Journal of Development Economics 155 (2022) 102781

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular Article

Unpacking a multi-faceted program to build sustainable income for the very poor☆

Abhijit Banerjee [a], Dean Karlan [b,*], Robert Osei [c], Hannah Trachtman [d], Christopher Udry [e]

[a] MIT, CEPR, NBER, Jameel Poverty Action Lab (J-PAL), United states
[b] Northwestern University, CEPR, NBER, IPA, J-PAL, United states
[c] University of Ghana, Legon, Ghana
[d] Hebrew University of Jerusalem, Israel
[e] Northwestern University, CEPR, NBER, J-PAL, United states

ARTICLE INFO

JEL classification:
O12
I38

Keywords:
Poverty alleviation
Returns to capital

ABSTRACT

A multi-faceted program comprising a grant of productive assets, training, unconditional cash transfers, coaching, and savings has been found to build sustainable income for those in extreme poverty. We focus on two important questions: whether a mere grant of productive assets would generate similar impacts (it does not), and whether access to a savings account with a deposit collection service would generate similar impacts (it does, but they are short-lived).

1. Introduction

One of the most exciting ideas in the fight against extreme poverty is the discovery that a focused multi-faceted intervention can durably unleash productive human potential, even in circumstances of severe economic hardship. Banerjee et al. (2015) and Bandiera et al. (2017) present impact results from seven countries for a multi-faceted "graduation" program that includes at its core a transfer of productive assets, between one and two years of training and coaching, weekly or monthly cash transfers for consumption support, and access to a saving account. This program successfully increased net worth, income and consumption three years after the productive assets were transferred, and in the two sites where long-term analysis is complete, impacts persisted (and indeed grew) after seven years (Banerjee, Duflo, and Sharma forthcoming; Bandiera et al., 2017; Balboni et al., 2021) and then remained at a similar level until the tenth year (Banerjee et al., forthcoming). Based on this evidence, many governments are implementing this program, often alongside further research to learn what model works best given their context and implementation capabilities.[1]

A better understanding of the underlying mechanisms through which the program works is critical, both for answering key theoretical questions about poverty traps and also for determining the ideal design for social protection programs. Here we explore further results from the Ghana site of Banerjee et al. (2015): we test whether two of the components, transfer of a productive asset and access to savings, are, by themselves enough to generate impacts comparable to the full package. The first intervention—the pure asset transfer—investigates whether lack of wealth is all that stops the poor, which would of course vastly

☆ Approval from the Yale University Human Subjects Committee, IRB #0705002656, 1002006308, 1006007026, and 1011007628; and from the Innovations for Poverty Action Human Subjects Committee, IRB Protocol #19.08 January-002, 09 December-003, 59.10 June-002, and 10 November-003.494. Thanks to the Ford Foundation, and 3ie for funding. Thanks to Nathan Barker, Caton Brewster, Abubakari Bukari, David Bullon Patton, Sébastien Fontenay, Angela Garcia, Yann Guy, Samantha Horn, Sana Khan, Hideto Koizumi, Matthew Lowes, Elizabeth Naah, Michael Polansky, Elana Safran, Sneha Stephen, Rachel Strohm, and Stefan Vedder for outstanding research assistance and project management, and in particular Bram Thuysbaert for collaboration. The authors would like to thank the leadership and staff at Presbyterian Agricultural Services (PAS) for their partnership. Thanks to Frank DeGiovanni of the Ford Foundation, Syed Hashemi of BRAC University, and Aude de Montesquiou and Alexia Latortue of CGAP for their support and encouragement of the research. No authors have any real or apparent conflicts of interest, except Karlan is on the Board of Directors of Innovations for Poverty Action, which participated in oversight of the implementation. All data and code are available at the IPA Dataverse (doi pending).

* Corresponding author.
E-mail address: dean.karlan@gmail.com (D. Karlan).

[1] This list includes Afghanistan, Brazil, Burkina Faso, Chad, Colombia, India, Indonesia, Kenya, Lebanon, Mali, Mauritania, Mozambique, Niger, Pakistan, Paraguay, Philippines, and Senegal.

https://doi.org/10.1016/j.jdeveco.2021.102781
Received 10 April 2020; Received in revised form 21 September 2021; Accepted 4 November 2021
Available online 8 December 2021
0304-3878/

simplify anti-poverty policy. The second—improved access to savings—examines whether the expensive wealth transfers are necessary or whether a good savings technology could suffice to help households accumulate their own wealth. Together, these two constitute obvious benchmarks against which the graduation program ought to be compared.

1.1. Background

The interest in multi-faceted approaches comes from the rather weak evidence of long-term impact on earnings from a number of well-thought of interventions, including microcredit, entrepreneurship training, cash transfers and savings promotion. The multi-faceted "graduation" program, is effectively an amalgam of these. Interestingly, given the often (but not always) discouraging track record of the individual interventions, the program combining them does yield consistent and positive long-term results. In six out of seven evaluated sites, the program generated economically meaningful, cost effective, and sustained positive average impacts on earnings, consumption and other welfare measures over at least three years. Moreover, the trajectories of the beneficiaries continue to diverge from that of the control group in the two places, Bangladesh and India, where there are data from a seven-year and a ten-year follow up.

BRAC, the organization that was instrumental in developing this program, has always argued that there are complementarities between the program's pieces. The weekly or monthly cash transfers are argued to help the families get through the initial setup phase for their business without feeling the pressure to sell or consume the asset, while the training and the hand-holding is argued to help them not make elementary mistakes and stay motivated during the same period. The savings accounts then help households save their earnings, and convert savings into future lumpy investments for the household or business.

However, while the complementarity argument is plausible based on the above evidence, it could also be that the locations where capital grants and business training were tested in the past were less conducive for the success of the program than the locations where the graduation program was implemented. Or it could be that the fact that the graduation programs deliberately target the poorest of the poor is key. Other programs are often more inclusive of a wider set of poor households. It therefore remains possible that the individual components would work by themselves if they were similarly targeted.

1.2. What we do here

We examine whether, for the population targeted by the graduation program, it is possible to get similar results with just one of the main components of the program. We use two additional experimental arms from the Ghana site of Banerjee et al. (2015) to examine whether the savings component alone or the grant of goats alone (the most common asset transferred in the graduation program) generate long-term improvements in income and consumption comparable to the graduation program in the same population.

The savings-only program has statistically significant positive effects on financial inclusion and consumption at two years, but both effects are much weaker by the three-year mark. The asset-only treatment has no evidence of any positive welfare effects after either two years or three years. These are important when contrasted with the full graduation program, which at the three-year mark yielded statistically significant positive effects on all five of our indicators.

