The (very) long-run impacts of cash grants during a crisis

Nathan Fiala[1,2], Julian Rose[2,3], Filder Aryemo[4], and Jörg Ankel-Peters[2,3]

[1]University of Connecticut, USA; [2]RWI – Leibniz Institute for Economic Research, Essen, Germany; [3]University of Passau, Germany; [4]Gaplink Uganda

June 2023

Abstract

We investigate the very long-run impacts of a randomized cash grant in Uganda during COVID-19 lockdowns. In 2008, the program supported young adults through a one-time entrepreneurial grant. Considerable effects materialized after four years, which vanished after nine years. For the 12-year follow-up, we pre-specified three outcomes, including a heterogeneity analysis by gender. We find positive effects on employment and income for men, but no effects on food security. We also find no effects for women. These re-surfacing effects suggest that the timing of a follow-up matters – an important insight for the growing literature on long-run studies.

Key words: cash transfers, long-run impacts, COVID-19.

JEL-Codes: C93, J24, O12, H53, I38

Acknowledgements: Fiala (corresponding author): nathan.fiala@uconn.edu. Rose: rose@rwi-essen.de. Aryemo: aryemofil@gmail.com. Ankel-Peters: peters@rwi-essen.de. Data collection was funded by the IZA Institute of Labor Economics (Grant Agreement GA-5-698), the German Federal Ministry of Education and Research (BMBF, funding code 011A1807A DECADE), and RWI – Leibniz Institute for Economic Research. We are grateful for valuable comments and suggestions by Christopher Blattman, Erin Kelley, David McKenzie, and participants at the G²LM|LIC COVID-19 Research Meeting, the 2022 German Development Economics Conference, 5th IZA/FCDO G²LM-LIC ONLINE Research Conference as well as seminars at IFPRI, UC Irvine, University of Connecticut, and the World Bank. The analysis underlying this paper is preregistered at: www.socialscienceregistry.org/trials/6158.

1. Introduction

Cash transfers have become a widely used tool to combat poverty, for example shaped as one-time entrepreneurial grants for equipment and training. The underlying theory is that the poor are trapped in poverty because, despite high returns to capital, they are subject to market imperfections preventing them from profitable investments. Short-term studies of entrepreneurial grants document mixed results (Banerjee et al. 2015; Blattman et al. 2014; Brudevold-Newman et al. 2017; De Mel et al. 2008; De Mel et al. 2012b; Fafchamps et al. 2014; Fiala 2018; Hussam et al. 2022). Recent long-term studies (ten and eleven years) show that the poor can sustainably increase their income (Banerjee et al. 2021; Balboni et al. 2022). In contrast, Blattman et al. (2020) find that substantial effects after two and four years observed in Blattman et al. (2014) do not seem to sustain nine years after the intervention.

The present paper examines the same intervention as in Blattman, Fiala, and Martinez (2014; 2020, henceforth BFM), the Youth Opportunities Program (YOP) in Uganda, a randomized one-time cash grant, now 12 years later and during the early months of the COVID-19 crisis in 2020. The economic consequences of the pandemic and restrictions to contain the spread of the virus were devastating in the Global South in general (Egger et al. 2021) and in Uganda in particular (Mahmud and Riley 2021; Hartwig and Lakemann 2020; Kansiime et al. 2021). More than a decade before, in 2007/2008, the Ugandan government implemented YOP to help poor and unemployed adults finance vocational training and equipment to become self-employed artisans.[1] We find that YOP participants' income is more resilient during the COVID-19 crisis, an average effect that is driven by the men in the treatment group.

YOP invited young adults to apply for cash grants averaging $380 (in 2008 USD)[2] per person. While the cash grant was labelled as a business creation subsidy, there was no enforcement after the disbursement. Individuals applied in groups with a proposal and grants were paid to the groups. Of 535 eligible groups, 265 were randomly assigned to treatment and 270 to control. A random selection of 2,598 YOP participants were originally surveyed for BFM (2014) and BFM (2020). In July

[1] The World Bank provided funding for the program, but the Ugandan government was responsible for its implementation.
[2] The transfer roughly amounted to the average annual income of the participants at baseline. The vast majority of recipients, 80 percent, received between 200 and 600 USD.

2020, in a first phase of data collection, we successfully retrieved 1,242 of them, using phone surveys because the pandemic situation did not allow in-person interviews. In the second phase, we tracked a random sample of those reached via phone, now in-person, and additionally reached 414 YOP participants. We implemented a survey experiment to check the data consistency across phone and in-person interviews and find no statistically significant differences. As the second round is a representative draw of those not reached in the first round, we have an *effective response* rate of 83.2 percent. Despite this high response rate compared to other studies implemented during COVID-19 and after such a long time (see Bouguen et al. 2019), we conduct thorough robustness checks to account for the conceivable non-randomness of attrition.

We pre-specified three primary outcomes before data collection in July 2020: *employment*, defined as whether the respondent worked for remuneration in the past seven days[3]; total *income,* defined as cash income in the last four weeks; and *food security*, defined as the number of days the household reduced the regular number of meals or portion size in the past seven days (see Table A10 for a detailed description of all outcomes)[4]. Additionally, we pre-specified six secondary outcomes which are discussed in the online appendix. For the primary outcomes, we find that the treatment group, on average, reports a 22 percent higher income (p=0.08). Figure 1 (Panel A) illustrates the income development across time for the treatment and control group. The decline in income for both groups between the 9-year follow-up in 2017 and our 12-year follow-up in 2020 is likely due to the COVID-19 lockdowns, yet for the treatment group the income shock was considerably alleviated.[5] Both groups are equally likely to be employed and these effects on income do not translate into higher food security. Moreover, a pre-specified heterogeneity analysis by gender reveals that the income effect is entirely driven by treated men who are also significantly more likely to be employed (p=0.02; Panel B in Figure 1) and report a 24 percent higher income than control men (p=0.08; Panel B in

[3] For employment and income, we consider daily labor, working for wages or in-kind, and self-employment including agricultural businesses. Participants that have produced crops or animal products for sale in the past four weeks are considered as being employed. See Table A11 or the pre-analysis plan for more details.

[4] Additionally, Table A11 provides the exact questions and coding of the outcomes. The pre-analysis plan can be accessed here: https://www.socialscienceregistry.org/trials/6158.

[5] Given that the effect for income is only statistically significant at the 10-percent level, we test whether the 9- and 12-year ITT effects are statistically different and cannot reject that they are the same. Additionally, we average the 9- and 12-year income at the individual level to assess whether seasonality or noise drive our results; using this new variable as outcome we still find a significant difference between treatment and control, which confirms our main result.

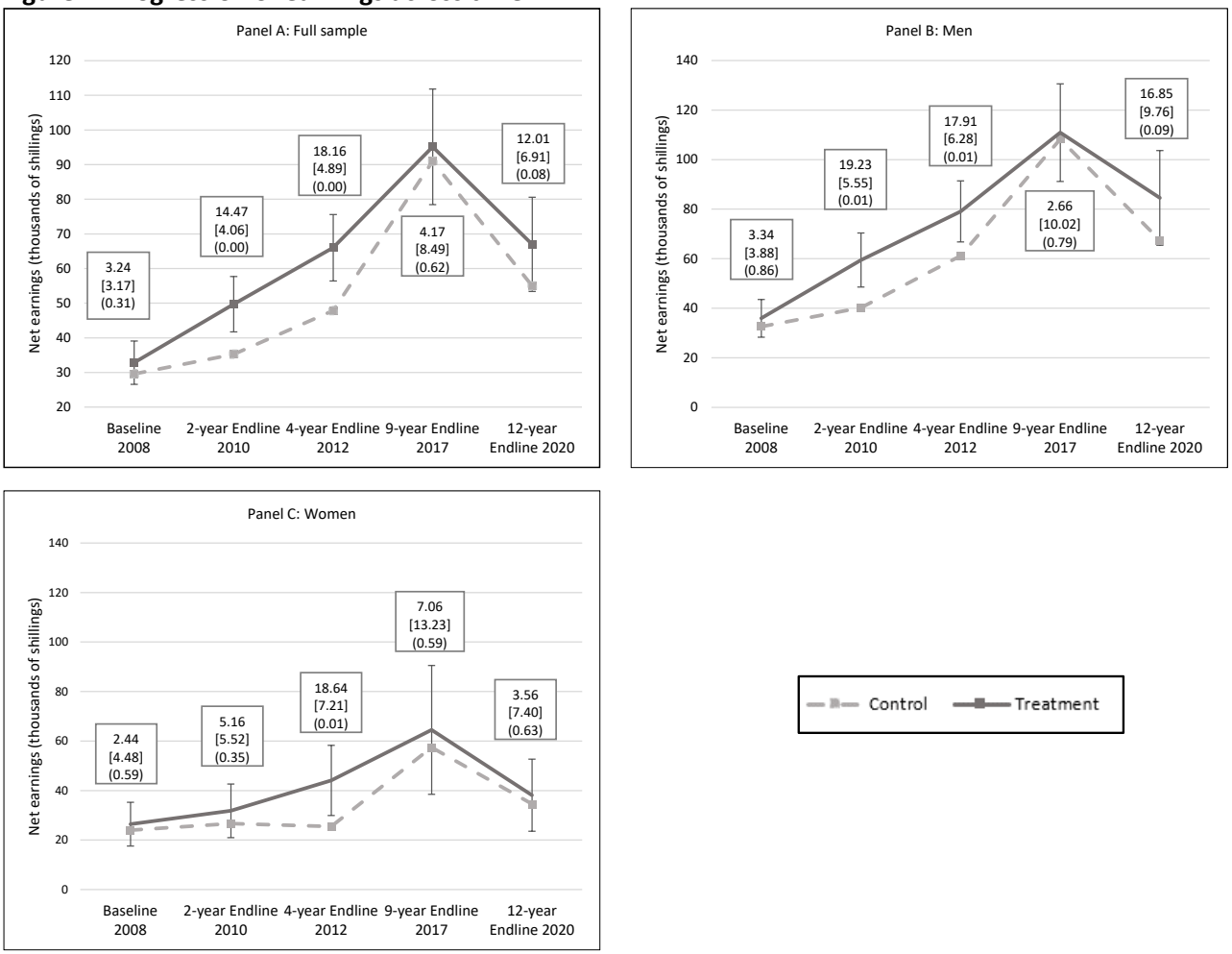
Figure 1). While our income estimates are relatively precise, given the long-term nature of the study, significance levels of our main specifications are marginally significant. We therefore use a specification curve analysis and, derived from this, a one-factor-at-a-time analysis (OFAT, see Figure 3) to scrutinize the robustness of results (see Simonsohn et al. 2019). The income effects are overall very robust. Only for some specifications the coefficients exceed the 10% threshold, for example very conservative attrition correction and one conservative top-coding scenario.

We explore (i.e., without having prespecified) potential mechanisms that might lead to the positive effects on income and find evidence that both the quality and the quantity of labor supply drive these results: like in the 4- and 9-year studies, men in the treatment group are substantially more engaged in what BFM called skilled trades. Consistently across all follow-up studies, skilled trades refer to all trades that were supported by YOP, which covers all occupations in this region that require skilled labor.[6] At the same time, our explorative analysis also shows that men in the treatment group report more working hours in the past four weeks. Hence, the higher income for men can stem from better jobs, more working hours, or a combination of the two. These plausible mechanisms enhance the credibility of our findings on income and employment.

The 4- and 9-year studies in BFM (2014, 2020) elicited a broad range of outcomes. For our 12-year study we partly resort to phone interviews (due to COVID-19 restrictions) and had to be more parsimonious by focusing on outcomes that are arguably affected immediately by the crisis. Table A2 in the online appendix offers a comprehensive overview of all outcomes in the 4-, 9-, and 12-year study, including whether the outcome was pre-specified for the respective study. BFM (2014) find substantial positive impacts on capital stock, income, consumption, and engagement in skilled trades after four years. After nine years, BFM (2020) observe that the control group has converged to the treatment group's income and consumption levels over time, arguing that, after a few years of economic stability and growth, the control group members found other profitable income sources. BMF (2020) also find that treatment group members have higher assets and are still systematically more engaged in skilled trades and work more hours – hence, the deeper structural changes we deem responsible for the income and employment effects during the COVID-19 crisis.

[6] Skilled trades comprise of tailoring, weaving, metal fabrication, blacksmith, carpentry, construction work, and running a hair salon.

Figure 1: Progression of earnings across time



Notes: We plot the mean value of reported monthly income for the control group and the sum of the control mean and the ITT estimate of YOPs impact. Income is in thousands of 2008 Ugandan shillings using the 2008 exchange rate of 1,720 shillings to $1. Results from 2008, 2010, and 2012 are published in BFM (2014); 2017 in BFM (2020); and 2020 in this paper. Inside the boxes, the ITT-effect (top), standard error (middle), and *p*-value (bottom) are shown.

Our paper contributes to the understanding of entrepreneurial cash transfer effects in the long run. According to Bouguen et al. (2019), the current prior in the small literature is that these interventions "initially help the poor to accumulate assets […] and these assets are generally gradually run down over time, generating little permanent impact on poverty". Persistent long-term effects only occur, as summarized in Bouguen (2019), if transfers are tied to very intensive support programs (as opposed to the light and untied vocational training in YOP). Examples are the multifaceted assistance in Bandiera et al. (2017) and Banerjee et al. (2021) where people received weekly consumption support and visits by trainers. However, as we show in the present paper, the *timing* of a long-term study matters for what results are obtained. Furthermore, Bouguen et al. (2019) emphasize that the effect heterogeneity along gender lines is large in the few existing studies. We

add another important observation by the considerable positive effects on men and the absence thereof among women.[7]

This paper also speaks to the general literature on social assistance and resilience during and after large shocks. De Mel et al. (2012a) randomly granted microenterprises cash and in-kind transfers after the 2003 Tsunami in Sri Lanka. They find that the additional capital increases the speed of recovery substantially. Rosas et al. (2022) study what they call a "cash plus" intervention, combining a regular cash stream with skills training during the Ebola crisis in Sierra Leone. Results suggest a positive impact on employment and profits. Similarly, Cañedo et al. (2023) show positive impacts of emergency cash transfers for informal workers during the COVID-19 pandemic on food security and psychological well-being. Premand and Stoeffler (2022) show that a small regular cash transfer can strengthen poor households' ability to mitigate the welfare effects of drought shocks in Niger. Bandiera et al. (2019) study the impact of a randomized women empowerment program implemented shortly before the Ebola outbreak in Sierra Leone and observe substantial positive effects during the epidemic. Similarly, Christensen et al. (2021) evaluate the impact of a randomized community monitoring program for health clinics implemented two years before the Ebola outbreak in Sierra Leone and show that the intervention successfully improved health clinics' performance before and during the epidemic. Casey et al. (2023) conduct an 11-year follow-up of a community-driven development program in Sierra Leone, finding suggestive evidence that the program improved communities' response to the Ebola epidemic in 2014.

