INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

CGIAR FOOD FRONTIERS AND SECURITY

CGIAR POLICY INNOVATIONS

IFPRI Discussion Paper 02356

September 2025

Displacement and Development

Evidence from a Graduation Program for Somalia's Ultra-Poor

Jessica Leight

Kalle Hirvonen

Naureen Karachiwalla

Deboleena Rakshit

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

Jessica Leight (j.leight@cgiar.org) is a Senior Research Fellow in the Poverty, Gender, and Inclusion (PGI) Unit at the International Food Policy Research Institute (IFPRI), Washington, DC.

Kalle Hirvonen (k.hirvonen@cgiar.org) is a Senior Research Fellow in IFPRI's PGI Unit, Washington, DC.

Naureen Karachiwalla (n.karachiwalla@cgiar.org) is a Research Fellow in IFPRI's PGI Unit, Nairobi, Kenya.

Deboleena Rakshit (d.rakshit@cgiar.og) is a Research Analyst in IFPRI's PGI Unit, Washington, DC.

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

While the population of internally displaced people around the world continues to grow, evidence around strategies to sustainably enhance livelihoods among IDPs remains extremely limited. We present findings from a randomized trial of an ultra-poor graduation program targeting IDPs in urban Baidoa, Somalia; the intervention pro-vided cash transfers, an asset transfer or technical training program, and facilitated savings groups. Our findings suggest that two years following program launch, the intervention has led to significant increases in consumption, assets, and savings; however, these effects seem to be driven almost exclusively by increased livestock production. An exploration of heterogeneous effect using generalized random forest methods further suggests that the positive effects of the treatment are dramatically larger for smaller households characterized by lower dependency ratios.

Keywords: Somalia, internally displaced people, ultra-poor graduation

Acknowledgments

Funding for this work was provided by the U.S. Agency for International Development (USAID), the CGIAR Initiative on Fragility, Conflict, and Migration as well as the CGIAR Policy Innovations Program. We thank the UPG team at World Vision and ACTED who have facilitated this ongoing collaboration with IFPRI, particularly Caitlin Whittemore, Andrew Mu-gobo, and Asrat Bekele Balcha. We have benefited from discussions with Asrat Bekele Balcha, Daniel O. Gilligan, Colton Parks, and Caitlin Whittemore. IRB approval for this study was granted by the International Food Policy Research Institute (IFPRI), protocol #00007490. The trial was registered with the AEA RCT Registry (AEARCTR-0009452).

1 Introduction

The global population of forcibly displaced people has doubled over the past decade to reach a record 120 million by 2024, and more than half of this population remain in their countries of origin as internally displaced persons (IDPs) (UNHCR, 2023a, 2024b). This sharp increase reflects the surging incidence of both violent conflict and climate-induced disasters (IDMC, 2024; Rustad, 2024; WMO, 2024). Many forcibly displaced people — 90% of whom reside in low- and middle-income countries —- then remain trapped in protracted displacement for years or even decades (UNHCR, 2024a), often living side by side with host communities facing their own economic and social challenges (UNHCR, 2023b).

Aid in forced displacement contexts has traditionally focused on providing regular transfers of cash, food, or vouchers to support household consumption (Aker, 2017; Hidrobo et al., 2014; Altındağ and O'Connell, 2023), without addressing deeper, interconnected drivers of poverty among the displaced, including lost assets, inadequate shelter, poor physical and mental health, uncertain prospects for future residence, and overlapping market failures. Effectively targeting these multiple constraints requires a multifaceted intervention, suggesting that graduation model programs — a sequenced set of interventions including consumption support, training, access to savings or credit, and an asset transfer — may be a promising strategy. There is strong evidence about the effects of these programs in reducing poverty and increasing investments in more stable settings (Banerjee et al., 2015; Bandiera et al., 2017), but very limited evidence in forced displacement contexts (Rozo and Grossman, 2025), where complementary public services are often lacking and high levels of ongoing uncertainty and sometimes violence could render households unwilling to invest.

This study presents findings from a randomized controlled trial in the city of Baidoa, Somalia, evaluating an ultra-poor graduation (UPG) program targeted to IDPs. Baidoa is one of Somalia's largest displacement hubs, hosting more than 600,000 IDPs who have fled protracted violence or droughts; broadly, in Somalia, over 70% of IDPs live below the extreme poverty line (Pape, 2017). Within this context, UPG provided six months of cash transfers, followed by a choice between an asset transfer —– such as livestock or agricultural inputs — or enrollment in technical and vocational education training (TVET) to support livelihoods development. Additional components included the formation of savings groups and group-level coaching (focused on financial literacy, business skills, and social capital).

The trial included a sample of 4,116 households identified as eligible for the intervention due to their baseline vulnerability to hunger and their residence in the targeted IDP sites; given that the number of eligible households exceeded those who could be served given resource constraints, households were randomly selected to enter the UPG program.[1] Using baseline data from 2022 and two follow-up surveys conducted in 2023 and 2024 and characterized by minimal attrition (less than 4% of households), we exploit the randomized design to estimate the program's causal impact on consumption, food security and livelihood outcomes.

Our primary findings suggest that two years following its launch, the UPG intervention had substantial effects on a range of consumption and livelihoods activities, including a 30% increase in consumption (consistent across both food and non-food consumption), a 300% increase in the value of assets (driven almost entirely by goats), and a nearly 50 percentage point increase in the probability of reporting any savings. Treated households are more likely to report income from a range of sources, though the largest increase is observed in livestock. There are also positive effects on respondent locus of control, and no adverse effects on local social cohesion despite the household-level randomization. Longitudinal data suggests that positive effects were evident even one year post-launch, and generally widened over the second year of implementation. Attrition in this trial was extremely low but concentrated in the control arm: accordingly, we also construct bounds on the treatment effects of interest following Lee (2005) and Kling et al. (2007) and find the estimated effects are generally robust to both Lee trimming and allowing attritors to deviate from the treatment arm-specific mean

[1] There were 6,323 eligible households and only 5,000 households could be served by the intervention; a subset of these intervention households entered the trial. Appendix A3 provides a detailed discussion of the ethical aspects of this trial, based on the structured ethics appendix suggested by Asiedu et al. (2021).

by up to two standard deviations.

We also explore heterogeneity in the estimated treatment effects following the generalized random forest method proposed by Athey et al. (2019). We find that there is substantial heterogeneity in the estimated conditional average treatment effects, primarily predicted by variation in household composition at baseline (the average household size is nearly seven, with a standard deviation of over two). Households that are smaller, and characterized by a lower ratio of dependents to prime aged adults, show significantly larger treatment effects. In fact, households with five or fewer members and a dependency ratio in the lowest quartile exhibit a gain in consumption that is around 30% larger than the average treatment effect, suggesting that households with more care responsibilities may not be able to effectively take advantage of new livelihoods opportunities. We further demonstrate that a comparable pattern of heterogeneity is not observed in replication data from Banerjee et al. (2015), an evaluation of a graduation model intervention implemented in six stable, non-displacement sites, suggesting that this pattern may be somewhat distinct to displacement settings, requiring consideration in program design in such settings.

Our paper contributes to the growing literature evaluating the effectiveness of graduation model programs (Banerjee et al., 2015, 2022, 2021; Bandiera et al., 2017; Balboni et al., 2022). While most previous studies have been conducted in stable, rural contexts, ours is among the first to evaluate a graduation program in a setting marked by forced displacement. Table A1 provides a overview of this literature and key characteristics of the setting, empirical design, and interventions, highlighting the absence of evidence from conflict-affected contexts. Moreover, whereas previous research primarily evaluates the performance of these programs in rural areas, this trial contributes evidence from an urban setting. This trial, together with a study from post-conflict Afghanistan (Bedoya et al., 2019), indicate that graduation models may be particularly effective in fragile or humanitarian settings, demonstrating relatively large impacts on household economic outcomes.

This pattern of relatively large impacts aligns with findings from a parallel literature on economic interventions targeting forcibly displaced populations. Despite the scale of global displacement, there remains a striking lack of rigorous evidence on the effectiveness of livelihood and poverty alleviation programs for these populations—particularly IDPs (Rozo and Grossman, 2025; Schuettler and Caron, 2020). Our study provides one of the first experimental evaluations of a multifaceted livelihoods program for IDPs. Related studies focusing on cash and cash+ interventions in refugee settings also report promising results. In Uganda, Gupta et al. (2024) conduct a randomized trial evaluating the economic impacts of a one-off unconditional $1,000 USD transfer directed at South Sudanese refugees residing in refugee camps, showing notable improvements in household consumption (11% relative to control group), asset accumulation (30%), and business revenue (64%) after 18 months. Also in Uganda, Baseler et al. (2024) experimentally evaluate a program combining $540 USD cash grants with mentorship for young urban microentrepreneurs, including both refugees and host community members. They find that the cash grant alone significantly boosts business profits and household earnings one year after implementation, but the mentorship component provides no additional benefits. Finally, in Kenya, MacPherson and Sterck (2021) use a regression discontinuity approach to compare a traditional humanitarian model of refugee assistance with a development-oriented approach, finding improvements in nutrition and food security after 16 months, likely driven by greater participation in small-scale agriculture.[2]

2 Context and Intervention

2.1 Context

Baidoa is a city in southwestern Somalia, serving as the capital of the South West State and located approximately 250 kilometers from the capital Mogadishu (see Figure A1a). It is one of Somalia's largest urban centers and an important hub for agriculture and livestock

[2]Another closely related strand of work focuses on assessing the impact of programs and legislation that facilitate refugee integration into local labor markets: see for example Sarvimäki and Hämäläinen (2016); Battisti et al. (2019); Fasani et al. (2021); Hussam et al. (2022).

trade (UN-Habitat, 2021). While population estimates vary widely due to the lack of recent census data and the large-scale influx of IDPs over the past decade, imputed estimates based on a 2014 survey suggest a population of around 300,000 in that year (UN-Habitat, 2021). However, ongoing conflict and recurring droughts have led to large-scale displacement, extensively increasing the city's population. In 2022, the number IDPs in Baidoa was estimated to be nearly 600,000 (UNHCR, 2022), up from 169,000 in 2017 (UNHCR, 2017) and constituting one of the largest IDP populations within Somalia. Though a number of humanitarian and development organizations operate in the city, security remains fragile, with Al-Shabaab, a designated terrorist group, maintaining influence in surrounding rural areas (UN-Habitat, 2021) from which households are often displaced.

2.2 Intervention

The project evaluated here was the Ultra-poor Graduation (UPG) Program, implemented by World Vision over three years and funded by the U.S. Agency for International Development (USAID) Bureau for Humanitarian Assistance (BHA).[3] UPG supported ultra-poor and vulnerable households in graduating from extreme poverty and moving toward self-reliance, targeting primarily IDPs as well as a small number of households from vulnerable host communities, returnees, and refugees.[4] Figure A1b maps the Baidoa IDP sites targeted for this program and thus included in the evaluation.

Launched in June 2022 and running until December 2024, the intervention consisted of four main components. First, UPG households received six monthly unconditional cash transfers of $42.50 to provide consumption support. Second, they participated in savings groups designed to encourage savings; these regular meetings also served as a platform for training on topics such as financial literacy and business management. Third, households

[3] The formal title was *Building Pathways Out of Poverty for Ultra-Poor IDPs and Vulnerable Host Communities in Baidoa*.

[4] Returnees are Somali households who had been refugees in other countries and then returned; Baidoa is not necessarily their home city, however. The very small number of international refugees in Somalia are generally from Ethiopia and Yemen.

received either a one-time asset transfer or funding to enroll in a six-month technical training course, based on their preference. Households opting for the asset transfer could choose from goats, chickens, sheep, cows, crop seeds, or tools; the technical training was provided through a local institute. Fourth, participants attended regular group-based coaching sessions focused on life skills and social integration. A summary of the intervention components compared to other similar interventions is provided in Table A1.

Program eligibility was determined based on household characteristics assessed in an initial vulnerability assessment, and eligible households had to meet two criteria: they were classified as experiencing moderate or severe hunger according to the Household Hunger Scale (Ballard et al., 2011), and they had resided in the IDP site for at least one month.[5] The initial assessment identified 6,323 eligible households.

3 Methods

3.1 Experimental Design

We employ a randomized controlled trial using randomization at the household level, given that UPG by design had resources to serve only 5,000 households among a larger number of eligible households. The control arm was expected to include 1,500 households, but was reduced to 1,323 households given the number of eligible households to ensure that the intervention target of 5,000 was met. Randomization was conducted by the research team in Stata prior to the baseline survey using data from the initial vulnerability assessment.[6]

[5]These criteria were identified as important for program eligibility by the program team and validated through focus group discussions with participants and community members. Households experiencing hunger were prioritized due to their higher level of need, while requiring at least one month of residency ensured greater stability, increasing the likelihood that participants would remain at the site.

[6]Randomization was stratified by four groups based on two binary variables: a binary variable for a household being above or below the median of an asset index constructed using vulnerability assessment data, and a binary variable equal to one if the household had resided in the IDP site for more than a year. The initial design also conducted random assignment to two treatment arms, to facilitate an analysis of two alternate household coaching strategies. However, ultimately only one coaching strategy — group-based coaching — was utilized, and thus the two treatment arms are pooled in analysis.

3.2 Ethics

Ethical approval for this study was granted by the International Food Policy Research Institute (IFPRI) Institutional Review Board (IRB), under protocol #00007490. Further ethical considerations, including policy equipoise, risks, and informed consent, are detailed in the structured ethics appendix (Appendix A3), following Asiedu et al. (2021).

3.3 Surveys

The baseline survey was conducted between May 18 and June 12, 2022 and was targeted to include 3,000 treatment households and all 1,323 control households; the remaining 2,000 treatment households were not targeted for inclusion in the survey or evaluation.

The realized sample included 4,116 households (2,872 treatment and 1,244 control).[7] The first follow-up survey was conducted 14 months post-baseline and one year after the launch of the UPG intervention and resurveyed 4,089 (99.3%) sample households. The second follow-up survey was conducted 28 months post-baseline and two years following UPG launch, and resurveyed 3,982 (96.7%) sample households. All interviews were conducted in person. Due to security risks and the limited survey capacity in the study area, all surveys were carried out by the implementing partner, World Vision, which was responsible for hiring and training the enumerators. As is common in insecure forced displacement settings (Pape and Mistiaen, 2020; Pape and Verme, 2023), interview durations had to be restricted (with a target of around 60 minutes per household), limiting the scope of the questionnaires. Figure A2 in the Appendix summarizes the study timeline.

[7] The sampling frame fielded included 2,980 treatment households and 1,323 control households (the latter comprised of all households in the target IDP sites not served by the intervention); the sampling frame of treatment households itself shrank slightly from the original planned 3,000 given that 20 households in the sample list were duplicates. Within the target sample, both treatment and control, 187 households were not interviewed because they either could not be reached during the designated survey period or declined to participate.