We then work to unpack these differences. We start by examining some of the mechanisms associated with changes in the full graduation treatment. We find that the graduation program's strong positive effect on income is driven by increased business income, crop income, and animal revenue, and the positive effect on assets is driven almost entirely by livestock. Furthermore, using the experimental variation between a full graduation program with the savings component and one without, we find that even graduation households without the savings component are saving considerably more than control households.

Next we turn to our detailed savings data in order to understand why participants in the savings-only intervention were not able to save to accumulate assets or start similarly profitable businesses. We show that the graduation program with the savings component is much more successful than the savings-only program in generating savings, even when the savings-only program had a 50% match rate (an additional experimental treatment arm). Perhaps this is saying that people need earnings in order to save, or that the coaching and handholding was critical for ensuring that the savings turn into investments. In sum, the savings-only component did not appear to generate savings that would enable households to start profitable businesses, or to generate persistent effects on a financial inclusion index.

We then ask why the households who only received assets do less well than graduation households in terms of accumulating assets or starting profitable businesses. Although asset-only households do own more goats than control households after both two and three years, they own fewer goats than graduation households, suggesting that they were unable to hold onto or breed their goats the way households in the graduation program did. Moreover, they own no more total livestock than control households, implying that they were more likely to get rid of other livestock. The evidence suggests that the additional training and consumption support enabled graduation households (perhaps through a capabilities effect) to accumulate more goats while keeping other livestock as well, ultimately making them more successful in building businesses that persistently generate income. Using consumption as the final, primary outcome measure for which to calculate benefits, the full program yields a 1.2x benefit-cost ratio, whereas we cannot reject the null hypotheses that benefit-cost ratio for the asset-only and the savings-only treatment arms is zero.

2. The graduation program and experimental methods

2.1. The graduation program

For the multi-faceted program in Ghana, Graduating from Ultra Poverty ("GUP"), implementers first identified poor communities in poor regions of the country. In each identified community, staff members then facilitated a Participatory Wealth Ranking (PWR), in which members of the community worked together to rank households by economic status. Finally, staff members returned for a verification of the households judged to be the poorest. The program was implemented by Presbyterian Agricultural Services, a local nongovernmental organization, in coordination with Innovations for Poverty Action, a non-profit research organization.

The basic GUP program involved six key components, delivered over two years from 2011 to 2013 via regular visits (typically weekly) by a field officer from the implementing organization (see Appendix Table 1 and Banerjee et al. (2015) for more details). The first component was a transfer of a productive asset. Households were permitted to choose a package of assets from a set list, which included combinations of goats, hens, pigs, maize inputs, shea nut inputs, paddy rice inputs, and sorghum inputs. The second component was skills training for the management of the asset, delivered by a Field Agent over the duration of the program. The third component was a weekly cash stipend for consumption support, worth between $6 and $9 PPP depending on family size, lasting for the duration of each lean season and extending to 14 months. The fourth component was access to a savings account at a local bank, and an option to make deposits with a Field Agent, who visited the household weekly (we provide more details below in the Experimental Methods section, asthis is one of the components unpacked.) The fifth component was some basic health services and health education. The sixth and final component was the regular visits themselves, which included encouragement and life coaching.

2.2. Unpacking mechanisms design

Beyond the full graduation program, the experiment included four additional experimental arms designed to unpack whether specific components were sufficient on their own, and included randomization at both the village and household level.

We implemented two additional treatment arms at the village level: "Asset-Only" and "Saving Out of Ultra Poverty" ("SOUP"). For each, a two-level design was maintained, thus creating treatment households in treatment villages, control households in treatment villages, and control households in control villages. We also implemented two additional sub-treatment arms at the household level within GUP and SOUP villages. Some GUP households in GUP villages were assigned to "GUP without savings," and some SOUP households in SOUP villages were assigned to "SOUP with match." Appendix Table 2 presents the experimental arms and sample sizes for each arm, and Appendix Table 3 presents more details on each component.

In Asset-Only villages, 50% of sample households were assigned to treatment, and received *only* a productive asset, without skills training on how to use it, or any of the other GUP components. These households were simply given four goats, since this was the most popular asset in GUP (71% of households chose a package of assets that included four goats). Goats were chosen because most households chose goats in the full program, and because most households either have had or currently have goats. We wanted an asset where households could succeed with little technical training, and one that was unlikely to be turned down by households due to lack of familiarity or experience.

In SOUP villages, 59% of sample households were assigned to the SOUP treatment, and received a visit from the field agent to collect savings. This treatment group is akin to the standard GUP (with savings) group, but without any other components of the graduation program.

In GUP villages, we introduced a slightly reduced version of the full graduation program, a "GUP without savings" treatment arm, to 50% of treatment households. The other 50% received the full graduation program, "GUP with savings," which included the collection of savings for deposit into a local bank by the field agent just as in SOUP.

Finally, in SOUP villages, we introduced a matched savings sub-treatment. Of the households assigned to treatment, half received savings accounts and deposit collection without a match ("SOUP without match") and half received savings accounts and deposit collection with a 50% match ("SOUP with match"). Specifically, for every GHC 1 deposited, households in this group received a matching contribution of GHC 0.50.[2] The remaining households in SOUP villages were assigned to the SOUP control group.

2.3. Data collection

We conducted household surveys at baseline, two years after the assets were transferred and training conducted (and shortly after the end of the household visits), and three years after. While the majority of the intervention took place in the first month of the program (the technical training and the productive asset transfer), the household visits and savings collection lasted almost two years. We conducted three additional short midline surveys after six months, one year, and one and a half years after the asset transfers; we include the latter two in our two-year analysis. We do not have a baseline survey for the asset-only treatment arm because at the time of starting the project it was not clear we had the funding to implement that arm. Those villages were included in the village level randomization, so as to preserve the option for including the treatment arm, but we did not conduct household-level baseline surveys.

Most measures were collected during the aforementioned household surveys from the primary respondent in the household (typically the female head). The health, mental health, political, time use, and gender measures were collected in a separate "adult" survey, typically administered to the same primary respondent but focusing on individual members within the household. Respondents were asked about the health of all household members, but only about his or her own mental health, political involvement, time use, and gender norms. We pool all of the data that we have for each indicator, which explains much of the variation for the number of observations across regressions. See Appendix Table 4 for attrition and the number of observations by survey round.

