Graduation from ultra poverty in Ghana

Abhijit Banerjee, Massachusetts Institute of Technology

Dean Karlan, Yale University

Robert Osei, University of Ghana

Bram Thuysbaert, Ghent University

Christopher Udry, Yale University

Grantee Final Report

Accepted by 3ie: August 2017

Note to readers

This final impact evaluation grantee report has been submitted in partial fulfilment of the requirements of grant OW2.206 awarded under Open Window 2. 3ie is making it available to the public in this final report version as we received it. This report, and the two accompanying quantitative and qualitative reports associated with it, do not comply with 3ie's reporting standards. What has been submitted is technically sound for publication as a grantee final report. No further work has been done.

All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was submitted. Please direct all comments or queries to the corresponding author, Dean Karlan at dkarlan@yale.edu

3ie received funding for the Open Window from our donors, which include UK aid, the Bill & Melinda Gates Foundation and the William and Flora Hewlett Foundation. A complete listing of all of 3ie's donors is available on the 3ie website.

Suggested citation: Banerjee, A, Karlan, D, Osei, R, Thuysbaert, B and Udry, C. 2017. *Graduation from ultra poverty in Ghana*, *3ie Grantee Final Report.* New Delhi: International Initiative for Impact Evaluation (3ie)

Contents

Acknowledgements

Thanks to the Ford Foundation and 3ie for funding and for technical review and support throughout the study. Thanks to N. Barker, C. Brewster, A. Bukari, M. Dieci, M. Husselman, S. Khan, E. Naah, M. Polansky, E.Strohm, H. Trachtman, and R. Strohm for outstanding research assistance and project management.

Executive Summary

More than one fifth of the world's population lives on less than US$1.25 per day. While many credit and training programs have not been successful at raising income levels for these ultra-poor households, recent support for livelihoods programs has spurred interest in evaluating whether comprehensive "big push" interventions may allow for a sustainable transition to self-employment and a higher standard of living. To test this theory, in six countries researchers evaluated a multi-faceted approach aimed at "graduating" the ultra-poor from poverty. They found that the approach had long-lasting economic and self-employment impacts and that the long-run benefits, measured in terms of household expenditures, outweighed their up-front costs. Here we summarize the Ghana site, which had similar effects as the other successful sites.

Policy Issue:

More than one fifth of the world's population lives on less than US$1.25 per day. Many of these families depend on insecure and fragile livelihoods, including casual farm and domestic labor. Their income is frequently irregular or seasonal, putting laborers and their families at risk of hunger. Self-employment is often the only viable alternative to menial labor for the ultra-poor, yet many lack the necessary cash or skills to start a business that could earn more than casual labor.

In the past, many programs that have provided ultra-poor households with either credit or training to alleviate these constraints have not been successful at raising household income levels on average. However, in recent years, several international and local nongovernmental organizations have renewed their support for programs that foster a transition to more secure livelihoods. Combining complementary approaches—the transfer of a productive asset, training, consumption support, and coaching— into one comprehensive program may help spur a sustainable transition to self-employment. To better understand the effect of these programs on the lives of the ultra-poor, researchers conducted six randomized evaluations in Ethiopia, Ghana, Honduras, India, Pakistan, and Peru.

Context of the Evaluation:

In Ghana, researchers partnered with implementing organizations Innovations for Poverty Action and Presbyterian Agricultural Services (PAS). The study took place in in the Northern and Upper East regions of Ghana, a region that is disproportionately poorer than the coastal south. Fifty-three percent of households in the study were living on US$1.25 a day or less when the study began, compared to 29 percent in Ghana as a whole.

To select the poorest members of the communities, the project team conducted a Participatory Wealth Ranking, in which villagers collectively ranked households according to their wealth during a community meeting. PAS conducted a short survey afterwards to verify the results of the ranking.

Details of the Intervention:
Researchers conducted a randomized evaluation to test the impact of a two-year comprehensive livelihoods program ("the Graduation approach") on the lives of the ultra-poor in northern Ghana. The approach was first developed by the Bangladeshi NGO BRAC in 2002 and has since been replicated in several countries. The Graduation program consisted of six complementary components, each designed to address specific constraints facing ultra-poor households.

In Ghana, researchers first randomly assigned villages composed of a total of 2,606 households, to one of two groups. One group served as a pure comparison group and was not offered the program. In the other group, 666 households were randomly assigned to receive the program. The other half of the households in that group did not receive the program, and served as a sub-comparison group to measure "spillover" effects on non-participating households living nearby. The program consisted of six complementary components, each designed to address specific constraints facing ultra-poor households:

1. Productive asset transfer: One-time transfer of a productive asset valued at GHS 300 (2014 PPP US$451). Forty-four percent of participants chose goats and hens, roughly a quarter picked goats and maize inputs, and small number picked shea nuts and hens (6 percent).

2. *Technical skills training*: Training on running a business and managing their chosen livelihood. For example, households who selected livestock were taught how to rear the livestock, including vaccinations, feed and treatment of diseases.

3. *Consumption support*: During the lean season (14 out of 24 months), households received weekly cash transfers of GHS 4-6 (2014 PPP US$6.02- 9.03), depending on household size.

4. *Health:* Households were enrolled in the National Health Insurance Scheme and received health and nutrition education.

5. Savings account: Half of the Graduation households received savings accounts through the Savings Out of Ultra Poverty (SOUP) program, also implemented by PAS. When PAS staff made their weekly visits, they collected deposits and households logged deposits.

6. Households visits: Weekly visits by PAS staff to provide to provide accountability, coaching, and encouragement.

In order to test the relative effectiveness of the savings and asset transfer component, the researchers also randomly assigned a portion of households (733 households) to only receive the SOUP program, while another portion (329 households) only received the asset transfer component of the program. Half of

those in the SOUP program (362 households) received a 50 percent match on their savings to test the impact of incentives to save.

Researchers conducted the first endline survey immediately after the two-year program ended, as well as a second endline survey around one year later.

Results and Policy Lessons:
Note: Results forthcoming from the relative effectiveness of the savings component (with and without incentives) and an asset transfer-only treatment.

Across all six countries, researchers found that the program caused broad and lasting economic impacts. Treatment group households consumed more, had more assets, and increased savings. The program also increased basic entrepreneurial activities, which enabled the poor to work more evenly across the year. While psychosocial well-being improved, these noneconomic impacts sometimes faded over time. In five of the six studies, long-run benefits outweighed their up-front costs. In Ghana, households that received the Graduation program saw similar effects one year after the program ended:

Economic impacts: Average total monthly consumption among treatment households was 2014 PPP US$33.62, an 11 percent increase over households in the comparison group. They spent $22.41 on food every month on average, 12 percent more than the comparison group. Households saw significant increases in asset holding and borrowed 58 percent more than those in the comparison group (2014 PPP US$35.60 monthly average), They also saved 2014 PPP US$16 a month on average, which was three times more than households in the comparison group**.**

Self-employment: Households experienced a 91 percent increase in non-farm income, earning 2014 PPP US$12.86 on average, as well as significant gains in livestock revenue, earning 2014 PPP US$40.60 a month on average, or 50 percent more than the comparison group.

Psychosocial wellbeing: Households that participated in the program did not report feeling significantly less stressed or happier than households in the comparison group.

Political involvement: Women in treatment households did not experience significant gains in empowerment in Ghana, and in fact experienced significantly less power in decisions about food in the household. However, treatment households did participate in more community meetings than those in the comparison group.

Cost-benefit analysis: Researchers calculated total implementation and program costs to be US$1,777 per household (2014 PPP US$5,408). However, estimated benefits of consumption and asset growth amount to 2014 PPP US$7,175 per household, representing an overall 133 percent return on investment.

List of Figures and Tables

Abbreviations and Acronyms

AO	Asset Only
CAI	Computer Assisted Interviewing
CDF	Cumulative Distribution Function
GUP	Graduation from Ultra Poverty
HFCs	High Frequency Checks
HH	Households
IPA	Innovations for Poverty Action
IRB	Institutional Review Board
LEAP	Livelihood Empowerment Against Poverty
LESDEP	Local Enterprise & Skills Development Program
NHIS	National Health Insurance Scheme
PAS	Presbyterian Agricultural Services
PII	Personally Identifiable Information
PPI	Progress Out of Poverty Index
PPP	Purchasing Power Parity
PRA	Participatory Rural Appraisal
PWR	Participatory Wealth Ranking
RCT	Randomized Controlled Trial
SADA	Savannah Accelerated Development Authority
SOUP	Savings Out of Ultra Poverty
ToC	Theory of Change
TUP	Targeting the Ultra Poor

1. Introduction

The *Graduation from Ultra Poverty* (GUP) program aims to improve the economic status of the very poor and move them towards self-sufficiency in Ghana. GUP is one of ten CGAP-Ford Foundation Graduation Pilots which adapted BRAC's Challenging the Frontiers of Poverty Reduction – Targeting the Ultra Poor program outside of Bangladesh.

The Ghana study is part of a larger six-country evaluation that assessed the impact of the Graduation program across different contexts. All six sites, which include Ghana, Ethiopia, Peru, Honduras, Pakistan and India, implemented and evaluated the Graduation program on similar timelines and with comparable instruments to facilitate cross-country comparisons.[1] The pooled results show strong, cost-effective impacts on livelihoods, living standards, and psychosocial status of the targeted households (Banerjee et al., 2015).

The Graduation Approach is modelled as a method of enabling the ultra-poor to build businesses and improve their lives. The GUP program first identifies the ultra-poor within a community and later intensively works with these families to improve business-oriented skills. The GUP households are provided a productive asset (such as a cow or goats) with which they will develop their enterprise. Overall, the program aims to improve the incomes of the ultra-poor and hopes to see positive changes in school attendance of children, food security, health, and increased assets among the ultra-poor.

Evaluating this program's efficacy in West Africa is particularly important, as over half of the population lives below the poverty line and requirements for emergency relief, food aid, and international funds are on the rise. The effectiveness of the intervention has been evaluated at multiple stages, including: (1) whether the ultra-poor have been successfully identified; and (2) whether the whole program is effective at boosting income and overall social welfare during the intervention period and one year after its conclusion.

The novelty of the Graduation Approach lies not in its constituent components, but in the way components are combined in a holistic way to lift the very poor out of extreme poverty while ensuring they do not slip backward from shocks along the way. While many of the program's components (e.g., consumption support) are relatively well-tested, to our knowledge, only one study was completed on the impact of the integrated set of Graduation components before this project. BRAC's Research and Evaluation Division conducted an impact evaluation of the original TUP program in Bangladesh (Ahmed et al., 2009). The BRAC study compared those identified as eligible for the TUP program to those above the poverty cut-off, and found the ultra-

[1] There were two additional randomized evaluations of CGAP-Ford Foundation Graduation pilots: one in Yemen was delayed due to civil conflict and another in India was conducted by a separate research team.

poor who participated in the program made improvements in several areas, including income and food shortages.

Our evaluation advances the existing knowledge in several important ways. First, it enhances the rigor of the original evaluation design by conducting a randomized trial, removing any selection biases from the comparison between groups. Second, it evaluates a replication of the program in West Africa, extending the external validity of evidence of the effectiveness of the graduation model outside of the original implementer and into a different region, one with key policy importance.

In Ghana, the research design allows for measuring the impact of the whole program as well as two of the central components: the asset drop and the savings component delivered without the rest of the supporting features of the GUP program. We are also able to measure 'spillovers' within communities, allowing for a measure of impact on households with neighbors who benefitted from the program.

The remainder of the report is structured as follows:

- Description of the intervention, theory of change and research hypotheses
- Context and timeline of program implementation and evaluation
- Description of evaluation design and methods
- Description of intervention design and methods
- Presentation of impact evaluation results
- Discussion of implications, policy-relevant findings, and direction for future work

2. Intervention, theory of change, and research hypothesis

2.1 Intervention

The Graduation from Ultra Poor (GUP) project in Ghana involves three treatment arms: the full Graduation from Ultra Poor (GUP) arm, the Savings Out of Ultra Poverty arm (SOUP), and the Asset Only arm (AO). GUP and SOUP also have sub-treatment arms within their interventions.

GUP
The households within the full GUP treatment arm received a comprehensive package of services, including:

(1) A productive asset transfer (such as a goat or guinea fowls)
(2) A consumption stipend of 4 to 6 Cedis (2014 PPP US$6.01-9.02) per week, according to household size. The stipend was provided during roughly the lean season (July – Sept 2011, April 1 to October 15, 2012 and April 1 to July 7, 2013)
(3) A healthcare component
(4) Weekly coaching on assets/enterprises
(5) Education on finances, health, and nutrition

The healthcare component involved registering all GUP clients and three dependents each on the National Health Insurance Scheme (NHIS) in the first year of the program and renewing them in the second year. Field agents provided the consumption stipend, weekly coaching, and education components. They also helped clients create aspiration plans and set concrete goals based on their needs (e.g. build a room, purchase a cloth, etc.).