3.4 Outcomes

Table A10 in the Appendix provides definitions of all outcome variables analyzed, as prespecified in a registered analysis plan.[8] The primary outcomes include the share of households characterized by moderate or severe hunger over the last 30 days based on the Household Hunger Scale; household per capita consumption in 2017 Purchasing Power Parity (PPP) dollars; and the estimated value of household assets, also valued in 2017 PPP dollars.[9] Table A4 reports the estimated minimum detectable effect sizes for the primary outcomes based on the mean and standard deviations observed in the main follow-up survey. The experiment is adequately powered to detect even relatively small treatment effects: a 0.05 change in the share of households with moderate or severe hunger (11.5% of the control mean), a 0.15 $PPP change in per capita consumption (5.7% of the control mean), and a 35 $PPP change in the total value of assets (14% of the control mean).

Secondary outcomes include additional, more detailed measures of assets and financial inclusion; income data capturing whether households report income from any one of six sources, and the amount of income; the livelihood coping strategies score, capturing strategies household may need to use to adapt to shortages of food or money; and measures of social cohesion and locus of control. The index of social cohesion was constructed using a series of questions about the individual's perception of the broader community (Humble et al., 2023; Catholic Relief Services, 2019); and locus of control was measured following Rotter (1966) and Malacarne (2024).

[8] The trial registration number is AEARCTR-0009452.

[9] Consumption was measured using modules adapted from the 2017 Somali High-Frequency Survey (Pape (2017)) conducted by the World Bank, but shortened to manage interview length; additional details on the construction of the consumption measure are provided in Appendix A4.2. The 2017 base year is chosen because PPP at the time of the baseline survey it was the latest available International Comparison Program (ICP) benchmark year. PP Conversion factors are most reliable in benchmark years when the ICP conducts comprehensive price surveys (e.g., 2011, 2017, 2021). PPPs for non-benchmark years are typically interpolated using domestic and U.S. CPIs and may be subject to greater uncertainty.

3.5 Econometric Specification

We estimate the following specification:

$$y_{i,t} = \alpha + \beta \, \text{Treatment}_i + \gamma \, y_{i,0} + \lambda_i + \epsilon_{i,t},$$

where $y_{i,t}$ is an outcome for household i in year t, Treatment_i is an indicator for assignment to the UPG program, $y_{i,0}$ is the baseline measure of the outcome (if available), and λ_i are fixed effects for randomization strata. For outcomes that were not measured at baseline, we estimate the same model excluding $y_{i,0}$.[10]

Since assignment to treatment was not clustered, standard errors are adjusted for heteroskedasticity following White (1980). We also report q-values corrected for multiple hypothesis testing (MHT) following Anderson (2008). We conduct MHT corrections within the set of primary outcomes, and within the set of each family of secondary outcomes. We also report average standard treatment effects for broader outcome families that pool across primary and secondary outcomes (consumption and food security, assets and savings, and income), following Kling et al. (2007). Although our prespecified main specification does not include additional baseline covariates, we also report an alternative specification that uses double lasso for covariate selection (Belloni et al., 2013; Cilliers et al., 2024).

4 Empirical Findings

4.1 Baseline Characteristics and Balance

To characterize the sample, Table 1 summarizes key demographic characteristics at baseline and reports balance across the control and treatment arms. Eighty-three percent of households are IDPs (with the remainder including 7% refugees, 1% returnees, and 9% host

[10]Baseline values are reported only for the Livelihoods Coping Scale, tropical livestock units, and any savings; baseline HHS was also measured, but there is no variation in this measure at baseline since all eligible households were characterized by moderate or severe hunger.

community members), and the average household includes nearly seven members, of whom four are children. Sixty percent of households report having a pregnant or lactating woman, and 20% report the presence of an individual with a disability. Unsurprisingly, sampled households were characterized by a high level of food insecurity and a high level of deprivation at baseline: the average HHS score was nearly four, consistent with the eligibility criteria of moderate to severe hunger; less than 1% of households reported any savings, and households owned around .2 tropical livestock units on average. Total baseline asset value was estimated to be at around $350.)[11]

Out of the 11 t-tests comparing baseline characteristics across the two study arms, only one shows a statistically significant difference: household size is modestly larger in the control group, by approximately 0.3 members (or 5%), and this difference is highly statistically significant. When we estimate a joint test of balance across covariates, the p-value corresponding to the null hypothesis of no significant imbalance is 0.053 when using conventional p-values or 0.079 using randomization inference following Kerwin et al. (2024); we will explore further in the robustness checks below the possibility of any bias in the estimated treatment effects due to the imbalance detected in baseline household size.

4.2 Implementation Fidelity

Table A2 summarizes findings around implementation fidelity that suggest that in general, the program was carefully implemented in line with the randomized design. 88% of treated households reported receiving cash transfers from a non-governmental organization in the past three years, compared to 14% of control households (who may plausibly also be reporting transfers received from other NGO programs in the same recall period). Those who do report receiving transfers reported six transfers of $42 each, for a total transfer of around $250.

Similarly, 87% of households assigned to treatment reported receipt of either assets or TVET training over the past three years, compared to only 3% of control households. The

[11] Note that as asset price data was not collected at baseline, assets are valued at prices collected in the second follow-up survey, using the median price measured within the sample.

assets track seems to have been slightly more popular than the TVET track, based on households' self-reports: 41% report receipt of assets, and 30% report receipt of TVET (9% report receipt of both, an allocation that was not generally allowed under program guidelines and suggests that households may be mis-identifying another service as one provided by UPG). For households reporting assets, goats were the dominant choice (reported by nearly 80% of households) as evident in Figure A3a, while for households reporting training, the most popular courses were tailoring, tie dying, and beauty salon services (Figure A3b). Reported participation in savings groups and coaching is, however, somewhat lower: 63% of treatment households report participation in savings groups and 49% in coaching, compared to minimal participation in the control arm.

We also separately assessed whether there were any cross-household spillovers from households in the treatment to the control arm, and observed these were very rare. Fewer than 2% of control households reported receiving cash remittances from any other household (whether a program beneficiary or not). Fewer than 20% of treatment households reported that they had transferred or loaned the asset they received to any other household. While informational spillovers may have been more common (a majority of treatment households stated they shared information received in training, though this is only vaguely defined), other direct forms of spillovers seem to be infrequent.

4.3 Primary Findings

The primary treatment effects are reported in Table 2; Panel A reports effects on consumption and food security, Panel B reports effects on assets and financial inclusion, Panel C reports effects on income, and Panel D reports effects on social cohesion and locus of control.[12] The findings in Panel A suggest the intervention led to large positive shifts in consumption, with an increase in per capita consumption of around 30% ($0.81) in absolute terms: the relative magnitude of this effect is consistent comparing across both food and

[12] Again, the prespecified primary outcomes are per capita consumption, household hunger scale status, and asset value, or Columns (1) and (4) of Panel A and Column (1) of Panel B.

non-food consumption. There is also a dramatic decline in the probability of households being characterized by moderate or severe food insecurity according to the household hunger scale: recall that at baseline, all households were identified as eligible based on this criteria. Two years later, 42% of households in the control arm continue to experience this high level of food insecurity, while this has declined to only 12% in the treatment arm. The livelihoods coping score also shows a decline of about a third, consistent with the previous findings and suggesting that many households exposed to UPG are not having to resort to adverse measures (such as selling assets, borrowing, or withdrawing children from school) to obtain food.

Panel B documents effects on assets and financial inclusion. The average asset value in dollars has roughly tripled in the treatment arm (reaching nearly $900, compared to under $300 in the control arm) and this is substantially driven by livestock, where the average number of tropical livestock units (TLUs) increases by nearly fourfold to around .49 TLUs. (In practice, this corresponds to a gain of roughly three goats at the second follow-up; given reported market prices, this is an increase in asset value of nearly $500.) Panel A of Table A5 in the Appendix reports effects on the count estimates for a whole range of assets, and it is evident that there are significant treatment effects on a large number of asset categories (mobile phones, various productive tools, and other livestock). These estimates are generally quite small in absolute terms (none exceeds .5 other than the estimate for goats, and most are under .1), though in multiple categories the increase is proportionately large: i.e, treatment households increase their reported inventory of spades, tarpaulins, solar panels, and sheep by around 70%; the number of donkey carts and poultry owned increases by around 50%; and there is a sixfold increase in the number of sewing machines.

Returning to the main table, Column (3) in Panel B in Table 2 suggests there are also substantial effects on savings, as virtually no households (4%) in the control arm report cash savings, compared to nearly half of treatment households. The effects on credit access, while still positive, are less dramatic in magnitude (eight percentage points relative to a mean of

57%).

Panel C reports effects on income. There is a large effect on income from livestock: the probability of any reported income from cropping or livestock production increases by 16 percentage points relative to a mean of only 5% in the control arm, and treatment households report an average of $34 of income from cropping and livestock in the reference period of one month, relative to only $6 for control households. The effects for non-farm businesses are, however, modest: there is an increase in the probability of having a non-farm business is five percentage points relative to a mean in the control arm of 15 percentage points, but the increase in the continuous measure of income is extremely small (2% of the control mean) and statistically insignificant. There is similarly no overall shift in the probability of wage income, though a further decomposition shows that there is some shift away from informal to formal wage labor.

Panel D then reports two variables capturing shifts in social cohesion and locus of control. Because the intervention was individually randomized in an urban setting, there was an increased risk of adverse effects on social cohesion—particularly if some households were aware that others were receiving support while they were not. However, there is no evidence of this phenomenon here; we cannot reject the null that the treatment affected social cohesion. The estimated treatment effect on locus of control is notably positive and significant, suggesting some shifts in the psychological outlook of households linked to their enhanced economic status. [[[Naureen can add evidence here from other studies that find similar effects.]]]

To capture some key effects graphically, Figure 1 shows group means and treatment effect estimates for the three primary outcomes as well as tropical livestock units across the two follow-up surveys, allowing us to track how impacts evolved over time. Panel A shows a steady decline in moderate or severe hunger, with the largest gains in the first year. Households in the control arm saw some early gains but little shift thereafter, leading to a widening gap between the two groups. Panel B presents total per capita consumption (not measured at baseline), which shows modest gains for the treatment group by the first follow-up and a much larger gap in the following year. Growth in asset value can be assessed by employing asset prices as measured in the second follow-up survey: while control households saw flat or declining asset values, treated households accumulated assets steadily, and similarly for tropical livestock units.

We report average standard treatment effects following Kling et al. (2007) to facilitate interpretation of general treatment effects across categories: consumption and food security, assets and financial inclusion, income, and social cohesion and locus of control.[13] These findings are reported in Table 3 and highlight the wide variation in effect sizes across outcome families. The positive effect on consumption and food security is fairly large at around .28 standard deviations, but the effect on income is only around .05 standard deviations and statistically insignificant; both are dwarfed by the positive effect on asset and financial inclusion, nearly 1.5 standard deviations. The effects on attitudinal variables are around 0.05 standard deviations, but insignificant.

We also report two additional exploratory analyses: the first is treatment effects on household size and composition. Although this analysis was not pre-specified, we consider it important given the evidence of baseline imbalance in household size, and because differential shifts in household composition have been documented in prior empirical evaluations of cash transfer programs for displaced households (Özler et al., 2021). As shown in Panel B of Table A5 in the Appendix, there is a small treatment effect for overall household size of .068 members, corresponding to less than a one percent increase in household size relative to the mean in the control arm at follow-up: the only (weakly) statistically significant shifts were in the number of young children (0-4 years of age) and adolescents (15-19 years of age). There was no significant change in the number of prime-age or older adults. Overall, these results suggest that the intervention did not meaningfully alter household composition, and it is unlikely that selective household entry or exit accounts for the main treatment effects.

Given the evidence of baseline imbalance in household composition, we also re-estimate

[13]The variable capturing any moderate or severe food insecurity is reverse-coded in this analysis.

our main specification using a double lasso routine to select baseline covariates as controls: the lasso uniformly selects a single variable, household size. (Baseline values of the outcome variable are still uniformly included as control variables, when available.)[14] Table A3 reports these specifications. While some treatment effects are slightly smaller (e.g., the coefficient on consumption), in general the differences are extremely minor. There is very little evidence that baseline imbalance led to any bias in the primary estimated effects.

4.4 Attrition

As previously noted, attrition was on average strikingly low in this trial, particularly for an IDP sample: fewer than 4% of households were lost to follow-up. However, despite this low rate, there is a meaningful difference across treatment arms as reported in Table A6: in Column (1), we observe that the rate of attrition is six percentage points lower among treatment households (1.4%, compared to 7.6% among control households). In Columns (2) and (3) of the same table, we then regress attrition on a set of baseline covariates (the same covariates previously reported in the balance table) and the interaction of these covariates with treatment. Only a few baseline characteristics predict attrition: households characterized by a higher HHS score at baseline (more intense hunger) are more likely to attrit, while households characterized by a higher LCS score (indicative of more intense use of coping strategies) and a higher dependency ratio are less likely to attrit. However, the interaction effects between baseline covariates and the treatment dummy are generally insignificant, implying that characteristics of the attrited do not appear to differ across arms, again with the exception of the dependency ratio: households with a higher dependency ratio are less likely to attrit in the control arm, but this relationship is zero in the treatment arm.

Nonetheless, to further explore the potential of any bias due to attrition, we also estimate bounds on the primary treatment effects using various strategies. For the continuous variables of interest, we first measure the attrition gap and construct bounds following Lee

[14]The fact that this routine selects only a single variable is by no means unusual, as extensively documented in Cilliers et al. (2014).

(2005): we estimate the difference in the proportion of non-missing observations between the treated and control groups and then create two counterfactual treated samples. To estimate the lower bound, we drop the treated units characterized by the highest outcome values until the attrition gap is exhausted, and to estimate the upper bound, we drop those characterized by the lowest outcome values. As an alternate strategy, we also follow Kling et al. (2007) to generate bounds assuming that attrited units in the treatment or control arm are characterized by outcomes N standard deviations above or below the treatment group specific mean, where N in this case is set to two: the upper bound is estimated by setting attrited units in the treatment arm to two standard deviations above the treatment arm mean and attrited units in the control arm to be two standard deviations below the control arm mean, and vice versa for the lower bound. This allows for wide disparities in outcomes comparing attrited and non-attrited individuals. For binary variables, we estimate simple Manski bounds.

The findings are presented in Table A7 and are generally consistent with our primary results. In Panel A, we can observe that the estimated treatment effects on consumption, assets, and the livelihoods coping score can be bounded away from zero, and with relatively tight variation in the estimated magnitude: i.e., the magnitude of the estimated effect for consumption ranges between 20% and 40%. In Panel B, we can see that even quite conservative Manski bounds allow us to reject the hypothesis of a null effect on moderate or severe hunger status, any savings, and any credit; we can also reject the hypothesis of a null effect for any agricultural or livestock income, though for the other income variables, the bounds cross zero. (Similarly, the estimated bounds cross zero for the continuous income variables, where the primary estimates were all somewhat noisy likely due to the large number of zeroes, and bounds also cross zero for locus of control; for concision, these are not reported in the table.)