2.4. Integrity of the experiment design

Appendix Table 5 provides descriptive statistics for key baseline indicators across treatment arms. Although no systematic pattern emerges, we reject the joint null hypothesis of orthogonality for three out of 14 variables. In analysis, we will show results with and without controls for baseline variables, and in each primary analysis table we report the p-value for the difference at baseline for each outcome variable (labelled "bsl p-value").

2.5. Analysis methods

We estimate OLS regressions of outcomes on treatments, where the omitted group is all control households, including the control households in treatment villages.[3] We include village-level fixed effects for all villages except pure control villages (constrained to have a common intercept),[4] and fixed effects for participation in each survey wave. We cluster standard errors at the unit of randomization: village-level for households in pure control villages, and household-level otherwise.

For regressions that do not involve the Asset-Only treatment group, we include additional controls for the outcome at baseline and the baseline variables that we used for stratification via a re-randomization procedure.[5] For regressions that include the Asset-Only treatment households (households for which we did not collect baseline data), we also estimate specifications with controls for three key endline variables that we assert are highly unlikely to have changed as a treatment effect from GUP or SOUP: average household age, household size, and whether

[2] At the onset of the program, there was a maximum match of GHC 1.50 GHC per week (for a GHC 3 deposit) but this cap was eventually removed.

[3] In Appendix Table 6 we exclude control households in treatment villages, and show our main three-year results (those shown in Table 1) relative to an omitted group of pure control households only. We still find statistically significant (though smaller) effects of GUP on asset value, financial inclusion, and income. We still find no evidence of effects of SOUP or Asset-Only.

[4] In pure control villages, all households are untreated, so we do not have variation in treatment/control within those villages, and thus cannot estimate separate intercepts for each village.

[5] The vector of baseline controls used in re-randomization Z_i^k includes: household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, the total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village.

or not the house has a metal roof.[6]

Finally, we control for the treatment status of a separate but related study (Banerjee et al., 2020) in which we created a cross-cutting, short-term intervention to make bags from traditional Ghanaian cloth (not implemented in the Asset-Only treatment villages).[7] In Appendix Table 7, we show that our results are robust to the exclusion of the households that received the "bags" program (but are estimated less precisely).

As mentioned above, there were three midline surveys administered to a fixed random subset of households, a survey administered to all households at two years (the end of the program), and a survey administered to all households at three years (a year later). We typically either report "two-year," "three-year," or "pooled" outcomes, as indicated in each table. Importantly, our two-year outcomes are an average of the outcome measured at two years and the outcomes measured in the two midline surveys administered within the 12 months prior to the two-year survey.

The most common specification is as follows:

$$Y_{it}^{k}=\alpha+\beta T_i+\gamma Z_i^k+W_i^{strat}+V_i^{short\ survey}+\theta_i^{village}+\mu_i^{emp}+\varepsilon_{it}$$

where Y_{it}^k is outcome k for household i at time t (where t is either two years or three years), T_i is a treatment dummy, Z_i^k is the baseline value of outcome k for household i, W_i^{strat} is a vector of controls that consists of the variables we used for re-randomization, $V_i^{short\ survey}$ is a vector of dummies for whether or not the household was surveyed in each of the three midlines, $\theta_i^{village}$ is a vector of village-level fixed effects, and μ_i^{emp} is a vector of controls for the employment (bags) program treatment arms. The village-level fixed effects $\theta_i^{village}$ are included for all villages with the exception of pure control villages.

We use the Benjamini and Hochberg (1995) step-up method and procedures put forward in Anderson (2008) to compute q-values that correct for the multiple hypotheses within each table (and sometimes within panels). We do not extend these corrections beyond the boundary of an individual table (or panel) because the substantive aspects of the hypotheses we test change dramatically across tables.[8] We organize the results by theoretically related hypotheses, which is reflected in the way our tables (panels) are structured.

3. Results

3.1. Impacts

Table 1 presents estimates of treatment effects on five indices that capture economic wellbeing three years after the productive asset transfer (i.e., three years after the start of the program, and one year after the end of the household visits). These indices are standardized with respect to baseline values; the components are listed in Appendix: Variable Definitions and Construction. Appendix Table 9 presents estimates of the same outcomes two years after the productive asset transfer.

At two years, GUP without savings shows statistically and economically significant effects on asset value, consumption, food security, and income; at three years, all of these effects persist and an effect on financial inclusion emerges as well. In Appendix Table 10 we show that the effects on financial inclusion is driven by an increase in self-reported savings balances. Thus, even GUP households without deposit collection services manage to save more than control households. GUP with savings shows statistically significant short-run effects on financial inclusion and income, both of which persist a year later, at which time an effect on asset value also emerges. In summary, with or without savings, GUP has long-run effects on income, assets, and financial inclusion; and without savings, the long-run consumption effect is significantly positive as well. The point estimate for the consumption index is 0.12 for GUP without savings and 0.05 for GUP with savings (p-value on difference across coefficients is 0.07). The lower impact on consumption could be a by-product of consumption being diverted into savings, for future consumption, durables or investment, but being borderline statistically significant we do not emphasize this comparison.

SOUP has a positive effect on consumption and financial inclusion at two-years; at three-years, the consumption effect disappears and the financial inclusion effect shrinks (and is no longer statistically significant once we account for multiple hypotheses). The positive two-year effects seem driven by higher savings balances. Appendix Table 10 shows that at two years SOUP participants have more than three times the savings balances as control participants; at three years the effect is smaller, with balances less than double those of control. Thus while SOUP does have important short-run impacts, they do not persist after the intervention and deposit-collecting visits to households end, and in the long-run we observe no substantial changes in household welfare.[9]

Critically, the Asset-Only treatment effects at both two years (Appendix Table 9, Panel B) and three years (Table 1, Panel B) are null for all five indices of economic wellbeing. We discuss below potential mechanisms behind this null effect.

We find only a few effects of GUP, SOUP and the Asset-Only treatment arms on secondary outcomes (physical health, mental health, political involvement, labor supply, and female empowerment), which we report in Appendix Tables 13 and 14. After two years, there are only four effects that come close to surviving multiple hypothesis correction: GUP with savings on political involvement, GUP without savings on mental health, Asset-Only on mental health (negative), and Asset-Only on time working. None of these effects persist until the year three measurement—indeed, the effect of Asset-Only on mental health appears to turn positive (but is not statistically significant after adjusting for multiple hypotheses). Overall, there is no evidence that these downstream impacts sustained to three years from any of the individual treatments (although note that in Banerjee et al. (2015), which uses data from multiple sites, downstream results from the full graduation program do persist at three years).