2. Program implementation, randomization, and data

2.1. Context, Program implementation and randomization

The Ugandan government implemented YOP to stabilize and develop Northern Uganda in 2007/2008 as part of the Northern Uganda Social Action Fund (NSUAF) after a violent insurgency that afflicted the region in the early 2000s. YOP targeted poor, rural, and unemployed young adults aged between 16 and 35 to facilitate self-employment as craftsmen and -women. The government designed YOP as a community-driven development approach where groups could apply for cash

[7] More long-term studies are underway in related areas. See for example Agte et al. (2022) for an 11-year follow-up of a microfinance intervention.

grants of up to $10,000 in total, labelled for vocational training and tools, to start a skilled trade. The groups, consisting of 22 members on average, submitted a written proposal specifying how they plan to divide and spend the grant for vocational training, tools, and enterprise start-up costs. Although the group submitted the proposal together, members usually applied to set up an independent business. Typically, groups applied for one trade and selected their trainers, e.g., local artisans or small institutes. In most cases, group members came from the same village, and one application per village was submitted. Most groups are mixed, with only 5 percent of groups being all-female and 12 percent all-male.

The proposal passed several governmental screening levels, from the village level up to the national NSUAF office, prioritizing early and disqualifying incomplete submissions. The successful groups then received the cash grant as a lump sum in a bank account. After that, no monitoring or enforcement mechanisms were implemented. On average, groups received $7,497 ($382 per individual). Due to group size and requested amounts, the individual amount varied, with 80 percent receiving between $200 and $600.[8] We account for grant size heterogeneity in our analysis.

For the randomization, the program was oversubscribed such that 535 eligible groups were nominated in 14 districts in Northern Uganda. Then, 265 groups were randomly assigned to treatment and 270 to control, stratified by districts.[9] Shortly after the randomization in 2008, BFM conducted a baseline survey sampling five members per group resulting in a total sample of 2,598 (see Figure 2 for a detailed timeline of all events). BFM (2014) comprehensively discuss the balancing and show that the differences that are statistically significant are small in size. Like BFM, we include baseline covariates in our analysis.

Our 12-year follow-up took place during a time of crisis indeed: In response to the spread of the COVID-19 virus, the Government of Uganda gradually imposed a strict lockdown with an overnight curfew from the end of March until the end of May. The measures also included a ban on public transport and restrictions on private movement, the closing of international borders, and a restriction on activities outside the food sector.[10] While the lockdown led to a standstill of economic

[8] See Figure 1 in BFM (2014) which shows the distribution of group size and the average grant size per person.
[9] For more details on the experimental design, see BFM (2014).
[10] See Online Appendix A for a description of COVID-19 measures in Uganda.

activities in the non-food sectors, farming, and food vendors were less affected (Mahmud and Riley 2021; Hartwig and Lakemann 2020). For Kampala, Hartwig and Lakemann (2020) document that 81 percent of businesses in their sample were closed during the lockdown. This is also reflected in a substantial drop in profits and income. Similarly, studies report an income shock to non-farm income and an increase in food insecurity (Mahmud and Riley 2021; Kansiime et al. 2021). Their findings suggest that households shifted labor supply to agriculture during this period, a finding that is echoed in our data (Figure 4). Our qualitative interviews with YOP participants and community leaders suggest that the lockdown was strictly enforced in our communities with implications for their businesses and income similar to what was observed in Kampala.[11]

2.2. Previous Studies

BFM (2014) and BFM (2020) then conducted follow-ups in 2010, 2012, and 2017 to evaluate the effects of YOP after 4- and 9-years, as shown in Figure 2.[12] Table A1 summarizes the main findings of the previous studies and Table A2 offers an overview of all analyzed outcomes in the three studies including an indication on whether the outcome was pre-specified. BFM (2014) demonstrate YOP's successful implementation and document substantial positive impacts after four years. The treatment group is 53 percentage points more likely to have enrolled in vocational training and received 340 more hours of vocational training. Moreover, the treatment group reports a higher capital stock, more durable assets and income, is more engaged in skilled work, and higher consumption.

Yet, after nine years, BFM (2020) document only minor sustained effects. The control group managed to find work with similar levels of productivity and earnings. The control group's income and consumption converge towards the treatment group over time. The authors still find positive impacts on durable assets, and the treatment group systematically worked more and more in skilled trades.

[11] Transcript of qualitative interviews can be obtained from authors upon request.
[12] In Online Appendix B, we provide a comprehensive summary of both papers.

Figure 2: Timeline of YOP, surveys, and COVID-19 measures

Note: The distance between events does not represent the time that has passed.

2.3. Data

For the first tracking phase in July 2020, the phone survey, we were able to successfully reach 47.8 percent of the sample. For the second tracking phase in September 2020, now in-person, we randomly selected 44.5 percent of those not found during the first phase.[13] The sampling was stratified by treatment status and district, ensuring that the selected subsample represents the entire sample of those not found in the first phase. We successfully reached 68.5 percent of those selected for the second phase. Given the random sampling for phase two, we give higher weight in the analysis to those reached in the second phase using inverse sampling weights and disregard those not selected for intensive tracking. Therefore, we end up with an effective response rate of 83.2 percent.[14] In the analysis, we account for the tracking strategy by weighting those found in phase 1 with unit weight, and those selected in phase 2 are weighted by the inverse of their selection probability (results are not sensitive to this weighting).

Our attrition rate is very modest for a 12-year follow-up, but in case observations dropped out selectively from treatment and control group it would nevertheless be a concern. Yet, our attrition analysis does not support selective attrition. Moreover, the correlates of attrition for baseline characteristics are very weak predictors for attrition (with districts as strong predictors being an exception), and attrition patterns look similar to the previous studies.[15] Nevertheless, to remedy against a potential bias of attrition, we weight participants by the inverse of their predicted probability of attrition.[16] Furthermore, we include attrition scenarios in the OFAT and specification curve analysis.

[13] The decision to track 44.5 percent intensively was based on budget constraints.
[14] See Table A3 for an overview of the tracking results.
[15] See Table A4 and Table A5.
[16] The sampling and attrition weights are multiplied such that retrieved members of the sample who are more similar to the attritors receive more weight in the estimations. As robustness check, we run all regressions without these weights and find no major differences (see OFAT analysis in Figure 3 and specification curve in online appendix).

To rule out that the difference in survey mode (phone and in-person interviews) affect our results, we implemented a survey experiment to ensure consistency. Concerns regarding data quality in phone surveys are widespread, because, the argument goes, phone surveys are error-prone and facilitate misreporting (Arthi et al. 2018; Garlick et al. 2020; Heath et al. 2020). We therefore randomly selected YOP participants for different questionnaires (short and long income section) and varied whether the interview took place in-person or via phone (see the online appendix for more details). Our findings suggest that our main outcome variables are consistent across survey modes and, hence, we have no indication for bias in the phone survey responses.

3. Results

3.1. Estimation Strategy

We are primarily interested in the effect of YOP on employment, income, and food security, our three pre-specified outcomes. We estimate a simple intent-to-treat effect (ITT) of the program impacts on outcome Y via the weighted least squares regression:

$$Y_{ij} = \beta_{ITT} T_{ij} + \delta X_i + \alpha_d + \varepsilon_{ij} \quad (1)$$

Where Y_{ij} denotes the outcome for individual i in group j. T_{ij} is a dummy variable equal to 1 if the individual was part of the treatment group; X_i is the set of baseline covariates; α_d are district fixed effects and ε_{ij} is an individual error term clustered by group.[17] In addition to baseline controls, we include controls for the timing of the survey and survey mode (phone or in-person). Below we use a simple OFAT analysis (see Ankel-Peters et al. 2023) and a full-fledged specification curve analysis in the online appendix to scrutinize the robustness of results to our various specification decisions.

Table A10 displays the pre-specified primary and secondary outcomes and the way we calculated each outcome. Additionally, Table A11 includes the exact question in the survey. The primary outcome *Income* is measured in UGX and is the sum of all reported income in the past four weeks. Since income exhibits some outliers in the upper tail, partly caused by obvious enumerator errors, we top-code income at the 99th percentile as pre-specified in our pre-analysis plan (PAP), affecting

[17] We present results without baseline controls in the OFAT analysis (Figure 3).

15 observations in total. We include different top-coding scenarios in the specification curve (see Figure 3).

Following a results chain logic, we expect that the impacts on our three primary outcomes for the 12-year study materialize in a sequenced manner (see Figure A4). First, we assumed the COVID-19 lockdowns to affect the labor market and accordingly defined *employment* as primary outcome (intermediary impact in results chain terms). Second, more or better employment translates into higher income (between intermediary and final impact) and, further down the road, also food security. Food insecurity was one of the major concerns in the Global South in the first pandemic period. The underlying theoretical argument is that YOP has made the beneficiaries more resilient to shocks; for example, by its positive effect on skills. Lastly, we hypothesized that the sustained 9-year impacts, namely assets and savings, led to a higher resilience to shocks in the treatment group, which is reflected in our secondary outcomes (subjective resilience, business resilience, farming resilience, safety net, savings, and remittances).

3.2. Primary Outcomes[18]

Table 1 shows ITT effects for the primary outcomes. The lower part of the table presents the only pre-specified heterogeneity analysis, distinguishing effects for men and women. The presented interaction terms can be directly interpreted as elasticities. We do not find a positive impact on *employed* for the full sample (column 1). The mean of 0.67 in the control group suggests that pursuing economic activities was generally possible during lockdowns. For *income*, we find an effect that is statistically significant at the 10% level and economically meaningful at 12,000 UGX (column 2), which translates to a 22 percent higher income than the control group. Panel A of Figure 1 illustrates the development of monthly earnings over the entire 12-year period for the full sample. While there is a clear increase in monthly earnings until the 9-year follow-up, the consequences of the pandemic are visible, with a drop in income for the control and treatment groups. Yet, the decline is less strong for the treatment group resulting in a statistically different income after 12-years.[19] Finally, for *food*

[18] In addition to the primary outcomes, we defined and pre-registered several secondary outcomes. We do not find any meaningful impacts (see online appendix for a full discussion).

[19] To rule out that changes in sample composition between waves drive the income effect, we reproduce Figure 1 in the online appendix only for participants we successfully in all five waves (see Figure A6). The results do not change; if anything, both the effect on income and the statistical significance increase.

security we do not find a significant difference, suggesting that the higher resilience of the treatment group, reflected in its higher income, does not translate into strong enough improvements for food security to be measurable in our set-up.

Table 1: ITT effects for primary outcomes – pre-specified analysis

	Employed	Income in UGX (thousands)	Food Security
	(1)	(2)	(3)
Treatment	0.04	12.01	-0.02
	(0.03)	(6.91)	(0.25)
District FE	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes
p-value	0.17	0.08	0.95
q-value	0.29	0.29	0.46
Control mean	0.67	54.97	4.14
N	1466	1525	1524
R2	0.174	0.11	0.17
Treatment x Women	-0.05	2.52	-0.02
	(0.04)	(7.34)	(0.42)
Treatment x Men	0.08	17.15	-0.02
	(0.03)	(9.73)	(0.30)
Men	-0.08	17.38	-0.36
	(0.04)	(9.66)	(0.40)
District FE	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes
p-value Treatment x Women	0.31	0.73	0.97
p-value Treatment x Men	0.02	0.08	0.95
q-value Treatment x Women	0.767	0.87	1.00
q-value Treatment x Men	0.095	0.29	1.00
N	1466	1525	1524
R2	0.178	0.12	0.17

Notes: Standard errors in parentheses. Sampling weights are applied. Standard errors clustered at the group level. In all regressions we control for timing of interview and mode of interview (phone vs. in-person). To correct for multiple hypothesis testing, we calculate *q*-values for the upper and lower panel separately using the Benjamini-Hochberg step-up method. The *q*-values indicate the smallest false discovery rate at which the null hypothesis of zero effect is rejected. As baseline controls we include: Age, age squared, age cubed, male (only full sample), urban, risk aversion, highest grade, literate, vocational training, digit recall test score, ADL index, distance to educational facilities, wealth index, savings, monthly income, could borrow $58, could borrow $580, weekly hours in low skill/business/agriculture, in school, grant amount applied for, group size, grant amount per member, group existed before application, group age in years, within-group heterogeneity, group dynamic, group committee member, chair or vice-chair. Pre-specification took place before the data collection.

The heterogeneity analysis in the lower panel of Table 1 reveals relevant gender differences. We find that men in the treatment group have a 25 percent higher income and are 8 percentage points more likely to pursue income generating activities than men in the control group, which is statistically significant at the ten- and five-percent level respectively (columns 2 and 1, lower panel). Figure 1 displays the evolvement of income over time for men (Panel B) and women (Panel C) and

underlines the contrasting development between 9- and 12-year endline. The heterogeneity analysis does not reveal any impact for food security.

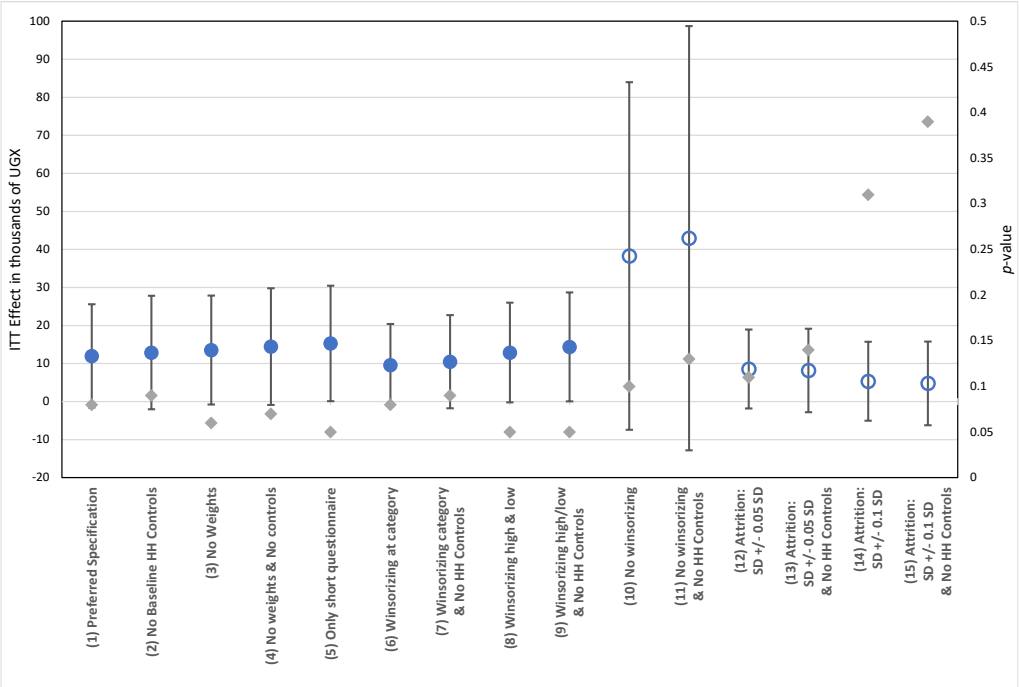
3.3. Robustness of Primary Outcomes

We emphasize that all estimates are somewhat imprecise, which is unsurprising given the long-term nature of our study. Both our full sample income finding and the one for men only are significant at the 10% level and thus the 95% confidence interval overlaps with zero (-1.57 to 25.59 and -1.98 to 36.27). While we observe decreasing standard errors from the 9- to the 12-year follow-up (Figure 1), the precision of our estimates has obvious robustness implications which we now scrutinize by means of a OFAT sensitivity graph (Ankel-Peters et al. 2023) in Figure 3, showing the most relevant specifications obtained from two full-fledged specification curves in the online appendix (Figure A1 and Figure A2). Given the borderline significance of our main specifications it is evident that some manipulations to this specification lead to lower *p*-values. Most specifications deliver very similar results in terms of effect size and *p*-values range between 0.05 and 0.14 for 13 of 15 specifications. Nine of 15 specifications deliver *p*-values below the 10% threshold.[20]

When we do not winsorize income in specifications (10) and (11), the effect becomes very large and imprecise. This notable change is primarily driven by a single observation reporting an unusually high income of over $13,000 in the last four weeks, in contrast to $61 for that same observation in the 9-year study. However, when we adjust only this specific observation, the effects closely align (18.98) and maintain statistical significance. Specifications (12) - (15) explore different attrition scenarios, assuming dropped out control observations (treated observations) have incomes 0.05/0.1 SDs higher (lower) than the control group's (treatment group's) mean. Even when using 0.1 SDs, the effects remain positive yet statistically insignificant. Yet, considering the absence of evidence for systematic we are confident that these scenarios are conservative.