Half of the households in the GUP treatment arm had a compulsory savings component, which required a minimum of 0.5 Cedis (2014 PPP US$0.75) weekly savings during the lean season while households were receiving a consumption stipend (savings were voluntary during the months households were not receiving a consumption stipend). Our partner organization, Presbyterian Agricultural Services (PAS), hired 23 field agents to help households open savings accounts and collect deposits from them each week, following a Susu collector model.[2] These deposits were placed by station team leaders in individual clients' rural bank accounts. In order to withdraw money, households were required to go to the bank themselves.

Households chose the asset that was transferred to them. Below is the breakdown of asset transfer types (666 is the total number of HHs that received the GUP treatment—we are missing the asset choices of 4 of them):

Table 1: Asset Breakdown

Livelihood Option	# of Clients	% of clients

[2] The Susu model is an informal savings model that allows people to save and access their money securely and gain a limited access to credit.

4 Goats/4 Hens	292	44.84
1 bag (100kg) shea nut/4 Hens	39	5.86
1 bag shea nut/1 acre maize production	34	5.11
4 goats/1 acre maize production	180	27.03
1 acre maize/4 hens	30	4.50
1 acre maize/2 pigs	24	3.60
1 bag paddy rice/4 hens	36	5.40
4 goats/1 bag sorghum	27	4.05
Missing Data		0.60
Total	666	100

SOUP

Similar to the savings component of the GUP treatment, field agents opened savings accounts for households receiving the SOUP treatment and visited each week to collect voluntary savings. In order to withdraw money, households were required to go to the bank themselves. Half of the households in the SOUP treatment arm had their savings matched at 50%. At the onset of the program, there was a maximum match of 1.5 cedis per week (for a 3 cedi deposit) but this cap was eventually removed. While we mention SOUP in the outline of the intervention, we do not have results at this time, as analysis is still ongoing.

Asset Only

The households within the AO treatment arm received goats as a productive asset without any related coaching or support. Unlike the other treatment arms, the AO group was not given the option of selecting their productive asset.

2.2 Outcomes of Interest

The primary research questions of interest include:

- What is the impact of the GUP intervention on social and economic outcomes (income, assets, school attendance of children, health, and food security)?
- What is the viability of “graduating” the ultra-poor to food security and/or microfinance?
- Is mandating savings necessary and sufficient for ensuring financial stability among the target group?

Overall, the program aims to improve the incomes of the ultra-poor and hopes to see positive changes in school attendance of children, food security, health, and increased assets among the ultra-poor.

2.3 Theory of Change

The GUP intervention is based on the premise that the ultra-poor need a more holistic approach to graduate from poverty. The intervention combines a productive asset transfer with an intensive period of training, financial education, consumption support

and saving, with the intention of providing participants all the inputs necessary to successfully start a business. Additionally, consumption support is provided during the lean season to prevent participants from consuming the income generated through the livelihood activity. The Graduation theory of change posits that a combination of all these factors will enable the ultra-poor to eventually graduate (after 24 months) into a more sustainable state. The use of a subsidy for an intensive, well-defined period with the goal of sustainable growth out of poverty could reduce long-term spending on safety nets.

The variations in the full program help us better understand what components of the program drive the effects. We are able to explore if the complementarities of the program's components make it effective or if the individual components, like savings or assets, drive the changes. By comparing the impact of the GUP, SOUP, and AO interventions, this study has aimed to determine the relative impacts and cost-effectiveness of the three variants in improving household economic and social outcomes in the short and medium term.

3. Context

Ghana has experienced steady growth for more than two decades. During this period, there have been several initiatives by various administrations to bring economic stability to the lives of the poor, from the Economic Recovery Programs of the 1980s and 90s to the Ghana Poverty Reduction Strategies of the last decade. In recent years, attempts have been made to target resources directly to the very poor. Examples of such programs are the Livelihood Empowerment Against Poverty (LEAP) and Local Enterprises & Skills Development Program (LESDEP). In order to prioritize social protection, the former Ministry of Women and Children's Affairs has been renamed Ministry of Gender and Social Protection and is in charge of coordinating activities of sector ministries that work to improve the quality of life for the poorest people.

Though Ghana has experienced some success in meeting certain Millennium Development Goals, the successes are not evenly distributed. The 2007 UNDP Ghana Human Development Report indicates that the three northern regions continue to "harbour the poorest of the poor" (United Nations Development Programme, 2007). In part, this may be explained by geography, lack of infrastructure and rainfall patterns. The northern regions are landlocked and further away from the ports and industrial centers. The poor in these regions lack many of the basic resources to engage in some kind of productive activity. Access to financial capital is another limiting factor. The government has acknowledged these conditions that disparately affect northern Ghana and has launched an initiative to bridge the gap between northern and southern Ghana. This initiative is largely driven by the Savannah Accelerated Development Authority (SADA), which was set up to direct and coordinate development projects in the north.

Given this background, northern Ghana was chosen as the site for the GUP project. In order to be representative of northern Ghana, the sample was selected from two of three administrative regions in the north: the Upper East and Northern Regions. These regions share similar characteristics with the Upper West Region in terms of climate, culture, economic activities, housing structure, and religion.

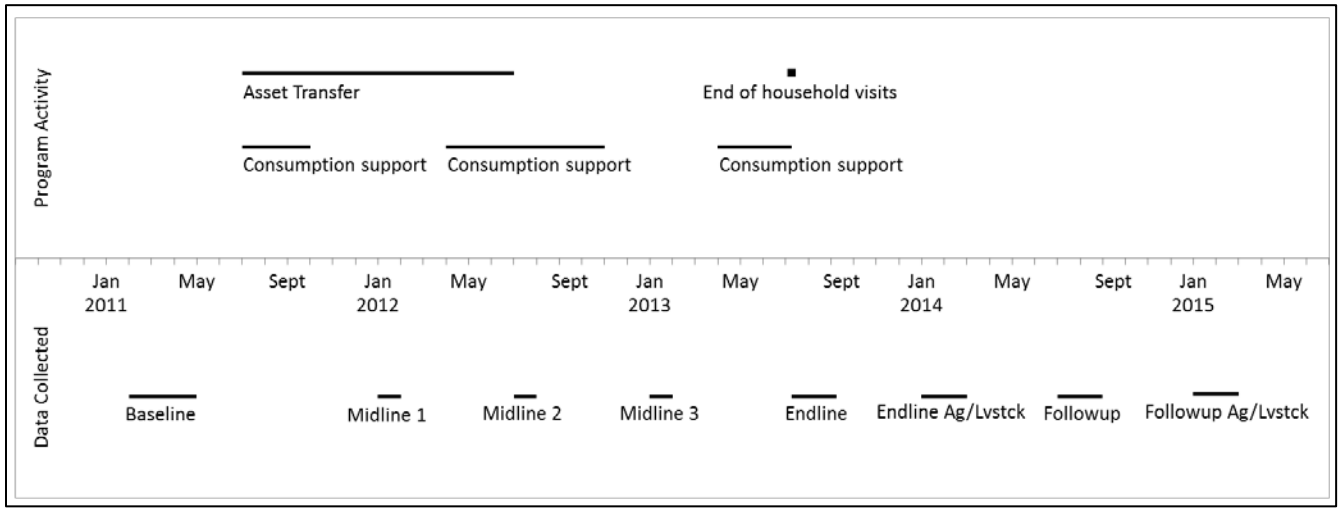
Two hundred and forty-one out of 300 listed communities were selected based on number of eligible project participants found in the communities. The size of the study area was partly influenced by the ability of the partner organization, Presbyterian Agricultural Services (PAS), to implement the program effectively at scale. As a result, areas where PAS had previous work experience and field offices were selected, while still ensuring we were selecting communities in the area that had little to no NGO intervention.

For selected communities, a participatory wealth ranking (PWR) activity was carried out. This involved community members ranking all households by wealth. To verify the ranking, the Ghana "Progress out of Poverty" scorecard was administered to households that were categorized as poor by community members. All households eligible for the GUP program had to be categorized as poor according to the Ghana Progress out of Poverty Index (PPI), the national criteria for poverty. With this level of

rigor in sample selection, we are confident that ultra-poor households selected from the 241 communities and subsequently randomized into treatment and control groups were largely similar to ultra-poor households in non-sample communities in northern Ghana at baseline.

4. Timeline

Table 2: Timeline of intervention and evaluation



Date	**Description**
August – September 2010	BRAC study tour by Implementation Coordinator and two PAS staff
August – September 2010	Area/village selection by PAS
September - October 2010	Field Staff recruitment/selection
October 2010	Training of field staff in PRA/PWR processes
November 2010 - January 2011	Identification of eligible households using PRA/PWR Tools
January – March 2011	Verification of eligible households
February – April 2011	Operational manuals and M&E tools developed
February – April 2011	Baseline survey
April 2011	Treatment/control households selected and final field staff hired
May – June 2011	Field staff trained to deliver GUP services/products
July 8, 2011 - July 7, 2013	Conduct weekly home visit to include:

	• Procurement and transfer of asset to 666 GUP clients • Coaching/training on asset and enterprise development • Provision of consumption stipends • Mobilization of savings and financial education • Education on essential health care and nutrition
January 2012 – January 2013	Short surveys (3 month-long surveys – January, July, January)
June 3 – July 7, 2013	General counseling
July 7, 2013	Five remedial trainings for field staff
July 2013	Stakeholder review and planning meeting
August 1 – 31, 2013	Conduct exit sensitization meetings at community/household levels
	Graduation/end of implementation weekly activities
	Mop-up program outstanding activities
	Write final Ford report
July - September 2013	Endline
January – March 2014	Endline Agriculture/Livestock
June – August 2014	Follow-up
January 2015	Follow-up Agriculture/Livestock
October 1, 2013 – January 31, 2014	GUP 2: Re-sensitization on NHIS/savings & distribution of boxes
January 15 – 31, 2014	Write final implementation Report

5. Evaluation: Design, methods and implementation

5.1 Ethical Measures

As per Institution Review Board (IRB) regulations, all Innovations for Poverty Action (IPA) surveys require informed consent from surveyed individuals. The consent form for GUP was approved by the IRB prior to the start of the study. Respondents had the option to decline to take the survey or to refuse to answer any questions within the survey at any point. Additionally, the consent form listed an IPA phone number which they could call in case of any questions or complaints. All surveyed individuals signed two copies of the consent form: one for IPA records and one for them to keep.

All survey questions and the study protocol were reviewed by the IPA and Yale IRB. Whenever documents, scripts, or protocol were altered, an amendment was submitted to the IRB for approval.

GUP villages (and households within villages) were randomly assigned to receive any given intervention. Every village in the sampling frame had an equal chance of being selected to receive the program. Given the positive results of the six-country evaluations, IPA is working with governments, donors and implementing partner organizations to disseminate the results and discuss scale-up efforts so that other eligible individuals may also benefit from the research findings.

In handling the data cleaning and analysis, all raw data was stripped of personally identifiable information (PII), and the PII was stored separately, encrypted using TrueCrypt encryption technology. We took careful measures to ensure participants' privacy.

5.2 Evaluation and Identification Strategy

IPA used a randomized controlled trial (RCT) to study the impact of the GUP program. Our evaluation strategy involved seven rounds of surveying: baseline, midline 1, midline 2, midline 3, endline, endline agriculture/livestock, follow-up, and follow-up agriculture/livestock. All rounds included a household-level survey. The household survey collected information on assets, savings, income and consumption, health and education, businesses and money transfers, livestock and agriculture, social capital, and household's placement in the community. The baseline survey also had a village and market survey to learn about community activities, institutions, and infrastructure. For the baseline, midline 1-3, endline, and follow-up surveys, there was an additional adult survey that asked a separate set of questions covering time-use, health, risk aversion, social capital, and time- and money-related preferences. Adult surveys were asked to one woman per household. The same woman was surveyed in each round that the adult survey was administered.

The baseline, endline, endline agriculture/livestock, follow-up, and follow-up agriculture/livestock surveys were administered to the full sample. Midline surveys 1-3 were administered to 30% of the sample, and the same subset was interviewed in all three midline rounds. The endline agriculture/livestock and follow-up

agriculture/livestock surveys only included agriculture and livestock modules and were administered after the harvest season. A summary of the surveys by round is below:

Table 3: Surveys by round

Survey	Date	Harvest/Lean	% of communities	Adult Survey	Other Notes
Baseline	Feb – April 2011	Lean	100	Yes	Village + Market Survey
Midline 1	Jan 2012	Harvest	30	Yes	
Midline 2	July 2012	Lean	30	Yes	No Agriculture questions
Midline 3	Jan 2013	Harvest	30	Yes	
Endline	Jun-Aug 2013	Lean	100	Yes	
Endline Ag/lvstck	Jan-Mar 2014	Harvest	100	No	Only ag/livestock questions
Follow-up	Jun-Aug 2014	Lean	100	Yes	
Follow-up Ag/lvstck	Jan 2015	Harvest	100	No	Only ag/livestock questions

In addition to our quantitative surveys, we also had a qualitative component. Forty families across six communities, representing a diverse mix of village and household characteristics, were chosen to participate.