3.2. Unpacking the effects

3.2.1. Unpacking sources of income

With or without the savings component, GUP has persistent effects on income: the effect of having any GUP treatment is 0.223 standard deviations (se = 0.063), and the effects are similar for both the GUP with

[6] At the time of the two-year survey the Asset-Only households are 18.5% smaller than the control households (shown in Appendix Fig. 1). Unfortunately, because the Asset-Only treatment was decided upon after the baseline was completed (due to logistics), we have no baseline measure of household size for Asset-Only households. We do however look at how GUP and SOUP affect household size, and find no evidence of a change for SOUP households but small, statistically significant increase for GUP households (Appendix Table 8). This fits with our expectations: these households are richer and probably need more labor, hence growth is plausible. Based on this, we would expect the treatment effect of Asset-Only on household size to also be positive though perhaps smaller. We therefore infer that the negative household size difference in the asset-only group between treatment and control is a pre-existing difference and not a treatment effect, and therefore control for it in our regressions.

[7] The vector of controls for the employment program treatment arms μ_i^{emp} includes dummies for simple bag assignment and complex bag assignment, as well as their interactions with assignment high unconditional consumption support.

[8] The boundaries of a set of tests over which one might correct for multiple hypotheses is arbitrary unless one takes a full Bayesian approach.

[9] In Appendix Tables 11–12, we report two- and three-year estimates, respectively, of differences between SOUP and SOUP match, and GUP and GUP with savings.

Table 1
Three-year effects of GUP, SOUP, and asset only on household-level economic indices.

		(1) Asset Value Index	(2) Consumption Index	(3) Financial Inclusion Index	(4) Food Security Index	(5) Income Index
PANEL A: GUP vs. SOUP						
SOUP	ITT	0.029	−0.013	0.129	0.002	−0.071
	SE	(0.076)	(0.034)	(0.073)	(0.044)	(0.062)
	p-val	0.701	0.700	0.078*	0.962	0.254
	q-val	0.825	0.825	0.156	0.963	0.373
	Bsl p-val	0.051*	0.993	0.555	0.098*(+)	0.277
GUP no sav.	ITT	0.280	0.124	0.204	0.114	0.202
	SE	(0.078)	(0.046)	(0.086)	(0.050)	(0.073)
	p-val	0.000***	0.007***	0.018**	0.024**	0.006***
	q-val	0.003***	0.022**	0.051*	0.059*	0.022**
	Bsl p-val	0.741	0.022**	0.014**	0.282	0.205
GUP sav.	ITT	0.318	0.050	0.532	0.092	0.243
	SE	(0.082)	(0.036)	(0.105)	(0.050)	(0.076)
	p-val	0.000***	0.169	0.000***	0.062*	0.001***
	q-val	0.002***	0.282	0.001***	0.139	0.008***
	Bsl p-val	0.592	0.100	0.632	0.704	0.794
GUP sav. - SOUP	Diff	0.289	0.063	0.402	0.090	0.314
	SE	(0.105)	(0.046)	(0.122)	(0.063)	(0.094)
	p-val	0.006***	0.173	0.001***	0.150	0.001***
	bsl p-val	0.240	0.176	0.426	0.138	0.353
any GUP	ITT	0.299	0.088	0.366	0.103	0.223
	SE	(0.068)	(0.036)	(0.077)	(0.043)	(0.063)
	p-val	0.000***	0.015**	0.000***	0.017**	0.000***
	Bsl p-val	0.624	0.025**	0.103	0.390	0.378
	Obs	3781	3597	3603	3603	3781
PANEL B: GUP vs. Asset Only						
asset	ITT	−0.022	−0.009	0.050	−0.079	−0.133
	SE	(0.103)	(0.055)	(0.073)	(0.070)	(0.085)
	p-val	0.832	0.867	0.490	0.261	0.119
	q-val	0.913	0.913	0.654	0.373	0.218
	ITT, ctrls	−0.043	−0.006	0.029	−0.075	−0.148
	p-val, ctrls	0.684	0.909	0.692	0.283	0.080*
	q-val, ctrls	0.866	0.910	0.866	0.708	0.399
GUP no sav. - asset	Diff	0.325	0.114	0.154	0.188	0.345
	SE	(0.135)	(0.073)	(0.113)	(0.086)	(0.114)
	p-val	0.016**	0.116	0.172	0.029**	0.002***
	ITT, ctrls	0.288	0.123	0.160	0.178	0.319
	p-val, ctrls	0.032**	0.091*	0.153	0.039**	0.004***
	Obs	4102	3883	3893	3893	4102

Estimates from OLS regressions of household-level economic indices at year three on treatments. The omitted group is control households in all villages. The regression in Panel A excludes the Asset-Only villages and includes controls for re-randomization variables and the baseline value of the outcome. The regression in Panel B includes the Asset-Only villages (without baseline controls) and reports only the coefficient on Asset-Only. At the bottom of each panel we report linear combinations of interest. Both panels include controls for employment program treatments. We include fixed effects for all villages except those assigned to pure control, and indicators for whether or not the household was surveyed in each midline. We cluster standard errors at the unit of randomization (village for pure control, individual otherwise). We use the Benjamini-Hochberg step-up method to compute q-values, considering the 20 independent hypotheses in the table. For regressions that include asset households, we also report p-values and q-values for a specification with three two-year variables as controls, since we have no baseline controls. Finally, we report p-values for the same specification using the baseline value of each outcome. We use a superscript (+) to indicate a positive t-statistic. Indices are centered around their baseline values. For detailed descriptions of index construction, see Appendix: Variable Definition and Construction.

savings and GUP without savings treatment arms. Table 2 examines the source of these three-year effects.[10] It appears that the GUP program boosted income from all three of the activities that are most profitable among control households: crops, businesses, and animals, although only animal revenue effects survive multiple hypothesis correction. Table 2, Column 1 suggests that for GUP with savings, this higher income may be driven in part by the creation of new businesses (though again this effect does not survive the multiple hypothesis correction). GUP households, irrespective of the inclusion of the savings treatment, were seemingly able to build or grow businesses, improve the profitability of their farms, and generate revenues from livestock as a result of the program (Columns 1–4, the differences between the GUP no savings and GUP with savings coefficients are small, and the estimates for "Any GUP" are all statistically significant).

Why did income rise for GUP households and not for households in the SOUP or Asset-Only treatments? In Section 4.2.2 we take a closer look at SOUP households, and in Section 4.2.3 we turn to Asset-Only households.