4.5 Heterogeneous Effects

The previous graduation literature has generally used quantile regression to explore variation in treatment effects (Banerjee et al., 2015; Bandiera et al., 2017; Bedoya et al., 2019). However, this strategy does not identify the underlying drivers of heterogeneity, and simpler linear analyses focusing on specific baseline characteristics as predictors of heterogeneity (e.g., wealth, education), have so far provided limited evidence of consistent or interpretable moderators (Bedoya et al., 2019; Bossuroy et al., 2022). To address this, we use relatively recent machine learning methods to explore treatment effect heterogeneity using a generalized random forest (GRF) (Athey et al., 2019), allowing for a data-driven exploration of heterogeneity across a rich set of baseline covariates.

The GRF algorithm builds a causal random forest (CRF) that allows for the estimation of conditional average treatment effects, conditional on observable baseline characteristics. This method is arguably well suited to our trial, characterized by a large sample and individual-level randomization —- both important conditions for the effective application of causal forest methods (Wager and Athey, 2018; Davis and Heller, 2017).[15]

The first step is simply to assess how much heterogeneity in treatment effects is evident, focusing on per capita consumption as the primary outcome variable of interest. We estimate what is known as the "out-of-bag" conditional average treatment effect (CATE) — in which the treatment effect for each observation is predicted using only the trees for which that observation was not used in the training set — and present the distribution in Figure A4a. It is evident that there is substantial mass for an effect on consumption between around \$0.6 and \$0.85 (relative to the estimated mean effect of \$0.81), but there is also a right tail of consumption effects of more than a dollar. The cumulative distribution function shown in Figure A4b suggests that the top quartile of consumption effects is above \$0.77.

We then probe the variable importance estimated by the GRF algorithm: this captures

[15]We use the GRF algorithm in R, and draw on the useful replication code and discussion of applications of GRF in an RCT context provided in Sylvia et al. (2021).

the percentage of importance for each baseline covariate in the forest, as measured by the frequency with which this variable is used as a splitting variable. The findings reported in Table A8 in the Appendix suggest that the most important variables predicting heterogeneity are both linked to household composition: the number of members in the household, and the dependency ratio (defined as the ratio of the number of children under 14 and elderly over 55 to the number of adults and adolescents). Other, weaker, predictors of heterogeneity include baseline asset value and tropical livestock units. Figure A5 then captures how the estimated out-of-bag CATEs vary with respect to the three most predictive characteristics. The first two scatter plots capture a notable pattern of treatment effect heterogeneity in which the largest effects are observed for smaller households (four or fewer members) and those characterized by lower dependency ratios (under around two dependents per productive adult). There is also some weak evidence of heterogeneity in which households that were worse-off at baseline as measured by asset value are characterized by somewhat larger treatment effects, but this effect is relatively flat.

To then test heterogeneity using a simpler specification, we follow Sylvia et al. (2021) and estimate a standard heterogeneous effects regression using these key indicators of interest. Given the graphical pattern suggesting that larger treatment effects are concentrated in the left tail of household size and dependency ratio, we generate binary variables equal to one if the household is characterized by household size, dependency ratio, or baseline asset value below the 25th percentile at baseline and estimate heterogeneous effects, reported in Table 4. We can observe that households characterized by a smaller size (under five members) and a lower dependency ratio (under 2.5) show dramatically larger treatment effects, but for baseline asset value, the heterogeneity is lower in magnitude and not statistically significant. Column (4) presents the joint specification including all three variables and both household size and the dependency ratio remain strongly significant: a household characterized by both small size and a low dependency ratio would have a predicted treatment effect that is more than double a household characterized by large size and a high dependency ratio.

Overall, this analysis suggests that in predicting treatment effects of this intervention, demographics is truly destiny — and dominates other observable characteristics at baseline. Importantly, the transfers and other material support provided by UPG were fixed in size at the household level and did not scale with respect to family size; this is a common feature of graduation model interventions, as evident (for example) in the BRAC program handbook or in the detailed program descriptions provided in Banerjee et al. (2015).[16] That being said, the magnitude of the treatment effect heterogeneity; the fact that it is observed both for overall household size and for the dependency ratio; and the fact that this heterogeneity is observed around a year following the conclusion of transfers suggests that this is not solely a mechanical effect of greater intervention intensity, and may also reflect the fact that large households with many dependents were less able to exploit new livelihoods opportunities.

To further explore the potential generalizability of this pattern of heterogeneity, we also conduct an exploratory analysis of the data from the original six-site graduation trial reported in Banerjee et al. (2015) using the same generalized random forest method; we focus primarily on the question of what baseline variables predict heterogeneity in the treatment effect for per capita consumption in the longer-term (three-year) follow-up.[17] Our goal is to understand whether the pattern in which household size and demographics dominate in predicting conditional average treatment effects in our sample of Baidoa is consistent in a sample drawn from stable, non-displacement, rural contexts.

We first conduct the GRF analysis using a large set of covariates available at baseline, including a number of variables (baseline consumption and income) that were not collected at baseline in the Baidoa trial; we then restrict to the set of covariates that are also available in the Baidoa baseline data, and report findings from the "full" and "reduced" model, respectively. The findings suggest that there is also meaningful heterogeneity in the estimated treatment effects in the Banerjee et al. trial: the simple intent-to-treat estimate on

[16]Only one site had a program feature that was adjusted for household size, consumption support as provided in Ghana.

[17]We follow the primary specification described in the paper, using both binary variables for country and randomization strata and clustering at the randomization unit level.

long-term consumption is \$3.36 per person per month, but the estimated CATEs range from \$1 to nearly \$6 in 2014 PPP terms, as captured in Figure A6 presenting presents the density functions of the conditional average treatment effect for the full and restricted model, respectively.

Table A9 then documents patterns of variable importance in the two models, focusing on the ten most important variables in each and sorted in accordance with their rank in the full model.[18] In the full model, baseline consumption, asset value, and income from revenue and agriculture are most predictive of variation in the conditional average treatment effect (weights of roughly 0.1 each); and in the restricted model, asset value remains the most predictive (0.24), with similar predictive power then from household size, food security, and perceived economic welfare (weights of 0.2 each).[19] The observed patterns of heterogeneity in the CATE with respect to these baseline characteristics seem somewhat nonlinear, as summarized in Figure A7, but for household size, a similar (though flatter) pattern is observed in the restricted model in which smaller households are characterized by larger treatment effects.[20] Overall, though, household size is not a dominant predictor of variation in the CATE; it has some predictive power, baseline economic characteristics seem to be somewhat more predictive.

This exploration suggests that the pattern observed in our data in which demographic characteristics are dramatically more predictive of variation in the conditional average treatment effect vis-a-vis baseline economic variables is distinct from the pattern observed in Banerjee et al. (2015) for a graduation program in a non-displacement settings. There are several potential explanations for this that are not mutually exclusive. First, household size may be uniquely important in a displacement setting, where households are large and have

[18] Binary variables corresponding to country and randomization strata are included in the random forest method, but parallel to the Baidoa analysis, the variable importance weights are re-normalized to exclude the weights estimated on randomization strata. "NA" in the final column denotes that that variable is not included in the restricted model. The weights in the full model do not sum to one because other variables outside the top ten are predictive.

[19] Data on household composition that would allow us to identify the number of dependents is not reported.

[20] The top 5% of outliers for each explanatory variable (consumption, assets, etc.) are truncated from the graph for clarity.

agglomerated or altered in composition either by necessity or as a coping mechanism in the face of shocks. Average household size at baseline in our sample is 6.7, not dramatically different from the mean (5.8) in the sample in Banerjee et al. (2015), though the dependency ratio is somewhat high: only 2.6 of these members, on average, are between the ages of 15 and 55.

Dependency ratios even in sub-Saharan Africa are typically below 100% (Cleland and Machiyama, 2017), though that is often using a different minimum age for elderly (65), and constructed using macro-level data that may not be directly comparable. In the Somali context, restrictive gender norms may pose a further challenge by limiting the economic engagement of women even of prime working age.

A second hypothesis is that baseline economic status (level of assets, food security) has little predictive power in a displacement context in which households have experienced a series of adverse shocks and thus where their current ownership of assets may be effectively random. Interestingly, the probability of owning any assets is estimated to be much higher in this sample compared to the Banerjee et al. sample (98% versus 42%), though the mean level of assets is higher in the Banerjee et al. sample due to higher asset ownership at the 75th percentile and above; Figure A8 summarizes kernel densities in both samples, with the important caveat that the specific questionnaire modules used to measure assets are not the same.[21] Accordingly, it is not the case that assets are non-predictive of treatment effects because no household in the Baidoa sample owns any assets; but it may be there is simply little predictive information captured in the existing distribution.

A third hypothesis is that constraints on economic activities linked to household size (and the number of dependents) may be more acute in an urban displacement context such as Baidoa, also characterized by some ongoing security risks. In a more typical, non-fragile, rural setting, the active care burden of both young and old dependents may be lower due to

[21] The continuous variables in the Banerjee et al. sample are expressed in 2014 purchasing power parity terms, while our variables are captured in 2017 purchasing power parity terms. We employ a simple adjustment using the U.S. GDP deflator and thus apply an adjustment factor of 1.05 to the Banerjee et al. estimates, bearing in mind this is only indicative.

reduced security risk for unsupervised dependents, and the existence of a typical informal network of care and support. That being said, it is somewhat surprising that this would pose a meaningful constraint given that the most important livelihoods activity for households benefiting from the UPG intervention in Baidoa is raising livestock, an activity typically undertaken at home and viewed as compatible with domestic and care responsibilities; one hypothesis is that marketing livestock for actual income generation is challenging for those constrained by the care of more dependents.

4.6 Cost-effectiveness

Given the evidence around the substantial positive effects of the UPG intervention in Baidoa, it is informative to consider its costs and explore cost-effectiveness. Here, we find the the intervention is in fact relatively expensive: the all-in cost per household is estimated at approximately $7,770 in 2017 PPP terms or $2,930 in 2017 nominal USD (to enable cross-country comparability, we express costs in both PPP and nominal terms).[22] This suggests the Baidoa UPG program is one of the highest-cost programs assessed in the literature (Table A1), though our cost estimate plausibly reflects an upper bound given that it is based on the total program budget including overhead, monitoring, and management costs often excluded from cost analyses.[23] For comparison, Banerjee et al. (2015) report 2017 PPP costs of approximately $5,600 in Ghana, $3,960 in Pakistan, and $4,420 in Peru. In Afghanistan, Bedoya et al. (2019) estimate the intervention cost at around $7,470 per household, a cost level comparable to Somalia (and perhaps not coincidentally, in a similarly challenging, conflict-affected setting).

[22] The original budgeted cost was $3,200 in 2022 nominal USD. We convert this estimate to Somali shillings, and adjust to the 2017 base year using IMF CPI data and World Bank exchange and PPP conversion factors.

[23] We were unable to conduct a detailed costing analysis, as this analysis was interrupted when many staff at the implementing partner were terminated following loss of USAID funding. Accordingly, the estimate here is based on the total program budget awarded to the World Vision-led consortium (with the IFPRI subaward removed) divided by the 5,000 households served. It includes all management, oversight, and monitoring costs—expenses often excluded from published cost estimates. Moreover, it assumes the entire budget was spent in Somalia, though in practice, some share likely supported international staff or external oversight.

In terms of effectiveness, the intervention effect size here for consumption (30% relative to the control mean) is among the largest effects observed in this literature as summarized in Table A1, larger than the effects observed in Banerjee et al. (2015) and comparable to Bedoya et al. (2019).[24] The effect on assets and financial inclusion — nearly 1.5 standard deviations — is much larger than the effect on comparable variables observed in Banerjee et al. (around 0.2 – 0.4 standard deviations). However, the effects in this trial are observed at relatively short duration (only two years post-program launch, rather than three or four), again as summarized in the same table. While this analysis should be interpreted cautiously given the absence of detailed cost data and thus our inability to verify that costs are measured comparably across these different trials, it suggests that UPG may be in a broadly similar (or slightly lower) range of cost-effectiveness vis-a-vis other parallel interventions, but only if the large effects observed in the short-term persist.[25]

5 Conclusions

Recognizing the inadequacy of regular consumption support in addressing the multiple challenges faced by displaced households, the 2016 New York Declaration for Refugees and Migrants called for a shift toward sustainable, development-oriented responses to forced displacement that promote economic self-reliance among refugees and IDPs (United Nations, 2016). Our study is among the first to experimentally assess an intervention closely aligned with this declaration, and our findings suggest that an integrated ultra-poor graduation model does lead to significant increases in consumption, assets, and financial inclusion for a vulnerable IDP sample in urban Somalia.

These effects are primarily driven by enhanced income from livestock—particularly goats—with little evidence of household livelihoods diversifying into non-agricultural activities. While

[24] Relative to standard deviations in the control arm, however, the effect of .1 standard deviations observed here is roughly similar to the pooled effect observed in the original Science trial (Banerjee et al., 2015).

[25] There are, of course, other examples of parallel interventions implemented at dramatically lower cost with high levels of effectiveness: for example, recent work in Niger by Bossuroy et al. (2022) assessing an intervention costing less than $600.

such gains offer a tangible path to improved welfare, they raise important questions about the sustainability of these effects. On the one hand, reliance on livestock may reflect a natural livelihood strategy in Baidoa, a major livestock trade hub in Somalia. On the other hand, it is unlikely that one can goat one's way out of poverty in the long run, to paraphrase Lant Pritchett (The New York Times, 2007), as sustained poverty reduction typically requires movement out of agriculture.

In addition, we demonstrate using a generalized random forest analysis that the positive effects of the treatment are notably larger for smaller households characterized by a smaller number of dependents; household size is the primary baseline variable predicting variation in the conditional average treatment effect. This is in contrast to the findings from an exploratory re-analysis of the data from Banerjee et al. (2015), a graduation program implemented in multiple stable settings, where baseline values of economic outcomes are more predictive of varying treatment effects. These findings suggest that livelihoods interventions implemented in displacement contexts should more carefully consider how household composition and dependent responsibilities may shape households' economic choices, and ultimately influence households' capacity to benefit from such programs.

This empirical pattern has implications for both targeting — some households might show significantly more positive treatment effects — and intervention design, in that it might be desirable to tailor any available intervention to households that have more care responsibilities. Given that large households do not show very positive effects from an integrated livelihoods intervention in this context, these households might alternatively be preferentially targeted for ongoing cash transfers to sustain consumption and human capital investment. At least in a fragile or displacement-affected context, a livelihoods intervention that has the explicit objective of generating a longer-term income stream might consider targeting households that are smaller and have reduced care responsibilities.

Table 1: Baseline balance

Variable	(1) Control Mean/(SE)	(2) Treatment Mean/(SE)	(1)-(2) Pairwise t-test Mean difference
IDP household	0.833	0.838	-0.005
	(0.011)	(0.007)	
Household size	6.901	6.573	0.328***
	(0.068)	(0.043)	
Dependency ratio	2.398	2.373	0.025
	(0.096)	(0.060)	
Any pregnant / lactating woman	0.589	0.601	-0.012
	(0.014)	(0.009)	
Any disabled member	0.204	0.192	0.013
	(0.011)	(0.007)	
Any adult male	0.871	0.859	0.011
	(0.010)	(0.006)	
Any cash savings	0.007	0.009	-0.002
	(0.002)	(0.002)	
Baseline asset value	358.784	337.792	20.992
	(19.428)	(11.100)	
Tropical livestock units	0.198	0.182	0.016
	(0.015)	(0.009)	
Household Hunger Scale	3.477	3.469	0.008
	(0.033)	(0.022)	
Livelihoods Coping Strategies index	1.657	1.587	0.069
	(0.061)	(0.038)	
Number of observations	1244	2872	4116
F-test of joint significance			0.053
F-test using randomization inference			0.079

Notes: All pair-wise regressions and F-tests are based on specifications including strata fixed effects and robust standard errors. Randomization inference p-values are based on 1,000 repetitions. Asterisks indicate significance at the ten, five, and one percent level.