[20] See Online Appendix D for a comprehensive specification curve analysis and discussion.

Figure 3: OFAT sensitivity analysis for income effects



Notes: The OFAT analysis shows the effect sizes (circles) and 95%-Confidence Intervals for alternative specifications. A filled circle indicates significance at least at the 10%-level. (1) represents our preferred specification as outlined in section 3.1. (2) does not include controls for baseline household characteristics. (3) performs the same analysis as (1) but without weights accounting for the tracking strategy. (4) combines (2) and (3). (5) only includes participants interviewed with the short questionnaire (see Appendix D for a description of the survey experiment). (6) Instead of winsorizing after aggregating all income sources (as pre-specified), we winsorize each income category before aggregating. (7) combines (6) and (2). (8) winsorizes income after aggregating but includes also the 0.01 percentile to account for extreme negative values. (9) combines (8) and (2). (10) uses income without any winsorizing. (11) combines (10) and (2). (12)-(15) are different attrition scenarios where we impute for all dropped out control observations a 0.05 and 0.1 SD higher income than the mean in the control group and for all dropped out treated observations a 0.05 and 0.1 SD lower income than the mean in the treatment group.

Another caveat is the correction for multiple hypothesis testing. When applying the procedure pre-specified in the PAP, the Benjamini-Krieger-Yekutieli sharpened *q*-values, all coefficients for primary outcomes turn insignificant. Yet, this is technically due to high p-value for food security. The Benjamini-Krieger-Yekutieli method sequentially adjusts *p*-values in order of increasing magnitude, that is the large *p*-value for food security inflates the sharpened *q*-values. This correction, we believe, does justice to outcomes that are independently and equally likely to be affected by the treatment – which our three outcomes are not. Food security is clearly a downstream outcome and can only materialize plausibly contingent on an effect on income and employment. Therefore, the Benjamini-Krieger-Yekutieli correction in our case is overly pessimistic. Our theoretical prediction is more in line with a sequential results chain logic: our findings on income and employment,

outcomes that are closer to the treatment, do not translate into (measurable) effects on food security. In other words, if we had observed effects on food security only, this would have been an implausible finding and likely been due to an artefact in the data.[21] Yet, since we did not pre-specify this sequential logic and for reasons of transparency, we nevertheless report the *q*-values in Table 1.

3.4. Exploration of mechanisms

The explorative (i.e. not pre-specified) analysis in this section shows that the treatment group's income seems to decline less for men because they engage more in skilled jobs – and they work generally more hours. The caveat of this finding is that for the explorative analysis in this section we can only use the subsample interviewed with the long questionnaire because only here we collected detailed information for 35 activities (instead of summarizing them into five broad categories in the short questionnaire). Figure 4 is hence only based 194 observations. The patterns are nevertheless interesting and coherent.

Panel A and B depict the extensive margin of occupational choice for men and women respectively, so the share of participants stating that their primary occupation is agriculture or some skilled trade. The panels nicely show that, while agricultural engagement varies over time, it does not differ between treatment and control, neither for men, nor for women. What differs between treatment and control for both men and women is the share of participants pursuing skilled work. Here, the treatment effect seems to persist over time. Yet, the difference between treatment and control is on a constant decline since the 4-year follow-up.

At the intensive margin, the sample is too small to meaningfully compare working hours in skilled jobs between treatment and control group for both men and women[22]. Drawing on the full sample, we compare total reported working hours and find no significant differences for working days and hours in the past four weeks between treatment and control group (see Table A12). The

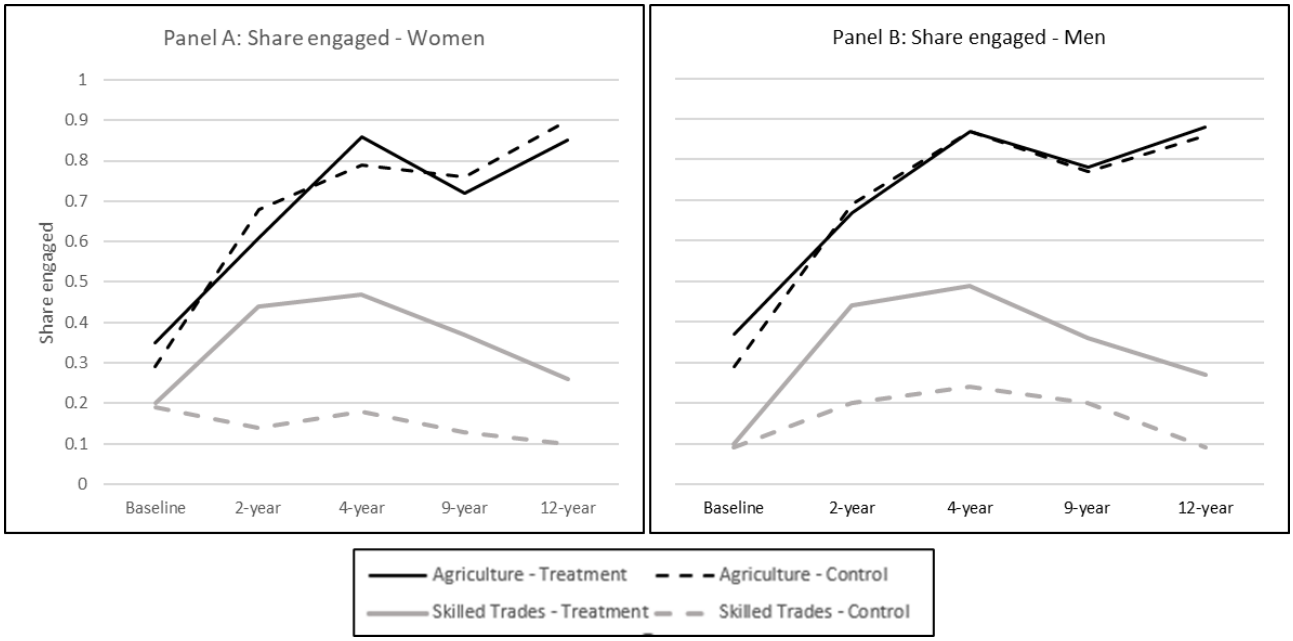
[21] This interpretation follows the concept of gatekeeping strategies in multiple hypothesis testing (Dmitrienko & Tamhane 2007). Here, families of analyses are examined sequentially and each one serves as a gatekeeper for the subsequent families of analyses. In our case, each step in the results chain from employment via income to food security represents a separate family of analysis.

[22] For the few observations, we do find that conditional on working in a skilled job, treated men (n=24) report substantially more hours than control men (n=5). For women, the working hours are very similar in treatment and control group (n=12).

heterogeneity analysis, though, shows that men work more in terms of days and hours. It is hence a consistent pattern that men engage more in skilled work and, in general, work more hours.

We also elicited data on a few other potential strategies to cope with the COVID-19 shock, suggesting that households mostly use their savings or borrow money to cover living expenses during the lockdown. Some households also sell assets. Yet, for none of the coping strategies we see a difference between the treatment and control group. We also see no difference in aid during the crisis from either the government or NGOs.

Figure 4: Labor supply in skilled trades and agriculture – not pre-specified



Notes: The data for baseline, 2-year, and 4-year is from BFM (2014); 9-year from BFM (2020); 12-year is from this data collection. The 12-year sample consists only of 194 participants from Groups 2 & 3 of the survey experiment (see online appendix for detailed description). Skilled trades are tailoring, weaving, metal fabrication, blacksmith, carpentry, construction work, and running a saloon. Agriculture includes subsistence and commercial farming. Sampling weights applied. We reproduce this using a consistent sample of participants that we successfully interviewed in all five waves, see Figure A7, showing very similar trends to the full sample.

4. Long-term RCTs: challenges and opportunities

Our paper is part of a nascent but rapidly growing literature that evaluates the long-term effects of randomized interventions. We reconcile our results with some dimensions put forward in Bouguen et al. (2019), a review of long-term studies of randomized interventions, by emphasizing that the timing of a long-run study matters: the positive 4-year YOP effects had largely vanished after nine years. Without our 12-year study the literature would have concluded that YOP had only short-term

effects. Yet, the deeper structural effects on the YOP treatment group that were also evidenced after nine years – on assets, occupational choice, and labor supply – apparently led to an increased resilience materializing in higher income and employment in the COVID-19 crisis, a period of economic decline.

Bouguen et al. (2019) also raise methodological concerns related to attrition and statistical power, which are potentially aggravated by the long-term character of these studies. While tracking our respondents was challenging during the lockdown and via phone, we managed to keep effective attrition rates at about 16 percent, which performs well vis-à-vis many other long-term studies reviewed in Bouguen et al. (2019). Moreover, Bouguen et al. (2019) voice concerns about a specific type of publication bias in long-term studies: If only interventions with very promising short-term or intermediary studies are followed up on in the very long run, this will lead to a misleading picture. This could be worsened in case statistical power is lower for the long-term follow-up than for the short– or mid-term study, leading to an Ioannidis et al. (2017) type power-of-bias problem. Statistical power might decrease for several reasons: attrition is one, but standard errors for outcome variables could also increase, for example because different subgroups of the study population are exposed to a changing environment to varying degrees. The standard error in our population indeed increased between the 4- and 9-year follow-up, but it decreased again in the 12-year follow-up. Regarding this type of publication bias, we believe our study is important because we decided to return despite a null effect for most indicators after nine years. Our findings thereby also confirm the concern raised by Bouguen et al. (2019) that systematic reviews of (very) long-term studies should be careful in assuming that null effects in the short term or after a longer period also imply null effects in the (very) long run.

Another potential caveat is that while RCTs ensure balance at baseline, treatment and control group might be differently influenced by subsequent policy changes, leading to an invalid counterfactual. For a welfare-to-work program in Canada, Riddell and Riddell (2020) showcase that a policy introduced after the treatment changed the counterfactual, leading to a biased policy conclusion. We have no indication from our repeated data collections for something similar in our sample. In

support of this, we collected data on whether the respondent has received support during the COVID-19 crisis and whether members of both groups are equally likely to have received support.

While pre-specifying outcomes for RCTs is standard practice, it is less clear whether and how to pre-specify for a long-term study with previous studies. For example, a non-prespecified statistically significant outcome in a mid-term study is typically suspected to be a statistical artefact. If that outcome is then prespecified prior to a long-term study and confirmed to be significant, that might well simply be the 'long-term effect' of the statistical artefact. On a similar note, one significant outcome among several insignificant outcomes that were all prespecified for the mid-term study would still require some sort of multiple-hypothesis correction in the long-term study, even if it is selected as the only outcome in the pre-specification of the long-term study. In our case, we selected the three outcomes and a heterogeneity analysis based on our predictions how the COVID-19 pandemic lockdown in Uganda would affect our sample – none of them were significant at the 9-year follow-up. In other words, as can be inferred from BFM (2020) and also Table A1, we did not cherry-pick those outcomes that were significant at the 9-year follow-up.

5. Conclusion

This paper has investigated the very long-term effects of a one-time entrepreneurial cash grant program that helped young adults become self-employed artisans twelve years prior to data collection – now in times of a crisis. Previous studies found substantial effects after four years (BFM 2014), which vanished after nine years (BFM 2020). Our key finding is that the treatment group reports significantly higher incomes and employment again after twelve years. The effect is driven by men; for women we do not find this effect. Food security, though, does not differ, neither for men nor for women. We transparently address challenges that are typical for long-term studies, most notably attrition, by conducting a OFAT analysis and a comprehensive specification curve analysis showing that the income effect is robust for virtually all specification choices. Only very conservative attrition corrections render results insignificant. In terms of mechanisms, we exploratively document that treatment group members work more hours and are engaged in better jobs. This, we believe, underpins the plausibility of our findings. Further mechanisms like different savings and selling assets are possible, but we find no evidence for this in our data.

Our result of a re-surfaced effect after twelve years is important as it suggests that the deeper structural changes induced by the treatment materialize in income and employment again in times of a crisis by making treatment group members more resilient. It generally also emphasizes the pertinence of the *timing* of a long-term follow-up: future long-term studies might particularly examine how effects develop after economic shocks like natural disasters. Indeed, our paper calls for more research on the long-term effects of RCTs because the heterogeneity of contexts and interventions we are anyway facing is certainly even more pronounced for long-term developments. More observations are needed to conclude which effects sustain – under which circumstances.

References

Agte, P., Bernhardt, A., Field, E. M., Pande, R., & Rigol, N. (2022). *Investing in the next generation: The long-run impacts of a liquidity shock* (No. w29816). National Bureau of Economic Research.

Ankel-Peters, J., Bensch, G., & Vance, C. (2023). *Spotlight on researcher decisions: Infrastructure evaluation, instrumental variables, and specification screening* (No. 991). Ruhr Economic Papers.

Arthi V., Beegle K., De Weerdt J., & Palacios-López A. (2018). Not your average job: Measuring farm labor in Tanzania. *Journal of Development Economics. 130*: 160–72.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., & Heil, A. (2022). Why do people stay poor?. *The Quarterly Journal of Economics*, *137*(2), 785-844.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I. & Sulaiman, M. (2017). Labor markets and poverty in village economies. *The Quarterly Journal of Economics*, *132*(2), 811-870.

Bandiera, O., Buehren, N., Goldstein, M. P., Rasul, I., & Smurra, A. (2019). *The economic lives of young women in the time of Ebola: Lessons from an empowerment program*. The World Bank.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., ... & Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, *348*(6236), 1260799.

Banerjee, A., Duflo, E., & Sharma, G. (2021). Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights*, *3*(4), 471-86.

Blattman, C., Fiala, N., & Martinez, S. (2014). Generating skilled self-employment in developing countries: Experimental evidence from Uganda. *The Quarterly Journal of Economics*, *129*(2), 697-752.

Blattman, C., Fiala, N., & Martinez, S. (2020). The long-term impacts of grants on poverty: Nine-year evidence from Uganda's Youth Opportunities Program. *American Economic Review: Insights*, *2*(3), 287-304.

Bouguen, A., Huang, Y., Kremer, M., & Miguel, E. (2019). Using randomized controlled trials to estimate long-run impacts in development economics. *Annual Review of Economics*, *11*, 523-561.

Brudevold-Newman, A. P., Honorati, M., Jakiela, P., & Ozier, O. W. (2017). A firm of one's own: Experimental evidence on credit constraints and occupational choice. *World Bank Policy Research Working Paper*, (7977).

Cañedo, A. P., Fabregas, R., & Gupta, P. (2023). Emergency cash transfers for informal workers: Impact evidence from Mexico. *Journal of Public Economics*, *219*, 104820.