3.2.2. *Unpacking the savings process, using transaction data*

In Fig. 1 we look at the weekly data from our savings collectors, which is, by its very nature, restricted to treatments where there was a savings intervention. We therefore use the pure savings treatment (SOUP no match) as the comparison group. The average SOUP no match household deposited $1 in a week on average; this effect rose 9% in the presence of a match, and more than doubled in the presence of GUP. GUP savings participants save much more during the lean season, which could be because they received consumption support during this time (the savings collector was also the individual responsible for bringing

[10] Appendix Table 15 reports the corresponding two-year results.

Table 2
Three-year effects of GUP, SOUP, and asset only on income sources.

		(1) Household has Business	(2) Business Income, Monthly (USD)	(3) Crop Income, Monthly (USD)	(4) Animal Revenue, Monthly (USD)	(5) Wage Income, Monthly (USD)
PANEL A: GUP vs. SOUP						
SOUP	ITT	−0.023	−1.716	−2.281	0.274	−0.384
	SE	(0.028)	(1.447)	(2.762)	(0.873)	(0.420)
	p-val	0.419	0.236	0.409	0.753	0.361
	q-val	0.559	0.429	0.559	0.772	0.559
	Bsl p-val	0.010***	0.192	0.751	.	0.544
GUP no sav.	ITT	0.051	2.840	5.263	2.873	0.180
	SE	(0.035)	(1.863)	(3.083)	(1.096)	(0.468)
	p-val	0.153	0.128	0.088*	0.009***	0.700
	q-val	0.338	0.320	0.252	0.089*	0.772
	Bsl p-val	0.434	0.054*	0.879	.	0.021**
GUP sav.	ITT	0.077	3.426	6.182	3.734	0.227
	SE	(0.034)	(1.789)	(3.144)	(1.062)	(0.534)
	p-val	0.026**	0.056*	0.049**	0.000***	0.671
	q-val	0.171	0.223	0.223	0.009***	0.772
	Bsl p-val	0.190	0.853	0.894	.	0.189
GUP sav. - SOUP	ITT	0.100	5.143	8.463	3.459	0.611
	SE	(0.042)	(2.168)	(4.012)	(1.306)	(0.635)
	p-val	0.017**	0.018**	0.035**	0.008***	0.336
	Bsl p-val	0.007***(+)	0.550	0.739	.	0.125
any GUP	ITT	0.064	3.131	5.730	3.309	0.203
	SE	(0.030)	(1.536)	(2.628)	(0.902)	(0.430)
	p-val	0.035**	0.042**	0.029**	0.000***	0.637
	Bsl p-val	0.103	0.359	0.988	.	0.038**
	Ctrl Mean	0.27	6.86	35.23	7.54	1.75
	Ctrl SD	0.44	20.10	45.33	14.62	6.76
	Obs	3605	3604	3698	3781	3604
PANEL B: GUP vs. Asset Only						
asset	ITT	−0.091	−1.978	−4.586	−0.411	−0.349
	SE	(0.051)	(2.197)	(3.331)	(1.414)	(0.906)
	p-val	0.071*	0.368	0.169	0.772	0.701
	q-val	0.237	0.559	0.338	0.772	0.772
	ITT, ctrls	−0.097	−2.349	−4.950	−0.651	−0.272
	p-val, ctrls	0.056*	0.289	0.130	0.646	0.762
	q-val, ctrls	0.282	0.483	0.325	0.762	0.762
GUP no sav. - asset	Diff	0.141	4.847	10.486	3.608	0.471
	SE	(0.062)	(2.881)	(4.588)	(1.821)	(1.021)
	p-val	0.024**	0.093*	0.022**	0.048**	0.644
	ITT, ctrls	0.143	5.043	9.301	3.470	0.371
	p-val, ctrls	0.022**	0.082*	0.040**	0.058*	0.714
	Ctrl Mean	0.27	6.86	35.23	7.54	1.75
	Ctrl SD	0.44	20.10	45.33	14.62	6.76
	Obs	3896	3895	3999	4102	3895

Estimates from OLS regressions of income sources from year three on treatments. The omitted group is control households in all villages. The regression in Panel A excludes the Asset-Only villages and includes controls for re-randomization variables and the baseline value of the outcome. The regression in Panel B includes the Asset-Only villages (without baseline controls) and reports only the coefficient on Asset-Only. At the bottom of each panel we report linear combinations of interest. Both panels include controls for employment program treatments. We include fixed effects for all villages except those assigned to pure control, and indicators for whether or not the household was surveyed in each midline. We cluster standard errors at the unit of randomization (village for pure control, individual otherwise). We use the Benjamini-Hochberg step-up method to compute q-values, considering the 20 independent hypotheses in the table. For regressions that include asset households, we also report p-values and q-values for a specification with three two-year variables as controls, since we have no baseline controls. Finally, we report p-values for the same specification using the baseline value of each outcome. We use a superscript (+) to indicate a positive t-statistic. For detailed descriptions of variables, see Appendix: Variable Definition and Construction.

them the cash they received as consumption support, so they could immediately save the cash if they wished).

In Appendix Table 10, Columns 1–3, we again look at the impact of the program over the long run using the deposits data, again using SOUP no match as the comparison group. In Column 4, we look at self-reported savings balances from the two-year household survey, conducted between 1 and 3 months after the end of savings collection, and in Column 5 we look at the same outcome a year later. Here, we use control households as the comparison group (since we have these data for the full sample) in order to look at the effects of SOUP (match and no-match) and GUP (saving and no-savings) on savings balances. Households in GUP savings both deposit much more and take out much more than both the SOUP no-match recipients and the SOUP match recipients, and by the end of the program they have 88% more in the "bank" than either group. The match has no additional effect on balances, a fact that is consistent with the self-reported data (Appendix Tables 11 and 12). The fourth column of Appendix Table 10 also confirms that the GUP no-savings intervention approximately doubles balances relative to the