Three qualitative surveyors spent several weeks of every month living in the chosen villages and conducting in-depth semi-structured interviews with GUP beneficiaries, as well as PAS workers and other community members in leadership positions. The interviews were designed to develop life histories, community histories, and experiences with the program. Each month, after three weeks in the villages, the qualitative researchers reconvened to debrief and discuss future possible lines of inquiry and generally adapt the research to the reality on the ground.

5.3 Sample Size Determination

The sample size for the GUP study was calculated to be in line with power calculations done for other sites in this set of studies. Power calculations for the GUP project suggest a total sample size of 2601 (25% treatment and 75% control) households yields a lower bound minimal detectable effect size of 0.127 standard deviations with analysis at the individual level (i.e., using the fact that we randomized at the individual level within treatment groups) and an upper bound minimal detectable effect size of 0.204 standard deviations with analysis at the village level (i.e., using the fact that the first level randomization was to assign villages to either treatment or control, and thus ignoring the second stage individual level randomization within treatment villages).

Our estimation of power utilizes the following assumptions:

1. Power is fixed at 0.8
2. Significance is fixed at 0.05
3. Portion of participants in treatment group is 0.25
4. The intracluster correlation for the group level randomization is 0.10
5. The total sample size is 2607. For the group level randomization, the number of cluster are 154 and the number of observations per cluster are 16.92

We use the following formula to calculate the MDES (from Bloom 2005):

$$MDE = \frac{M_{J-2}}{\sqrt{P(1-P)J}}\sqrt{\rho + \frac{1-\rho}{n}\sigma}$$

Table 4: Power Calculations

	Randomization at Individual Level (random assignment to treatment and control, within treatment villages)	Randomization at group level (random assignment of villages to treatment and control)
p (intracluster correlation)	0	0.1
Power	0.8	0.8
Statistical significance	0.05	0.05
Multiplier	2.8	2.8
Portion of participants in Treatment	0.25	0.25
Portion of participants in Control	0.75	0.75
J (number of clusters)	2607	154
n (number of observations per cluster)	1	16.92
MDE (in standard deviations)	0.127	0.204

5.4 Sampling Design

Eligibility Criteria

Participatory Rural Appraisal (PRA) tools and questionnaires were used to target the ultra-poor households in our sample communities. PAS first provided IPA a list of 300 communities in which no sustained NGO work was ongoing. A Participatory Wealth Ranking (PWR) process was then used to rank the well-being of households within each community. Communities were divided into groups by gender, to allow a more frank assessment of wealth ranking within the community. Each group identified relevant characteristics and household indicators that represent wealth within that specific community. The focus groups then categorized households based on these identified characteristics into the following buckets: very rich, rich, poor, and very poor (ultra-poor). After households were sorted, the groups met together and reconciled discrepancies in their respective categorizations.

This was followed by a survey modeled after the set of questions in the Progress out of Poverty Index (PPI). The PPI survey was administered to over 6,000 poor and ultra-poor households. The exclusion criteria applied using this survey were as follows:

- Ownership of > 30 small ruminants

- Ownership of > 50 fowl
- Roof made of manufactured materials (i.e. corrugated iron) that were purchased by the household (not provided by an NGO)
- Ownership of a mobile phone

The last two criteria caused sample size problems, especially in the Tamale area, and were eventually dropped as exclusion criteria. Any communities with fewer than eight eligible households were dropped. Communities with fewer than eight but more than three eligible households were later used as our AO sample.

During the survey, auditors conducted back checks and field managers followed up with enumerators to ensure surveys were conducted properly. The evaluation team later cleaned the PPI data and removed households which met the following exclusion criteria:

- No eligible female lives in the household (an eligible female is 18-64 years of age)
- The household owns more than 50 fowl
- The household owns more than 30 small ruminants
- The household has a roof made of non-natural materials (such as corrugated iron) that was purchased independently by household members

Once these households were removed from the sample, a list of more than 4,000 potential ultra-poor households was drawn up.

This list was then passed on to the verification team. In communities, verification teams first confirmed the categorizations of each poor and ultra-poor household with the chief and elders. In addition to verifying the PPI-like questionnaire, the following questions were asked:

- Is any member of the HH a drug addict or alcoholic?
- Does any HH member have disability (blindness, physical disability, and/or mental illness)?
- Is the HH living in a roofed house with a roof made from manufactured sheets (e.g. aluminum sheets) that were purchased by the HH?
- Does a member of the HH own a mobile phone?

Households for which community leaders raised doubts about any of the above issues were revisited for confirmation.

Once all households had been ranked and approved, any household which did not have a person meeting the following criteria was removed:

- Female
- Between the ages of 18 and 64 (to reduce attrition in the sample through marriage or death). This age bracket was finally opened to include 64 years and above.
- Individual meeting the above two criteria must be the household head, wife of the household head, or daughter-in-law of the household head

After all of the exclusion criteria were applied, 4,000 households were identified as eligible for our study.

Quantitative Sampling Design

In addition to pure control communities, there were control households within treatment communities (to identify spillovers). Within GUP and SOUP villages, we randomly assigned households to the following groups:

- For GUP communities
    - GUP with savings
    - GUP without savings
    - Control
- For SOUP communities
    - Matched SOUP
    - Ordinary SOUP
    - Control
- For AO communities
    - AO
    - Control

For household selection, we ensured balance on the following variables:

(1) number of compounds
(2) distance to market
(3) household size
(4) an asset index created using principle component analysis
(5) age of primary respondent
(6) a livestock index using principle component analysis
(7) whether the primary respondent operates business
(8) total plot area owned by household
(9) whether someone in the household is a member of a savings group

All households in the same compound had to have the same treatment status.

Quantitative Sample Size
The table below details the various stages of targeting, sample selection and the survey rounds. 1,394 clients (666 GUP/732 SOUP) received weekly home visits during the program and 666 GUP clients were supported with asset and enterprise development of their choice. Our final analysis is done using our sample at endline and follow-up. Tests for differential attrition can be viewed in Section 8.

Table 4: Evolution of sample size

300 grey Communities								
PWR								
PPI Survey 6,000+ HHs								
Verification Process								
Baseline Survey 231 Communities 3,850 HHs								
GUP 78 communities 1,308 HHs			SOUP 77 communities 1,243 HHs			Pure Control 76 commuities 1,299		
GUP - Savings 333 HHs	GUP - No Savings 333 HHs	GUP - Control 642 HHs	SOUP - Not Matched 371 HHs	SOUP - Matched Savings 362 HHs	SOUP - Control 510 HHs			
Midline 1 Survey 69 Communities 1,071 HHs								
GUP 23 communities 341 HHs			SOUP 23 communities 356 HHs			Pure Control 23 commuities 374 HHs		
GUP - Savings 88 HHs	GUP - No Savings 92 HHs	GUP - Control 161 HHs	SOUP - Not Matched 98 HHs	SOUP - Matched Savings 101 HHs	SOUP - Control 157 HHs			
Midline 2 Survey 69 Communities 1,073 HHs								
GUP 23 communities 338 HHs			SOUP 23 communities 353 HHs			Pure Control 23 commuities 382 HHs		
GUP - Savings 88 HHs	GUP - No Savings 91 HHs	GUP - Control 159 HHs	SOUP - Not Matched 96 HHs	SOUP - Matched Savings 102 HHs	SOUP - Control 155 HHs			
Midline 3 Survey 69 Communities 1,020 HHs								
GUP 23 communities 318 HHs			SOUP 23 communities 338 HHs			Pure Control 23 commuities 364 HHs		
GUP - Savings 84 HHs	GUP - No Savings 82 HHs	GUP - Control 152 HHs	SOUP - Not Matched 91 HHs	SOUP - Matched Savings 98 HHs	SOUP - Control 149 HHs			
Use PPI Survey data to identify 45 additional communities to use in an asset only treatment arm								
Endline Survey 275 Communities 3,978 HHs								
GUP 78 communities 1,272 HHs			SOUP 77 communities 1,172 HHs			Asset Only 44 communities 299 HHs		Pure Control 76 commuities 1,235 HHs
GUP - Savings 325 HHs	GUP - No Savings 321 HHs	GUP - Control 626 HHs	SOUP - Not Matched 351 HHs	SOUP - Matched Savings 345 HHs	SOUP - Control 476 HHs	Asset Only 152 HHs	Asset Only - Control 147 HHs	
Endline Agriculture and Livestock Survey 274 Communities 4,005 HHs								
GUP 78 communities 1,271 HHs			SOUP 77 communities 1,184 HHs			Asset Only 43 communities 298 HHs		Pure Control 76 commuities 1,252 HHs
GUP - Savings 329 HHs	GUP - No Savings 319 HHs	GUP - Control 623 HHs	SOUP - Not Matched 356 HHs	SOUP - Matched Savings 344 HHs	SOUP - Control 484 HHs	Asset Only 154 HHs	Asset Only - Control 144 HHs	
Followup Survey 273 Communities 3,901 HHs								
GUP 77 communities 1,240 HHs			SOUP 77 communities 1,167 HHs			Asset Only 43 communities 291 HHs		Pure Control 76 commuities 1,203 HHs
GUP - Savings 313 HHs	GUP - No Savings 319 HHs	GUP - Control 608 HHs	SOUP - Not Matched 355 HHs	SOUP - Matched Savings 338 HHs	SOUP - Control 474 HHs	Asset Only 150 HHs	Asset Only - Control 141 HHs	
Followup Agriculture and Livestock Survey 273 Communities 3,999 HHs								
GUP 77 communities 1,262 HHs			SOUP 77 communities 1,184 HHs			Asset Only 43 communities 301 HHs		Pure Control 76 commuities 1,252 HHs
GUP - Savings 326 HHs	GUP - No Savings 317 HHs	GUP - Control 619 HHs	SOUP - Not Matched 355 HHs	SOUP - Matched Savings 344 HHs	SOUP - Control 485 HHs	Asset Only 152 HHs	Asset Only - Control 149 HHs	

Qualitative Sampling Design

Six communities were selected for our qualitative sample – two from each of the three stations. The communities were split between GUP and SOUP, were of varying sizes, and were all located far from major roads and towns. Village leaders in eligible communities agreed to the long-term presence of the qualitative researchers and each participating family gave oral consent to take part in the qualitative study.

Within each community, the qualitative research team sought to maximize variation between participant households, along the following demographic variables:

- Number of families within compound (single nuclear family versus multiple nuclear families in an extended family compound)
- Female headed household (widows) versus households with support from a male (either family member or spouse)
- Households with co-wives versus households with only one wife

Qualitative Sample Size

The sample contained 40 families across six communities. Specific characteristics of the families are very context-dependent, including livelihood choices due to geographic location, (lack of) infrastructure, and access to particular resources. The table below shows the distribution of sample families across intervention groups over the entire qualitative sample. In each household, at least one individual was interviewed on a regular basis to monitor their experiences over the course of the GUP program. In total, 57 individuals in 40 households were interviewed.

Table 5: Qualitative Sample - Households

Intervention Group		
GUP community	GUP Matched Savings	9
	GUP Unmatched Savings	5
	GUP Control	7
SOUP community	SOUP Matched Savings	9
	SOUP Unmatched Savings	5
	SOUP Control	5

Most households farmed as their primary economic activity, with a range of additional supplemental activities. Interviews were predominantly conducted with women, although male household heads were also interviewed. Participants ranged in age from early 20s to mid-70s. All households had at least one child living in the household or compound; these were predominantly the household's own children, but some were grandchildren or nephews and nieces.

The community members belonged to a number of ethnic groups including Dagomba, Mamprusi, Mosi, Builsa, Hausa, and Fulani. Ethnic identity often determines which families share resources and labor due to common bonds, and some ethnic groups specialize in a particular livelihood. For example, traditionally, the Fulani were nomadic pastoralists who herded cattle, sheep, and goats in the region. The Fulani in this sample largely came from Burkina Faso or another region of Ghana and settled in these

communities. They care for the cattle of other community members in exchange for the milk from the cows, which they consume or sell in larger town markets. The Fulani also described being left out of development aid in the past, either from the government or from NGOs and remarked on the strength of the GUP program for helping the households in their community.

5.5 Treatment Assignment

The Ghana experiment was a clustered randomized trial, with randomization at both the village and the household level. In Ghana "villages were randomly selected to be treatment or control villages, and then treatment house-holds were randomly selected within the set of eligible households in treatment villages. The goal of this design was to be able to measure spillovers." (Banerjee et al., 2015) Randomization was carried out remotely by the research team (using a Stata script). The Ghana site had "two additional treatment groups (savings only, and productive asset grant only) to 'unpack' those aspects of the intervention." (Banerjee et al., 2015). The savings treatment and asset-only treatment has not been fully analyzed yet.

