, K., Levtov, R. G., Barker, G., Bastian, G. G., Bingenheimer, J. B., Kazimbaya, S., Nzabonimpa, A., Pulerwitz, J., Sayinzoga, F., Sharma, V., & Shattuck, D. (2018). Gender-transformative bandebereho couples' intervention to promote male engagement in reproductive and maternal health and violence prevention in Rwanda: Findings from a randomized controlled trial. *PLoS ONE*, *13*(4). https://doi.org/10.1371/journal.pone.0192756.

Duflo, E., & Udry, C. (2004). *Intrahousehold Resource Allocation in Côte d'Ivoire: Social Norms, Separate Accounts and Consumption Choices*. http://www.nber.org/papers/w10498.

Glinski, A., Schwenke, C., O'Brien-Milne, L., & Farley, K. (2018). *Gender Equity and Male Engagement: It only works when everyone plays.*

Hossain, M., Zimmerman, C., Kiss, L., Abramsky, T., Kone, D., Bakayoko-Topolska, M., Annan, J., Lehmann, H., & Watts, C. (2014). Working with men to prevent intimate partner violence in a conflict-affected setting: a pilot cluster randomized controlled trial in rural Côte d'Ivoire. *BMC Public Health*, *14*(1), 339. https://doi.org/10.1186/1471-2458-14-339.

Jayachandran, S. (2021). Social Norms as a Barrier to Women' s Employment in Developing Countries. *IMF Economic Review*, *69*, 576–595.

Jewkes, R., Nduna, M., Levin, J., Jama, N., Dunkle, K., Puren, A., & Duvvury, N. (2008). Impact of Stepping Stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: cluster randomised controlled trial. *BMJ*, *337*(aug07 1), a506–a506. https://doi.org/10.1136/bmj.a506.

Kling, J., Liebman, J., & Katz, L. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119.

Mabsout R, and van Staveren I. 2010. "Disentangling Bargaining Power from Individual and Household Level to Institutions: Evidence on Women's Position in Ethiopia." World Development 38 (5): 783–96.

Mccullough, E., Mcgavock, T., Magnan, N., Assefa, T., Mcgavock, T. J., & Assefa, T. W. (2022). *Experimental Effects of Updating Husbands' Beliefs on Participation in Women's Household Work.*

Nguyen, C. V., & Tarp, F. (2022). Changing male perceptions of gender equality: Evidence from a randomised controlled trial study. *World Development*, *158*. https://doi.org/10.1016/j.worlddev.2022.106019.

Raj, A., Ghule, M., Ritter, J., Battala, M., Gajanan, V., Nair, S., Dasgupta, A., Silverman, J. G., Balaiah, D., & Saggurti, N. (2016). Cluster Randomized Controlled Trial Evaluation of a Gender Equity and Family Planning Intervention for Married Men and Couples in Rural India. *PLOS ONE*, *11*(5), e0153190. https://doi.org/10.1371/journal.pone.0153190.

Ranganathan, M., Pichon, M., Hidrobo, M., Tambet, H., Sintayehu, W., Tadesse, S., & Buller, A. M. (2022). Government of Ethiopia's public works and complementary programmes: a mixed-methods study on pathways to reduce intimate partner violence. Social Science & Medicine (1982), 294, 114708–114708.

Sharma, V., Leight, J., Verani, F., Tewolde, S., & Deyessa, N. (2020). Effectiveness of a culturally appropriate intervention to prevent intimate partner violence and HIV transmission among men, women, and couples in rural Ethiopia: Findings from a cluster-randomized controlled trial. *PLoS Medicine*, *17*(8). https://doi.org/10.1371/JOURNAL.PMED.1003274.

Shattuck, D., Kerner, B., Gilles, K., Hartmann, M., Ng'ombe, T., & Guest, G. (2011). Encouraging contraceptive uptake by motivating men to communicate about family planning: The Malawi Male Motivator project. *American Journal of Public Health*, *101*(6), 1089–1095. https://doi.org/10.2105/AJPH.2010.300091.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, *73*(3), 751–754. https://doi.org/10.1093/biomet/73.3.751.