Table 2: Primary experimental effects

	(1)	(2)	(3)	(4)	(5)

Panel A: Consumption and food security

	Total consumption	Food consumption	Non-food consumption	Moderate or severe HHS	Livelihood coping score
UPG	0.81***	0.59***	0.22***	-0.30***	-0.57***
	(0.06)	(0.04)	(0.02)	(0.02)	(0.09)
q-value	0.001***	0.001***	0.001***	0.001***	0.001***
Control mean	2.63	2.03	0.61	0.42	1.71
Observations	3964	3964	3964	3982	3982

Panel B: Assets and financial inclusion

	Asset value	TLUs	Any savings	Any credit
UPG	603.14***	0.35***	0.42***	0.08***
	(21.25)	(0.02)	(0.01)	(0.02)
q-value	0.001***	0.001***	0.001***	0.001***
Control mean	263.32	0.14	0.04	0.57
Observations	3982	3982	3966	3974

Panel C: Income

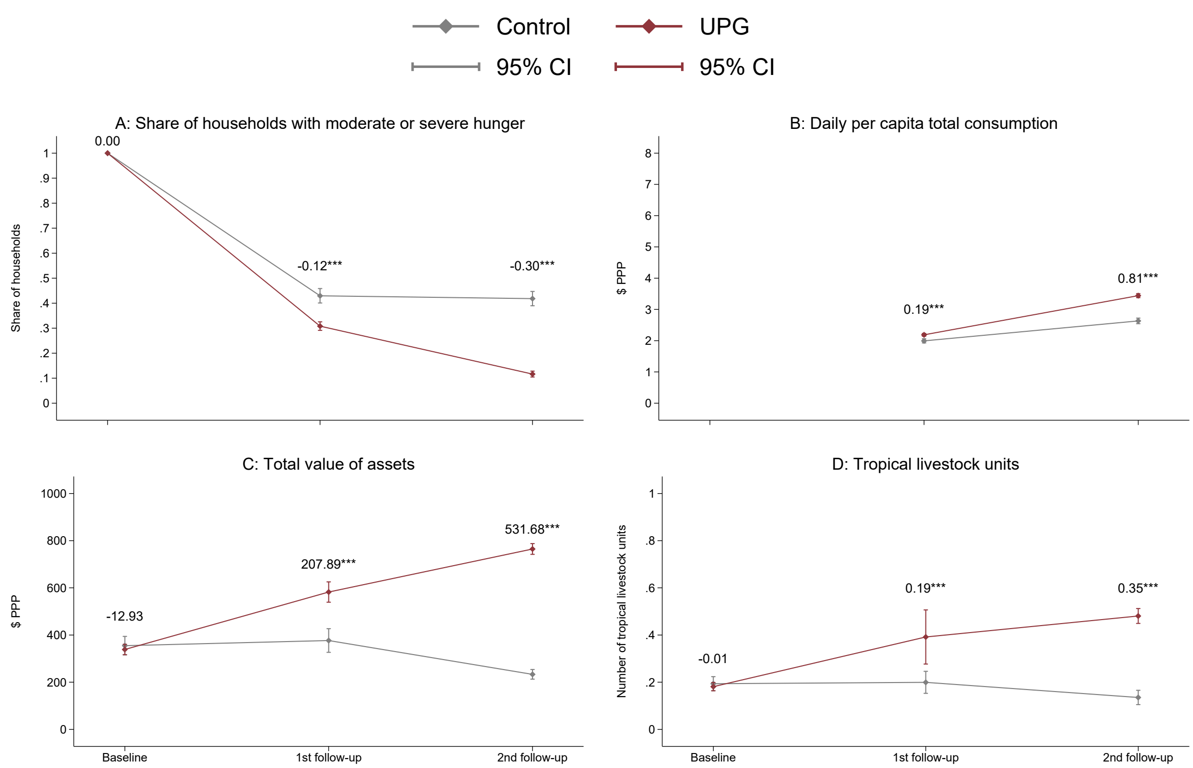
	Any ag. + livestock	Ag. + livestock income	Any non-farm business	Non-farm income	Any wage	Wage income
UPG	0.16***	28.11***	0.05***	3.81	-0.00	21.62
	(0.01)	(8.47)	(0.01)	(25.63)	(0.02)	(36.51)
q-value	0.001***	0.001***	0.001***	0.317	0.317	0.206
Control mean	0.05	6.08	0.15	125.84	0.42	586.59
Observations	3982	3982	3982	3982	3982	3982

Panel D: Social cohesion and locus of control

	Social cohesion	Locus of control
UPG	-0.04	0.10***
	(0.03)	(0.03)
q-value	0.085*	0.001***
Control mean	0.03	-0.07
Observations	3982	3982

Notes: The primary outcomes are per capita total consumption (col 1); moderate or severe household hunger scale (HHS) (col 4); and value of assets (col 1, Panel B). All regressions control for strata fixed effects and for the baseline outcome variable, if available (baseline values are available for the livelihood coping score, tropical livestock units, and any savings; there is no baseline variation in the household hunger score). Robust standard errors are reported in parentheses and the endline control mean is reported. Asterisks indicate significance at the ten, five, and one percent level. The reported q-values are p-values adjusted for multiple inference based on the false discovery rate correction procedure outlined in Anderson (2008).

Figure 1: Primary outcomes: Longitudinal effects



Notes: These graphs report treatment effects at both 1st follow-up (one year following program enrollment) and 2nd follow-up (two years). Solid dots show the control and treatment (UPG) group means, and capped bars represent the corresponding 95% confidence intervals. The reported numbers are treatment effects, estimated using an ANCOVA at first and second follow-up rounds that controls for strata dummies and baseline values of the variable, if available. Statistical significance is indicated as * 0.1, ** 0.05, and *** 0.01. Daily per capita consumption was not measured at baseline.

Table 3: Average standard treatment effects

	(1) Consumption and food security	(2) Assets and financial inclusion	(3) Income	(4) Social cohesion and locus of control
UPG	0.287***	1.436***	0.047	0.041
	(0.035)	(0.043)	(0.037)	(0.035)
Constant	0.000	0.000	0.001	0.000
	(0.030)	(0.030)	(0.029)	(0.029)
Observations	3964	3961	3982	3982

Notes: This table reports the average standard treatment effect for each family of outcomes following Kling et al. 2007, normalized with respect to standard deviations in the control arm. (The variable corresponding to any moderate or severe food insecurity is reverse-coded in this analysis.) Robust standard errors in parentheses. Asterisks indicate significance at the ten, five, and one percent level.

Table 4: Heterogeneous effects

	(1)	(2)	(3)	(4)
	Per capita consumption			
UPG	0.625***	0.680***	0.703***	0.493***
	(0.054)	(0.062)	(0.092)	(0.090)
UPG X Household under 5 members	0.369***			0.296**
	(0.130)			(0.133)
Household under 5 members	1.229***			1.178***
	(0.110)			(0.113)
UPG X Dependency ratio under 2.5		0.367***		0.240**
		(0.127)		(0.120)
Dependency ratio under 2.5		0.491***		0.168*
		(0.104)		(0.099)
UPG X Low baseline assets			0.171	0.126
			(0.116)	(0.106)
Low baseline assets			-0.181*	-0.143
			(0.101)	(0.093)
Observations	3964	3964	3964	3964

Notes: This table reports heterogeneous effects for the primary outcome of per capita daily consumption with respect to baseline covariates identified using the generalized random forest method. We define three binary variables using cutoffs derived from the 25th percentile of the baseline distribution: household size under five members, a dependency ratio (defined as the ratio of individuals under 15 or over 55 to those aged 16–54) under one, and baseline asset value below the 25th percentile or $74. Asterisks indicate significant at the ten, five and one percent level.

References

Aker, Jenny C., "Comparing Cash and Voucher Transfers in a Humanitarian Context: Evidence from the Democratic Republic of Congo," *The World Bank Economic Review*, 2017, *31* (1), 44–70.

Altındağ, O. and S. D. O'Connell, "The short-lived effects of unconditional cash transfers to refugees," *Journal of Development Economics*, 2023, *160*, 102942.

Anderson, Michael L, "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* (484), 1481–1495.

Asiedu, Edward, Dean Karlan, Monica Lambon-Quayefio, and Christopher Udry, "A call for structured ethics appendices in social science papers," *Proceedings of the National Academy of Sciences*, 2021, *118* (29), e2024570118.

Athey, Susan, Julie Tibshirani, and Stefan Wager, "Generalized random forests," *The Annals of Statistics*, 2019, *47* (2), 1148 – 1178.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil, "Why do people stay poor?," *The Quarterly Journal of Economics*, 2022, *137* (2), 785–844.

Ballard, Terri, Jennifer Coates, Anne Swindale, and Megan Deitchler, "Household hunger scale: indicator definition and measurement guide," *Washington, DC: Food and nutrition technical assistance II project, FHI*, 2011, *360*, 23.

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

Banerjee, Abhijit, Dean Karlan, Robert Osei, Hannah Trachtman, and Christopher Udry, "Unpacking a multi-faceted program to build sustainable income for the very poor," *Journal of Development Economics*, 2022, *155*, 102781.

_ , Esther Duflo, and Garima Sharma, "Long-term effects of the targeting the ultra poor program," *American Economic Review: Insights*, 2021, *3* (4), 471–486.

_ , _ , Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry, "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 2015, *348* (6236), 1260799.

Baseler, Travis, Thomas Ginn, Ibrahim Kasirye, Belinda Muya, and Andrew Zeitlin, "Mentoring Small Businesses: Evidence from Uganda," *mimeo*, 2024.

Battisti, Michele, Yvonne Giesing, and Nadzeya Laurentsyeva, "Can job search assistance improve the labour market integration of refugees? Evidence from a field experiment," *Labour Economics*, 2019, *61*, 101745.

Bedoya, G. et al., "No household left behind: Afghanistan targeting the ultra-poor impact evaluation," Technical Report w25981, National Bureau of Economic Research 2019.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on Treatment Effects after Selection among High-Dimensional Controls†," *The Review of Economic Studies*, 11 2013, *81* (2), 608–650.

Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand, Catherine C Thomas, Christopher Udry, Julia Vaillant et al., "Tackling psychosocial and capital constraints to alleviate poverty," *Nature*, 2022, *605* (7909), 291–297.

Brune, L. et al., "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.

Carletto, Gero, Marco Tiberti, and Alberto Zezza, "Measure for measure: Comparing survey based estimates of income and consumption for rural households," *The World Bank Research Observer*, 2022, *37* (1), 1–38.

Catholic Relief Services, "Social Cohesion Indicators Bank," `https://www.crs.org/our-work-overseas/research-publications/social-cohesion-indicators-bank` 2019. Accessed: 2024-03-07.

Cilliers, Jacobus, Nour Elashmawy, and David McKenzie, "Using post-double selection Lasso in field experiments," Technical Report, World Bank 2024.

Cleland, John and Kazuyo Machiyama, "The challenges posed by demographic change in sub-Saharan Africa: A concise overview," *Population and Development Review*, 2017, *43*, 264–286.

Davis, Jonathan MV and Sara B Heller, "Using causal forests to predict treatment heterogeneity: An application to summer jobs," *American Economic Review*, 2017, *107* (5), 546–550.

Deaton, Angus and Salman Zaidi, *Guidelines for constructing consumption aggregates for welfare analysis*, Vol. 135, World Bank Publications, 2002.

Fasani, Francesco, Tommaso Frattini, and Luigi Minale, "Lift the ban? Initial employment restrictions and refugee labour market outcomes," *Journal of the European Economic Association*, 2021, *19* (5), 2803–2854.

Gibson, John and Scott Rozelle, "Prices and unit values in poverty measurement and tax reform analysis," *The World Bank Economic Review*, 2005, *19* (1), 69–97.

Gupta, Prankur, Daniel Stein, Kyla Longman, Heather Lanthorn, Rico Bergmann, Emmanuel Nshakira-Rukundo, Noel Rutto, Christine Kahura, Winfred Kananu, Gabrielle Posner et al., "Cash transfers amid shocks: A large, one-time, unconditional cash transfer to refugees in Uganda has multidimensional benefits after 19 months," *World Development*, 2024, *173*, 106339.

Hidrobo, Melissa, John Hoddinott, Amber Peterman, Amy Margolies, and Vanessa Moreira, "Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador," *Journal of Development Economics*, 2014, *107*, 144–156.

Humble, Steve, Aditya Sharma, Baladevan Rangaraju, Pauline Dixon, and Mark Pennington, "Associations between neighbourhood social cohesion and subjective well-being in two different informal settlement types in Delhi, India: a quantitative cross-sectional study," *BMJ open*, 2023, *13* (4), e067680.

Hussam, R. et al., "The psychosocial value of employment: Evidence from a refugee camp," *American Economic Review*, 2022, *112* (11), 3694–3724.

IDMC, "2024 Global Report on Internal Displacement," 2024.

Kaplan, Lennart, Utz Pape, and James Walsh, "Eliciting Accurate Consumption Responses from Vulnerable Populations," *Data Collection in Fragile States: Innovations from Africa and Beyond*, 2020, pp. 193–206.

Kerwin, Jason, Nada Rostom, and Olivier Sterck, "Striking the Right Balance: Why Standard Balance Tests Over-Reject the Null, and How to Fix It," Technical Report, IZA Discussion Papers 2024.

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

Lee, David S, "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," 2005.

Leight, Jessica, Harold Alderman, Daniel Gilligan, Melissa Hidrobo, and Michael Mulford, *Can a light-touch graduation model enhance livelihood outcomes? Evidence from Ethiopia*, Intl Food Policy Res Inst, 2023.

MacPherson, C. and O. Sterck, "Empowering refugees through cash and agriculture: A regression discontinuity design," *Journal of Development Economics*, 2021, *149*, 102614.

Malacarne, Jonathan G, "The farmer and the fates: Locus of control and investment in rainfed agriculture," *Applied Economic Perspectives and Policy*, 2024, *46* (2), 534–552.

Mancini, Giulia and Giovanni Vecchi, "On the construction of a consumption aggregate for inequality and poverty analysis," *World Bank Group, Washington, DC*, 2022.

Özler, Berk, Çiğdem Çelik, Scott Cunningham, P Facundo Cuevas, and Luca Parisotto, "Children on the move: Progressive redistribution of humanitarian cash transfers among refugees," *Journal of Development Economics*, 2021, *153*, 102733.

Pape, Utz, "Somali Poverty Profile: Findings from Wave 1 of the Somali High Frequency Survey," Technical Report, World Bank Group, Washington, D.C. 2017.