5.6 Data Collection and Construction

As mentioned previously, the GUP study included several rounds of primary data collection, including a baseline survey, three midlines (administered to 30% of the study sample), an endline, and a follow-up survey one year after completion of the study. Details of data collection activities can be found in section 5.2.

All data collection activities were carried out by teams of enumerators using netbooks. The surveys were programmed using Blaise software. The full data collection team was comprised of 46 individuals: 30 enumerators (10 per station), 6 team leaders (1 team leader for every 5 enumerators), 3 field managers (1 per station, overseeing 2 team leaders and their enumerators), 3 auditors and data editors each, and 1 associate field manager responsible for monitoring activities alongside the project coordinator and manager. Each auditor was assigned to one station and was responsible for conducting back-checks, booking interview appointments for subsequent days of surveying, and conducting market surveys, when relevant. The data editors (1 per station) were responsible for scrutinizing the collected data and reconciling the survey and back-check data to produce discrepancy reports.

Prior to the launch of the survey, all team members attended five days of classroom training on survey content. The first three days of training were spent going over all of the questions (and relevant translations) on the survey. Role play activities were incorporated into the training to help team members practice administering the survey. On the fourth day, the teams were introduced to electronic data collection and practiced using the netbooks to conduct the survey. Classroom training was followed by two days of field training where the team conducted surveys in an area that was not a part of the sample. This allowed the enumerators to improve their knowledge of the survey and practice surveying with real respondents. Baseline aside, every round of survey work involved a false launch (i.e., survey work began in a community that

was not a part of the sample, but the data collection team was not aware of this and carried out all work as if the actual survey had begun). The work of the team was closely supervised and feedback/clarifications were provided in real time to ensure operations improved the following day (actual launch). All data collection activities and trainings for enumerators were consistent across treatment and control groups. The enumerators were not aware of the household or community-level treatment assignment.

All trainings were led by the project and field managers. The project coordinator provided additional support for the second half of the training when Computer Assisted Interviewing (CAI) technology was introduced.

The data collection team was paid a daily wage through the IPA office in Ghana. Respondents were compensated in-kind for completing the survey. The compensation for the GUP survey was a bar of soap per questionnaire.

The roles and responsibilities of the implementation and evaluation teams were clearly defined and the two teams were kept separate in order to minimize association between the program and the evaluation. The implementation team always identified themselves as PAS staff, while the evaluation team always identified themselves as IPA staff.

During the data cleaning process, treatment status was not incorporated into the data until immediately before analysis, reducing the risk of research staff introducing any biases in the cleaning and management of the data.

5.7 Data Quality Controls

The following data quality measures were in place prior to and during the data collection process:

1. Design & piloting of survey instruments. All survey instruments were extensively piloted for comprehension, cultural appropriateness, and length (to avoid survey fatigue). The electronic programming was bench-tested and field-tested to iron out any errors in programming and data format.
2. Use of CAI software. CAI software helped reduce survey errors in a variety of ways. This included limiting the range of answer choices (i.e. age); automating complex calculations (i.e. total income); pre-populating tables with information presented earlier in the survey (i.e. household member names); and conducting logic checks. The GUP surveys utilized CAI software for all rounds of data collection.
3. Field management and editing. During the course of data collection, surveyors were managed in the field by a supervisor. The supervisor typically was a former surveyor with extensive experience in field research. The supervisor ensured that surveyors contacted the correct respondents and completed the surveys in a thorough but timely manner.

4. Accompaniments. The supervisors accompanied each enumerator for approximately 10% of their surveys. The goal of the accompaniments was for the supervisor to listen to the enumerator administer the survey to make sure the enumerator's understanding of the survey questions was correct and that they were administering the survey properly.
5. Backchecks. The auditors on the data collection team were tasked with revisiting 10% of the respondents who had already been surveyed to re-administer some selected questions one day after the original survey had been administered. The editors then compared the original answers to the backcheck survey answers to check for discrepancies. Backchecks were conducted to ensure accuracy of data. If the original surveyor failed to find the correct respondent (a challenge in areas where many people share first names and surnames), or falsified data, the backcheck would identify this, and the survey would be conducted again.
6. Scrutiny & high frequency checks (HFCs): All the surveys were regularly checked for completion, missing fields, inconsistencies, etc. For each survey type, a HFC program was written in advance and run at the end of the day, checking all incoming surveys for common errors. HFCs were also used to detect patterns of behavior and performance for surveyors (e.g. amount of time spent on a survey). This information was used to monitor the performance of the enumerators.

There was no data entry for GUP surveys since all data collection was done using CAI. All personally identifiable information (PII) associated with the respondents was saved in an encrypted folder. Only the Principal Investigators and other research staff members approved by the IRB had access to this information.

During the data cleaning and coding process, no changes to the data and variables were hardcoded: all changes were made using .do files in Stata. All incoming data was checked for adherence with skip patterns and for logical consistency. Finally, all .do files related to cleaning and analysis were code-checked by a second staff member for errors.

6. Programme or policy: Design, methods and implementation

6.1 Key programme elements and programmatic activities

GUP was one of the ten CGAP-Ford Graduation pilots, all adaptations of BRAC's "Challenging the Frontiers of Poverty Reduction/Targeting the Ultra Poor program." The program was adapted for northern Ghana by IPA and PAS. GUP staff and PAS leadership traveled to Bangladesh for training from BRAC.

As outlined in section two, the GUP program consisted of the following components:

i. Productive asset transfer (such as goats or guinea fowls)
ii. Consumption stipend of 4-6 cedis (2014 PPP US$6.01-9.02) per week according to household size. This was provided only during the lean season
iii. Weekly coaching on asset/enterprise
iv. Education on finances, health and nutrition
v. Healthcare component: all GUP clients and three dependents each were registered on the National Health Insurance Scheme (NHIS) for the first two years of the program
vi. Savings. All treatment households were encouraged to save as a part of the GUP program. Half of the total treatment households had a compulsory savings component (min of 0.5 cedis) weekly saving during the lean season when the households were receiving a consumption stipend from the program. The program helped all households open savings account and the field agents collected deposits from these households on a weekly basis.

IPA worked with PAS in northern Ghana to deliver the program content for GUP. IPA was responsible for the design of the content, the monitoring of implementation activities, and the evaluation. PAS is a regional organization that has been providing a range of services including technical agricultural training, financial advising, and health services in the Northern, Upper East, and Upper West Regions of Ghana since 1967. When exploring options for partner organizations for GUP, IPA prioritized prior experience in service delivery to rural communities and existing presence in the northern regions of Ghana. PAS was the best fit for these requirements given their involvement in delivery of rural development services to a wide area of northern Ghana. Furthermore, they had existing field offices in the study area with permanent staff (from managers to field agents) stationed in these field offices.

Details on the setting in which the program content was delivered are provided in Section 3 of this report. The GUP program was implemented in the Upper East and Northern Regions.

Program exposure lasted for a period of two years, from July 2011-2013. All asset transfer related activities were completed within the first year of the program. Coaching/training on the productive livelihood was provided prior to the transfer of

assets. Weekly visits to the households were continued through the entire length of the project. All GUP clients (and three dependents each) were registered for the national health insurance in the first year and the registration was renewed for a second year. Consumption support was only provided during the lean season (July – Sept 2011, April 1 to October 15, 2012 and April 1 to July 7, 2013). Field agents collected savings from households on a weekly basis for the length of the program.

No incentives were provided to increase participation in the program. There was compensation, a bar of soap, provided to all participants (treatment and control) for completing the surveys.

The program did not require any special materials or technology. All participants were given lockboxes to hold their savings at the end of the program to encourage GUP clients to keep saving after the program ended. All data collection utilized CAI technology.

6.2 Monitoring system and activities

IPA had a designated implementation team that monitored all project-related implementation activities. The implementation team comprised of the implementation coordinator and the implementation manager. This team worked closely with PAS throughout the project life cycle to ensure that implementation activities went according to project design. The implementation team was responsible for monitoring community selection, the PWR process, and the asset transfer process. During the asset transfer process, the implementation team visited 10% of all treatment households to ensure delivery of assets. Additionally, the implementation coordinator and manager held weekly debriefing sessions with all field agents to discuss progress and challenges related to field activities.

The IPA team developed the implementation manual for the project. Before the roll-out of the project, there was a training session for all PAS field agents on the implementation manual and field protocols. PAS was aware that the study was a RCT. The IPA team conducted several rounds of trainings (prior to the start of the program and quarterly after that) for the PAS staff to familiarize them with the concept of RCTs and the implications for implementation. The trainings also included material on how to avoid violating the research design during implementation.

In addition to this, specific content training was provided by various agencies depending on the topic. For example, the Ministry of Food and Agriculture staff trained PAS staff on livestock rearing and care. Similarly, a consultant counselor was hired to train the staff on how to coach and counsel the GUP clients during the weekly visits. These professional counselors were also invited to provide refresher trainings during the quarterly trainings held by IPA.

There were no major changes to the core program during the course of the study period. There were some minor changes made that we list below:

1. The initial plan was to conduct 4 midline surveys, with the fourth between endline and follow-up, but this was replaced by the livestock/agriculture surveys.
2. Monitoring activities indicated that the SOUP-Matching clients were not saving differently from the regular SOUP clients. The PIs decided to raise the cap for matching savings to see if this would result in additional savings. The original product offered a 50% match up to 3 cedis. The revised product raised the matching limit to 20 cedi.

The recruitment strategy and selection process is explained in greater detail above in sections 5.3 (Sample size determination) and 5.4 (Sample design). The aim of the program was to target the poorest population in rural settings in northern Ghana. As mentioned earlier, the eligibility criteria were altered slightly after observing the reality on the ground. The World Bank reports that the share of households in Ghana living below PPP US $1.25/day is 29%. In our sample, 54% of households lived below PPP US $1.25/day. Additionally, for our sample, the median daily per capita consumption (including health and durable goods spending) is $1.45, median monthly per capita consumption is $43.35 and mean monthly per capita consumption is $53.97. This indicates that the targeting for GUP was successful. Changes to eligibility criteria during the targeting and selection process have been noted above. There were no adverse events during the course of GUP implementation.

7. Impact Analysis and Results of the Key Evaluation Questions

7.1 Primary Quantitative Specifications

Primary Equation

Our primary equation for GUP analysis is as follows:

$$Y_i^k = \alpha + \beta_1 assignment_i + \beta_2 Z_i^k + \boldsymbol{U_{shortsurveys}} + \boldsymbol{V_{stratification}} + \boldsymbol{G_{geographicblock}} + \varepsilon_i$$

where Y_i^k is the outcome *k* of interest for either household or adult *i*, $assignment_i$ is an indicator for having been randomly selected into the program, Z_i^k is the household or adult's baseline value of the outcome variable *k* (coded as zero, with an indicator for missing baseline, whenever it was not available), $\boldsymbol{U_{shortsurveys}}$ is a vector of dummy variables, one for midline 2 and one for midline 3 (we averaged together midline 2, midline 3, and endline for our endline results since they were all done within the same calendar year), and $\boldsymbol{V_{stratification}}$ is the vector of all variables included in stratification. During randomization, we re-randomized to ensure balance on a set of variables. These variables (community level: number of compounds, distance to market; household level: hhsize, asset pca index, age of primary respondent, livestock pca index, primary respondent operates business, total plot area owned by hh, household is member of savings group) are included as controls. We also include geographic block strata, which are included here as dummies for each block. In other words, we include a dummy for each treatment village but none for pure control villages, since randomization didn't occur at the individual level for households in control villages.

For the AO analysis, the equation is slightly different:

$$Y_i^k = \alpha + \beta_1 assignment_i + \beta_2 Z_i^k + \boldsymbol{R_{round}} + \boldsymbol{G_{geographicblock}} + \varepsilon_i$$

We no longer have controls for stratification (as these were baseline values, which we did not have for our asset only households) and we no longer control for the inclusion of midline 2 and midline 3, as AO households were not identified at the time of these surveys. We pooled endline and endline agriculture/livestock survey data, in addition to follow-up and follow-up agriculture/livestock survey data, so $\boldsymbol{R_{round}}$ is a dummy to control for the round.

Balance Tables

Below is a table showing tests on orthogonality for households that received any GUP treatment (either with or without savings). "Control" refers to control households within GUP villages and all households within control villages. For AO households, no baseline was performed so we are unable to test for balance.