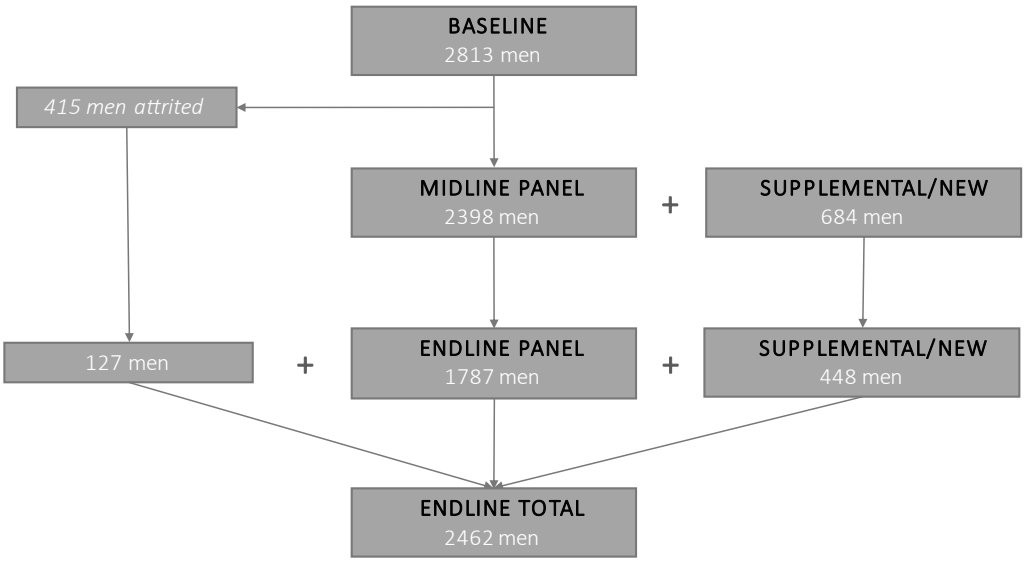
Tefera, N. (2020). *Gender-Differentiated Time Allocations on Daily Activities in Rural Ethiopia*.

Udry, C. (1996). Gender, Agricultural Production, and the Theory of the Household. In *Source: Journal of Political Economy* (Vol. 104, Issue 5). https://about.jstor.org/terms.

Vaillant, J., Koussoubé, E., Roth, D., Pierotti, R., Hossain, M., & Falb, K. L. (2020). Engaging men to transform inequitable gender attitudes and prevent intimate partner violence: a cluster randomised controlled trial in North and South Kivu, Democratic Republic of Congo. *BMJ Global Health*, *5*(5). https://doi.org/10.1136/bmjgh-2019-002223.