_ **and Johan Mistiaen**, "Household expenditure and poverty measures in 60 minutes: a new approach with results from Mogadishu," *World Bank Policy Research Working Paper*, 2018, (8430).

_ **and** _ , "Rapid Consumption Surveys," in Johannes Hoogeveen and Utz Pape, eds., *Data Collection in Fragile States: Innovations from Africa and Beyond*, Palgrave Macmillan, 2020, chapter 9, pp. 153–171.

_ **and Paolo Verme**, "Measuring Poverty in Forced Displacement Contexts," *GLO Discussion Paper*, 2023.

_ **and Philip Randolph Wollburg**, "Estimation of poverty in Somalia using innovative methodologies," *World Bank Policy Research Working Paper*, 2019, (8735).

Rotter, Julian B, "Generalized expectancies for internal versus external control of reinforcement.," *Psychological monographs: General and applied*, 1966, *80* (1), 1.

Rozo, Sandra V and Guy Grossman, "Refugees and Other Forcibly Displaced Populations," *VoxDevLit*, 2025, *14* (1).

Rustad, Siri Aas, *Conflict Trends: A Global Overview, 1946–2023*, Peace Research Institute Oslo (PRIO), 2024.

Sarvimäki, Matti and Kari Hämäläinen, "Integrating immigrants: The impact of restructuring active labor market programs," *Journal of Labor Economics*, 2016, *34* (2), 479–508.

Schuettler, Kirsten and Laura Caron, *Jobs interventions for refugees and internally displaced persons*, World Bank Group, 2020.

Sylvia, Sean, Nele Warrinnier, Renfu Luo, Ai Yue, Orazio Attanasio, Alexis Medina, and Scott Rozelle, "From quantity to quality: Delivering a home-based parenting intervention through China's family planning cadres," *The Economic Journal*, 2021, *131* (635), 1365–1400.

The New York Times, "Should We Globalize Labor Too?," *The New York Times Magazine*, June 10, 2007 2007.

UN-Habitat, "Baidoa Urban Profile: Working Paper and Spatial Analyses for Urban Planning and Durable Solutions," `https://reliefweb.int/report/somalia/baidoa-urban-profile-working-paper-and-spatial-analyses-urban-planning-and-durable` 2021. Accessed: 2024-03-07.

UNHCR, "Location and populations of IDP sites in Baidoa as at 28 April 2017," `https://data.unhcr.org/en/documents/details/56361` 2017. Accessed: 2024-03-07.

_ , "Location and populations of IDP sites in Baidoa as at July 2022," `https://data.unhcr.org/en/documents/details/94414` 2022. Accessed: 2024-03-07.

_ , "Global Trends: Forced Displacement," `https://www.unhcr.org/global-trends` 2023. Accessed: 2024-03-07.

_ , "Mid-Year Trends," `https://www.unhcr.org/mid-year-trends` 2023. Accessed: 2024-03-07.

_ , "Internally Displaced People," 2024. Accessed: 2024-03-07.

_ , "Refugee Statistics," `https://www.unhcr.org/refugee-statistics` 2024. Accessed: 2024-03-07.

United Nations, "New York Declaration for Refugees and Migrants (A/RES/71/1)," 2016. Resolution adopted by the General Assembly on 19 September 2016.

Wager, Stefan and Susan Athey, "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

White, Halbert, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica: Journal of the Econometric Society*, 1980, pp. 817–838.

WMO, "Extreme Weather," 2024. Accessed: 2024-03-07.

Appendix

A1 Graduation literature

Table A1 provides an overview of key studies evaluating graduation-style interventions, summarizing characteristics of the trial design, intervention components, and estimated impacts on consumption. The final two rows of the table report the estimated per-household or per-participant intervention costs in both 2017 Purchasing Power Parity (PPP) dollars and 2017 nominal U.S. dollars, allowing for cross-country and cross-study comparison. These cost estimates have been harmonized using a standardized approach described in the following paragraphs.

The original intervention costs are reported in U.S. dollar terms in Banerjee et al. (2015) (Ethiopia, Ghana, Honduras, India, Pakistan, and Peru), Bedoya et al. (2019) (Afghanistan), and in this study (Somalia). In this case, the procedure follows three steps: first, we recover the local currency value by applying the USD exchange rate for the base year of the original cost data. Second, we adjust the local currency values to 2017 levels using changes in domestic consumer price indices (CPI). Finally, we convert the 2017 local currency values into 2017 PPP dollars using PPP conversion factors for individual consumption. All CPI, exchange rate, and PPP data are sourced from the World Bank, except in the case of Somalia, where CPI data were drawn from the IMF due to gaps in the World Bank series.

Bandiera et al. (2017) (Bangladesh), Bossuroy et al. (2022) (Niger), and Brune et al. (2022) (Yemen) report costs in PPP dollar terms. Here, to convert these costs to 2017 PPP and nominal U.S. dollars, we follow a two-step process: first, we recover the local currency value by multiplying the original PPP figure by the relevant PPP exchange rate in the base year. Then, we follow the inflation and conversion steps described above starting from step two.

Yemen is a special case. The intervention cost was reported in 2010 PPP dollars, but no reliable CPI data exist for Yemen beyond 2014. To address this, we estimate an inflation adjustment factor based on the change in the nominal exchange rate between 2010 and 2017, scaled by U.S. inflation over the same period. This approach assumes that relative movements in the exchange rate reflect changes in local price levels in the absence of reliable domestic inflation data.

Table A1: Overview of graduation literature

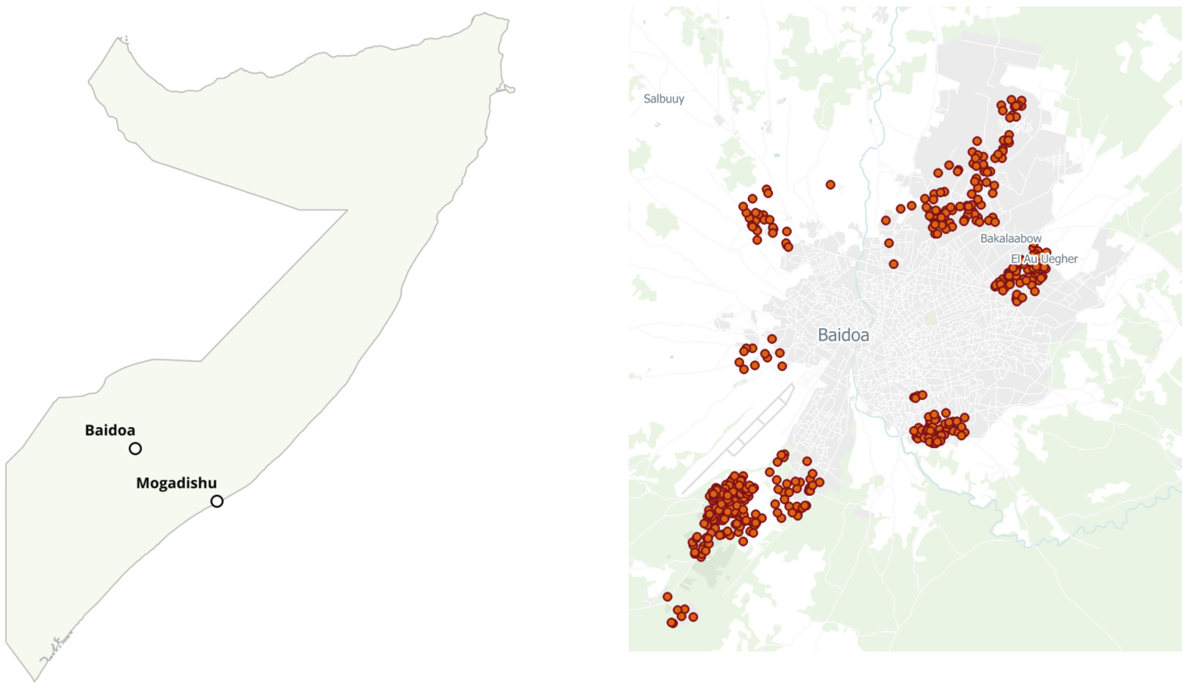
Study	(a)	(b)	(b)	(b)	(b)	(b)	(b)	(c)	(d)	(e)	(f)	(g)
Trial characteristics												
[12pt] Country	BGD	ETH	GHA	HND	IND	PAK	PER	AFG	NER	YEM	ETH	SOM
Year launched	2007	2010	2011	2009	2007	2008	2011	2016	2016	2010	2019	2022
Years elapsed	4	3	3	3	3	3	3	2	1.5	4	3	2
Context	Rural	Rural	Rural	Rural	Rural	Rural	Rural	Rural	Rural	Both	Rural	Urban
Forced displacement	No	No	No	No	No	No	No	No	No	No	No	IDP
Ongoing conflict	No	No	No	No	No	No	No	Yes	No	Yes	No	Yes
Consumption ITT	10.9%	18.2%	10.6%	-5.8%	10.7%	7.0%	5.2%	29.7%	14.7%	-2.4%	0.7%	30.8%
p-value	0.003	0.000	0.007	0.152	0.001	0.079	0.085	0.000	0.000	0.644	$> .2$	0.000
Intervention characteristics												
Asset (A) or grant (G)	A	A	A	A	A	A	A	A	G	A	A / G	A
Consumption support	Cash	Food *	Cash	Food	Cash	Cash	Cash *	Cash	Cash *	Cash *	Cash*	Cash
Mentoring	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Skills training	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	No
Savings groups	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Access to finance	No	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	No	No
Psychosocial support	No	No	No	No	No	No	No	No	Yes	No	Yes†	No
Social/behavioral support	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cost ($PPP 2017)	1,489	3,251	5,600	3,046	1,234	3,959	4,422	7,474	595	1,101	NR	7,768
Cost ($USD 2017)	541	1,118	2,201	1,385	382	1,231	2,591	1,768	250	332	NR	2,927

Notes: [a] Bandiera et al. (2017), [b] Banerjee et al. (2015), [c] Bedoya et al. (2019), [d] Bossuroy et al. (2022), [e] Brune et al. (2022), [f] Leight et al. (2023), [g] This study. BGD = Bangladesh, ETH = Ethiopia, GHA = Ghana, HND = Honduras, IND = India, PAK = Pakistan, PER = Peru, AFG = Afghanistan, NER = Niger, YEM = Yemen, SOM = Somalia. Years elapsed refer to the years elapsed between program launch and the collection fo follow-up data. Costs are converted to purchasing power parity terms in the base year employed and then converted to PPP terms in 2017. For Leight et al. in Ethiopia, the reported coefficient is the mean effect on consumption across three arms examined; none were statistically significant, hence the notation of the p-value. * Consumption support was provided to all individuals in the sample (i.e., there was no experimental variation.) † Psychosocial support was provided only to those identified as eligible based on symptoms of depression

A2 Appendix Exhibits

Figure A1: Map of Somalia and the IDP survey sites in Baidoa

(a) Somalia (b) Baidoa



Notes: Figure (a) shows the location of Baidoa in Somalia and with respect to the capital, Mogadishu. Shapefile from UNOCHA Somalia. Figure (b) shows the locations of IPD households at the second follow-up survey. Basemap from OpenStreetMap.

Figure A2: Study timeline

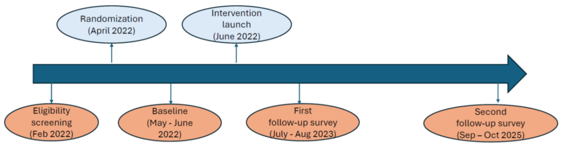
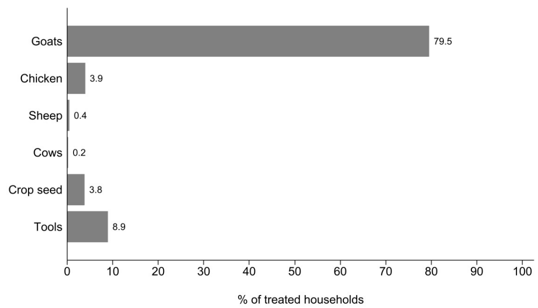
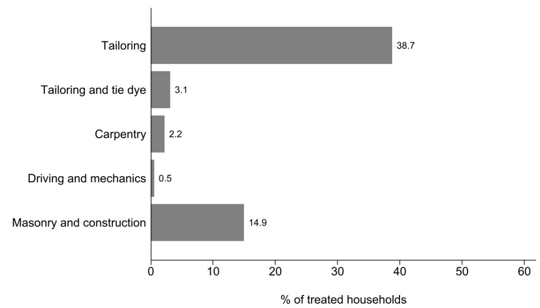


Figure A3: Asset and TVET Choices

(a) Asset transfers



(b) TVET training

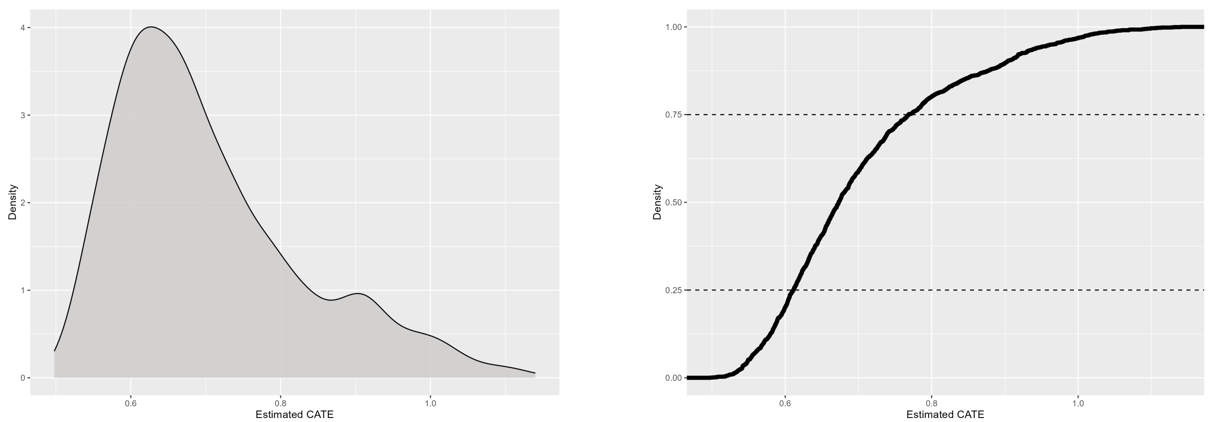


Notes: The figures capture the reported choices of productive assets and Technical and Vocational Education and Training (TVET) in the treatment arm.