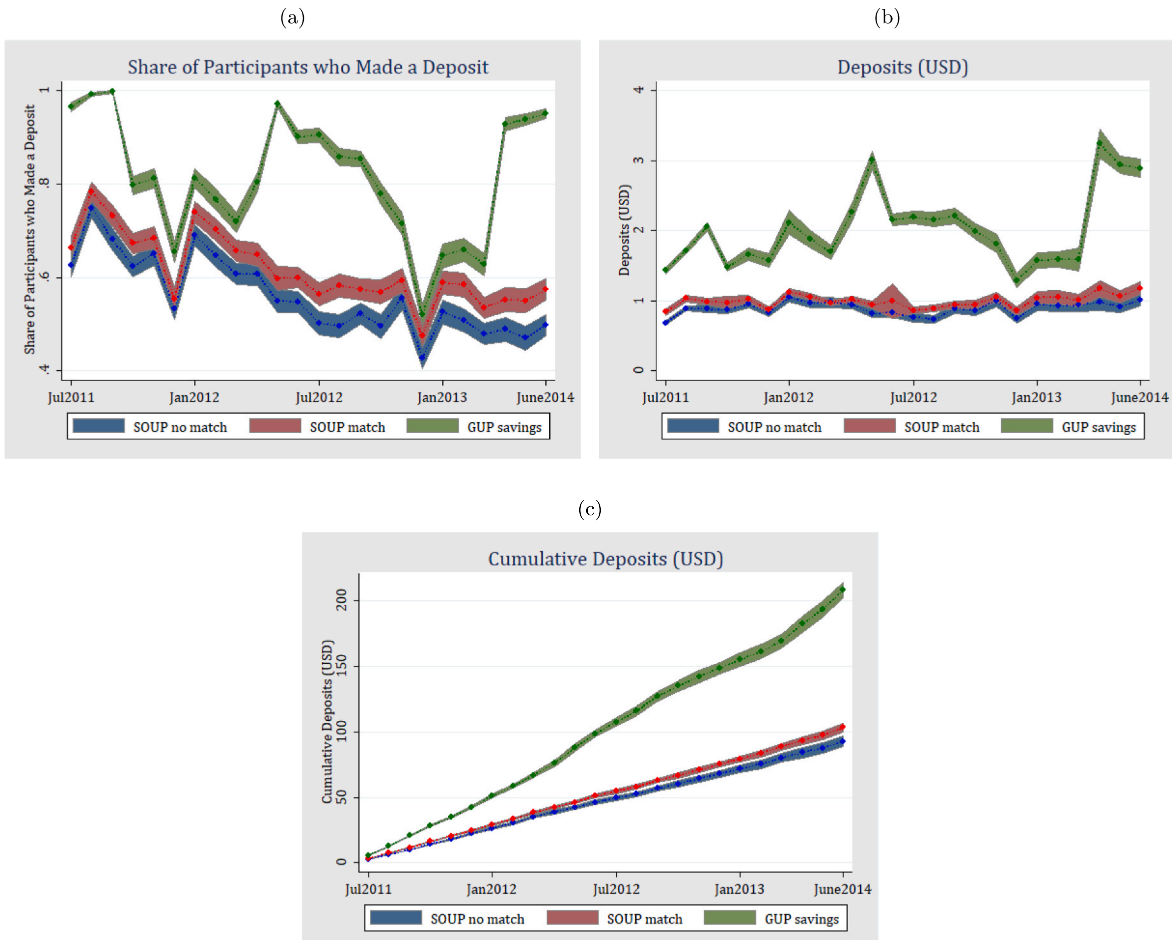

Fig. 1. Data on monthly deposits for treatment groups with the savings component. Panel (a) shows the share of participants who made a deposit, Panel (b) shows the flow of deposits (USD), and Panel (c) shows cumulative deposits (USD).

control group, the SOUP treatments triple it, and GUP savings raises it more than fivefold.[11] At three years, the treatment effect for GUP no-savings has remained the same (double the control group), and the other treatment arms still generate positive effects but are smaller in magnitude.

The main takeaway seems to be that the availability of savings collectors matters a lot, but the rate of return on savings less so. There also seems to be an income effect—GUP by itself almost doubles savings, even in the absence of savings collectors. There is also an interaction effect between income and savings collection services—at two years GUP savings households save \$12.9 more than the sum of the independent treatments of GUP no-savings and SOUP no-match, a difference that is statistically significant at the 1% level ($p = 0.003$).

3.2.3. *Unpacking the livestock effect*

In Table 3 we compare GUP no-savings with the Asset-Only treatment to pinpoint the differences in asset accumulation that they generate. The main difference between the two treatments was the combination of handholding and consumption support, both of which were intended to encourage the recipient to further invest in the asset rather than consume it. The handholding provided know-how on how to take care of the asset (such as when to vaccinate it, given that goats were the most commonly chosen asset by GUP households) and nudges to help the household to focus on building productive assets to generate positive change in long-term outcomes. The consumption support was explicitly intended to help this process in the short-run, by absorbing short-run shocks that could lead to households consuming the transferred assets.

The question of interest here is whether there are differences in the investment patterns. Column 1 shows that both treatments raise the value of goats owned by the household, though the effect of GUP is higher by \$34. This is despite the fact that, unlike the Asset-Only treatment, not at all GUP households had received goats—they were given a choice between several asset bundles that included goats, fowl, pigs, inputs for maize farming, inputs for rice farming, inputs for sorghum farming, and inputs to begin a shea-butter business. It seems that the GUP households were better at holding onto or growing their goats.[12] GUP households also accumulate more fowl, which makes sense since many of them chose an asset bundle that included fowl.

Asset-only households do not accumulate any other livestock apart

[11] The self-reported savings balance data do not match precisely with the transaction data, as demonstrated by the differences between columns Appendix Table 10 columns 3 and 4. Survey data were collected between one and three months after the end of the transaction data, thus some of the discrepancy could be due to withdrawals in that period; but undoubtedly is also due to accuracy of self-reported savings data. The difference is consistent across all three treatment groups for which we have transaction data.

[12] Appendix Table 16 examines the flow of goats between rounds conditional on owning goats in the current round, and finds that GUP households have more goat births and sales than Asset-Only. We cannot construct stock estimates from the flows, in part because we only collected flow data for households that owned at least one goat.

Table 3
Pooled two-year and three-year effects of GUP and asset only on household-level livestock values.