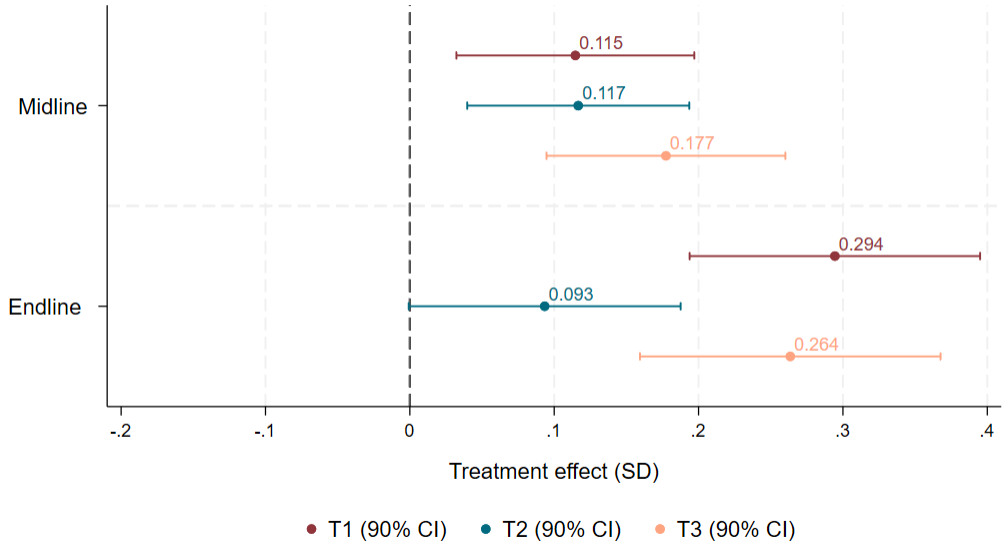
Figures

Figure 1: Flow chart of primary males in study



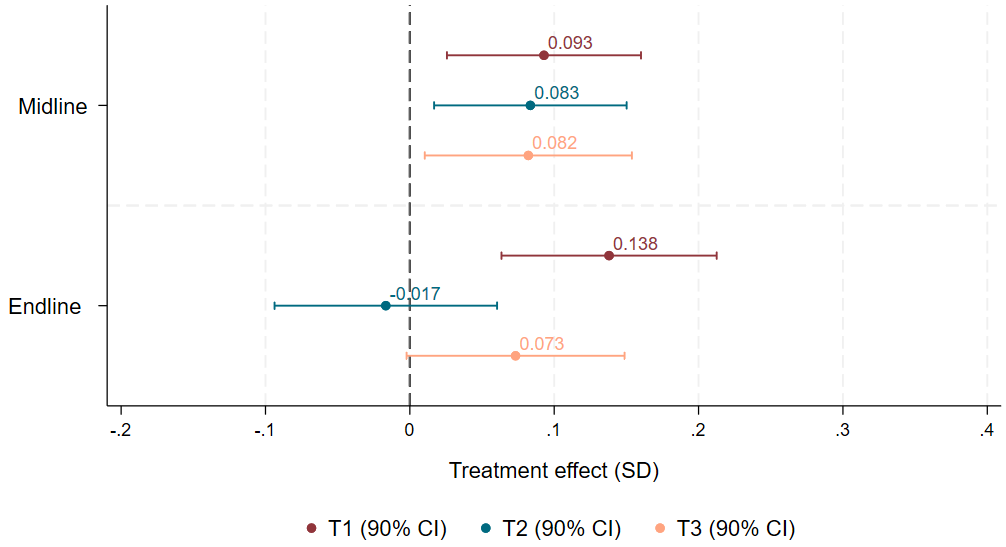
*Note: Supplemental and new men refers to primary males from the supplemental sample and primary males who were not surveyed at baseline but whose household was surveyed at baseline.

Figure 2: Impacts on summary index for men's involvement in domestic tasks



Notes: The model specifications control for the woreda-level fixed effects and the binary indicators of being eligible for the livelihood transfers and being in the supplemental sample. Endline specification also controls for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. Standard errors are clustered at the kebele level.

Figure 3: Impacts on summary index for men's gender equitable attitudes



Notes: The model specifications control for the woreda-level fixed effects and the binary indicators of being eligible for the livelihood transfers and being in the supplemental sample. Endline specification also controls for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. Standard errors are clustered at the kebele level.

Tables

Table 1: Study intervention arms and livelihood and nutrition components

Intervention description	T1: L*+N*	T2: L*+N	T3: L+N*	T4: PSNP only
PSNP consumption-support transfers	✓	✓	✓	✓
L: VESA+ gender dialogue +trainings	✓	✓	✓	
N: Nutrition BCC + WASH	✓	✓	✓	
L*: + livelihood transfers	✓	✓		
N*: + TTC + CPNP + men's engagement + IPT-G	✓		✓	

*VESA is Village Economic and Social Associations. BCC is behavior change and communication. WASH is water, sanitation, and hygiene. TTC is time and targeted nutrition counseling. CPNP is community-based participatory nutrition promotion, IPT-G is interpersonal psychotherapy in groups.

Table 2: Primary male's involvement in domestic tasks in the past 3 days

	Midline			Endline			Midline vs endline		
	(1) Helped with household chores	(2) Helped with cooking or meal preparation	(3) Helped with collecting firewood and water	(4) Helped with household chores	(5) Helped with cooking or meal preparation	(6) Helped with collecting firewood and water	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.046	0.049*	0.066**	0.151***	0.127***	0.208***	0.019	0.020	0.001
	(0.028)	(0.026)	(0.033)	(0.046)	(0.035)	(0.035)			
	[0.103]	[0.079]	[0.065]	[0.003]	[0.001]	[0.000]			
T2	0.046	0.054**	0.062**	0.030	0.003	0.104***	0.643	0.113	0.306
	(0.028)	(0.024)	(0.031)	(0.034)	(0.027)	(0.033)			
	[0.163]	[0.076]	[0.084]	[0.464]	[0.923]	[0.011]			
T3	0.069**	0.097***	0.079**	0.146***	0.138***	0.144***	0.048	0.236	0.108
	(0.028)	(0.024)	(0.036)	(0.038)	(0.029)	(0.035)			
	[0.018]	[0.000]	[0.027]	[0.000]	[0.000]	[0.000]			
Test: T1 = T2	(0.992)	(0.845)	(0.911)	(0.007)	(0.000)	(0.001)			
	[0.992]	[0.992]	[0.992]	[0.014]	[0.002]	[0.002]			
Test: T2 = T3	(0.439)	(0.091)	(0.637)	(0.001)	(0.000)	(0.188)			
	[0.527]	[0.182]	[0.637]	[0.004]	[0.000]	[0.282]			
Test: T1 = T3	(0.414)	(0.078)	(0.723)	(0.902)	(0.706)	(0.039)			
	[0.828]	[0.234]	[0.867]	[0.902]	[0.867]	[0.234]			
Mean of control	0.342	0.233	0.506	0.294	0.227	0.551			
N	3,067	3,067	3,068	2,452	2,450	2,451			