Figure A4: Out-of-bag CATE estimates on consumption from GRF algorithm

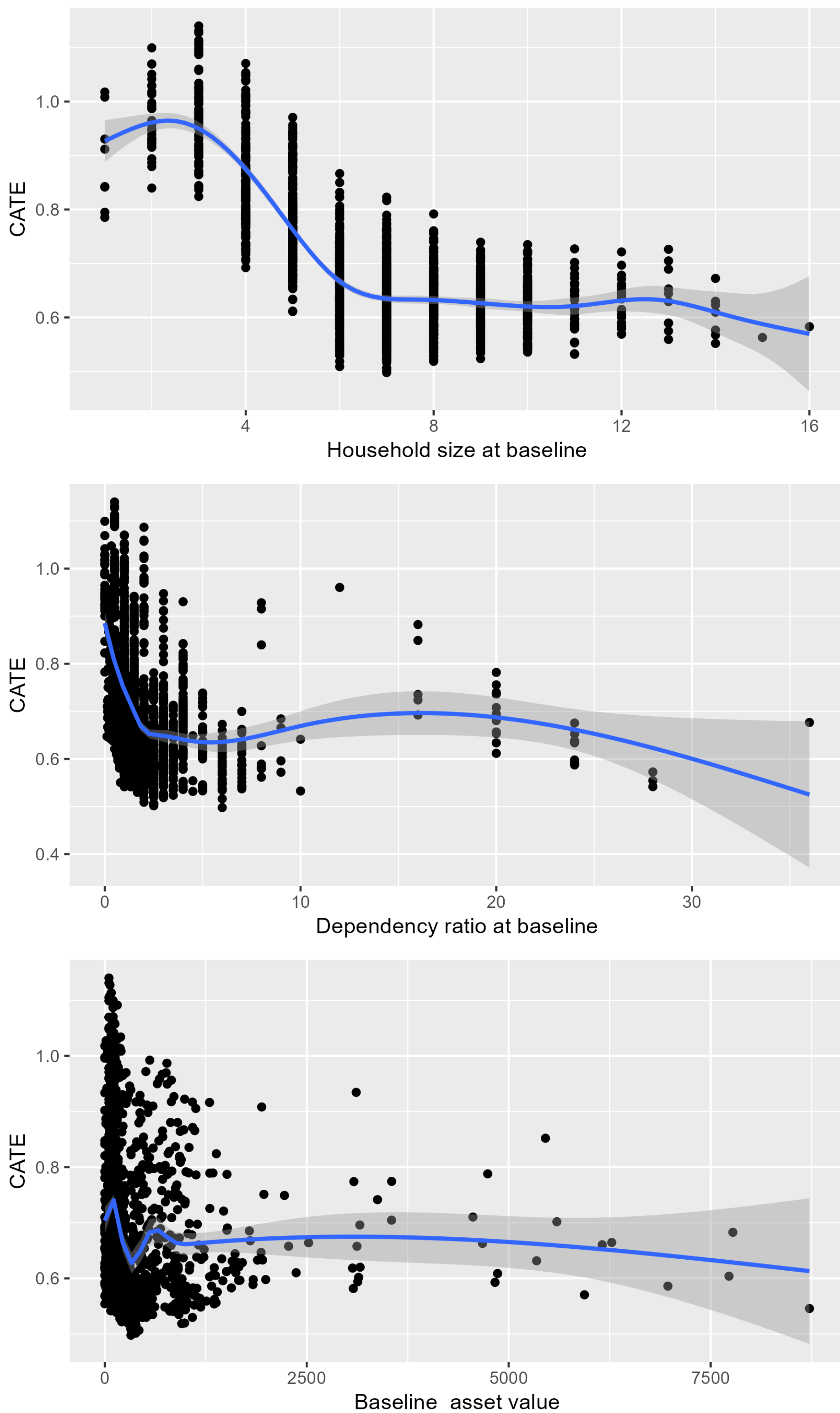
(a) Kernel density function

(b) Cumulative distribution function



Notes: These figures report the out-of-bag conditional average treatment effects (CATE) from the generalized random forest (GRF) method following Athey et al. (2019). Figure A4a reports the kernel; density function, and Figure A4b reports the cumulative density function.

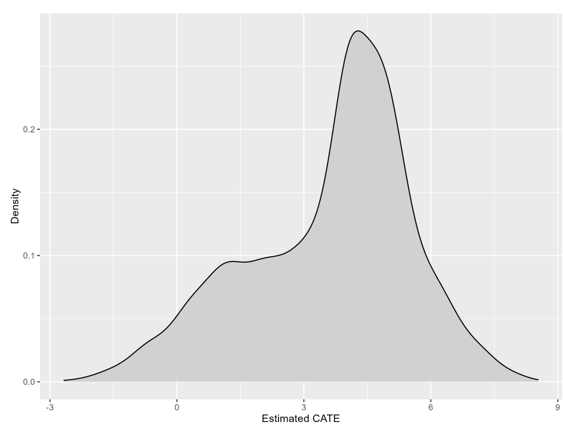
Figure A5: Scatter plots of out-of-bag CATE estimates and observable baseline characteristics



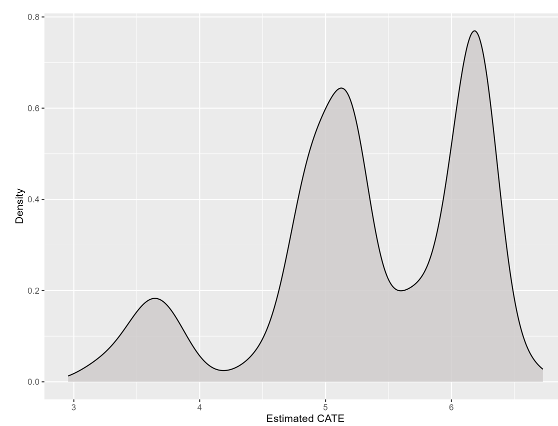
Notes: This graph reports the correlation between the conditional average treatment effect (CATE) and the three baseline covariates identified as most predictive of variation in the conditional average treatment effect, as reported in Table A8.

Figure A6: Out-of-bag CATE estimates on consumption: Banerjee et al. (2015)

(a) Model including full set of baseline covariates



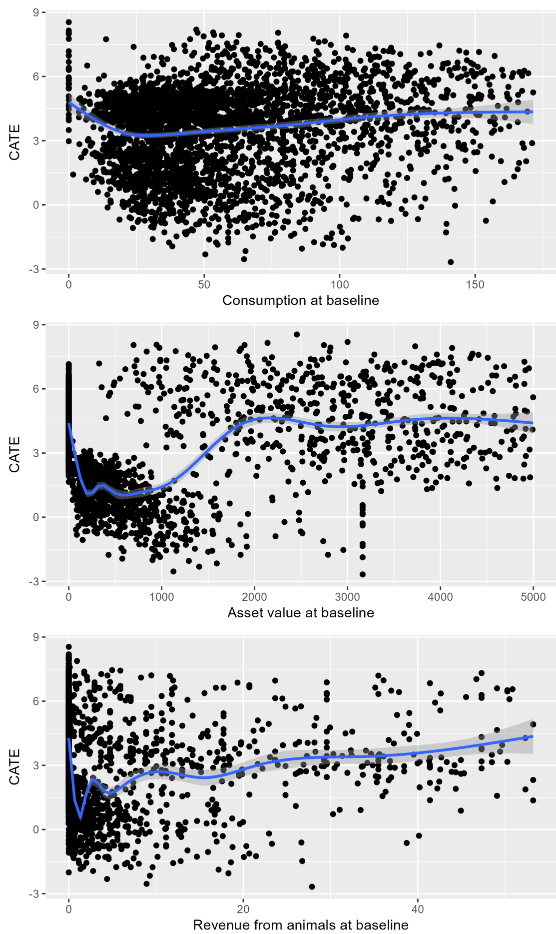
(b) Model including restricted set of baseline covariates



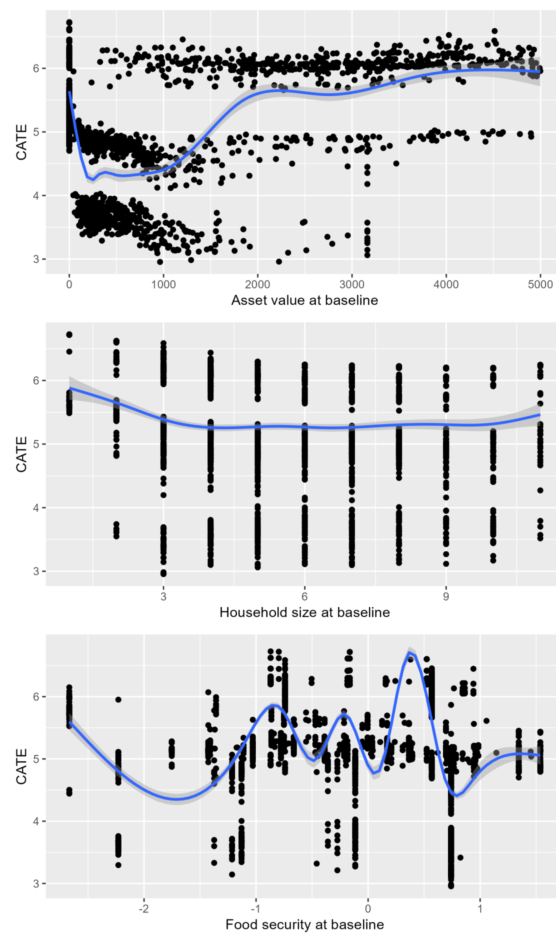
Notes: These figures report the out-of-bag conditional average treatment effects (CATE) from the generalized random forest method following Athey et al. (2019), estimated using the replication data from the Banerjee et al. (2015) trial for the outcome variable of consumption in the long-term follow-up.

Figure A7: Scatter plots of out-of-bag CATE estimates and observable baseline characteristics: Banerjee et al. (2015)

(a) Model including full set of baseline covariates

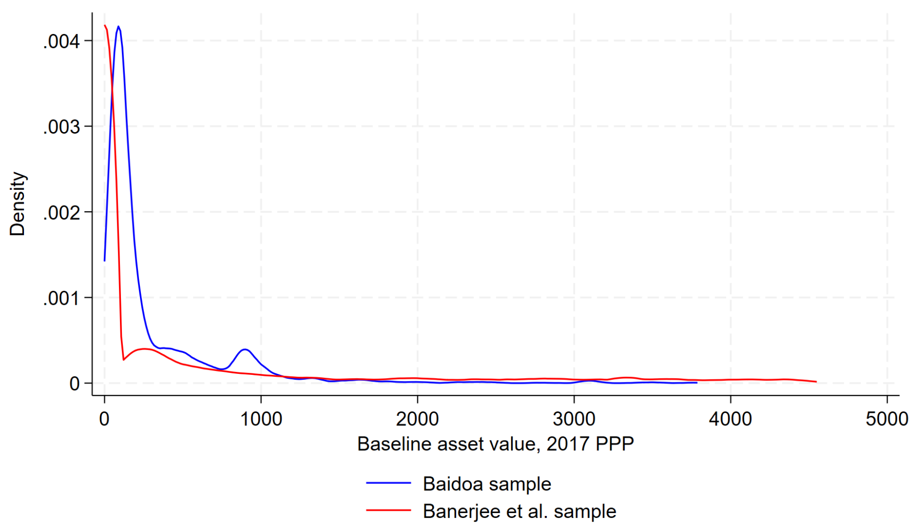


(b) Model including restricted set of baseline covariates



Notes: This graph reports the correlation between the conditional average treatment effect (CATE) and the three baseline covariates identified as most predictive of variation in the conditional average treatment effect, as reported in Table A8.

Figure A8: Baseline assets across samples



Notes: This graph presents the kernel density of baseline asset value in both the Baidoa sample analyzed in this paper and the sample in Banerjee et al. (2015); the density figure is truncated at the 95th percentile of asset value in the combined sample. All value estimates are in 2017 purchasing power parity-adjusted dollars.

Table A2: Program exposure

	(1)	(2)	(3)	(4)
	Cash transfers	Asset transfer or TVET training	Savings groups	Coaching
UPG	0.738***	0.844***	0.569***	0.310***
	(0.012)	(0.008)	(0.011)	(0.015)
Constant	0.143***	0.024***	0.056***	0.182***
	(0.010)	(0.005)	(0.007)	(0.011)
Observations	3982	3982	3982	3982

Notes: This table reports treatment effects on reported intervention receipt for the four main intervention elements: cash transfers, asset transfers or Technical and Vocational Education and Training (TVET), savings groups, and coaching. All regressions include strata fixed effects; asterisks indicate significance at the ten, five, and one percent level.

Table A3: Primary experimental effects: Additional baseline controls

	(1)	(2)	(3)	(4)	(5)	
Panel A: Consumption and food security						
	Total consumption	Food consumption	Non-food consumption	Moderate or severe HHS	Livelihood coping score	
UPG	0.71***	0.51***	0.19***	-0.30***	-0.57***	
	(0.05)	(0.04)	(0.02)	(0.02)	(0.09)	
q-value	0.001***	0.001***	0.001***	0.001***	0.001***	
Control mean	2.63	2.03	0.61	0.42	1.71	
Observations	3964	3964	3964	3982	3982	
Panel B: Assets and financial inclusion						
	Asset value	TLUs	Any savings	Any credit		
UPG	603.97***	0.35***	0.42***	-0.00		
	(21.34)	(0.02)	(0.01)	(0.01)		
q-value	0.001***	0.001***	0.001***	0.372		
Control mean	263.32	0.14	0.04	0.18		
Observations	3982	3982	3966	3982		
Panel C: Income						
	Any livestock	Livestock income	Any non-farm business	Non-farm income	Any wage	Wage income
UPG	0.15***	34.11***	0.05***	7.99	-0.00	22.37
	(0.01)	(7.84)	(0.01)	(25.59)	(0.02)	(36.57)
q-value	0.001***	0.001***	0.001***	0.372	0.396	0.291
Control mean	0.02	-1.05	0.15	125.84	0.42	586.59
Observations	3982	3982	3982	3982	3982	3982
Panel D: Social cohesion and locus of control						
	Social cohesion	Locus of control				
UPG	-0.04	0.10***				
	(0.03)	(0.03)				
q-value	0.111	0.001***				
Control mean	0.03	-0.07				
Observations	3982	3982				

Notes: The primary outcomes are per capita total consumption (col 1); moderate or severe household hunger scale (HHS) (col 4); and value of assets (col 1, Panel B). All regressions control for strata fixed effects and for the baseline outcome variable, if available (baseline values are available for the livelihood coping score, tropical livestock units, and any savings; there is no baseline variation in the household hunger score); a control variable for household size, selected by a double lasso, is also included. Robust standard errors are reported in parentheses and the endline control mean is reported. Statistical significance is indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The reported q-values are p-values adjusted for multiple inference based on the false discovery rate correction procedure outlined in Anderson (2008).

Table A4: Minimum detectable effects

	Share of households moderate or severe HHS	Per capita consumption in USD-PPP	Asset value in USD-PPP
N, Control	1,150	1,146	1,150
N, Treatment	2,832	2,818	2,832
Mean, Control	0.42	2.63	263
SD, Control	0.49	1.54	482
Adjusted SD, Control	0.49	1.53	481
MDES	0.05	0.15	47.15
MDES relative to mean (%)	11.50	5.70	17.90
MDES relative to SD (%)	9.80	9.80	9.80

Notes: HHS = Household hunger score, USD = United States dollar, PPP = Purchasing power parity, N = Number of observations, SD = Standard deviation, MDES = Minimum Detectable effect size.