Table 6: Orthogonality

Orthogonality, Comparison of Means

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Total per capita consumption, standardized	Food Security Index	Asset Index	Financial Inclusion Index	Total Time Spent Working, Standardized	Incomes and Revenues Index	Physical Health Index	Mental Health Index	Political Involvement Index	Women's Empowerment Index
Ghana										
Control Mean (standard error)	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000
	(-0.0227)	(-0.0227)	(-0.0227)	(-0.0227)		(-0.0227)	(-0.0228)		(-0.0229)	(-0.0227)
Treatment Mean (standard error)	-0.0368	0.0957	-0.0131	-0.0247		0.0044	0.0488		0.0005	0.0218
	(-0.0357)	(-0.043)	(-0.0254)	(-0.0255)		(-0.051)	(-0.0394)		(-0.039)	(-0.0395)
p-value from t-test of equality of means	0.4044	0.0387	0.7518	0.5537		0.9289	0.281		0.9914	0.6293
Orthogonality, Baseline Characteristics on Treatment										
Ghana	-0.0370	0.096**	-0.0130	-0.0250	0.0490		0.0044	0.0120	0.0005	0.0220
	(0.0440)	(0.0460)	(0.0420)	(0.0420)	(0.0450)		(0.0490)	(0.0450)	(0.0450)	(0.0450)
Observations	2602	2604	2606	2601	2579		2605	2560	2572	2617
R-Squared	0.000	0.002	0.000	0.000	0.000		0.0000	0.0000	0.0000	0.0000

The table below shows a joint test of significance on all of the family outcome variables we present results on later within this section. We looked at a number of welfare variables as well as ten indices and aggregate variables generated from these welfare component variables. The table below tests the joint significance of these ten indices and aggregate variables.

Table 7: Joint test of significance

	Ghana
Joint F-test of treatment on all family outcome variables	0.846
p-value	0.573

Notes on Data

During the data cleaning process, some outliers were recoded, but there was no systematic approach, such as through winsorizing or truncating values. For our analysis on the asset only households, variables were winsorized at the 99th percentile.

In terms of our data collection, it should be noted that all outcomes are self-reported as all of our data was collected via surveys.

7.2 Pre-Analysis Plan

We did not file a pre-analysis plan. Instead, we include all results from the surveys, regardless of their statistical significance. Additionally, we do not have any empirical results of interest outside those that we present. Our comparison control group is made up of all households in the control villages and control households within GUP villages.

7.3 Empirical Results

Below we include the outcomes for several welfare variables for households that received any GUP treatment (i.e. received either GUP with a savings component or GUP without a savings component) and for households that were part of the asset only intervention. Results presented are intent-to-treat estimates.

All values are reported in 2014 USD, Purchasing Power Parity (PPP) terms. Standard errors are reported in parentheses below the estimates.

The estimates that are reported reflect the increase in the given variable associated with a household that was included in the given program. For example, in the consumption table below, being included in the GUP program is associated with a PPP US $2.82 increase from the control mean in total consumption per capita per month.

Table 8: Effects on per capita consumption

Any GUP

	Endline				*Followup*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Consumption per capita, month	Food consumption per capita, month	Nonfood consumption per capita, month	Durable good expenditure per capita, month	Total Consumption per capita, month	Food consumption per capita, month	Nonfood consumption per capita, month	Durable good expenditure per capita, month
Any GUP	2.82**	2.18**	0.44	0.05	3.22***	2.41***	0.55	0.03
	(1.42)	(1.04)	(0.76)	(0.037)	(1.19)	(0.81)	(0.66)	(0.04)
Observations	2525	2525	2525	2525	2434	2434	2438	2442
R-squared	0.19	0.18	0.10	0.07	0.16	0.16	0.08	0.10
Control mean	40.8	30.1	10.3	0.34	30.40	20.00	8.63	0.34
Baseline mean	54.4	44.90	9.5	0.47	54.40	44.90	9.50	0.47

Asset Only

	Endline				*Followup*			
	Total Consumption per capita, month	Food consumption per capita, month	Nonfood consumption per capita, month	Durable good expenditure per capita, month	Total Consumption per capita, month	Food consumption per capita, month	Nonfood consumption per capita, month	Durable good expenditure per capita, month
Asset Only	-0.64	-0.06	-0.64	-0.03	-0.14	-0.45	0.54	0.02
	(2.48)	(2.01)	(0.91)	(0.055)	(2.18)	(1.57)	(0.86)	(0.07)
Observations	1528	1528	1528	1528	1482	1482	1486	1491
R-squared	0.04	0.04	0.04	0.03	0.06	0.06	0.05	0.05
Control mean	38.10	27.80	8.95	0.32	32.40	21.60	8.53	0.36

Note:
1. All values reported in 2014 USD, Purchasing Power Parity (PPP) terms
2. Column (1) is the sum of columns (2) through (4).

Table 9: Effects on food security

Any GUP

	Endline				*Followup*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Everyone in the household gets enough food everyday	No adults skipped meals	No one in the household went a whole day without food	No children skipped meals	Everyone in the household gets enough food everyday	No adults skipped meals	No one in the household went a whole day without food	No children skipped meals
Any GUP	0.03	0.02	0.039*	0.00	0.01	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	2501	2501	2499	2500	2438	2438	2437	2436
R-squared	0.26	0.20	0.14	0.15	0.15	0.20	0.18	0.21
Control mean	0.27	0.26	0.71	0.45	0.25	0.23	0.73	0.44
Baseline mean	0.15	0.12	0.50	0.41	0.15	0.12	0.50	0.41

Asset Only

	Endline				*Followup*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Everyone in the household gets enough food everyday	No adults skipped meals	No one in the household went a whole day without food	No children skipped meals	Everyone in the household gets enough food everyday	No adults skipped meals	No one in the household went a whole day without food	No children skipped meals
Asset Only	0.00	0.01	-0.04	-0.01	0.01	-0.01	-0.04	-0.07
	(0.03)	(0.03)	(0.04)	(0.05)	(0.04)	(0.03)	(0.05)	(0.05)
Observations	1528	1528	1526	1527	1489	1489	1488	1487
R-squared	0.10	0.09	0.09	0.09	0.10	0.12	0.08	0.09
Control mean	0.27	0.27	0.72	0.45	0.26	0.24	0.74	0.45

Table 10: Effects on asset ownership

Any GUP

	Endline			*Followup*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Asset Index	Productive Asset Index	Household Asset Index	Asset Index	Productive Asset Index	Household Asset Index
Any GUP	0.25***	0.25***	0.07	0.34***	0.34***	0.16***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)
Observations	2584	2584	2583	2442	2442	2442
R-squared	0.32	0.28	0.42	0.30	0.27	0.37
Control mean	0.00	0.00	0.00	0.00	0.00	0.00
Baseline mean	0.00	0.00	0.00	0.00	0.00	0.00

Asset Only

	Endline			*Followup*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Asset Index	Productive Asset Index	Household Asset Index	Asset Index	Productive Asset Index	Household Asset Index
Asset Only	0.00	-0.07	0.01	0.06	0.03	0.00
	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)
Observations	3081	3081	1528	3042	3042	1489
R-squared	0.07	0.06	0.09	0.05	0.04	0.07
Control mean	-0.04	-0.03	-0.05	-0.03	-0.03	-0.04

Note:
1. Columns (2), (4) and (6) are reported in 2014 USD, Purchasing Power Parity (PPP) terms.
2. For Any GUP results, indices weight assets based on their relative values using several other countries' values we collected in similar surveys. They are standardized to Ghana's control group mean in every time period. For Asset Only, we show results for z-scores indices on ownership of certain assets.

Table 11: Effects on financial inclusion

Any GUP

	Endline		*Followup*	
	(1)	(2)	(3)	(4)
	Total amount borrowed, last 12 months	Total savings	Total amount borrowed, last 12 months	Total savings
Any GUP	5.43	16.8***	13.1*	10.5***
	(4.80)	(3.09)	(7.51)	(2.39)
Observations	2522	2520	2437	2435
R-squared	0.099	0.074	0.048	0.098
Control mean	17.30	7.15	22.50	5.49
Baseline mean	22.90	3.85	22.90	3.85

Asset Only

	Endline		*Followup*	
	(1)	(2)	(3)	(4)
	Total amount borrowed, last 12 months	Total savings	Total amount borrowed, last 12 months	Total savings
Asset Only	1.01	0.83	2.42	0.63
	(3.73)	(1.36)	(4.43)	(1.34)
Observations	1526	1525	1488	1488
R-squared	0.022	0.034	0.034	0.038
Control mean	13.50	3.84	18.00	4.19

Note:
1. All values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.
2. Total amount borrowed includes formal borrowing, informal borrowing and borrowing where the source was not specified.

Table 12: Effects on income and revenues

Any GUP

	Endline				*Followup*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Household livestock revenue, month	Household agricultural income, month	Household Non-farm Micro-enterprise Income, month	Household Income from Paid Labor, month	Household livestock revenue, month	Household Non-farm Micro-enterprise Income, month	Household Income from Paid Labor, month
Any GUP	1.31	0.61	5.32***	0.84*	13.6***	6.13***	0.24
	(1.23)	(0.62)	(1.82)	(0.48)	(2.43)	(1.92)	(0.67)
Observations	2519	2531	2522	2525	2436	2438	2438
R-squared	0.13	0.38	0.06	0.03	0.21	0.08	0.04
Control mean	9.2	6.5	6.90	1.88	27.00	6.73	2.39
Baseline mean		31.8	9.73	3.12		9.73	3.12

Asset Only

	Endline				*Followup*		
	Household livestock revenue, month	Household agricultural income, month	Household Non-farm Micro-enterprise Income, month	Household Income from Paid Labor, month	Household livestock revenue, month	Household Non-farm Micro-enterprise Income, month	Household Income from Paid Labor, month
Asset Only	-2.19*	2.12	0.31	-0.07	-0.85	-2.43	-0.35
	(1.31)	(2.06)	(1.62)	(0.82)	(1.42)	(1.89)	(0.92)
Observations	1547	1549	1523	1528	3038	1490	1490
R-squared	0.03	0.05	0.03	0.06	0.02	0.05	0.02
Control mean	8.1	21.9	5.28	1.66	8.54	6.24	1.91

Note:
1. Agriculture income was not available in followup
2. Columns (1) to (4) are reported in 2014 USD, Purchasing Power Parity (PPP) terms.
3. Agriculture prices were not recorded in Followup, thus agricultural income is not reported

Table 13: Effects on use of time

Any GUP

	Endline					*Followup*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(6)	(7)
	Total minutes spent on productive activities in last day	Minutes spent on agriculture in last day	Minutes spent tending livestock in last day	Minutes spent on own business in last day	Minutes spent on paid labor in last day	Total minutes spent on productive activities in last day	Minutes spent on agriculture in last day	Minutes spent tending livestock in last day	Minutes spent on own business in last day	Minutes spent on paid labor in last day
Any GUP	5.28	0.82	1.99***	1.75	-0.24	10.00	10.90	0.97	2.65	-3.73
	(11.20)	(9.40)	(0.63)	(8.00)	(1.89)	(11.70)	(10.80)	(0.77)	(6.18)	(2.77)
Observations	2377	2377	2377	2377	2376	2279	2279	2279	2279	2279
R-squared	0.17	0.14	0.09	0.10	0.10	0.19	0.20	0.06	0.08	0.05
Control mean	147.00	102.00	0.36	38.50	6.69	171.00	136.00	0.86	27.20	6.91
Baseline mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Asset Only

	Endline					*Followup*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(6)	(7)
	Total minutes spent on productive activities in last day	Minutes spent on agriculture in last day	Minutes spent tending livestock in last day	Minutes spent on own business in last day	Minutes spent on paid labor in last day	Total minutes spent on productive activities in last day	Minutes spent on agriculture in last day	Minutes spent tending livestock in last day	Minutes spent on own business in last day	Minutes spent on paid labor in last day
Asset Only	61.8**	45.5*	1.95	9.94	-0.80	-6.97	-0.45	0.00	-5.25	0.99
	(26.10)	(24.00)	(1.19)	(12.60)	(0.78)	(28.70)	(27.20)	(0.00)	(10.70)	(5.59)
Observations	1406	1406	1406	1406	1406	1387	1387	1387	1387	1387
R-squared	0.12	0.12	0.03	0.05	0.00	0.11	0.12	0.00	0.04	0.08
Control mean	148.00	109.00	0.65	29.40	4.42	173.00	139.00	1.00	24.30	5.11

Note:
1. Column (1) is the sum of columns (2) through (5).
2. In Ghana, the adult survey was asked almost exclusively to women. Very few respondents in Ghana tended to livestock.