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Table 3: Primary Male's gender equitable attitudes

	Midline			Endline			Midline vs endline		
	(1) A husband is not justified in beating his wife in any of these situations [1]	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(3) Index of support for equitable gender norms	(4) A husband is not justified in beating his wife in any of these situations [1]	(5) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(6) Index of support for equitable gender norms	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.034	0.034	0.914***	0.057*	0.051	0.800**	0.565	0.689	0.794
	(0.033)	(0.028)	(0.319)	(0.033)	(0.043)	(0.380)			
	[0.306]	[0.284]	[0.028]	[0.181]	[0.284]	[0.109]			
T2	0.043	0.043	0.557*	-0.005	0.008	-0.252	0.156	0.343	0.042
	(0.030)	(0.027)	(0.336)	(0.032)	(0.035)	(0.347)			
	[0.312]	[0.312]	[0.312]	[0.876]	[0.876]	[0.703]			
T3	0.044	0.028	0.520	0.051	0.017	0.380	0.847	0.782	0.731
	(0.031)	(0.029)	(0.336)	(0.033)	(0.041)	(0.339)			
	[0.309]	[0.390]	[0.309]	[0.309]	[0.680]	[0.390]			
Test: T1 = T2	(0.778)	(0.730)	(0.245)	(0.047)	(0.258)	(0.002)			
	[0.778]	[0.778]	[0.386]	[0.142]	[0.386]	[0.009]			
Test: T2 = T3	(0.960)	(0.579)	(0.909)	(0.057)	(0.802)	(0.026)			
	[0.960]	[0.960]	[0.960]	[0.171]	[0.960]	[0.157]			
Test: T1 = T3	(0.749)	(0.845)	(0.201)	(0.852)	(0.327)	(0.098)			
	[0.852]	[0.852]	[0.604]	[0.852]	[0.654]	[0.590]			
Mean of control	0.602	0.356	8.696	0.666	0.426	10.333			
N	3,070	3,076	3,079	2,450	2,458	2,460			

[1] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband.

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Tables

Appendix Table 1: Attrition rates by treatment arm of primary male respondents

	(1) Primary male respondent attrited baseline to midline	(2) Primary male respondent attrited baseline to endline	(3) Primary male respondent attrited midline to endline
T1	-0.002	-0.060**	-0.048*
	(0.021)	(0.030)	(0.027)
T2	-0.021	-0.068**	-0.058**
	(0.021)	(0.032)	(0.028)
T3	-0.003	-0.056*	-0.034
	(0.020)	(0.032)	(0.029)
Test: T1 = T2	0.358	0.777	0.692
Test: T2 = T3	0.336	0.685	0.379
Test: T1 = T3	0.978	0.885	0.609
Mean of control	0.132	0.313	0.264
N	2,813	2,813	3,082

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

Appendix Table 2: Selective attrition by baseline characteristics
Attrition from baseline to midline