Casey, K., Glennerster, R., Miguel, E., & Voors, M. (2023). Long-Run Effects of Aid: Forecasts and Evidence from Sierra Leone. *The Economic Journal*, *133(652),* 1348-1370.

Christensen, D., Dube, O., Haushofer, J., Siddiqi, B., & Voors, M. (2021). Building resilient health systems: Experimental evidence from Sierra Leone and the 2014 Ebola outbreak. *The Quarterly Journal of Economics, 136(2),* 1145-1198.

De Mel, S., McKenzie, D., & Woodruff, C. (2008). Returns to capital in microenterprises: Evidence from a field experiment. *The Quarterly Journal of Economics*, *123*(4), 1329-1372.

De Mel, S., McKenzie, D., & Woodruff, C. (2012a). Enterprise recovery following natural disasters. *The Economic Journal*, *122*(559), 64-91.

De Mel, S., McKenzie, D., & Woodruff, C. (2012b). One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka. *Science*, *335*(6071), 962-966.

Dmitrienko, A., Wiens, B. L., Tamhane, A. C., & Wang, X. (2007). Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Statistics in medicine*, *26*(12), 2465-2478.

Egger, D., Miguel, E., Warren, S. S., Shenoy, A., Collins, E., Karlan, D., ... & Vernot, C. (2021). Falling living standards during the COVID-19 crisis: Quantitative evidence from nine developing countries. *Science Advances*, *7*(6), eabe0997.

Fafchamps, M., McKenzie, D., Quinn, S., & Woodruff, C. (2014). Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana. *Journal of Development Economics, 106*, 211-226.

Fiala, N. (2018). Returns to microcredit, cash grants and training for male and female microentrepreneurs in Uganda. *World Development*, *105*, 189-200.

Garlick, R., Orkin, K., & Quinn, S. (2020). Call me maybe: Experimental evidence on frequency and medium effects in microenterprise surveys. *The World Bank Economic Review*, *34*(2), 418-443.

Hartwig, R., & Lakemann, T. (2020). *When the going gets tough: Effects of the COVID-19 pandemic on informal entrepreneurs in Uganda*. German Institute for Global and Area Studies. Retrieved from: https://www.giga-hamburg.de/de/publikationen/22075331-when-going-gets-tough-effects-covid-19-pandemic-informal-entrepreneurs-uganda/.

Heath, R., Mansuri, G., Rijkers, B., Seitz, W., & Sharma, D. (2020). Measuring employment: Experimental evidence from urban Ghana. *The World Bank Economic Review*, *35*(3), 635-651.

Hussam, R., Rigol, N., & Roth, B. N. (2022). Targeting high ability entrepreneurs using community information: Mechanism design in the field. *American Economic Review, 112*(3), 861-98.

Ioannidis, J., Stanley, T. and Doucouliagos, C. (2017). The power of bias in economics research. *Economic Journal*, *127*(605), F236-F265.

Kansiime, M. K., Tambo, J. A., Mugambi, I., Bundi, M., Kara, A., & Owuor, C. (2021). COVID-19 implications on household income and food security in Kenya and Uganda: Findings from a rapid assessment. *World development*, *137*, 105199.

Mahmud, M., & Riley, E. (2021). Household response to an extreme shock: Evidence on the immediate impact of the Covid-19 lockdown on economic outcomes and well-being in rural Uganda. *World Development*, *140*, 105318.

Premand, P., & Stoeffler, Q. (2022). Cash transfers, climatic shocks and resilience in the Sahel. *Journal of Environmental Economics and Management*, *116*, 102744.

Riddell, C., & Riddell, W. C. (2020). Interpreting experimental evidence in the presence of postrandomization events: A reassessment of the Self-Sufficiency Project. *Journal of Labor Economics*, *38*(4), 873-914.

Rosas, N., Acevedo, M. C., & Zaldivar, S. (2022). Starting points matter: Cash plus training effects on youth entrepreneurship, skills, and resilience during an epidemic. *World Development*, *149*, 1056.

Online Appendix

The (very) long-run impacts of cash grants during a crisis

Nathan Fiala
Julian Rose
Filder Aryemo
Jörg Ankel-Peters

Contents

A. Covid-19 in Uganda

On 18 March 2020, the Ugandan government started implementing the first measure to combat the pandemic's spread even before the first confirmed case. The first measures included the ban of public gatherings and the closing of all educational facilities (Museveni 2020). In the subsequent weeks, the measures became more rigorous, including a ban of public transport and closing of borders until eventually the Ugandan government imposed a nationwide lockdown with a curfew from 7 pm to 6.30 am on 30 March 2020. Initially, the government planned the lockdown for 31 days, but it was not until late May 2020 before the government eased the measures.

The lockdown from March to May affected income-generating activities differently. Farming and food vendors were less affected by the lockdown; farming was possible throughout the strict lockdown without any restrictions. Non-food businesses, though, were severely affected by the

complete shutdown of non-essential activities and weekly markets. Hartwig and Lakemann (2020) report for Kampala that 81 percent of businesses in their sample were closed from late March until the end of May. Our qualitative interviews confirm this finding in our sample. Yet, by the time of our interviews, 80 percent of the businesses were re-opened, and the recovery process had started. The complete shutdown of non-food businesses is also reflected in profits and income. Hartwig and Lakemann (2020) show that profits dropped substantially during the lockdown in Kampala. Findings by Mahmud and Riley (2021) show that the non-farm income of rural households in western Uganda dropped substantially (60 percent) during May 2020. Furthermore, they document a shift of labor supply to agriculture during the period of the strict lockdown. Our qualitative interviews with participants and community leaders are in line with those findings. During the strict lockdown from March to May, people focused on farming activities since other businesses were not allowed to operate. Additionally, the qualitative interviews underline that the lockdown was strictly enforced in the communities (transcripts are available upon request from the authors).

In response to businesses' distress, the government launched several programs to assist: (i) obligatory payments to the National Social Security Fund were put on hold; (ii) deferment of tax payments; (iii) economic stimulus package. However, the qualitative and quantitative evidence in our data suggests that these programs hardly reached our communities. Only 13 percent in our sample stated that they received support from the government or an NGO.

B. Previous Studies: BFM (2014) & BFM (2020)

BFM (2014) provide evidence for the YOP's successful implementation and positive impacts after four years. Of the treated groups, 89 percent received the grant.[1] The program successfully increased investments into human capital (skills training) and capital stocks (tools and materials). Table A1 provides an overview of the main 4- and 9-year findings, while Table A2 list all outcomes in all studies. In the treatment group, 68 percent enrolled in vocational training between 2008 and 2010, compared to 15 percent in the control group. Accordingly, the treatment group received on average 340 more hours of vocational training. Moreover, treatment successfully increased capital stocks.

[1] BFM (2014) note two reasons for groups not receiving the grant: (i) 21 groups could not access the funds due to unsatisfactory proposals, bank complications, or collection delays; (ii) 8 groups reported having never received funds due to theft. Baseline characteristics are generally very similar between receivers and non-receivers.

After four years, the treatment group reported 57 percent higher capital stocks. BFM (2014) also document a shift in occupation toward skilled work and increased total labor supply.

Table A1 shows that total hours worked a week increased by 17 percent relative to the control group. This increase is mainly in skilled trades leading to greater participation and hours in a skilled trade that is 2.5 times greater than in the control group. Lastly, BFM (2014) report effects on income and consumption. Income in the treatment group increased by 39 percent relative to controls after four years. For durable and nondurable consumption, they find 0.18 standard deviations larger consumption for both over the control group. Overall, YOP successfully set the treatment group on a growth trajectory after four years, translating into higher income and consumption levels and more durable assets.

After nine years, BFM (2020) finds overall only minor sustained effects on income. The authors explain the vanishing of effects by the fact that the control group found other kinds of work with similar levels of productivity and earnings. Over time, the control group's income converges towards the treatment group. After nine years, the impact on income is just 4.6 percent of the control mean and not statistically significant. For non-durable consumption, the effect is only 1.4 percent of the control mean, also statistically insignificant. Meanwhile, the 9-year impact on durable assets is significant, with 0.145 standard deviations greater durable assets in the treatment group. Yet, YOP did have lasting effects on occupational choices. The treatment group spent twice as much time in a skilled trade and was twice as likely to be working at least thirty hours per week in skilled trades.

Table A1: Summary of 2-, 4-, and 9-year impacts

	(1) ITT Coefficient	(2) Std. Err.	(3) Control Mean	(4) Obs.	(5) Comment
			BFM (2014): 2 and 4-year impacts		
Investments					
Enrolled in vocational training	0.532	0.023	0.152	1,999	Impact 2010 (2 years)
Hours of vocational training	340.5	22.521	49.0	1,999	Impact 2010 (2 years)
Business assets (000s 2008 UGX)	225.0	62.601	392.8	1,868	Impact 2012 (4 years)
Employment					
Avg. Employment hours per week	5.5	1.284	32.2	1,864	Impact 2012 (4 years)
Engaged in any skilled trade	0.261	0.026	0.22	1,868	Impact 2012 (4 years)
Works >= 30 hours a week in skilled trade	0.037	0.013	0.03	1,868	Impact 2012 (4 years)
Income					
Monthly cash earnings (thousands)	18.19	4.898	47.8	1,868	Impact 2012 (4 years)
Durable assets (z-score)	0.181	0.055	0.15	1,853	Impact 2012 (4 years)
Nondurable consumption (z-score)	0.18	0.051	-0.011	1,862	Impact 2012 (4 years)
			BFM (2014): 9-year impacts		
Employment					
Avg. Employment hours per week	0.513	1.593	44.68	1,981	
Engaged in any skilled trade					
Works >= 30 hours a week in skilled trade	0.029	0.011	0.03	1,981	
Income					
Monthly cash earnings (thousands)	4.172	8.491	90.97	1,981	
Durable assets (z-score)	0.145	0.047	0.25	1,981	
Nondurable consumption (thousands)	2.726	6.298	190.56	1,981	

Notes: The 2- and 4-year impacts are obtained from Table 3 in BFM (2014). 9-year impacts are obtained from Table 1 in BFM (2020). implies $p < .1$, implies $p < .05$, implies $p < .01$.

Table A2: Summary of all outcomes in the 4-, 9-, and 12-year follow-up

	4-year		9-year		12-year	
	Effect size	Std. Err.	Effect size	Std. Err.	Effect size	Std. Err.
Business assets	225	62.601	-	-	-	-
Food Security	-	-	-	-	-0.02[1]	0.25
Employment						
Avg. Employment hours per week	5.5	1.284	0.513	1.593	-	-
Engaged in any skilled trade	0.26	0.026	-	-	-	-
Works >= 30 hours a week in skilled trade	0.04	0.013	0.029[3]	0.011	-	-
Agricultural hrs/wk	0.4	0.945	0.08[3]	0.856	-	-
Nonagricultural hrs/wk	5.1	0.998	0.43[3]	1.488	-	-
Skilled Trades only hrs/wk	3.8	0.548	2.8[3]	0.529	-	-
No employment hours in past month	-0.02	0.009	0.03[3]	0.011	-	-
Casual labor, low skill hrs/wk			-1.21[3]	0.99	-	-
Petty business, low skill hrs/wk			-1.6[3]	1.069	-	-
High-skill wage hrs/week			0.91[3]	0.582	-	-
Employment (pursued income generating activity in past week)	-	-	-	-	0.04[1]	0.03
Labor supply – total days					0.58	0.93
Labor supply – total hours					6.87	6.62
Income						
Standardized Income Index	0.22	0.05	0.08[1]	0.048	-	-
Monthly cash earnings (thousands)	18.19	4.898	4.172[1]	8.491	21.22[1]	12.418
Durable assets (z-score)	0.18	0.055	0.145[1]	0.047	-	-
Nondurable consumption (z-score)	0.18	0.051	2.726[1]	6.298	-	-
Migration and urbanization						
Has changed parish since baseline	-0.07	0.026	-	-	-	-
Lives in large town or city	0.01	0.019	0.01	0.021	-	
Moved from village to town/city	-	-	0.02	0.017	-	-
Business formality						
Maintains formal records	0.12	0.023	0.03[2]	0.021	-	-
Enterprise is formally registered	0.06	0.019	-0.01[2]	0.013	-	-
Pays business taxes	0.09	0.023	0.02[2]	0.018	-	-
Hired labor						
No. of paid/unpaid laborers in past month	0.64	0.243	0.32[3]	0.165	-	-
Nonagricultural activities only	0.21	0.108	0.15[3]	0.083	-	-
Skilled Trade only	0.09	0.038	0.14[3]	0.042	-	-
Total hours of paid/unpaid laborers in past month	210.6	63.915	20.75[3]	8.909	-	-
Nonagricultural activities only	34.3	24.711	6.31[3]	6.83	-	-
Skilled Trade only	7.3	3.895	9.7[3]	3.763	-	-
No. of paid laborers in past month	0.26	0.148	0.26[3]	0.136	-	-
Nonagricultural activities only	0.08	0.05	0.11[3]	0.067	-	-
Skilled Trade only	0.05	0.026	0.1[3]	0.034	-	-
Total pay to others on typical working day	2.28	1.414	-	-	-	-

Nonagricultural activities only	0.82	0.743	-	-	-	-
Skilled Trade only	0.42	0.26	-	-	-	-
Estimated total pay to others in past month	32.3	33.99	0.85[3]	6.162	-	-
Nonagricultural activities only	7.2	15.65	0.4[3]	5.02	-	-
Skilled Trade only	5.5	3.174	3.05[3]	2.326	-	-
Number of family employees	-	-	0.09[3]	0.096		
Number of non-family employees	-	-	0.25[3]	0.144		
Savings						
Has savings account/savings group	-	-	-0.03	0.024	-	-
Amount of savings in 000s	-	-	-9.17	12.362	109.9[2]	66.035
Log savings	-	-	-0.02	0.106	-	-
Social Outcomes						
Kin integration	0.04	0.047	-	-	-	-
Community participation	0	0.05	-	-	-	-
Public goods contribution	0.01	0.049	-	-	-	-
Antisocial behavior	0.013	0.046	-	-	-	-
Protest and attitudes and participation	-0.02	0.043	-	-	-	-
Own health Outcomes						
Respondent passed away	-	-	-0.004	0.006	-	-
Physical health index (z-score)	-	-	-0.03[2]	0.047	-	-
Mental health index (z-score)	-	-	-0.06[2]	0.047	-	-
Fertility, HH size, and child expenditures						
Number of pregnancies 2007 or later	-	-	0.1[2]	0.101	-	-
Percent of births that were live 2007 or later	-	-	0.01[2]	0.01	-	-
Percent of pregnancies 2007 or later where child still living	-	-	0.01[2]	0.012	-	-
Percent of successful pregnancies 2007 or later where child still living	-	-	-0.01[2]	0.006	-	-
Number of biological children alive born 2007 or later	-	-	0.08[2]	0.083	-	-
Size of household	-	-	-0.13	0.162	-	-
Mean age of children (0–15)	-	-	0.01	0.138	-	-
Mean age of biological children (0–15)	-	-	0.1	0.147	-	-
Child educational outcomes						
Child age-adjusted educational attainment (6–24)	-	-	-0.01[2]	0.037	-	-
Child age-adjusted educational attainment (6–24), biological	-	-	-0.05[2]	0.045	-	-
Mean of child enrollment	-	-	-0.02[2]	0.013	-	-
Mean of child enrollment, biological	-	-	-0.02[2]	0.013	-	-
Current child expenditures (clothes and school)	-	-	0.41	2.784	-	-
Current child expenditures per child	-	-	0.5	1.071	-	-
Child health outcomes						
Mean health index per child, ages 3–9, family average	-	-	0.08	0.043	-	-
Mean parent-reported health score per child, ages 3–9, family average	-	-	0.07	0.047	-	-
Mean malaria cases in past year, ages 3–9, family average	-	-	-0.13	0.087	-	-