Table A5: Additional outcomes

Variable	(1) Control mean: binary owned	(2) Control mean	(3) Treatment effect	(4) Std. error	(5) N
Panel A: Assets					
Mobile phones	0.908	1.013	0.068***	0.016	3954
Axe	0.742	0.802	-0.014	0.019	3872
Hammer	0.233	0.260	0.021	0.018	3752
Grain bag	0.182	0.596	0.329***	0.064	3745
Panga	0.167	0.188	0.026	0.016	3761
Pick axe	0.159	0.181	0.030*	0.016	3776
Spade or shovel	0.138	0.143	0.098***	0.014	3718
Plough (oxen-pulled)	0.132	0.183	0.074***	0.020	3679
Hoe	0.130	0.160	0.048***	0.017	3711
Hand mattock	0.125	0.157	0.075***	0.019	3793
Goats	0.091	0.264	2.879***	0.076	3912
Sickle	0.089	0.116	0.026*	0.016	3692
Donkey	0.076	0.076	0.020**	0.010	3763
Poultry	0.070	0.185	0.102***	0.034	3899
Hand saw	0.068	0.100	0.011	0.015	3729
Tarpaulin	0.065	0.083	0.064***	0.014	3702
Wheelbarrow	0.062	0.066	0.063***	0.011	3751
Donkey cart	0.060	0.064	0.034***	0.010	3758
Solar panels	0.060	0.060	0.042***	0.009	3896
Sheep	0.057	0.093	0.073***	0.018	3893
Panel B: Household composition					
Household size	.	7.152	0.068**	0.032	3982
Ages 0–4	.	0.925	0.049*	0.026	3982
Ages 5–14	.	2.768	0.052	0.034	3982
Ages 15–19	.	0.964	-0.047*	0.024	3982
Ages 20–54	.	2.057	0.008	0.021	3982
Ages 55+	.	0.438	-0.011	0.008	3982

Notes: This table presents supplementary regression findings for ownership by asset category, in Panel A, and household composition, in Panel B. Column (1) reports the mean of a binary variable for any asset in that category owned, in the control arm. Asterisks indicate significance at the ten, five, and one percent level.

Table A6: Attrition

	(1)	(2)	(3)
[1em] Treated household	-0.062***	-0.062***	-0.061
	(0.008)	(0.008)	(0.047)
IDP household		-0.009	-0.030
		(0.008)	(0.022)
IDP X Treat			0.030
			(0.023)
Household size		0.000	0.004
		(0.001)	(0.004)
Size X Treat			-0.005
			(0.004)
Dependency ratio at baseline		-0.002***	-0.005***
		(0.001)	(0.001)
Dependency X Treat			0.005***
			(0.001)
Any pregnant or lactating woman		0.004	0.013
		(0.006)	(0.015)
PLW X Treat			-0.011
			(0.016)
Any disabled member		0.002	-0.006
		(0.007)	(0.018)
Disability X Treat			0.011
			(0.020)
Any adult male		-0.005	-0.021
		(0.009)	(0.025)
Male X Treat			0.022
			(0.026)
Any cash savings		-0.000	0.000
		(0.000)	(0.001)

Any savings X Treat			-0.000
			(0.001)
Baseline asset value		-0.001	-0.000
		(0.001)	(0.003)
Asset value X Treat			-0.001
			(0.003)
Tropical livestock units		0.019	0.017
		(0.017)	(0.037)
TLUs X Treat			0.001
			(0.041)
Household Hunger Scale		0.006**	0.010
		(0.003)	(0.007)
HHS X Treat			-0.006
			(0.007)
Livelihoods Coping Strategies index		-0.002**	-0.006**
		(0.001)	(0.003)
LCS X Treat			0.005
			(0.003)
[1em] Observations	4116	4116	4116

Notes: Each column regresses a binary variable equal to one for households that attrite at the 2nd follow-up on a binary variable for treated; baseline characteristics; and the interaction between the two. Baseline asset value is expressed in hundreds of dollars. All regressions includes strata fixed effects; asterisks indicate significance at the ten, five, and one percent level.

Table A7: Treatment effects bounds correcting for attrition

Panel A: Lee and Kling-Liebman bounds

	Lee lower		Lee upper		Kling-Liebman lower		Kling-Liebman upper	
Total cons.	.523	(.053)	.97	(.056)	.526	(.058)	1.09	(.058)
Food cons.	.366	(.042)	.719	(.044)	.37	(.046)	.814	(.046)
Non-food cons.	.119	(.018)	.263	(.019)	.12	(.02)	.311	(.02)
Asset value	603.137	(21.252)	651.061	(21.558)	507.03	(21.938)	699.675	(21.938)
LCS score	-.948	(.079)	-.57	(.085)	-1.019	(.089)	-.123	(.089)
TLUs	.204	(.017)	.346	(.022)	.247	(.023)	.447	(.023)

Panel B: Manski bounds

	Lower		Upper	
Moderate ./ severe HHS	-.377	(.015)	-.228	(.015)
Any savings	.32	(.011)	.456	(.013)
Any credit	.008	(.017)	.163	(.017)
Ag. / livestock inc.	.058	(.01)	.19	(.012)
Non-farm bus. inc.	-.043	(.012)	.098	(.014)
Wage inc.	-.085	(.017)	.07	(.017)

Notes: This table reports bounds correcting for attrition for the outcomes of interest. For the continuous variables in Panel A, we report Lee bounds and Kling-Liebman bounds, allowing for attrited individuals to be characterized by outcomes two standard deviations above (below) the mean observed in the relevant treatment arm. For the binary variables in Panel B, we report Manski bounds. Standard errors are in parentheses.

Table A8: Baseline characteristics used in GRF analysis and variable importance

Baseline characteristics	**Variable importance**
Household size	0.31
Dependency ratio	0.18
Baseline asset value	0.17
Tropical Livestock Units	0.06
Livelihoods Coping Score	0.05
Strata 3	0.04
Strata 2	0.03
Any pregnant or lactating woman	0.03
Household Hunger Score	0.02
Strata 1	0.02
Engaged in farming	0.02
Any disabled member	0.02
Any adult male	0.01
Receiving transfers	0.01
IDP household	0.01
Self-employed	0.01
Other income	0.01
Any primary education	0.01
Wage labor	0.01
Any savings	0
Asset income	0

Notes: This table reports the variables employed in the generalized random forest (GRF) algorithm and their estimated importance, defined as the frequency with which each observable characteristic is used as a splitting variable. The weights are re-scaled such that the total weights for baseline characteristics excluding binary variables for randomization strata add up to one.

Table A9: Baseline characteristics used in GRF analysis and variable importance: Banerjee et al. 2015

Characteristic	**Importance (Full)**	**Importance (Restricted)**
Consumption per capita	0.1	NA
Total asset value	0.1	0.24
Income from agriculture	0.09	NA
Revenue from animals	0.09	NA
Income from paid labor	0.08	NA
Household size	0.07	0.2
Food security	0.06	0.2
Income from business	0.05	NA
Perception of economic welfare	0.05	0.2
Any savings	0.04	0.16

Notes: This table reports the variables employed in the generalized random forest (GRF) algorithm and their estimated importance, defined as the frequency with which each observable characteristic is used as a splitting variable, in the analysis conducted using the replication data from Banerjee et al (2015). We report the analysis conducted for a full set of available baseline characteristics, and a restricted set of available baseline characteristics, corresponding to those also available in the Baidoa sample; only the top ten most predictive variables in the "full" analysis are reported in the table, and thus for this model, the reported weights do not sum to one. "NA" denotes a variable not included in the restricted model. The weights are re-scaled such that the total weights for baseline characteristics excluding binary variables for randomization strata and country add up to one.

A3 Structured Ethics Appendix

The structured ethics appendix is based on Asiedu et al. (2021).

Policy Equipoise

Is there policy equipoise? That is, is there uncertainty regarding participants' net benefits from each arm of the study relative to the other arms and to the best possible policy to which participants could have access? If not, ethical randomization requires two conditions related to scarcity: (1) Was there scarcity, i.e., did the inclusion of multiple arms change the expected aggregate value of the programs delivered? (2) Do all ex-ante identifiable participants have equal moral or legal claims to the scarce programs?

While evidence on the effectiveness of graduation programs in forced displacement settings remains limited, prior academic literature from other contexts suggests that receiving the graduation package —which includes cash grants and either a TVET opportunity or an asset transfer — is likely to yield greater benefits than not receiving it. Given this, and despite the scarcity of evidence in displacement settings, the study may not fully meet the condition of policy equipoise. However, the randomization is ethically justified under conditions of scarcity. First, the available resources were insufficient to provide the package to all eligible households in this context. Second, eligibility was determined using pre-specified household characteristics assessed ex ante: households were classified as experiencing moderate or severe hunger based on the Household Hunger Scale and had resided in the IDP site for at least one month. These criteria were applied consistently, and no ex-ante identifiable group with a stronger moral or legal claim to the program was excluded. Third, randomization did not shift the aggregate benefit of the intervention (provided to 5,000 households). On this basis, the study satisfies commonly accepted ethical conditions for randomization in settings where policy equipoise may be uncertain or absent.

Role of researchers with respect to implementation

Are researchers "active" researchers, i.e. did the researchers have direct decision making power over whether and how to implement the program? If YES, what was the disclosure to participants and informed consent process for participation in the program? Providing IRB approval details may be sufficient but further clarification of any important issues should be discussed here. If NO, i.e., implementation was separate, explain the separation.

The researchers did not play an "active" role in the implementation of this project. Implementation was carried out independently by World Vision, and the researchers were not involved in delivering the intervention or interacting directly with participants. The program was funded by USAID, and the researchers had no control over funding allocations or implementation decisions other than conducting the randomization within the sample identified as eligible by the implementing partner. Final authority over all aspects of implementation rested with the implementing partner, not the research team.

Potential harms to participants or nonparticipants from the interventions or policies

Does the intervention, policy or product being studied pose potential harm to participants or non-participants? Related, are participants or likely affected non-participants particularly vulnerable? Also related, are participants' access to future services or policies changed because of participation in the study? If yes to any of the above, what is being done to mitigate such risks?

The graduation intervention was not expected to pose material harm to participants or non-participants. The program aimed to improve household welfare through a cash grant paired with either vocational training or an asset transfer, and participants retained autonomy over how they engaged with the support. However, as with many livelihood-focused programs, there were potential opportunity costs — for example, time spent attending training or managing a new income-generating activity — which may have redirected time away from other responsibilities. These risks were considered minor relative to the expected benefits.

One potential concern was that the intervention was not cluster-randomized, raising the possibility that untreated households might perceive the allocation as unfair or experience social tension as a result. To monitor this, the study included questions on social cohesion and trust at the second follow-up round. Results showed no significant differences in reported cohesion between treatment and control groups, suggesting that the intervention did not weaken community relationships.

Participants were internally displaced persons (IDPs), a population with heightened vulnerability due to displacement, food insecurity, and limited access to services. The program did not reduce or restrict access to other forms of aid, and participation did not affect households' eligibility for future support from World Vision or other humanitarian actors.

Implementation was carried out independently by World Vision, and the researchers were not "active" in implementation. While this limits their direct responsibility for any harms, the research team designed the study to identify and assess any unintended negative effects.

Potential harms to research participants or research staff from data collection (e.g., surveying, privacy, data management) or research protocols (e.g., random assignment)

Are data collection and/or research procedures adherent to privacy, confidentiality, risk-management, and informed consent protocols with regard to human subjects? Are they respectful of community norms, e.g., community consent not merely individual consent, when appropriate? Are there potential harms to research staff from conducting the data *collection that are beyond "normal" risks?*

IRB approval for this study was granted by the International Food Policy Research Institute (IFPRI), protocol #00007490. Data collection followed standard ethical procedures, including informed consent, confidentiality protections, and secure data management. Respondents were informed that participation was voluntary and that they could decline to answer any question. The survey instrument did not include questions likely to cause distress or discomfort, and no personal identifiers were retained in the final dataset.

Financial and reputational conflicts of interest

Do any of the researchers have financial conflicts of interest with regard to the results of the research? Do any of the researchers have potential reputational conflicts of interest?

No. The researchers have no financial conflicts of interest to declare.

Intellectual freedom

Were there any contractual limitations on the ability of the researchers to report the results of the study? If so, what were those restrictions, and who were they from?

The researchers maintained complete autonomy over the analysis and dissemination of study findings.

Feedback to participants or communities

Is there a plan for providing feedback on research results to participants or communities? If yes, what is the plan? If not, why not?

There is no plan to provide direct feedback on research results to participants or communities. While the research team recognizes the value of sharing findings with participants, it was not practical in this context due to cost constraints. The study took place in urban IDP settings where participants were relatively accessible, but available resources were allocated to ensuring high-quality data collection and analysis. Findings will be shared with implementing partners and other stakeholders to support evidence-based decision-making.

Foreseeable misuse of research results

Is there a foreseeable and plausible risk that the results of the research will be misused and/or deliberately misinterpreted by interested parties to the detriment of other interested parties? If yes, please explain any efforts to mitigate such risk.

No.

Other ethics issues to discuss

Are there any other issues to discuss?

None.