Table 14: Effects on political involvement and women's empowerment

Any GUP

	Endline					*Followup*				
	Political Involvement	*Women's Empowerment*				*Political Involvement*	*Women's Empowerment*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Attended village meeting in last year	Woman has major say on food decisions	Woman has major say on education decisions	Woman has major say on health decisions (personal and family)	Woman has major say in how to manage household finances	Attended village meeting in last year	Woman has major say on food decisions	Woman has major say on education decisions	Woman has major say on health decisions (personal and family)	Woman has major say in how to manage household finances
Any GUP	0.089***	0.01	0.04	0.01	0.04	0.049*	-0.048*	0.03	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Observations	2335	2319	2266	2617	2264	2294	2272	2220	2617	2195
R-squared	0.13	0.09	0.09	0.07	0.17	0.12	0.14	0.11	0.10	0.08
Control mean	0.57	0.51	0.37	0.58	0.41	0.45	0.37	0.23	0.44	0.28
Baseline mean	0.55	0.45	0.33	0.50	0.36	0.55	0.45	0.33	0.50	0.36

Asset Only

	Endline					*Followup*				
	Political Involvement	*Women's Empowerment*				*Political Involvement*	*Women's Empowerment*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Attended village meeting in last year	Woman has major say on food decisions	Woman has major say on education decisions	Woman has major say on health decisions (personal and family)	Woman has major say in how to manage household finances	Attended village meeting in last year	Woman has major say on food decisions	Woman has major say on education decisions	Woman has major say on health decisions (personal and family)	Woman has major say in how to manage household finances
Asset Only	0.03	0.07	-0.07	-0.05	0.13	-0.01	0.02	0.02	0.03	-0.06
	(0.06)	(0.08)	(0.09)	(0.06)	(0.09)	(0.06)	(0.07)	(0.07)	(0.06)	(0.09)
Observations	1415	1401	1364	1390	1363	1396	1389	1355	1374	1346
R-squared	0.07	0.04	0.03	0.05	0.07	0.06	0.08	0.07	0.06	0.08
Control mean	1.42	1.60	1.77	0.65	1.85	1.54	1.77	1.94	0.50	2.05

Table 15: Effects on physical and mental health

Any GUP

	Endline					*Followup*				
	Physical Health		*Mental Health*			*Physical Health*		*Mental Health*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Member has not missed any days due to illness, last month	Activities of Daily Living Score	Self-reported happiness (1-10)	Stress Index	Member has not experienced a period of worry in last year	Member has not missed any days due to illness, last month	Activities of Daily Living Score	Self-reported happiness (1-10)	Stress Index	Member has not experienced a period of worry in last year
Any GUP	0.01	0.041***	0.16***	0.12**	0.03	-0.01	0.01	-0.01	0.05	0.02
	(0.03)	(0.02)	(0.05)	(0.05)	(0.02)	(0.03)	(0.01)	(0.05)	(0.06)	(0.02)
Observations	2277	2336	2373	2333	2332	2215	2291	2287	2292	2291
R-squared	0.09	0.20	0.14	0.10	0.09	0.07	0.20	0.09	0.11	0.13
Control mean	0.63	0.85	2.83	0.00	0.21	0.68	0.88	3.06	0.00	0.29
Baseline mean	0.71	0.88	2.61	0.00	0.00	0.71	0.88	2.61	0.00	0.00

Asset Only

	Endline					*Followup*				
	Physical Health		*Mental Health*			*Physical Health*		*Mental Health*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Member has not missed any days due to illness, last month	Activities of Daily Living Score	Self-reported happiness (1-10)	Stress Index	Member has not experienced a period of worry in last year	Member has not missed any days due to illness, last month	Activities of Daily Living Score	Self-reported happiness (1-10)	Stress Index	Member has not experienced a period of worry in last year
Asset Only	0.02	-0.04	-0.26**	-0.18	0.03	-0.03	-0.03	0.21**	0.01	-0.03
	(0.05)	(0.09)	(0.11)	(0.12)	(0.05)	(0.06)	(0.09)	(0.10)	(0.10)	(0.04)
Observations	1378	1415	1409	1412	1412	1341	1395	1391	1395	1395
R-squared	0.05	0.05	0.07	0.10	0.05	0.02	0.05	0.07	0.09	0.13
Control mean	0.65	1.63	2.93	0.00	0.21	0.68	1.52	3.06	0.00	0.30

Note: The adult survey was asked exclusively to women.

Below is a table of GUP outcomes for several welfare indices and aggregate variables generated from the components of each of the preceding tables. We report the standardized mean treatment effect and the q-value, which has been adjusted for all 10 hypotheses. Results presented are mean standardized intent-to-treat estimates, including controls for the household's value at baseline and controls for geographic units used for block stratification and variables used in randomization to ensure balance. Dummy variables are included for endline 1 regressions for whether the values include each wave of short survey data. Standard errors are reported in parentheses below the estimates.

Because these estimates are standardized, they are interpreted differently than our results presented above: the estimates are the number of standard deviations (measured on the control group) the treatment is above the control mean. For example, looking at column 1, we see that receiving the GUP treatment increases per capita consumption by 0.097 standard deviations.

Table 16: Indexed family outcome variables and aggregates

	Endline 1		*Endline 2*	
	(1)	(2)	(3)	(4)
Indexed Outcomes	Standardized Mean Treatment Effect	q-value for all 10 hypotheses	Standardized Mean Treatment Effect	q-value for all 10 hypotheses
Total per capita consumption, standar	0.097**	0.068	0.136***	0.018
	(0.049)		(0.050)	
Food security index (4 components)	0.065	0.168	0.077*	0.144
	(0.044)		(0.045)	
Asset index	0.247***	0.001	0.342***	0.001
	(0.049)		(0.057)	
Financial inclusion index (2 component	0.261***	0.001	0.341***	0.001
	(0.060)		(0.080)	
Total time spent working, standardized	0.026	0.638	0.042	0.558
	(0.056)		(0.049)	
Incomes and revenues index (4 compo	0.156***	0.004	0.330***	0.001
	(0.049)		(0.063)	
Physical health index (2 components)	0.114**	0.053	-0.011	0.847
	(0.053)		(0.055)	
Mental health index (3 components)	0.177***	0.003	0.035	0.665
	(0.052)		(0.057)	
Political Involvement index (1 compone	0.179***	0.003	0.099*	0.144
	(0.054)		(0.055)	
Women's empowerment index (4 comp	0.045	0.438	0.029	0.665
	(0.053)		(0.054)	

7.4 Heterogeneities of impacts

We present cumulative distribution function (CDF) graphs for three primary outcome variables for GUP households, AO households, and control households to gain insight into the heterogeneity of impacts. The outcomes we present are monthly consumption per capita, monthly revenue from livestock, and total savings.