Mean normalized ADL score per child, ages 3–9, family average	-	-	0.05	0.041	-	-
Political Behavior						
Index of political action (z-score)	-	-	0.06[2]	0.05	-	-
Attended voter education meeting	-	-	0.03[2]	0.024	-	-
Discussed Vote	-	-	-0.003[2]	0.024	-	-
Reported campaign malpractice or incident	-	-	-0.013[2]	0.012	-	-
Voted in presidential election	-	-	0.01[2]	0.013	-	-
Attended political rally	-	-	0.01[2]	0.025	-	-
Participated in political primary	-	-	0.014[2]	0.024	-	-
Worked to get a candidate/party elected	-	-	0.04[2]	0.025	-	-
Member of a political party	-	-	0.05[2]	0.024	-	-
Index of NRM/Presidential support (z-score)	-	-	0.03[2]	0.05	-	-
Would vote for NRM if election were tomorrow	-	-	-0.01[2]	0.021	-	-
Like or strongly like NRM	-	-	0.01[2]	0.02	-	-
Worked to get the NRM elected	-	-	0.02[2]	0.03	-	-
Member of the NRM	-	-	0.04[2]	0.024	-	-
Voted or supported the president in the last election	-	-	-0.01[2]	0.02	-	-
Index of opposition support (z-score)	-	-	0.08[2]	0.044	-	-
Would vote for opposition if election were tomorrow	-	-	0.02[2]	0.016	-	-
Like or strongly like any opposition party	-	-	0.03[2]	0.022	-	-
Worked to get the opposition elected	-	-	0.01[2]	0.01	-	-
Member of an opposition party	-	-	0.04[2]	0.024	-	-
Voted or supported an election party in the past election	-	-	-0.01[2]	0.022	-	-
Resilience						
Subjective Resilience	-	-	-	-	0.18[2]	0.06
Business Resilience	-	-	-	-	0.07[2]	0.07
Farming Resilience	-	-	-	-	-0.001[2]	0.03
Economic Wellbeing	-	-	-	-	0.02[2]	0.03
Safety Net	-	-	-	-	-0.02[2]	0.03
Remittances						
Remittances sent in 000s	-	-	-	-	29.97[2]	43.531
Remittances received in 000s	-	-	-	-	-2.38[2]	13.641

Note: [1] outcome was pre-specified as primary outcome for respective follow-up. [2] outcome was pre-specified as secondary outcome for respective follow-up. [3] outcome was pre-specified as other outcome for respective follow-up. implies $p < .1$, implies $p < .05$, implies $p < .01$.

C. Attrition Analysis

Table A3: Tracking and survey response rate for 12-year follow-up

Selection and tracking by survey phase						Effective response rate			
Total sought	Found phase 1 (%)	Selected phase 2 (%)	Found phase 2 (%)	Final # of obs.	All (%)	Control (%)	Treated (%)	Weighted difference (%)	p-value
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2,598	47.8	44.5	68.5	1,656	83.2	82.0	84.3	0.3	0.895

Notes: Column (1) includes only observations that were found at baseline. Columns (6)–(9) report the effective response rates overall, by treatment status, and the treatment-control difference (calculated via regression, controlling for baseline district). Columns (6)–(9) are weighted by the inverse probability of selection in phase 2 of the survey and are referred to as ''effective'' response rates. Unfound respondents randomly dropped in phase 2 receive zero weight. Column (10) reports *p*-values on the difference term, using robust standard errors clustered at the group level. The weighted difference between attrited participants from the control and treatment group is calculated using weights.

Table A4: Correlates of survey attrition

	4-year endline		9-year endline		12-year endline	
	Coeff	SE	Coeff	SE	Coeff	SE
	(1)	(2)	(3)	(4)	(5)	(6)
Assigned to treatment	-0.03	0.02	0.04	0.02	0.02	0.02
Age at baseline	-0.00	0.04	0.01	0.04	-0.07	0.04
Age squared	-0.00	0.00	-0.00	0.00	0.00	0.00
Age cubed	0.00	0.00	0.00	0.00	-0.00	0.00
Male	0.05	0.02	-0.01	0.03	-0.04	0.03
Large town / urban area	0.18	0.04	0.12	0.03	0.13	0.03
Risk Aversion (z-score)	0.02	0.01	0.04	0.01	-0.01	0.01
Found at baseline	-0.70	0.06	0.16	0.06	0.00	0.00
Highest grade reached in school	-0.00	0.00	0.00	0.00	0.00	0.00
Able to read and write minimally	0.02	0.03	-0.04	0.02	0.02	0.03
Received prior vocational training	-0.04	0.04	-0.10	0.03	0.00	0.04
Digit recall test score	0.02	0.01	-0.00	0.01	0.00	0.01
ADL index	-0.00	0.00	-0.00	0.00	-0.01	0.00
Durable Assets (z-score)	-0.01	0.01	-0.01	0.01	-0.03	0.01
Savings (000s 2008 UGX)	0.00	0.00	0.00	0.00	0.00	0.00
Monthly gross earnings (000s 2008 UGX)	-0.00	0.00	0.00	0.00	-0.00	0.00
Could obtain 100,000 UGX (58 USD) loan	0.01	0.02	0.03	0.02	-0.01	0.02
Could obtain 1,000,000 UGX (580 USD) loan	0.01	0.04	-0.06	0.03	0.00	0.03
Weekly work hours: low skill	0.00	0.00	0.00	0.00	0.01	0.00
Weekly work hours: other business	-0.00	0.00	-0.00	0.00	-0.00	0.00
Weekly work hours: skilled trade	0.00	0.00	-0.00	0.00	0.00	0.00
Weekly work hours: high skilled trade	-0.01	0.01	-0.00	0.01	0.00	0.01
Weekly work hours: other non-agricultural	-0.00	0.00	-0.00	0.00	0.00	0.00
Weekly work hours: agricultural	-0.00	0.00	0.00	0.00	-0.00	0.00
Weekly household chores, hours	0.00	0.00	0.00	0.00	0.00	0.00
Zero employment hours in past month	-0.02	0.03	0.06	0.03	0.02	0.03
Main occupation is non-agricultural	0.03	0.05	0.09	0.04	0.00	0.04
Engaged in a Skilled Trade	-0.03	0.05	0.03	0.05	-0.02	0.05

Currently in School	-0.04	0.05	-0.06	0.05	0.03	0.05
Grant amount applied for (USD)	0.00	0.00	-0.00	0.00	0.00	0.00
Group Size	-0.00	0.00	-0.00	0.00	-0.01	0.00
Grant Amount per Member, USD	-0.00	0.00	-0.00	0.00	-0.00	0.00
Group existed before application	-0.01	0.02	-0.01	0.03	0.05	0.03
Group age, in years	-0.00	0.01	-0.00	0.01	-0.00	0.01
Within-group heterogeneity (z-score)	0.03	0.01	-0.00	0.01	0.03	0.01
Quality of in-group dynamic (z-score)	-0.02	0.02	-0.01	0.01	-0.01	0.01
Management committee member	-0.02	0.02	-0.01	0.03	-0.02	0.03
Chairperson or vice-chairperson	0.02	0.04	-0.05	0.03	-0.03	0.04
Distance to educational facilities (km)	0.00	0.00	0.00	0.00	0.00	0.00
Lives in Adjumani	-0.09	0.09	-0.06	0.06	0.35	0.09
Lives in Apac	-0.07	0.08	0.07	0.06	0.15	0.05
Lives in Arua	-0.03	0.08	0.16	0.07	0.16	0.06
Lives in Kaberamaido	0.03	0.10	0.03	0.08	0.14	0.07
Lives in Kotido	0.09	0.10	0.06	0.08	-0.11	0.05
Lives in Kumi	-0.07	0.08	-0.01	0.06	0.11	0.06
Lives in Lira	-0.09	0.08	0.10	0.08	0.07	0.06
Lives in Moroto	0.14	0.10	0.06	0.08	-0.02	0.06
Lives in Moyo	-0.16	0.08	0.12	0.09	0.52	0.08
Lives in Nakapiripirit	0.01	0.10	0.11	0.08	-0.08	0.05
Lives in Nebbi	-0.10	0.09	-0.00	0.06	0.05	0.05
Lives in Pallisa	-0.16	0.08	0.04	0.06	0.10	0.05
Lives in Soroti	-0.05	0.09	0.06	0.07	0.09	0.06
Mean	0.18		0.12		0.17	
P-value of F-test	0.00		0.00		0.00	
N	2,111		2,086.00		1,846.00	
	.00					
R-squared	0.27		0.11		0.17	

Notes: Each pair of columns report the results from a WLS regression of an attrition indicator on baseline covariates and district fixed effects. Standard errors are clustered at the group level. implies p < .1, implies p < .05, implies p < .01. Observations are weighted by the probability into selection of endline tracking, and errors are clustered by group.

Table A5: Correlates of survey attrition using 9-year outcomes (full sample)

	Attrited in 2020
Employment Outcomes	
Average employment hrs/wk	-0.000
	(0.96)
Agricultural hrs/wk	-0.000
	(0.96)
Non-agricultural hrs/wk	0.000
	(.)
Casual labor, low skill hrs/wk	0.001
	(0.50)
Petty business, low skill hrs/wk	-0.000
	(0.93)
Skilled Trades hrs/wk	-0.002
	(0.40)
High-skill wage labor hrs/wk	-0.002
	(0.25)
No employment hours in past month	0.000
	(.)
Main occupation is non-agricultural	0.011
	(0.77)
Engaged in any skilled trade	-0.021

	(0.53)
Works over 30 hrs/wk in skilled trade	0.124
	(0.23)
Average hours of chores per week	0.001
	(0.86)
Income and Consumption	
Average earnings/hr (000s of 2008 UGX)	0.033
	(0.05)
Standardized Income Index	0.000
	(.)
Monthly net earnings (000s of 2008 UGX)	-0.000
	(0.18)
Nondurable Consumption (000s of 2008 UGX)	0.000
	(0.80)
Durable assets	-0.024
	(0.12)
Politics	
Index of political action (z-score)	-0.050
	(0.37)
Attended voter education meeting	0.021
	(0.55)
Discussed Vote	0.047
	(0.22)
Reported campaign malpractice or incident	-0.005
	(0.93)
Voted in presidential election	0.000
	(.)
Attended political rally	-0.004
	(0.93)
Participated in political primary	-0.024
	(0.54)
Worked to get a candidate/party elected	0.062
	(0.47)
Member of a political party	-0.086
	(0.23)
Index of NRM/Presidential support (z-score)	0.093
	(0.18)
Would vote for NRM if election were tomorrow	-0.066
	(0.37)
Like or strongly like NRM	-0.053
	(0.45)
Worked to get the NRM elected	-0.061
	(0.51)
Member of the NRM	0.088
	(0.23)
Voted or supported the president in the last election	0.000
	(.)
Index of opposition support (z-score)	0.008
	(0.90)
Would vote for opposition if election were tomorrow	0.004
	(0.97)
Like or strongly like any opposition party	0.014
	(0.81)
Worked to get the opposition elected	0.000
	(.)
Member of an opposition party	-0.071
	(0.45)
Voted or supported an election party in the past election	-0.006
	(0.93)

Observations	1928

Notes: P-values in parentheses. Results from a WLS regression of an attrition indicator for attrited in 2020 on 9-year outcome and district fixed effects. Includes all observations found in 9-year follow-up. Standard errors are clustered at the group level. implies p < .1, implies p < .05, implies p < .01. Observations are weighted by the probability into selection of endline tracking, and errors are clustered by group.

D. Robustness Analysis

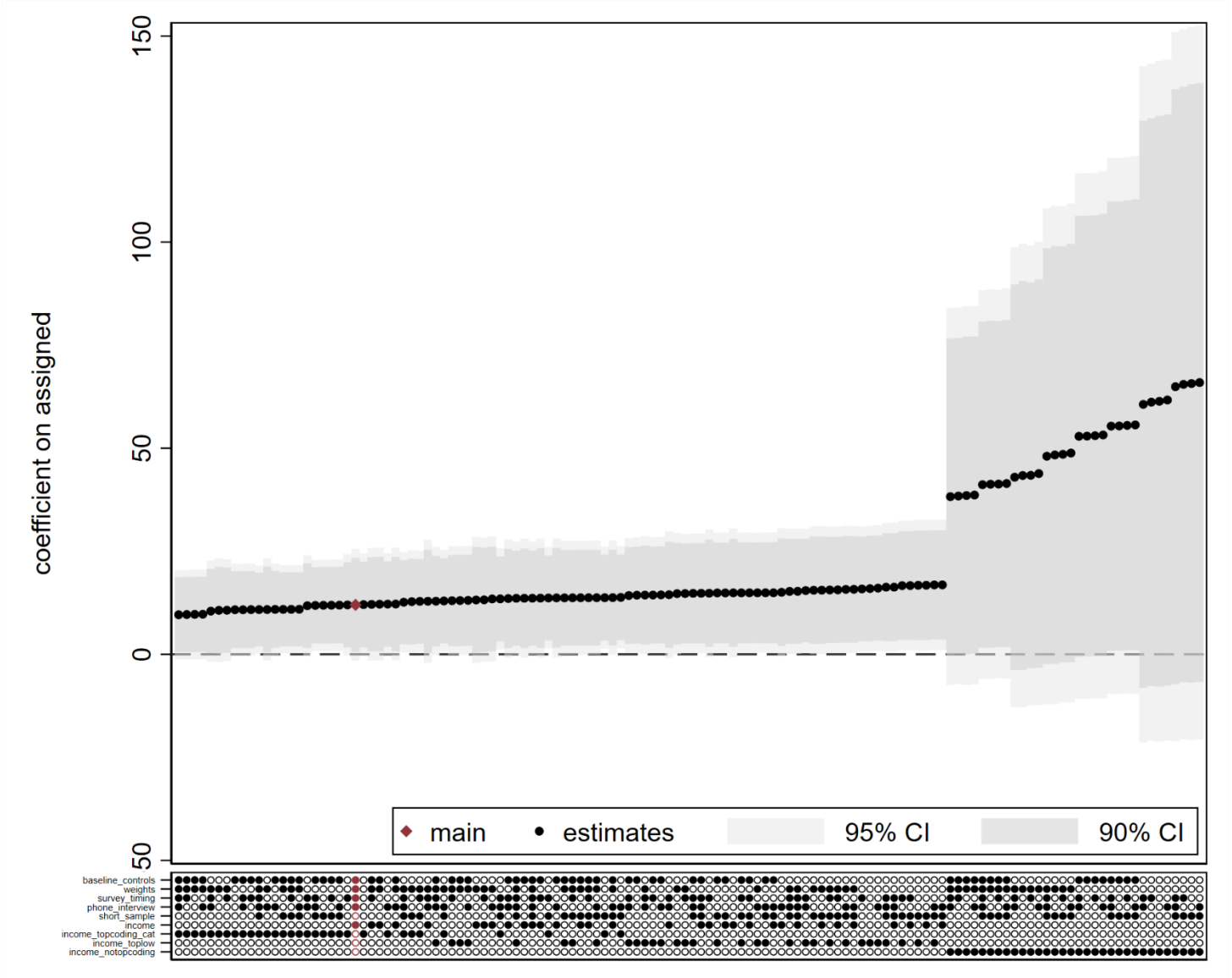
This section gives a more detailed overview of the multiple robustness checks. Figure A1 presents a specification curve, showing the influence of several researcher decision. It includes 128 different specifications. We include the following researcher decisions:

- Inclusion of baseline controls (baseline_control): whether to include a full set of household characteristics.
- Inclusion of weights (weights): whether to weight observations according to our tracking strategy.
- Control for timing of interview (timing_survey): the duration in weeks after the lockdown when the interview took place.
- Control for phone interview (phone_interview): whether a respondent was interviewed via phone or in person.
- Control for questionnaire (short_sample): exclude observations that were interviewed with the long questionnaire (see Appendix E for a description).
- Winsorize scenarios: we apply for different ways to winsorize income:
    - 99th percentile *after* aggregating all income sources (income)
    - 99th percentile *before* aggregating all income sources (income_topcoding_cat)
    - 99th and 1st percentile after aggregating (income_toplow)
    - No winsorizing (income_notopcoding).