	(1) N	(2) T1	(3) T2	(4) T3	(5) T1 x Attrition	(6) T2 x Attrition	(7) T3 x Attrition
Household size	2,916	0.211*	0.064	0.023	0.754	0.050	0.240
		(0.127)	(0.116)	(0.126)	(0.486)	(0.587)	(0.530)
Number of children under the age of 5	2,916	0.010	-0.010	0.004	0.074	0.240	0.127
		(0.032)	(0.032)	(0.034)	(0.139)	(0.210)	(0.151)
Log of total monthly consumption expenditure per adult equivalent	2,888	-0.003	-0.006	0.031	-0.242	-0.028	-0.127
		(0.057)	(0.060)	(0.062)	(0.198)	(0.331)	(0.204)
Age of primary male	2,813	0.582	0.229	0.198	7.144**	1.252	5.020
		(0.571)	(0.533)	(0.491)	(3.085)	(3.309)	(3.978)
Primary male is the household head	2,813	-0.021	0.006	0.010	0.225**	0.098	0.138
		(0.025)	(0.020)	(0.018)	(0.112)	(0.130)	(0.116)
Primary male has some education	2,812	0.006	-0.001	0.010	-0.152	0.248	-0.027
		(0.030)	(0.034)	(0.031)	(0.154)	(0.188)	(0.156)
Primary male's main activity is crop production	2,811	-0.000	0.012	0.030	0.019	0.050	-0.115
		(0.037)	(0.034)	(0.036)	(0.141)	(0.134)	(0.143)
Primary male is married and lives with their spouse	2,800	0.002	-0.012	0.014	0.050	-0.072	0.049
		(0.020)	(0.021)	(0.018)	(0.104)	(0.133)	(0.101)
Primary male finds it acceptable for a woman to travel alone to market, health center, or visit a friend	2,646	0.031	0.030	0.036	0.006	0.180	0.205
		(0.040)	(0.041)	(0.042)	(0.160)	(0.174)	(0.150)
Primary male doesn't find violence justified in any situation asked about	2,646	-0.040	-0.034	-0.064**	0.037	0.028	0.101
		(0.030)	(0.029)	(0.028)	(0.131)	(0.128)	(0.131)
Total number of childcare activities primary male participated in (0-10)	2,800	-0.060	0.025	0.060	1.273	1.041	0.278
		(0.157)	(0.173)	(0.161)	(1.070)	(0.961)	(0.740)

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

Appendix Table 3: Selective attrition by baseline characteristics
Attrition from baseline to endline

	(1) N	(2) T1	(3) T2	(4) T3	(5) T1 x Attrition	(6) T2 x Attrition	(7) T3 x Attrition
Household size	2,916	0.191	0.082	0.036	0.870**	-0.231	0.041
		(0.127)	(0.116)	(0.128)	(0.416)	(0.344)	(0.365)
Number of children under the age of 5	2,916	0.008	-0.000	0.011	0.071	-0.074	-0.038
		(0.032)	(0.032)	(0.035)	(0.132)	(0.125)	(0.107)
Log of total monthly consumption expenditure per adult equivalent	2,888	-0.005	-0.011	0.032	-0.046	0.073	-0.062
		(0.057)	(0.061)	(0.062)	(0.174)	(0.148)	(0.137)
Age of primary male	2,813	0.654	0.279	0.396	2.111	-0.318	-0.730
		(0.567)	(0.538)	(0.504)	(2.973)	(2.070)	(2.230)
Primary male is the household head	2,813	-0.016	0.006	0.011	-0.003	0.047	0.021
		(0.025)	(0.020)	(0.018)	(0.082)	(0.051)	(0.052)
Primary male has some education	2,812	0.001	-0.001	0.008	0.047	0.108	0.045
		(0.030)	(0.034)	(0.031)	(0.156)	(0.127)	(0.102)
Primary male's main activity is crop production	2,811	-0.003	0.012	0.027	0.080	0.042	0.013
		(0.038)	(0.035)	(0.036)	(0.111)	(0.107)	(0.099)
Primary male is married and lives with their spouse	2,800	0.006	-0.012	0.015	-0.069	-0.013	-0.004
		(0.020)	(0.022)	(0.018)	(0.062)	(0.051)	(0.043)
Primary male finds it acceptable for a woman to travel alone to market, health center, to visit a friend	2,646	0.036	0.022	0.036	-0.195	0.239*	0.105
		(0.040)	(0.042)	(0.043)	(0.125)	(0.140)	(0.132)
Primary male doesn't find violence justified in any situation asked about	2,646	-0.040	-0.036	-0.063**	0.018	0.062	0.030
		(0.030)	(0.029)	(0.029)	(0.105)	(0.088)	(0.107)
Total number of childcare activities primary male participated in (0-10)	2,800	-0.043	0.054	0.115	0.475	-0.213	-0.817
		(0.160)	(0.176)	(0.164)	(0.772)	(0.617)	(0.589)

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

Appendix Table 4: Primary male's involvement in domestic tasks in the past 3 days
Midline to endline panel sample