Figure 1: Endline per capita consumption CDF

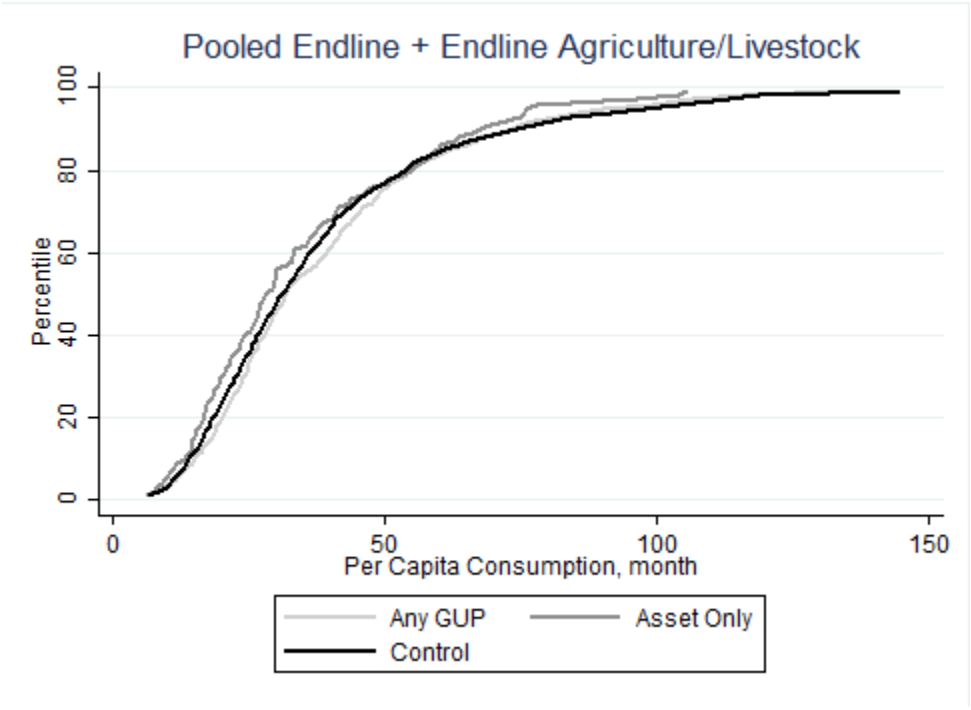


Figure 2: Followup per capita consumption CDF

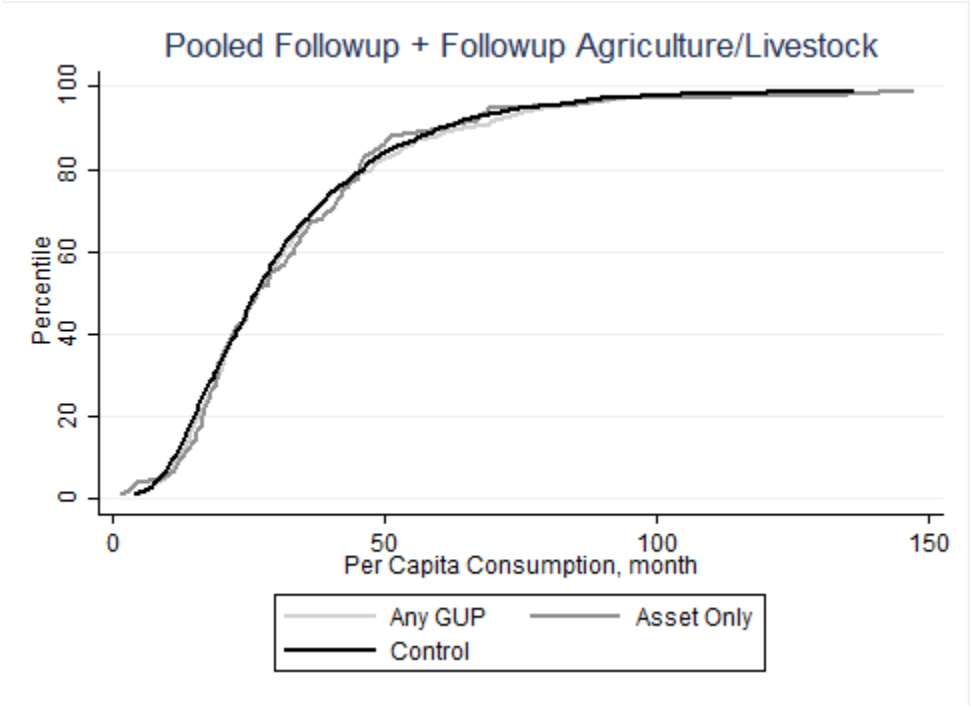


Figure 3: Endline livestock revenue CDF

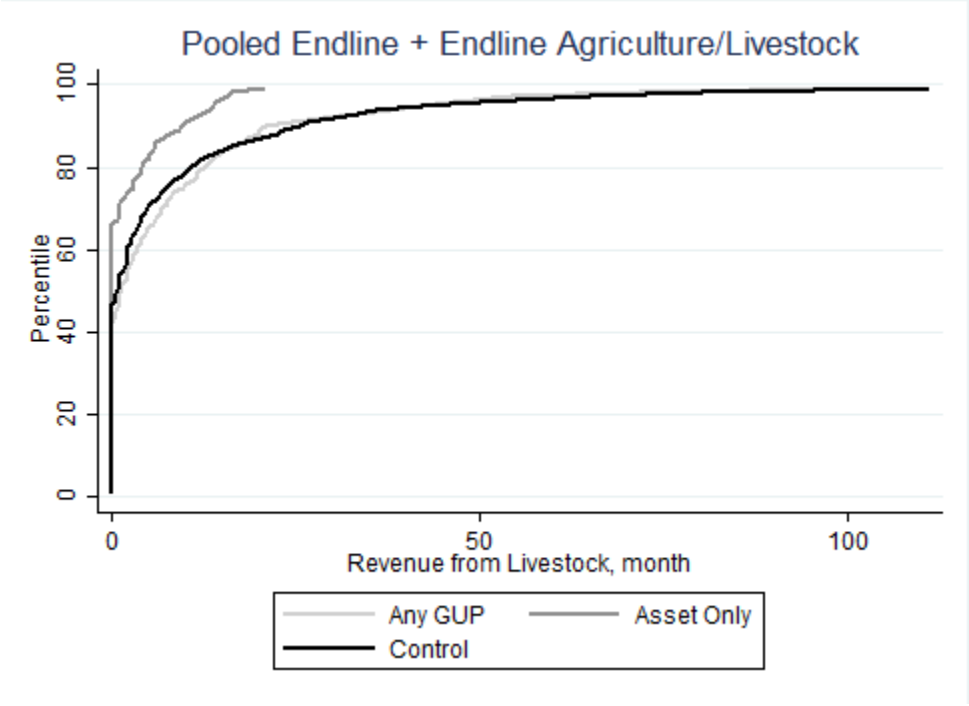


Figure 4: Followup livestock revenue CDF

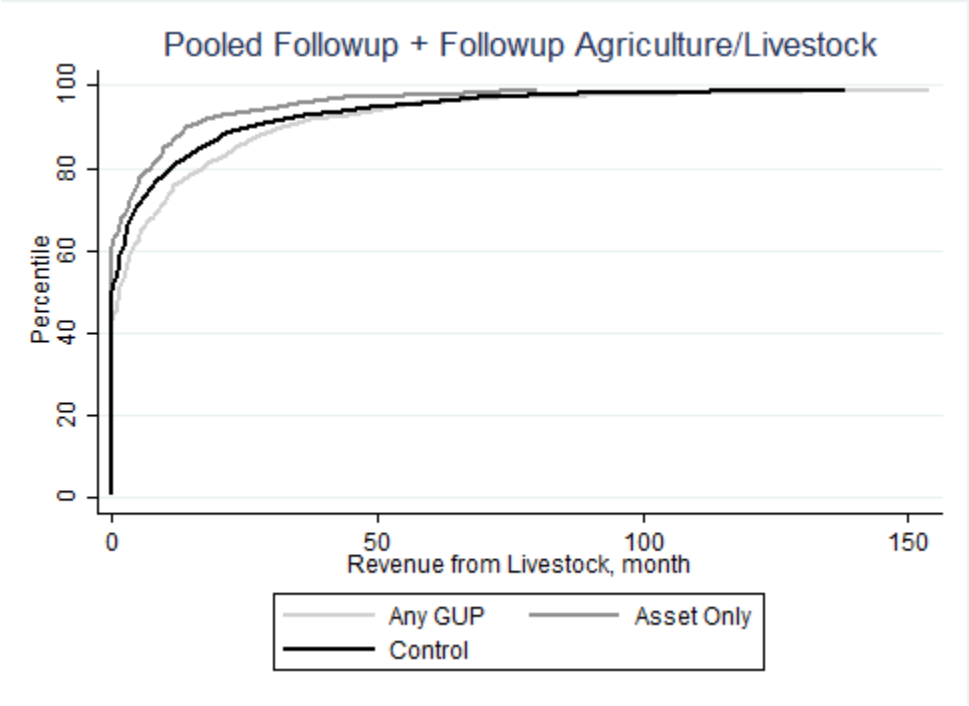


Figure 5: Endline total savings CDF

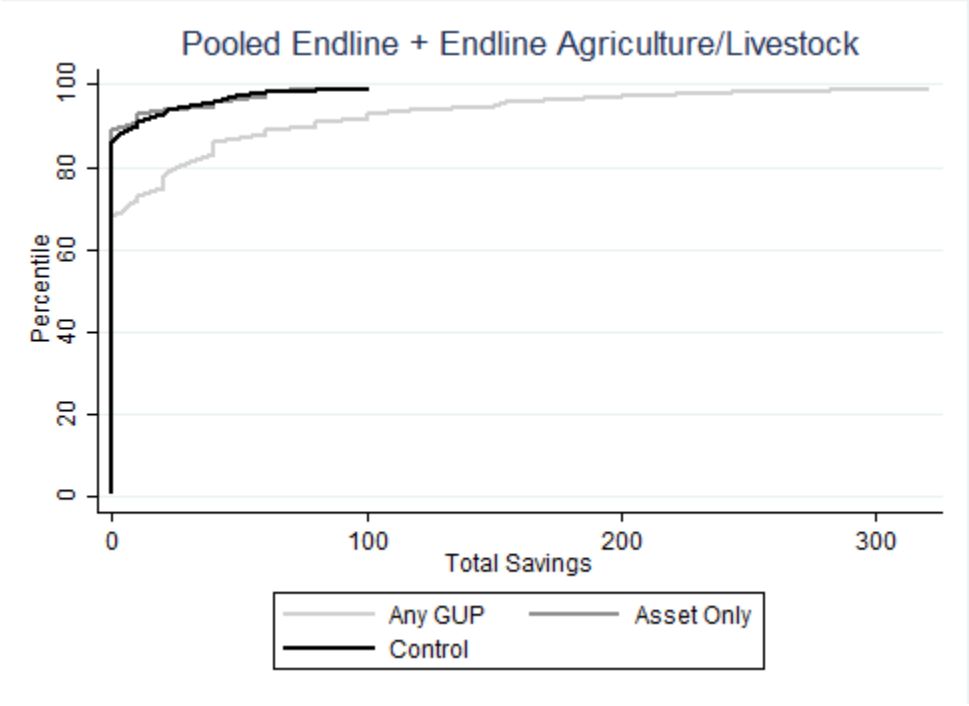
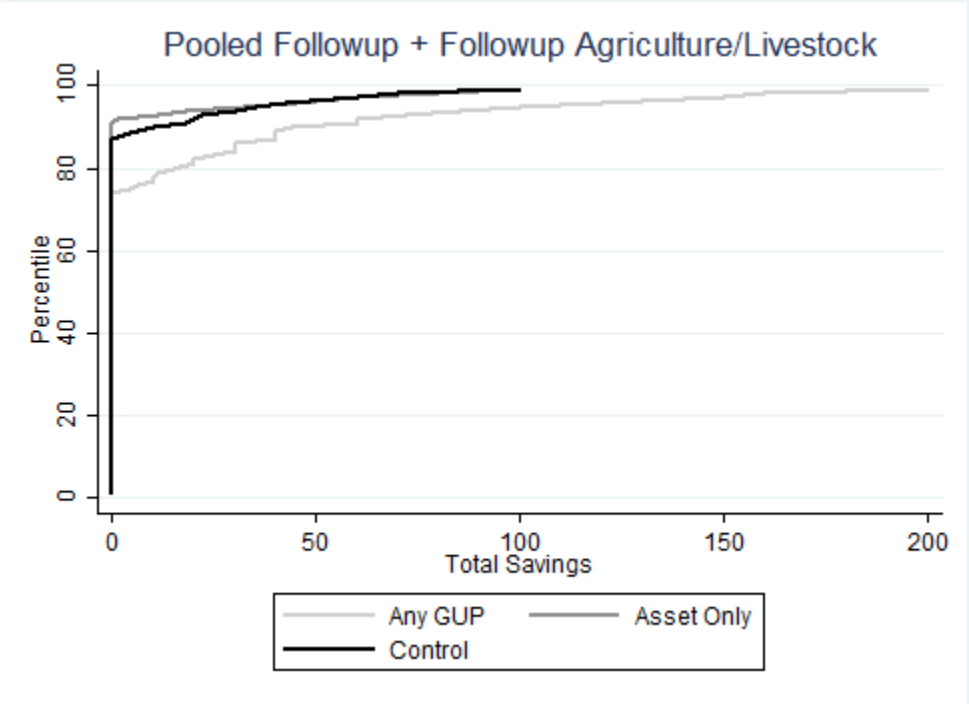


Figure 6: Followup total savings CDF



7.5 Costs of the Program

The table below outlines the cost-benefit analysis done on the full GUP program.

Table 17: Cost-Benefit Analysis

	Panel A: Program Costs per Household, USD PPP 2014	Ghana
	Direct Transfer Costs	680
(1)	Asset Cost	451
	Food stipend	229
	Total Supervision Costs	2832
	Salaries of Implementing Organization Staff	1994
	Materials	119
	Training	44
	Travel Costs	293
	Other Supervision Expenses	382
	Total Direct Costs	3513
	Start-up expenses	133
	Indirect Costs	1026
	Total Costs, calculated as if all incurred immediately at beginning of Year 0	4672
(2)	Total Costs, Inflated to Year 3 at 5% annual discount rate	5408
	Exchange rate to PPP Adjustment Scalar	2.19
	Panel B: Benefits per Household, USD PPP, All Values Inflated or Deflated to Year 3 at 5% annual social dis	
(3)	Year 1 Annual Non-durable Consumption ITT, assuming treatment effect equal to Year 2	293
(4)	Year 2 Annual Non-durable Consumption ITT Treatment Effect	293
(5)	Year 3 Household Asset ITT Treatment Effect	15
(6)	Year 3 Non-durable Annual Consumption ITT Treatment Effect	332
(7)	Year 4 Onward Total Consumption ITT Treatment Effect, assuming Year 3 gains persist in p	6241
(8)	Total Benefits: (3) + (4) + (5) + (6) + (7) = (8)	7175
(9)	Year 3 Productive Asset ITT Treatment Effect	118
(10)	Year 3 Savings Balance ITT Treatment Effect	11
	Panel C: Benefit/Cost Ratios	
(11)	Total Benefits/Total Costs Ratio: (8) / (2) = (11)	133%
(12)	Increase in Asset Value in Year 3 (Household, Productive and Financial) / Cost of Asset Transfers: ((5) + (9) + (10)) / (1) = (13)	32%
(13)	Increase in Asset Value/Transfers, 10th percentile	5%
(14)	Increase in Asset Value/Transfers, 25th percentile	12%
(15)	Increase in Asset Value/Transfers, 50th percentile	20%
(16)	Increase in Asset Value/Transfers, 75th percentile	29%
(17)	Increase in Asset Value/Transfers, 90th percentile	37%
	Sensitivity Analysis	
(18)	Internal Rate of Return (IRR)	6.9%
(19)	Annual Rate of dissipation of the treatment effect such that Costs = Benefits	1.8%
(20)	Benefit/Cost Ratio, at discount rate of 7%	93%
(21)	Benefit/Cost Ratio, at discount rate of 10%	64%

Note:

1.. In Ghana, individuals did not provide an estimate of the value of all assets. We use the relative value of assets across sites where a similar program was implemented and the average purchase prices available in each country (e.g. goats and cattle) to provide an estimate of the asset ITT. We calculate non-durable consumption equal to the total of Columns (2) and (3) in Tables S5a-1 and -2 (i.e. total consumption less durable good expenditures) multiplied by the average household size times 12. The average household sizes used is 8.34 in endline and 8.48 in followup.

2. Cost/benefit: To estimate the breakeven dissipation rate (i.e., the rate of decline of the impact on consumption from one time period to the next), we calculate the net present value of consumption in perpetuity beginning in year 4 with the equation [*(ITT consumption*(1 - dissipation rate))/1.05]/(discount rate + dissipation rate)*. We then solve for the level of dissipation such that the net present value of the costs equals the net present value of the benefits.

8. Discussion

The evaluation shows that the ultra-poor Graduation program in Ghana improves the lives of the very poor along many dimensions. The program's primary goal, to increase consumption, is achieved by the end of the program and maintained one year later. Furthermore, the increase in productive assets and income and revenues (intermediate and downstream outcomes) validates the overall theory of change. This section discusses the validity, limitations, and implications of this study for future research and policy work.

8.1 Internal Validity

From the study design to the implementation to the evaluation strategy, care was taken to ensure high data quality. The random assignment creates the necessary identification assumptions to establish causality.

The table below presents an analysis of survey attrition for both endlines for the GUP program. The follow-up rate was excellent: 99% at endline 1 and 94% at follow-up (Panel A). Panel B presents analysis on the type of people that were more likely to be re-surveyed at each round. Panel C presents a test of whether the treatment affected the type of person who completed the endline surveys (i.e., whether the treatment induced sample composition bias). The p-values on a full set of baseline characteristics interacted with treatment are 0.99 (endline 1) and 0.58 (endline 2), thus supporting the conclusion that the survey attrition did not lead to a markedly different sample frame across treatment and control groups.