Figure A1 shows that our main estimate is well placed within the different possible specifications. Moreover, most specifications produce larger effects as our reported effect. The only clear outliers occur when we do not winsorize income at all leading to a much larger yet, insignificant effect. A closer inspection reveals that one observation reported an income of over $ 13,000 in the last four weeks, compared to $61 in the 9-year study. When only changing this value, the effects are very close to our preferred specification (18.98) and remain statistically significant.

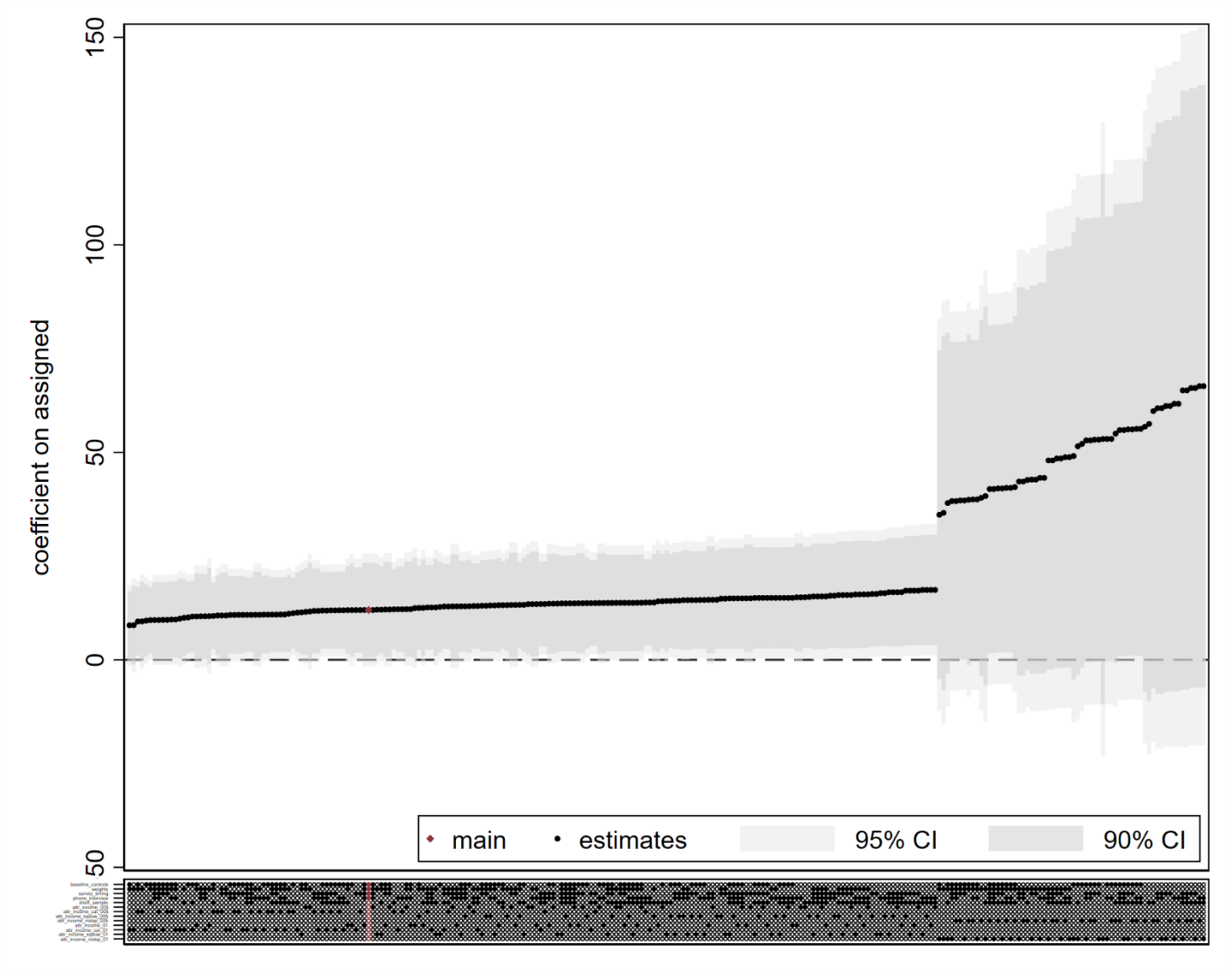
Figure A2 shows the specification curve for two attrition scenarios. In both scenarios we impute for attrited participants an income. For atttrited participants from the control group, we impute a 0.05 (0.1) standard deviations higher income than the average in the control group and for the treated attritors we impute a 0.05 (0.1) standard deviations lower income than the average in the treatment group. The pattern looks very similar to Figure A1 with effect sizes being somewhat smaller.

Figure A1: Specification curve for income effect



Notes: Description of the different models: "Baseline_controls" includes a full set of baseline household characteristics. "Weights" includes weights to account for the tracking strategy. "Survey_timing" includes a variable for the weeks since lockdown when the interview took place. "phone_interview" includes a dummy whether a respondent was interviewed via phone. "short_sample" includes only respondents interviewed with the short questionnaire. "income" is winsorized at the 99th percentile after aggregating all income sources. "income_topcoding_cat" winsorizes each income source before aggregating at the 99th percentile. "income_toplow" winsorizes income after aggregating at the 1st and 99th percentile. "income_notopcoding" does not winsorize income at all.

Figure A2: Specification curve for attrition scenarios (income)



Notes: This specification curve presents the results for two attrition scenarios: we impute for all attrited controls a 0.05 (0.1 SD) higher income than the mean in the control group and for all attrited treated a 0.05 (0.1 SD) lower income than the mean in the treatment group. The first scenario is denoted by "_005", the second by "_01". We apply the same specifications as in Figure A1 but just to the altered sample.

E. Survey Experiment

This follow-up study was initiated right after the pandemic started, and, by definition, we had to deal with a dynamic and unclear public health situation. Our response to this was to administer the first wave of interviews via phone surveys in July 2020. The response rate was at 47.8 percent and we, therefore, conducted a second wave of interviews among those we could not reach via phone,

now in-person, once the public health situation allowed. Since integrating phone- and in-person data requires that the survey mode does not systematically affect responses, we implemented a survey experiment in the second phase to compare phone and in-person responses to test for this consistency.

Figure A3 depicts how our surveys were sequenced, including the survey experiment. In the first phase, we sought to interview the entire sample via phone with a short questionnaire focusing on employment, income, and coping strategies during the COVID-19 pandemic. Of those we did not reach in the first phase, we randomly selected 603 for the second phase of in-person tracking.[2] For the experiment, we split the in-person interview sample into two groups and used two different questionnaires: a short questionnaire, identical to the one used in the phone survey phase, and a long questionnaire containing a more detailed income section.[3] The long questionnaire elicits employment and income outcomes for over thirty activities separately. In contrast, the short questionnaire contains only five broad activity categories, requiring respondents to add up numbers for working hours, days, and income. We used the long questionnaire for a randomly selected group (n=100) out of those not reached in the first phase, henceforth group *Long Questionnaire Group*. Another group (henceforth *Short Questionnaire Group*) was also randomly selected for in-person tracking (n=503) from those not reached in the first phase and interviewed with the short questionnaire.

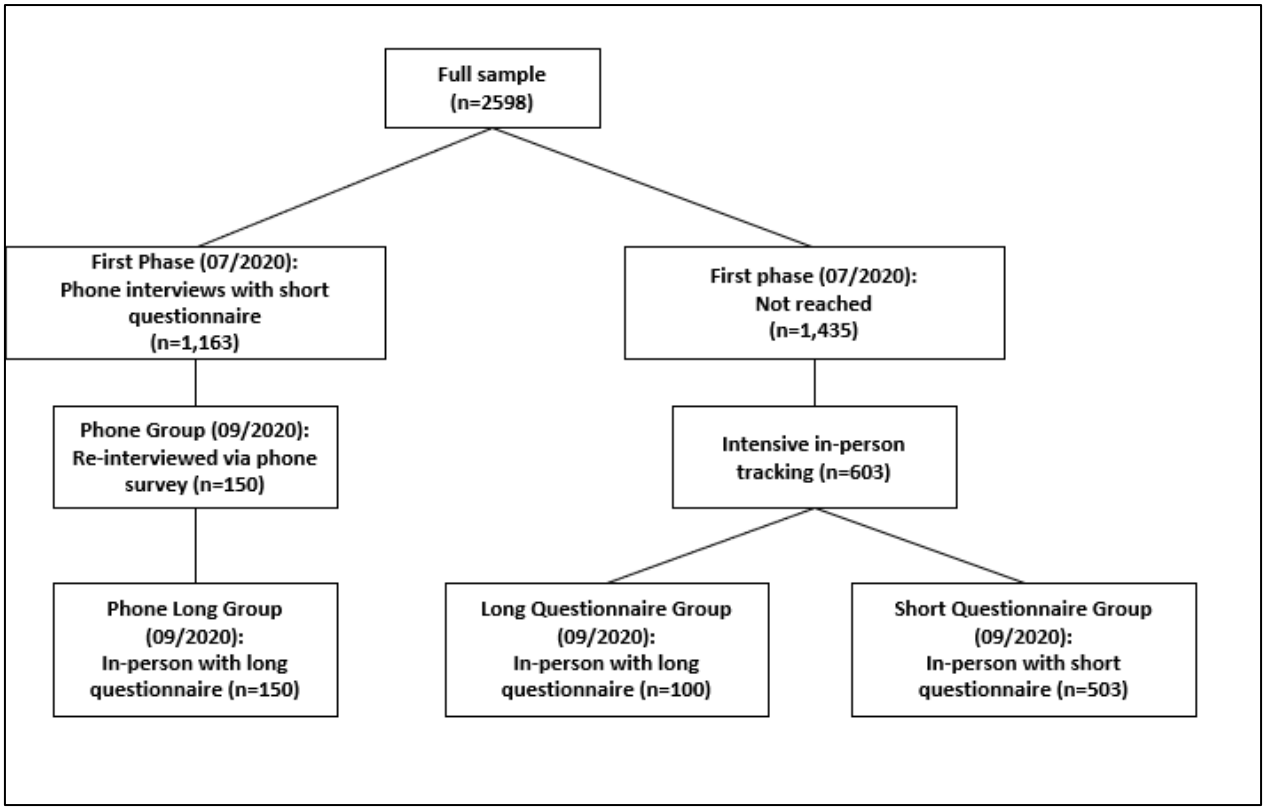
Moreover, we took a random sample of those reached in the first phase and conducted another two interviews with them (henceforth *Phone Group*). First, we re-interviewed *Phone Group* in the first week of September again with the short questionnaire via phone (so, precisely as in the first phase). We need this to correct for potential seasonality effects between the phone and the in-person group. Second, we re-interviewed the same group of people one week later in person with the long questionnaire (*Phone Long Group*).

The random allocation of the short and long questionnaires allows comparing the responses between the groups directly. We can, therefore, causally evaluate whether the questionnaire length influences

[2] Due to budget restrictions, we were not able to conduct intensive in-person tracking for the entire sample.
[3] The income section in the long questionnaire was taken from the 2017 data collection.

labor market outcomes in our sample.[4] Yet, the comparison between *Phone* vs. *Short Questionnaire*, which shows the effect of phone vs. in-person, rests on the assumption that there is no self-selection into participating in the first phase. Hence, we must assume that respondents reached in the first phase have similar characteristics to those not reached. Table A6 compares baseline characteristics of *Phone* versus *Long Questionnaire* and *Short Questionnaire* and finds some statistically significant differences. It seems that participants in *Phone* are more likely to consist of men employed in the non-agricultural sector. We include the full set of baseline controls and use district fixed effects in our analysis to control for these differences. The comparison between *Phone* and *Short Questionnaire* could also be influenced by repetition effects since the *Phone* group was interviewed twice in a short period. We cannot rule out that this is happening and are therefore unable to disentangle effects from the survey mode and repetition.

Figure A3. Design of the survey experiment



[4] Questionnaire length could also affect probability of responding, yet, we do not see that the response rate differs between the *Short Questionnaire* and *Long Questionnaire Group*.

Results of the survey experiment:

To test whether the data collection method affects our results, we compare data from phone vs. in-person and short questionnaire vs. long questionnaire. The latter is randomized through the survey experiment, allowing us to test for a systematic bias. We interpret the in-person long questionnaire interviews as the ground-truthed benchmark. Yet, for the comparison between phone and in-person data, we assume that there is no selection bias.

Table A7 presents the difference of in-person interviews with short and long questionnaire, hence, between the groups *Short Questionnaire Group* and *Long Questionnaire Group*. The results suggest that reported income is higher in the short questionnaire, yet the difference is not statistically significant.[5] For food security, we find a substantial and significant difference. This finding is surprising since there was no difference in the questions on food security in the short and long questionnaires. The only difference between the groups is the lengthier income section for *Long Questionnaire Group* leading to a longer overall interview time. In columns 3 and 4, we find that days/hours worked are substantially smaller in the short questionnaire.

Table A8 presents the results for comparing phone and in-person outcomes using the groups *Phone Group* and *Short Questionnaire Group*. To reduce concerns of self-selection, we include the full set of baseline controls and district fixed effects. We do not detect any significant differences between phone and in-person surveys for our primary outcomes (columns 1 to 3). There are also no differences regarding total days and hours worked. The findings of our survey experiment indicates that rather than the survey mode, the level of detail of the questionnaire plays a role. Overall, the results of the survey experiment increase our confidence in pooling phone- and in-person data.

Table A6: Baseline Balance for survey experiment groups

	Group 1 (=2)		Group 3 and 4			
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	SD	Mean	SD	Difference	p-value
Age at baseline	26.28	5.92	24.97	5.12	1.32	0.02
Male	0.74	0.44	0.58	0.49	0.16	0.001
Urban	0.18	0.38	0.15	0.36	0.02	0.55

[5] To ensure that our findings for income are not driven by this difference, we estimate the effect excluding all participants interviewed with the long-questionnaire and find similar effects. This analysis is available upon request.