		(1) Goat Value (USD)	(2) Fowl Value (USD)	(3) Pig Value (USD)	(4) Sheep Value (USD)	(5) Cow Value (USD)	(6) Total Livestock Value (USD)
GUP no sav.	ITT	71.511	12.527	4.550	1.377	12.975	134.776
	SE	(7.984)	(3.722)	(2.002)	(9.006)	(12.953)	(27.636)
	p-val	0.000***	0.001***	0.023**	0.878	0.317	0.000***
	q-val	0.001***	0.003***	0.056*	0.879	0.423	0.001***
	ITT, ctrls	68.681	11.047	4.460	−2.649	10.330	120.363
	p-val, ctrls	0.000***	0.003***	0.026**	0.768	0.424	0.000***
	q-val, ctrls	0.001***	0.009***	0.063*	0.769	0.512	0.001***
asset	ITT	37.217	−3.221	1.861	−22.715	−19.320	−13.798
	SE	(10.230)	(4.708)	(1.806)	(13.920)	(15.536)	(38.521)
	p-val	0.000***	0.494	0.303	0.103	0.214	0.720
	q-val	0.002***	0.593	0.423	0.206	0.367	0.786
	ITT, ctrls	36.178	−3.799	2.372	−24.377	−18.285	−16.406
	p-val, ctrls	0.000***	0.426	0.230	0.083*	0.215	0.665
	q-val, ctrls	0.002***	0.512	0.345	0.167	0.345	0.726
GUP no sav. - asset	Diff	34.294	15.748	2.690	24.092	32.296	148.574
	SE	(12.978)	(6.007)	(2.705)	(16.596)	(20.199)	(47.433)
	p-val	0.008***	0.009***	0.320	0.147	0.110	0.002***
	ITT, ctrls	32.502	14.846	2.088	21.727	28.616	136.769
	p-val, ctrls	0.012**	0.014**	0.457	0.193	0.145	0.003***
	Ctrl Mean	80.0	47.8	3.5	68.0	38.1	263.9
	Ctrl SD	115.4	60.7	21.4	149.4	198.1	475.7
	Obs	8217	8222	8217	8217	8217	8222

Estimates from OLS regressions of livestock values on treatments. The omitted group is control households in all villages. We pool outcomes from the two-year (averaging over the two-year outcome and midline outcomes that were collected at least one year after treatment start) and three-year surveys. We include the Asset-Only villages (without baseline controls) and report only the coefficients on Asset-Only and GUP-no-savings. At the bottom we report linear combinations of interest. We control for employment program treatments. We include fixed effects for all villages except those assigned to pure control, indicators for the survey round (two-year or three-year), and indicators for whether or not the household was surveyed in each midline. We cluster standard errors at the unit of randomization (village for pure control, individual otherwise). We use the Benjamini-Hochberg step-up method to compute q-values, considering the 12 independent hypotheses in the table. We also report p-values and q-values for a specification with three two-year variables as controls (average age, metal roof, household size), since we have no baseline controls. For detailed descriptions of variables, see Appendix: Variable Definition and Construction.

from goats, and indeed appear to have reduced the number of sheep, though this effect does not survive the multiple hypothesis correction. The point estimate on cow value is negative as well. Ultimately, GUP without savings increases the total value of livestock by $149 more than the Asset-Only intervention without controls and by $137 with controls.

Thus, it seems that graduation households were able to use the additional training and consumption support to accumulate more goats while keeping other livestock as well. This explains why GUP produced sustained effects on assets and animal revenue, and may also have contributed to the rise in business income, by enabling households to undertake riskier projects and investments.[13]

4. Cost-benefit analysis

Our results thus far suggest that neither savings nor assets alone are sufficient to produce the kinds of persistent impacts on assets, income, and financial inclusion that the full graduation program generated. However, the graduation program cost $288 per capita versus about $40 for the SOUP and Asset-Only programs.

Table 4 examines the cost-effectiveness of the three programs. We take our point estimates from Appendix Tables 18, 19, and 20, which show treatment effects on the values of nondurable consumption and assets. We conduct our analysis at the per capita level, and not per household, to be consistent with Table 1 (which shows impacts on a consumption index based on consumption per capita). That said, we report both per capita and per household effects in Appendix Tables 18, 19, and 20 so that any differences can be taken into account. For the Asset-Only households we do not have data from the first year, so we run two versions of the analysis: one where we assume that year-one effects were the same as year-two effects, and another where we assume that year-one effects were equal to the value of the asset transferred. We then assume that three-year effects persist in perpetuity, assuming a 5% annual discount rate (a defendable assumption, given the evidence from elsewhere on the long-term persistence of the results).

We find a benefit-cost ratio of 1.21 for the graduation program. For the SOUP and Asset-Only programs (under the assumption that year-one effects were equal to the year-two effects), we interpret the benefit-cost ratios for SOUP and Asset-Only as effectively zero. Thus, even when the high costs of GUP are taken into account, the program is cost-effective relative to SOUP or Asset-Only. In the final row of Table 4 we relax the assumption that year three gains persist in perpetuity and compute the annual rate at which the treatment effect would need to dissipate in order for each program to break even. We find that the GUP effect would need to dissipate at 1.3% per year for the program to break even.

5. Discussion

While earlier work (Banerjee et al., 2015; Bandiera et al., 2017) found that a multi-faceted program was sufficient for generating economically meaningful and sustainable impacts for those in extreme poverty, the analysis did not establish whether the multi-faceted approach was necessary. Here we show that neither transferring a productive asset (in this case, goats) nor providing access to a savings account, on their own, generate similar economically meaningful and sustainable impacts in the same population.

Many questions remain regarding the underlying mechanisms of poverty traps. Our results do not ascertain whether complementarities were essential, or whether GUP's success was due to one of the remaining components of GUP (cash transfers, training, and coaching). Our results also cannot tell us *which* components or complementarities were indispensable. But given the importance of identifying programs

[13] Appendix Table 17 examines productive assets, household assets, and agricultural stocks separately, but finds no GUP treatment effect on productive assets. The impact on assets is driven entirely by livestock.

Table 4
Cost benefit analysis.

		GUP	SOUP	Asset only
COSTS PER CAPITA				
(1)	Program costs, calculated as if all incurred immediately at beginning of year 0	288	40	40
(2)	Program costs, USD PPP 2014, calculated as if all incurred immediately at beginning of year 0	631	260	88
(3)	Program costs, USD PPP 2014, inflated to year 3 at *5%* annual discount rate	731	301	102
BENEFITS PER CAPITA, USD PPP				
(5)	Year 1 annual nondurable consumption ITT (assuming treatment effect equal to	17	10	−8
(6)	year 2 for asset only)	(34)	(40)	(201)
(7)	Year 1 annual nondurable consumption ITT (assuming treatment effect equal to value of asset for asset only)			36
(8)	Year 2 annual nondurable consumption ITT treatment effect	53**	65**	−8
		(21)	(22)	(34)
(9)	Year 3 household asset ITT treatment effect	17**	−1	−1
		(5)	(5)	0)
(10)	Year 3 nondurable annual consumption ITT treatment effect	40**	−5	−1
		(16)	(15)	(24)
(11)	Year 4 onward total consumption ITT treatment effect, assuming year 3 gains persist in perpetuity (discount rate 5%)	760	−95	−19
(12)	Total benefits, assuming year 1 nondurable consumption effect equal to year 2 for asset only: (5) + (8) + (9) + (10) + (11)	887	−26	−37
(13)	Total benefits, assuming year 1 nondurable consumption effect equal to value of asset for asset only: (7) + (8) + (9) + (10) + (11)			7
BENEFIT/COST RATIOS				
(14)	Total benefits/total costs ratio, assuming year 1 nondurable consumption effect equal to year 2 for asset only: (12)/(3)	1.21	−0.09	−0.36
(15)	Total benefits/total costs ratio, assuming year 1 nondurable consumption effect equal to value of asset for asset only: (13)/(3)			0.07
(16)	Annual rate of dissipation of the treatment effect such that costs = benefits	1.3%	−7.3%	−5.8%