	Midline			Endline			Midline vs endline		
	(1) Helped with household chores (e.g. cleaning)	(2) Helped with cooking or meal preparation	(3) Helped with collecting firewood and water	(4) Helped with household chores (e.g. cleaning)	(5) Helped with cooking or meal preparation	(6) Helped with collecting firewood and water	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.035	0.056*	0.083**	0.151***	0.123***	0.193***	0.010	0.053	0.017
	(0.034)	(0.031)	(0.036)	(0.046)	(0.035)	(0.034)			
	[0.299]	[0.083]	[0.033]	[0.003]	[0.001]	[0.000]			
T2	0.048	0.057**	0.077**	0.029	0.001	0.102***	0.606	0.095	0.554
	(0.035)	(0.027)	(0.034)	(0.034)	(0.028)	(0.033)			
	[0.250]	[0.076]	[0.070]	[0.467]	[0.982]	[0.014]			
T3	0.082**	0.095***	0.095**	0.154***	0.130***	0.125***	0.077	0.376	0.478
	(0.032)	(0.029)	(0.038)	(0.040)	(0.031)	(0.035)			
	[0.013]	[0.002]	[0.013]	[0.000]	[0.000]	[0.001]			
Test: T1 = T2	(0.715)	(0.977)	(0.870)	(0.006)	(0.000)	(0.003)			
	[0.977]	[0.977]	[0.977]	[0.011]	[0.002]	[0.010]			
Test: T2 = T3	(0.322)	(0.184)	(0.640)	(0.001)	(0.000)	(0.444)			
	[0.483]	[0.368]	[0.640]	[0.002]	[0.000]	[0.533]			
Test: T1 = T3	(0.152)	(0.223)	(0.761)	(0.941)	(0.809)	(0.024)			
	[0.447]	[0.447]	[0.941]	[0.941]	[0.941]	[0.144]			
Mean of control	0.342	0.227	0.487	0.292	0.227	0.558			
N	2,260	2,260	2,261	2,260	2,260	2,261			

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 5: Primary male's gender equitable attitudes
Midline to endline panel sample

	Midline			Endline			Midline vs endline		
	(1) A husband is not justified in beating his wife in any of these situations [1]	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(3) Index of support for equitable gender norms	(4) A husband is not justified in beating his wife in any of these situations [1]	(5) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(6) Index of support for equitable gender norms	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.041	0.048	0.713**	0.071**	0.055	0.921**	0.465	0.863	0.650
	(0.036)	(0.031)	(0.345)	(0.033)	(0.041)	(0.404)			
	[0.252]	[0.183]	[0.081]	[0.081]	[0.218]	[0.081]			
T2	0.039	0.035	0.426	-0.009	0.008	-0.289	0.169	0.468	0.073
	(0.036)	(0.032)	(0.352)	(0.032)	(0.035)	(0.363)			
	[0.554]	[0.554]	[0.554]	[0.817]	[0.817]	[0.639]			
T3	0.031	0.060*	0.302	0.055	0.014	0.440	0.545	0.239	0.733
	(0.037)	(0.033)	(0.360)	(0.033)	(0.042)	(0.361)			
	[0.484]	[0.301]	[0.484]	[0.301]	[0.742]	[0.450]			
Test: T1 = T2	(0.940)	(0.682)	(0.356)	(0.011)	(0.203)	(0.001)			
	[0.940]	[0.819]	[0.534]	[0.034]	[0.406]	[0.004]			
Test: T2 = T3	(0.820)	(0.432)	(0.707)	(0.037)	(0.881)	(0.017)			
	[0.881]	[0.865]	[0.881]	[0.110]	[0.881]	[0.102]			
Test: T1 = T3	(0.768)	(0.694)	(0.205)	(0.585)	(0.224)	(0.079)			
	[0.768]	[0.768]	[0.447]	[0.768]	[0.447]	[0.447]			
Mean of control	0.607	0.340	8.800	0.668	0.429	10.354			
N	2,260	2,270	2,275	2,260	2,270	2,275			

[1] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband.