Table 18: Attrition analysis

Dependent Variable: Completed Survey, OLS		
	Endline	*Follow-up*
Panel A.		
Treatment Status	0.00071	0.0051
	(0.0052)	(0.0130)
N	2606	2606
R-Squared	0.0340	0.138
Surveyed mean	0.99	0.94
Panel B.	0.00	0.00
Treatment Status	0.00150	0.00560
	(0.0052)	(0.0130)
Consumption per capita, standardized	-0.00120	-0.0046
	(0.00200)	(0.0049)
Food Security Index	-0.0014	-0.0066
	(0.00200)	(0.0048)
Asset index	0.00230	-0.00210
	(0.0020)	(0.0050)
Financial Inclusion Index	0.00	-0.012**
	(0.00200)	(0.0050)
Time spent working, standardized	0.00000	0.0000
	0.00000	0.0000
Incomes and Revenues Index	0.000690	0.0074*
	(0.00180)	(0.0044)
Physical Health Index	0.0010	0.0038
	(0.00190)	(0.0048)
Mental Health Index	-0.00150	-0.00380
	(0.00200)	(0.0049)
Political Involvement Index	0.0028000	0.013***
	(0.00190)	(0.0048)
Women's Empowerment Index	-0.00260	-0.00700
	(0.0020)	(0.0049)
N	2606	2606
R-Squared	0.06	0.15
Surveyed mean	0.99	0.94
Panel C.	0.00	0.00
Treatment Status	-0.00074	0.0078
	(0.005)	(0.0130)
Baseline characteristics	Yes	Yes
Baseline characteristics interacted with Treatment	Yes	Yes
N	2606	2606
R-Squared	0.09	0.16
Surveyed mean	0.99	0.94
p-value from test that Treatment and all other variables interacted with Treatment are jointly 0	0.99	0.58

Dependent Variable: Completed Survey, OLS

	Endline	*Follow-u*
Panel A.		
Treatment Status	0.00071	0.0051
	(0.0052)	(0.0130)
N	2606	2606
R-Squared	0.0340	0.138
Surveyed mean	0.99	0.94
Panel B.	0.00	0.00
Treatment Status	0.00150	0.00560
	(0.0052)	(0.0130)
Consumption per capita, standardized	-0.00120	-0.0046
	(0.00200)	(0.0049)
Food Security Index	-0.0014	-0.0066
	(0.00200)	(0.0048)
Asset index	0.00230	-0.0021(
	(0.0020)	(0.0050)
Financial Inclusion Index	0.00	-0.012**
	(0.00200)	(0.0050)
Time spent working, standardized	0.00000	0.0000
	0.00000	0.0000
Incomes and Revenues Index	0.000690	0.0074*
	(0.00180)	(0.0044)
Physical Health Index	0.0010	0.0038
	(0.00190)	(0.0048)
Mental Health Index	-0.00150	-0.0038(
	(0.00200)	(0.0049)
Political Involvement Index	0.0028000	0.013**
	(0.00190)	(0.0048)
Women's Empowerment Index	-0.00260	-0.0070(
	(0.0020)	(0.0049)
N	2606	2606
R-Squared	0.06	0.15
Surveyed mean	0.99	0.94
Panel C.	0.00	0.00
Treatment Status	-0.00074	0.0078
	(0.005)	(0.0130)
Baseline characteristics	Yes	Yes
Baseline characteristics interacted with Treatment	Yes	Yes
N	2606	2606
R-Squared	0.09	0.16
Surveyed mean	0.99	0.94

p-value from test that Treatment and all other variables interacted with Treatment are jointly 0	0.99	0.58

The study's first-stage clustered randomization at the village level made it so there was no contamination of the control group. All treatment households received the program, and no control group households did. The second-stage randomization within treatment villages assigned households to either receive the program or not. By comparing the households who did not receive the program but lived in treatment villages to households who did not receive the program but lived in control villages, one can see whether or not there were any within-village spillover effects of the GUP program. This analysis is in progress.

Both treatment and control groups were administered the same survey, and the implementation and evaluation team were kept separate. The evaluation team did not know the treatment status of the households at the time of surveying. The fact that many of the treatment effects observed after one year of program implementation persist a year later (and a year after the end of the program) lends confidence to the internal validity of the study. As with any survey, there is the possibility that respondents would tailor their answers based on what they think the surveyor wants to hear. For example, participants might under-report income if they think that by doing so they will be more likely to benefit from a future program. The fact that the evaluation team and implementation teams were kept separate, and that the evaluation team clearly stated that it was not linked to any program were ways that this risk was reduced. There is no evidence to suggest systematic over- or under-reporting.

The qualitative results draw interesting insights that are consistent with the quantitative analysis. The interviews conclude that the households that received the productive asset treated it as an investment – a finding that is supported by the increase in productive asset holdings and increased livestock revenue at both endline and follow-up. The qualitative findings suggest that households did not substitute out of activities such as farming and microenterprise to focus on livestock-rearing, but added livestock-rearing to their portfolio of income-generating activities. The quantitative data supports this finding that households do not see the different activities as substitutes, but rather view them as additional sources of income. The fact that agricultural, business and paid labor income do not see a decline provides evidence for this.

8.2 External Validity

The GUP intervention's sample, which was randomly selected, is reflective of the population that would be considered when thinking about scaling up the GUP program: the rural ultra-poor. Furthermore, the GUP intervention was one of six different interventions carried out across six countries to test the same basic program. These replications, which had very similar components, were carried out by various implementing partners in different cultural contexts. All interventions targeted the

same population in each respective country: the ultra-poor. The promising results that these combined interventions yield are robust and can be used to inform similar programs.

There are, nonetheless, some questions that arise when thinking about scaling up the GUP intervention and thinking about its applicability to other contexts. The first is the target sample: this intervention yields positive results among the rural poor. The GUP communities were primarily ones where residents worked in agriculture and animal husbandry. The intervention may need to be adapted if it were to be implemented in urban settings or in settings with access to more sources of labor. The second consideration is to make sure that the productive asset is one that is contextually appropriate and sustainable. In the Graduation program replication in Honduras, the chickens that were given to beneficiaries were not native to the area, and as a result, many of the chickens died over the course of the intervention.

8.3 Dissemination and future research

The GUP program results have been shared as part of the larger set of Graduation interventions with implementing agencies in Ghana and policymakers. The studies were received well, and have sparked discussion around scale-up and directions for future research, both in Ghana and elsewhere. The results are aligned with the expectations of the implementing partners. The participants at the two local dissemination conferences (in Tamale & Accra) were excited by the results and the events drew substantial media attention. The Ministry of Gender and Social Protection and its collaborating partners (World Bank and UNICEF) are interested in exploring possibilities to include components of the graduation program to the LEAP cash transfer program. See the PIP report for more details on dissemination events. Although there has been no client targeted dissemination, the results are aligned with the aspiration plans that the participants set at the start of the project.

In terms of key lessons to help inform the design of monitoring plans, IPA found that giving PAS short term deliverables combined with frequent field visits by the IPA implementation team helped ensure that activities were going as planned. The combination of the two also helped ensure adherence to research design. Finally, quarterly refresher trainings (focused around research design and planning activities for the next quarter) helped build the implementing partner's capacity and strengthened IPA's relationship with them.

Based on the positive results of the GUP study, along with results from the other graduation evaluations, researchers are now working to learn how best to scale the program: measure the value of each program component to determine whether the program can be delivered at lower cost. The GUP evaluation design allowed testing the effectiveness of just the asset drop and of just the savings component. This report has presented results of the asset drop group, and these preliminary results suggest that the asset drop is not sufficient to see the same level of positive results in consumption, income, and asset holdings. Results from the savings-only group are pending.

GUP was one site out of ten CGAP-Ford Graduation Pilots. With strong impacts across six randomized evaluations in six countries, it is clear the model can work in varied settings. Perhaps more important is to consider how the program can work best for different types of beneficiaries. There was substantial heterogeneity of impacts with the graduation sample.

9. Specific findings for policy and practice

The GUP findings, along with the results from the other CGAP-Ford graduation RCTs clearly demonstrate the graduation model to be effective. The graduation model can now be an important part of an evidence-based toolkit for helping the extreme poor develop livelihoods and improve their standard of living. Some of the sites, including Ghana, show positive but relatively modest benefit-cost ratios while others have very large effects on income, consumption, assets, and food security per dollar invested in the program. Questions remain about how to optimize the program and bring it to scale such that the greatest number of households can benefit. This will require collaboration between governments, evaluators, and donors to carry out a learning agenda as the program is scaled in multiple countries.

Appendix A: Field Notes

Who is Ultra poor: Determining how best to select ultra poor households in a context where typically everyone in the community is very poor, was one of the key challenges that we faced. It was realized early that very few assets are owned by most households and therefore what divides ultra poor household from the rest of the community is very little. As a result, the team focused on finding what was considered the bottom third (ultra poor) of the community and worked closely with PAS to determine what assets or number of assets a household needs in order to be secured. Also, since sampled communities were very small and far apart it was important during selection to ensure that there were enough program participants to make implementation of the program feasible for field agents.

Asset procurement: PAS' experience in procurement helped immensely in acquiring the asset necessary for the program. However, due to the large numbers of animals needed, PAS faced some challenges in procuring all of the goats and poultry needed at the three stations. Some of these challenges are unavoidable (e.g. rain washed out several key market days), but others required creative solutions. The limited numbers of goats available at local markets were overcome by contacting community butchers and farmers. Similarly, vendors would often raise the price when a large number of animals were being purchased by one person, PAS, thus, used multiple buyers including middlemen at each market. These challenges, delayed procurement. However, PAS and IPA's teamwork paved the way for it to be finished within schedule.

Security of asset: In the early part of the first year, some clients had their asset stolen by some community members. These assets were retrieved by the Community GUP committee members set up in each of the 78 GUP communities to primarily, provide security for clients' asset and the program in general. The committee met monthly to discuss issues and concerns that would improve implementation at their level without interfering with the research/implementation design.

Inaccessible Communities: Poor access to communities was one key challenge field staff faced. That led to three FAs involvement in motor accidents and had their arms broken and stayed at home for almost two months. Fortunately, three Monitoring Agents (one per station) who were hired to support the station Team Leaders would stand in for the accident victim to have implementation go on without a halt.

Interference of rains: Other challenges were unavoidable such as rains during rainy season that created or flooded streams and made it extremely difficult for an FA to visit a HH at the expected day and or time. Such visits were sometimes re-scheduled for the weekend or the clients swimming across the stream to meet with the FA for the weekly services so as not the break the chain of the weekly protocol.

Disability of clients: Another notable challenge during the implementation period was the disability of some clients that affected their input in the program. Their HH members who promised at the beginning to fully support them declined leaving the

disabled clients to implement the program alone. When they were identified and talked to, some started supporting but a few still did not put in much.

Interference of husbands: Yet, another challenge also worth mentioning is the interruption of some clients' husbands in the program. For instance, some husbands sold out some asset of the HH and refused to use the money to support the HH up keep. Such men were again identified and re-sensitized. One client husband married a second wife because they harvested their maize and had over 7 bags which to the Ultra poor husband placed them in another class of rich. The community Chief and elders, however, supported the man's idea saying that no man in the village has one wife except an Ultra poor.

Late update of passbooks: We also addressed how best to keep clients passbooks regularly updated by dialoging with bank officials who eventually released point officer per bank to update clients savings passbook. Until the end of implementation, the individual savings at the household level continued to be a chosen option for the GUP Pilot program. Nevertheless, some passbooks balance did not balance the clients bank balance and had to take PAS staff one whole month to correct the abnormally.

Resignation of staff: In the course of implementation four partner staff (2 Team Leaders, 1 Monitoring Agent and 1 Field Agent) and one IPA GUP Pilot staff (Implementation Manager) resigned. These persons were all replaced.

Poor management of finances -both clients and management levels: IPA Team had to conduct several monitoring visits, on the spot remedial trainings and several audits to help correct most challenges.

Inadequate road worthy vehicles/monitoring: In the course of implementation, the partner staff faced major constraints/challenges in the areas of inadequate road worthy vehicles and too much work load of partner staff to the neglect of regular monitoring. That really affected the full involvement of FAs from the beginning who would have loved to have regular monitoring visit by partner staff as motivation. To address this short fall, IPA had two field staff (Implementation Manager and a Monitoring Agent) to beef up the weekly monitoring.

Appendix B: Survey Instruments

Please see the folder titled, “Survey Instruments” in the attachments for copies of the survey instruments used.

Appendix C: Descriptive Statistics

Baseline Summary Stats

	GUP + Control
Total per capita consumption, mean	53.97
	-0.77
Total per capita consumption, median	43.35
Share of households below $1.25/day	0.53
World Bank share in country below $1.25/day line	0.29
Year of World Bank poverty data	2006
Percentage of primary age children enrolled in school, mean	0.63
Percentage of primary age children enrolled in school, median	1
Average level of education received in household, mean	8.28
Average level of education received in household, median	8.58
Total savings amount, mean	6.03
Total savings amount, median	0

Note: All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Baseline Univariate Tabulations: Any GUP + Control

There were days without enough food in the last 12 months	Freq.	Percent
No	350	13.44
Yes	2,254	86.56
Total	*2,604*	*100*
months	Freq.	Percent
No	1,321	51.02
Yes	1,268	48.98
Total	*2,589*	*100*
Household has a business	Freq.	Percent
No	1,669	64.09
Yes	935	35.91
Total	*2,604*	*100*
Household owns a television	Freq.	Percent
No	2,575	98.89
Yes	29	1.11
Total	*2,604*	*100*
Any household member ill in the past 30 days	Freq.	Percent
No	894	34.36
Yes	1,708	65.64
Total	*2602*	*100*

Appendix D: .do Files

The GUP data and code used in the publication of this paper is housed on the Harvard University Dataverse:

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NHIXNT

The Readme file provides an analysis map for replication. The original datasets used in analysis are pooled_hh and pooled_mb. They contain data from the household survey (hh) and the adult survey (mb) at three time periods (baseline, endline 1, endline 2). Ghana's country code is 2.

References

Ahmed, A., Rabbani, M., Sulaiman, M., Das, N., 2009. Series No. 39 The Impact of Asset Transfer on Livelihoods of the Ultra Poor in Bangladesh [WWW Document]. URL http://research.brac.net/new/component/k2/livelihoods (accessed 7.31.15).

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

United Nations Development Programme, 2007. The Ghana Human Development Report.