Estimates of benefits come from Appendix Tables 18, 19, and 20. Stars report significance of estimates according to the q-values from our Bonferonni procedure, taking into account all the hypotheses in each table. Note that the negative estimates driving benefit/cost ratios to be negative for SOUP and asset only are not statistically different from zero, and in fact, these negative estimates are not robust to a specification that uses per household measures of benefits, as evidenced by column 2 of Appendix Table 20. (This is because asset only households are slightly larger than asset spillover households, and our specification includes village fixed effects.) Since we don't have data from year 1 for asset only, we calculate benefits under two assumptions: (1) that year 1 consumption effects were equal to those of year 2; (2) that year 1 consumption effects were equal to the value of the asset. The annual rate of dissipation of the treatment effect such that costs equal benefits (16) is computed as in Banerjee et al. (2015), see SOM Text 5 for details.

simpler than GUP (i.e., ones with reduced implementation complexity and lower costs) for nationwide social protection policies, ruling out savings-only and asset-only interventions as drivers in this context is a critical finding.

Additional questions remain regarding the optimal policy for social protection at scale. For example, cash transfers are a natural alternative (because of lower transaction costs, lower probability of moving prices when implemented at scale, and higher flexibility the cash affords the recipient to choose their own investment). However cash transfers also have been shown to be less likely to be invested (Fafchamps et al., 2014). Lump-sum cash transfers do better than constant smaller streams of cash flow for encouraging investment (rather than immediate consumption), but still much of the funds get used for durable consumption goods, such as home improvements (Haushofer and Shapiro 2016). These may generate long-term benefits for households, but perhaps not higher long-term income. More research is needed to understand whether cash transfers implemented in other locations or alongside some form of behavioral intervention, e.g. a "nudge" in which individuals form a simple non-binding plan before receiving the cash, would lead to higher levels of investment and thus longer term impact on income.

The household visits serve multiple roles, including providing information and behavioral support. Implementing these at scale poses a real challenge, as they require a vast network of field agents who are both well informed about the range of productive assets that might be transferred to help households when problems arise, and also well versed in how to engage households in life coaching, to help build hope and encourage the aspirations of the households and guide them to stay on track with a long term plan of building productive assets. Some have suggested technological solutions to this problem, for example a mobile device that provides videos with information and mobile applications which facilitate communication between households and field agents (for example, that generate a regular stream of text messages at predefined or appropriately triggered times). Such a technology may make it easier to implement the program at scale without losing implementation fidelity, yet may put at risk the impact if direct human interaction is necessary.

On the other hand, perhaps rather than looking for components to shed, an even richer program would be more effective. Despite the success on average, not everyone benefits from the program. Those in extreme poverty suffer from high levels of depression (Sipsma et al., 2013). Perhaps those with poor mental health are not able to embrace the opportunity fully, and thus a mental health intervention that precedes the multi-faceted program would generate even bigger impacts. Among a highly selected population of youth engaged in street crime in Liberia, cognitive behavioral therapy in conjunction with cash has led to important positive economic changes a year later (Blattman et al. 2015). In Ghana, this is now being tested in a new sample frame of ultra-poor households similar to the population studied here.

Data availability

Data will be posted on the IPA & JPAL Dataverse

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jdeveco.2021.102781.

References

Anderson, Michael, 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the abecedarian, perry preschool, and early training projects. J. Am. Stat. Assoc. 103 (484), 1481–1495. https://doi.org/10.1198/016214508000000841.

Balboni, Clare, Bandiera, Oriana, Burgess, Robin, Ghatak, Maitreesh, Heil, Anton, 2021. Why do people stay poor? Q. J. Econ. https://www.dropbox.com/s/4tfuhclfvh6ynnd/povertyTraps.pdf?dl=0.

Bandiera, Oriana, Burgess, Robin, Das, Narayan, Gulesci, Selim, Rasul, Imran, Sulaiman, Munshi, 2017. Labor markets and poverty in village economies. Q. J. Econ. 132 (2), 811–870. https://doi.org/10.1093/qje/qjx003.

Banerjee, Abhijit, Duflo, Esther, Goldberg, Nathanael, Karlan, Dean, Osei, Robert, Parienté, William, Shapiro, Jeremy, Thuysbaert, Bram, Udry, Christopher, 2015. A multifaceted program causes lasting progress for the very poor: evidence from six countries. Science 348 (6236), 1260799. https://doi.org/10.1126/science.1260799.

Banerjee, Abhijit, Esther Duflo, and Garima Sharma. forthcoming. "Long-Term Effects of the Targeting the Ultra Poor Program." *Am. Econ. Rev.: Insights*. https://doi.org/10.1257/aeri.20200667.

Banerjee, Abhijit, Dean, Karlan, Trachtman, Hannah, Udry, Christopher, 2020. Does Poverty Change Labor Supply? Evidence from Multiple Income Effects and 115,579 Bags. Working Paper).
Benjamini, Yoav, Hochberg, Yosef, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 289–300.
Blattman, Christopher, Jamison, Julian, Sheridan, Margaret, 2015. Reducing Crime and Violence: Experimental Evidence on Adult Noncognitive Investments in Liberia.
Fafchamps, Marcel, McKenzie, David, Quinn, Simon, Woodruff, Christopher, 2014. Microenterprise growth and the flypaper effect: evidence from a randomized experiment in Ghana. J. Dev. Econ. 106 (January), 211–226. https://doi.org/10.1016/j.jdeveco.2013.09.010.
Haushofer, Johannes, Shapiro, Jeremy, 2016. The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. Q. J. Econ. 131 (4).
Sipsma, Heather, Ofori-Atta, Angela, Canavan, Maureen, Osei-Akoto, Isaac, Udry, Christopher, Bradley, Elizabeth H., 2013. Poor mental health in Ghana: who is at risk? BMC Publ. Health 13 (1), 288.