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 6: Primary male's involvement in domestic tasks in the past 3 days as reported by primary female and primary male at endline

	Female's report			Male's report			Female's vs male's reports		
	(1) Spouse helped with household chores (e.g. cleaning)	(2) Spouse helped with cooking or meal preparation	(3) Spouse helped with collecting firewood and water	(4) Helped with household chores (e.g. cleaning)	(5) Helped with cooking or meal preparation	(6) Helped with collecting firewood and water	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.119***	0.104***	0.215***	0.151***	0.127***	0.208***	0.364	0.497	0.828
	(0.040)	(0.036)	(0.039)	(0.046)	(0.035)	(0.035)			
	[0.004]	[0.005]	[0.000]	[0.002]	[0.001]	[0.000]			
T2	-0.002	0.012	0.056*	0.030	0.003	0.104***	0.315	0.716	0.111
	(0.032)	(0.025)	(0.029)	(0.034)	(0.027)	(0.033)			
	[0.947]	[0.920]	[0.155]	[0.773]	[0.947]	[0.011]			
T3	0.090**	0.101***	0.166***	0.146***	0.138***	0.144***	0.073	0.233	0.500
	(0.036)	(0.030)	(0.037)	(0.038)	(0.029)	(0.035)			
	[0.013]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]			
Test: T1 = T2	(0.002)	(0.007)	(0.000)	(0.007)	(0.000)	(0.001)			
	[0.002]	[0.007]	[0.000]	[0.007]	[0.001]	[0.001]			
Test: T2 = T3	(0.006)	(0.001)	(0.002)	(0.001)	(0.000)	(0.188)			
	[0.008]	[0.002]	[0.002]	[0.002]	[0.000]	[0.188]			
Test: T1 = T3	(0.411)	(0.907)	(0.173)	(0.902)	(0.706)	(0.039)			
	[0.822]	[0.907]	[0.519]	[0.907]	[0.907]	[0.237]			
Mean of control	0.198	0.153	0.366	0.294	0.227	0.551			
N	3,070	3,070	3,069	2,452	2,450	2,451			

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 7: Primary Male's involvement in domestic tasks in the past 3 days
Excluding households where a female was eligible for IPT-G

	Midline			Endline			Midline vs endline		
	(1) Male reports that he helped with household chores (e.g. cleaning)	(2) Male reports that he helped with cooking or meal preparation	(3) Male reports that he helped with collecting firewood and water	(4) Male reports that he helped with household chores (e.g. cleaning)	(5) Male reports that he helped with cooking or meal preparation	(6) Male reports that he helped with collecting firewood and water	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.035	0.048	0.047	0.146***	0.137***	0.185***	0.009	0.014	0.003
	(0.030)	(0.030)	(0.035)	(0.044)	(0.035)	(0.035)			
	[0.246]	[0.166]	[0.215]	[0.002]	[0.000]	[0.000]			
T2	0.047	0.060**	0.059*	0.027	0.023	0.097***	0.567	0.290	0.385
	(0.031)	(0.026)	(0.033)	(0.036)	(0.029)	(0.034)			
	[0.197]	[0.062]	[0.163]	[0.460]	[0.460]	[0.027]			
T3	0.076**	0.103***	0.074*	0.131***	0.140***	0.125***	0.144	0.344	0.248
	(0.030)	(0.026)	(0.038)	(0.037)	(0.031)	(0.036)			
	[0.015]	[0.000]	[0.052]	[0.001]	[0.000]	[0.001]			
Test: T1 = T2	(0.691)	(0.677)	(0.744)	(0.006)	(0.001)	(0.005)			
	[0.744]	[0.744]	[0.744]	[0.011]	[0.006]	[0.011]			
Test: T2 = T3	(0.348)	(0.096)	(0.685)	(0.003)	(0.000)	(0.373)			
	[0.448]	[0.193]	[0.685]	[0.010]	[0.001]	[0.448]			
Test: T1 = T3	(0.169)	(0.062)	(0.493)	(0.686)	(0.917)	(0.052)			
	[0.338]	[0.186]	[0.740]	[0.823]	[0.917]	[0.186]			
Mean of control	0.339	0.238	0.521	0.312	0.228	0.579			
N	2,629	2,628	2,629	2,083	2,082	2,083			

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 8: Primary Male's gender equitable attitudes
Excluding households where a female was eligible for IPT-G