Risk aversion (z-score)	-0.03	0.94	-0.01	1.06	-0.02	0.84
Highest grade reached in school	7.86	3.00	7.34	3.00	0.52	0.09
Able to read	0.74	0.44	0.71	0.46	0.03	0.48
Received prior vocational training	0.07	0.25	0.07	0.26	0.00	0.93
Digit recall test score	4.16	2.00	3.93	1.96	0.22	0.27
ADL index	8.63	2.22	8.80	2.67	-0.17	0.52
Durable assets (z-score)	-0.12	0.92	-0.07	1.01	-0.04	0.65
Savings (000s 2008 UGX)	39.04	157.13	22.26	100.72	16.78	0.16
Monthly gross earnings (000s 2008 UGX)	66.61	117.01	65.25	129.82	1.36	0.92
Could obtain 100,000 UGX (58 USD) loan	0.37	0.49	0.38	0.49	-0.01	0.91
Could obtain 1,000,000 UGX (580 USD) loan	0.11	0.31	0.11	0.32	-0.01	0.87
Weekly work hours: low skill	0.72	4.06	0.86	4.40	-0.14	0.75
Weekly work hours: other business	2.57	7.39	2.41	6.88	0.16	0.82
Weekly work hours: skilled trade	3.16	11.33	1.17	6.66	1.99	0.02
Weekly work hours: high skilled trade	0.24	1.38	0.05	0.60	0.20	0.03
Weekly work hours: other non-agricultural	0.65	3.78	0.44	3.49	0.21	0.57
Weekly work hours: agricultural	4.79	10.33	5.54	10.38	-0.75	0.48
Weekly household chores, hours	7.37	14.82	9.48	16.97	-2.11	0.21
Zero employment hours in past month	0.39	0.49	0.45	0.50	-0.06	0.24
Main occupation is non-agricultural	0.37	0.49	0.24	0.43	0.14	0.00
Engaged in a skilled trade	0.11	0.31	0.06	0.24	0.04	0.10
Currently in school	0.03	0.17	0.04	0.20	-0.01	0.59
Grant amount applied for (USD)	7159.00	2159.08	7525.72	2016.51	-366.73	0.08
Group size	20.76	6.24	22.33	7.26	-1.58	0.03
Grant amount per member, USD	372.47	165.52	375.93	172.05	-3.46	0.84
Group existed before application	0.44	0.50	0.43	0.50	0.01	0.83
Group age, in years	3.89	2.03	3.71	1.65	0.17	0.34
Within-group heterogeneity (z-score)	-0.07	1.00	-0.08	0.96	0.01	0.91
Quality of in-group dynamic (z-score)	-0.05	0.99	-0.10	1.08	0.05	0.65
Management committee member	0.33	0.47	0.25	0.43	0.08	0.08
Chairperson or vice-chairperson	0.16	0.37	0.09	0.29	0.07	0.03
Distance to educational facilities (km)	5.81	3.42	8.69	8.18	-2.88	0.00
Lives in Adjumani	0.00	0.00	0.04	0.20	-0.04	0.02
Lives in Apac	0.18	0.39	0.23	0.42	-0.05	0.25
Lives in Arua	0.08	0.27	0.08	0.27	0.00	0.92
Lives in Kaberamaido	0.00	0.00	0.03	0.18	-0.03	0.04
Lives in Kotido	0.06	0.24	0.09	0.29	-0.03	0.30
Lives in Kumi	0.05	0.23	0.11	0.31	-0.05	0.07
Lives in Lira	0.08	0.27	0.17	0.37	-0.09	0.01
Lives in Moroto	0.03	0.17	0.04	0.19	-0.01	0.69
Lives in Moyo	0.01	0.09	0.02	0.16	-0.02	0.24
Lives in Nakapiripirit	0.05	0.21	0.07	0.25	-0.02	0.42
Lives in Nebbi	0.08	0.27	0.01	0.07	0.07	0.00

Lives in Pallisa	0.19	0.39	0.04	0.20	0.15	0.00
Lives in Soroti	0.10	0.30	0.04	0.20	0.06	0.01

Notes: implies p < .1, implies p < .05, implies p < .01.

Table A7: Survey experiment (short questionnaire group vs long questionnaire group)

	(1) Income	(2) Food Security	(3) Total Days	(4) Total Hours
Short-Questionnaire	52,180	1.01	-13.29	-102.15
	(0.22)	(0.01)	(0.00)	(0.00)
District FE	Yes	Yes	Yes	Yes
Control Mean	66,784	3.07	40.13	191.78
N	366	366	366	366
R2	0.06	0.28	0.35	0.22

Notes: P-values in parentheses. Standard errors clustered at the group level. Due to a coding error in the long questionnaire, we do not have the employment outcome for this group. Control variable for timing of survey included. implies p < .1, implies p < .05, implies p < .01.

Table A8: Survey experiment (phone group vs short questionnaire group)

	(1) Employed	(2) Income	(3) Food Security	(4) Total Days	(5) Total Hours
In-Person	-0.09	-9,760	0.19	2.20	6.01
	(0.22)	(0.80)	(0.82)	(0.40)	(0.75)
District FE	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes
Control Mean	0.79	186,437	3.09	25.80	98.55
N	435	435	359	435	435
R2	0.28	0.20	0.37	0.35	0.22

Notes: P-values in parentheses. Standard errors clustered at the group level. Control variable for timing of survey included. As baseline controls we include: Age, age squared, age cubed, male (only full sample), urban, risk aversion, highest grade, literate, vocational training, digit recall test score, ADL index, distance to educational facilities, wealth index, savings, monthly income, could borrow $58, could borrow $580, weekly hours in low skill/business/agriculture, in school, grant amount applied for, group size, grant amount per member, group existed before application, group age in years, within-group heterogeneity, group dynamic, group committee member, chair or vice-chair. implies p < .1, implies p < .05, implies p < .01.

F. Secondary Outcomes

In addition to the primary outcomes, we defined and pre-registered several secondary outcomes. Table A10 displays the pre-specified secondary outcomes and Table A11 provides the exact questions. Table A9 presents the ITT effects for different measurements for resilience (column 1-3), economic well-being (column 4), safety nets (column 5), savings (column 6), and remittances (columns 7 and 8). The ITT effects in Table A9 suggest minimal effects of YOP on the secondary outcomes. In terms of resilience, we do not find that businesses or farming activities in the treatment and control group are differently affected (columns 2 and 3). Yet, we find that the treatment group

is subjectively more resilient in that they are more confident in coming up with UGX 100,000 in seven days (column 1).[6] The lower panel of column 1 suggests that this effect is driven entirely by men. In line with this finding, effects in column 6 suggest that the treatment group accumulated significantly more savings than the control, and again in particular, men are driving this effect. Having a high amount of savings can lead to more certainty in coming up with UGX 100,000. For remittances, we do not detect any meaningful transfers and no difference between the groups. A potential concern for our primary findings is that the treatment group received more aid during the crisis since they might be better connected to governmental programs due to their experience with YOP. Yet, findings on the safety net in column 5 suggest that the treatment group is not more likely to have received any support from the government or any NGO than the control group.

Table A9: ITT effects for secondary outcomes

	(1) Subjective Resilience	(2) Business Resilience	(3) Farming Resilience	(4) Economic Wellbeing	(5) Safety Net	(6) Total savings	(7) Remittances sent	(8) Remittances received
Assigned to treatment	0.18	0.07	-0.09	0.02	-0.02	132,102	29,968	-2,379
	(0.06)	(0.07)	(0.28)	(0.03)	(0.03)	(62,991)	(43,531)	(13,641)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value	0.01	0.35	0.75	0.51	0.49	0.04	0.49	0.86
q-value	0.02		1.00	1.00		0.207	1.00	1.00
Control Mean	1.241	0.330	0.335	1.452	0.340	448,185	187,149	98,958
N	1524	483	1524	1524	1466	1525	588	365
R2	0.162	0.228	0.105	0.125	0.108	0.101	0.125	0.257
Treatment x Women	0.06	0.17	0.31	-0.05	-0.04	23,483	105,531	-4,371
	(0.10)	(0.12)	(0.50)	(0.05)	(0.56)	(97,340)	(101,821)	(26,565)
Treatment x Men	0.25	-0.01	-0.29	0.06	-0.01	190,944	-1,295	-1,220
	(0.08)	(0.09)	(0.32)	(0.04)	(0.03)	(72,392)	(37,578)	(15,916)
Male	0.15	0.00	0.69	-0.01	-0.05	104,708	52,550	8,911
	(0.11)	(0.10)	(0.47)	(0.05)	(0.05)	(72,002)	(46,990)	(21,828)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value Treatment x Women	0.56	0.16	0.53	0.35	0.5	0.81	0.30	0.87
p-value Treatment x Men	0.00	0.94	0.36	0.16	0.72	0.01	0.97	0.94
q-value Treatment x Women	1.00		1.00	1.00		1.00	1.00	1.00
q-value Treatment x Men	0.03		1.00	1.00		0.154	1.00	1.00
N	1524	483	1524	1524	1466	1525	588	365
R2	0.164	0.230	0.105	0.127	0.11	0.102	0.128	0.257

Notes: Standard errors in parentheses. Sampling weights are applied. Standard errors clustered at the group level. In all regressions we control for timing of interview and mode of interview (phone vs. in-person). To correct for multiple hypothesis testing, we calculate q-values for the upper and lower panel separately using the Benjamini-Hochberg step-up method. The q-values indicate the smallest false discovery rate at which the null hypothesis of zero effect is rejected. As baseline controls we include: Age, age squared, age cubed, male (only full sample), urban, risk aversion, highest grade, literate, vocational training, digit recall test score, ADL index, distance to educational facilities, wealth index, savings, monthly income, could borrow $58, could borrow $580, weekly hours in low

[6] UGX 100,000 correspond to $28 which is around 1/20th of the GNI per capita.

skill/business/agriculture, in school, grant amount applied for, group size, grant amount per member, group existed before application, group age in years, within-group heterogeneity, group dynamic, group committee member, chair or vice-chair. implies $p < .1$, implies $p < .05$, implies $p < .01$.

G. Additional Tables and Figures

Table A10: Pre-specified outcomes

Outcome	Indicator	Coding
Primary Outcomes		
Employment	Respondent worked for remuneration last 7 days	Binary with 1 if respondent worked and 0 otherwise
Income	Income of respondent last month	Sum of respondent income in the past month. Coded as zero if respondent did not earn any income in the last month. Coded as missing if one of the subcategories is missing. Top censored at the 99th percentile to contain outliers
Food Security	Number of days with reduced number of meals or reduced portion size (household)	Additive index
Secondary Outcomes		
Subjective Wellbeing	Subjective Economic Status	Index constructed as average of the two ordinal variables
Business Resilience	Change in business operations	Question E 8 will be coded as: 0 business remains open as usual, 1 temporarily closed by government mandate, 2 business temporarily closed, 3 business permanently closed
Farming Resilience	Change in farming practices	Additive standardized index of 6 ordinal variables. All farming variables are coded to missing if off season or if household does not grow crops
Safety Net	Amount of savings	Sum of respondent savings in bank accounts and saving groups. Coded as zero if the respondent does not have any savings
Remittances Received	Amount of remittances received	Total amount of remittances received. Coded as zero if the respondent has not received any remittances
Remittances Sent	Amount of remittances sent	Total amount of remittances sent. Coded as zero if the respondent has not sent any remittances

Notes: We pre-specified the outcomes before the data collection started. The registration and the PAP can be accessed here: https://www.socialscienceregistry.org/trials/6158.

Table A11: Variable description for primary and secondary outcomes

Outcome	Indicator	Question	Coding
		Primary Outcomes	
Employment	Respondent worked for remuneration last 7 days	In the past 7 days, have you worked for remuneration for at least one hour? By "work for renumeration" we mean any activities you undertook for remuneration, including daily labor, working for wages or in-kind, or working on your own account or running a business, including an agricultural business.	Binary with 1 if respondent worked and 0 otherwise
Income	Income of respondent last month	Q1: For casual labor/salaried employment, what was your wage/salary in the last 4 weeks? By salary I mean the cash that you earned related to activity. Q2: For commercial farming/self-employed business owner, what was your profit from this farm in the last month? By profits I mean the cash that you earned minus all expenses related to activity.	Sum of respondent income in the past month. Coded as zero if respondent did not earn any income in the last month. Coded as missing if one of the subcategories is missing. Top censored at the 99th percentile to contain outliers
Food Security	Number of days with reduced number of meals or reduced portion size (household)	Q1: In the past 7 days, how many days have you or someone in your household had to… Limit portion size at mealtimes? Q2: In the past 7 days, how many days have you or someone in your household had to… Reduce number of meals eaten in a day?	Additive index
		Secondary Outcomes	
Subjective Wellbeing	Subjective Economic Status	Q1: Compared to last year, would you say the economic situation of your household this year has improved, stayed the same or worsened? Q2: Compared to your neighbors, would you say the economic situation of your household is better than average, about average or worse than average?	Index constructed as average of the two ordinal variables
Business resilience	Change in business operations	What is the current status of your business?	Question E8 will be coded as: 0 business remains open as usual, 1 temporarily closed by government mandate, 2 business temporarily closed, 3 business permanently closed)
Farming resilience	Change in farming practices	For your main crop... Q1: Relative to the same season in the last year, how many days did you and your household members spend on this activity on your farm? Q2: Relative to the same season in the last year, how many days did you hire workers to work on this activity on your farm? Q3: Relative to the same season in the last year, how many seeds and inputs (e.g. fertilizer, chemicals) have you used (do you plan to use) for your farm for this crop? Q4: Relative to the same season in the last year, how much have you harvested (do you expect to harvest) for your farm for this crop? Q5: Relative to the same season in the last year, how are /do you expect prices for this crop? Q6: Are you/do you expect to be able to sell your crop in the locations/markets where you usually sell it?	Additive standardized index of 6 ordinal variables All farming variables are coded to missing if off season or if household does not grow crops
Safety Net	Amount of savings	Q1: How much of your own money do you have saved in this bank account now?	Sum of respondent savings in bank accounts and saving