A4 Outcomes of Interest

A4.1 Overview

Table A10: Description of the outcome variables

Outcome Variable	Definition	Survey Round(s)
Primary outcomes:		
Share of households experiencing moderate or severe hunger	Based on the household hunger scale (HHS), which is a 30-day recall, experience-based food insecurity indicator focused on hunger. Responses to three frequency-of-occurrence questions are recoded into rarely/sometimes (1) and often (2), with no responses assigned 0. Household scores range from 0 (no hunger) to 6 (severe hunger), where 0–1 indicates little to no hunger, 2–3 indicates moderate hunger, and 4–6 indicates severe hunger. Binary variable obtaining 1 if HHS ≥ 2.	All survey rounds
Total consumption	The sum of 'Food consumption' and 'Non-food consumption' (see below for definitions). Measured in per capita terms in 2017 $PPP.	1st and 2nd follow-up
Total value of assets	Based on 43 asset categories, including durable, productive, and livestock assets. Calculated by multiplying the number of assets owned by the household by the self-reported asset price.	All survey rounds[1]
Secondary outcomes:		
Food consumption	Aggregated consumption based on 69 food items with a 7-day recall, initially measured in quantities, then converted to monetary values using price estimates from a food price opinion module. Measured in per capita terms in 2017 $PPP. For more details about consumption measurement, see Appendix section A4.2.	1st and 2nd follow-up
Non-food consumption	Aggregated expenditures based on 58 non-food items or services with a 7-day, 1-month, 3-month, or 12-month recall, depending on expected purchasing frequency. Measured in per capita terms in 2017 $PPP. For more details about consumption measurement, see Appendix section A4.2.	1st and 2nd follow-up

Continued on next page

Outcome Variable	Definition	Survey Round(s)
Income from formal wage work	Earnings from formal wage work in the last 30 days, reported at the household level. Respondents specified whether the salary was received in cash or in-kind, providing amounts for each. For in-kind payments, they estimated the value in monetary terms. Measured in monthly terms and in 2017 $PPP. For more details about income measurement, see Appendix section A4.3.	2nd follow-up
Income from casual wage work	Earnings from casual wage work in the last 30 days, reported at the household level. Respondents specified whether the salary was received in cash or in-kind, providing amounts for each. For in-kind payments, they estimated the value in monetary terms. Measured in monthly terms and in 2017 $PPP. For more details about income measurement, see Appendix section A4.3.	2nd follow-up
Income from nonfarm business	Earnings from all household-operated nonfarm businesses in the last 30 days. For each business, respondents reported total expenses and revenues, with net income calculated as the difference. The final income value is the sum across all businesses, measured in monthly terms and in 2017 $PPP. For more details about income measurement, see Appendix section A4.3.	2nd follow-up
Income from livestock keeping	Net income from household livestock activities, calculated separately for each livestock category and summed across all categories. Revenues include inferred sales in the last 12 months and sales of livestock products in the last 30 days, while expenses cover inferred purchases in the last 12 months and upkeep in the past 30 days. Net income is measured in monthly terms and in 2017 $PPP. For more details about income measurement, see Appendix section A4.3.	2nd follow-up
Income from farming	Net income from household crop agriculture, calculated separately for up to three key crops and summed across all crops. Revenues include earnings from crop sales and the estimated value of unsold harvests in the last 12 months, while expenses cover production costs over the same period. Net income is measured in monthly terms and in 2017 $PPP. For more details about income measurement, see Appendix section A4.3.	2nd follow-up

Continued on next page

Outcome Variable	Definition	Survey Round(s)
Livelihoods coping score	Raw count of negative coping strategies used by the household in the past 30 days due to insufficient food or money to buy food. The indicator sums binary responses (yes = 1, no = 0) to ten specific negative coping strategies, including selling household assets, reducing non-food expenditures, selling productive assets, spending savings, borrowing money or food, selling house or land, withdrawing children from school, selling last female animals, begging, and selling more animals than usual. Higher values reflect greater reliance on negative coping strategies and indicate more severe food insecurity.	All survey rounds
Any cash savings	Binary variable based on the question "Does anyone in the household regularly save cash?"	Baseline and 2nd follow-up
Access to credit	Binary variable indicating whether any household member borrowed in cash or in-kind from any source in the past 12 months. In households with both a household head and spouse, borrowing was assessed for each of several sources (NGOs, informal/formal lenders, friends/relatives, VSLAs, etc.), with follow-up questions on who decided to borrow and how borrowed funds were used. In households without both a head and spouse, a simplified question asked whether any household member had borrowed in the past year, followed by a question about the source.	2nd follow-up Social cohesion
Factor score derived from selected community-level social cohesion items, using factor analysis with regression scoring. The score is based on responses to items about trust among members, the capacity to manage social problems peacefully, mutual help in times of need, fairness in public resource management, fair treatment by public officials, and whether people are heard when sharing concerns. Higher values reflect stronger perceived community cohesion.	2nd follow-up	
Locus of control	Factor score derived from selected locus of control items using factor analysis with regression scoring. The score reflects the respondent's sense of control over life events, where higher values indicate a stronger internal locus of control.	2nd follow-up

Notes: [1]The number of assets owned by the households were collected in all survey rounds, asset prices only at the 2nd follow-up. $PPP = Purchasing power parity dollars.

A4.2 Measuring Consumption

Introduction

This appendix outlines our approach to measuring daily per capita consumption expenditures in the first and second follow-up surveys.

We begin by describing two common challenges in conducting consumption surveys in IDP settings and the strategies we used to address them. Next, we provide an overview of the food and non-food consumption sub-modules included in the first and second follow-up survey questionnaires. We then explain our approach to price measurement and deflation. Finally, we describe the construction of the final consumption aggregate measures used in this study.

Challenge (1): Interview length

Administering standard consumption modules, such as those used in household consumption expenditure surveys (HCES) or the World Bank's Living Standards Measurement Study (LSMS) surveys, typically takes well over 60 minutes to complete. In insecure and fragile contexts, implementing such lengthy modules is often not feasible, necessitating alternative approaches (Pape and Mistiaen (2020)).[26] We opted for a reduced item list, focusing on the most frequently consumed items among households in the Bay region and in IDP settlements across Somalia, as identified in the 2017 World Bank survey (Pape (2017)).

Challenge (2): Misreporting

Another challenge relates to potential misreporting. Previous research on consumption measurement in IDP settings in Somalia and Sudan highlights concerns that respondents may intentionally under-report consumption in the hope of receiving larger aid allocations (Kaplan et al. (2020)). To trigger unconscious stimuli that encourage truthful reporting,

[26]One option is the split-questionnaire design (see Pape and Mistiaen (2018)), where survey items are divided into core and optional modules. Each household responds to the core module and one optional module, with within-survey imputation methods used to estimate total consumption and poverty status for each household. This approach was used in Wave 2 of the Somali High Frequency Survey, see Pape and Wollburg (2019).

researchers in this space have included an *honesty prime* at the beginning of the consumption module to activate a norm of truthfulness. For example, in wave 2 of the World Bank's 2019 Somalia survey, the consumption module began by asking respondents their opinion about the following scenario:

> *Khalid asks his good friend Bashir if he has some money that he can lend him to help pay for medicine for his sick son. Bashir has money but was planning to buy cigarettes with it. He lies and tells Khalid that he has none. Is it okay for Bashir to lie to Khalid?*

Response options were: (1) *Yes, it is okay for Bashir to lie to Khalid*; (2) *No, it is not okay for Bashir to lie to Khalid*; (3) *Refused to respond.*

Survey research suggests that incorporating such simple honesty primes can improve the accuracy of responses, including in IDP settings (Kaplan et al. (2020)). Accordingly, we adopted the same honesty prime from the World Bank Somalia survey in both our first and second follow-up surveys.

Consumption sub-modules

The consumption module administered in first and second follow-up surveys had three sub-modules: Household food consumption over the past 7 days; Household expenditures on food consumed away from home in the past 7 days; Household expenditures on non-food items and services. Below we provide more details about each consumption component.

Household food consumption over the past 7 days

The household food consumption module includes 69 food items. These items are selected using the World Bank 2017 data. We restricted the sample to households residing in the Bay region in which Baidoa is located to capture the typical foods consumed in this region. In addition, since households residing in IDP settlements may have different food consumption patterns, we also considered the food consumption of the 468 IDP households in the World Bank 2017 survey. The food consumption module in the 2017 World Bank survey had 114 food items. For the Baidoa IDP survey, we included food items that were consumed by at least 3 percent of the households residing in the Bay region or IDP settlements in the World Bank 2017 survey.

During the interview, the enumerators first went through the list of 69 food items asking whether the household consumed the item in the past seven days or not. The survey instrument then be programmed to carry forward all food items that the household reported to have consumed in the past seven days to the next section that asks about the quantity ('amount consumed') within the 7 days.

Households were allowed to report the consumed amounts in any unit. In the analysis stage, we converted all amounts to kilograms using the conversion factors provided by the World Bank in the 2017 survey. All consumed amounts were then be valued in Somali shilling terms using the food price data collected as a part of the survey (for more details, see below).

Household expenditures on non-food items and services

Following the 2017 World Bank survey design, we have four different recall periods for non-food items and services: 7-day, 1-month, 3-month, and 12-month. (All will ultimately be converted to daily estimates for the purposes of aggregation.) Non-food items that are expected to be purchased weekly (e.g., charcoal, public transport), the recall period is 7 days while items purchased more infrequently have longer recall periods. In total, we have 58 non-food items in the consumption module. As before, we selected these items by analyzing the 2017 World Bank survey data. The non-food consumption module in this survey had 90 non-food items. Out of these, 60 items were consumed by at least 3 percent of the households residing in the Bay region or by IDP households. We then omitted two items related to expenditures on pets that do not seem relevant in the IDP settlement context (expenditures on these two items were reported by 5-7 percent of the Bay area household, but none of the IDP households).

The enumerators first went through the list of 58 non-food items asking whether the household purchased the item during the recall period or not. The survey instrument was then be programmed to carry forward all non-food items that the household reported to have purchased to the next section that asks about the amount spent during the recall period.

At the first follow-up, respondents were asked to report expenditure values in Somali shillings only. In the second follow-up survey, they were given the option to report values either in Somali shillings or in US dollars. Since official exchange rates in Somalia are not frequently updated, we asked all respondents to provide their estimate of the prevailing exchange rate in Baidoa. We then used the median of these responses to convert all values reported in US dollars to Somali shillings.

Food price data and deflation

To value the reported food consumption in Somali shillings, we used the price data collected both at first and second follow-up. We opted for using a price opinion module embedded in the household questionnaire (see Gibson and Rozelle (2005)) in which we asked households to estimate food prices in their locality. The list of food items in the price module was matched the food items in the food consumption modules. However, households were not asked to estimate the price of all food items. Instead, each household will be asked to estimate only 4 food items, resulting in more than 200 price estimates for each food item. To reduce the influence of extreme values, we took the median estimate for each item to represent the price in Baidoa.

After ensuring that all reported consumption expenditure values were expressed in Somali shillings, we deflated them to 2017 purchasing power parity (PPP) terms. For both follow-up survey rounds, we first used the food consumer price index (CPI) series from the International Monetary Fund (IMF) to adjust prices to 2017 levels. Then, we applied the World Bank's PPP conversion factor for individual consumption expenditure by households to express values in 2017 PPP terms.

Consumption aggregate

The consumption aggregate was then formed as the sum of the two consumption components listed above. All consumption amounts were transformed into daily terms by dividing the reported amount by the number of days in the recall period. Daily per capita household consumption expenditures were then constructed by dividing the consumption expenditure by household size.

A4.3 Measuring Income

Measuring income is challenging in low-income contexts characterized by large informal economies (Deaton and Zaidi (2002); Mancini and Vecchi (2022)), typically leading to sizable underreporting in household surveys (Carletto et al. (2022)). As described by Pape and Verme (2023), these challenges are even more pronounced for forcibly displaced persons (FDPs), who often lack regular labor income and instead rely on informal or sporadic earnings, in-kind aid, and cash assistance. FDPs may receive various forms of support, including household goods, food vouchers, and direct cash transfers, often from multiple donors. These irregular and diverse income sources are difficult to capture in surveys in FDP contexts, particularly since lengthy questionnaires are not feasible due to security concerns, which necessitate keeping interviews short (Pape and Wollburg (2019)).

Moreover, FDPs may supplement their income by producing and exchanging goods, with some items consumed rather than sold (Pape and Verme (2023)). Many also store wealth in portable assets such as jewelry or gold, though the extent to which these are used for daily or occasional expenses remains unclear (Pape and Verme (2023)). Given these complexities, estimating income in context like ours is highly challenging.

With these caveats in mind, the second follow-up questionnaire included questions on households' engagement in various income-generating activities over the past 30 days and the past 12 months. These activities covered formal wage employment, casual wage labor, non-farm businesses, livestock keeping, and crop agriculture. For each activity from which a household reported earning income in the past 30 days (12 months for crop agriculture), follow-up questions captured details on the amount earned. Below, we outline the measurement and aggregation approach used for each income component.

Formal wage work

A the second follow-up, less than 3% of households in our sample reported earning income from formal wage work in the past 30 days. For these households, we asked whether they received their salary in cash or in-kind. For each payment modality, we then asked about the amount received. If the payment was not in cash, we requested that the respondent estimate its value in monetary terms.

Casual wage work

At the second follow-up, 36% of households reported earning income from casual wage work. For these households, we asked whether they received their wages in cash or in-kind. For each payment modality, we then asked about the amount received. If the payment was not in cash, we requested that the respondent estimate its value in monetary terms.

Non-farm business work

At the second follow-up, 16% of households reported operating a non-farm business. For each business, we asked about total expenses and revenues over the past 30 days to calculate net income from non-farm business activities.

Income from livestock

In the second follow-up survey, households were asked to report: a) the income they received from the sales of livestock or livestock products during the past 30 days, and b) the expenses they incurred from keeping livestock during the same period.

However, this questionnaire design (employed for concision) also presents some key limitations. First, while it asks about income from livestock sales, it does not account for costs related to livestock purchases. Second, income from livestock sales and income from livestock products are pooled into a single question, limiting the ability to distinguish between these two sources of income. Third, livestock transactions are often irregular and involve large sums of money. For such infrequent transactions, a longer recall period than 30 days would reduce noise, though reporting transactions distant in time can also lead to recall bias.

To address these limitations in the questionnaire design, we construct the livestock revenue and expense variables by leveraging data from livestock asset modules fielded at first and second follow-up rounds that asks about the number of different livestock types owned by the household at the time of the interview. In the second follow-up, we also asked the price of livestock, which we can use to value the change in livestock counts. In addition, we also know the number of livestock that was transferred by NGOs since the first follow-up survey.

We can use these data to factor in the change in the value of stock when we calculate net income from livestock keeping. However, a key drawback is the need to make some strong assumptions about the sources of changes in livestock counts. Specifically, we assume that positive changes are due to either purchases or transfers from NGOs or births (which comes with additional assumptions), while negative changes are attributed to sales or home consumption (not mortality). To value changes in livestock between first and second follow-up rounds we follow these steps:

Step 1: Calculate expected livestock at second follow-up round

We calculate: expected second follow-up number of livestock = first follow-up number of livestock + NGO Transfers + expected births. We assume that for each livestock owned at the first follow-up or received as a NGO transfer since the first follow-up, on average, one new livestock should have been born by the second follow-up round. For example, for a household that owned one goat and received one through UPG, the expected stock at second follow-up would be four (one original goat, one transferred, two goat offspring).

Step 2: Infer change in stock

We then calculate the inferred change = expected second follow-up round number of livestock - actual second follow-up round number of livestock. (For example, if the hypothetical household described above then reports owning three goats at the second follow-up, we infer that one was sold.)

Step 3: Calculate Revenues from Sales and Expenses from Purchases

We calculate estimated livestock sales revenue, or expenses, as the number of livestock estimated bought or sold times the average livestock price. (For livestock purchases, we assume that purchased livestock are young kids or calves, and are bought at 50% of the price of an adult qnimql.)

Livestock revenues are then calculated as: "inferred livestock sales revenue between the first and second follow-up rounds" plus "revenues from the sales of livestock and livestock products in the past 30 days", annualized. Livestock expenses are calculated as "inferred livestock purchase costs between the first and second follow-up rounds" and "total costs from keeping livestock in the past 30 days, annualized."

Revenues from livestock product sales over the past 30 days are reported separately from inferred livestock sales revenue, which is calculated for the period between the first and second follow-up rounds. This creates a risk of double counting if high-value livestock sales occurred during the past 30 days and are reflected in both measures. To address this, we apply top-coding (winsorization) to the 30-day revenue measure, effectively removing potential overlap with the inferred livestock sales revenue.

After annualizing all revenues and expenses, the net income from livestock is then calculated by subtracting total expenses from total revenues.

Income from crop agriculture

Only 4% of households reported engaging in crop agriculture in the past 12 months. For these households, we asked them to identify the three most important crops they cultivated over the past year. For each crop, we then asked about the revenue earned from selling the harvest, the estimated value of the portion they did not sell, and the costs associated with production over the past 12 months. Net income from crop agriculture is calculated by subtracting total farming expenses from the total value of the harvest.

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