	Midline			Endline			Midline vs endline		
	(1) A husband is not justified in beating his wife in any of these situations [1]	(2) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(3) Index of support for equitable gender norms	(4) A husband is not justified in beating his wife in any of these situations [1]	(5) It is acceptable for a woman to travel alone to market, health center, and to visit friends	(6) Index of support for equitable gender norms	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.001	0.053*	0.805**	0.052	0.027	0.823**	0.248	0.556	0.967
	(0.033)	(0.030)	(0.347)	(0.036)	(0.044)	(0.380)			
	[0.965]	[0.163]	[0.094]	[0.226]	[0.659]	[0.094]			
T2	0.037	0.049*	0.454	0.006	0.001	-0.248	0.377	0.237	0.070
	(0.030)	(0.028)	(0.357)	(0.034)	(0.039)	(0.349)			
	[0.440]	[0.440]	[0.440]	[0.970]	[0.970]	[0.715]			
T3	0.032	0.028	0.286	0.070**	0.004	0.486	0.335	0.589	0.617
	(0.030)	(0.029)	(0.347)	(0.034)	(0.043)	(0.339)			
	[0.493]	[0.493]	[0.493]	[0.267]	[0.920]	[0.459]			
Test: T1 = T2	(0.289)	(0.894)	(0.309)	(0.168)	(0.520)	(0.001)			
	[0.464]	[0.894]	[0.464]	[0.464]	[0.624]	[0.006]			
Test: T2 = T3	(0.856)	(0.448)	(0.626)	(0.040)	(0.939)	(0.007)			
	[0.939]	[0.895]	[0.939]	[0.120]	[0.939]	[0.039]			
Test: T1 = T3	(0.372)	(0.413)	(0.123)	(0.579)	(0.536)	(0.180)			
	[0.579]	[0.579]	[0.539]	[0.579]	[0.579]	[0.539]			
Mean of control	0.613	0.347	8.970	0.657	0.430	10.412			
N	2,630	2,636	2,639	2,084	2,088	2,090			

[1] Situations asked about: if wife goes out without telling her husband, if wife neglects the children, if wife argues with husband, if wife burns the food, if wife to have sex with husband.

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

Appendix Table 9: Primary Male's childcare activities

	Midline			Endline			Midline vs endline		
	(1) Number of childcare activities (out of 7) in the last 3 days	(2) Took part in any childcare activities in the last 3 days	(3) Took part in any child feeding activities in the last 3 days[1]	(4) Number of childcare activities (out of 7) in the last 3 days	(5) Took part in any childcare activities in the last 3 days	(6) Took part in any child feeding activities in the last 3 days[1]	(7) Test of difference between (1) and (4)	(8) Test of difference between (2) and (5)	(9) Test of difference between (3) and (6)
T1	0.138	0.011	0.034	0.092	0.018	-0.015	0.787	0.805	0.509
	(0.099)	(0.024)	(0.026)	(0.157)	(0.013)	(0.074)			
	[0.397]	[0.773]	[0.397]	[0.773]	[0.397]	[0.836]			
T2	-0.021	0.007	0.021	0.053	-0.004	0.075	0.645	0.703	0.345
	(0.106)	(0.026)	(0.026)	(0.113)	(0.014)	(0.058)			
	[0.840]	[0.840]	[0.840]	[0.840]	[0.840]	[0.840]			
T3	-0.037	0.019	-0.011	-0.002	-0.011	-0.009	0.829	0.345	0.974
	(0.097)	(0.026)	(0.026)	(0.140)	(0.018)	(0.064)			
	[0.990]	[0.990]	[0.990]	[0.990]	[0.990]	[0.990]			
Test: T1 = T2	(0.172)	(0.876)	(0.676)	(0.775)	(0.132)	(0.201)			
	[0.401]	[0.876]	[0.876]	[0.876]	[0.401]	[0.401]			
Test: T2 = T3	(0.887)	(0.656)	(0.262)	(0.639)	(0.709)	(0.170)			
	[0.887]	[0.851]	[0.786]	[0.851]	[0.851]	[0.786]			
Test: T1 = T3	(0.108)	(0.761)	(0.137)	(0.406)	(0.070)	(0.914)			
	[0.273]	[0.914]	[0.273]	[0.608]	[0.273]	[0.914]			
Mean of control	2.218	0.794	0.678	1.728	0.983	0.469			
N	2,076	2,076	1,842	750	750	457			

[1] Excludes children who are exclusively breastfed.

Notes: Standard errors (in parentheses) are clustered at the kebele level. All models control for woreda level fixed effects, supplemental sample, and sample eligible for livelihood transfers. Endline specifications also control for whether the kebele was randomly assigned to the first or second phase rollout of the men's engagement program. False Discovery Rate corrected q-values are reported in brackets and computed by pooling all specifications included in the table. Asterisks indicate significance at the 10, 5 and 1 percent level and are calculated with respect to the standard errors.

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available **here**

They can be downloaded free of charge.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS

1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org