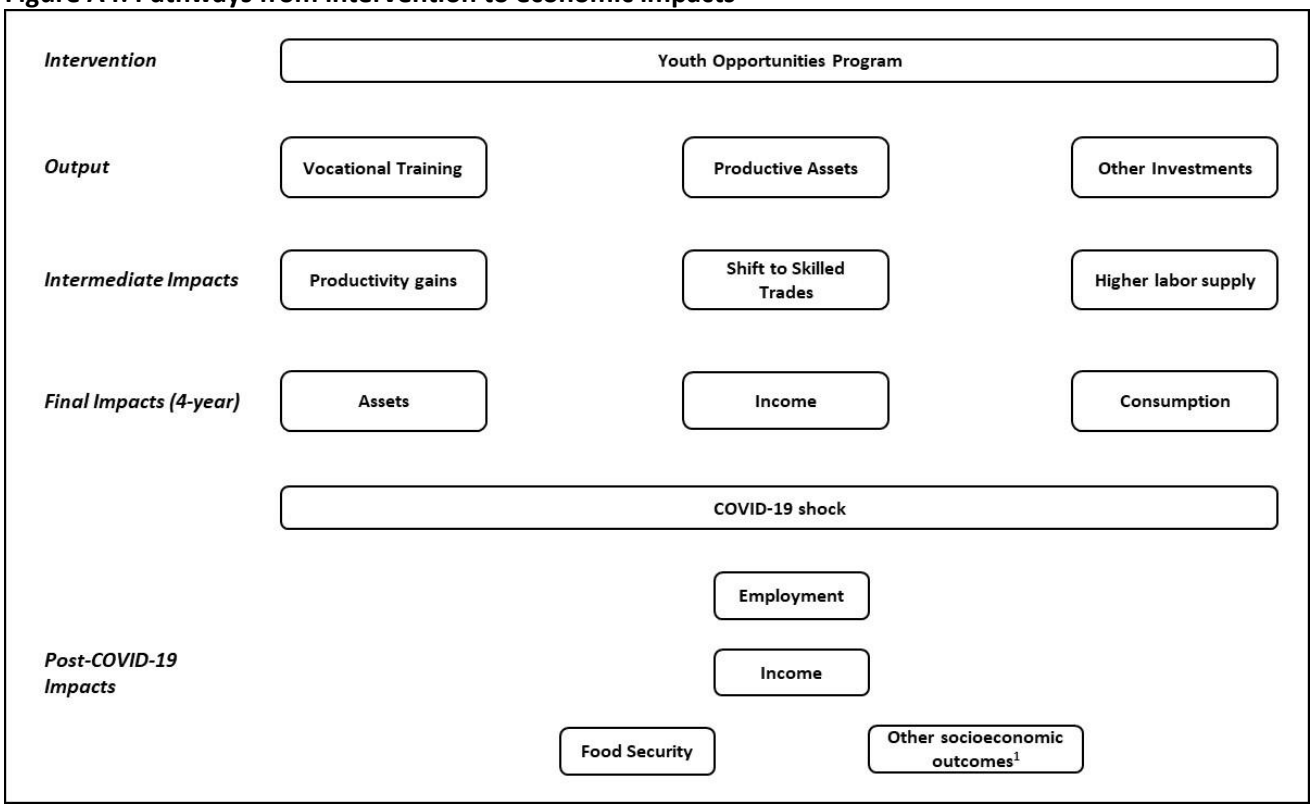
		Q2: How much of your own money do you have saved with these groups? Q3: How much money do you have saved in other locations (Just to clarify, savings do not have to be deposited in an account or formal institution, and they may or may not gain interest. They can be somewhere at home, hidden in a safe place, or with a friend or family member)?	groups. Coded as zero if the respondent does not have any savings.
Remittances Received	Respondent received remittances	How much (remittances received) in total since the lockdown (March 17th)?	Total amount of remittances received. Coded as zero if the respondent has not received any remittances
Remittances Sent	Respondent sent remittances	How much (remittances sent) in total since the lockdown (March 17th)?	Total amount of remittances sent. Coded as zero if the respondent has not sent any remittances

Table A12: ITT effects on labor supply – not pre-specified

	(1) Total Days	(2) Total Hours
Treatment	0.58	6.87
	(0.93)	(6.62)
District FE	Yes	Yes
Baseline Controls	Yes	Yes
P-values	0.53	0.30
Control Mean	27.92	93.29
N	1525	1525
R2	0.24	0.13
Treatment x Women	-1.72	-6.10
	(1.77)	(10.65)
Treatment x Men	1.83	13.90
	(1.03)	(7.97)
Men	-3.07	-4.16
	(1.81)	(10.93)
District FE	Yes	Yes
Baseline Controls	Yes	Yes
P-values Treatment x Women	0.33	0.57
P-values Treatment x Men	0.08	0.08
R2	0.24	0.13
N	1525	1525

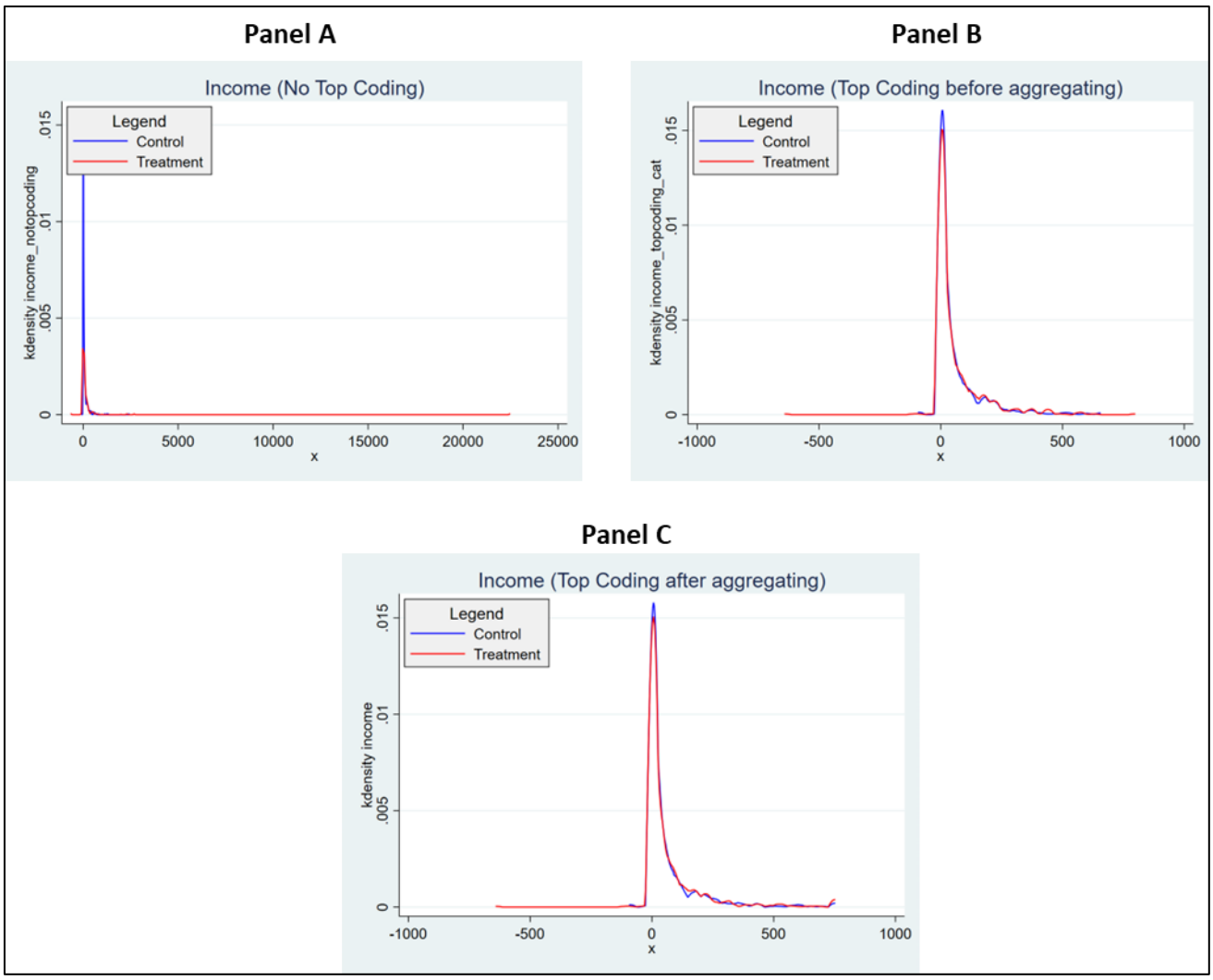
Notes: Standard errors in parentheses. Sampling weights are applied. Standard errors clustered at the group level. In all regressions we control for timing of interview and mode of interview (phone vs. in-person). As baseline controls we include: Age, age squared, age cubed, male (only full sample), urban, risk aversion, highest grade, literate, vocational training, digit recall test score, ADL index, distance to educational facilities, wealth index, savings, monthly income, could borrow $58, could borrow $580, weekly hours in low skill/business/agriculture, in school, grant amount applied for, group size, grant amount per member, group existed before application, group age in years, within-group heterogeneity, group dynamic, group committee member, chair or vice-chair. implies $p < .1$, implies $p < .05$, implies $p < .01$.

Figure A4: Pathways from intervention to economic impacts



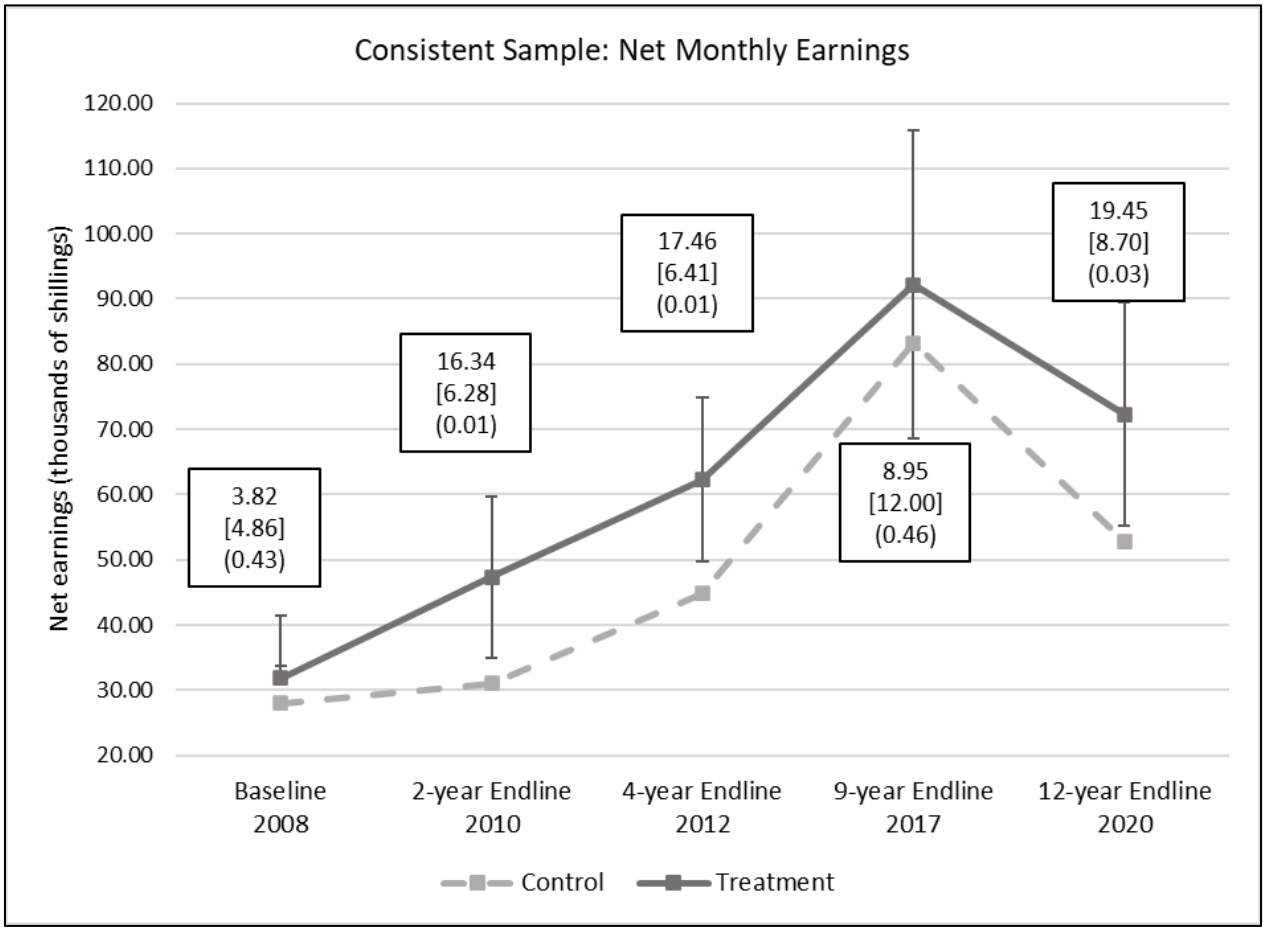
Notes: [2]We pre-specified several secondary outcomes capturing resilience: subjective resilience, business resilience, farming resilience, safety net, savings, remittances (see Table A11 for an overview). [#] BFM (2014 & 2020) document several additional socioeconomic outcomes, see Table A2 for a full list.

Figure A5: Distribution of income with different top-coding scenarios



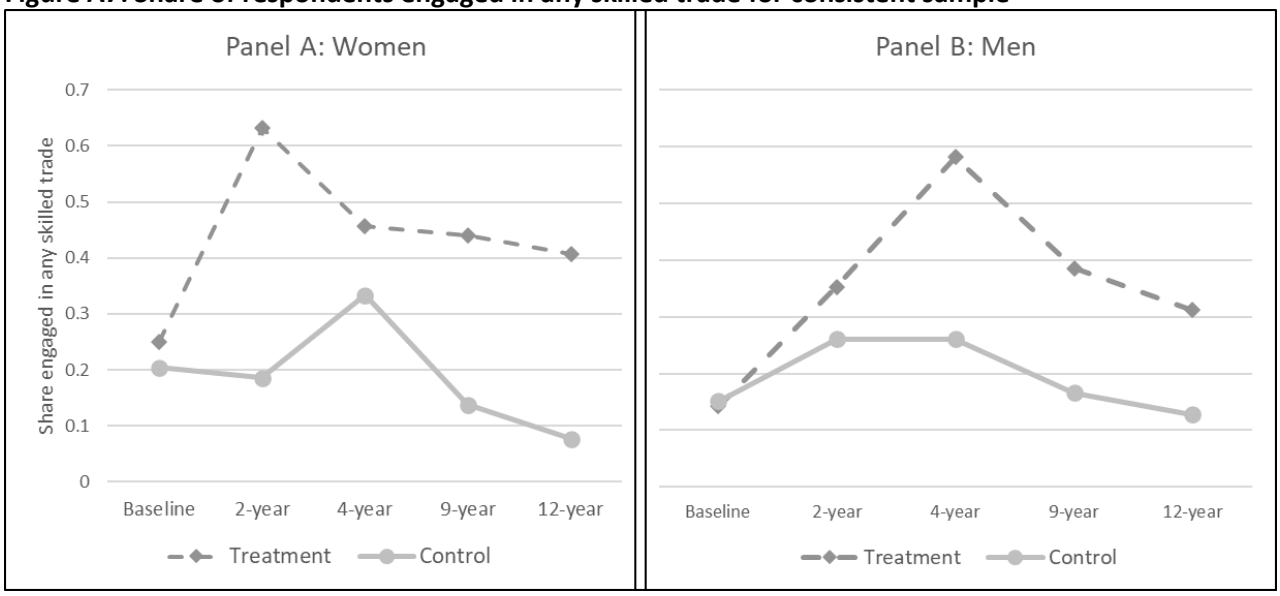
Notes: Income is in thousands of 2008 Ugandan shillings using the 2008 exchange rate of 1,720 shillings to $1.

Figure A6: Progression of earnings over time for a consistent sample



Notes: The consistent sample includes only participants that were successfully interviewed in all five waves. It includes 867 participants.

Figure A7: Share of respondents engaged in any skilled trade for consistent sample



Notes: The consistent sample includes only participants that were successfully interviewed in all five waves. It includes 122 participants.

References Online Appendix

Karlan, D., & Valdivia, M. (2011). Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Review of Economics and Statistics, 93*(2), 510-527.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, *76*(3), 1071–1102.

Museveni, Y. (2020). *Address on the Corona virus (COVID 19): Guidelines on avoiding the pandemic.* Retrieved from: https://www.yowerikmuseveni.com/address-corona-virus-covid-19-guidelines-avoiding-pandemic